

The melodic mind:
Neural bases of intonation in speech and music

Impressum

Max Planck Institute for Human Cognitive and Brain Sciences, 2018



Diese Arbeit ist unter folgender Creative Commons-Lizenz lizenziert:
<http://creativecommons.org/licenses/by-nc/3.0>

Druck: Sächsisches Druck- und Verlagshaus Direct World, Dresden

Titelbild: © Sergey Nivens/Shutterstock.com; © kvartur/Shutterstock.com;
© Hannu Viitanen/Colourbox.de; Daniela Sammler adapted by MPI-CBS

ISBN 978-3-941504-81-3

The melodic mind:
Neural bases of intonation in speech and music

Der Fakultät für Lebenswissenschaften
der Universität Leipzig

eingereichte

HABILITATIONSSCHRIFT

zur Erlangung des akademischen Grades
doctor rerum naturalium habilitatus
Dr. rer. nat. habil.

vorgelegt von
Dr. rer. nat. Daniela Sammler
geboren am 9. September 1978 in Meerane, Deutschland

Datum des Verleihungsbeschlusses:
15.01.2018

Gutachter dieser Habilitationsschrift:

Prof. Dr. **Eckart Altenmüller**

Hochschule für Musik, Theater und Medien
Hannover, Deutschland

Prof. Dr. **Isabelle Peretz**

Université de Montréal
Montréal, Kanada

Prof. Dr. **David Poeppel**

Max-Planck-Institut für Empirische Ästhetik
Frankfurt am Main, Deutschland

Datum der Einreichung

24. Februar 2017

Datum der Verteidigung

3. November 2017

Datum der öffentlichen Probevorlesung

5. Dezember 2017

Datum des Verleihungsbeschlusses

15. Januar 2018

Bibliographische Beschreibung

Name: Dr. rer. nat. Daniela Sammler

Titel: The melodic mind: Neural bases of intonation in speech and music

Universität Leipzig, Habilitationsschrift

235 Seiten, 34 Abbildungen, 31 Tabellen, 887 Referenzen

Referat

Melodie und Tonfall spielen eine Schlüsselrolle in unserer Gesellschaft und Kultur: Nicht nur beeinflussen sie Sprachverstehen und musikästhetischen Genuss; sie sind auch in der Gestaltung persönlicher Beziehungen und im zwischenmenschlichen Umgang ausschlaggebend. Die vorliegende Arbeit beleuchtet die wesentlichen *linguistischen* und *sozialen* Funktionen von Melodie und Tonfall in Sprache und Musik und eruiert deren neuronale Grundlagen. Dabei verfolgt die Arbeit ein dreistufiges Forschungsprogramm, das bestehendes Wissen in mehrfacher Hinsicht erweitert: Die Arbeit zeigt in einer Sammlung von acht Veröffentlichungen, (A) dass Melodie in beiden Domänen entlang *multipler dorsaler und ventraler fronto-temporaler Pfade* der (rechten) Hemisphäre verarbeitet wird, (B) dass diese Netzwerke mit (nicht-melodischen) linguistischen und sozio-kognitiven neuronalen Netzwerken *interagieren*, um (C) den konventionalisierten Gebrauch von Melodie und Tonfall *innerhalb und zwischen* (Inter)Akteuren zu steuern. Diese Ergebnisse wurden mit jeweils angemessenen psychoakustisch-behavioralen, elektrophysiologischen (z.B. Elektroenzephalographie), bildgebenden (z.B. funktionelle Magnetresonanztomographie) und nichtinvasiven Stimulationsverfahren (z.B. transkranielle Magnetstimulation) bei gesunden Erwachsenen, Patienten mit neurologischen Störungen und professionellen Musikern gewonnen. Zusammengefasst erweitert die vorliegende Habilitation frühere Arbeiten zu Melodie um Kernkonzepte der *neuronalen Konnektivität* und *sozialen Verbundenheit* und trägt somit der sozio-biologischen Komplexität eines scheinbar einfachen, zwischenmenschlichen Kommunikationskanals Rechnung, der gesellschaftliche Harmonie und intellektuelles Erleben gleichermaßen beeinflusst.

Acknowledgements

First and foremost I wish to thank Angela Friederici and the Max Planck Society for giving me an intellectual home and inspiration over the years, for their support of the Otto Hahn Group and their faith in my scientific creativity and growth. I am also grateful to Séverine Samson, Pascal Belin, and Marie-Hélène Grosbras for hosting and supporting me in Paris and Glasgow, on my way to becoming an independent researcher, and for introducing me into their strong French scientific community. Enchantée! Thank you also to Stefan Koelsch who influenced my scientific leadership more than he might think. Best judges on that matter are certainly Roberta Bianco and Nele Hellbernd—I am proud of you and the achievements we have celebrated together in the OHG. I am truly grateful to Arno Villringer who gave the music research in our group enormous support and swing. My special thanks go to Peter Keller and Giacomo Novembre for hosting me in Sydney and for the most energetic transcontinental collaboration one can wish to have. My gratitude goes also to Erich Schröger for leading the commission of this habilitation and to Ingrid Schmude for her competent advice in many administrative decisions. For most enlightening and enjoyable discussions I wish to thank Bob Turner and Tom Bever. There were many more people on the way who were of hard-to-quantify help. I thank them all. Irene Alonso, Eckart Altenmüller, Amee Baird, Patricia Bestelmeyer, Sylvain Clément, Katrin Cunitz, Vera Dinkelacker, Sophie Dupont, Korinna Eckstein, Sarah Gierhan, Thom Gunter, Philipp Kanske, “SG” Kim, Sonja Kotz, Marianne Latinus, Phil McAleer, Jürgen Meixensberger, Derek Ott, Tamara Rathcke, Yves von Cramon—thank you for your scientific inspiration and collaboration. Alfred Anwander, Eric Bardinet, Eric Bertasi, Heike Böthel, Bernard Bouchard, Frances Crabbe, Fritz Deubel, Kerstin Flake, Anne-Kathrin Franz, Andrea Gast-Sandmann, Maren Grigutsch, Sven Gutekunst, Jöran Lepsien, Stephan Liebig, Burkhard Maess, Ramona Menger, Florian Scharf, and Romain Valabrègue—thank you for your technical, graphical and culinary support. Ein letztes Dankeschön geht an meine Eltern, Ilona und Dieter Sammler, die mich vor Jahren auf den richtigen, guten Weg gebracht haben, und an Olav, der einfach alles möglich macht.

Contents

| | |
|---|-----------|
| Introduction | 3 |
| I Framework and summary | 5 |
| 1 Use of pitch in speech and music | 7 |
| 1.1 Prior clarifications: What is melody in speech and music? | 8 |
| 1.2 Towards a testable framework | 12 |
| 1.3 Summary of the empirical evidence | 14 |
| 1.3.1 Neural networks for melody | 14 |
| 1.3.2 Melody and fact | 18 |
| 1.3.3 Melody in (inter)action | 21 |
| 1.4 General discussion and outlook | 24 |
| 1.4.1 Dual streams for melody in speech and music | 24 |
| 1.4.2 Melody's social communicative value | 29 |
| 1.4.3 Conclusion | 30 |
| II Experiments | 31 |
| 2 Neural networks for melody | 33 |
| 2.1 Study I – Neural networks for speech prosody | 33 |
| 2.2 Study II – Neural networks for speech prosody | 58 |
| 2.3 Study III – Neural networks for musical harmony | 69 |
| 3 Melody and fact | 81 |
| 3.1 Study IV – Speech prosody and syntax | 81 |
| 3.2 Study V – Song melody and lyrics | 105 |

| | | |
|-----|---|-----|
| 3.3 | Study VI – Song melody and lyrics | 117 |
| 4 | Melody in (inter)action | 129 |
| 4.1 | Study VII – Speech prosody in (inter)action | 129 |
| 4.2 | Study VIII – Musical harmony in (inter)action | 147 |
| III | Appendix | 163 |
| | Zusammenfassung | 177 |
| | Bibliography | 179 |
| | List of abbreviations | 197 |
| | List of figures | 197 |
| | Curriculum vitae | 199 |
| | Publications | 205 |

Introduction

It don't mean a thing if it ain't got that swing.

—Duke Ellington—

Our auditory world is populated with melodies—modulations of pitch over time—that influence our lives more diversely than we might think. Just imagine a world without melodies: no perfect fourth of fire engines that signals us to instantly clear the road, no music that inspires us to sing, and no tone in a speaker's voice that makes us feel welcomed (or not). The present thesis investigates various ways in which we use melody and pitch in our everyday lives and their neural underpinnings, with particular focus on speech and music.

One of the probably greatest achievements of man is the use of speech and language (Darwin, 1871). Linguists have filled books about it, debated the preeminent role of syntax (Chomsky, 1957) and the architecture of the mental lexicon (Collins & Quillian, 1969; Smith, Shoben, & Rips, 1974), and more recently started to dissect the language faculty's neural implementation (Friederici, 2011; Hickok & Poeppel, 2007). But how important is language and the meaning of words in declared “post-factual” times? The present thesis will follow Wittgenstein's intuition and will argue that there is more to language than words—that a main ingredient to conversational success is the *why* and *how* things are said. As will be shown, the vocal tone and melodic aspects of speech—i.e., speech prosody—can be decisive for mutual understanding: Prosody not only shapes the verbal, factual message of an utterance (Cutler, Dahan, & Van Donselaar, 1997); it also plays a major role in conveying “unspoken” messages on its own. While intonational phonologists (Bolinger, 1986) and developmental psychologists (Malloch & Trevarthen, 2009a) have been aware of the communicative power of prosody for quite some time, comparably little is known about the neurocognitive bases of the melodic communicative channel. The **first goal** of the thesis is, therefore, to describe the neurocognitive architecture that en-

ables listeners to make sense of others' melodic-prosodic signals and to show how they are linked to verbal language comprehension.

One recurring theme in the thesis will be the relationship between melodic and verbal aspects in speech and vocal music, including their neural integration and potential segregation. Some insights on the “melodic mind” will be clearly, and sometimes boldly, gained through extrapolation of verbal-linguistic theory to the musical-prosodic domain (a dual-stream model, the notion of syntax) and vice versa. This cross-domain translation of concepts is not counter-intuitive given the numerous parallels between language and music in terms of design features (Fitch, 2006) and neuro-cognitive organization (Patel, 2008). Proclaimed common evolutionary origins of language and music—both potentially rooted in a song-like protolanguage (Darwin, 1871; Fitch, 2010; Mithen, 2005)—not only further underline the relevance of melody and pitch for mankind but also license the integrated investigation of verbal factual and non-verbal melodic processing. The **second goal** of the thesis is, therefore, to showcase persistent neural analogies between verbal and melodic information processing and to demonstrate their entwined organization in human cognition.

Finally, the use of pitch—both in speech and music—follows culturally determined conventions (if not universals) that are deeply carved into our minds. In music, series of tones have to follow specific compositional rules to sound harmonious to listeners of the respective culture, even to those without musical training (Zatorre & Krumhansl, 2002). In speech, distinct prosodic contours convey different intentions, such as rising pitch at the end of a sentence typically signals the speaker's search for an answer (Schneider, Lintfert, Dogil, & Möbius, 2006). Knowledge about these conventions provides common ground for both receivers and senders and fosters seamless interactions (Pickering & Garrod, 2013). Crucially, it not only guides listeners' perception, but equally determines melodic and prosodic choices in the sender to make herself properly understood, within the realms of aesthetics and social adequacy. While melody and prosody *perception* have been addressed by cognitive neuroscientists in the past (Koelsch, 2011a; Schirmer & Kotz, 2006), available data on neural correlates of their *production* remain sparse (Pichon & Kell, 2013; Zatorre, Chen, & Penhune, 2007). The **third goal** of the thesis is, therefore, to lend neuro-behavioural evidence for the role of pitch-based conventions in driving performers' actions leading them to produce melodically well-formed signals that facilitate social interaction and communication.

To reach these goals, the experiments summed up here interrogate the neural bases of melody and pitch in speech and music, in perception and production, with various psychoacoustic, electrophysiological, and neuroimaging techniques in healthy adults, patients with neurological disorders, and professional musicians. Chapter 2 specifies neural networks for different uses of pitch: It introduces a right-hemispheric dual-stream account of prosody perception in speech (**Studies I & II**) and draws parallels to networks of harmony perception and production in music (**Study III**). Chapter 3 focuses on the interaction of melodic and verbal information: It presents cross-talk between right-hemispheric prosodic and left-hemispheric linguistic processes in speech (**Study IV**), and interrogates the binding of melody and lyrics in the perception of songs (**Studies V & VI**). Chapter 4 finally turns to the use of pitch in (inter)action: It shows that prosodic conventions in speech allow speakers to convey their intentions without words to drive listeners' conversational reactions (**Study VII**), and that harmonic conventions in music—sometimes referred to as “syntax”—determine the way how musicians produce music, even without sound, suggesting an abstraction of pitch-based processes to the motor domain (**Study VIII**). Overall, the combined data make a strong case for the centrality of melody and pitch in driving human behaviour, be it receptive or active, similarly in speech and in music.

Part I

Framework and summary

Chapter 1

Use of pitch in speech and music

Modulation of pitch in speech and music is a springboard for humans to grow into socio-emotionally and linguistically competent individuals. Infants are born with a remarkable sensitivity for pitch and melody (Perani et al., 2011; Stefanics et al., 2009): From early on, they pick up on the emotional tone in lullabies (Trehub & Trainor, 1998) and maternal speech (Grossmann, Oberecker, Koch, & Friederici, 2010) that regulate their own emotions. 4-day old newborns can already recognize (Mehler et al., 1988) and imitate pitch patterns of their native language (Mampe, Friederici, Christophe, & Wermke, 2009), which they start to use intentionally at the end of their first year of life to regulate cooperative social interactions with their parents while their verbal abilities are still limited (Esteve-Gibert & Prieto, 2013; Prieto, Estrella, Thorson, & del Mar Vanrell, 2012). Ultimately, infants benefit in various ways from intonation in infant-directed speech for language acquisition (Graf Estes & Hurley, 2013; Männel & Friederici, 2013; Thiessen, Hill, & Saffran, 2005). Hence, it seems no coincidence that parents around the world sing for their children (Trehub & Trainor, 1998) and engage in “musical” conversations with their babies (Malloch & Trevarthen, 2009a). They use expanded pitch contours in music and speech to match the little listeners’ perceptual capabilities and preferences (Cruttenden, 1994; Fernald & Kuhl, 1987), fostering social bonding and the language faculty as main pillars of human companionship.

Early child development is perhaps the most obvious example for the centrality of melody and pitch for human society and culture—the present thesis will make a strong case on their persistent role in adulthood, as continuous support for language (Chapter 3) and

social interaction (Chapter 4).¹ The following sections shall sketch a research agenda on the neurocognitive foundations of using pitch in speech and music that will then be tackled in the second part of the thesis. Beforehand, three fundamental questions will be addressed: what are melodies in speech and music, what do they have in common, and what are they good for?

1.1 Prior clarifications: What is melody in speech and music?

Melodies are modulations of acoustic frequency over time that are perceived as variations in pitch and are both musically and linguistically relevant. The importance of melody for music needs little explanation—it is hard to imagine a song with only a single pitch (perhaps apart from rap music). The matter is more complicated in speech. Other than musical melody that can be an aesthetic object in itself, melody in speech—termed intonation²—constitutes one of the many vocal modulations that are subsumed under the term prosody (Warren, 1999) and typically *modify* conversational and information structure of an utterance rather than standing alone (Cole, 2015; Cutler et al., 1997). Even though this view of speech intonation as mere ‘means to an end’ can and will be challenged in the present thesis (Study VII; see also Bolinger, 1986), it is true that (linguistic) prosody conveys semantic relationships (Cole, 2015; Cutler et al., 1997; Wagner & Watson, 2010), disambiguates syntactic constituent structure (Frazier, Carlson, & Clifton Jr., 2006), and marks declarative vs. interrogative sentence mode (Schneider et al., 2006; Srinivasan & Massaro, 2003), but rarely gets “caught in one’s head” as is frequently the case with musical melody.

One reason for this seeming autonomy of musical melody—despite its frequent coupling with verbal material in songs—may be its construction along a fixed interval grid that has no counterpart in speech intonation. Music uses discrete pitches in specific frequency relations—take a C and an F that together form a major fourth—as opposed to the continuous rise and fall of pitches devoid of predefined bounds in spoken language (for an

¹Low-level decoding of pitch from periodicities in the auditory signal is at the basis of the processes in question, but not topic of the present thesis. Reviews on the neural coding of pitch along the auditory pathway can be found in Plack, Oxenham, and Fay (2005).

²The present thesis will use the term prosody synonymously with speech melody and intonation to refer to pitch fluctuations in the speech stream. Rhythmic aspects of prosody and their roles in both speech and music are reviewed in Liberman and Prince (1977) and Patel (2008).

overview, see Patel, 2008). Notably, these fixed frequency relations apply both vertically and horizontally, i.e., to consecutive as well as simultaneously sounding pitches (chords), yet another point that finds no correspondence in speech and is denoted as harmony.³ Most importantly, pitches in music, but not speech, differ in their *function* that depends on the tonal key they belong to and determines their melodic and harmonic relationship with preceding, concurrent and forthcoming elements of the musical piece. One tone or chord (the tonic), for example, acts as central reference point of a tonal key and is characterized by highest stability in a piece, while stability of the remaining tones or chords of the scale varies hierarchically (Krumhansl, 1979). These (and other) relationships are sometimes referred to as “musical syntax” (Swain, 1995) that governs the rule-based arrangement of discrete pitches into musically meaningful sequences (Lerdahl & Jackendoff, 1983) and allows listeners (and performers) to build up expectations upon forthcoming elements (Studies III and VIII). Confirmation and violation of these expectations when moving through tonal pitch space is an archetypical musical process and the source of musical pleasure (Huron, 2006). Altogether, musical use of pitch compared to speech intonation engenders a much richer set of vertical and horizontal pitch relations that hold interest and are memorable on their own, even if fused with words (see Studies V and VI).

Nevertheless, to put it in Dwight Bolinger’s words “since intonation is synonymous with speech melody, and melody is a term borrowed from music, it is natural to wonder what connection there may be between music and intonation” (Bolinger, 1986:28; cf. Patel, 2008:238). There are (at least) three aspects that melodies in music and speech have in common and that are relevant for the present thesis: the processing of pitch contour (as opposed to intervals), the grouping of elements into phrases, and the power to convey interpersonal meaning without pointing to items in the objective world.⁴

Pitch contour refers to the global direction of pitch changes irrespective of their exact size. Other than fixed intervals, contour is processed from early infancy on (Fernald & Kuhl, 1987; Stefanics et al., 2009; Trainor & Trehub, 1992), is more readily remembered and reproduced (Dowling, 1978, 1982) and can be preserved despite deficits in musical interval processing after brain damage (Liégeois-Chauvel, Peretz, Babai, Laguitton, & Chauvel,

³The present thesis will use the terms melody and harmony even-handedly to refer to the arrangement of pitch elements into musical sequences, irrespective of whether elements are single tones or simultaneously sounding pitches (i.e., chords). Note that both types of sequences underlie similar compositional principles denoted as musical “syntax”.

⁴Note that more than half of the world’s languages are tone languages, such as Mandarin Chinese, Thai, or Yoruba, in which pitch contours determine the lexical meaning of speech sounds (Haspelmath, Dryer, Gil, & Comrie, 2005). Lexical use of pitch and its neural processing are not topic of the present thesis, but are reviewed in Gandour and Krishnan (2015).

1998; Peretz, 1990; Schuppert, Münte, Wieringa, & Altenmüller, 2000) suggestive of dissociated neural substrates (Stewart, Overath, Warren, Foxton, & Griffiths, 2008). Hence, compared to fixed intervals, global contour perception seems to be the more basic process that may be common to both speech and music. Several findings have fuelled this idea: spoken speech, when repeated, can be perceived as sung (Deutsch, Henthorn, & Lapidis, 2011; S. Falk, Rathcke, & Dalla Bella, 2014) and contours of infants' cries can be transcribed as musical notation (cf. Wermke & Mende, 2009) suggesting perceptual continuity between domains (see also Merrill, Bangert, Sammler, & Friederici, 2016); musical training with emphasis on pitch improves prosodic contour perception suggesting mutual benefit through shared cognitive resources (M. Besson, Schön, Moreno, Santos, & Magne, 2007; Magne, Schön, & Besson, 2006; Moreno & Besson, 2006; Schön, Magne, & Besson, 2004); finally, fronto-temporal brain damage (Patel, Peretz, Tramo, & Labreque, 1998) or congenital conditions of tone deafness (also termed amusia; Hutchins, Gosselin, & Peretz, 2010; Liu, Patel, Fourcin, & Stewart, 2010; Patel, Foxton, & Griffiths, 2005) can lead to contour processing deficits in both music and speech arguing for dysfunction of common neural bases (but see Peretz & Hyde, 2003). Together these findings lend indeed ample reason to assume that the processing of coarse melodic ups and downs is a default when perceiving tone sequences that may be used for similar purposes in speech and music.

One of these purposes is the **grouping** of elements into phrases. Both speech and music typically package information into smaller, syntactic units; in both domains, salient pitch movements align with this grouping structure to facilitate proper decoding of the sequential input. In speech, this becomes particularly clear when regarding ambiguous sentences like "The queen said the king is singing.". Noticeable pitch events—termed phrase or boundary tones—together with duration and intensity cues help to determine who is singing, the queen or the king (de Pijper & Sanderman, 1994; Ladd, 1988; Steinhauer, Alter, & Friederici, 1999; Streeter, 1978). Just in the same way, boundary tones guide how music—an intrinsically ambiguous stimulus—is perceived (Juszyk & Krumhansl, 1993; Knösche et al., 2005; Lerdahl & Jackendoff, 1983); they mark beginnings and endings of phrases in a musical piece, which helps to determine the functions of tones or chords and their harmonic relationships (just as they disclosed the thematic roles in the royal example above). More generally, the melodic grouping of syntactic units proves beneficial within developmental realm in both domains: pitch movements lend salient perceptual anchors that contour-competent infants use to segment the sea of sound they

are exposed to, as a first step into decoding the structure of their language or musical tonal system (Brandt, Gebrian, & Slevc, 2012).

Finally, beyond grouping, melody and pitch are key to **interpersonal meaning**—they are concerned with sharing feelings and regulating social behaviour. It is undisputed that both speech intonation (Banse & Scherer, 1996; Bänziger & Scherer, 2005; Frick, 1985) and music (Gabrielsson & Juslin, 2003; Juslin, 2013) display basic emotions such as joy, tenderness, or sadness that share a common acoustic code across domains (Juslin & Laukka, 2003) and are recognized cross-culturally (music: Fritz et al., 2009; Sievers, Polansky, Casey, & Wheatley, 2013) (speech: Bryant & Barrett, 2008; Laukka et al., 2013; Pell, Paulmann, Dara, Allasseri, & Kotz, 2009). Likewise, caregivers around the world use specific intonation patterns in speech when soothing, warning or playfully interacting with their babies (Fernald, 1989; M. Papoušek, Bornstein, Nuzzo, Papoušek, & Symmes, 1990), patterns that are mirrored in lullabies and play songs for children (S. Falk, 2011) and are understood universally (Bryant & Barrett, 2007; Trehub, Unyk, & Trainor, 1993). Altogether, although neither musical melody nor speech intonation refers to specific concepts—such as words do in language—they are meaningful, in more general, socially relevant ways. It is this communicative capacity that is thought to give speech intonation (D. Falk, 2004) and music (Dissanayake, 1992) survival value, as they can forge interpersonal bonds and guide social interaction, in infancy and beyond (Study VII).

In sum, despite differences in how pitches are arranged in music and speech—along a rule-based grid of fixed pitches or in continuous flux—the present section has highlighted functional commonalities in the use of pitch and melodic contour: First, pitch movements, together with duration and intensity cues, segment the musical and speech stream, helping listeners to navigate through the structure of a musical piece or discourse. Second, pitch information is used to communicate a performers' state of mind, guiding social interaction. On a more abstract level, it seems that pitch can either *assist* higher-order processes in speech and music perception, or acts as the *main character* in social co-ordination that can but does not have to be fused with verbal information to be effective.

These reflections identify three challenges for a neural approach to the use of pitch in speech and music: It not only has to (A) specify the neural implementation of melody and pitch perception in its own rights; it also has to (B) account for its interplay with non-melodic, verbal processes, and to (C) lend explanations for its social communicative power in musical and verbal (inter)actions. The present thesis will address these challenges in a three-layered research program as depicted in Figure 1.1.

1.2 Towards a testable framework

Starting point 1 of this thesis is the well-accepted and intuitive notion that neither music (Koelsch, 2011a; Peretz & Coltheart, 2003) nor speech (Friederici, 2011; Poeppel, 2014) are *monolithic* capacities: Like many higher-order cognitive functions (Mesulam, 1990), their perception and production entails **multiple subroutines** (e.g., processing of rhythm and harmony in music, or syntax and semantics in speech) that are organized in parallel, but **inter-dependent large-scale neural networks**. Anatomically, such networks comprise functionally specialized brain areas (network nodes) that dynamically interact via white matter fibre bundles (network connections) to pass information through consecutive processing stages of **gradually increasing complexity** (Kiebel, Daunizeau, & Friston, 2008; Koechlin & Jubault, 2006; Obleser & Eisner, 2009).

Modern views on the neurobiological foundations of **language** have assimilated the neural network idea to considerable extent: Over the past decade it has become largely undisputed that the cortical language system is organized along anatomically and functionally distinct dorsal and ventral processing streams in the left hemisphere (Bornkessel-Schlesewsky, Schlesewsky, Small, & Rauschecker, 2015; Friederici, 2011; Gierhan, 2013; Hickok & Poeppel, 2007; Rauschecker & Scott, 2009; Ueno, Saito, Rogers, & Lambon Ralph, 2011). The **dorsal stream** connects temporoparietal with frontal/premotor areas via the arcuate/superior longitudinal fascicles (AF/SLF) and is involved in the comprehension of syntax (Friederici, 2012; S. M. Wilson et al., 2011), as well as in sound-to-motor mapping required for normal speech production (Hickok, 2012; Saur et al., 2008), and potentially supporting speech perception via invocation of a motor code (Skipper, Devlin, & Lametti, 2017). The **ventral stream** connects temporal with prefrontal regions via the extreme capsule (EmC) and uncinate fascicle (UF; amongst others) and is proposed to map sound to meaning required for normal language comprehension (Galantucci et al., 2011; Kümmerer et al., 2013; Scott, Blank, Rosen, & Wise, 2000).

Chapter 2 (“Neural networks for melody”) of the present thesis will adopt this dual stream view and will scrutinize whether or not the processing of speech intonation and musical harmony—two reportedly right-hemispheric faculties—involves analogous fronto-temporal processing streams. Chapter 3 (“Melody and fact”) will then look into cross-stream interactions between melodic and verbal-factual information in speech and song.

Starting point 2 of this thesis is the fact that neither music nor speech are *solitary* phenomena: They are typically used in groups—i.e., **social networks**—in which musicians

(D'Ausilio, Novembre, Fadiga, & Keller, 2015) or dialogue partners (Pickering & Garrod, 2013) dynamically interact to mutually drive and coordinate each other's behaviours.

Social neuroscience has reached consensus that successful interactive behaviour resides on the capacity to recognize and infer others' **action goals** and **social intentions** by virtue of two complementary mechanisms: motor simulation and mentalizing (for review, see Van Overwalle & Baetens, 2009). During observable, bodily interactions such as clinking glasses or shaking hands, individuals represent the interlocutor's actions through **simulation**, based on their own motor repertoire (M. Wilson & Knoblich, 2005). Online forward models of the observed action (Wolpert, 1997) allow prediction of the other's immediate action goal and, hence, the timely preparation of one's own (re)action (Pickering & Garrod, 2013; Sebanz & Knoblich, 2009). During more complex social scenarios with abstract motives not immediately derivable from bodily motion, individuals construct the interlocutor's state of mind through **mentalizing**, based on a theory of mind (Frith & Frith, 2006). This high-level social inference allows tracking others' long-term intentions and dispositions to inform one's own interactive behaviour.

Chapter 4 ("Melody in (inter)action") will apply these concepts to the use of pitch in speech and music. Specifically, it will scrutinize the communication of social intentions through intonation in adult speech, probe simulation mechanisms during music performance, and interrogate both their reliance on culturally grounded conventions.

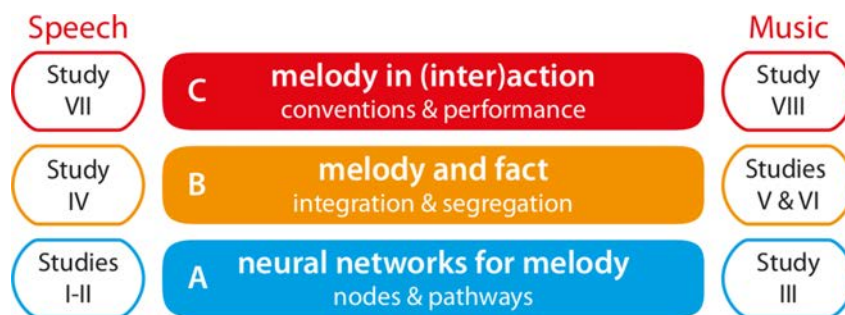


Figure 1.1: Three-layered research agenda towards a comprehensive description of the neural bases of processing pitch in speech and music. Note that the term melody is used here to broadly refer to *pitch patterns*, irrespective of whether they are composed of single (fixed or gliding) or simultaneously sounding pitches (chords).

In sum, the present thesis adopts a three-layered research agenda towards a comprehensive description of the neural bases of using pitch in speech and music (Figure 1.1). It first identifies the large-scale networks of the "melody subroutine" in both domains, including the specification of network nodes and pathways (blue panel in Figure 1.1; Chapter 2).

It then interrogates the interaction between large-scale networks for melodic and verbal information (orange panel in Figure 1.1; Chapter 3), before adopting a social point of view seeking to sketch how pitch information can guide interpersonal interaction and performance (red panel in Figure 1.1; Chapter 4). These matters will be addressed by gathering behavioural evidence of how healthy individuals, patient populations, and professional musicians process pitch in speech and music, and by investigating the underlying neural substrates using structural, functional and diffusion-weighted magnetic resonance imaging (MRI), electroencephalography (EEG), and transcranial magnetic stimulation (TMS).

1.3 Summary of the empirical evidence

Following the research agenda outlined above, the empirical part of this thesis has been sectioned into three chapters presenting eight publications. The following paragraphs will summarize the aims, approaches, and results of these studies.

1.3.1 Neural networks for melody

Chapter 2 (“Neural networks for melody”) showcases three studies that take very different approaches to reach a common goal: the identification of network nodes and—most importantly—network connections that are relevant for the use of pitch in speech and music.

Studies I & II focus on prosodic contour perception in **speech**. They were inspired by the established notion that speech prosody perception is a hierarchical multi-step process with consecutive stages of auditory abstraction along the (right) superior temporal lobe followed by linguistic/socio-emotional evaluation and integration of the prosodic signal in (bilateral) frontal areas (for reviews, see Brück, Kreifelts, & Wildgruber, 2011; Friederici & Alter, 2004; Kotz & Paulmann, 2011; Schirmer & Kotz, 2006; Wildgruber, Ackermann, Kreifelts, & Ethofer, 2006; Wildgruber, Ethofer, Grandjean, & Kreifelts, 2009). What remained, however, a matter of debate so far is how prosodic information is transferred between these sub-processes. Functional connectivity studies left no doubt about the existence of fronto-temporal processing streams (Ethofer et al., 2006; Frühholz & Grandjean, 2012; Leitman et al., 2010), but opinions diverged on whether information trav-

els along dorsal (Glasser & Rilling, 2008; Wildgruber et al., 2006) or ventral pathways (Ethofer et al., 2012; Schirmer & Kotz, 2006), and with which hemispheric lateralization (Kotz, Meyer, & Paulmann, 2006; Witteman, van Ijzendoorn, van de Velde, van Heuven, & Schiller, 2011).

In **Study Ia** we addressed this question with multimodal neuroimaging and showed that prosodic contour perception takes dual routes along both dorsal *and* ventral pathways in the right hemisphere (Sammler, Grosbras, Anwander, Bestelmeyer, & Belin, 2015). We used a novel audio-morphing paradigm (Kawahara, 2006) which gradually manipulated mono-syllabic words along two orthogonal dimensions: prosodic contour ranging from statement (falling) to question (rising contour) and voice-onset time of the word-initial phoneme ranging from /b/ to /p/. fMRI activation clusters that were stronger during judgment of prosody (did the speaker *name* the object or *ask for* it?) than phoneme (did the speaker say *bear* or *pear*?) were taken as network nodes relevant for prosodic contour perception and were used as seeds in multi-fibre probabilistic tractography to identify network connections. These connections comprised a ventral stream along the right middle longitudinal fascicle (MdLF) connecting posterior (pSTS) and anterior superior temporal sulcus (aSTS), and two parallel dorsal streams along the right arcuate/superior longitudinal fascicles (AF/SLF) connecting pSTS with dorsal posterior portions of inferior frontal gyrus (IFG) and premotor cortex (PMC), respectively. Right IFG and PMC were interconnected via short U-fibres; no connections were found between right IFG and aSTS. Overall, this dual stream architecture is reminiscent of left dorsal and ventral streams known as the neural core of the language faculty, yet with relative rightward asymmetry (Bornkessel-Schlesewsky et al., 2015; Friederici, 2011; Hickok & Poeppel, 2007; Rauschecker & Scott, 2009; Saur et al., 2008; Ueno et al., 2011).

Particularly notable was the involvement of the right *dorsal* stream whose relevance for speech and language has been frequently denied (Hickok, 2012; Hickok & Poeppel, 2007). Its involvement was therefore substantiated in Studies Ib & II, both lending evidence for its functional necessity in prosodic contour perception. In Study Ib (Sammler et al., 2015), inhibitory TMS of right PMC as a key node of the dorsal stream decreased participants' performance in prosody categorization, compared to phoneme categorization, sham stimulation and stimulation of left PMC. Importantly, this result integrates into a research tradition that considers PMC essential for facilitating perception of ambiguous speech sounds via simulation of articulatory gestures (for review, see Skipper et al., 2017). In line with this assumption, our PMC activation (Study Ia) and TMS target region (Study

lb) was located at the somatotopic level of the dorsal larynx representation that controls (simulation of) vocal pitch (Brown, Ngan, & Liotti, 2008; Houde & Chang, 2015).

Study II (Sammler et al., 2018) extended these findings by showing causal contribution of the right dorsal stream to sentence-level prosody perception. It reports the case of a patient in whom tissue properties of the right dorsal stream were transiently compromised—as measured with diffusion-weighted MRI—due to a vasogenic peritumoural oedema. This special condition allowed assessment of prosodic deficits and their potential post-surgical recovery in the same individual. Stimuli were auditorily presented sentences with a manipulation of prosodic pitch contour that signalled closure (falling pitch contour) at sentence-final (congruent) or non-final positions (incongruent; Eckstein & Friederici, 2006). The patient's performance in detecting prosodic incongruities significantly improved between pre- and post-surgical test sessions, more than in ten matched healthy controls and more than performance in a non-prosodic sentence comprehension task. Importantly, behavioural recovery was accompanied by an increase in average fractional anisotropy in right AF/SLF, corroborating the assumption of *causal* right dorsal stream involvement in processing vocal prosodic information.

Together, these data draw a dual-stream picture of prosodic contour perception that shows analogies to the established left-hemispheric multi-stream architecture of language, but with relative rightward asymmetry. For the time being, the computational roles of the right ventral and dorsal stream were proposed to parallel those in the left hemisphere (Bornkessel-Schlesewsky et al., 2015), including the *time-sensitive* tracking and evaluation of prosodic contours in audio and/or motor format along dorsal pSTS-IFG and pSTS-PMC-IFG connections (Rauschecker, 2011), and the gradual formation of *time-invariant* “prosodic Gestalts” along the ventral pathway with links to communicative meaning (Petkov, Logothetis, & Obleser, 2009; Schirmer & Kotz, 2006; Scott et al., 2000). The potential parallel involvement yet flexible weighting of such complementary mechanisms may be particularly suited to meet the demands of multiple (para)linguistic functions of prosody at word and sentence level and in various listening situations.

Whether or not a similar dual stream organization applies to the use of pitch in **music** was addressed in **Study III** (Bianco et al., 2016). Inspired by the proposed division of labour of ventral comprehension- and dorsal action-related pathways (Hickok & Poeppel, 2007; Rauschecker & Scott, 2009; Saur et al., 2008), we combined functional connectivity at rest with task-based fMRI while professional pianists listened to (without playing) or played (without sound) harmonically regular or irregular chord sequences. Results

showed dissociable ventral and dorsal streams in the right hemisphere when contrasting irregular > regular musical events: temporal auditory areas identified in the listening task interacted with sub-regions of IFG (BA44/45) that are anatomical endpoints of EmC and UF, as well as AF/SLF. Parietal motor areas identified in the production task were linked with dorsal sub-regions of IFG (BA44/6) that are anatomical endpoints of AF/SLF only. These findings lend empirical support for previously hypothesized right fronto-temporal connectivity as prerequisite for normal processing of melodic and harmonic structure in music (Hyde, Zatorre, & Peretz, 2011; Tillmann, L  v  que, Fornoni, Albouy, & Caclin, 2016). Notably, the present data highlight the potential role of right dorsal and *ventral* connections in the perception of musical structure, while previous studies mainly assumed relevance of the dorsal route (Loui, Alsop, & Schlaug, 2009; Loui, Li, & Schlaug, 2011) or focused on the left hemisphere only (Musso et al., 2015). Dorsal connections are known to play a role in music production, supporting audio-motor mapping (Chen, Rae, & Watkins, 2012; Engel et al., 2014) and vocal control in singing (Halwani, Loui, R  ber, & Schlaug, 2011). The novel finding of the present study is dorsal parieto-frontal connectivity during music production *without* sound that does not involve temporal auditory areas, arguing for an abstraction of pitch information to the motor system. Functionally we proposed that both streams are likely to provide the infrastructure that allows frontal computational areas to keep track of abstract pitch relations in music via dynamic exchange with modality-specific auditory/motor systems of knowledge, that are likely to act in concert during real music performance (i.e., playing with sound).

Overall, the data of Studies I-III highlight the idea that multiple processing streams are a ubiquitous feature of neuro-cognitive architectures that applies to language, music, prosody—and beyond as corroborated by dual-stream accounts of visuo-spatial processing (Kravitz, Saleem, Baker, & Mishkin, 2011) or action understanding (Kilner, 2011). Importantly, these studies break the myth that the right dorsal stream has no say in speech perception, and put the right ventral stream in the limelight of research on music perception, both leading towards a dynamic network view on the use of pitch in speech and music with relative right-dominance. Future studies will have to address mechanisms of cross-stream interaction—within and between domains (Bornkessel-Schlesewsky et al., 2015; Cloutman, 2013). This question will be addressed in Chapter 3.

1.3.2 Melody and fact

Chapter 3 (“Melody and fact”) presents three studies that move beyond the neural bases of pitch processing as such. Instead, they use multiple approaches to interrogate the *integration* of melodic/prosodic information with the concomitant verbal channel during perception of speech (Study IV) and song (Studies V & VI). They, hence, address the pressing question of cross-stream interaction as a logical next step towards a comprehensive neural model of pitch processing in speech and music.

Study IV (Sammler, Kotz, Eckstein, Ott, & Friederici, 2010) probed the neural interaction between intonation and syntax in **speech** via the corpus callosum (CC). As mentioned above, intonational phrasing in speech aligns with syntactic grouping structure, facilitating amongst others the proper decoding of syntactically ambiguous sentences such as “The queen says the king is singing.” (for reviews, see Buxó-Lugo & Watson, 2016; Cole, 2015; Cutler et al., 1997). The neural bases of these prosody-syntax interactions are, however, far from clear, especially when considering that syntax-related processes are thought to involve fronto-temporal processing streams in the *left* hemisphere (Friederici, 2012; Griffiths, Marslen-Wilson, Stamatakis, & Tyler, 2013; Hagoort, 2009; S. M. Wilson et al., 2011), whereas analogous *right*-hemispheric streams may have primacy for the processing of prosodic contour (Friederici & Alter, 2004; Sammler et al., 2015). If there is indeed such a hemispheric division of labour, the lateralized speech streams need to dynamically interact to yield fluent comprehension (Steinmann & Mulert, 2012). In Study IV, we employed an EEG lesion approach and showed cross-stream interactions between the left and right temporal (but not frontal) nodes of the syntax and prosody streams connected via the posterior CC (Hofer & Frahm, 2006; Huang et al., 2005). Patients with lesions in the posterior CC, i.e., disconnection of temporal areas, did not show an early (~200ms) anterior negativity that is typically evoked by misalignments of prosodic and syntactic boundaries (Eckstein & Friederici, 2006). This negativity was robustly found in matched healthy controls as well as patients with lesions in the anterior CC, i.e., disconnection of frontal areas. Given that hearing thresholds and simple syntax processing (i.e., detection of word-category violations) were intact in all participants, including those with posterior CC lesions, we concluded that their deficit must reflect the missing fusion of prosodic and syntactic information due to disrupted transcallosal exchange, making their brains unaware of the prosody-syntax mismatch (for similar results, see Friederici, von Cramon, & Kotz, 2007). Altogether, the data argue for a *rapid* integration of local syntactic and pitch-related features at *early* processing stages in the temporal lobe via the

posterior CC, plausibly meeting the requirements for the seamless coordination of complementary linguistic information during auditory speech perception.

While there is no doubt about the natural integration of melodic and verbal information in speech, the picture is different in music: The relative autonomy of musical melody mentioned earlier led to the on-going debate whether or not melody is processed as a separate component in vocal music (M. Besson, Faïta, Peretz, Bonnel, & Requin, 1998; Hébert & Peretz, 2001; Hébert, Racette, Gagnon, & Peretz, 2003; Peretz, 1996; Racette, Bard, & Peretz, 2006; Samson & Zatorre, 1991) or is tightly bound to the verbal contents, similar as in speech (Bigand, Tillmann, Poulin, D'Adamo, & Madurell, 2001; Crowder, Serafine, & Repp, 1990; Gordon, Schön, Magne, Astésano, & Besson, 2010; Lidji, Jolicœur, Moreau, Kolinsky, & Peretz, 2009; Poulin-Charronnat, Bigand, Madurell, & Peereman, 2005; Samson & Zatorre, 1991; Serafine, 1984; Serafine, Davidson, Crowder, & Repp, 1986).

In **Studies V & VI** (Alonso et al., 2014; Sammler, Baird, et al., 2010), we sought to reconcile these conflicting findings by means of fMRI and showed that both positions are mutually not exclusive, in two respects: The degree of integration or segregation of melody and lyrics depends (i) on the cognitive stage of song processing (Study V; Sammler, Baird, et al., 2010), and (ii) on the neural integrity of the participants tested (Study VI; Alonso et al., 2014). Both studies capitalized on the phenomenon of repetition-suppression (Grill-Spector, Henson, & Martin, 2006; Krekelberg, Boynton, & van Wezel, 2006) and used a novel variant of these paradigms to induce fMR adaptation effects for lyrics and/or melodies during passive listening to series of unfamiliar songs. In both studies, healthy participants presented over-additive adaptation effects for lyrics and melodies in (bilateral) mid-STS and PMC, taken as index for integrated processing of the two components. The degree of lyrics-melody integration, however, continuously decayed along the ventral stream towards anterior STS, where the lack of interaction and stronger adaptation to lyrics than melodies indicated an independent processing of lyrics. One way of interpreting these data is that an integrated percept of song is formed at early auditory, prelexical processing stages in mid-STS that can then be transmitted both along an anterior pathway to anterior STS for an independent analysis of the linguistic content and structure, and along a posterior pathway to PMC for an integrated audio-to-motor conversion of the song stimuli. In other words, lyrics and melodies may split up in the ventral stream for semantics and comprehension (Friederici, 2011; Hickok & Poeppel, 2007; Saur et al., 2008; Scott et al., 2000; Spitsyna, Warren, Scott, Turkheimer, & Wise, 2006; Vandenberghe, Nobre, & Price, 2002) but stay integrated in the dorsal stream for emergence of a unified motor code for subvocal singing (Zarate, 2013). On a more abstract level, these

results demonstrate that lyrics and melody are processed at varying degrees of integration (and segregation) depending on the cognitive process along the ventral and dorsal stream.

Building on these findings in healthy listeners, **Study VI** finally showed that early auditory integration of lyrics and melodies in the ventral stream depends on the neural integrity of the hippocampus: Adaptation to lyrics and neural integration of lyrics and melodies in STG/STS were diminished in patients with left (or right) unilateral hippocampal sclerosis, compared to healthy controls. We had chosen to test these patients for two reasons: First, lyrics-melody segregation has been mostly observed in patients with temporal lobe lesions (Hébert & Peretz, 2001; Hébert et al., 2003; Peretz, 1996; Racette et al., 2006) including hippocampal damage (Samson & Zatorre, 1991), rather than in healthy participants (Crowder et al., 1990; Peretz, Gagnon, Hébert, & Macoir, 2004; Serafine, 1984; Serafine et al., 1986). Second, hippocampus plays a key role in the **binding** of sensory input into conjunctive representations (Davachi, 2006; Opitz, 2010) during implicit memory formation (Chun & Phelps, 1999; Graham, Barense, & Lee, 2010) via recurrent hippocampal-cortical loops (Eichenbaum, 2000; McClelland, McNaughton, & O'Reilly, 1995). Taking lateral temporal adaptation effects as index for incidental memory formation through hippocampal top-down modulation during repeated exposure to song, we propose that hippocampal malfunction in our patients and/or deficient hippocampal-cortical connectivity (P. Besson et al., 2014) weakened the emergence of memory traces for lyrics and their binding with melody into a unified representation of song. Together with earlier lesion evidence (Samson & Zatorre, 1991) and recent support from healthy participants (Alonso et al., 2016), this study, hence, highlighted the central role of hippocampus and cortico-subcortical circuits, beyond established cortico-cortical networks, for normal integration of concomitant melodic and verbal information in song. More generally, this study clarified that neurological disorder may entail segregation of otherwise integrated processes, bearing potential for conflicting results.

Overall, Studies IV-VI demonstrate that the integration of melodic and verbal information during perception of speech and song is a complex network capacity that requires dynamic inter- and intrahemispheric as well as cortico-subcortical information exchange within and between large-scale neural networks. Despite their different approaches, all three studies seem to converge on cross-stream interactions at early auditory processing stages in the mid-posterior temporal lobe, although it is certainly premature to rule out interactions at later stages (for a corresponding proposal, see Schirmer & Kotz, 2006). Future studies should zoom into the neurophysiological underpinnings of this auditory

cross-talk between melodic and verbal information, explore their possible grounding in nested oscillatory activity (Giraud & Poeppel, 2012; Kreiner & Eviatar, 2014), and scrutinize potential cross-stream synergies through melodic and/or linguistic enrichment (e.g., musical training or bilingualism; Kraus & Chandrasekaran, 2010) and their benefits for everyday music and speech perception.

1.3.3 Melody in (inter)action

After having probed the way by which pitch accompanies, enriches and interacts with verbal-linguistic information, Chapter 4 (“Melody in (inter)action”) takes a seemingly contrary perspective turning to the communicative power of melody *in its own rights*, beyond verbal content. Particularly, instead of focusing on individual *neural* networks for melody, this chapter turns to the use of melody in *social* networks summarizing two studies on the role of pitch-based conventions in driving conversations (Study VII) and coordinating musical (inter)actions (Study VIII).

Study VII (Hellbernd & Sammler, 2016) was inspired by pragmatic theory positing that **speech** communication is not only about *what* is said, but essentially about *why* things are said (Bühler, 1934; Grice, 1957; Wittgenstein, 1953). Depending on the latter, we may understand the same question “Do you know what time it is?” as criticism or simple search for information. It is particularly the *why*—the communicative intention of the speaker—that drives the recipient’s behaviour and is the true motive of communication (Austin, 1962; Grice, 1957; Searle, 1969). Yet, speakers do not always code their intentions literally (Holtgraves, 2005). In Study VII, we sought to establish that conventionalized prosodic patterns contribute to the communication of “unspoken” social intentions in everyday adult conversation. In three experiments, we combined perceptual judgments of listeners with acoustic analyses of single (non-)words that were intoned to express six different intentions (criticism, doubt, naming, suggestion, warning, wish; for a similar approach to vocal emotions, see Banse & Scherer, 1996; Sauter, Eisner, Calder, & Scott, 2010). Discriminant analyses of the acoustics showed characteristic prosodic feature configurations for different intentions that were used *consistently* across speakers and utterances, pointing to the existence of conventionalized prosodic signatures that are appropriate for different communicative goals. Correspondingly, these patterns were reliably recognized by listeners, despite lack of contextual information (i.e., in single words), irrespective of emotional connotations (i.e., perceived valence and arousal), and irrespective of

whether the utterance carried lexical meaning or not (i.e., in words and non-words alike). Hence, the combined data demonstrate that prosody *itself* conveys communicative concepts, beyond verbal content, and can, thus, determine the course and success of interpersonal interactions. Notably, the present data highlight the communicative value of prosody in everyday *adult* conversation, extending earlier findings from preverbal infants who use prosodic patterns as primitive devices to express simple intents (Esteve-Gibert & Prieto, 2013; Prieto et al., 2012) (see also Fernald, 1989). In other words, Study VII may disclose the persistence of an onto- and perhaps phylogenetically early capacity to communicate without words (Levinson, 2006; Oller & Griebel, 2014) via vocal prosodic tone that complements—advertently or inadvertently—adults’ mature verbal abilities.

What remains to be shown in future studies is the neural implementation of this communicative prosodic capacity. Our current working hypothesis is that the comprehension of speaker’s intentions from prosody relies on a weighted contribution of auditory-prosodic circuits (as investigated in Chapter 2) and other, socio-cognitive networks whose exact nature and ways of interaction we are currently clarifying (Hellbernd & Sammler, 2018). At this point, it shall be revealed that our data lend evidence for *mentalizing* processes, i.e., involvement of theory of mind areas (Frith & Frith, 2006), during comprehension of speakers’ intent—in line with recent work on (non-prosodic) social communicative inference (Bašnáková, Weber, Petersson, van Berkum, & Hagoort, 2013; Egorova, Pulvermüller, & Shtyrov, 2014; Egorova, Shtyrov, & Pulvermüller, 2016), and mentalizing during recognition of abstract intentions that are not derivable from bodily motion (Canessa et al., 2012; Van Overwalle & Baetens, 2009).

Finally, **Study VIII** (Sammler, Novembre, Koelsch, & Keller, 2013) turns to the recognition of intentions and action goals in **music**, as a prerequisite for fluent ensemble performance. The study drew inspiration from theories of joint action (Knoblich, Butterfill, & Sebanz, 2011; Sebanz & Knoblich, 2009) and dialogue (Pickering & Garrod, 2013) positing that successful interpersonal interaction requires precise temporal coordination that can only be achieved if participants *predict* co-performers’ actions. These predictions are purportedly generated by internal models that serve to *simulate* co-actors’ on-going movements (Wolpert, Doya, & Kawato, 2003) and run slightly ahead of action execution thereby allowing anticipation of co-performers’ action goals and timely preparation and adaption of one’s own actions (Bekkering et al., 2009). It is firmly established that these internal models allow musicians to predict *when* an event will occur (for review, see Keller, 2012). In Study VIII, we sought to establish that prediction of *what* comes next is equally important for coordinated (musical) action planning. Notably, we showed that

these predictions occur—non-trivially—in unrehearsed pieces by drawing on formalized harmonic conventions in music, sometimes referred to as “syntax” (Swain, 1995).

To pinpoint harmony-based prediction of others’ actions in (quasi) joint musical behaviour, we used a real-time imitation task (Novembre & Keller, 2011) in which trained classical pianists watched and simultaneously imitated chord sequences performed by a pianist’s hand on screen. All sequences followed the rules and conventions of Western tonal harmony (Riemann, 1877/1971, see also 1971), which musicians are highly familiar with. We probed predictive processes by registering behavioural and EEG responses to more or less surprising (mis)matches between anticipated and observed action goals. Therefore, we manipulated the predictability and harmonic fit of sequence-final chords by placing them at the end of 5- or 2-chord sequences (to induce high/low predictability, respectively) and by rendering them harmonically (in)congruent with the preceding context. Performance of both pianists (i.e., video and participant) was muted to avoid auditory interference. Imitation was faster and more accurate for chords that matched the harmonically predicted action goal. Mismatches induced an early anterior and late posterior negativity in the EEG, previously associated with observation of performance errors in others (van Schie, Mars, Coles, & Bekkering, 2004), and inhibition and reprogramming of one’s own action (Leuthold, 2004). These neuro-behavioural effects were strongest during long, i.e., highly predictable sequences, demonstrating that pianists indeed used their knowledge of harmony—despite absence of sound—to anticipate the co-performers’ action goal and prepare their own movement ahead of time. Altogether, this exposes harmonic conventions as a (culturally grounded) scaffold that is co-represented by inter-actors and may, hence, provide common ground for (motoric) predictions and behavioural coordination in joint musical performance.

Taken together, Studies VII and VIII demonstrate—despite their different scopes—that melodic conventions are socially relevant, in that they disclose performers’ intentions and action goals. Whether coded in abstract sound patterns (Study VII) or concrete bodily movements (Study VIII), pitch-based conventions can be used by receivers and senders alike for coordination of interleaved or concurrent (inter)actions. Future studies should delve further into potential cross-domain similarities that are derivable from recent developments in psycholinguistics putting syntax- and pitch-based predictions as well as decoding of speakers’ “unspoken” intentions (speech acts) at the basis of turn-coordination in spontaneous spoken interactions (Holtgraves, 2008; Levinson & Torreira, 2015; Pickering & Garrod, 2013). Furthermore, future efforts should elucidate neural networks for pitch-related categorization and prediction and their interaction with socio-

cognitive neural networks (Van Overwalle & Baetens, 2009), ideally by using (truly) interactive setups (Novembre, Sammler, & Keller, 2016).

1.4 General discussion and outlook

Melody and pitch are central to human society and culture: Not only do they shape language comprehension and music-aesthetic appreciation; they are also key to social bonding and regulation of interactive behaviour. The presented studies took a fresh look onto these linguistic and social functions of pitch use in speech and music, seeking to dissect their inevitably complex neural organization. The thesis' main tenets and directions for future research will be summarized and discussed in the following.

1.4.1 Dual streams for melody in speech and music

In the absence of any particularly communicative verbal content or social meaning, the processing of melody rests on a **dual stream architecture** in both speech (Studies I & II) and music (Studies III, V & VI). This architecture shows a **relative right-dominance** and is depicted in Figure 1.2 (right panel), side-by-side with established left-hemispheric networks for (non-prosodic) language processing (left panel; Friederici, 2011) (see also Hickok & Poeppel, 2007; Rauschecker & Scott, 2009; Ueno et al., 2011), highlighting the *structural* similarities of (multiple) dorsal and ventral streams in both domains and hemispheres. For the time being, we assign analogous *functional* roles to corresponding left- and right-hemispheric streams (see below) that they may, however, exert preferably on verbal or pitch-related information, respectively. A relative (though not absolute) right dominance of melody processing in speech and music is in line with prevailing cue-dependent models of hemispheric specialization. These propose a processing benefit of the right hemisphere for spectral information (Zatorre, 2002) that unfolds over extended timescales (Giraud et al., 2007; Poeppel, 2003). The following paragraphs will briefly outline the dual streams' hypothesized roles and discuss outstanding questions, before turning to the role of pitch-based conventions in guiding conversations and coordinating musical (inter)actions.

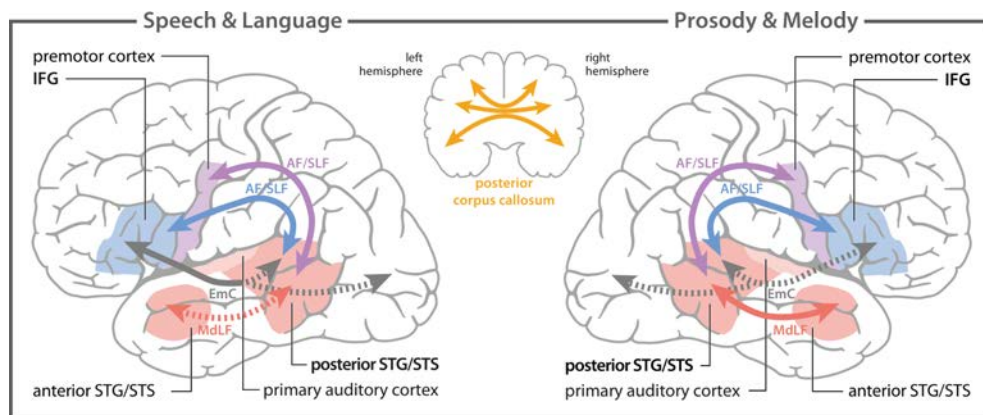


Figure 1.2: Schema of processing streams in the left and right hemisphere. Left-hemispheric dorsal and ventral streams for speech and language are shown as described by (Friederici, 2011), including temporal-(pre)motor connections for audio-motor conversion of syllables and phonemes (purple), temporal-inferior frontal connections for hierarchical syntactic structure building (blue), ventral connections for semantic processes (red) and local syntactic structure building (grey). Right-hemispheric streams for prosody and musical melody follow the findings of Studies I-III, V & VI of the present thesis, showing temporal-(pre)motor connections for pitch contour tracking and auditory enhancement via vocal simulation (purple), temporal-inferior frontal connections for musical and prosodic structure building (blue), ventral connections for local musical structure building (grey), and integration of auditory “Gestalts” that may link to meaning (red). The posterior corpus callosum (orange) is depicted as inter-hemispheric interface between prosodic and non-prosodic processing streams. AF/SLF: arcuate/superior longitudinal fascicle, EmC: extreme capsule, MdLF: middle longitudinal fascicle, IFG: inferior frontal gyrus, STG/STS: superior temporal gyrus/sulcus.

Right ventral stream I – Sound to meaning. As argued in Chapter 2, the ventral stream along the temporal lobe may map sound to meaning (Galantucci et al., 2011; Kümmeler et al., 2013; Weiller, Bormann, Saur, Musso, & Rijntjes, 2011) by virtue of gradual auditory feature integration and formation of auditory “Gestalts” (Chevillet, Riesenhuber, & Rauschecker, 2011; Rauschecker & Scott, 2009; Schirmer & Kotz, 2006). This idea resonates with earlier findings describing progressive grouping of single tones into melodic sequences (Patterson, Uppenkamp, Johnsrude, & Griffiths, 2002), and gradual emergence of complex sound categories along the anterior temporal lobe (Belin & Zatorre, 2003; Leaver & Rauschecker, 2010; Merrill et al., 2016; Norman-Haignere, Kanwisher, & McDermott, 2015; Tierney, Dick, Deutsch, & Sereno, 2013). One point worth reflecting upon is, however, **what “meaning” means** when it comes to melody in speech and music. Arguably, neither prosodic nor musical pitch patterns convey messages as unambiguously as words refer to objects in the world. Rather, both pragma-linguists (Wichmann, 2002) and musicologists (Meyer, 1956) agree that melodies *connote* a set of conceptually

related images that become particularized in the experience of the individual listener by virtue of *inference*. Notably, in both speech and music, the acoustic composition of a melody limits the connotations possible: Low terminal tones in musical or prosodic contours trigger, for example, concepts of rest, closure and fulfilment, rather than strain, openness and yearning. One novel finding of this thesis is that connotations in prosody can further pertain to abstract *interpersonal* concepts (even beyond the display of emotions, Juslin & Laukka, 2003): In Study VII, we showed that prosodic signatures are conventionalized to the extent that listeners can infer interlocutors' social communicative *intentions*, as decisive factor in everyday interactions (Hellbernd & Sammler, 2016) (for perception of intentionality in music, see Steinbeis & Koelsch, 2009). Importantly, the considerable consistency with which speakers used distinct prosodic signatures for different intents may allow for a *norm-based coding* of intonational meaning: Speakers may vocally target entries in an experience-dependent inventory of idealized acoustic representations against which listeners, in turn, reference the perceived prosodic "Gestalt" (see Study Ia) to identify the implied meaning (Latinus, McAleer, Bestelmeyer, & Belin, 2013). The neural reality and implementation of such a mechanism, its likely probabilistic nature (Tanenhaus, Kurumada, & Brown, 2015), as well as its certainly inevitable linkage with higher-order inferential processes are important topics for future research.

Right ventral stream II – Simple structure. Next, it should be noted that in music, unlike in prosody, connotations go beyond socio-emotional qualities; additionally, meaning can emerge through *structural* relationships between tones and chords (Meyer, 1956)⁵ that have no counterpart in prosody's gliding contours. This intra-musical way to meaning may be one reason why our harmonic manipulation in Study III induced activity in the ventral stream connecting frontal and temporal areas (Bianco et al., 2016). The gradual build-up of a structural-harmonic ("syntactic") representation of the musical sequence as such would be an alternative (yet closely related) interpretation that parallels proposals of local syntactic structure building in language along (left) ventral fronto-temporal fibre tracts (Friederici, 2011). Irrespective of which interpretation holds, the present data allow the insight that the right *ventral* stream certainly plays a role in the perception of musical structure, thereby rectifying the emerging impression in the literature (Loui et al., 2011; Peretz, 2016, see also below) that integration of structured pitch over time falls *solely* within right dorsal stream remit. Not only is the number of structural connectivity studies ($N = 3$) still too limited (Chen et al., 2015; Loui et al., 2009, 2011) to draw firm

⁵In music, the concept of closure is, for example, not only conveyed by falling contour and "cessation or silence" (Meyer, 1956, p. 129) but reinforced (or diluted) by a terminal tonic (or its absence, respectively; see Chapter 1.1; for further ways to meaning in music, see Koelsch, 2011b).

conclusions on the primacy of the dorsal route. Also would a single-stream organization be at odds with the apparent ubiquity of multiple processing streams across domains (Friederici, 2011; Hickok & Poeppel, 2007; Kilner, 2011; Kravitz et al., 2011). Future studies will need to explicitly address connectivity in the musical brain, remaining open to the intuition that—similar to processing of syntax (Friederici, 2011) or prosody in speech (Study I)—also parsing of melodic structure in music may involve more complex neural connectivity patterns than previously believed.

Right dorsal stream I – Sound to articulation. In prosody, we illustrated the complexity of network connections not only by showing a distinction between dorsal and ventral streams (Study Ia), but also by revealing two dorsal tracts interfacing temporal with premotor and inferior frontal areas, respectively. As argued in Chapter 2, the perception of vocal pitch contours may employ temporal-premotor connections, i.e., the dorsal audio-(pre)motor system (see also Lévêque & Schön, 2015), as a suitable device to *track* pitch contours over time (like in auditory feedback control during speech and singing; Guenther & Vladusich, 2012; Halwani et al., 2011; Hickok, 2012; Houde & Chang, 2015; Zarate, 2013), and to possibly *enhance* melody perception by invocation of a sub-vocal motor-code (vocal simulation; Hickok & Rogalsky, 2011; Rauschecker, 2011). The former aspect is in line with Studies V and VI that showed premotor cortex involvement during passive listening to melody in songs, particularly when stimuli were on repeat allowing to sub-vocally “sing along” (see also Tierney et al., 2013). The latter aspect found specific support in Study I (Sammler et al., 2015) in which inhibitory stimulation of (laryngeal) PMC temporarily decreased participants’ performance in prosodic categorization (for converging findings, see D’Ausilio, Bufalari, Salmas, Busan, & Fadiga, 2011), in line with (non-prosodic) motor theories of speech perception (A. M. Liberman & Mattingly, 1985) (for recent reviews, see Schomers & Pulvermüller, 2016; Skipper et al., 2017). One key question of our on-going research is whether premotor involvement in pitch contour perception is genuine or strategic (i.e., task-dependent), with the goal to specify the audio-premotor pathway as *principal* or *ancillary* constituent of the neural network for melody perception.

Right dorsal stream II – Structure building. Beyond basic (pre)motor pitch monitoring, right dorsal connections between temporal and inferior frontal areas may support higher-level parsing of musical or speech melodic structure (as discussed in Studies II & III). As mentioned above, the latter idea has been lingering for a while in the field of *music* cognition (Loui et al., 2009, 2011; Peretz, 2016) (but see Chen et al., 2015) and finds partial support in our Study III (Bianco et al., 2016). Notably, however, our data seem to suggest both dorsal *and* ventral routes for music-structural parsing (for converging results, see Musso

et al., 2015), drawing an interesting analogy to models of auditory language processing in the left hemisphere (Friederici, 2011; S. M. Wilson et al., 2011). Those ascribe hierarchical and local syntactic structure building to dorsal and ventral tracts, respectively, two processing modes that were both targeted—yet inseparably entwined—in Study III, clearly showing leads for future empirical research (for hierarchical structure building in music, see Koelsch, Skouras, & Jentschke, 2013). Interestingly, recent proposals have also started to associate the dorsal stream with structure processing in *prosody*, including segmentation of the speech input into basic prosodic units (e.g., prosodic words or intonational phrases) and their combination into larger syntactic structures (Bornkessel-Schlesewsky & Schlewsky, 2013). The results of our Study II are compatible (or at least not incompatible)⁶ with this hypothesis: The employed sentence-level prosody task required accurate prosodic structure building, a function that we found to be compromised when dorsal stream tissue properties were perturbed (Sammler et al., 2018). It should be noted here that segmentation of prosodic units is certainly not trivial and unlikely to draw on prosodic cues alone. Rather we propose that it is additionally informed by syntactic structure that is constantly integrated with prosody during natural sentence comprehension (Buxó-Lugo & Watson, 2016; Cole, 2015; Cutler et al., 1997). This brings us to the question of cross-stream interactions.

Cross-stream interactions. As argued in Chapter 3 (Study IV), syntactic priors for prosodic structure building most likely chime in via the posterior corpus callosum (Sammler, Kotz, et al., 2010) as a crucial interface between left- and right-hemispheric speech processing streams (Friederici et al., 2007; Steinmann & Mulert, 2012). The two streams would, thus, interact at early (yet possibly top-down controlled) processing stages in the temporal lobes (similar as in music; Studies V & VI), granting smooth and rapid coordination of complementary linguistic information. Beyond these transcallosal interhemispheric interactions, future work will need to further specify where, when and how reciprocal interactions occur (i) intrahemispherically between dorsal and ventral streams (see Study V and Cloutman, 2013), and (ii) more generally between melody and *linguistic* (Sammler, Baird, et al., 2010; Sammler, Kotz, et al., 2010) as well as *social* networks (Hellbernd & Sammler, 2018). This requires strong multi-methods approaches combining high temporal and spatial resolution (as obtained with MEG or chronometric TMS) with functional and structural connectivity measures (e.g., Granger causality and DWI) and clever exper-

⁶The limited focality of oedema-induced disruptions does not allow discriminating between dorsal sub-pathways (because both were infiltrated), such that deficits could equally emerge from impaired lower-level pitch contour tracking that then affects prosodic structure building, or impaired higher-level prosodic structure building only.

imental designs (e.g., repetition suppression or interference paradigms). At this point, it may turn out that our so far cortico-centric model of melody perception must be extended to include subcortical structures. First signs of this can be seen, for example, in the *hippocampal* involvement when binding lyrics and tunes in songs (Study VI; Alonso et al., 2014, 2016) (see also Samson & Zatorre, 1991), the assumed *thalamic* or *striatal* involvement when interfacing syntax and prosody (David, Maess, Eckstein, & Friederici, 2011), or the emerging insight that the *amygdala* acts as a computational hub that interlinks auditory-melodic and socio-affective processes (Frühholz, Trost, & Grandjean, 2014; Hellbernd & Sammler, 2018; Koelsch, 2014).

In sum, the processing of melody in speech and music rests on a dynamic ‘give and take’ between distributed, yet integrated cortical and/or subcortical systems. While future work is certainly needed to firmly assess the proposed divisions of labour—between streams and between hemispheres—the outlined model stresses the joint contribution of multiple, functionally *complementary* and perhaps *partially redundant* processing streams for melody perception that may make the system particularly versatile and robust to noise, i.e., suited to meet the changing requirements of various listening situations and idiosyncrasies of different speakers.

1.4.2 Melody's social communicative value

While the relevance of *neural connectivity* within and between circuits in the individual brain cannot be stressed enough, any model of melody perception would remain incomplete if it neglected the factor of *social connectedness* that arises from our use of melody in everyday life. As argued in Chapter 4, melodic conventions provide a scaffolding for interpersonal coordination and collaboration—in speech and music alike: Rules and customs of how to assemble pitches and shape melodic contours are shared by performers and perceivers (Studies III & VII), fostering mutual *understanding*. Notably, we use the term “conventions” to imply their likely acquisition through exposure and interaction between individuals of a specific culture. Whether or not melodic customs rest on a number of universal primitives—e.g., our preference for consonance (Fritz et al., 2009; McDermott, Schultz, Undurraga, & Godoy, 2016), a grounding in body dynamics (Sievers et al., 2013), or intuitive parenting (H. Papoušek & Papoušek, 1987)—remains an interesting question for future cross-cultural research. It may turn out that some melodic expressions tend towards universality, such as those rich in basic affective cues (Fritz et al., 2009, their study

l) that are typically used to guide behaviour of preverbal infants (including approval, prohibition, or comfort; Bryant & Barrett, 2007), while others are more specific to culture, such as music-structural systems (Cross, 2012), or vocal expressions of politeness or scepticism (Rilliard, Shochi, Martin, Erickson, & Aubergé, 2009), entailing the potential risk of cross-cultural *misunderstandings*.

Altogether, we emphasize that melody in both speech and music is a tool with social relational value that interactors use beyond infancy to regulate interpersonal encounters. This insight has profound implications on neural models of music and language that now face the challenge to identify the *social* mechanisms of music or language *use* rather than comprehension (or production) of the stimulus alone. The present thesis closes with the proposal that musical and linguistic research may benefit from core concepts of social cognition: The human capacities of action simulation (Pickering & Garrod, 2013; Sebanz & Knoblich, 2009; M. Wilson & Knoblich, 2005) and mental state attribution (Frith & Frith, 2006; Schaafsma, Pfaff, Spunt, & Adolphs, 2015) are highlighted as mechanisms that contribute to the prediction and recognition of others' goals and communicative intentions expressed in musical actions (Studies III & VIII) or speech-melodic contours (Studies I & VII). Future research will have to further scrutinize the involvement of these socio-cognitive neural systems and their interaction with the above-described auditory-prosodic and -melodic circuits (Hellbernd & Sammler, 2018), ideally by using truly interactive setups (Bögels, Barr, Garrod, & Kessler, 2015; Novembre et al., 2016; Stolk et al., 2013).

1.4.3 Conclusion

Our auditory world is populated with melodies that are instrumental in structuring sonic streams of information and in achieving social goals—both in speech and music. The present thesis demonstrates that we are only beginning to sense the complexity of the neural machinery that we are equipped with by nature to make use of this rich channel of information, for linguistic and interpersonal purposes alike. To what extent the use of melody is part of our evolutionary heritage (Cross & Woodruff, 2009; Darwin, 1871; Fitch, 2010; Mithen, 2005) or subject to cultural transmission, is topic for further research. Yet, it becomes clear that the socio- and biocultural status of melody requires an integrated view on distributed auditory-melodic, linguistic, and socio-cognitive neural systems and their dynamic interaction, to account for the diversity of melody use in society.

Part II

Experiments

Chapter 2

Neural networks for melody

2.1 Study I – Neural networks for speech prosody

Sammler, D., Grosbras, M.-H., Anwander, A., Bestelmeyer, P. E. G., Belin, P. (2015). Dorsal and ventral pathways for prosody. *Current Biology*, 25, 3079-3085.*

[Author contributions by D.S.: designed research with P.B. and M.-H.G.; performed research; analysed data with A.A. and P.E.G.B., wrote the paper, incorporating edits by co-authors]

*Reprinted with permission from Cell Press.

Report

Current Biology

Dorsal and Ventral Pathways for Prosody

Highlights

- fMRI and categorization of prosody highlight a right-hemispheric neural network
- DWI and fiber tractography reveal dual right fronto-temporal streams for prosody
- TMS suggests a motor contribution to prosody perception in the right dorsal stream

Authors

Daniela Sammler,
Marie-Hélène Grosbras,
Alfred Anwander,
Patricia E.G. Bestelmeyer, Pascal Belin

Correspondence

sammler@cbs.mpg.de

In Brief

Left dorsal and ventral streams are established as the neural core of the language faculty. Sammler et al. show a similar architecture, but with relative rightward asymmetry, for the perception of prosody. The computational roles of the dual streams may complement each other when decoding subtle vocal prosodic cues in interpersonal communication.



Sammler et al., 2015, Current Biology 25, 3079–3085
December 7, 2015 ©2015 Elsevier Ltd All rights reserved
<http://dx.doi.org/10.1016/j.cub.2015.10.009>

CellPress

Dorsal and Ventral Pathways for Prosody

Daniela Sammler,^{1,2,*} Marie-Hélène Grosbras,^{2,3} Alfred Anwander,⁴ Patricia E.G. Bestelmeyer,⁵ and Pascal Belin^{2,6,7}¹Otto Hahn Group Neural Bases of Intonation in Speech, Max Planck Institute for Human Cognitive and Brain Sciences, 04103 Leipzig, Germany²Institute of Neuroscience and Psychology, University of Glasgow, Glasgow G12 8QB, UK³Laboratoire de Neurosciences Cognitives, FR 3C, Aix-Marseille Université, 13331 Marseille, France⁴Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, 04103 Leipzig, Germany⁵School of Psychology, Bangor University, Gwynedd LL57 2AS, UK⁶Institut des Neurosciences de La Timone, UMR 7289, CNRS and Aix-Marseille Université, 13005 Marseille, France⁷International Laboratory for Brain, Music and Sound Research, University of Montréal and McGill University, Montréal H3C 3J7, Canada

*Correspondence: sammler@cbs.mpg.de

<http://dx.doi.org/10.1016/j.cub.2015.10.009>

SUMMARY

Our vocal tone—the prosody—contributes a lot to the meaning of speech beyond the actual words. Indeed, the hesitant tone of a “yes” may be more telling than its affirmative lexical meaning [1]. The human brain contains dorsal and ventral processing streams in the left hemisphere that underlie core linguistic abilities such as phonology, syntax, and semantics [2–4]. Whether or not prosody—a reportedly right-hemispheric faculty [5, 6]—involves analogous processing streams is a matter of debate. Functional connectivity studies on prosody leave no doubt about the existence of such streams [7, 8], but opinions diverge on whether information travels along dorsal [9] or ventral [10, 11] pathways. Here we show, with a novel paradigm using audio morphing combined with multimodal neuroimaging and brain stimulation, that prosody perception takes dual routes along dorsal and ventral pathways in the right hemisphere. In experiment 1, categorization of speech stimuli that gradually varied in their prosodic pitch contour (between statement and question) involved (1) an auditory ventral pathway along the superior temporal lobe and (2) auditory-motor dorsal pathways connecting posterior temporal and inferior frontal/premotor areas. In experiment 2, inhibitory stimulation of right premotor cortex as a key node of the dorsal stream decreased participants’ performance in prosody categorization, arguing for a motor involvement in prosody perception. These data draw a dual-stream picture of prosodic processing that parallels the established left-hemispheric multi-stream architecture of language, but with relative rightward asymmetry.

RESULTS

Language is a uniquely human trait that is supported by parallel but interdependent large-scale neural networks. Currently, the

work of 19th century neuroscientists undergoes a revival inasmuch as timely neural language models focus on dorsal and ventral pathways between left frontal and temporal brain areas as the neural backbone of the language faculty [2, 3]. Despite their remarkable appeal and success, these models remain incomplete in one important way: they rarely include prosody [9, 12]. The term prosody subsumes rhythmic and melodic variations in speech that convey conversational and information structure or the speaker’s emotions and intentions—thereby making it an important tool in social interaction. Until today, neurolinguistic research has sought to formalize discrete hierarchical levels of prosody perception from sensory processing via auditory integration toward evaluative judgments of prosody within fronto-temporal cortical and subcortical gray-matter structures [11, 13–15] and to determine the conditions that account for the frequent (but not constant) right-hemispheric dominance of prosody perception [6]. What is still missing, however, is a network approach to prosody that describes how prosodic information passes through these processing stages.

Experiment 1—Dual Streams for Prosody Perception

Here, we sought to fill this gap by using combined diffusion-weighted imaging (DWI) and fMRI during a prosody categorization task, compared to a non-prosodic control task. To study how the prosodic form of an utterance changes its meaning [1], the prosody task used mono-syllabic words intoned as statements or questions, i.e., conventionalized pitch contours that fulfill distinct communicative functions in everyday life. The use of single words capitalizes on their reduced dependency from left-hemispheric segmental processes (e.g., syntax) that interact with prosody perception at sentence level [16] and may obscure the boundaries of the prosody network. For stimulus construction, four adult Scottish speakers (two females) uttered the words “bear” and “pear” with either falling (statement) or rising pitch contour (question; Figure 1A, central panel). These recordings were fed into an audio morphing algorithm to obtain seven-step prosody continua in which pitch contour gradually changed from statement to question (stepwise increase of offset minus onset F0 by 15.3 Hz; Figure 1A and Audio S1). Orthogonal to that, PRAAT 5.3.01 (<http://www.praat.org>) was used to create seven-step phoneme continua in which pitch contour was fixed but voice onset time (VOT) of the word-initial consonant gradually changed from /b/ to /p/ (stepwise increase of VOT by



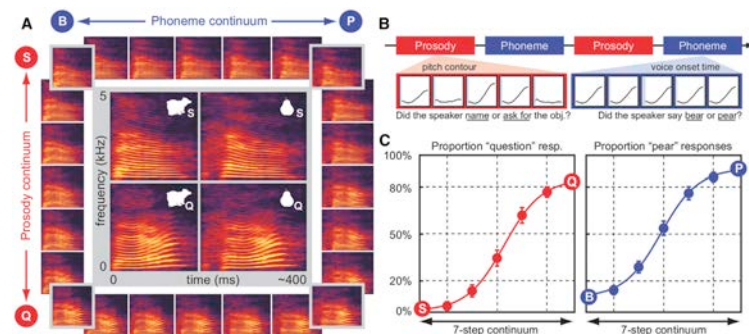


Figure 1. Experimental Design and Stimuli

(A) The words “bear” and “pear” spoken as question (Q) or statement (S) (central panel) were used to construct continua along two dimensions: prosody (vertical; Audio S1) and word-initial phoneme (horizontal; Audio S2).

(B) Participants categorized these stimuli in terms of prosody or phoneme, in separate blocks. Tasks alternated across blocks. Each block contained seven repetitions of seven prosody or seven phoneme steps, arranged according to a serially balanced sequence.

(C) Psychophysical identification functions in the prosody (red) and the phoneme (blue) task in the fMRI experiment. The similar slopes indicate that the two tasks were matched in difficulty.

See also Table S1 and Audio S1 and S2.

3 ms; Audio S2). Step size was chosen such that both tasks were of equal difficulty. Furthermore, continua were centered on each participant’s individual points of subjective equality as assessed in a pre-fMRI test (see the Supplemental Experimental Procedures).

We first used these stimuli to identify brain areas that are involved in the processing of prosodic pitch contour in speech. We used fMRI to measure neural activity in the brains of 23 healthy native English participants while they completed 16 task blocks in a 3T Siemens TIM TRIO scanner (Figure 1B). Half of the blocks required a two-alternative forced-choice judgment of prosody (did the speaker *name* or *ask* for the object?; experimental task) or of phoneme (did the speaker say *bear* or *pear*?; non-prosodic control task). Tasks were comparable in difficulty as demonstrated by the similar average regression slopes of the psychophysical identification functions in both prosody (mean slope \pm SEM: 0.35 ± 0.03) and phoneme categorization (0.34 ± 0.03 ; two-tailed paired-sample t test: $t_{22} = -0.682$, $p > 0.502$; Figure 1C).

fMRI random-effects analyses revealed stronger activations during prosody than phoneme blocks in the right posterior (pSTS) and anterior (aSTS) superior temporal sulcus, the right inferior frontal gyrus (IFG) at the border between pars opercularis and pars triangularis, and the right premotor cortex (PMC) at the level of the “larynx representation” [17] (Table 1; see also Figure S1). No such activations reached significance in the left hemisphere, and we found a relative right dominance of these clusters when examining their lateralization indices (LIs; one-sample t tests against zero; pSTS: $t_{22} = -5.156$, $p < 0.001$; aSTS: $t_{22} = -2.376$, $p < 0.027$; IFG: $t_{22} = -2.300$, $p < 0.031$; PMC: $t_{22} = -2.102$, $p < 0.047$; see Figure S1). LIs were calculated as $LI = (L - R) / (L + R)$, with R and L representing the effect sizes

of the right-hemisphere clusters’ peak voxel and its left-hemisphere homolog (negative x coordinate), respectively (see the Supplemental Experimental Procedures). Apart from these regions, we found prosody-related activations in bilateral supplementary motor area (SMA; LI: $t_{22} = -1.251$, $p > 0.224$) and left intraparietal sulcus (IPS; LI: $t_{22} = 1.967$, $p > 0.062$; see Figure S1). No area showed stronger activation during phoneme than prosody blocks (but see the Supplemental Information for a more fine-grained analysis of the phoneme data).

We then used our fMRI activation clusters in right pSTS, aSTS, PMC, and IFG as seed and target regions in multi-fiber probabilistic tractography to estimate the most likely white-matter pathways that connect these prosody-relevant nodes. No tractography was carried out in the left hemisphere given that no left fronto-temporal activation clusters were found in the fMRI experiment. This approach revealed robust ventral and dorsal fiber tracts in the right hemisphere (Figure 2). Ventrally, the pSTS and aSTS clusters were connected via the middle longitudinal fascicle (MdLF; 23/23 participants). No connections were found between the aSTS and IFG clusters. Dorsally, the pSTS-IFG connection followed the direct branch of the arcuate/superior longitudinal fascicle (AF/SLF; 19/23 participants). The pSTS-PMC connection took a similar but slightly more dorsal and posterior trajectory along the AF/SLF (18/23 participants). Finally, PMC was strongly interconnected with IFG via short U fibers (23/23 participants; see Figure S2 for individual data and Figure S3 for a 3D view of the results).

Dorsal and ventral pathways have been subject to speculation in prosody research for a while [11, 15, 18], but the available data are sparse [9, 10, 19]. They often suggest that prosodic information travels either dorsally [9, 15] or ventrally [11] between superior temporal areas and IFG. Our tractography data argue for a



Table 1. Overview of Significant Clusters in the Prosody > Phoneme Random-Effects Contrast

| Site | BA | Cluster Size | MNI Coordinate | | | Z Value |
|-----------------|-------|--------------|----------------|-----|-----|---------|
| | | | x | y | z | |
| L IPS | 40 | 84 | -33 | -46 | 34 | 4.34 |
| | | | -24 | -37 | 28 | 3.80 |
| | | | -36 | -40 | 40 | 3.67 |
| R pSTS | 21/22 | 156 | 42 | -37 | 4 | 4.36 |
| | | | 48 | -28 | 1 | 4.22 |
| R TP | 38 | 62 | 57 | 11 | -17 | 4.06 |
| R aSTS | 21/22 | | 57 | 2 | -8 | 3.75 |
| | | | 45 | 5 | -20 | 3.51 |
| R PMC | 6 | 120 | 45 | 5 | 40 | 4.00 |
| R IFG (p. op.) | 44 | | 42 | 8 | 31 | 3.50 |
| | | | 33 | 5 | 46 | 3.21 |
| ACC/SMA | 32/6 | 62 | 9 | 17 | 49 | 3.80 |
| R IFG (p. op.) | 44 | 44 | 60 | 20 | 25 | 3.71 |
| R IFG (p. tri.) | 45 | | 48 | 20 | 13 | 3.46 |
| | | | 60 | 23 | 10 | 3.39 |

Thresholded at $p < 0.001$, cluster extent ≥ 20 voxels (~ 540 mm³; equaling whole-brain $p < 0.01$). Specifications refer to peak voxels. BA, Brodmann area; MNI, Montréal Neurological Institute; L, left; R, right; IPS, intraparietal sulcus; pSTS, posterior superior temporal sulcus; TP, temporal pole; aSTS, anterior STS; PMC, premotor cortex; IFG, inferior frontal gyrus; p. op., pars opercularis; p. tri, pars triangularis; ACC, anterior cingulate cortex; SMA, supplementary motor area. See also Figure S1.

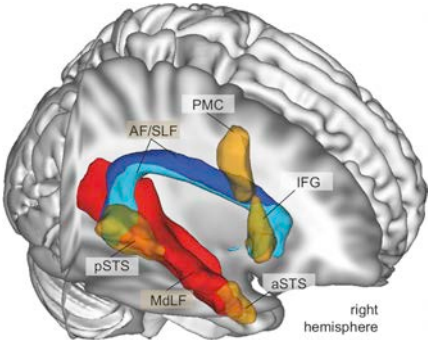


Figure 2. Functional and Diffusion MRI Data
Brain areas with a stronger blood-oxygen-level-dependent response during prosody than phoneme task (yellow; $p_{\text{voxel}} < 0.001$, $p_{\text{cluster}} < 0.01$) and group overlay of fiber tracts connecting the functional clusters: pSTS with aSTS (red), PMC (blue), and IFG (cyan). U fibers between the PMC and IFG are not depicted. Only voxels with fibers in more than 50% of participants are highlighted. Activation clusters and fiber tracts were slightly smoothed (1 mm full width at half maximum) for visualization. AF, arcuate fascicle; SLF, superior longitudinal fascicle; MdLF, middle longitudinal fascicle. Other abbreviations are as in Table 1. See also Figures S2 and S3.

more complex connectivity pattern in two respects: First, they show that information travels along both dorsal and ventral pathways in the right hemisphere. This invites hypotheses on a computational division of labor between streams as will be outlined in more detail in the Discussion below. Second, our data show the involvement of an additional dorsal auditory-motor circuit between the right pSTS and PMC. Prevailing neural prosody models [5, 11, 14, 15, 18] typically focus on superior temporal areas and IFG as established prosody processors according to numerous neuroimaging [18, 20], brain stimulation [21], and lesion [5, 13] studies. Only more recently have publications started to report PMC activations during prosody perception [22, 23]. Interestingly, PMC activity is central to an entire research tradition on (non-prosodic) motor theories of speech perception [24]. One recurrent debate in this field pertains to the question whether PMC is essential for tuning-up the processing of speech sounds, e.g., via simulation of articulatory gestures [25, 26]. While our results invite speculations about a similar mechanism in prosody perception, the inclusion of an auditory-motor pathway into a network model of prosody requires more evidence for the necessity of PMC in prosody perception in the first instance.

Experiment 2—Motor-Involvement in Prosody Perception
Here, we used repetitive transcranial magnetic stimulation (rTMS) to assess changes in prosody perception after temporary inhibition of PMC in a new group of 32 healthy native English participants. They were invited to do a shortened version of the

prosody and phoneme tasks as used in experiment 1, once after 15 min of inhibitory offline 1-Hz rTMS of the PMC, and once after 15 min of sham stimulation of the same area (Figure 3A). The order of task (prosody/phoneme) and stimulation type (rTMS/sham) was counterbalanced across participants. In half of the participants, rTMS was applied to the right PMC (as identified in experiment 1), and in the other half, the coil was placed over the homolog coordinate in the left hemisphere (negative x coordinate). No rTMS was applied to the pSTS, aSTS, or IFG because their involvement in prosody perception is firmly established [11, 13, 18, 20, 21]. The two participant groups were carefully matched in age, handedness, years of musical training, stimulator output during rTMS, and performance during sham stimulation ($p_s > 0.107$; Table S3). We ensured that both tasks were challenging and therefore sensitive to possible disruption by rTMS by centering the stimuli on each participant's individual point of subjective equality obtained before the actual TMS session. If the right PMC is functionally relevant for prosody perception, its rTMS (but not sham stimulation) should induce a performance drop that is stronger in the prosody than in the phoneme task, and that might be stronger after rTMS of the right than of the left PMC based on our fMRI results.

Slopes of the psychophysical identification functions served as dependent variable: the shallower the slope the weaker the participant's ability to identify tokens as belonging to one or the other category (statement versus question or /b/ versus /p/). The results showed a clearly reduced performance in the prosody task after rTMS of the right PMC (compared to sham stimulation; $t_{1,15} = -3.888$, $p < 0.005$, paired-sample t test, Bonferroni corrected) that was not seen after rTMS of the left PMC

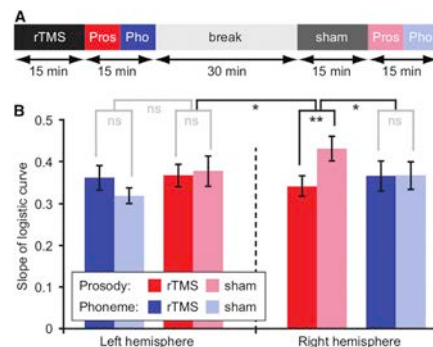


Figure 3. Repetitive TMS Experiment

(A) Participants completed one prosody (Pros) and one phoneme (Pho) block after rTMS and sham stimulation of right or left PMC, in counterbalanced order. (B) Performance in the prosody (but not phoneme) task decreased after rTMS of the right (but not left) PMC compared to sham stimulation. Bars depict mean slopes \pm 1 SEM.

See also Tables S2 and S3.

($t_{1,15} = -0.342$, $p > 0.737$; pink and red bars in Figure 3B). Furthermore, rTMS in either hemisphere did not affect performance in the phoneme task ($t_{1,15} < 2.351$, $ps > 0.132$; light and dark blue bars in Figure 3B). This is not surprising given that the stimulation site was chosen to specifically interfere with prosody, not with phoneme perception. Statistically, the interaction between task (prosody/phoneme) and stimulation type (rTMS/sham) was significant in the right hemisphere ($F_{1,15} = 7.328$, $p < 0.017$, $\eta_p^2 = 0.328$; for full statistical details, see Table S2), lending strong evidence for the functional relevance of the right PMC in prosody perception. No such interaction was found in the left hemisphere ($F_{1,15} = 3.755$, $p > 0.072$, $\eta_p^2 = 0.200$). Furthermore, the interaction between stimulation (rTMS/sham) and hemisphere (left/right) was significant in the prosody task ($F_{1,30} = 4.252$, $p < 0.049$, $\eta_p^2 = 0.124$), in line with the relative right dominance observed in experiment 1. No such interaction was found for the phoneme task ($F_{1,30} = 1.672$, $p > 0.206$, $\eta_p^2 = 0.053$). In fact, the stimulation of right (but not left) PMC consistently decreased participants' prosody perception, while phoneme categorization remained unimpaired (see the Supplemental Information for more details on the phoneme data). Finally, this effect cannot be driven by nominal performance differences in the sham conditions, as demonstrated by the absence of a task \times hemisphere interaction ($F_{1,30} = 0.02$, $p > 0.901$, $\eta_p^2 = 0.001$) in an ANOVA with the factors task (prosody sham versus phoneme sham; within subject) and hemisphere (left versus right; between subjects).

DISCUSSION

In language research, the notion of dorsal and ventral streams—as the neural core of the language faculty—has become largely

undisputed [2–4, 27]. Functionally, the left dorsal stream has been associated with time-dependent mechanisms of speech processing that afford, among others, the conversion of the auditory speech signal into a motor format to support articulation (“how”) [3, 4, 27] and that allow at the same time a motor influence on perception, i.e., the simulation of articulatory gestures to support comprehension [26, 28]. The ventral stream, in turn, has been proposed to gradually transform complex acoustic feature combinations into abstract time-invariant representations of speech sounds that interface with conceptual systems and, thus, are linked to meaning (“what”) [4, 27, 29]. Whether a similar organization also pertains to prosodic aspects of speech is a matter of debate. Here we provide evidence for a multi-stream architecture of prosodic contour perception and propose a functional division of labor that parallels prevailing language models, but in the right hemisphere.

Dorsal Pathways—Sound to Articulation and Evaluation

The identification of statements and questions by our participants evoked activity in two dorsal stream regions: right PMC and IFG. The PMC activation was located at the somatotopic level of the dorsal larynx representation [17, 30]. This area controls the pitch of voluntary vocalizations in humans. Our participants did not vocalize (according to self-report) but may have simulated the laryngeal gesture that the speaker used to produce the vocal prosodic contour to sharpen their perception [26]. This proposal is based on the notion of sensorimotor loops that provide auditory feedback control for speech production but that can be reversed to tune up speech perception by recovering the articulatory gesture that generated the speech sound [25, 28, 31, 32]. Although it is arguable whether the motor system is essential for speech perception [25, 26], there is little doubt that passive listening to speech sounds activates (pre)motor areas [26] and that up- or downregulation of lip- or tongue-related motor regions alters performance in phonetic language tasks [31, 32]. The novel implication of our combined fMRI and TMS results is that this motor involvement also applies to prosody: temporary downregulation of the right PMC led to a performance decrease that was specific to prosody categorization, in line with similar findings on emotional prosody perception [33] and vocal pitch discrimination [34]. Importantly, vocal pitch productions are known to rely on sensorimotor loops: self-vocalizations of pitch suppress auditory activity, while transient perturbations of vocal pitch feedback enhance auditory activity and induce instantaneous vocal compensation (for a review, see [35]). It is plausible to assume that these loops can also motor modulate the perception of prosodic contours, similar to how they augment the perception of non-prosodic speech sounds [31, 32]. Note that we do not rule out that motor involvement in prosody perception mainly applies to subtle modulations of vocal tone as those in our ambiguous stimuli [25]. Furthermore, we are not claiming to prove that the PMC is an area of prosodic processing per se. Rather, we favor the view that it is part of a larger action-perception network, including PMC interlinked with temporal auditory areas, the dynamics of which were unbalanced after TMS of the PMC. Crucially, our tractography results highlight the AF/SLF as the most likely anatomical link between the pSTS and PMC to support this mechanism, in keeping with influential models of auditory feedback control [35, 36].

One further node of the dorsal stream was located in the right IFG, at the border between dorsal Brodmann area 44 (BA44) and BA45. Following prevailing models of prosody perception [9, 14, 15, 18], we propose that IFG contributes task-dependent cognitive resources via the dorsal pathways to parse and explicitly label the dynamic prosodic contour. Notably, the anatomical links of IFG with pSTS and PMC via AF/SLF and U fibers, respectively, raise the interesting hypothesis that IFG may evaluate prosodic patterns in both audio and motor formats and eventually supports their integration. More generally, the time-sensitive monitoring and assessment of evolving pitch contours [28] via pSTS-IFG and pSTS-PMC(-IFG) connections may be relevant beyond single words and extend to sentence-level prosody (see below).

Overall, our results show an interesting consistency of the right dorsal tracts in 21 of our 23 participants. Increasing evidence indeed indicates that the prevalence of the right AF is higher than previously believed [37], although its volume is typically lower than in the left hemisphere [38]. Importantly, recent work on aphasia recovery [39] has started to advocate the right AF as a mediator of right-hemispheric language functions, including verbal recall [40] and prosody [39].

Ventral Pathway—Sound to Auditory Object

Categorizing the prosody of speech sounds further involved two ventral stream regions: the right posterior and anterior STS. Following models of auditory (speech) perception [4], the pSTS is an established processor of non-verbal complex acoustic forms and the aSTS a known seat of invariant sound characteristics that identify auditory objects [41]. Together, they may afford a gradual segregation and abstraction of the prosodic signal from a granular, speaker-dependent representation of speech sounds in pSTS to a speaker-invariant “prosodic Gestalt” in the aSTS [11]. A similar process of abstraction along the ventral auditory stream has been described as the basis for perceptual constancy in speech perception in the left hemisphere [4, 29, 42], as well as for the perception of voice identity [43] in the right hemisphere.

Notably, our data suggest the MdLF [44] as direct anatomical link that would allow prosody to follow these hierarchical processing steps along the axis of the temporal lobe [4]. The left MdLF has been previously described in the context of language comprehension [27, 44, 45], although many authors—when talking about the ventral pathway—refer to the extreme capsule (EmC) or uncinate fascicle (UF) connecting the temporal and inferior frontal lobe [2, 19, 46]. Indeed, a ventral projection to IFG would make IFG the most plausible interface to integrate complementary information provided by ventral and dorsal pathways. In the present study, we did not find the EmC and UF, most likely because their frontal terminations are typically more anterior (BA45/47; see Figures S2 and S3) [38] than our IFG target region (peak in BA44). However, considering the widespread notion that the ventral stream maps sound to meaning [3, 27, 46], the absence of EmC and UF involvement in the present study might reflect an essential difference of how our stimuli were processed: unlike sentences [2, 46] or prototypical emotional vocalizations [19], linguistic prosodic contours of single words may not easily map onto a semantic space, even if they can be processed as a

signal of more general interpersonal significance, to convey a concept of the speaker’s communicative intention or attitude [1]. The mechanisms that link “prosodic Gestalts” and communicative concepts are an interesting topic for future research.

Right Lateralization

Our fMRI results highlight a relative right dominance of brain activations, in line with prevailing cue-dependent models of prosody [6]. These suggest a processing benefit of the right hemisphere for spectral information [47] that unfolds over extended timescales [48]. As a consequence, this implies that white-matter tracts in the right hemisphere qualify as main (although perhaps not only) routes of prosody perception, despite the frequently reported larger volume of left-hemispheric fiber bundles [38]. Note that we do not exclude the involvement of left-hemispheric pathways *per se*, but rather favor the view that our experimental design was particularly sensitive to the right-hemispheric core components of a probably more extended bilateral network [19]. In particular, it remains to be specified in which way the present results generalize to sentence-level prosodic functions such as phrasing or accentuation. One conceivable scenario is that time-sensitive prosodic processes of the right dorsal stream interact transcallosally [16] with sentence-level syntactic and semantic operations in the left hemisphere [12]. Interestingly, the rightward asymmetry included the PMC. Although surprising, given a bilateral organization of voice control [17, 30, 49], this finding is in keeping with mounting evidence of a right premotor advantage in auditory-feedback control [35, 36] and might indicate a projection of auditory asymmetries up to premotor areas [48]. Exceptions to the right dominance were (1) a bilateral (but right-trended) activation in the SMA and (2) a left-dominant activation in the IPS, both of which may relate to the cortical sensorimotor control of (simulated) vocalizations [4, 36].

To sum up, our study indicates dual processing streams for prosody in the right hemisphere. The associated computational roles are proposed to parallel those in the left hemisphere, including the time-sensitive evaluation of prosodic contours in audio and/or motor format along the dorsal pathways (“how”) [28] and the gradual formation of time-invariant “prosodic Gestalts” (“what”) along the ventral pathway [4, 29]. While future work is needed to firmly assess these functional hypotheses, the potentially parallel involvement yet flexible weighting and ultimate fusion of both streams could provide a robust basis for prosody perception that is particularly suited to meet the demands of multiple functions and levels of prosody in various listening situations.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, three figures, three tables, and two audio files and can be found with this article online at <http://dx.doi.org/10.1016/j.cub.2015.10.009>.

AUTHOR CONTRIBUTIONS

D.S., M.-H.G., and P.B. designed the study. D.S. performed the research. D.S., A.A., and P.E.G.B. analyzed the data. D.S., M.-H.G., A.A., P.E.G.B., and P.B. co-wrote the manuscript.



ACKNOWLEDGMENTS

The research protocols were approved by the ethics committee of the College of Science and Engineering at the University of Glasgow (CSE00960; CSE01245). The authors are grateful to Tamara Rathcke, Marianne Latinus, David Fleming, Phil McAleer, Judith L. Stevenson, Rebecca Watson, Frances Crabbe, Merle-Marie Ahrens, Sibylle Mohr, Elizabeth Bucher, Lotta Ranta, and Christoph Scheepers for their help during experiments. Furthermore, the authors are thankful for insightful comments and contributions of the anonymous reviewers. The project was funded by the Otto Hahn award of the Max Planck Society to D.S. and by BBSRC grant BB/1006494/1 and FRM grant AJE201214 to P.B.

Received: December 20, 2014

Revised: September 2, 2015

Accepted: October 5, 2015

Published: November 5, 2015

REFERENCES

- Bolinger, D. (1986). *Intonation and Its Parts - Melody in Spoken English* (Stanford University Press).
- Friederici, A.D. (2011). The brain basis of language processing: from structure to function. *Physiol. Rev.* 91, 1357–1392.
- Hickok, G., and Poeppel, D. (2007). The cortical organization of speech processing. *Nat. Rev. Neurosci.* 8, 393–402.
- Rauschecker, J.P., and Scott, S.K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat. Neurosci.* 12, 718–724.
- Ross, E.D., and Monnot, M. (2008). Neurology of affective prosody and its functional-anatomic organization in right hemisphere. *Brain Lang.* 104, 51–74.
- Witteman, J., van Ijzendoorn, M.H., van de Velde, D., van Heuven, V.J., and Schiller, N.O. (2011). The nature of hemispheric specialization for linguistic and emotional prosodic perception: a meta-analysis of the lesion literature. *Neuropsychologia* 49, 3722–3738.
- Ethofer, T., Anders, S., Erb, M., Herbert, C., Wiethoff, S., Kissler, J., Grodd, W., and Wildgruber, D. (2006). Cerebral pathways in processing of affective prosody: a dynamic causal modeling study. *Neuroimage* 30, 580–587.
- Frühholz, S., and Grandjean, D. (2012). Towards a fronto-temporal neural network for the decoding of angry vocal expressions. *Neuroimage* 62, 1658–1666.
- Glasser, M.F., and Rilling, J.K. (2008). DTI tractography of the human brain's language pathways. *Cereb. Cortex* 18, 2471–2482.
- Ethofer, T., Bretscher, J., Gschwind, M., Kreifelts, B., Wildgruber, D., and Vuilleumier, P. (2012). Emotional voice areas: anatomic location, functional properties, and structural connections revealed by combined fMRI/DTI. *Cereb. Cortex* 22, 191–200.
- Schirmer, A., and Kotz, S.A. (2006). Beyond the right hemisphere: brain mechanisms mediating vocal emotional processing. *Trends Cogn. Sci.* 10, 24–30.
- Friederici, A.D., and Alter, K. (2004). Lateralization of auditory language functions: a dynamic dual pathway model. *Brain Lang.* 89, 267–276.
- Baum, S.R., and Pell, M.D. (1999). The neural bases of prosody: insights from lesion studies and neuroimaging. *Aphasiology* 13, 581–608.
- Brück, C., Kreifelts, B., and Wildgruber, D. (2011). Emotional voices in context: a neurobiological model of multimodal affective information processing. *Phys. Life Rev.* 8, 383–403.
- Wildgruber, D., Ackermann, H., Kreifelts, B., and Ethofer, T. (2006). Cerebral processing of linguistic and emotional prosody: fMRI studies. *Prog. Brain Res.* 156, 249–268.
- Sammler, D., Kotz, S.A., Eckstein, K., Ott, D.V.M., and Friederici, A.D. (2010). Prosody meets syntax: the role of the corpus callosum. *Brain* 133, 2643–2655.
- Brown, S., Ngan, E., and Liotti, M. (2008). A larynx area in the human motor cortex. *Cereb. Cortex* 18, 837–845.
- Frühholz, S., and Grandjean, D. (2013). Processing of emotional vocalizations in bilateral inferior frontal cortex. *Neurosci. Biobehav. Rev.* 37, 2847–2855.
- Frühholz, S., Gschwind, M., and Grandjean, D. (2015). Bilateral dorsal and ventral fiber pathways for the processing of affective prosody identified by probabilistic fiber tracking. *Neuroimage* 109, 27–34.
- Frühholz, S., and Grandjean, D. (2013). Multiple subregions in superior temporal cortex are differentially sensitive to vocal expressions: a quantitative meta-analysis. *Neurosci. Biobehav. Rev.* 37, 24–35.
- Alba-Ferrara, L., Ellison, A., and Mitchell, R.L. (2012). Decoding emotional prosody: resolving differences in functional neuroanatomy from fMRI and lesion studies using TMS. *Brain Stimulat.* 5, 347–353.
- Warren, J.E., Sauter, D.A., Eisner, F., Willand, J., Dresner, M.A., Wise, R.J., Rosen, S., and Scott, S.K. (2006). Positive emotions preferentially engage an auditory-motor “mirror” system. *J. Neurosci.* 26, 13067–13075.
- Beaucousin, V., Lacheret, A., Turbelin, M.R., Morel, M., Mazoyer, B., and Tzourio-Mazoyer, N. (2007). fMRI study of emotional speech comprehension. *Cereb. Cortex* 17, 339–352.
- Liberman, A.M., and Mattingly, I.G. (1985). The motor theory of speech perception revised. *Cognition* 21, 1–36.
- Hickok, G., Houde, J., and Rong, F. (2011). Sensorimotor integration in speech processing: computational basis and neural organization. *Neuron* 69, 407–422.
- Pulvermüller, F., and Fadiga, L. (2010). Active perception: sensorimotor circuits as a cortical basis for language. *Nat. Rev. Neurosci.* 11, 351–360.
- Saur, D., Kreher, B.W., Schnell, S., Kümmerer, D., Kellmeyer, P., Vry, M.S., Umarova, R., Musso, M., Glauche, V., Abel, S., et al. (2008). Ventral and dorsal pathways for language. *Proc. Natl. Acad. Sci. USA* 105, 18035–18040.
- Rauschecker, J.P. (2011). An expanded role for the dorsal auditory pathway in sensorimotor control and integration. *Hear. Res.* 271, 16–25.
- Scott, S.K., Blank, C.C., Rosen, S., and Wise, R.J. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain* 123, 2400–2406.
- Bouchard, K.E., Mesgarani, N., Johnson, K., and Chang, E.F. (2013). Functional organization of human sensorimotor cortex for speech articulation. *Nature* 495, 327–332.
- Möttönen, R., and Watkins, K.E. (2009). Motor representations of articulators contribute to categorical perception of speech sounds. *J. Neurosci.* 29, 9819–9825.
- D'Ausilio, A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C., and Fadiga, L. (2009). The motor somatotopy of speech perception. *Curr. Biol.* 19, 381–385.
- Banissy, M.J., Sauter, D.A., Ward, J., Warren, J.E., Walsh, V., and Scott, S.K. (2010). Suppressing sensorimotor activity modulates the discrimination of auditory emotions but not speaker identity. *J. Neurosci.* 30, 13552–13557.
- D'Ausilio, A., Bufalari, I., Salmas, P., Busan, P., and Fadiga, L. (2011). Vocal pitch discrimination in the motor system. *Brain Lang.* 118, 9–14.
- Houde, J.F., and Chang, E.F. (2015). The cortical computations underlying feedback control in vocal production. *Curr. Opin. Neurobiol.* 33, 174–181.
- Guenther, F.H., and Hickok, G. (2016). Neural models of motor speech control. In *Neurobiology of Language*, G. Hickok, and S.L. Small, eds. (Academic Press), pp. 725–740.
- Fernández-Miranda, J.C., Wang, Y., Pathak, S., Stefaneau, L., Verstynen, T., and Yeh, F.-C. (2015). Asymmetry, connectivity, and segmentation of the arcuate fascicle in the human brain. *Brain Struct. Funct.* 220, 1665–1680.
- Thiebaut de Schotten, M., Ffytche, D.H., Bizzi, A., Dell'Acqua, F., Allin, M., Walshe, M., Murray, R., Williams, S.C., Murphy, D.G., and Catani, M. (2011). Atlas location, asymmetry and inter-subject variability of white

- matter tracts in the human brain with MR diffusion tractography. *Neuroimage* 54, 49–59.
39. Forkel, S.J., Thiebaut de Schotten, M., Dell'Acqua, F., Kalra, L., Murphy, D.G., Williams, S.C., and Catani, M. (2014). Anatomical predictors of aphasia recovery: a tractography study of bilateral perisylvian language networks. *Brain* 137, 2027–2039.
 40. Catani, M., Allin, M.P.G., Husain, M., Pugliese, L., Mesulam, M.M., Murray, R.M., and Jones, D.K. (2007). Symmetries in human brain language pathways correlate with verbal recall. *Proc. Natl. Acad. Sci. USA* 104, 17163–17168.
 41. Petkov, C.I., Logothetis, N.K., and Obleser, J. (2009). Where are the human speech and voice regions, and do other animals have anything like them? *Neuroscientist* 15, 419–429.
 42. Obleser, J., and Eisner, F. (2009). Pre-lexical abstraction of speech in the auditory cortex. *Trends Cogn. Sci.* 13, 14–19.
 43. Belin, P., and Zatorre, R.J. (2003). Adaptation to speaker's voice in right anterior temporal lobe. *Neuroreport* 14, 2105–2109.
 44. Makris, N., Papadimitriou, G.M., Kaiser, J.R., Sorg, S., Kennedy, D.N., and Pandya, D.N. (2009). Delineation of the middle longitudinal fascicle in humans: a quantitative, in vivo, DT-MRI study. *Cereb. Cortex* 19, 777–785.
 45. Wong, F.C., Chandrasekaran, B., Garibaldi, K., and Wong, P.C. (2011). White matter anisotropy in the ventral language pathway predicts sound-to-word learning success. *J. Neurosci.* 31, 8780–8785.
 46. Weiller, C., Bormann, T., Saur, D., Musso, M., and Rijntjes, M. (2011). How the ventral pathway got lost: and what its recovery might mean. *Brain Lang.* 118, 29–39.
 47. Zatorre, R.J., Belin, P., and Penhune, V.B. (2002). Structure and function of auditory cortex: music and speech. *Trends Cogn. Sci.* 6, 37–46.
 48. Giraud, A.L., Kleinschmidt, A., Poeppel, D., Lund, T.E., Frackowiak, R.S., and Laufs, H. (2007). Endogenous cortical rhythms determine cerebral specialization for speech perception and production. *Neuron* 56, 1127–1134.
 49. Cogan, G.B., Thesen, T., Carlson, C., Doyle, W., Devinsky, O., and Pesaran, B. (2014). Sensory-motor transformations for speech occur bilaterally. *Nature* 507, 94–98.

Current Biology

Supplemental Information

Dorsal and Ventral Pathways for Prosody

**Daniela Sammler, Marie-Hélène Grosbras, Alfred Anwander, Patricia E.G. Bestelmeyer,
and Pascal Belin**

Supplemental information

SUPPLEMENTAL FIGURES AND TABLES:

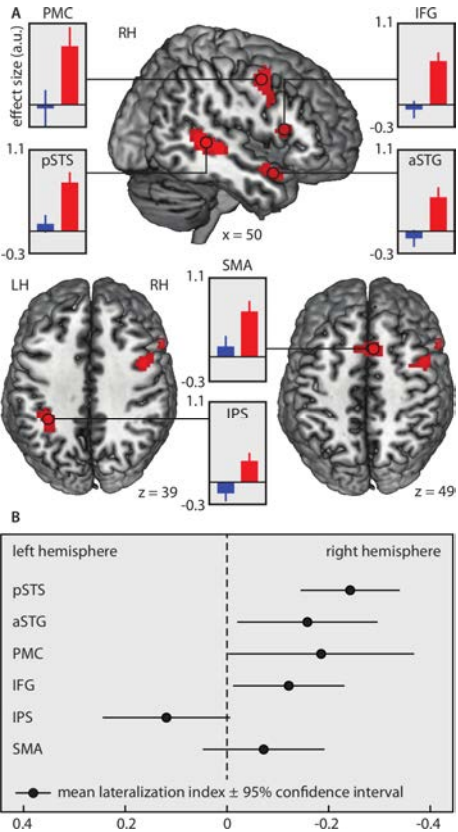


Figure S1 Related to Table 1. Functional MRI data. **(A)** Brain areas involved in the processing of prosody. SPMs depict T -values of the contrast of prosody > phoneme task thresholded at voxel-level $P < .001$, cluster-level $P < .01$. Bar graphs depict mean effect sizes (a.u.) ± 1 SEM at peak voxels during the prosody (red) and the phoneme task (blue). **(B)** Mean lateralization index LI $\pm 95\%$ confidence interval. Posterior STS, aSTS, PMC and IFG activations were right-lateralized, while the IPS showed a trend towards left-hemispheric dominance. pSTS, posterior superior temporal sulcus; aSTS, anterior superior temporal sulcus; PMC, premotor cortex; IFG, inferior frontal gyrus; SMA, supplementary motor area; IPS, intraparietal sulcus.

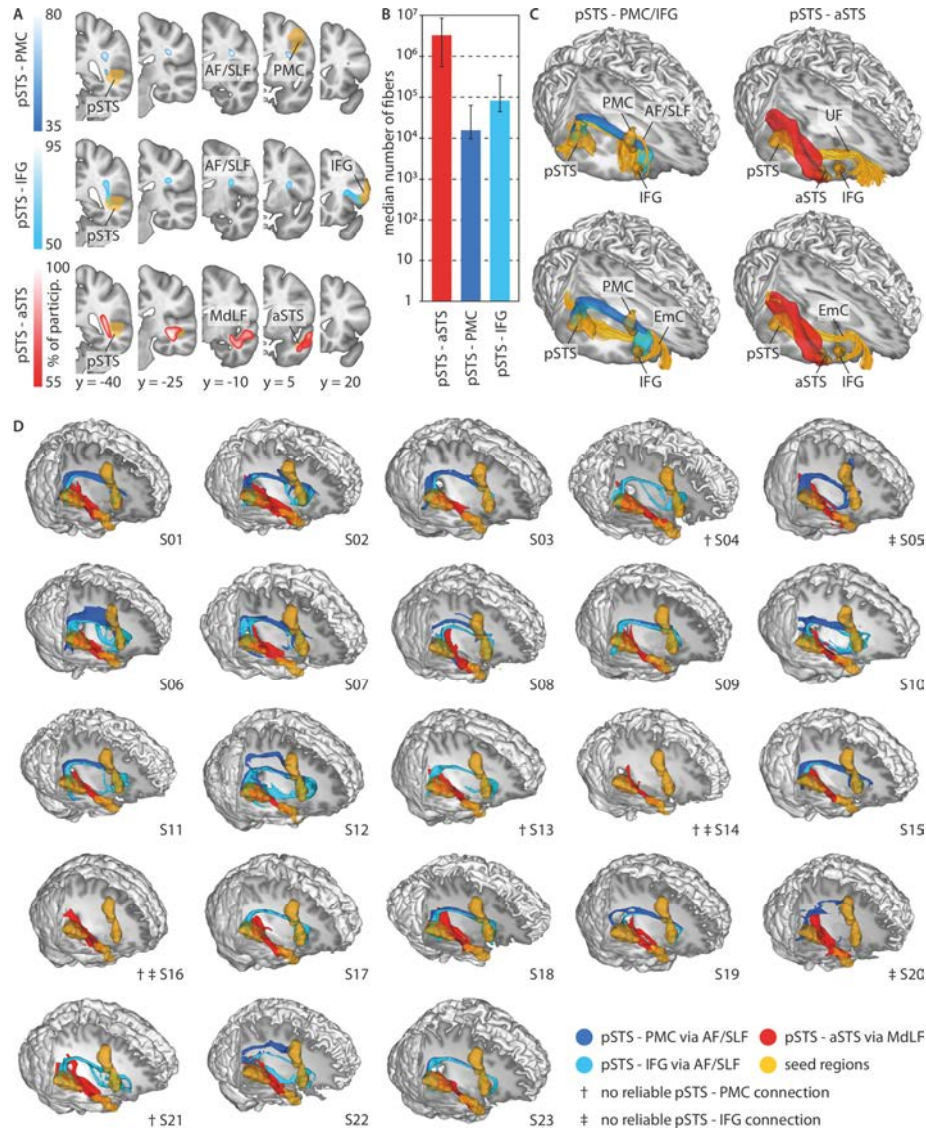


Figure S2 Related to Figure 2 and Figure S3. Diffusion MRI data. **(A)** Coronal slices through the group overlay of fiber tracts between pSTS and PMC (top row), pSTS and IFG (middle row), and pSTS and aSTS (bottom row). Y-coordinates are provided in MNI space. The color code indicates the percentage of participants with fibers in the respective voxels. Seed regions are displayed in yellow. **(B)** Quantitative connection strengths of the fiber tracts connecting pSTS with aSTS (red), PMC (blue), and IFG (cyan) across participants. The ventral connection (red) was significantly stronger than both dorsal connections, and the dorsal connection to IFG

was stronger than the dorsal connection to PMC (Wilcoxon signed rank tests: W 's > 4.107, P 's < .001, Bonferroni corrected). Bars indicate the median number of fibers contained in each of the three pathways. Error bars show quartiles. **(C)** Group overlay of dorsal (blue and cyan, left panels) and ventral fiber tracts (red, right panels) plotted onto the brain of a single representative participant. Seed regions are displayed as yellow clusters. As anatomical landmarks, the participant's AF/SLF (left upper panel), the uncinate fascicle (UF; right upper panel) and the extreme capsule (EmC; lower panels) are shown in yellow [S1, S2]. Note that both the UF and the EmC project to frontal regions that are more anterior than our IFG seed region (see Figure S3 for a 3D view). BrainGL was used for the segmentation of these fiber bundles (<http://braingl.googlecode.com>). Seed regions were placed according to the literature in the posterior temporal lobe and ventral pars triangularis of the IFG to segment the EmC [S2], the temporal pole and ventral pars triangularis/orbitalis of the IFG to segment the UF [S3], the posterior temporal lobe and dorsal pars opercularis / ventral premotor cortex to segment the direct branch of the AF/SLF [S4, S5]. **(D)** Tractography results in single subjects. AF, arcuate fascicle; SLF, superior longitudinal fascicle; EmC, extreme capsule; MdLF, middle longitudinal fascicle; UF, uncinate fascicle; pSTS, posterior superior temporal sulcus; aSTS, anterior superior temporal sulcus; IFG, inferior frontal gyrus; PMC, premotor cortex.

Table S1 Related to Figure 1. Acoustic properties of the continua in the fMRI experiment after adjustment to participants' individual PSE's.

| Continuum | VOT (ms) | Duration (ms) | F0 glide ^b (Hz) | F0 glide ^b (Hz/s) | mean F0 (Hz) | HNR (dB) | Jitter (μs) | Shimmer (dB) |
|----------------------|-------------|------------------|-------------------------------|---------------------------------|-----------------|-------------|----------------|-----------------|
| Bear ↑ ↓ | 8.1 ± 0.8 | 359.7 ± 0.9 | 40.7 ± 2.0 | 25.2 ± 0.3 | 179.0 ± 0.6 | 11.5 ± 0.0 | 0.58 ± 0.00 | 1.00 ± 0.02 |
| | 11.2 ± 0.8 | 362.9 ± 0.9 | 40.3 ± 1.9 | 25.1 ± 0.3 | 179.1 ± 0.6 | 11.7 ± 0.1 | 0.58 ± 0.00 | 1.00 ± 0.01 |
| | 14.4 ± 0.8 | 366.1 ± 0.8 | 40.4 ± 2.1 | 25.2 ± 0.3 | 179.2 ± 0.6 | 11.8 ± 0.1 | 0.58 ± 0.00 | 1.00 ± 0.01 |
| | 17.5 ± 0.8 | 369.1 ± 0.8 | 40.2 ± 2.1 | 25.1 ± 0.3 | 179.6 ± 0.6 | 11.9 ± 0.0 | 0.59 ± 0.01 | 1.01 ± 0.01 |
| | 20.6 ± 0.8 | 372.0 ± 0.8 | 39.4 ± 2.0 | 25.3 ± 0.4 | 179.4 ± 0.6 | 12.0 ± 0.1 | 0.60 ± 0.01 | 1.02 ± 0.01 |
| Pear | 23.7 ± 0.8 | 375.0 ± 0.8 | 39.4 ± 2.0 | 24.9 ± 0.3 | 179.6 ± 0.6 | 12.0 ± 0.0 | 0.59 ± 0.01 | 1.07 ± 0.02 |
| | 26.8 ± 0.8 | 378.1 ± 0.9 | 38.3 ± 1.9 | 24.9 ± 0.3 | 179.5 ± 0.6 | 12.1 ± 0.0 | 0.60 ± 0.01 | 1.07 ± 0.02 |
| Average | 17.5 ± 2.5 | 369.0 ± 2.5 | 39.8 ± 0.3 | 25.1 ± 0.1 | 179.3 ± 0.1 | 11.9 ± 0.1 | 0.59 ± 0.00 | 1.02 ± 0.01 |
| Statement ↑ ↓ | 17.4 ± 0.8 | 369.3 ± 0.9 | -6.4 ± 1.6 | 15.5 ± 0.2 | 164.7 ± 0.5 | 13.4 ± 0.0 | 0.42 ± 0.01 | 1.09 ± 0.02 |
| | 17.4 ± 0.8 | 369.2 ± 0.9 | 6.9 ± 1.7 | 16.8 ± 0.3 | 168.6 ± 0.5 | 13.1 ± 0.1 | 0.45 ± 0.01 | 1.05 ± 0.02 |
| | 17.4 ± 0.8 | 369.1 ± 0.9 | 21.2 ± 1.9 | 19.6 ± 0.4 | 172.9 ± 0.5 | 12.6 ± 0.1 | 0.51 ± 0.01 | 0.99 ± 0.01 |
| | 17.4 ± 0.8 | 368.9 ± 0.9 | 36.6 ± 2.0 | 22.9 ± 0.5 | 177.6 ± 0.6 | 11.4 ± 0.1 | 0.57 ± 0.01 | 0.99 ± 0.01 |
| | 17.4 ± 0.8 | 368.7 ± 0.9 | 51.7 ± 2.1 | 26.7 ± 0.5 | 182.6 ± 0.6 | 10.8 ± 0.1 | 0.64 ± 0.01 | 1.03 ± 0.01 |
| | 17.4 ± 0.8 | 368.6 ± 0.9 | 68.9 ± 2.3 | 30.5 ± 0.5 | 187.9 ± 0.7 | 10.8 ± 0.1 | 0.70 ± 0.01 | 1.02 ± 0.02 |
| Question | 17.4 ± 0.8 | 368.4 ± 0.9 | 85.4 ± 2.3 | 34.6 ± 0.5 | 193.8 ± 0.8 | 10.2 ± 0.1 | 0.75 ± 0.01 | 0.98 ± 0.02 |
| Average | 17.4 ± 0.0 | 368.9 ± 0.1 | 37.8 ± 12.5 | 23.8 ± 2.7 | 178.3 ± 4.0 | 11.8 ± 0.5 | 0.58 ± 0.05 | 1.02 ± 0.01 |
| P-value ^a | .735 | .735 | .866 | .735 | .799 | .932 | .735 | .799 |

Values depict mean ± SEM. ^aWilcoxon signed rank tests were used to statistically compare acoustic properties between prosody and phoneme continua. ^bF0 glide was calculated by subtracting values at pitch onset from pitch offset; PSE, point of subjective equality; VOT, voice onset time; HNR, harmonics-to-noise ratio. All values were extracted using PRAAT 5.3.01 (<http://www.praat.org>) [S6].

Table S2 Related to Figure 3. ANOVAs comparing the slopes of the psychophysical identification functions obtained after rTMS or sham stimulation, split by Hemisphere and by Task.

| By HEMISPHERE ^a | | | | By TASK ^b | | | |
|----------------------------|--------------|--------------|-------------|----------------------|--------------|--------------|-------------|
| Effect | $F_{1,15}$ | P -value | η^2_p | Effect | $F_{1,30}$ | P -value | η^2_p |
| <i>Right hemisphere</i> | | | | <i>Prosody task</i> | | | |
| TASK | 0.402 | .535 | .026 | HEM | 0.149 | .702 | .005 |
| STIM | 5.199 | .038* | .257 | STIM | 6.808 | .014* | .185 |
| TASK × STIM | 7.328 | .016* | .328 | HEM × STIM | 4.252 | .048* | .124 |
| <i>Left hemisphere</i> | | | | <i>Phoneme task</i> | | | |
| TASK | 1.216 | .288 | .075 | HEM | 0.461 | .503 | .015 |
| STIM | 0.566 | .464 | .036 | STIM | 1.530 | .226 | .049 |
| TASK × STIM | 3.755 | .072 | .200 | HEM × STIM | 1.672 | .206 | .053 |

^aANOVAs with factors TASK (prosody/phoneme) and STIMULATION (rTMS/sham) calculated separately for the right and left hemisphere. ^bANOVAs with the factors HEMISPHERE and STIMULATION calculated separately for the prosody and the phoneme task. * P -value < .05.

Table S3 Related to Figure 3. Description of the participant groups who received rTMS of the left or the right premotor cortex.

| Group (Stimulation site) | Left hemisphere | Right hemisphere | t_{30} | P -value |
|------------------------------|-----------------|------------------|----------|------------|
| N | 16 | 16 | - | - |
| Age | 22.19 ± 0.98 | 24.00 ± 1.38 | 1.067 | .294 |
| Handedness (LQ) ^a | 70.86 ± 4.41 | 82.61 ± 5.52 | 1.664 | .107 |
| Musical training | 8.88 ± 1.50 | 7.25 ± 1.31 | 0.815 | .422 |
| Stimulation intensity | 52.69 ± 1.41 | 51.31 ± 1.09 | 0.770 | .447 |
| Slope Prosody sham | 0.38 ± 0.04 | 0.43 ± 0.03 | 1.153 | .258 |
| Slope Phoneme sham | 0.32 ± 0.02 | 0.37 ± 0.03 | 1.263 | .216 |

Values indicate means ± SEM. ^aLQ: lateralization quotient according to the Edinburgh Handedness Inventory [S7].

SUPPLEMENTAL EXPERIMENTAL PROCEDURES:

Ethics approval

The research protocols were approved by the ethics committee of the College of Science and Engineering at the University of Glasgow, UK (CSE00960; CSE01245). All participants gave written consent and were paid £6 / hour.

Stimulus preparation

To study how the prosodic form of an utterance changes its communicative meaning, we chose to focus on declarative vs. interrogative mode, i.e. statements and questions. These are prime categories of linguistic prosody that are expressed in conventionalized pitch contours to fulfill distinct communicative functions, e.g. naming an object (statement) vs. asking for it (question). We furthermore chose to use single-word utterances as this reduces the influence of focus and grammatical form on prosody (although it does not fully exclude that the single-word prosody may implicitly activate the syntactic structure typically associated with statements and questions). The interaction between left-hemispheric language streams and prosody perception – that is well-known from studies with sentence-level material [S8, S9] – was thus kept to a minimum and allowed us to identify the core network of prosodic contour perception. Single-word utterances occur naturally in every-day communications and from early on in ontogeny [S10]. Importantly, the pitch contrast between statement and question is primarily coded by the fundamental frequency in the *last syllable* of an utterance [S11]. Hence, single (potentially sentence-final) nouns or syllables can be reliably perceived as statement or question.

Recordings (16 kHz, 16 bit, mono) from 2 male and 2 female adult Scottish speakers uttering the word “bear” or “pear” with either falling (statement) or rising pitch contour (question) were used to construct continua of prosody (statement → question; Audio clip S1) and word-initial phoneme (/b/ → /p/; Audio clip S2) by means of audio morphing (Figure 1A). Separately for each speaker, 21 × 21 continua covering the space between the 4 original utterances (i.e. “Bear.”, “Bear?”, “Pear.”, “Pear?”) in 5% steps were created using STRAIGHT [S12] in MatlabR2009 (The MathWorks, Inc., Natick, MA, USA). Temporal anchor points for morphing were set to the onsets and offsets of the phonation and the word-final “r”. Spectro-temporal anchors were set to the 1st to 4th formants at onsets and offsets of formant transition as well as the pitch rise/drop in the question and statement. Based on a logarithmic interpolation of the anchor templates and spectrogram, morphed stimuli were then resynthesized in steps of 5%. The stop-consonants /b/ and /p/ were clipped prior to morphing and subsequently spliced back in by means of PRAAT 5.3.01 (<http://www.praat.org>) to create the phoneme continua from “bear” to “pear”. The first token in the continuum had a VOT of 0 ms; VOT of subsequent tokens was increased by inserting 3-ms increments of

aspiration noise from the corresponding “pear” between the release-burst and the phonation onset of the “bear”, resulting in a maximal VOT of 60 ms in the last token of the continuum.

Stimulus choice based on individual category boundaries

To account for perceptual variability between listeners, stimuli in the fMRI and TMS experiment were chosen individually for each participant from the 21×21 continua, adapted to his/her individual point of subjective equality (PSE) for prosody and phoneme. To identify the individual PSE, each participant took part in a pre-test during which he/she was presented with 49 stimuli representing all possible combinations of stimuli with 5/95% – 20/80% – 35/65% – 50/50% – 65/35% – 80/20% – 95/5% question/statement and “bear”/“pear” ratio. The test consisted of two blocks with the identical stimulus material, requiring a two-alternative forced choice (2-AFC) judgement of prosody or phoneme, respectively, in counterbalanced order across subjects. Each block contained 7 continua à 7 morph steps for each speaker, i.e. 28 continua allowing for a reliable estimation of the psychometric function. Stimuli within each continuum were presented in random order with a variable inter-stimulus interval (ISI) of 2.5 ± 0.5 seconds. Participants gave their response by pressing one of two buttons with their right index or middle finger. Key assignment was counterbalanced across participants. Stimuli were presented with the Psychtoolbox3 via DT 770 PRO, 250 Ohms headphones (beyerdynamic, Heilbronn, Germany). The test lasted approximately 20 minutes.

Based on these data, 4 continua of 7 morph steps, centred on the participant’s individual PSE, were chosen from the stimulus space of each speaker: 2 prosody continua (“Bear. → “Bear?”, “Pear. → “Pear?”), and 2 phoneme continua (“Bear.” → “Pear.”, “Bear?” → “Pear?”; see Figure 1A), resulting in 16 continua in total for presentation during the fMRI or TMS experiment.

Based on piloting with 14 native English listeners (mean age \pm SEM: 23.21 ± 1.54 years, 9 female, none of them participated in the fMRI or TMS experiment), the step size within the stimulus continua was set to 10% in the prosody and 5% in the phoneme continua to obtain parity in task demand (prosody: stepwise increase of offset minus onset F0 by an average of 15.3 Hz; phoneme: stepwise increase of voice onset time by 3 ms). Sound characteristics of the final set of stimuli used in the fMRI session are displayed in Table S1. Note that the average sound properties of the prosody and phoneme continua were virtually identical (P ’s $> .633$; Mann-Whitney U -tests). Therefore, any difference in brain activation between the two tasks cannot be due to physical differences in the stimulus material.

Experiment 1 – Dorsal and ventral pathways for prosody

Participants

Data were obtained from 23 healthy native English speakers (10 females, mean age \pm SEM: 25.7 ± 1.4 years). This group size is sufficient to detect large effects as estimated with GPower3 (<http://www.psych.uni-duesseldorf.de/aap/projects/gpower>). All participants were right-handed according to the Edinburgh Handedness Inventory [S7] (mean laterality quotient \pm SEM: 76.3 ± 3.6) and had normal hearing as assessed with the absolute threshold subtest of the STAIRCASE toolbox (<http://www.psy.unipd.it/~grassi/psypsy.html>) at 0.125, 0.25, 0.5, 1, 2, 4 and 8 kHz following a simple up-down procedure. Furthermore, using the same toolbox, participants showed an average pitch discrimination threshold of 3.12 ± 0.51 Hz (mean \pm SEM) for complex tones and 150 Hz / 500 Hz pure tones, ensuring that they would be principally able to perceive the pitch glide differences of 15.3 Hz in the prosody continua (see above). Auditory stimuli were delivered in a soundproof cabin via DT 770 PRO, 250 Ohms headphones (beyerdynamic, Heilbronn, Germany) calibrated with a TENMA 72-6635 digital sound level meter.

Procedure

Participants were presented with 16 stimulus blocks, each containing 7 repetitions of the 7 morph steps of one continuum (Figure 1B). Stimulus onset asynchrony (SOA) was set to 2.4 seconds. Stimuli within each block were presented following a type-1 index-1 sequence [S13] to control for carry-over effects in fast event-related designs, i.e. each stimulus was preceded by itself and every other within-continuum stimulus in a balanced manner. Stimulus blocks were separated by 19.2 seconds of silence and each block started with a reminder about the current task and key assignment. Stimuli were presented using Psychtoolbox3 and delivered through in-ear headphones (NNL – NordicNeuroLab, Inc., Bergen, Norway). To acquaint participants with the procedure, they were presented with a 10 minutes version of the experiment, including pre-recorded scanner noise. Overall duration of the session was ~45 minutes.

MR image acquisition

Functional MRI was performed in a 3T Siemens TIM Trio scanner at the Centre for Cognitive Neuroimaging (CCNi) in Glasgow, UK. A series of 1082 T_2^* -weighted functional images were acquired with a single-shot echo-planar gradient-echo (EPI) pulse sequence (TR = 2400 ms, TE = 30 ms, iPAT = 2, the first 4 s were discarded to allow for steady state magnetization). Thirty-one ascending interleaved slices ($3 \times 3 \times 3$ mm³ voxel size, 15% interslice gap) aligned to the Sylvian Fissure were collected with a 32-channel head coil. The field of view (FOV) was 210×210 mm² with an in-plane resolution of 70×70 pixels and a flip angle of 77°.

After each scan (TA = 1700 ms), scanner noise was interrupted for 700 ms. Stimuli were centred in these silent gaps between scans.

After the functional scans, a high-resolution T_1 -weighted dataset ($1 \times 1 \times 1 \text{ mm}^3$ voxel size) was acquired for anatomical co-registration using a magnetization-prepared rapid acquisition gradient echo (MPRAGE) sequence (TR = 2300 ms, TE = 2.96 ms, 192 axial slices, FOV = $256 \times 256 \text{ mm}^2$, matrix size = 256×256 pixels, flip angle = 9°).

Finally, a diffusion-weighted data set was acquired with a twice-refocused spin echo EPI sequence (TR = 12000 ms, TE = 100 ms, 88 axial slices without gap, FOV = 220 mm, matrix size = 128×128 pixels, iPAT = 2) with a voxel size of $1.72 \times 1.72 \times 1.7 \text{ mm}^3$. Diffusion-weighting was isotropically distributed along 60 diffusion-encoding gradient directions with a b-value of 1000 s/mm^2 . Acquisition of diffusion-weighted images was interspersed with seven images without diffusion-weighting (b_0) serving as anatomical reference for offline motion correction. Total duration of the diffusion MRI scan was about 15 minutes.

Functional MRI data analysis

All functional data were analyzed using SPM8 (Wellcome Trust Centre for Neuroimaging, University College, London, UK). Preprocessing included spatial realignment, co-registration of the functional and anatomical data, spatial normalization into the MNI stereotactic space (using the unified segmentation-based procedure), and spatial smoothing using an isotropic 8 mm^3 Gaussian kernel. Low frequency drifts were eliminated by means of a temporal high-pass filter with a cutoff of 200 s. Two onset regressors were modeled using boxcar functions convolved with a canonical hemodynamic response function (HRF): prosody task and phoneme task. Onsets of task instructions and motion parameters were included as regressors of no interest in the single-subject GLMs. The linear contrast of prosody task > phoneme task was calculated for each participant and submitted to a one-sample *t*-test during random effects group analysis. For statistical thresholding, a Monte Carlo simulation implemented in Matlab (10000 iterations, no volume mask) suggested a cluster extent threshold of at least 19 resampled voxels and a voxel-level uncorrected *P*-value of .001 to ensure whole-volume type I error probability smaller than .01 (code available at <http://www2.bc.edu/~slotnics/scripts.htm>). All coordinates refer to standard MNI space. Lateralization indices were calculated as $LI = (L-R)/(L+R)$, with R and L representing the effect sizes of the right-hemisphere clusters' peak voxel and its left-hemisphere homologue (negative x-coordinate), respectively. The analysis was based on anatomical rather than functional homologues, given that no functional homologues could be identified, e.g., for pSTS and aSTS, not even at a threshold as lenient as $P_{\text{voxel}} < .05$ uncorrected. Due to the anatomical symmetry of the MNI template brain, all flipped coordinates were located in the same

anatomical structure in the left and right hemisphere. We assume that these regions are the best possible choice of candidates that support potentially similar functions.

Diffusion MRI data analysis

Diffusion MRI data were analyzed using LIPSIA (Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany), FSL (FMRIB, University of Oxford, UK, www.fmrib.ox.ac.uk/fsl), and SPM8. During preprocessing of the diffusion weighted images, T_1 -weighted structural scans were skull-stripped and aligned to AC-PC in Talairach space. Motion correction parameters for the 60 diffusion-weighted images were computed based on the 7 (b0) reference images distributed over the entire sequence using rigid-body transformations implemented in FSL. Interpolated motion correction parameters as well as a global registration to the T_1 -anatomy were applied to all 60 volumes that were, thus, resampled to an isotropic resolution of 1 mm. The gradient direction for each volume was corrected using the rotation parameters.

After preprocessing, anatomical connectivity between the fMRI activation maxima in right pSTS, aSTS, PMC, and IFG was obtained with probabilistic fiber tracking using FDT (FMRIB's Diffusion Toolbox) in FSL. Seed and target regions were obtained by morphing the four relevant functional group activation clusters to each participant's diffusion MRI space with 1 x 1 x 1 mm resolution. Then, fiber orientation was estimated in each voxel by means of the software module BEDPOSTX using a crossing fiber model with up to two directions per voxel. Finally, probabilistic tractography between pSTS–aSTS, pSTS–PMC, pSTS–IFG, aSTS–IFG, and PMC–IFG was computed bidirectionally using the standard parameters of the PROBTRACKX module, i.e., 5000 streamlines per seed region voxel, a curvature threshold of 0.2, a step length of 0.5, and a maximum number of steps of 2000.

To obtain an overlay of fiber tracts, the tractography images of each participant were first screened for unreliable connections. A connection was considered unreliable, if the number of fibers that reached the target (in both directions) did not exceed one third of the number of voxels contained in the pair of seed and target regions (i.e., thresholds were the following: pSTS–aSTS: 3952, pSTS–PMC: 4643, pSTS–IFG: 3722, aSTS–IFG: 2693, PMC–IFG: 3384 fibers). This threshold was determined empirically based on the data properties and the consistency of the pathways across participants. It was aimed at reducing false-positive connections while staying sensitive to true connections. The pSTS–IFG and pSTS–PMC connections were considered unreliable in four and five participants, respectively (for individual data, see Figure S2). These tractography images were set to zero. All participants showed reliable pSTS–aSTS and PMC–IFG connections. After that, all tractography images were cleaned from random connections beyond the fiber tract of interest (binarization threshold at 5% of the image's maximum intensity value), normalized to the template brain in MNI standard

space, slightly smoothed (Gaussian filter with 0.5mm FWHM) and corrected for filter-induced blurring at the rim (binarization threshold at 0.0001). Plots were created with in-house software (brainGL).

Behavioral data analysis

For each of the seven morph steps, the proportion of “question” and “pear” responses was calculated. Then, non-linear least-squares regressions were used in each participant to fit psychometric curves to the categorization data (separately for the prosody and the phoneme task) using a cumulative logistic function of the following form: $y = a + \frac{b}{1 + \exp(\frac{c-x}{s})}$ with a and b representing the lower and upper asymptotes, c the center of symmetry of the curve (i.e. the PSE; point of subjective equality), and s the slope of the curve at c . To avoid overfitting, a and b were set to the lower and upper asymptote obtained in an unconstrained fit on the averaged data. The slopes s in the prosody and the phoneme task were taken as indices for participants’ categorization performance (and thus, task difficulty) [S14] and were compared by means of a t -test for paired samples (two-tailed) using SPSS (PASW Statistics 18). Data met the criterion of normal distribution as assessed with Komolgorov-Smirnov tests (P ’s > .495).

Experiment 2 – Motor-involvement in prosody perception

Participants

Thirty-two right-handed volunteers with normal hearing and no personal/family history of neurological disorder or seizure gave written consent to participate (mean age \pm SEM: 23 \pm 0.9 years; mean laterality quotient \pm SEM: 76.7 \pm 3.6 [S7]). None of them had participated in Experiment 1. They were split into two groups of 16 participants (8 females in each) in whom we targeted either the left or the right hemisphere with TMS. The two groups were carefully matched in age, handedness, years of musical training (to rule out advantages in pitch perception), stimulator output during rTMS, and performance in the prosody and phoneme task during sham stimulation (t -tests for independent samples: P ’s > .107; see Supplemental Table S3).

Procedure

First, participants’ resting hand motor threshold was assessed, defined as the lowest stimulator output that evoked at least 5 of 10 visible motor responses in the first dorsal interosseus muscle and thumb while the participant’s arm was relaxed [S15]. Thresholds ranged from 43 – 61% (mean \pm SEM: 52.0 \pm 0.9%) of the maximum stimulator output and did not differ between groups (right: 51.31 \pm 1.09; left: 52.69 \pm 1.41; $t_{30} = .770$, $P > .446$). Thereafter, participants completed one prosody and one phoneme block after rTMS and sham stimulation of right or left PMC, in counterbalanced order (Figure 3A). Stimuli were the same as in

Experiment 1 and were presented with Psychtoolbox3 via DT 770 PRO, 250 Ohms headphones (beyerdynamic, Heilbronn, Germany). Participants gave their response with the index and middle finger of the hand ipsilateral to the stimulation site. Each of the four task blocks contained 8 continua (two for each of the four speakers, no breaks) with a total of 176 randomized trials and a block duration of 7 minutes. SOA was set to 2.4 seconds. To maximize detectability of slope changes around the PSE, stimuli in the ambiguous centre of the continua (i.e. morph steps 3-5) were presented more often (32 times, each) than steps 2 and 6 (24 times) or steps 1 and 7 (16 times). rTMS and sham stimulation were separated by a 30 minutes break.

rTMS

Stimulation intensity was set to 90% of each participant's resting motor threshold. 900 biphasic TMS pulses were generated by a Magstim Rapid² (Magstim Company Limited, Whitland, UK) and delivered through a 70 mm figure-of-eight coil. To avoid overheating, the coil was replaced after 450 pulses. During rTMS, the coil was placed tangentially to the skull with the handle pointing posterior at 45° to the intra-hemispheric fissure. During sham stimulation, the coil was angled at 90° to the skull. Frameless stereotaxy (Brainsight 2, Rogue Research Inc., Montréal, Quebec, Canada) was used for coil positioning over the stimulation site. The coordinates of each participant's stimulation site were calculated by converting the MNI coordinates of the right-premotor functional cluster obtained in the fMRI experiment or its flipped coordinate ($x = +/-54$, $y = 2$, $z = 46$) to the participant's individual structural space using the inverse-normalization parameters in SPM 8. Accurate coil position was monitored throughout stimulation to keep movement induced deviations below 3 mm.

Data analysis

To show that the right PMC is necessary for prosody perception, the slopes of the logistic curves (obtained in the same way as in Experiment 1) were compared by means of two 2-way ANOVAs. First, an ANOVA with the within-subject factors STIMULATION (rTMS vs. sham) and TASK (prosody vs. phoneme) was used to test whether rTMS of right PMC affected prosody perception more strongly than phoneme perception (i.e. comparison with a control task). Second, an ANOVA with the factors STIMULATION (rTMS vs. sham; within-subject) and HEMISPHERE (left vs. right; between-subjects) was used to test whether rTMS of right PMC affected prosody perception more strongly than rTMS of left PMC (i.e. comparison with a "control site"). For the sake of completeness, we also report the two complementary ANOVAs that focus on the left hemisphere and the phoneme control task (Table S2). No 3-way ANOVA was computed because the triple interaction does not discriminate between the hypothesized prosody effects in the right hemisphere and potential phoneme effects in the left hemisphere [S16, S17].

SUPPLEMENTAL DATA:

Several studies have shown the involvement of left motor/premotor cortex in phoneme perception. Although it is a matter of debate in how far the motor system is essential for language comprehension [S18–S20], it is well established that (pre)motor cortex is activated during the perception of speech sounds [S21–S26] and contributes to performance in phoneme tasks [S16, S17, S19, S27]. These effects can occur pre-attentively [S28, S29] but are particularly strong when stimuli are ambiguous, embedded in noise [S30], or difficult to produce [S31]. While the present study was not primarily designed to investigate phoneme perception, the use of a phoneme task as control condition naturally raises the question whether some of the previous results would surface in the present data. We first explored this in the fMRI data. Assuming that the ambiguity of the speech signal increases (pre)motor cortex involvement in phoneme perception [S30], we contrasted ambiguous phoneme trials (morph step 4) with clear phoneme trials (morph steps 1 and 7). This analysis revealed one single cluster for ambiguous > clear phonemes that was located in the left PMC ($x = -48$, $y = 8$, $z = 28$; $Z = 4.47$; 74 voxels; $P_{\text{voxel}} < .001$; $P_{\text{cluster}} < .05$ FWE-corrected), close to the lip-area [S32]. The parameter estimates extracted at the peak voxel for every morph step exhibited an inverted U-shape in line with the idea of (pre)motor involvement as a function of phoneme clarity [S30].

Note that the phoneme cluster was located more ventrally than the stimulation site in our TMS experiment ($x = -45$, $y = 5$, $z = 40$) that was designed to interfere with prosody perception. This precludes an interpretation of the TMS data in terms of phoneme perception.

SUPPLEMENTAL REFERENCES:

- S1. Catani, M., and Thiebaut de Schotten, M. (2008). A diffusion tensor imaging tractography atlas for virtual in vivo dissections. *Cortex* 44, 1105–1132.
- S2. Makris, N., and Pandya, D. N. (2009). The extreme capsule in humans and rethinking of the language circuitry. *Brain Struct Funct* 213, 343–358.
- S3. Thiebaut de Schotten, M., Dell’Acqua, F., Valabrègue, R., and Catani, M. (2012). Monkey to human comparative anatomy of the frontal lobe association tracts. *Cortex* 48, 82–96.
- S4. Anwender, A., Tittgemeyer, M., von Cramon, D. Y., Friederici, A. D., and Knösche, T. R. (2007). Connectivity-based parcellation of Broca’s area. *Cereb. Cortex* 17, 816–825.
- S5. Wakana, S., Caprihan, A., Panzenboeck, M. M., Fallon, J. H., Perry, M., Gollub, R. L., Hua, K., Zhang, J., Jiang, H., Dubey, P., et al. (2007). Reproducibility of quantitative tractography methods applied to cerebral white matter. *Neuroimage* 36, 630–644.
- S6. Boersma, P., and Weenink, D. (2006). Praat, a system for doing phonetics by computer (Version 5.3.03) [Computer software]. Available at <http://www.praat.org>.
- S7. Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113.
- S8. Sammler, D., Kotz, S. A., Eckstein, K., Ott, D. V. M., and Friederici, A. D. (2010). Prosody meets syntax: The role of the corpus callosum. *Brain* 133, 2643–2655.
- S9. Friederici, A. D., von Cramon, D. Y., and Kotz, S. A. (2007). Role of the corpus callosum in speech comprehension: Interfacing syntax and prosody. *Neuron* 53, 135–145.
- S10. Dore, J. (1975). Holophrases, speech acts and language universals. *J Child Lang* 2, 21–40.
- S11. Bolinger, D. (1986). *Intonation and its parts - Melody in spoken English* (Stanford: Stanford University Press).
- S12. Kawahara, H. (2006). STRAIGHT, exploitation of the other aspect of vocoder: perceptually isomorphic decomposition of speech sounds. *Acoust. Sci. Technol.* 27, 349–353.
- S13. Nonyane, B. A., and Theobald, C. M. (2007). Design sequences for sensory studies: achieving balance for carry-over and position effects. *Br. J. Math. Stat. Psychol.* 60, 339–349.
- S14. Kingdom, F. A. A., and Prins, N. (2010). *Psychophysics: A practical introduction* (London: Academic Press).
- S15. Varnava, A., Stokes, M. G., and Chambers, C. D. (2011). Reliability of the “observation of movement” method for determining motor threshold using transcranial magnetic stimulation. *J. Neurosci. Methods* 201, 327–332.
- S16. D’Ausilio, A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C., and Fadiga, L. (2009). The motor somatotopy of speech perception. *Curr. Biol.* 19, 381–385.

- S17. Möttönen, R., and Watkins, K. E. (2009). Motor representations of articulators contribute to categorical perception of speech sounds. *J. Neurosci.* 29, 9819–9825.
- S18. Pulvermüller, F., and Fadiga, L. (2010). Active perception: sensorimotor circuits as a cortical basis for language. *Nat. Rev. Neurosci.* 11, 351–360.
- S19. Krieger-Redwood, K., Gaskell, M. G., Lindsay, S., and Jefferies, E. (2013). The selective role of premotor cortex in speech perception: a contribution to phoneme judgements but not speech comprehension. *J. Cogn. Neurosci.* 25, 2179–2188.
- S20. Hickok, G., Houde, J. F., and Rong, F. (2011). Sensorimotor integration in speech processing: computational basis and neural organization. *Neuron* 69, 407–422.
- S21. Wilson, S. M., Saygin, A. P., Sereno, M. I., and Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nat. Neurosci.* 7, 701–702.
- S22. Pulvermüller, F., Huss, M., Kherif, F., Martin, F. M. D. P., Hauk, O., and Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proc. Natl. Acad. Sci. U. S. A.* 103, 7865–7870.
- S23. Yuen, I., Davis, M. H., Brysbaert, M., and Rastle, K. (2010). Activation of articulatory information in speech perception. *Proc. Natl. Acad. Sci. U. S. A.* 107, 592–597.
- S24. Fadiga, L., Craighero, L., Buccino, G., and Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *Eur. J. Neurosci.* 15, 399–402.
- S25. Murakami, T., Restle, J., and Ziemann, U. (2011). Observation-execution matching and action inhibition in human primary motor cortex during viewing of speech-related lip movements or listening to speech. *Neuropsychologia* 49, 2045–2054.
- S26. Roy, A. C., Craighero, L., Fabbri-Destro, M., and Fadiga, L. (2008). Phonological and lexical motor facilitation during speech listening: a transcranial magnetic stimulation study. *J. Physiol. Paris* 102, 101–105.
- S27. Meister, I. G., Wilson, S. M., Deblieck, C., Wu, A. D., and Iacoboni, M. (2007). The essential role of premotor cortex in speech perception. *Curr. Biol.* 17, 1692–1696.
- S28. Chevillet, M. A., Jiang, X., Rauschecker, J. P., and Riesenhuber, M. (2013). Automatic phoneme category selectivity in the dorsal auditory stream. *J. Neurosci.* 33, 5208–5215.
- S29. Möttönen, R., Dutton, R., and Watkins, K. E. (2013). Auditory-motor processing of speech sounds. *Cereb. Cortex* 23, 1190–1197.
- S30. D’Ausilio, A., Bufalari, I., Salmas, P., and Fadiga, L. (2012). The role of the motor system in discriminating normal and degraded speech sounds. *Cortex* 48, 882–887.
- S31. Bartoli, E., D’Ausilio, A., Berry, J., Badino, L., Bever, T., and Fadiga, L. (2015). Listener-speaker perceived distance predicts the degree of motor contribution to speech perception. *Cereb. Cortex* 25, 281–288.
- S32. Brown, S., Ngan, E., and Liotti, M. (2008). A larynx area in the human motor cortex. *Cereb. Cortex* 18, 837–845.

2.2 Study II – Neural networks for speech prosody

Sammler, D., Cunitz, K., Anwander, A., Gierhan, S. M. E., Adermann, J., Meixensberger, J., Friederici, A. D. (2018). White matter pathways for prosodic structure building: A case study. *Brain and Language*, 183, 1-10.*

[Author contributions by D.S.: designed research with J.M. and A.D.F.; received materials from S.M.E.G. and J.A.; performed research with K.C., analysed data with A.A.; wrote the paper, incorporating edits by co-authors]

*Note: At the time of submission of the habilitation thesis, this study was under review. The manuscript version of the study was replaced by the final published version here. Reprinted with permission from Elsevier.



White matter pathways for prosodic structure building: A case study



Daniela Sammler^{a,b,*}, Katrin Cunitz^{b,c}, Sarah M.E. Gierhan^{b,d}, Alfred Anwander^b, Jens Adermann^e, Jürgen Meixensberger^e, Angela D. Friederici^{b,d}

^a Otto Hahn Group "Neural Bases of Intonation in Speech and Music", Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstraße 1a, 04103 Leipzig, Germany

^b Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstraße 1a, 04103 Leipzig, Germany

^c Department of Child and Adolescent Psychiatry and Psychotherapy, University Hospital Ulm, Steinhövelstraße 5, 89075 Ulm, Germany

^d Berlin School of Mind and Brain, Humboldt University Berlin, Unter den Linden 6, 10099 Berlin, Germany

^e University Hospital Leipzig, Clinic and Policlinic for Neurosurgery, Liebigstraße 20, 04103 Leipzig, Germany

ARTICLE INFO

Keywords:
Language
Prosody
Voice
Pitch
DTI
Auditory pathways
Arcuate fascicle
Dorsal stream
Corpus callosum

ABSTRACT

The relevance of left dorsal and ventral fiber pathways for syntactic and semantic comprehension is well established, while pathways for prosody are little explored. The present study examined linguistic prosodic structure building in a patient whose right arcuate/superior longitudinal fascicles and posterior corpus callosum were transiently compromised by a vasogenic peritumoral edema. Compared to ten matched healthy controls, the patient's ability to detect irregular prosodic structure significantly improved between pre- and post-surgical assessment. This recovery was accompanied by an increase in average fractional anisotropy (FA) in right dorsal and posterior transcallosal fiber tracts. Neither general cognitive abilities nor (non-prosodic) syntactic comprehension nor FA in right ventral and left dorsal fiber tracts showed a similar pre-post increase. Together, these findings suggest a contribution of right dorsal and inter-hemispheric pathways to prosody perception, including the right-dorsal tracking and structuring of prosodic pitch contours that is transcallosally informed by concurrent syntactic information.

1. Introduction

White-matter fiber bundles connecting left fronto-temporal (and parietal) 'language areas' have become a centerpiece of modern language models (Friederici, 2011) and their divide into functionally specialized dorsal and ventral routes is largely undisputed (Hickok & Poeppel, 2007; Rauschecker & Scott, 2009). However, the established roles of left dorsal and ventral fiber tracts in speech production and semantic comprehension (Fridriksson et al., 2018; Kümmerer et al., 2013; Saur et al., 2008), as well as syntactic parsing (Friederici, 2012; Griffiths, Marslen-Wilson, Stamatakis, & Tyler, 2013; Wilson et al., 2011) ignore one important component of spoken language: Speech prosody, the rhythmic-melodic variations in speech that serve linguistic functions¹ (Cutler, Dahan, & Van Donselaar, 1997). The notable involvement of right-hemispheric fronto-temporal brain areas in linguistic prosodic processing (for reviews, see Baum & Pell, 1999; Belyk & Brown, 2014; Paulmann, 2016; Witteman, van Ijzendoorn, van de Velde, van Heuven, & Schiller, 2011) calls for reflection upon the contribution of right-hemispheric (Sammler, Grosbras, Anwander,

Bestelmeyer, & Belin, 2015) as well as inter-hemispheric pathways to natural language comprehension (Friederici & Alter, 2004). Here, we present a case study that lends evidence for the functional necessity of right dorsal and transcallosal pathways in linguistic prosodic structure building.

Speakers naturally vary prosodic features such as intonation and rhythm in their utterances to package information into meaningful units and to accentuate thematically relevant words (Cole, 2015; Cutler et al., 1997; Wagner & Watson, 2010). Indeed, we do well in relying on prosodic cues in ambiguous sentences like "Wave at the girl with the hat." to understand at which girl we should wave and in which way (Lehiste, 1973; Snedeker & Trueswell, 2003). This example is one of many to illustrate how the prosodic structure of an utterance, i.e., its organization into smaller phonological or intonational phrases (Selkirk, 1996), can assist language comprehension: The systematic alignment of prosodic phrase boundaries with syntactic and semantic structure (Cooper & Paccia-Cooper, 1980; Selkirk, 1984; Watson & Gibson, 2004) allows listeners to use prosody in their syntactic and semantic interpretation, and vice versa (Buxó-Lugo & Watson, 2016; Cole, Mo, & Baek,

* Corresponding author at: Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstraße 1a, 04103 Leipzig, Germany.

E-mail address: sammler@cbs.mpg.de (D. Sammler).

¹ Prosody also conveys paralinguistic information about speakers' emotions, attitudes, and intentions (Hellbernd & Sammler, 2016; Scherer, 1986) which will not be addressed in the present study.

<https://doi.org/10.1016/j.bandl.2018.05.001>

Received 3 May 2017; Received in revised form 14 March 2018; Accepted 3 May 2018
Available online 11 May 2018

0093-934X/ © 2018 Elsevier Inc. All rights reserved.

2010). Changes in pitch contour, pre-boundary lengthening and pauses are amongst the most important acoustic cues that signal prosodic boundaries (Ladd, 2008; Pierrehumbert & Hirshberg, 1990) and constrain parsing possibilities.

The tracking of these cues has often been associated with the right hemisphere, in line with cue-dependent models of auditory speech perception (Friederici & Alter, 2004; McGettigan & Scott, 2012; Poeppel, 2003; Zatorre, Belin, & Penhune, 2002). These models argue for a relative processing benefit of right auditory cortices for pitch and spectral information (Jamison, Watkins, Bishop, & Matthews, 2006; Johnsrude, Penhune, & Zatorre, 2000; Obleser, Eisner, & Kotz, 2008; Schönwiesner, Rübsamen, & von Cramon, 2005; Zatorre et al., 2002) that unfolds over extended timescales (Giraud et al., 2007; Poeppel, 2003). As a consequence, the right hemisphere may optimally track suprasegmental prosodic features and complement left-hemispheric syntactic and semantic processes, as proposed in the *Dynamic Dual Pathway Model* of Friederici and Alter (2004). In keeping with this hemispheric division of labor, fMRI and dichotic listening studies reported predominant right fronto-temporal activations (Kyong et al., 2014; Meyer, Alter, Friederici, Lohmann, & von Cramon, 2002; Meyer, Steinhauer, Alter, Friederici, & von Cramon, 2004; Plante, Creusere, & Sabin, 2002) and a left ear advantage (i.e., right hemisphere involvement; Blumstein & Cooper, 1974; Shipley-Brown, Dingwall, Berlin, Yeni-Komshian, & Gordon-Salant, 1988) when listening to filtered or degraded speech with high demands on prosodic processing. Likewise, explicit attention to prosodic pitch contours in statements and questions (compared to processing of phonemes and lexical meaning) induced right-lateralized activity in fronto-temporal regions (Kreitewolf, Friederici, & von Kriegstein, 2014; Sammler et al., 2015). The right-lateralization of prosody is less clear-cut in studies with natural language material (e.g., Perkins, Baran, & Gandour, 1996; Tang, Hamilton, & Chang, 2017), prosody production (Kellmeyer et al., 2013; Peschke, Ziegler, Eisenberger, & Baumgaertner, 2012), and tasks that go beyond the processing of low-level acoustic-prosodic cues such as pitch contour (for reviews showing bilateral involvement, see Baum & Pell, 1999; Belyk & Brown, 2014; Paulmann, 2016; Witteman et al., 2011). This indicates the inevitable interaction of prosodic information with concurrent syntactic (den Ouden, Dickey, Anderson, & Christianson, 2016) or lexical-semantic processes (Domahs, Klein, Huber, & Domahs, 2013; Gandour et al., 2004; van Lancker, 1980) that are hard to separate during natural language comprehension.

The present study focuses on sentence-level prosodic structure building, i.e., the gradual emergence of a (hierarchical) representation of prosodic constituency that aligns with syntactic structure. As outlined above and implied by previous psycholinguistic research, the prosodic parser most likely draws both on acoustic markers for prosodic boundaries (Ladd, 2008; Pierrehumbert & Hirshberg, 1990; Snedeker & Trueswell, 2003) as well as concurrent syntactic structure (Buxó-Lugo & Watson, 2016; Cole et al., 2010) to build prosodic representations. At the neural level, this implies involvement of both right-hemispheric fronto-temporal networks that track relevant prosodic features over time as well as inter-hemispheric exchange to map syntactic and prosodic structure onto each other (Friederici & Alter, 2004).

This assumption naturally raises the question *how* information is transferred between relevant brain areas. Syntactic structure building in the left hemisphere is known to involve ventral fronto-temporal connections via the inferior fronto-occipital (IFOF) and uncinate fascicles (UF) for simple syntactic parsing, while dorsal connections via the arcuate and superior longitudinal fascicles (AF/SLF) support parsing of complex syntactic structures (Friederici, 2012; Griffiths et al., 2013) (for reviews, see Friederici, 2011; Gierhan, 2013b). Correspondingly, damage to left dorsal fiber tracts (Meyer, Cunitz, Obleser, & Friederici, 2014; Wilson et al., 2011) or their developmental immaturity (Skeide, Brauer, & Friederici, 2016) coincide with reduced comprehension of syntactically complex sentences.

Recently, we demonstrated a similar multi-pathway architecture in

the right hemisphere for the perception of prosodic pitch contours in statements and questions (Sammler et al., 2015). This finding was remarkable because the relevance of right-hemispheric and particularly right *dorsal* tracts in speech and language has been questioned until very recently (Hickok, 2012). Indeed, direct right dorsal fronto-temporal connections were often found to be anatomically weaker than their left-hemispheric counterparts (Fernández-Miranda et al., 2015; Glasser & Rilling, 2008; Parker et al., 2005; Powell et al., 2006; Thiebaut de Schotten, Ffytche, et al., 2011) and have been studied nearly exclusively in the context of atypical language lateralization (Duffau, Leroy, & Gatignol, 2008; Vassal, Le Bars, Moritz-Gasser, Menjot, & Duffau, 2010), e.g., during aphasia rehabilitation (Forkel et al., 2014; Schlaug, Marchina, & Norton, 2009). What has remained unexplored so far is the potential contribution of right dorsal tracts to the processing of suprasegmental prosodic information in speech. Our data on statement and question discrimination lend initial evidence for that, albeit only for single words (Sammler et al., 2015); for converging evidence in emotional prosody perception, see Frühholz, Gschwind, & Grandjean, 2015; Glasser & Rilling, 2008). It seems plausible, though, that the capacity of (right) AF/SLF and temporal-premotor loops to constantly monitor sound and pitch (Guenther & Vladusich, 2012; Houde & Chang, 2015; Zarate, 2013) may benefit the acoustic detection of prosodic boundaries in sentences. A yet bolder proposal that awaits testing is the potential involvement of right dorsal fronto-temporal tracts in more advanced prosodic structuring, beyond basic pitch tracking (Bornkessel-Schlesewsky & Schlesewsky, 2013).

The interaction between the lateralized syntax and prosody streams requires a dynamic exchange between the two hemispheres (Friederici & Alter, 2004; Steinmann & Mulert, 2012). Several studies suggest that syntax-prosody alignment hinges particularly on the cross-talk between the temporal lobes via commissural fibers that cross through the posterior third of the corpus callosum (CC; Friederici, von Cramon, & Kotz, 2007; Sammler, Kotz, Eckstein, Ott, & Friederici, 2010; for the anatomy of CC, see Hofer & Frahm, 2006; Huang et al., 2005). Accordingly, patients with permanent lesions in the posterior CC no longer processed prosodic (or syntactic) irregularities that were only detectable if the syntactic (or prosodic) context was taken into account (Friederici et al., 2007; Sammler et al., 2010). The present study extends these findings to a new case with temporary dysfunctions of relevant white matter tracts.

We report the case of a patient in whom right dorsal and transcallosal connectivity were transiently compromised due to a vasogenic peritumoral edema, allowing assessment of potential prosodic deficits and their recovery in the same individual. Vasogenic edemas are extracellular edemas; other than cytotoxic edemas they infiltrate white matter, not cell bodies (Stokum, Gerzanich, & Simard, 2016), i.e., leave the neurons largely intact if the edema is medically treated to induce its reabsorption. Nevertheless, vasogenic edemas can compromise function in that they compress tissue and disturb information flow along the infiltrated white matter tracts. While the underlying neurophysiological mechanisms are still not fully understood, resulting physical or cognitive deficits are typically alleviated over the course of edema reabsorption (e.g., Bizzi et al., 2012). We capitalized on this phenomenon to probe the involvement of right dorsal and inter-hemispheric tracts in prosody perception in a patient with vasogenic edema infiltrating right AF/SLF and the posterior corpus callosum.

One important consideration for our investigation is that vasogenic edemas are seen around brain tumors. Our patient was diagnosed with a benign convexity meningioma (grade I) in the right parietal region. This type of meningioma is a slow-growing tumor on the surface of the brain, i.e., not invading grey matter. In line with a slow growth rate, symptoms typically have an insidious onset such as slowly evolving headache, suggestive of increased intracranial pressure, or a protracted history of partial seizures (Rockhill, Mrugala, & Chamberlain, 2007). Complete excision of the meningioma is often curative. Slow growth allows for functional compensation; among intracranial tumors, meningiomas are the ones with the highest incidental discovery rate, and

can remain neurologically and cognitively asymptomatic, especially if located in the right hemisphere (Nishizaki, Ozaki, Kwak, & Ito, 1999). Overall, the long-term course of meningioma with potential for compensation make it likely that post-surgical reversal of pre-surgical deficits in our patient are due to edema absorption, i.e., relief from edema-induced compression and recovery of white matter tracts, rather than removal of the tumor.

The present study combined pre- and post-surgical diffusion-weighted neuroimaging with behavioral assessment of linguistic prosody perception in a male patient with a right parietal edema infiltrating AF/SLF and the posterior corpus callosum. Specifically, we assessed the patient's ability to detect prosodic phrase boundaries that mismatched syntactic phrase structure (Eckstein & Friederici, 2006), requiring both right-hemispheric prosodic contour processing as well as syntax-prosody alignment across both hemispheres. To rule out general left-hemispheric syntax processing deficits, a non-prosodic control task was designed that tested comprehension of sentences with canonical and non-canonical syntactic structure (Gierhan, 2013a). The patient's performance was compared with performance of ten matched healthy controls (HC) that were tested twice at the same interval as the patient, to control for learning effects. General cognitive functioning was assessed with a standard neuropsychological test battery. If right dorsal and/or transcallosal pathways are essential for prosodic structure building and vasogenic edemas compromise white matter function, the patient's prosody perception should (i) be deficient in the pre-surgical but normal in the post-surgical session compared to HC, should (ii) increase more strongly between pre- and postsurgical session than in HC (who could show learning effects), and should (iii) increase more strongly than in the non-prosodic syntax task and neuropsychological tests.

2. Materials and methods

2.1. Participants

Our patient (male, 43 years, right-handed) presented for assessment of two generalized epileptic seizures at the neurosurgical department of the University Hospital Leipzig. A computer tomography (CT) of the brain revealed a vasogenic edema (37393 mm³) surrounding a benign convexity meningioma (25519 mm³) in the right parietal lobe (Fig. 1A; see also Supplementary Fig. 1). The edema was immediately admitted to treatment with glucocorticoids/dexamethasone (Meixensberger & Jaeger, 2005; 48 mg intravenous on the day of diagnosis, then 3 × 8 mg daily up until surgery). Such a treatment leads to the continuous absorption of the edema within a period of 2–3 weeks, usually associated with a considerable functional recovery that illustrates that the edema causes transient deficits beyond those caused by the tumor. Twenty-one days after diagnosis, the tumor was microsurgically resected. We tested the patient on day 2 and acquired MR images on day 3 after beginning of medication, i.e., when the edema still compromised the dorsal fiber tracts. The second acquisition was 110 days after neurosurgical resection of the tumor and full absorption of the edema (i.e., 134 days after the first session). Ten healthy male control participants (HC) matched in age (mean ± SEM: 42 ± 0.6 years), years of school education (10.2 ± 0.2 years), and handedness were tested and re-tested with a mean interval of 145 ± 12 days between sessions to account for potential learning effects in a test-retest setup. Neither the patient, nor the controls were musicians; none of the controls reported hearing deficits as verified with a MAICO MA 33 audiometer (MAICO Diagnostics GmbH, Berlin, Germany). The patient displayed slightly lower hearing on the left than the right ear (125 Hz: left 17.5/right 32.5 dB HL; 250 Hz: 20.0/37.5 dB HL; 1000 Hz: 25.0/45.0 dB HL; 4000 Hz: 17.5/55 dB HL). This was accounted for by adjusting the volume of the experimental stimuli to well audible level. All participants gave written informed consent. The study was approved by the ethics committee of the University of Leipzig (017-10-180112009).

2.2. MRI data acquisition

Anatomical and diffusion MRI data were obtained pre- and post-surgically in the patient and once for seven of the HC. Three HC were not scanned due to MR incompatibility. Data were acquired with a 32-channel head coil in a 3 Tesla TIM TRIO scanner (Siemens Healthineers, Erlangen, Germany) at the Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany. High-resolution T₁-weighted anatomical datasets (1 × 1 × 1 mm³ voxel size) were acquired using a 3D magnetization-prepared rapid gradient echo (MPRAGE) sequence (repetition time TR = 1300 ms, echo time TE = 3.46 ms, 176 sagittal slices, field of view FOV = 240 × 256 mm², flip angle = 10°). The diffusion-weighted data sets were acquired with a twice-refocused spin-echo EPI sequence (TR = 12.9 s, TE = 100 ms, 88 axial slices without gap, FOV = 220 × 220 mm², flip angle = 90°, GRAPPA acceleration factor 2) with a voxel size of 1.72 × 1.72 × 1.7 mm³. Diffusion-weighting was isotropically distributed along 60 diffusion-encoding gradient directions with a b-value of 1000 s/mm². Acquisition of diffusion-weighted images was interspersed with seven images without diffusion-weighting (b0), one at the beginning of the sequence and one after each block of 10 diffusion-weighted images, serving as anatomical reference for offline motion correction. Total duration of both MRI scans was about 25 min.

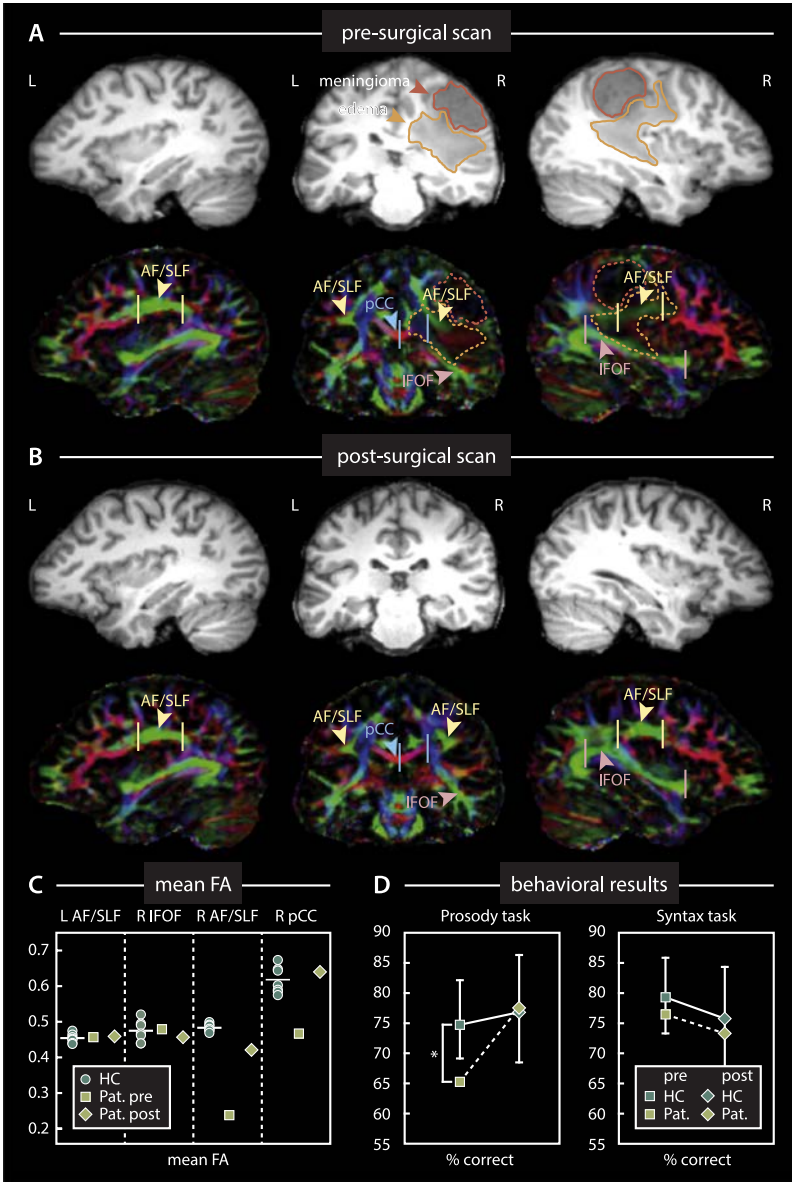
2.3. Behavioral assessment

2.3.1. Prosody task

Participants were presented with 96 German sentences spoken by a female native speaker. Half of the sentences contained a prosodic irregularity in that pitch contour signaled sentence closure before all obligatory syntactic elements had occurred (Fig. 2). All sentences consisted of a matrix clause including a proper name ('Steffen') and a verb ('steht'/'sees'), and a subordinate clause including a complementizer ('dass'/'that'), a noun phrase ('der Lehrer'/'the teacher'), a prepositional phrase ('beim Tadel'/'during the reproof'), and a verb ('schmunzelt'/'smiles'). Sentences were rendered prosodically irregular by inserting a prosodic boundary tone with falling pitch contour on the penultimate noun (i.e., 'Tadel') by means of cross splicing (for details on stimulus preparation and acoustic properties, see Eckstein & Friederici, 2006). This acoustic manipulation induced a mismatch between syntactic and prosodic structure: Syntax predicted sentence continuation beyond the noun given that at least the obligatory verb ('smiles') was still to follow. As a consequence, the noun's falling prosodic contour violated syntax-driven expectancies of prosodic form (auditory examples are provided as Supplementary Material). Prosodically regular and irregular sentences were presented with equal probability across the experiment in pseudo-random order with no more than three consecutive trials of the same type. Sentences had an average (± SEM) duration of 3368 ± 19 ms and were presented binaurally at a comfortable volume via loudspeakers (Eltax HT-1; Eltax, Aulum, Denmark) in a silent room using Presentation 12.2. (Neurobehavioral Systems, Inc., Albany, Canada). Participants were asked to judge the prosodic regularity of the sentences by pressing one of two buttons. The experiment started with ten practice trials (with feedback) to acquaint participants with the task and lasted approximately 15 min.

2.3.2. Syntax task

To ensure that putative deficits in detecting prosodic irregularities are not due to a general deficit in processing syntactic information, a non-prosodic sentence comprehension task was adapted from Gierhan (2013a) that is known to involve left (not right) dorsal fiber tracts (Friederici & Gierhan, 2013). Participants were presented with 90 sentences with canonical and non-canonical syntactic structure spoken by a female native speaker of German. Sentences were composed of a pronoun ('Dann'/'Then'), a verb ('grüßt'/'greet'), an animate subject ('der Soldat'/'the soldier'), and an animate object ('den Major'/'the



(caption on next page)

Fig. 1. Pre- and post-surgical results. (A) Pre-surgical anatomical (grey) and directionally colored diffusion MR FA-images of the patient. The peritumoral edema (orange) infiltrated right dorsal fronto-temporal/parietal pathways leading to lower fractional anisotropy (FA; dark green) in right compared to left AF/SLF (white arrows). The meningioma (red) did not invade but displaced brain tissue as can be seen in the position of the corpus callosum (CC). White bars indicate borders for extraction of mean FA in AF/SLF as depicted in (C). Edema and tumor were manually segmented with itk-SNAP 3.2 (<http://www.itksnap.org/pmwiki/pmwiki.php>). (B) Post-surgically, the edema was fully reabsorbed and the tumor resected, allowing for a recovery of tissue properties (measured by FA) in right AF/SLF (higher green saturation than in (A)). (C) Pre-surgically low FA in right AF/SLF and posterior CC (pCC) of the patient (green square) approached mean FA values of HC (green circles) in the post-surgical session (green diamond), while left AF/SLF as well as right IFOF showed similar FA values as in HC across both sessions. Horizontal lines indicate mean FA of HC. (D) Behavioral results in the prosody and syntax task. Plots show pre- and post-surgical performance of the patient (dashed line) compared to the 1st and 2nd test session of ten matched healthy controls (solid line). Error bars indicate bootstrapped 95% confidence intervals. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

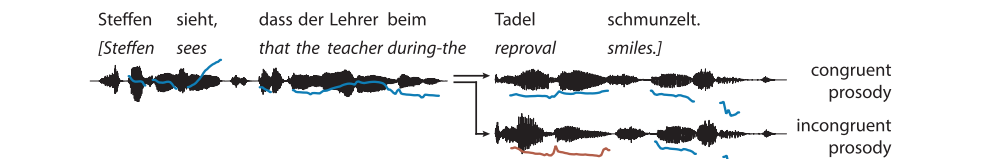


Fig. 2. Stimulus examples of the prosody task. The penultimate word of German sentences (here: ‘Tadel’) was rendered prosodically incongruent (falling instead of rising pitch contour) by means of cross-splicing (Eckstein & Friederici, 2006).

major’). Half of the sentences presented the subject first, followed by the object (S-O sentences, e.g., ‘Dann grüßt der_[S] Soldat_[S] den_[O] Major_[O].’/ ‘Then greets the_[S] soldier_[S] the_[O] major_[O].’). The other half presented the object before the subject (O-S sentences, e.g., ‘Dann grüßt den_[O] Major_[O] der_[S] Soldat_[S].’/ ‘Then greets the_[O] major_[O] the_[S] soldier_[S].’), which is a legal construction in German. (Note that in both examples, it is the soldier who greets the major.) Sentences were spoken with neutral, non-accentuated prosody (auditory examples are provided as [Supplementary Material](#)). After each sentence, participants were asked a comprehension question (spoken by the same female speaker) of the sort ‘who did what to whom’ to assess their syntactic processing. These questions required either a ‘yes’ response (50%), or a ‘no’ response because it reversed subject and object (25%; e.g., it was asked whether the major greets the soldier), or introduced a new action or protagonist (25%; e.g., it was asked whether the soldier blames the major; for examples, see [Table 1](#)). Sentences had an average (\pm SEM) duration of 2619 ± 12 ms and were presented in the same way as stimuli in the prosody task. The experiment started with six practice trials (with feedback) and lasted approximately 15 min.

2.3.3. Neuropsychological assessment – general cognitive functioning

Furthermore, to ensure that the results in the prosody task were not due to general cognitive deficits, we applied a brief battery of standard neuropsychological tests. This battery assessed the ability to store and rehearse verbal and spatial contents using the digit-span and block-span test of the Wechsler Memory Scale (WMS-R; [Wechsler, 1987](#)), the

ability to focus attention using the d2 test that requires speeded detection of target symbols amongst distractors ([Brickenkamp, 1994](#)), and visuo-spatial reasoning using the third subtest of the Leistungsprüfungssystem (LPS-3) that requires identification of rule violations in series of symbols ([Horn, 1983](#)). At the outset as well as post-surgically, the patient’s performance was within normal average-low to above-average range relative to the age matched normative samples provided by the standard neuropsychological tests (pre-/post-surgical percentile rank of forward digit-span: 35/35; backward digit-span: not assessed; forward block-span: 93/32; backward block-span: 23/23; LPS-3: 69/93; d2: 34/54; standard diagnostic cutoffs of normal average performance: 16–84).

2.4. Data analysis

2.4.1. Diffusion MRI data

Diffusion MRI data were analyzed using LIPSI (Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany), and FSL (FMRIB, University of Oxford, UK, www.fmrib.ox.ac.uk/fsl). First, T₁-weighted structural scans were reoriented to the sagittal intercommissural plane and the brain was segmented in LIPSI. A trimmed brain mask was created by shrinking the inner skull surface by 7 mm to later reduce noisy endpoints of streamlines. Then, motion correction parameters for the 60 diffusion-weighted images were computed based on the 7 (b0) reference images distributed over the entire sequence using rigid-body transformations implemented in FSL. Interpolated motion correction parameters were combined with a global registration to the T₁-anatomy and applied to all 60 volumes that were resampled to an isotropic resolution of 1.7 mm. The gradient direction for each volume was corrected using the rotation parameters. Finally, the diffusion tensor, the three eigenvectors, and the fractional anisotropy (FA) value were computed for each voxel.

The diffusion tensor image was used for full-brain deterministic fiber tracking using an in-house implementation of the tensor detection algorithm ([Lazar et al., 2003](#)) in all voxels of the trimmed brain mask. Together with a lowered FA threshold of 0.075, this algorithm allowed for robust tracking in areas of low anisotropy (i.e., the edema) while excluding the tumor area and the ventricles. Lowering the FA threshold without any additional processing might, however, introduce false positive connections. To eliminate spurious streamlines while retaining the bundles of interest, tracking was performed in an adapted two-step approach. First, right AF/SLF and posterior CC were selected from the full-brain tracking using inclusion and exclusion masks. The volume of

| Table 1 |
|---|
| Stimulus examples of the syntax task (with literal translations). |
| (1) Example S-O sentence (subject-first) Dann grüßt der _[S] Soldat _[S] den _[O] Major _[O] . Then greets the _[S] soldier _[S] the _[O] major _[O] . |
| (2) Example O-S sentence (object-first) Dann grüßt den _[O] Major _[O] der _[S] Soldat _[S] . Then greets the _[O] major _[O] the _[S] soldier _[S] . |
| (3) Examples of questions that require a ‘yes’ response Grüßt der _[S] Soldat _[S] den _[O] Major _[O] ?/Grüßt den _[O] Major _[O] der _[S] Soldat _[S] ? Greets the _[S] soldier _[S] the _[O] major _[O] ?/Greets the _[O] major _[O] the _[S] soldier _[S] ? |
| (4) Examples of questions that require a ‘no’ response Grüßt der _[S] Major _[S] den _[O] Soldat _[O] ?/Rügt der _[S] Soldat _[S] den _[O] Major _[O] ? Greets the _[S] major _[S] the _[O] soldier _[O] ?/Blames the _[S] soldier _[S] the _[O] major _[O] ? |

S: subject, O: object.

the extracted bundle was computed by selecting all voxels that were crossed by at least two streamlines. In this way, spurious single streamlines were excluded. Inclusion masks for right AF/SLF were placed in the temporal lobe near the temporo-parietal junction and in the posterior frontal lobe (Catani, Jones, & Ffytche, 2005; Mori, 2007). Streamlines crossing the extreme capsule and the thalamus were excluded. Inclusion masks for posterior CC were placed in isthmus and splenium of the CC and right posterior temporal lobe. Streamlines reaching parietal or occipital lobe were excluded. Masks were individually adapted for every participant. In the second step, the extracted bundle volume was dilated by 1 mm and streamline tracking was performed again, restricted to this volume. Then, the same inclusion and exclusion masks as described above were applied again and the final bundle was extracted.

The same procedure was applied to define left AF/SLF and right IFOF, as contralateral dorsal and ipsilateral ventral control tracts, respectively. Masks for left AF/SLF were analogous to those used in the right hemisphere. Inclusion masks for the right IFOF were placed in right IFG and right angular gyrus (Makris & Pandya, 2009). Streamlines targeting areas dorsal of angular gyrus or crossing right AF/SLF or the corpus callosum were excluded.

To assess recovery of FA, mean FA values of right AF/SLF and posterior CC were extracted from pre- and post-surgical scans, limited to the central horizontal parts of the extracted bundle volumes, i.e., deep white matter (see borders in Fig. 1A and B) where tractography is most robust and reproducible and FA values are less affected by partial volume effects than in the fanning ends of the bundle. To demonstrate general test/re-test stability of FA (compare green squares and diamonds in Fig. 1C), mean FA values of left AF/SLF and right IFOF were extracted in the same way, also excluding the fanning ends of the tract. The IFOF runs ventral of the edema but it was made sure that the chosen segment fully covered the extent of the edema along its y-dimension. The patient's FA values were statistically compared to those of the HC by means of bootstrapped confidence intervals (see Section 2.4.2).

2.4.2. Behavioral data

Prosody perception and syntax processing abilities were quantified as %correct responses. Cognitive abilities were compared based on raw test scores (Table 2; raw scores rather than percentile ranks were used for statistical comparison to preserve original distribution and differences between scale units; Thorndike & Thorndike-Christ, 2013). Change scores were calculated for each test by subtracting pre-surgical (1st session) from post-surgical (2nd session) performance measures. To specifically compare performance changes in the two matched language tasks—prosody vs. syntax—the difference between the two change scores $([Prosody_{post} - Prosody_{pre}] - [Syntax_{post} - Syntax_{pre}])$ was calculated. For statistical comparison of patient's and HC's data, we chose a non-parametric bootstrapping approach as in Meyer et al. (2014) to account for the limited number of HC and the resulting violations of the sphericity assumption (Mauchly, 1940). We generated two-tailed 95% confidence intervals (CIs) based on 10,000 random

draws from the values of HC (Efron, 1979) as implemented in MATLAB (The MathWorks, Inc., Natick, MA, USA). Patient's scores were considered significantly different from those of HC when they were outside these bootstrapped CIs. HC's performance in 1st and 2nd session was compared by means of two-tailed t-tests for paired samples.

3. Results

3.1. Anatomy and diffusion MRI data

Post-surgically, the tumor was fully resected with no remaining cavity, and the edema was entirely reabsorbed (Fig. 1B; see also Supplementary Fig. 2). As shown in Fig. 1C and Table 2, the patient's right AF/SLF showed significantly reduced mean FA in the pre-surgical session (green square; $M = 0.238$) compared to HC (green circles; $M \pm SD = 0.481 \pm 0.010$) that recovered to near-normal values after surgery (green diamond; $M = 0.421$), although it remained below HC's mean FA values. Similarly, the patient's posterior CC connecting the temporal lobes showed significantly reduced pre-surgical FA ($M = 0.470$) compared to HC ($M \pm SD = 0.619 \pm 0.178$) that fully recovered after surgery ($M = 0.643$). In turn, the patient's left AF/SLF showed nearly identical FA values in the pre- ($M = 0.457$) and post-surgical session ($M = 0.459$), both not significantly different from mean FA in HC ($M \pm SD = 0.454 \pm 0.013$). Likewise, the patient's right IFOF showed stable FA values in the pre- ($M = 0.478$) and post-surgical scan ($M = 0.454$) that were both within the bounds of HC's bootstrapped CI ($M \pm SD = 0.472 \pm 0.029$; for statistical values, see Table 2).

3.2. Behavioral data

In the **prosody task**, our patient showed a performance increase from pre- to post-surgical session (+12.50%) that was significantly stronger than the nominal performance change in HC ($M \pm SD = 2.08\% \pm 2.33\%$; see Fig. 1D and Table 3 for statistical details). Pre-surgically, the patient showed significantly lower performance (65.63%) than HC ($M \pm SD = 75.11\% \pm 3.91\%$), whereas his performance reached normal levels in the post-surgical session (78.13%; HC: $M \pm SD = 77.19\% \pm 5.41\%$). HC performed equally well across the two sessions (paired samples t-test: $t(9) = -0.89$, $p > .395$).

In the non-prosodic **syntax task**, our patient showed stable performance across sessions (pre: 76.67%; post: 73.33%), similar to HC (1st: $M \pm SD = 79.22\% \pm 3.36\%$; 2nd: $75.33\% \pm 4.99\%$). Performance did not differ between patient and HC, neither pre- nor post-surgically nor in terms of performance change over time (see Table 3 for CIs). HC showed no performance differences between the two sessions ($t(9) = 1.29$, $p > .229$; see Fig. 1D).

The comparison of the **change scores between the two tasks** revealed that the patient's pre-post-surgical performance gain was stronger (by 15.84%) in the prosody than in the syntax task, which was significantly more than in HC (CI: 0.15–12.99%). Note that this result cannot be due to general differences in task difficulty or ceiling and floor effects. HC's results confirmed that the two tasks were perfectly matched in difficulty (no significant main effect of TASK in a repeated-measures ANOVA with factors TASK and SESSION: $F(1,9) = 0.07$, $p > .791$) and were well above chance level (one-sample t-tests against 50% per session and task: $t's > 5.02$, $p's < .002$) and well below ceiling (one-sample t-tests against 100% per session and task: $t's < -4.21$, $p's < .003$).

Likewise, the patient's performance gain in the prosody task cannot be explained by a putative recovery of **general cognitive functions**. In none of the cognitive tasks, our patient showed a pattern of pre-surgical deficit and post-surgical recovery (as observed in the prosody task). In the span tests, the patient showed perfectly *stable* (forward digit-span, backward block-span) or post-surgically *decreased* (instead of increased)

Table 2
Statistical comparison of fractional anisotropy in patient and healthy controls.

| Fiber tract | Healthy Controls | | 95% CI | | Patient | | |
|--------------|------------------|-------|--------|-------|--------------|--------------|--------|
| | M | SEM | Lower | Upper | Pre | Post | Change |
| Right AF/SLF | 0.481 | 0.010 | 0.476 | 0.491 | 0.238 | 0.421 | +0.183 |
| Posterior CC | 0.619 | 0.178 | 0.596 | 0.648 | 0.470 | 0.643 | +0.173 |
| Left AF/SLF | 0.454 | 0.013 | 0.446 | 0.464 | 0.457 | 0.459 | +0.002 |
| Right IFOF | 0.472 | 0.029 | 0.453 | 0.494 | 0.478 | 0.454 | -0.024 |

AF/SLF: arcuate/superior longitudinal fascicle, CC: corpus callosum, IFOF: inferior fronto-occipital fascicle. Bold values represent significant lower fractional anisotropy in patient than healthy controls.

Table 3
Statistical comparison of behavioral results of patient and healthy controls in the two sessions.

| Test | Healthy Controls | | 95% CI | | Patient | |
|---|------------------|------------|--------|--------|--------------|-----------|
| | <i>M</i> | <i>SEM</i> | Lower | Upper | Score | Direction |
| <i>Critical tests (% correct)</i> | | | | | | |
| Prosody perception | | | | | | |
| 1st session | 75.11 | 3.91 | 68.23 | 82.82 | 65.63 | – |
| 2nd session | 77.19 | 5.41 | 67.25 | 87.12 | 78.13 | = |
| Difference (2nd – 1st session) | 2.08 | 2.33 | –2.61 | 6.04 | 12.50 | + |
| Syntax processing | | | | | | |
| 1st session | 79.22 | 3.36 | 73.33 | 85.67 | 76.67 | = |
| 2nd session | 75.33 | 4.99 | 66.17 | 84.45 | 73.33 | = |
| Difference (2nd – 1st session) | –3.89 | 3.01 | –9.17 | 2.22 | –3.34 | = |
| <i>General cognitive functions (raw scores)</i> | | | | | | |
| d2 concentration performance | | | | | | |
| 1st session | 175.20 | 10.30 | 158.00 | 196.20 | 141 | – |
| 2nd session | 197.50 | 10.40 | 178.50 | 217.00 | 163 | – |
| Difference (2nd – 1st session) | 22.30 | 2.92 | 17.60 | 28.60 | 22 | = |
| Digit span forward | | | | | | |
| 1st session | 8.10 | 0.48 | 7.10 | 8.90 | 7 | – |
| 2nd session | 8.10 | 0.60 | 7.00 | 9.30 | 7 | = |
| Difference (2nd – 1st session) | 0.00 | 0.49 | –1.00 | 0.80 | 0 | = |
| Block span backward | | | | | | |
| 1st session | 8.40 | 0.58 | 7.20 | 9.40 | 7 | – |
| 2nd session | 9.60 | 0.67 | 8.10 | 10.60 | 7 | – |
| Difference (2nd – 1st session) | 1.20 | 0.53 | 0.40 | 2.40 | 0 | – |
| Block span forward | | | | | | |
| 1st session | 8.60 | 0.37 | 8.00 | 9.50 | 12 | + |
| 2nd session | 9.30 | 0.37 | 8.60 | 9.90 | 8 | – |
| Difference (2nd – 1st session) | 0.70 | 0.37 | 0.10 | 1.50 | –4 | – |
| LPS-3 | | | | | | |
| 1st session | 27.50 | 1.92 | 21.80 | 30.00 | 23 | = |
| 2nd session | 29.90 | 1.22 | 27.40 | 32.00 | 30 | = |
| Difference (2nd – 1st session) | 2.40 | 1.38 | 0.10 | 5.50 | 7 | + |

Bold values represent significant differences between patient and healthy controls. Symbols indicate whether patient's performance (change) was higher (+), similar to (=) or lower than (–) in controls.

performance (forward block-span). In the attention task (d2), patient's performance increased over time, however, to a similar extent as in HC ($t(9) = -7.64, p < .001$), i.e., denoting a general learning effect rather than recovery. Overall, in both the attention task and the span tests, our patient scored mostly slightly below HC, constantly across both sessions, making it unlikely that these abilities contributed to performance recovery in the prosody test. Only exception was the spatial-reasoning task (LPS-3). Here, the patient showed a pre-post-surgical performance increase that was significantly stronger than in HC, although his pre-surgical performance did not significantly differ from HC's scores. This finding is compatible with right dorsal pathway involvement in spatial relational reasoning (Krawczyk, 2012; Shokri-Kojori, Motes, Rypma, & Krawczyk, 2012; Watson & Chatterjee, 2012).

4. Discussion

Following the hypotheses that right dorsal (Sammler et al., 2015) and posterior transcallosal fiber tracts (Friederici et al., 2007; Sammler et al., 2010) support prosodic structure building, and that vasogenic edemas can induce reversible deficits when infiltrating white matter tracts (Bizzi et al., 2012; Gierhan et al., 2012), we tested an edema patient on his abilities to process linguistic prosody, both before and after edema treatment and neurosurgery. Before treatment, i.e., when the edema was infiltrating and compressing right AF/SLF and posterior CC, the patient presented deficits in recognizing irregular prosody that mismatched syntactic structure. Reabsorption of the edema and resection of the benign meningioma were associated with a reversal of these deficits. This behavioral recovery was accompanied by increased average FA values in right AF/SLF and posterior CC, while FA was

constant in left AF/SLF and right IFOF. Neither short term and working memory nor syntactic comprehension exhibited a similar pre-to-post-surgical performance gain. Consequently, the reversal of the prosodic deficit cannot be due to a nonspecific recovery of cognitive or verbal abilities. Rather, the findings invite the discussion of *causal* involvement of right dorsal and posterior transcallosal fiber tracts in the processing of prosodic structure.

What remains to be resolved is exactly *how* these pathways may contribute to the processing of prosodic information. In cognitive terms, it seems plausible to think of prosodic structure building as a hierarchical multi-step process with information passing through consecutive stages of basic acoustic analyses, higher-level auditory grouping and integration with syntactic and semantic information (for similar multi-step models of emotional prosody, see Brück, Kreifelts, & Wildgruber, 2011; Kotz & Paulmann, 2011; Schirmer & Kotz, 2006; Wildgruber, Ackermann, Kreifelts, & Ethofer, 2006). Intra- and inter-hemispheric pathways secure rapid and bidirectional information exchange within and between these stages represented in distributed neural networks. Although the present study cannot dissociate the functions supported by right dorsal and transcallosal pathways (because both of them were infiltrated by the edema), models of auditory, speech and language processing (Bornkessel-Schlesewsky, Schlesewsky, Small, & Rauschecker, 2015; Friederici & Alter, 2004; Friederici, 2011; Hickok & Poeppel, 2007; Rauschecker & Scott, 2009) may lend a basis to start reflecting upon possible mechanisms, as will be done in the remainder of the text. Three capacities should be considered: (i) the time-sensitive tracking of prosodic features and (ii) their grouping into higher order structures, and (iii) the integration of syntactic and prosodic information.

4.1. Right dorsal pathways – time-sensitive tracking and structuring of prosodic features

The decision upon a sentence's prosodic form and regularity, as in the present study, requires processing of auditory-prosodic cues, e.g., the recognition of pitch contours as either rising (here: regular) or falling (irregular). This requirement may tap into the capacity of dorsal pathways to **track auditory information**—including pitch—over time. In its accepted role to map sound to articulation (Hickok & Poeppel, 2007; Saur et al., 2008), dorsal pathways are typically assumed to host auditory-motor loops that continuously monitor the sound and pitch of one's own vocalizations during speech production (Guenther & Vladusich, 2012; Hickok, 2012; Houde & Chang, 2015; Zarate, 2013). The same auditory-motor system in reversed processing mode has been proposed to serve speech perception (Hickok, Houde, & Rong, 2011; Rauschecker, 2011), including perception of prosodic contour (Sammler et al., 2015) and discrimination of vocal pitch (D'Ausilio, Bufalari, Salmas, Busan, & Fadiga, 2011). Altogether, it seems that **dorsal connections between auditory temporal and (pre)motor areas** could provide a basic computational building block necessary to track prosodic pitch contours over time, that may have been temporarily disrupted in our patient.

From a higher-order linguistic point of view, successful **parsing of prosodic structure** involves the segmentation of sentence-level prosodic information into constituent elements, e.g., intonational phrases (Nespor & Vogel, 1986; Selkirk, 1984). This requirement resonates with recent proposals that the dorsal stream may identify basic prosodic units (e.g., prosodic words) and combine them into successively larger linguistic chunks over time (e.g., intonational phrases; Bornkessel-Schlesewsky & Schlewsky, 2013). Right AF/SLF involvement in the structuring of auditory pitch information in music (Loui, Alsop, & Schlaug, 2009; Peretz, 2016) (but see Chen et al., 2015) may provide cross-domain support for this idea. To date, only little is known about the neuroanatomical bases of sentence-level prosodic structure building. However, several findings are compatible with fronto-temporal information exchange during prosodic phrasing (Geiser, Zaehle, Jancke, & Meyer, 2008; Ischebeck, Friederici, & Alter, 2008; Strelnikov, Vorobyev, Chernigovskaya, & Medvedev, 2006) (but see den Ouden et al., 2016). Whether or not the fronto-temporal information exchange happens via a **dorsal posterior temporal to inferior frontal pathway**, as possibly suggested by our patient's performance, is an interesting topic for future research.

Irrespective of whether dorsal white matter tracts constitute building blocks for basic pitch monitoring or advanced prosodic structuring (or both in interaction), their computational characteristics may be described in terms of internal (forward) models that serve to **predict forthcoming sensory events** (here: the to-be-perceived pitch contour) on the basis of previous input (Bornkessel-Schlesewsky & Schlewsky, 2013; Rauschecker, 2011). Notably, these predictions are likely to arise from both prosodic and non-prosodic priors—including continuity of pitch contour and prosodic phrase but also syntactic structure, as will be discussed in the following.

4.2. Posterior transcallosal pathways – syntax-prosody interface

Syntactic structure guides prosodic parsing (Buxó-Lugo & Watson, 2016; Cole et al., 2010; Cutler et al., 1997) and *vice versa* (Lehiste, 1973; Snedeker & Trueswell, 2003) and the posterior corpus callosum has been proposed as the relevant interface (Friederici et al., 2007; Sammler et al., 2010). The present task drew on this interface by violating prosodic expectancies that were established through syntactic structure (see Section 2). The patient's sentence-level syntactic processing was arguably intact in both sessions as indicated by his unimpaired performance in the non-prosodic syntax task. Yet, online syntactic processes may have no longer triggered the build-up of prosodic expectancies because of the temporary disruption of the necessary

crosstalk between syntax and prosody processing streams. This transient deficit would mark a further case for the relevance of the posterior CC for information exchange between the temporal lobes in syntax-prosody alignment.

Taken together, our patient's performance pattern may be interpreted as temporary deficit in tracking and/or predicting prosodic contour at lower and/or higher linguistic levels. Although we cannot isolate the relative contribution of intra- and inter-hemispheric fiber tracts in the present case, both the capacity of dorsal pathways to **process auditory-prosodic information in a time-sensitive manner** (Bornkessel-Schlesewsky & Schlewsky, 2013; Frühholz & Grandjean, 2013; Kreiner & Eviatar, 2014; Rauschecker, 2011) and the role of transcallosal pathways in **interfacing prosody and syntax** may satisfy crucial requirements for prosodic parsing: the analysis *how* prosodic information evolves over time and relative to concurrent syntactic information.

4.3. Focality of edema-induced disruptions

Vasogenic edemas are extracellular edemas; they spread along white matter tracts and are assumed to temporarily disturb information flow along these tracts (Bizzi et al., 2012; Gierhan et al., 2012). It is, however, difficult to reliably estimate the extent of the affected region. First, the edema invaded the posterior CC as well as several dorsal sub-pathways, including those supporting visuo-spatial attention and relational integration (Krawczyk, 2012; Shokri-Kojori et al., 2012; Watson & Chatterjee, 2012) such as right SLF III (Thiebaut de Schotten, Dell'Acqua, et al., 2011). The latter may account for the patient's post-surgical performance increase in visuo-spatial reasoning. A second obvious concern is that space-occupying edemas may compromise not only fiber tracts but may also reduce functionality of surrounding grey matter through compression, i.e., of right parietal areas in the present study. Prevailing models of prosody perception do not typically include the parietal lobe (Brück et al., 2011; Friederici & Alter, 2004; Schirmer & Kotz, 2006; Wildgruber et al., 2006; Wildgruber, Ethofer, Grandjean, & Kreifelts, 2009; Witteman, Van Heuven, & Schiller, 2012); nevertheless, right inferior parietal activations (BA 40/7) have been occasionally reported in prosody studies (Belyk & Brown, 2014; Merrill et al., 2012) and were proposed to reflect rehearsal-based working memory processes for prosodic pitch contours (Kreitewolf et al., 2014; Perrone-Bertolotti et al., 2013). The present study cannot discriminate between neural mass effects and changes in white matter conductivity. Yet, it is notable that the recovery of prosodic deficits was accompanied by increased average FA values in those tracts that had passed through the edema before surgery. This relationship between recovered diffusivity values and recovered cognitive functions makes it likely that white matter connectivity contributed to the behavioral effects. The potential contribution of IPL to the observed effects can be probed in future studies with healthy participants by means of transcranial magnetic stimulation.

5. Conclusion

In summary, the present case study provides new insights into right dorsal and posterior transcallosal pathway functions in auditory language comprehension by demonstrating that temporary edema-induced dysfunction of right AF/SLF and posterior CC perturbs sentence-level prosody perception. The sensitivity of dorsal pathways to temporal dynamics of auditory information may constitute the decisive computational feature that provides the dorsal pathways with the capacity to track, predict and/or evaluate prosodic contour over time. Future research should look into potential divisions of labor between different dorsal sub-pathways and their interaction. Assuming functional parallels to left dorsal pathways, posterior temporal to premotor connections might be particularly suited to track pitch over time by virtue of the time-sensitivity of the motor system (Houde & Chang, 2015;

Rauschecker, 2011). Posterior temporal to inferior frontal connections, in turn, might be involved in prosodic structure building by virtue of the combinatorial capacities of IFG (Friederici, 2011; Koelsch, 2005) and a linkage with non-prosodic syntactic information provided by the left hemisphere via the posterior corpus callosum (Friederici et al., 2007; Sammler et al., 2010).

6. Statement of significance

Dorsal and ventral pathways in the left hemisphere play established roles in syntactic parsing and semantic comprehension, while pathways for linguistic prosody perception remain little explored. The present single case lesion study suggests a causal role of right dorsal pathways in prosodic structure building that further depends on the integrity of the posterior corpus callosum to interact with concurrent syntactic information.

Acknowledgements

We thank Johanna Reichert and Katharina Fath for help with testing healthy controls. This research was funded by the Max Planck Society.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.bandl.2018.05.001>.

References

- Baum, S. R., & Pell, M. D. (1999). The neural bases of prosody: Insights from lesion studies and neuroimaging. *Aphasiology*, 13(8), 581–608.
- Belyk, M., & Brown, S. (2014). Perception of affective and linguistic prosody: An ALE meta-analysis of neuroimaging studies. *Social Cognitive and Affective Neuroscience*, 9(9), 1395–1403.
- Bizzi, A., Nava, S., Ferre, F., Castelli, G., Aquino, D., Ciaraffa, F., ... Piacentini, S. (2012). Aphasia induced by gliomas growing in the ventrolateral frontal region: Assessment with diffusion MR tractography, functional MR imaging and neuropsychology. *Cortex*, 48(2), 255–272.
- Blumstein, S., & Cooper, W. E. (1974). Hemispheric processing of intonation contours. *Cortex*, 10(2), 146–158.
- Borkessel-Schlesewsky, I., & Schlesewsky, M. (2013). Reconciling time, space and function: A new dorsal-ventral stream model of sentence comprehension. *Brain and Language*, 125(1), 60–76.
- Borkessel-Schlesewsky, I., Schlesewsky, M., Small, S. L., & Rauschecker, J. P. (2015). Neurobiological roots of language in primate audition: Common computational properties. *Trends in Cognitive Sciences*, 19(3), 142–150.
- Brickenkamp, R. (1994). *Test d2*. Göttingen: Hogrefe.
- Brück, C., Kreifelts, B., & Wildgruber, D. (2011). Emotional voices in context: A neurobiological model of multimodal affective information processing. *Physics of Life Reviews*, 8(4), 383–403.
- Buxó-Lugo, A., & Watson, D. G. (2016). Evidence for the influence of syntax on prosodic parsing. *Journal of Memory and Language*, 90, 1–13.
- Catani, M., Jones, D. K., & Ffytche, D. H. (2005). Perisylvian language networks of the human brain. *Annals of Neurology*, 57(1), 8–16.
- Chen, J. L., Kumar, S., Williamson, V. J., Scholz, J., Griffiths, T. D., & Stewart, L. (2015). Detection of the arcuate fasciculus in congenital amusia depends on the tractography algorithm. *Frontiers in Psychology*, 6, 1–11.
- Cole, J. (2015). Prosody in context: A review. *Language, Cognition and Neuroscience*, 30(1–2), 1–31.
- Cole, J., Mo, Y., & Baek, S. (2010). The role of syntactic structure in guiding prosody perception with ordinary listeners and everyday speech. *Language and Cognitive Processes*, 25(7–9), 1141–1177.
- Cooper, W. E., & Paccia-Cooper, J. (1980). *Syntax and speech*. Cambridge, MA: Harvard University Press.
- Cutler, A., Dahan, D., & Van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech*, 40, 141–201.
- D'Ausilio, A., Bufalari, I., Salmas, P., Busan, P., & Fadiga, L. (2011). Vocal pitch discrimination in the motor system. *Brain and Language*, 118(1–2), 9–14.
- den Ouden, D.-B., Dickey, M. W., Anderson, C., & Christianson, K. (2016). Neural correlates of early-closure garden-path processing: Effects of prosody and plausibility. *Quarterly Journal of Experimental Psychology*, 69(5), 926–949.
- Domahs, U., Klein, E., Huber, W., & Domahs, F. (2013). Good, bad and ugly word stress – fMRI evidence for foot structure driven processing of prosodic violations. *Brain and Language*, 125(3), 272–282.
- Duffau, H., Leroy, M., & Gatignol, P. (2008). Cortico-subcortical organization of language networks in the right hemisphere: An electrostimulation study in left-handers. *Neuropsychologia*, 46(14), 3197–3209.
- Eckstein, K., & Friederici, A. D. (2006). It's early: Event-related potential evidence for initial interaction of syntax and prosody in speech comprehension. *Journal of Cognitive Neuroscience*, 18(10), 1696–1711.
- Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7, 1–26.
- Fernández-Miranda, J. C., Wang, Y., Pathak, S., Stefaneau, L., Verstynen, T., & Yeh, F.-C. (2015). Asymmetry, connectivity, and segmentation of the arcuate fascicle in the human brain. *Brain Structure & Function*, 220(3), 1665–1680.
- Forke, S. J., Thiebaut de Schotten, M., Dell'Acqua, F., Kalra, L., Murphy, D. G., Williams, S. C., & Catani, M. (2014). Anatomical predictors of aphasia recovery: A tractography study of bilateral perisylvian language networks. *Brain*, 137, 2027–2039.
- Fridriksson, J., den Ouden, D.-B., Hillis, A. E., Hickok, G., Rorden, C., Basilakos, A., ... Bonilha, L. (2018). Anatomy of aphasia revisited. *Brain*.
- Friederici, A. D. (2011). The brain basis of language processing: From structure to function. *Physiological Reviews*, 91(4), 1357–1392.
- Friederici, A. D. (2012). The cortical language circuit: From auditory perception to sentence comprehension. *Trends in Cognitive Sciences*, 16(5), 262–268.
- Friederici, A. D., & Alter, K. (2004). Lateralization of auditory language functions: A dynamic dual pathway model. *Brain and Language*, 89(2), 267–276.
- Friederici, A. D., & Gierhan, S. M. E. (2013). The language network. *Current Opinion in Neurobiology*, 23(2), 250–254.
- Friederici, A. D., von Cramon, D. Y., & Kotz, S. A. (2007). Role of the corpus callosum in speech comprehension: Interfacing syntax and prosody. *Neuron*, 53(1), 135–145.
- Frühholz, S., & Grandjean, D. (2013). Processing of emotional vocalizations in bilateral inferior frontal cortex. *Neuroscience and Biobehavioral Reviews*, 37, 2847–2855.
- Frühholz, S., Gschwind, M., & Grandjean, D. (2015). Bilateral dorsal and ventral fiber pathways for the processing of affective prosody identified by probabilistic fiber tracking. *NeuroImage*, 109, 27–34.
- Gandour, J. T., Tong, Y. X., Wong, D., Talavage, T., Dziedzic, M., Xu, Y. S., ... Lowe, M. (2004). Hemispheric roles in the perception of speech prosody. *NeuroImage*, 23(1), 344–357.
- Geiser, E., Zaehle, T., Jancke, L., & Meyer, M. (2008). The neural correlate of speech rhythm as evidenced by metrical speech processing. *Journal of Cognitive Neuroscience*, 20(3), 541–552.
- Gierhan, S. M. E. (2013b). Connections for auditory language in the human brain. *Brain and Language*, 127, 205–221.
- Gierhan, S. M. E., Rhone, P., Anwender, A., Jost, L., Frydrychowicz, C., Hoffmann, K.-T., ... Friederici, A. D. (2012). Tracking the language pathways in edema patients: Preliminary results. In U. Goltz, M. Magnor, H.-J. Appellrath, H. Mathies, W.-T. Balke, & L. Wolf (Eds.), *Informatik 2012* (pp. 1102–1111). Bonn: Köllen Verlag.
- Gierhan, S. M. E. (2013). *Brain networks for language: Anatomy and functional roles of neural pathways supporting language comprehension and repetition*. MPI Series (Vol. 144). Leipzig: Max Planck Institute for Human Cognitive and Brain Sciences.
- Giraud, A. L., Kleinschmidt, A., Poeppel, D., Lund, T. E., Frackowiak, R. S., & Laufs, H. (2007). Endogenous cortical rhythms determine cerebral specialization for speech perception and production. *Neuron*, 56(6), 1127–1134.
- Glasser, M. F., & Rilling, J. K. (2008). DTI tractography of the human brain's language pathways. *Cerebral Cortex*, 18(11), 2471–2482.
- Griffiths, J. D., Marslen-Wilson, W. D., Stamatakis, E. A., & Tyler, L. K. (2013). Functional organization of the neural language system: Dorsal and ventral pathways are critical for syntax. *Cerebral Cortex*, 23(1), 139–147.
- Guenther, F. H., & Vladusich, T. (2012). A neural theory of speech acquisition and production. *Journal of Neurolinguistics*, 25(5), 408–422.
- Hellbernd, N., & Sammler, D. (2016). Prosody conveys speaker's intentions: Acoustic cues for speech act perception. *Journal of Memory and Language*, 88, 70–86.
- Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience*, 13(2), 135–145.
- Hickok, G., Houde, J. F., & Rong, F. (2011). Sensorimotor integration in speech processing: Computational basis and neural organization. *NeuroImage*, 69(3), 407–422.
- Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8(5), 393–402.
- Hofer, S., & Frahm, J. (2006). Topography of the human corpus callosum revisited – Comprehensive fiber tractography using diffusion tensor magnetic resonance imaging. *NeuroImage*, 32(3), 989–994.
- Horn, W. (1983). *Leistungsprüfungssystem*. Göttingen: Hogrefe.
- Houde, J. F., & Chang, E. F. (2015). The cortical computations underlying feedback control in vocal production. *Current Opinion in Neurobiology*, 33, 174–181.
- Huang, H., Zhang, J., Jiang, H., Wakana, S., Poetscher, L., Miller, M. L., ... Mori, S. (2005). DTI tractography based parcellation of white matter: Application to the mid-sagittal morphology of corpus callosum. *NeuroImage*, 26(1), 195–205.
- Ischebeck, A. K., Friederici, A. D., & Alter, K. (2008). Processing prosodic boundaries in natural and hummed speech: An fMRI study. *Cerebral Cortex*, 18(3), 541–552.
- Jamison, H. L., Watkins, K. E., Bishop, D. V. M., & Matthews, P. M. (2006). Hemispheric specialization for processing auditory nonspeech stimuli. *Cerebral Cortex*, 16(9), 1266–1275.
- Johnsrude, I. S., Penhune, V. B., & Zatorre, R. J. (2000). Functional specificity in the right human auditory cortex for perceiving pitch direction. *Brain*, 123, 155–163.
- Kellmeyer, P., Ziegler, W., Peschke, C., Juliane, E., Schnell, S., Baumgaertner, A., ... Saur, D. (2013). Fronto-parietal dorsal and ventral pathways in the context of different linguistic manipulations. *Brain and Language*, 127(2), 241–250.
- Koelsch, S. (2005). Neural substrates of processing syntax and semantics in music. *Current Opinion in Neurobiology*, 15(2), 207–212.
- Kotz, S. A., & Paulmann, S. (2011). Emotion, language, and the brain. *Language and Linguistics Compass*, 5(3), 108–125.
- Krawczyk, D. C. (2012). The cognition and neuroscience of relational reasoning. *Brain Research*, 1428, 13–23.

- Kreiner, H., & Eviatar, Z. (2014). The missing link in the embodiment of syntax: Prosody. *Brain and Language*, 137, 91–102.
- Kreitewolf, J., Friederici, A. D., & von Kriegstein, K. (2014). Hemispheric lateralization of linguistic prosody recognition in comparison to speech and speaker recognition. *NeuroImage*, 102, 332–344.
- Kümmerer, D., Hartwigsen, G., Kellmeyer, P., Glauche, V., Mader, I., Klöppel, S., ... Saur, D. (2013). Damage to ventral and dorsal language pathways in acute aphasia. *Brain*, 136(2), 619–629.
- Kyong, J. S., Scott, S. K., Rosen, S., Howe, T. B., Agnew, Z. K., & McGettigan, C. (2014). Exploring the roles of spectral detail and intonation contour in speech intelligibility: An fMRI study. *Journal of Cognitive Neuroscience*, 26(8), 1748–1763.
- Ladd, D. R. (2008). *Intonational phonology*. Cambridge: Cambridge University Press.
- Lazar, M., Weinstein, D. M., Tsuruda, J. S., Hasan, K. M., Arfanakis, K., Meyerand, M. E., ... Alexander, A. L. (2003). White matter tractography using diffusion tensor deflection. *Human Brain Mapping*, 18(4), 306–321.
- Lehiste, I. (1973). Phonetic disambiguation of syntactic ambiguity. *Glossa*, 7(2), 107–121.
- Loui, P., Alsop, D., & Schlaug, G. (2009). Tone deafness: A new disconnection syndrome? *The Journal of Neuroscience*, 29(33), 10215–10220.
- Makris, N., & Pandya, D. N. (2009). The extreme capsule in humans and rethinking of the language circuitry. *Brain Structure & Function*, 213(3), 343–358.
- Mauchly, J. W. (1940). Significance test for sphericity of a normal n-variate distribution. *The Annals of Mathematical Statistics*, 11, 204–209.
- McGettigan, C., & Scott, S. K. (2012). Cortical asymmetries in speech perception: What's wrong, what's right and what's left? *Trends in Cognitive Sciences*, 16(5), 269–276.
- Meixensberger, J., & Jaeger, M. (2005). Relevance of ICP and ptiO(2) for indication and timing of decompressive craniectomy in patients with malignant brain edema. *Acta Neurochirurgica*, 147(9), 952.
- Merrill, J., Sammler, D., Bangert, M., Goldhahn, D., Lohmann, G., Turner, R., & Friederici, A. D. (2012). Perception of words and pitch patterns in song and speech. *Frontiers in Psychology*, 3, 76.
- Meyer, M., Alter, K., Friederici, A. D., Lohmann, G., & von Cramon, D. Y. (2002). fMRI reveals brain regions mediating slow prosodic modulations in spoken sentences. *Human Brain Mapping*, 17(2), 73–88.
- Meyer, L., Cunitz, K., Obleser, J., & Friederici, A. D. (2014). Sentence processing and verbal working memory in a white-matter-disconnection patient. *Neuropsychologia*, 61, 190–196.
- Meyer, M., Steinhauer, K., Alter, K., Friederici, A. D., & von Cramon, D. Y. (2004). Brain activity varies with modulation of dynamic pitch variance in sentence melody. *Brain and Language*, 89(2), 277–289.
- Mori, S. (2007). *Introduction to diffusion tensor imaging*. Amsterdam: Elsevier.
- Nespor, M., & Vogel, I. (1986). *Prosodic phonology*. Dordrecht: Foris.
- Nishizaki, T., Ozaki, S., Kwak, T., & Ito, H. (1999). Clinical features and surgical outcome in patients with asymptomatic meningiomas. *British Journal of Neurosurgery*, 13(1), 52–55.
- Obleser, J., Eisner, F., & Kotz, S. A. (2008). Bilateral speech comprehension reflects differential sensitivity to spectral and temporal features. *The Journal of Neuroscience*, 28(32), 8116–8123.
- Parker, G. J. M., Luzzi, S., Alexander, D. C., Wheeler-Kingshott, C. A. M., Clecarelli, O., & Ralph, M. A. L. (2005). Lateralization of ventral and dorsal auditory-language pathways in the human brain. *NeuroImage*, 24(3), 656–666.
- Paulmann, S. (2016). The neurocognition of prosody. In G. Hickok, & S. L. Small (Eds.). *Neurobiology of Language* (pp. 1109–1120). Amsterdam: Elsevier.
- Peretz, I. (2016). Neurobiology of congenital amusia. *Trends in Cognitive Sciences*, 20(11), 857–867.
- Perkins, J. M., Baran, J. A., & Gandour, J. (1996). Hemispheric specialization in processing intonation contours. *Aphasiology*, 10(4), 343–362.
- Perrone-Bertolotti, M., Doherty, M., Laxenbruck, H., Sato, M., Pichat, C., & Baci, M. (2013). Neural correlates of the perception of contrastive prosodic focus in French: A functional magnetic resonance imaging study. *Human Brain Mapping*, 34(10), 2574–2591.
- Peschke, C., Ziegler, W., Eisenberger, J., & Baumgaertner, A. (2012). Phonological manipulation between speech perception and production activates a parieto-frontal circuit. *NeuroImage*, 59(1), 788–799.
- Pierrehumbert, J., & Hirschberg, J. (1990). The Meaning of Intonational Contours in the Interpretation of Discourse. In P. R. Cohen Morgan, J., & Pollack, M. E. (Eds.), *Intentions in Communication*. Cambridge: MIT Press.
- Plante, E., Creusere, M., & Sabin, C. (2002). Dissociating sentential prosody from sentence processing: Activation interacts with task demands. *NeuroImage*, 17(1), 401–410.
- Poeppel, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as “asymmetric sampling in time”. *Speech Communication*, 41(1), 245–255.
- Powell, H. W. R., Parker, G. J. M., Alexander, D. C., Symms, M. R., Boulby, P. A., Wheeler-Kingshott, C. A. M., ... Duncan, J. S. (2006). Hemispheric asymmetries in language-related pathways: A combined functional MRI and tractography study. *NeuroImage*, 32(1), 388–399.
- Rauschecker, J. P. (2011). An expanded role for the dorsal auditory pathway in sensorimotor control and integration. *Hearing Research*, 271(1–2), 16–25.
- Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nature Neuroscience*, 12(6), 718–724.
- Rockhill, J., Mrugala, M., & Chamberlain, M. C. (2007). Intracranial meningiomas: An overview of diagnosis and treatment. *Neurosurgical Focus*, 23(4), E1.
- Sammler, D., Grosbras, M.-H., Anwander, A., Bestelmeyer, P. E. G., & Belin, P. (2015). Dorsal and ventral pathways for prosody. *Current Biology*, 25(23), 3079–3085.
- Sammler, D., Kotz, S. A., Eckstein, K., Ott, D. V. M., & Friederici, A. D. (2010). Prosody meets syntax: The role of the corpus callosum. *Brain*, 133(9), 2643–2655.
- Saur, D., Kreher, B. W., Schnell, S., Kümmerer, D., Kellmeyer, P., Vry, M. S., ... Weiller, C. (2008). Ventral and dorsal pathways for language. *Proceedings of the National Academy of Sciences of the United States of America*, 105(46), 18035–18040.
- Scherer, K. R. (1986). Vocal affect expression – A review and a model for future research. *Psychological Bulletin*, 99(2), 143–165.
- Schirmer, A., & Kotz, S. A. (2006). Beyond the right hemisphere: Brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences*, 10(1), 24–30.
- Schlaug, G., Marchina, S., & Norton, A. (2009). Evidence for plasticity in white-matter tracts of patients with chronic Broca's aphasia undergoing intense intonation-based speech therapy. *Annals of the New York Academy of Sciences*, 1169, 385–394.
- Schönwiesner, M., Rüsamen, R., & von Cramon, D. Y. (2005). Hemispheric asymmetry for spectral and temporal processing in the human antero-lateral auditory belt cortex. *European Journal of Neuroscience*, 22(6), 1521–1528.
- Selkirk, E. (1984). *Phonology and syntax: The relation between sound and structure*. Cambridge, MA: MIT Press.
- Selkirk, E. (1996). Sentence prosody: Intonation, stress, and phrasing. In J. Goldsmith (Ed.), *Handbook of phonological theory* (pp. 550–569). New York: Blackwell Publishers.
- Shiple-Brown, F., Dingwall, W. O., Berlin, C. I., Yeni-Komshian, G., & Gordon-Salant, S. (1988). Hemispheric processing of affective and linguistic intonation contours in normal subjects. *Brain and Language*, 33(1), 16–26.
- Shokri-Kojori, E., Motes, M. A., Rypma, B., & Krawczyk, D. C. (2012). The network architecture of cortical processing in visuo-spatial reasoning. *Scientific Reports*, 2, 411.
- Skeide, M. A., Brauer, J., & Friederici, A. D. (2016). Brain functional and structural predictors of language performance. *Cerebral Cortex*, 26(5), 2127–2139.
- Snedeker, J., & Trueswell, J. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language*, 48(1), 103–130.
- Steinmann, S., & Mülert, C. (2012). Functional relevance of interhemispheric fiber tracts in speech processing. *Journal of Neurolinguistics*, 25(1), 1–12.
- Stokum, J. A., Gerzanich, V., & Simard, J. M. (2016). Molecular pathophysiology of cerebral edema. *Journal of Cerebral Blood Flow and Metabolism*, 36(3), 513–538.
- Strelnikov, K. N., Vorobyev, V. A., Chernigovskaya, T. V., & Medvedev, S. V. (2006). Prosodic clues to syntactic processing – A PET and ERP study. *NeuroImage*, 29(4), 1127–1134.
- Tang, C., Hamilton, L. S., & Chang, E. F. (2017). Intonational speech prosody encoding in the human auditory cortex. *Science*, 357(6353), 797–801.
- Thiebaut de Schotten, M., Dell'Acqua, F., Forkel, S. J., Simmons, A., Vergani, F., Murphy, D. G. M., & Catani, M. (2011). A lateralized brain network for visuospatial attention. *Nature Neuroscience*, 14(10), 1245–1246.
- Thiebaut de Schotten, M., Fyfe, D. H., Bizi, A., Dell'Acqua, F., Allin, M., Walshe, M., ... Catani, M. (2011). Atlas location, asymmetry and inter-subject variability of white matter tracts in the human brain with MR diffusion tractography. *NeuroImage*, 54(1), 49–59.
- Thorndike, R. M., & Thorndike-Christ, T. M. (2013). *Measurement and evaluation in psychology and education* (8th ed.). Harlow: Pearson Education Limited.
- van Lancker, D. (1980). Cerebral lateralization of pitch cues in the linguistic signal. *Papers in Linguistics*, 13(2), 201–277.
- Vassal, M., Le Bars, E., Moritz-Gasser, S., Menjot, N., & Duffau, H. (2010). Crossed aphasia elicited by intraoperative cortical and subcortical stimulation in awake patients. *Journal of Neurosurgery*, 113(6), 1251–1258.
- Wagner, M., & Watson, D. G. (2010). Experimental and theoretical advances in prosody: A review. *Language and Cognitive Processes*, 25(7–9), 905–945.
- Watson, C. E., & Chatterjee, A. (2012). A bilateral frontoparietal network underlies visuospatial analogical reasoning. *NeuroImage*, 59(3), 2831–2838.
- Watson, D. G., & Gibson, E. (2004). The relationship between intonational phrasing and syntactic structure in language production. *Language and Cognitive Processes*, 19(6), 713–755.
- Wechsler, D. (1987). *Wechsler memory scale – revised (WMS-R)*. San Antonio, CA: Psychological Corporation/Harcourt Brace Jovanovich.
- Wildgruber, D., Ackermann, H., Kreifelts, B., & Ethofer, T. (2006). Cerebral processing of linguistic and emotional prosody: fMRI studies. *Progress in Brain Research*, 156, 249–268.
- Wildgruber, D., Ethofer, T., Grandjean, D., & Kreifelts, B. (2009). A cerebral network model of speech prosody comprehension. *International Journal of Speech-Language Pathology*, 11(4), 277–281.
- Wilson, S. M., Galantucci, S., Tartaglia, M. C., Rising, K., Patterson, D. K., Henry, M. L., ... Gorno-Tempini, M. L. (2011). Syntactic processing depends on dorsal language tracts. *Neuron*, 72(2), 397–403.
- Witteman, J., Van Heuven, V. J. P., & Schiller, N. O. (2012). Hearing feelings: A quantitative meta-analysis on the neuroimaging literature of emotional prosody perception. *Neuropsychologia*, 50(12), 2752–2763.
- Witteman, J., van Ijzendoorn, M. H., van de Velde, D., van Heuven, V. J., & Schiller, N. O. (2011). The nature of hemispheric specialization for linguistic and emotional prosodic perception: A meta-analysis of the lesion literature. *Neuropsychologia*, 49(13), 3722–3738.
- Zarate, J. M. (2013). The neural control of singing. *Frontiers in Human Neuroscience*, 7, 237.
- Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: Music and speech. *Trends in Cognitive Sciences*, 6(1), 37–46.

2.3 Study III – Neural networks for musical harmony

Bianco, R., Novembre, G., Keller, P. E., Kim, S.-G., Scharf, F., Friederici, A. D., Villringer, A., **Sammler, D.** (2016). Neural networks for harmonic structure in music perception and action. *NeuroImage*, 142, 454-464.*

[Author contributions by D.S.: designed research with R.B., G.N., P.E.K.; supervised research and data analysis by R.B., S.-G.K. and F.S.; co-wrote the paper with R.B., incorporating edits by co-authors]

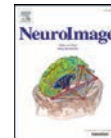
*Reprinted with permission from Elsevier.

NeuroImage 142 (2016) 454–464



Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Neural networks for harmonic structure in music perception and action

R. Bianco^{a,*}, G. Novembre^b, P.E. Keller^b, Seung-Goo Kim^a, F. Scharf^a, A.D. Friederici^a,
A. Villringer^a, D. Sammler^a^a Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany^b The MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Sydney, Australia

ARTICLE INFO

Article history:
Received 13 April 2016
Accepted 15 August 2016
Available online 16 August 2016

Keywords:
Music
Harmony
Syntax
IFG
Functional connectivity
Prediction

ABSTRACT

The ability to predict upcoming structured events based on long-term knowledge and contextual priors is a fundamental principle of human cognition. Tonal music triggers predictive processes based on structural properties of harmony, i.e., regularities defining the arrangement of chords into well-formed musical sequences. While the neural architecture of structure-based predictions during music perception is well described, little is known about the neural networks for analogous predictions in musical actions and how they relate to auditory perception. To fill this gap, expert pianists were presented with harmonically congruent or incongruent chord progressions, either as musical actions (photos of a hand playing chords) that they were required to watch and imitate without sound, or in an auditory format that they listened to without playing. By combining task-based functional magnetic resonance imaging (fMRI) with functional connectivity at rest, we identified distinct sub-regions in right inferior frontal gyrus (rIFG) interconnected with parietal and temporal areas for processing action and audio sequences, respectively. We argue that the differential contribution of parietal and temporal areas is tied to motoric and auditory long-term representations of harmonic regularities that dynamically interact with computations in rIFG. Parsing of the structural dependencies in rIFG is co-determined by both stimulus- or task-demands. In line with contemporary models of prefrontal cortex organization and dual stream models of visual-spatial and auditory processing, we show that the processing of musical harmony is a network capacity with dissociated dorsal and ventral motor and auditory circuits, which both provide the infrastructure for predictive mechanisms optimising action and perception performance.

© 2016 Elsevier Inc. All rights reserved.

1. Introduction

The brain shows a fine sensitivity to patterns and regularities that afford the prediction of incoming events in different domains (Tenenbaum et al., 2011). The theory of predictive coding (Friston, 2010) constitutes a unifying framework for human cognition and considers the brain as a “hypothesis tester” with the goal to optimise perception and action by constantly matching incoming sensory inputs with top-down predictions. Within a multi-level cascade of neural processes at different time scales, higher-level predictions act as priors for lower-level processes based on contextual information, previous exposure and acquired long-term knowledge. Recently, predictive coding theory has been used to explain predictions in the action domain (Kilner et al., 2007), as well as in music perception based on priors related to melodic (pitch) content (Pearce et al., 2010), metric structure (Vuust and Witek, 2014), or harmony (Rohrmeier and Koelsch, 2012). The present study takes a comparative stance on predictions in both

music perception and action, with a specific focus on Western tonal harmony.

Theoretical accounts refer to harmony as combinatorial arrangement of chords within musical sequences characterized by local and non-local dependencies (Swain, 1995). An instance of these dependencies is that a typical chord progression in Western tonal harmony starts and ends with a reference chord to which some chords are overwhelmingly likely to move to, while they rarely move to others (Tymoczko, 2003). Psychologically, these dependencies are predicted and perceived as tension-resolution patterns by listeners who have been sufficiently exposed to the prevailing musical system (Krumhansl, 1983; Lerdahl and Jackendoff, 1983). Convention in the field of music cognition has that the harmonic principles that govern musical structure are considered as part of a musical “syntax” (Bharucha and Krumhansl, 1983; Koelsch and Siebel, 2005; Patel, 2003), that also includes melodic and/or rhythmic principles of music (Large and Palmer, 2002; Rohrmeier and Koelsch, 2012). Here, we consider “syntax” generally as the knowledge of regularities that control the integration of smaller units into larger musical phrases (Swain, 1995) and thereby support predictions. It is well established that tacit knowledge about structural regularities of music 1) is acquired implicitly (Loui et al., 2009; Rohrmeier and

* Corresponding author at: Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstr. 1a, 04103 Leipzig, Germany.
E-mail address: bianco@chs.mpg.de (R. Bianco).

Rebuschat, 2012; Tillmann et al., 2000), 2) largely shapes our musical competence across different musical systems and cultures (Eerola et al., 2006; Lartillot and Ayari, 2011), and 3) enables listeners to cognitively link current auditory items to past events and to generate predictions on forthcoming events (Patel, 2003; Tillmann, 2012). In the present study we will focus on harmonic regularities and investigate how they govern predictions during (auditory) music perception and (silent) musical actions.

Harmony not only defines the sequence of musical sounds but also co-determines the associated chain of musical actions. Therefore, the implicit knowledge of harmonic regularities might influence not only listeners' predictions, but also musicians' action planning during performance (Palmer and van de Sande, 1995). While regularity-based predictions during music listening have already been thoroughly investigated (Rohmeier and Koelsch, 2012; Tillmann, 2012), the neural basis of motor predictions in musical actions has not been explored in depth (Maidhof et al., 2009; Ruiz et al., 2009). Recent behavioural (Novembre and Keller, 2011) and electrophysiological studies on music production (Bianco et al., 2016; Sammler et al., 2013b) revealed slower response times, higher number of errors and neural processing costs (a centro-parietal negativity) in expert pianists when asked to silently execute harmonically incongruent compared to congruent chord progressions. These costs were associated with the motor reprogramming of a pre-planned, congruent, action in face of an unexpected incongruity, and were taken as indirect evidence that pianists' action planning was based on musical context and internalised knowledge of harmony. In other words, these findings imply that harmonic structure might implicitly regulate mechanisms of motor control to improve music performance beyond fine movement optimization (Bianco et al., 2016; Novembre and Keller, 2011).

The goal of the present study is to identify the brain areas involved in motor planning based on the regularities of Western tonal harmony, to explore the connectivity between these areas and to compare this network with the neural network sub-serving analogous processes in auditory music perception. The rationale behind this study is that expert pianists have internalised the rules of harmony not only auditorily but also in the hand action domain. Their substantial motor training should enable them to parse harmonic dependencies also in sequences of silent musical actions to facilitate prediction and planning of forthcoming motor acts during performance. This is because the same harmonic structure in sequences of sounds or sequences of actions without sound (i.e., those movements typically employed for producing these sounds) should trigger cognitive processes that are analogous with regard to the structural information. At the same time, processing should differ between perception and action with regard to the associated sensory and memory retrieval processes (i.e., auditory sound vs. motoric act). Here, we sought to isolate and compare the neural networks involved in harmony processing during either perception or (silent) actions, i.e. to probe the potential contribution of auditory and motor prediction of harmony that are otherwise co-occurring during real music production.

Neural hypotheses for musical syntax processing (i.e., harmony) in music perception (Koelsch, 2011; Patel, 2003; Tillmann, 2012) posited a special role of frontal computational regions that successively integrate incoming information into higher-order structures by drawing on knowledge about regularities stored in posterior brain regions. Neuroimaging research points to the inferior frontal gyrus (IFG) as the critical computational area that, together with a repository of regularities in posterior auditory regions superior temporal gyrus (STG), affords the prediction of future musical sounds based on the context and listener's long-term music structural knowledge (Kim et al., 2011; Koelsch et al., 2005; Maess et al., 2001; Musso et al., 2015; Sammler et al., 2011; Tillmann et al., 2006). Interestingly, IFG has been associated not only with structural integration and prediction of musical sequences, but also with structuring of complex actions (Fuster, 2001; Koehlin and Summerfield, 2007) outside the music domain. Lesions of the left IFG

cause impairment in sequencing pictures representing human actions (Fazio et al., 2009), and bilateral IFG are involved in evaluating whether constituent acts belong to the same or separate sub-goals (Farag et al., 2010). Moreover, bilateral IFG activations have been reported during execution of series of motor acts that were organised according to hierarchical action plans (Koehlin and Jubault, 2006). In sum, IFG has become central to hypotheses on processing of structured sequential information in perception and action (Fitch and Martins, 2014; Fiebach and Schubotz, 2006, for various perspectives see Cortex, 2006, vol.2, issue 42), making it conceivable that IFG is also involved in parsing and predicting structural information embedded in musical actions.

What has received less attention than the role of IFG, however, is its interaction with task-relevant posterior systems of knowledge during structural processing. In other words, apart from frequently reported co-activations of IFG and auditory temporal regions during music listening (Koelsch and Siebel, 2005), the characterization of other 'modality-specific regions', e.g., in musical action, and particularly their connectivity with frontal 'computational regions' remains uncertain. In this study, we tested whether pianists' action planning based on knowledge of Western tonal harmony involves (i) IFG in interaction with (ii) posterior visual-motor areas. Furthermore, we (iii) compared the functional connectivity profiles of IFG during the processing of musical actions and auditory sequences that contained similar harmonic violations.

We acquired resting state fMRI data from expert pianists, and then fMRI data during an audio and an action task in which the same harmonic sequences were either auditorily presented or had to be motorically imitated. In the audio task, pianists listened to 5-chord sequences (similar to Koelsch et al., 2005) in which the last chord was either harmonically congruent or incongruent with the preceding musical context. In the action task, in total absence of musical sound, participants were presented with series of photos of a pianist's hand performing the same congruent/incongruent chord progressions on a piano (Bianco et al., 2016). To engage the motor system in the processing of musical actions, pianists had not only to watch the movements, but also to manually reproduce them on a glass-board. The contrasts of incongruent minus congruent chords during listening or imitation were used to functionally segregate modality-specific areas and to isolate frontal computational areas. To demonstrate crosstalk between these regions, we used the latter as seeds in a functional connectivity analysis of the resting state fMRI data.

If harmonic violations of audio sequences activate IFG, then violations of action sequences with the same musical structure should also activate IFG as parser of harmonic regularities and top-down generator of predictions. On the other hand, we expected to find divergent activity in temporal auditory or parietal visual-motor regions associated with item identification and storage of knowledge in their modality-specific format. Finally, by mirroring task-based activation (Smith et al., 2009), the resting-state data should reveal processing streams involved in processing harmonic regularities in music perception and action.

2. Materials and methods

2.1. Participants

29 pianists (17 female) aged 20–32 years (mean age: 24.7, SD = 2.9) took part in the experiment. They had a minimum of 5 years of piano training in classical Western tonal music (range = 5–27 years, mean years of training = 17.2, SD = 4.8) and had started to play the piano at an average age of 7.3 years (SD = 3.08). None of the pianists had training in improvisation or other musical styles. All participants were naïve with regard to the purpose of the study. Written informed consent was obtained from each participant before the study that was approved by the local ethics committee.

2.2. Stimuli

Stimuli (see Fig. 1) consisted of 60 different chord sequences that were presented as piano sounds in the listening task (similar to Koelsch, 2005), and as photos of a hand playing chords on a piano in the action imitation task (Bianco et al., 2016). The sequences were composed of 5 chords according to the rules of classical harmony and had various melodic contours. The first chord always represented the tonic (based on the first degree of the scale in the relevant musical key). The second chord could be tonic, mediant (based on the third scale degree) or subdominant (based on the fourth scale degree). Chords at the third position were subdominant, dominant, or dominant six-four chords, and chords at the fourth position were dominant seventh chords. At the last position, the target chord of each sequence was manipulated in terms of harmonic congruency (CONG), so that the last chord could be either congruent (a Tonic chord typically used to resolve a musical sequence) or incongruent (a Neapolitan chord that sounds normal when played in isolation but constitutes a violation when used at the end of a standard harmonic progression). Both the Tonic and Neapolitan are consonant major chords built on the 1st and lowered 2nd scale degree, respectively (i.e., A for Tonic and Bb for Neapolitan in A-major). Consequently, and due to the relationship of the tonalities within the circle of fifths, the exact same chord that acts as a Tonic in one tonality (e.g., A – #C – E in A-major), acts as a Neapolitan in another tonality (i.e., Bbb (=A) – Db (= #C) – Fb (=E) in Ab-major). We exploited this relationship and presented five sequences from each of six different tonalities (D, E, Bb, Ab, A and Eb major), such that the majority of final chords were presented as both Tonic and Neapolitan across the experiment. Therefore, potential neural differences in processing congruent and incongruent chords cannot be due to chord identity but more likely reflect harmony-related processes. With regard to the stimuli of the action block, the choice of different tonalities further allowed us to balance the visual appearance (i.e., number of black/white keys) and difficulty of execution of the target chord (i.e., movement distance from second last to target chord) in congruent and incongruent conditions (for visual appearance: average of 1.3 ± 0.5 black keys in the congruent and 1.2 ± 0.7 in the incongruent chords; for difficulty: average of 1 ± 0 key distance in the congruent and 1.5 ± 0.2 in the incongruent sequences). All sequences were played with normal fingering that was rated as being similarly conventional for congruent and incongruent endings (see Bianco et al., 2016).

The audio stimuli were created with Logic Pro 8 (Apple Inc.), normalised for loudness (RMS, root mean square) with Adobe Audition CS 6 and had a total duration of 6 s (1 s for each of the first four chords and 2 s for the target chord). In the action block, the same chord sequences were presented as photo series showing a male pianist's right hand pressing three keys forming each chord on a piano in conventional

fingering (Yamaha Clavinova CLP150, Yamaha Music Europe GmbH, Rellingen, Germany). Red circles were superimposed on top of each pressed key (cf. Bianco et al., 2016) for the whole duration of the photo to facilitate the recognition of the pressed keys. Each photo was presented for 2 s (total sequence duration: 10 s).

2.3. Procedure

The experimental session started with 14 min resting state fMRI data acquisition in which participants were instructed to keep their eyes open and not to fall asleep. To prevent any task-related bias in the measures of functional connectivity, pianists were asked not to practice piano on the scanning day. Thereafter, the task session started and lasted for approximately 25 min.

In the scanner, participants were required to imitate musical actions or to listen to musical sequences in two separate blocks with a counterbalanced order across the group (Fig. 1). Stimulus presentation was controlled in an event-related design with Presentation software (version 14.9, Neurobehavioural Systems, Inc.). In both blocks, congruent and incongruent sequences were intermixed in a way that no >3 sequences of the same condition followed each other. The inter-trial interval (ITI) ranged from 3 to 9 s and during this period participants saw a black screen. During action imitation, no sound was played. Pianists were asked to watch the performing hand in the photos and to simultaneously copy the presented hand postures on a 5×15 cm² glass-board with their right hand (Fig. 1, left panel). To motivate participants to follow the sequence accurately, they were told that their performance was monitored with a camera (MR-compatible camera, 12M camera, MRC Systems, Heidelberg Germany). Since it is a common way for pianists to mentally practice by motorically simulating piano performance, the playing along was meant to maximally involve the motor system during the processing of musical actions. In the audio block, pianists were asked to carefully listen to the sequences without playing along (Fig. 1, right panel).

Only to ensure that participants paid attention to the stimuli and to assess their awareness of the violations, 10 trials (1/6 of the trials) in both tasks were followed by a prompt that asked participants to judge the harmonic correctness of the last presented sequence. The judgement required a button response performed with the index or middle finger of the left hand (key assignment was counterbalanced across participants). These judgement trials were equally distributed over congruent and incongruent trials, and required 50% yes and 50% no responses.

After the scanning session, participants filled out a questionnaire to assess the degree to which they had imagined the sound of the chord sequences during the action imitation task and the movements to produce the sequences during the listening task.

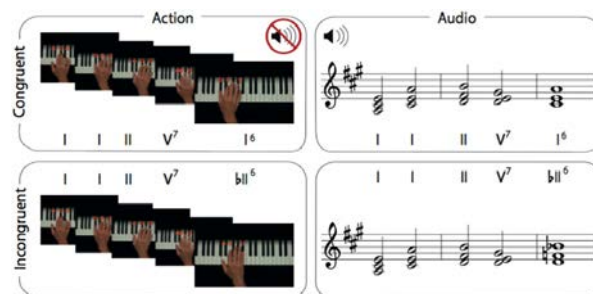


Fig. 1. Experimental design: expert pianists were presented with harmonically congruent or incongruent chord progressions, presented either as muted musical actions (photos of a hand playing chords) that they were required to imitate on a glass-board (left panel), or in an auditory format that they listened to (right panel).

2.4. Data acquisition

The experiment was carried out on a 3.0-Tesla Siemens TIM Trio whole body magnetic resonance scanner (Siemens AG, Erlangen, Germany) using a 32-radiofrequency-channel head coil. Functional magnetic resonance images were acquired using a T2*-weighted 2D echo planar imaging (EPI) sequence. During 14 min of acquisition (TE = 36.5 ms, TR = 1400 ms) at rest (eyes open, instructed not to fall asleep) 410 volumes were acquired with a square FOV of 64 axial slices of 2.3 mm thickness and no gap ($2.3 \times 2.3 \times 2.3$ mm³ voxel size) with a flip angle of 69°. Functional images during the two tasks were acquired using an EPI sequence with TE = 30 ms and TR = 2000 ms. 456 and 377 volumes were acquired in the action and audio block, respectively, with a square FOV of 210 mm, with 37 interleaved slices of 3.2 mm thickness and 15% gap ($3 \times 3 \times 3.68$ mm³ voxel size) aligned to the AC-PC plane, and a flip angle of 77°. For anatomical registration, high-resolution T1-weighted images were acquired using a 3D MP2RAGE sequence (T1₁ = 700 ms, T1₂ = 2500 ms, TE = 2.03 ms, TR = 5000 ms) with a matrix size of $240 \times 256 \times 176$, with 1 mm isotropic voxel size, flip angle₁ of 4°, flip angle₂ of 8°, and GRAPPA acceleration factor of 3.

2.5. Data analysis

2.5.1. Task-based fMRI

fMRI data of 29 participants were analysed with statistical parametric mapping (SPM8; Wellcome Trust Centre for Neuroimaging; <http://www.fil.ion.ucl.ac.uk/spm/software/spm8/>) using standard spatial pre-processing procedures. These consisted of: slice time correction (by means of cubic spline interpolation method), spatial realignment, co-registration of functional and anatomical data (uniform tissue-contrast image masked with the 2nd inversion image from the MP2RAGE sequence), spatial normalisation into the MNI (Montreal Neurological Institute) stereotactic space, that included resampling to $2 \times 2 \times 2$ mm voxel size. Finally, data were spatially low-pass filtered using a 3D Gaussian kernel with full-width at half-maximum (FWHM) of 8 mm and temporally high-pass filtered with a cut-off of 1/128 Hz to eliminate low-frequency drifts.

Statistical parametric maps for the whole brain data were generated in the context of the general linear model (GLM) separately for the action imitation and the listening task. The evoked hemodynamic response to the onset of the final chord was modelled for the congruent and incongruent conditions as boxcars convolved with a hemodynamic response function (HRF). To this design, we added estimated motion realignment parameters as covariates of no interest to regress out residual motion artefacts and increase statistical sensitivity. To identify hemodynamic responses related to the processing of harmonic violations, we computed the first level contrast CONG (i.e., incongruent > congruent chords), separately for the action imitation and the listening task. For random effects group analyses, the resulting contrast images were submitted to one-sample *t*-tests. Additionally, to identify areas that are modality-specific to either action or audio representation of the harmonic structure, we compared the CONG contrasts of the two tasks by means of paired *t*-tests. We controlled family-wise error rate (FWER) of clusters below 0.05 with a cluster-forming height-threshold of 0.001. Anatomical labels are based on Harvard-Oxford cortical structural atlas implemented in FSL (<http://neuro.debian.net/packages/fsl-harvard-oxford-atlases.html>).

2.5.2. Resting-state fMRI

In order to investigate intrinsic connectivity of the peak regions from task-based fMRI datasets (Bressler and Menon, 2010), independent resting state fMRI datasets were obtained from 28 of the pianists that participated in the task-fMRI session (one r-fMRI data set was not acquired due to technical problems). The pre-processing of the resting state data (realignment, unwarping, slice-timing correction) was done

using SPM8 by means of DPARSF (<http://rfmri.org/DPARSF>) SPM-based toolboxes. We applied a GLM to regress out non-neuronal signal changes due to physiological noise and, most importantly, head motions. The regressors included six rigid-body motion parameters, five principle components extracted by the “anatomical CompCor” (Behzadi et al., 2007) (i.e., signal from white matter and cerebral fluid masks defined from anatomical scans), and finally global signal (Power et al., 2015). Thereafter, band-pass-filtering (0.009 and 0.08 Hz), spatial normalisation of functional data into MNI stereotactic space (with resampling to $2 \times 2 \times 2$ mm³ resolution), and finally a minimal spatial smoothing with the FWHM of 3 mm were applied to the residual time-series.

Resting-state functional connectivity (RSFC) was defined by Pearson's correlations between a time-series of a seed region and time-series of whole brain voxels. Spherical seed regions (5 mm radius) were centred in IFG at the peak coordinates of the CONG contrasts obtained in the task-based analyses of the action imitation and the listening task. In order to match the smoothness of noise in task-based and resting-state analyses, the correlation maps were further smoothed with the FWHM of 2 mm, resulting in an effective FWHM of about 8 mm.

Voxel-wise paired *t*-tests were performed to identify differences between the two seed-based correlation maps (i.e., action and audio seeds). The normality assumption based upon the difference between the two correlation coefficients across subjects was fulfilled, as confirmed by Kolmogorov-Smirnov tests. We controlled FWER of clusters below 0.05 with a cluster-forming height-threshold of 0.001 in all reported results. Harvard-Oxford cortical structural atlas was used to assign anatomical labels.

3. Results

3.1. Behavioural

To ensure that participants paid attention to the stimuli in both modalities and that they were generally able to recognise the harmonic structure underlying the sequences, they were required to overtly judge harmonic congruency in 1/6 of the trials. They performed significantly above chance level in these explicit judgments both in the action (mean \pm SD: 68.96 \pm 27.06% correct, $p < 0.001$) and in the audio block (mean \pm SD: 91.03 \pm 16.40% correct, $p < 0.001$), as tested with one-sample *t*-tests against 50% chance level. Action block performance was lower than audio block performance ($t(28) = 4.704$, $p < 0.001$), partly due to a response bias towards “congruent” answers in the action block (32% of incongruent trials misclassified as congruent vs. 15% of congruent trials misclassified as incongruent: $t(28) = -2.95$, $p = 0.007$). These differences in explicit judgement might indicate that pianists were less consciously aware of the harmonic violations during action imitation than during passive listening, possibly because it is more taxing to copy sequences of actions on-line than to just listen. This may have led pianists to focus on the motor-executive task rather than harmonic relationships in the action block (see Discussion).

3.2. fMRI

3.2.1. Fronto-parietal vs. fronto-temporal areas for musical action vs. perception

In the action imitation task, the CONG contrast (incongruent vs. congruent chords) yielded larger hemodynamic responses in frontal and parietal areas, comprising the dorsal portion of rIFG (BA44) bordering precentral sulcus, and bilateral clusters extending from superior parietal cortex (SPL: BA7) to the inferior parietal and middle occipital gyrus (MOG: BA19) (Fig. 2 left-upper panel, Table 1).

In the listening task, the same contrast evoked stronger activity in frontal and temporal areas, including right IFG (BA44/45, peak in ventral

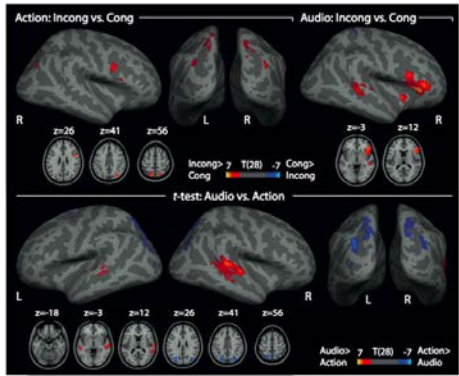


Fig. 2. Harmonic violations elicited activations in fronto-parietal areas during action imitation (upper left panel) and in fronto-temporal areas during listening (upper right panel). Areas involved in structural processing specifically for the action and the audio sequences were identified in bilateral posterior parietal regions (cold colours) and in bilateral temporal regions (hot colours), respectively (lower panel).

BA45) and the right posterior superior temporal gyrus and sulcus (pSTG/STS: BA22) (Fig. 2 right-upper panel, Table 1).

To identify areas exclusively recruited depending on stimulus format (photos of actions or audio), the CONG contrasts of both tasks were compared using a paired *t*-test (Table 2). Incongruent actions elicited greater activity in bilateral SPL, MOG, and in the left superior/middle frontal gyrus (frontal eye fields, FEF: BA8) (Fig. 2 lower panel, cold colours). Conversely, auditory violations yielded larger BOLD responses in bilateral STS/STG, compared to the action task (Fig. 2 lower panel, hot colours).

To identify areas commonly recruited during both audio and action task, we masked the audio CONG contrast with the action contrast. This analysis yielded a cluster in rIFG (BA44, $x = 64$, $y = 18$, $z = 24$, cluster extent = 28 voxels, $Z = 3.53$, $p_{\text{voxel}} < 0.001$) that, however, did not survive the cluster-level FWER correction.

Table 1
Congruency effect (incongruent > congruent) in the action imitation and listening tasks.

| Region | Hem. | BA | k | x | y | z | Z-value |
|--|------|--------|------|-----|-----|-----|---------|
| <i>Action: CONG incongruent > congruent</i> | | | | | | | |
| Precentral/inferior frontal gyrus | R | 44 | 182 | 44 | 6 | 26 | 4.29 |
| | | 44 | | 64 | 18 | 24 | 3.53 |
| | | 44 | | 54 | 14 | 16 | 3.49 |
| Middle occipital gyrus | R | 19 | 352 | 40 | −80 | 36 | 4.66 |
| Superior parietal lobe | R | 7P | | 32 | −78 | 42 | 4.32 |
| | | 7P | | 16 | −70 | 58 | 3.54 |
| Superior parietal lobe | L | 7P | 510 | −16 | −74 | 58 | 4.27 |
| | | 7P | | −20 | −70 | 50 | 3.97 |
| Middle occipital gyrus | L | 19 | | −30 | −80 | 34 | 3.72 |
| <i>Audio: CONG incongruent > congruent</i> | | | | | | | |
| Inferior frontal gyrus | R | 45 | 1667 | 44 | 34 | 2 | 5.12 |
| | | 45 | | 44 | 18 | 16 | 4.98 |
| | | Insula | | 36 | 10 | −2 | 4.49 |
| Superior temporal sulcus, post. | R | 22 | 256 | 48 | −32 | 0 | 3.92 |
| Superior temporal gyrus, post. | R | 22 | | 70 | −24 | 6 | 3.59 |
| | | 22 | | 60 | −34 | 8 | 3.46 |
| Cerebellum (crus II) | L | − | 132 | −14 | −76 | −36 | 4.37 |

Whole-brain activation cluster sizes (k), MNI coordinates (x, y, z), and Z-scores for the CONG contrast in action imitation and listening tasks ($p_{\text{voxel}} < 0.001$; $p_{\text{cluster}} < 0.05$, FWE corrected). BA: Brodmann area, Hem.: hemisphere, post.: posterior.

Table 2
t-test between the CONG contrasts in the action imitation and listening task.

| Region | | Hem. | BA | k | x | y | z | Z-value |
|--------------------------------|---|-------|------|-----|-----|-----|------|---------|
| <i>Action > audio</i> | | | | | | | | |
| Superior parietal lobe | R | 7 | | 806 | 26 | −76 | 46 | 4.46 |
| | | 7 | | | 24 | −58 | 44 | 4.22 |
| Middle occipital gyrus | R | 19 | | | 34 | −80 | 38 | 4.08 |
| Middle occipital gyrus | L | 19 | 1436 | | −32 | −76 | 24 | 4.99 |
| Superior parietal lobe | L | 7 | | | −30 | −74 | 36 | 4.75 |
| | | 7 | | | −20 | −72 | 48 | 4.41 |
| Superior frontal gyrus | L | 8 | 186 | | −20 | 2 | 60 | 3.97 |
| | | 8 | | | −22 | −2 | 72 | 3.60 |
| <i>Audio > action</i> | | | | | | | | |
| Superior-middle temporal gyrus | R | 21/22 | 1100 | 64 | −16 | 8 | 4.79 | |
| | | | | 62 | −32 | 4 | 4.64 | |
| | | | | 52 | −30 | −8 | 4.20 | |
| Superior-middle temporal gyrus | L | 21/22 | 166 | −52 | −32 | −2 | 4.11 | |
| | | | | −52 | −16 | 2 | 3.91 | |
| | | | | −62 | −30 | 6 | 3.51 | |

Whole-brain activation cluster sizes (k), MNI coordinates (x, y, z), and Z-scores for the paired-samples *t*-test comparison of the action > audio and audio > action CONG contrast ($p_{\text{voxel}} < 0.001$; $p_{\text{cluster}} < 0.05$, FWE corrected). BA: Brodmann area, Hem.: hemisphere.

3.2.2. Seed-based functional connectivity from the IFG peak maxima

To gather evidence for the communication between IFG and modality-specific areas, we conducted a seed-based functional connectivity analysis on the resting-state fMRI data acquired from the same pianists. The activation peaks in the IFG clusters in the action imitation and audio task were chosen as seed regions, which were located between the right dorsal BA44 and the pre-central sulcus (action-seed) and in right BA45 (audio-seed). The results are depicted in Fig. 3 (upper and middle panels) and show positive functional connectivity (hot colours) between IFG and (amongst others) regions that were functionally specific to the action or audio musical task.

In line with activity in IFG and posterior parietal regions in the action task, the action-seed in IFG (BA44) exhibited positive correlations within a dorsal motor network comprising bilateral parietal cortex, extending from the anterior ventral supramarginal gyrus (BA40) to the posterior superior parietal lobes (BA7) (Table 3). Notably, there were no significant correlations with temporal regions that were specific to the audio modality. A large cluster peaking in bilateral precentral gyrus showed positive correlations with the action-seed, including subclusters in bilateral BA44 extending to insular regions, ventral premotor cortex (BA6), middle frontal gyrus (BA9) bordering the superior frontal gyrus and the inferior portion of the frontal pole (BA10). Medially, the action seed exhibited positive correlations with the right posterior border of the supplementary motor cortex (BA6) and anterior cingulate (BA24). Finally, there were positive correlations with right inferior temporal gyrus at the temporo-occipital junction (BA20), and bilateral occipitotemporal areas (BA37), cerebellum and thalamus.

Consistent with activity in IFG and temporal areas in the audio task, the audio-seed in IFG (BA45) exhibited positive correlations within the auditory network comprising the posterior part of the right superior temporal gyrus (BA22) and left Heschl's gyrus (including BA41/42) (Table 3). Additionally, there were positive correlations with frontal areas in the right hemisphere including orbitofrontal (BA47/11/12) and frontopolar regions (BA10), superior (BA8), middle frontal areas (BA9), and anterior cingulate gyrus (BA24), and in the left hemisphere, including BA45, BA47, BA9, BA10, and BA12. In the parietal cortex, positive correlations were restricted to bilateral anterior ventral supramarginal gyrus (BA40), without extending to more posterior parietal regions. Finally, there were positive correlations with thalamus and right putamen.

Apart from positive correlations, activity in both action and audio seeds was negatively correlated (Fig. 3 upper and middle panels, cold colours) with activity in areas belonging to the default mode network (DMN), namely the cingulate gyrus and the superior portion of bilateral

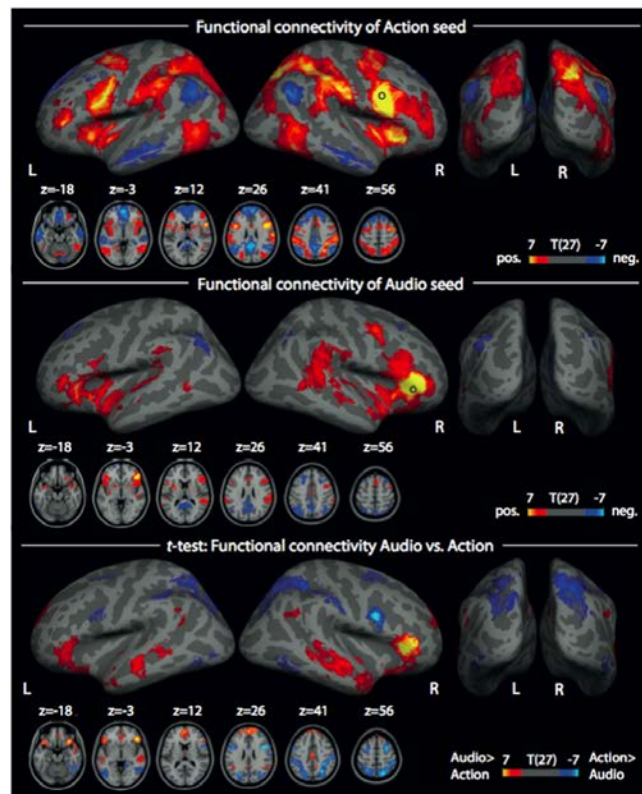


Fig. 3. Upper and middle panels: seed-based functional connectivity maps of resting-state data from the action seed in dorsal BA44 and audio seed in BA45, respectively. Seeds are depicted as black circles. Hot and cold colours indicate positive and negative functional connectivity, respectively. Consistent with the task-based activations, the topographical connectivity patterns include posterior parietal regions from the action-seed and temporal regions from the audio-seed. Lower panel: *t*-test between the connectivity maps of the action- and audio seed. Correlation values in posterior parietal areas were significantly higher for the action-seed than the audio-seed, whereas temporal regions were more strongly correlated to the audio- than the action-seed.

lateral occipital cortex extending into angular gyrus. These regions typically show a decrease of activation during attention-demanding tasks and goal-directed behaviours (Uddin et al., 2009). Additionally, negative correlations were found between the action-seed and bilateral anterior middle temporal gyrus and medial prefrontal cortex, anti-correlations that have been associated with highly difficult goal-directed tasks, as could apply in the case of our action-task (McKiernan et al., 2003).

Finally, a paired-samples *t*-test comparing the connectivity maps of the action- and audio-seed (Table 4) confirmed their differential predominant connectivity to parietal and temporal areas, respectively. Specifically, connectivity of the action-seed (compared to audio-seed) was stronger to bilateral posterior parieto-occipital areas, as well as to bilateral cerebellum, right frontal pole, frontal medial cortex and anterior cingulate gyrus, left superior frontal and precentral gyrus. Conversely, the connectivity of the audio-seed (compared to action-seed) was stronger to bilateral superior or middle temporal gyrus, as well as to bilateral cerebellum, right superior frontal gyrus, posterior cingulate and angular gyrus, thalamus, and left frontal operculum.

4. Discussion

The present study investigated the neural bases of action planning and prediction based on long-term knowledge of harmonic regularities and compared them with those involved in auditory prediction. Functional neuroimaging data of expert pianists were acquired at rest, during imitation of (without sound) or listening to (without imitation) harmonically congruent or incongruent chord sequences presented as photos of musical actions or sounds, respectively. Violations in both musical actions and sounds recruited distinct sub-regions (BA 44 and BA 45, respectively) in right IFG (rIFG) interconnected with parietal visual-motor and temporal auditory areas, respectively. We propose that motoric and auditory long-term representations of harmonic regularities are likely to account for the differential involvement of parietal and temporal areas that enter into dynamic interactions with computations in rIFG. Moreover, the involvement of rIFG in parsing musical action and sound sequences is sensitive to stimulus properties and task – production or perception – accounting for the divergent peak localizations, in line with prevailing models of general prefrontal cortex organization

Table 3
Resting-state functional connectivity from the action and audio seed in right inferior frontal gyrus.

| Region | BA | Action-seed | | | | | Audio-seed | | | | |
|---|----------|-------------|-----|-----|-----|---------|------------|-----|-----|-----|---------|
| | | k | x | y | z | Z-value | k | x | y | z | Z-value |
| <i>Right hemisphere (positive correlations)</i> | | | | | | | | | | | |
| Frontal pole | 10 | 125 | 26 | 38 | −16 | 4.94 | | | | | |
| Middle frontal gyrus | 9 | | | | | | 310 | 42 | 4 | 46 | 5.22 |
| Superior frontal gyrus | 8 | | | | | | 349 | 4 | 18 | 60 | 4.79 |
| Supplementary motor cortex | 6 | 734 | 6 | 14 | 52 | 5.92 | | | | | |
| Cingulate gyrus, ant. | 24 | 133 | 4 | 6 | 28 | 5.76 | 324 | 4 | 32 | 22 | 4.86 |
| Supramarginal gyrus/superior parietal lobe | 40/7 | 5418 | 52 | −30 | 48 | 6.89 | | | | | |
| Superior temporal gyrus, post. | 22 | | | | | | 119 | 50 | −14 | −8 | 4.33 |
| Superior temporal gyrus, post. | 22 | | | | | | 1148 | 52 | −30 | 6 | 4.72 |
| Middle temporal gyrus | 20/21/37 | 1304 | 52 | −56 | −12 | 5.92 | | | | | |
| Putamen | − | | | | | | 36 | 32 | −12 | −8 | 4.52 |
| Thalamus | − | 120 | 8 | −14 | 8 | 5.84 | | | | | |
| Cerebellum (VIIb) | − | 350 | 18 | −68 | −48 | 5.36 | | | | | |
| Cerebellum (VI) | − | 95 | 8 | −70 | −22 | 4.98 | | | | | |
| <i>Left hemisphere (positive correlations)</i> | | | | | | | | | | | |
| Frontal pole | 10 | 531 | −44 | 38 | 8 | 6.28 | | | | | |
| Frontal operculum cortex | 45 | | | | | | 2198 | −38 | 26 | 0 | 6.29 |
| Inferior frontal gyrus | 44 | | | | | | 50 | −46 | 12 | 22 | 4.29 |
| Precentral gyrus | 6/44/Ins | 2415 | −44 | 6 | 24 | 6.89 | | | | | |
| Middle frontal gyrus | 9 | 626 | −26 | 0 | 50 | 6.15 | | | | | |
| Cingulate gyrus, ant. | 24 | | | | | | 52 | −2 | −12 | 42 | 3.98 |
| Heschl's gyrus (H1 and H2) | 41/42 | | | | | | 173 | −52 | −14 | 4 | 4.07 |
| Supramarginal gyrus/superior parietal lobe | 40/7 | 3814 | −60 | −30 | 42 | 5.92 | 143 | −66 | −38 | 26 | 4.21 |
| Middle temporal gyrus | 20/21/37 | 1162 | −60 | −60 | −6 | 5.60 | 63 | −62 | −60 | 8 | 4.03 |
| Thalamus | − | 116 | −12 | −14 | 6 | 4.45 | 93 | −6 | −14 | 2 | 5.43 |
| Cerebellum (VIIb) | − | 969 | −26 | −66 | −52 | 6.06 | | | | | |
| Cerebellum (VI) | − | 95 | −22 | −62 | −28 | 4.98 | | | | | |
| Cerebellum (crus II) | − | | | | | | 58 | −16 | −78 | −34 | 4.40 |
| <i>Right hemisphere (negative correlations)</i> | | | | | | | | | | | |
| Frontal pole | 10 | 11,860 | 6 | 60 | 22 | 7.08 | | | | | |
| Superior frontal gyrus | 8 | | | | | | 384 | 28 | 30 | 54 | 4.97 |
| Middle temporal gyrus | 21 | 2367 | 60 | 4 | −24 | 6.88 | | | | | |
| Cingulate gyrus, post. | 24 | 5798 | 10 | −50 | 34 | 7.66 | 3682 | 10 | −46 | 12 | 5.35 |
| Cerebellum (IX) | − | 187 | 4 | −50 | 44 | 5.81 | | | | | |
| Cerebellum (crus I) | − | 1799 | 26 | −88 | −30 | 5.99 | 326 | 36 | −52 | −34 | 4.99 |
| <i>Left hemisphere (negative correlations)</i> | | | | | | | | | | | |
| Frontal pole | 10 | | | | | | 105 | −20 | 64 | −6 | 4.37 |
| Superior frontal gyrus | 8 | | | | | | 601 | −20 | 28 | 38 | 5.67 |
| Middle temporal gyrus | 21 | 2367 | −62 | −24 | −12 | 6.88 | | | | | |
| Inferior temporal gyrus, post. | 20 | | | | | | 70 | −60 | −44 | −14 | 4.17 |
| Lateral occipital cortex, sup. | 39 | 1764 | −40 | −50 | 26 | 6.27 | 1268 | −36 | −66 | 38 | 5.57 |
| Hippocampus | − | 62 | −34 | −34 | −8 | 5.01 | 65 | −30 | −34 | −12 | 4.07 |
| Cerebellum (crus I) | − | 27 | −44 | −56 | −42 | 4.275 | | | | | |

Results of the whole-brain functional connectivity analysis from IFG activation maxima in action imitation and listening tasks. k: cluster size, MNI coordinates (x, y, z), and Z scores. ($p_{\text{voxel}} < 0.001$; $p_{\text{cluster}} < 0.05$, FWE corrected), BA: Brodmann area, ant.: anterior, post.: posterior, sup.: superior.

(e.g., Fuster, 2001), and dual stream models of the visuo-spatial (e.g., Goodale and Milner, 1992) and auditory system (e.g., Rauschecker and Scott, 2009). Altogether, our results emphasise dissociable, neural action and audio networks in which modality-specific long-term knowledge and contextual information act as priors for the prediction of forthcoming events. In this respect, predictive coding models (Friston, 2010) may yield a unifying explanatory framework for information processing across both action and perception.

4.1. Musical action

The imitation of incongruent actions elicited activations in fronto-parietal areas (see Table 1), including the right inferior frontal gyrus (IFG; dorsal BA44 extending to the border of the precentral sulcus) and bilateral posterior parietal cortex (pSPL; BA7; MOG; BA19).

This activation pattern resembles the typical dorsal fronto-parietal network for visually guided behaviour that integrates sensory information with action-goals through sensorimotor transformations (Gallivan and Culham, 2015; Kravitz et al., 2011). Accordingly, MOG is known as an area involved in capturing relevant visual-spatial dimensions of

objects and visually-guided actions (Lingnau and Downing, 2015). SPL has been associated with high-level aspects of motor behaviour, such as the formation of intentions and early movement plans. These processes are aided by critical operations of multisensory integration and visuomotor transformation in SPL (Andersen and Buneo, 2002). Activations in pSPL have been reported during motor imagery of action-goals and trajectories (Aflalo et al., 2015), attentional spatial remapping/reprogramming of pre-selected actions (O'Reilly et al., 2013), and transformation of spatial target information into corresponding actions (Barany et al., 2014; Schon et al., 2002).

One crucial finding was the recruitment of the rIFG (dorsal BA44) when the final chord, predicted by the harmonic structure of the given musical sequence, was violated. This is consistent with the role of IFG in processing high-level aspects of motor behaviours (Grafton and Hamilton, 2007). Experimental evidence emphasises the role of bilateral IFG in processing hierarchical relationships within action sequences either when judging complex familiar activities (Farag et al., 2010) or when executing abstract hierarchically organised patterns of action sequences (Koechlin and Jubault, 2006). Altogether, these combined results suggest that the right IFG supports the structural integration of

Table 4
Comparison of rs-functional connectivity from the action- and audio-seed in the right IFG.

| Region | BA | Action > Audio seed | | | | | Audio > Action seed | | | | |
|------------------------------------|-------|---------------------|-----|-----|-----|---------|---------------------|-----|-----|-----|---------|
| | | k | x | y | z | Z-value | k | x | y | z | Z-value |
| <i>Right hemisphere</i> | | | | | | | | | | | |
| Frontal pole | 10 | 93 | 48 | 42 | 14 | 3.56 | | | | | |
| Superior frontal gyrus | 9 | | | | | | 4263 | 4 | 56 | 42 | 5.28 |
| Frontal medial cortex | 11 | 59 | 4 | 44 | −18 | 3.88 | | | | | |
| Cingulate gyrus | 23/24 | 52 | 2 | 6 | 30 | 4.58 | 200 | 2 | −14 | 38 | 4.89 |
| Superior temporal gyrus | 22 | | | | | | 645 | 52 | −8 | −8 | 5.16 |
| Angular gyrus | 40 | | | | | | 161 | 44 | −46 | 32 | 4.20 |
| Lingual gyrus | 27 | 72 | 16 | −42 | −6 | 4.87 | | | | | |
| Infer. temporal gyrus, temp-occ.j. | 37 | 996 | 52 | −56 | −14 | 5.53 | | | | | |
| Lateral occipital cortex, sup. | 7 | 4448 | 24 | −68 | 50 | 6.03 | | | | | |
| Precuneus cortex | 17 | 373 | 24 | −54 | 18 | 4.80 | | | | | |
| Thalamus | − | | | | | | 50 | 2 | −12 | 10 | 4.49 |
| Cerebellum (crus II) | − | 153 | 4 | −78 | −44 | 4.24 | 68 | 30 | −88 | −36 | 3.86 |
| Cerebellum (XI) | − | 51 | 16 | −46 | −48 | 4.46 | | | | | |
| <i>Left hemisphere</i> | | | | | | | | | | | |
| Frontal operculum cortex | 47 | | | | | | 976 | −40 | 26 | 0 | 5.35 |
| Superior frontal gyrus | 8 | 640 | −24 | 4 | 52 | 5.17 | | | | | |
| Precentral gyrus | 6 | 549 | −52 | 6 | 40 | 5.43 | | | | | |
| Middle temporal gyrus (middle) | 20 | | | | | | 206 | −56 | −20 | −12 | 4.88 |
| Middle temporal gyrus (post.) | 21 | | | | | | 64 | −54 | −38 | 0 | 4.38 |
| Temporal occipital fusiform cortex | 7 | 66 | −24 | −58 | −12 | 4.10 | | | | | |
| Lateral occipital cortex, sup. | 7 | 3678 | −26 | −76 | 30 | 5.66 | | | | | |
| Lateral occipital cortex, inf. | 19 | 567 | −50 | −76 | −4 | 4.87 | | | | | |
| Cerebellum (crus I/II) | − | 114 | −6 | −76 | −40 | 4.30 | 168 | −24 | −76 | −34 | 4.44 |

Results of the t-test between whole-brain functional connectivity from IFG activation maxima in action imitation and listening task. BA: Brodmann area, k: cluster size, MNI coordinates (x, y, z), and Z scores. ($P_{\text{voxel}} < 0.001$; $P_{\text{cluster}} < 0.05$, FWE corrected), temp-occ.j.: temporo-occipital junction, BA: Brodmann area, post.: posterior, sup.: superior, inf.: inferior.

simple acts into more complex combinatorial action sequences. The greater BOLD response during incongruent (compared to congruent) chords may be due to a mismatch with the predicted musical motor act that leads to higher computational costs during structural integration. Importantly, these findings indirectly show that pianists' knowledge of harmonic regularities transfers to the motor domain and enables them to predict and plan forthcoming musical acts during performance.

The absence of auditory activation in the incongruent vs. congruent contrast suggests that pianists relied more on their action knowledge recalled by the execution of the preceding chords than on auditory mechanisms (Bianco et al., 2016; Novembre and Keller, 2011; Sammler et al., 2013b). Note that this finding does not conflict with the large body of experimental evidence for action-perception coupling in trained musicians (for review, see Novembre and Keller, 2014; Zatorre et al., 2007). Our unusual and taxing imitation task on unhears sequences may have led pianists to focus on the motor part of the task, possibly suppressing unhelpful auditory images (cf. Pfordresher, 2012; van der Steen et al., 2014) (cf. Pfordresher, 2012). Alternatively, auditory feed-forward mechanisms may not discriminate between congruent and incongruent chords such that auditory activations cancelled out.

Overall, these fronto-parietal activations complement and support our previous behavioural (Novembre and Keller, 2011) and EEG studies on expert pianists (Bianco et al., 2016; Sammler et al., 2013a,b): silent production of harmonically incongruent chords elicited response time costs and a centro-parietal negativity that was associated with mechanisms of motor reprogramming of a pre-planned action in face of the violation. The activations of SPL and MOG match and support our interpretation of the posterior negativity as a correlate of the spatial remapping and reprogramming of pre-planned actions, and the activation of IFG lends evidence that these mechanisms stand under frontal control.

Within the predictive-coding framework (Friston, 2010), a bidirectional flow of information can be suggested to occur in the two hemispheres between parietal areas, processing visual-motor inputs, and the IFG, performing structural integration of incoming items. Indeed,

the functional connectivity analysis of our resting state data revealed strong positive correlations between right BA44 and, amongst others, bilateral superior parietal lobes, also revealed by the task-based analysis. A fronto-parietal network relying on the route of the dorsal visual stream has been associated with sensorimotor transformation during visually guided action planning (Goodale and Milner, 1992). According to motor control theory, these operations might be supported by "forward models", through which the expected outcome of an action is compared with actual sensory feedback (Wolpert and Flanagan, 2001). In this framework, posterior parietal regions simultaneously represent potential actions whose pre-selection is biased by the influence of internal models from prefrontal regions (Cisek, 2006). The novel finding is that these internal models may be shaped by the musician's knowledge of harmonic regularities and musical context. We propose that, on the one hand, visual-motor information about the current act is forwarded from posterior regions to the IFG that integrates the items and builds up an internal model of the sequence's harmonic structure. On the other hand, this internal model affords predictions of visual-spatial surface features of the next chord in MOG and may bias the pre-selection/representation of harmonically appropriate forthcoming motor acts in SPL. The generated model would be continually validated/updated via the matching between the expected action and the combined visual and proprioceptive signals from the current input (Wolpert and Flanagan, 2001). Interestingly, the combined findings raise the hypothesis that (musical) action knowledge, internal visual-motor models and fronto-parietal information flow may provide the basis on which the motor system contributes to visual perception and prediction of human behaviour (Novembre and Keller, 2014).

4.2. Music perception

In line with previous findings (Koelsch et al., 2005), listening to harmonically incongruent compared to congruent chords elicited activations in fronto-temporal areas: right inferior frontal gyrus (IFG: BA44, BA45) extending into the insular cortex, and right posterior superior temporal gyrus and sulcus (pSTG/STS: BA22).

The IFG and the posterior STG have been associated with structural analysis of auditory musical sequences based on internalised knowledge of harmonic regularities (Koelsch et al., 2005; Maess et al., 2001; Sammler et al., 2013a; Tillmann et al., 2006). The IFG has been proposed to support integration of discrete items into higher-order structures, based on which top-down predictions on forthcoming items can be generated. Greater BOLD responses in IFG may reflect the higher computational demand to integrate incongruent chords that are weakly related to the harmonic context and do not fulfil the prediction. Compared to these higher-order computations in IFG, pSTG/STS has been proposed to support lower-level matching processes between the actually perceived and the predicted sensory information (Sammler et al., 2013a). Indeed, posterior superior temporal areas have been associated with physical feature analysis and short-term representation of sounds (Seeger et al., 2013), as well as with the identification of the harmonic functions of chords within musical sequences (Musso et al., 2015).

Our connectivity analysis showed a functional coupling between IFG (BA44/BA45) and pSTG/STS, making it plausible to assume bidirectional dynamic fronto-temporal interactions during structural integration processes (Friston, 2010). On the one hand, early sensory analysis of chord functions may be forwarded from temporal to frontal regions where information is structurally integrated and harmonic predictions are established. On the other hand, these predictions may in turn inform the identification process in pSTG/STS where perceived and predicted items are matched to validate or revise the frontal prediction.

Overall, these data emphasise the crucial role of not just one area, but of a dynamic exchange of information between fronto-temporal areas in providing resources for the parsing of complex harmonically organised sounds (Hyde et al., 2011). Neuroanatomically, the fronto-temporal information exchange may be implemented along dorsal or ventral auditory pathways (see further below) (Loui et al., 2011; Musso et al., 2015; Rauschecker, 2011). The anatomical specification of these pathways, their functional relevance and dependency on musical training are interesting topics for future research.

4.3. Dorsal and ventral streams for musical action and perception

As discussed above, harmonic processing in musical actions and auditory perception relied on dissociable fronto-parietal and fronto-temporal neural networks, respectively. Representations of harmonic regularities in either visual-motor or auditory format are likely to account for the differential involvement of parietal and temporal areas, respectively, that both dynamically interact with computational processes in IFG. Interestingly, these interactions involved distinct posterior-dorsal and anterior-ventral rIFG sub-regions, i.e., BA6/44 in the action imitation task vs. BA44/45 in the audio task. This dissociation may either reflect (i) a task-unspecific sensitivity of IFG to structural processing demands in line with models of general prefrontal cortex specialization, or (ii) a task-specific involvement of dorsal and ventral IFG sub-regions as endpoints of different processing streams.

(i) Investigating harmonic structure processing in perception and action necessarily entails differences in experimental setup that alone suffice to induce different processing demands and shift activation peaks within IFG – even if both peaks may reflect similar structural computations. For example, recent theories propose anterior-posterior (Badre and D'Esposito, 2009; Fuster, 2001; Koehlin and Summerfield, 2007) and/or rostral-caudal (Friederici, 2011) gradients of prefrontal cortex organization along which similar functions, e.g., the “integration” of discrete items over time, operate at different levels of abstraction (Makuuchi et al., 2012). Along these lines, the more demanding imitation task might have triggered integration over shorter segments in the action sequences (i.e., integration at a lower level of complexity), limiting the activation to dorsal BA44 in the action contrast.

(ii) Alternatively, the divergence of dorsal and ventral rIFG peaks and connectivity profiles may arise from the intrinsically different nature of the tasks – silent musical action imitation vs. listening –

in line with dual stream models of the visuo-spatial (Goodale and Milner, 1992; Kravitz et al., 2011) and auditory system (Rauschecker and Scott, 2009; Rauschecker, 2011). According to these models, dorsal portions of IFG are interconnected with the parietal and temporal lobe within dorsal processing streams for time-dependent mechanisms that afford transformation between sensory input (visuo-spatial or sound) and motor representations, thereby supporting action. Dorsal stream involvement has been shown previously for goal-related actions (Kravitz et al., 2011), speech production (Hickok and Poeppel, 2007) and singing (Loui, 2015; Zarate, 2013) and is compatible with our fronto-parietal network observed in pianists during musical action imitation. Moreover, since the audio contrast comprised frontal activation extending to dorsal IFG, it is plausible that also during listening (although without imitation) a dorsal stream of auditory information might have been involved for mapping sound to action simulated by pianists (Zatorre et al., 2007). Notably, the present study adds two new insights: first, we demonstrate that frontal and parietal areas along the dorsal stream provide the neural resources for sequential structure processing during production of musical sequences; second, unlike in singing or speech production, our action imitation task eliminated auditory feedback during self-produced actions, hence, leading us to conclude that music-structural predictions can be grounded in the visual-motor control system.

Ventral IFG, in turn, is known as endpoint of the auditory ventral stream that, in concert with posterior temporal areas, is classically thought to process pitch information during singing (Berkowska and Dalla Bella, 2009; Zarate, 2013) and to map sound to meaning (Hickok and Poeppel, 2007; Rauschecker and Scott, 2009), compatible with our fronto-temporal network observed during listening. Although musical harmony does not have referential meaning as language, harmonic incongruities do have musical significance to listeners – i.e., intra-musical meaning as framed by Koelsch (2011) – in that the harmonic context leads towards a target chord that can be classified as more or less appropriate for musical closure.

Although the current findings do not speak to the causal role of the nodes or streams, they altogether highlight the relevance of considering structural integration in music production and perception as a network capacity by taking into account the connectivity between frontal computational and posterior modality-specific regions. Flexible and proficient music performance is likely to benefit from the dynamic weighting of these dissociable visual-motor and auditory circuits for prediction and motor planning based on internalised knowledge of harmony.

5. Conclusion

The present data provide first neuroimaging evidence that expert pianists predict forthcoming musical chords not only in auditory perception, but also in the processing of actions independently of auditory information. Remarkably, this suggests that, after intensive training, knowledge of structural regularities influences experts' action planning via implicit mechanisms of motor prediction/control, and might in turn increase proficiency of performance on top of fine movement optimization.

Our paradigm, in which pianists acted without listening to sound and listened without acting, dissociated a dorsal action and a ventral audio network for harmonic prediction, potentially acting in concert during real production (i.e., playing with sound). The dorsal and ventral networks both involve frontal computational sub-regions in rIFG, interconnected with parietal and temporal posterior systems of knowledge, respectively. These networks are likely to provide the infrastructure that allows frontal areas to keep track of abstract dependencies in sequential information via dynamic exchange with progressively lower-level modality-specific systems of knowledge. Predictive coding is proposed as an explanatory framework that

unifies both networks' functional roles: to optimise predictions in action and perception based on previous exposure and knowledge of harmony.

Acknowledgments

We are grateful to the pianists who participated in this experiment and to Bruno Pace who lent a hand in stimulus preparation. We also thank Mauricio Martins, Tomás Goucha, Gabriele Chierchia and Emiliano Zaccarella for inspiring discussions and Sven Gutekunst and Jöran Lepsien for technical support.

References

- Aflalo, T., Kellis, S., Klaes, C., Lee, B., Shi, Y., Pejsa, K., Shanfield, K., Hayes-jackson, S., Aisen, M., Heck, C., Liu, C., Andersen, R.A., 2015. Decoding motor imagery from the posterior parietal cortex of a tetraplegic human. *Science* 348 (80), 906–910.
- Andersen, R.A., Buneo, C.A., 2002. Intentional maps in posterior parietal cortex. *Annu. Rev. Neurosci.* 25, 189–220.
- Badre, D., D'Esposito, M., 2009. Is the rostro-caudal axis of the frontal lobe hierarchical? *Nat. Rev. Neurosci.* 10, 659–669.
- Barany, D.A., Della-Maggiore, V., Viswanathan, S., Cieslak, M., Grafton, S.T., 2014. Feature interactions enable decoding of sensorimotor transformations for goal-directed movement. *J. Neurosci.* 34, 6860–6873.
- Behzadi, Y., Restom, K., Liu, J., Liu, T.T., 2007. A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *NeuroImage* 37, 90–101.
- Berkowska, M., Dalla Bella, S., 2009. Acquired and congenital disorders of sung performance: a review. *Adv. Cogn. Psychol.* 5, 69–83.
- Bharucha, J., Krumhansl, C.L., 1983. The representation of harmonic structure in music: hierarchies of stability as a function of context. *Cognition* 13, 63–102.
- Bianco, R., Novembre, G., Keller, P.E., Scharf, F., Friederici, A.D., Villringer, A., Sammler, D., 2016. Syntax in action has priority over movement selection in piano playing: an ERP study. *J. Cogn. Neurosci.* 28, 41–54.
- Bressler, S.L., Menon, V., 2010. Large-scale brain networks in cognition: emerging methods and principles. *Trends Cogn. Sci.* 14, 277–290.
- Cisek, P., 2006. Integrated neural processes for defining potential actions and deciding between them: a computational model. *J. Neurosci.* 26, 9761–9770.
- Eerola, T., Himberg, T., Toivaniemi, P., Louhiuori, J., 2006. Perceived complexity of Western and African folk melodies by Western and African listeners. *Psychol. Music* 34, 337–371.
- Farag, C., Troiani, V., Bonner, M., Powers, C., Avants, B., Gee, J., Grossman, M., 2010. Hierarchical organization of scripts: converging evidence from fMRI and frontotemporal degeneration. *Cereb. Cortex* 20, 2453–2463.
- Fazio, P., Cantagallo, A., Craighero, L., D'Ausilio, A., Roy, A.C., Pozzo, T., Calzolari, F., Granieri, E., Fadiga, L., 2009. Encoding of human action in Broca's area. *Brain* 132, 1980–1988.
- Fiebach, C.J., Schubotz, R.I., 2006. Dynamic anticipatory processing of hierarchical sequential events: a common role for Broca's area and ventral premotor cortex across domains? *Cortex* 42, 499–502.
- Fitch, W.T., Martins, M.D., 2014. Hierarchical processing in music, language, and action: Lashley revisited. *Ann. N. Y. Acad. Sci.* 1316, 87–104.
- Friederici, A.D., 2011. The brain basis of language processing: from structure to function. *Physiol. Rev.* 91, 1357–1392.
- Friston, K., 2010. The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138.
- Fuster, J.M., 2001. The prefrontal cortex – an update: time is of the essence. *Neuron* 30, 319–333.
- Gallivan, J.P., Culham, J.C., 2015. Neural coding within human brain areas involved in actions. *Curr. Opin. Neurobiol.* 33, 141–149.
- Goodale, M.A., Milner, A.D., 1992. Separate visual pathways for perception and action. *Trends Neurosci.* 15, 20–25.
- Grafton, S.T., Hamilton, A.F.D.C., 2007. Evidence for a distributed hierarchy of action representation in the brain. *Hum. Mov. Sci.* 26, 590–616.
- Hickok, G., Poeppel, D., 2007. The cortical organization of speech processing. *Nat. Rev. Neurosci.* 8, 393–402.
- Hyde, K.L., Zatorre, R.J., Peretz, I., 2011. Functional MRI evidence of an abnormal neural network for pitch processing in congenital amusia. *Cereb. Cortex* 21, 292–299.
- Kilner, J.M., Friston, K.J., Frith, C.D., 2007. Predictive coding: an account of the mirror neuron system. *Cogn. Process.* 8, 159–166.
- Kim, S.-G., Kim, J.S., Chung, C.K., 2011. The effect of conditional probability of chord progression on brain response: an MEG study. *PLoS One* 6, e17337.
- Koechlin, E., Jubault, T., 2006. Broca's area and the hierarchical organization of human behavior. *Neuron* 50, 963–974.
- Koechlin, E., Summerfield, C., 2007. An information theoretical approach to prefrontal executive function. *Trends Cogn. Sci.* 11, 229–235.
- Koelsch, S., 2005. Neural substrates of processing syntax and semantics in music. *Curr. Opin. Neurobiol.* 15, 207–212.
- Koelsch, S., 2011. Toward a neural basis of music perception – a review and updated model. *Front. Psychol.* 2, 110.
- Koelsch, S., Siebel, W.A., 2005. Towards a neural basis of music perception. *Trends Cogn. Sci.* 9, 578–584.
- Koelsch, S., Fritz, T., Schulze, K., Alsop, D., Schlaug, G., 2005. Adults and children processing music: an fMRI study. *NeuroImage* 25, 1068–1076.
- Kravitz, D.J., Saleem, K.S., Baker, C.I., Mishkin, M., 2011. A new neural framework for visuospatial processing. *Nat. Rev. Neurosci.* 12, 217–230.
- Krumhansl, C.L., 1983. Perceptual structures for tonal music. *Mus. Percept.* 1, 28–62.
- Large, E.W., Palmer, C., 2002. Perceiving temporal regularity in music. *Cogn. Sci.* 26, 1–37.
- Lartillot, O., Ayari, M., 2011. Cultural impact in listeners' structural understanding of a Tunisian traditional modal improvisation, studied with the help of computational models. *J. Interdiscip. Music Stud.* 5, 85–100.
- Lerdahl, F., Jackendoff, R., 1983. *A Generative Theory of Tonal Music*. MIT Press, MA (Cambridge, MA: MIT Press).
- Lingnau, A., Downing, P.E., 2015. The lateral occipitotemporal cortex in action. *Trends Cogn. Sci.* 19, 268–277.
- Loui, P., 2015. A dual-stream neuroanatomy of singing. *Mus. Percept.* 32, 232–241.
- Loui, P., Wu, E.H., Wessel, D.L., Knight, R.T., 2009. A generalized mechanism for perception of pitch patterns. *J. Neurosci.* 29, 454–459.
- Loui, P., Li, H.C., Schlaug, G., 2011. White matter integrity in right hemisphere predicts pitch-related grammar learning. *NeuroImage* 55, 500–507.
- Maess, B., Koelsch, S., Gunter, T.C., Friederici, A.D., 2001. Musical syntax is processed in Broca's area: an MEG study. *Nat. Neurosci.* 4, 540–545.
- Maidhof, C., Vavatzanis, N., Prinz, W., Rieger, M., Koelsch, S., 2009. Processing expectancy violations during music performance and perception: an ERP study. *J. Cogn. Neurosci.* 22, 2401–2413.
- Makuuchi, M., Bahlmann, J., Friederici, A.D., 2012. An approach to separating the levels of hierarchical structure building in language and mathematics. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 367, 2033–2045.
- McKiernan, K.A., Kaufman, J.N., Kucera-Thompson, J., Binder, J.R., 2003. A parametric manipulation of factors affecting task-induced deactivation in functional neuroimaging. *J. Cogn. Neurosci.* 15, 394–408.
- Musso, M., Weiller, C., Horn, A., Glauche, V., Umarova, R., Hennig, J., Schneider, A., Rijntjes, M., 2015. A single dual-stream framework for syntactic computations in music and language. *NeuroImage* 117, 267–283.
- Novembre, G., Keller, P.E., 2011. A grammar of action generates predictions in skilled musicians. *Conscious. Cogn.* 20, 1232–1243.
- Novembre, G., Keller, P.E., 2014. A conceptual review on action-perception coupling in the musicians' brain: what is it good for? *Front. Hum. Neurosci.* 8, 603.
- O'Reilly, J.X., Schuffelgen, U., Cuell, S.F., Behrens, T.E.J., Mars, R.B., Rushworth, M.F.S., 2013. Dissociable effects of surprise and model update in parietal and anterior cingulate cortex. *Proc. Natl. Acad. Sci. U. S. A.* 110, E3660–E3669.
- Palmer, C., van de Sande, C., 1995. Range of planning in music performance. *J. Exp. Psychol. Hum. Percept. Perform.* 21, 947–962.
- Patel, A.D., 2003. Language, music, syntax and the brain. *Nat. Neurosci.* 6, 674–681.
- Pearce, Marcus T., et al., 2010. Unsupervised statistical learning underpins computational, behavioural, and neural manifestations of musical expectation. *NeuroImage* 50 (1), 302–313.
- Pfordresher, P.Q., 2012. Musical training and the role of auditory feedback during performance. *Ann. N. Y. Acad. Sci.* 1252, 171–178.
- Power, J.D., Schlaggar, B.L., Petersen, S.E., Power, J.D., Schlaggar, B.L., Petersen, S.E., 2015. Recent progress and outstanding issues in motion correction in resting state fMRI. *NeuroImage* 105, 536–551.
- Rauschecker, J.P., 2011. An expanded role for the dorsal auditory pathway in sensorimotor control and integration. *Hear. Res.* 271, 16–25.
- Rauschecker, J.P., Scott, S.K., 2009. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat. Neurosci.* 12, 718–724.
- Rohrmeier, M.A., Koelsch, S., 2012. Predictive information processing in music cognition: A critical review. *Int. J. Psychophysiol.* 83, 164–175.
- Rohrmeier, M., Rebuschat, P., 2012. Implicit learning and acquisition of music. *Top. Cogn. Sci.* 4, 525–553.
- Ruiz, M.H., Jabusch, H.-C., Altenmüller, E., 2009. Detecting wrong notes in advance: neuronal correlates of error monitoring in pianists. *Cereb. Cortex* 19, 2625–2639.
- Sammler, D., Koelsch, S., Friederici, A.D., 2011. Are left fronto-temporal brain areas a prerequisite for normal music-syntactic processing? *Cortex* 47, 659–673.
- Sammler, D., Koelsch, S., Ball, T., Brandt, A., Grigutsch, M., Huppertz, H.-J., Knösche, T.R., Wellmer, J., Widman, G., Elger, C.E., Friederici, A.D., Schulze-Bonhage, A., 2013a. Co-localizing linguistic and musical syntax with intracranial EEG. *NeuroImage* 64, 134–146.
- Sammler, D., Novembre, G., Koelsch, S., Keller, P.E., 2013b. Syntax in a pianist's hand: ERP signatures of 'embodied' syntax processing in music. *Cortex* 49, 1325–1339.
- Schon, D., Anton, J.L., Roth, M., Besson, M., 2002. An fMRI study of music sight-reading. *Neuroreport* 13, 2285–2289.
- Seger, C.A., Spiering, B.J., Quraini, A.G.S.S.I., Alpetter, C., David, J., Thaut, M.H., 2013. Corticostriatal contributions to musical expectancy perception. *J. Cogn. Neurosci.* 25 (7), 1062–1077.
- Smith, S.M., Fox, P.T., Miller, K.L., Glahn, D.C., Fox, P.M., Mackay, C.E., Filippini, N., Watkins, K.E., Toro, R., Laird, A.R., Beckmann, C.F., 2009. Correspondence of the brain's functional architecture during activation and rest. *PNAS* 106, 13040–13045.
- Swain, J.P., 1995. The concept of musical syntax. *Mus. Q.* 79, 281–308.
- Tenenbaum, J.B., Kemp, C., Griffiths, T.L., Goodman, N.D., 2011. How to grow a mind: statistics, structure, and abstraction. *Science* 331, 1279–1285.
- Tillmann, B., 2012. Music and language perception: expectations, structural integration, and cognitive sequencing. *Top. Cogn. Sci.* 4, 568–584.
- Tillmann, B., Bharucha, J.J., Bigand, E., 2000. Implicit learning of tonality: A self-organizing approach. *Psychol. Rev.* 107, 885–913.

- Tillmann, B., Koelsch, S., Escoffier, N., Bigand, E., Lalitte, P., Friederici, A.D., von Cramon, D.Y., 2006. Cognitive priming in sung and instrumental music: activation of inferior frontal cortex. *NeuroImage* 31, 1771–1782.
- Tymoczko, D., 2003. Local Harmonic Grammar in Western Classical Music Available at <http://www.dmitri.mycpanel.princeton.edu/mozart.pdf>.
- Uddin, L.Q., Kelly, A.M.C., Biswal, B.B., Castellanos, F.X., Milham, M.P., 2009. Functional connectivity of default mode network components: correlation, anticorrelation, and causality. *Hum. Brain Mapp.* 30, 625–637.
- van der Steen, M.C., Molendijk, E.B.D., Altenmueller, E., Furuya, S., 2014. Expert pianists do not listen: the expertise-dependent influence of temporal perturbation on the production of sequential movements. *Neuroscience* 269, 290–298.
- Vuust, P., Witek, M.A.G., 2014. Rhythmic complexity and predictive coding: a novel approach to modeling rhythm and meter perception in music. *Front. Psychol.* 5, 1–14.
- Wolpert, D.M., Flanagan, J.R., 2001. Motor prediction. *Curr. Biol.* 11, R729–R732.
- Zarate, J.M., 2013. The neural control of singing. *Front. Hum. Neurosci.* 7, 237.
- Zatorre, R.J., Chen, J.L., Penhune, V.B., 2007. When the brain plays music: auditory-motor interactions in music perception and production. *Nat. Rev. Neurosci.* 8, 547–558.

Chapter 3

Melody and fact

3.1 Study IV – Speech prosody and syntax

Sammler, D., Kotz, S. A., Eckstein, K., Ott, D. V. M., Friederici, A. D. (2010). Prosody meets syntax: the role of the corpus callosum. *Brain*, 133, 2643-2655.*

[Author contributions by D.S.: designed research with S.A.K., A.D.F.; received materials from K.E. and D.V.M.O.; performed research with K.E.; analysed data; wrote the paper, incorporating edits by co-authors]

*Reprinted with permission from Oxford University Press.

doi:10.1093/brain/awq231

Brain 2010; 133; 2643–2655 | 2643

BRAIN
A JOURNAL OF NEUROLOGY

Prosody meets syntax: the role of the corpus callosum

Daniela Sammler, Sonja A. Kotz, Korinna Eckstein, Derek V. M. Ott and Angela D. Friederici

Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

Correspondence to: Daniela Sammler,
Department of Neuropsychology,
Max Planck Institute for Human Cognitive and Brain Sciences,
04103 Leipzig, Germany
E-mail: sammler@cbs.mpg.de

Contemporary neural models of auditory language comprehension proposed that the two hemispheres are differently specialized in the processing of segmental and suprasegmental features of language. While segmental processing of syntactic and lexical semantic information is predominantly assigned to the left hemisphere, the right hemisphere is thought to have a primacy for the processing of suprasegmental prosodic information such as accentuation and boundary marking. A dynamic interplay between the hemispheres is assumed to allow for the timely coordination of both information types. The present event-related potential study investigated whether the anterior and/or posterior portion of the corpus callosum provide the crucial brain basis for the online interaction of syntactic and prosodic information. Patients with lesions in the anterior two-thirds of the corpus callosum connecting orbital and frontal structures, or the posterior third of the corpus callosum connecting temporal, parietal and occipital areas, as well as matched healthy controls, were tested in a paradigm that crossed syntactic and prosodic manipulations. An anterior negativity elicited by a mismatch between syntactically predicted phrase structure and prosodic intonation was analysed as a marker for syntax–prosody interaction. Healthy controls and patients with lesions in the anterior corpus callosum showed this anterior negativity demonstrating an intact interplay between syntax and prosody. No such effect was found in patients with lesions in the posterior corpus callosum, although they exhibited intact, prosody-independent syntactic processing comparable with healthy controls and patients with lesions in the anterior corpus callosum. These data support the interplay between the speech processing streams in the left and right hemispheres via the posterior portion of the corpus callosum, building the brain basis for the coordination and integration of local syntactic and prosodic features during auditory speech comprehension.

Keywords: syntax; prosody; interhemispheric transfer; corpus callosum

Abbreviations: antCC = lesions in the anterior two-thirds of the corpus callosum; ELAN = early left anterior negativity; postCC = lesions in the posterior third of the corpus callosum

Introduction

Contemporary neural models of language comprehension and production testify to a growing interest in the role of cortico-cortical pathways and the information flow within parallel, but inter-dependent large-scale networks for the processing of different

linguistic aspects such as phonology, syntax and semantics (Price *et al.*, 2005; Vigneau *et al.*, 2006; Hickok and Poeppel, 2007; Friederici, 2009). A great deal of work has concentrated on the intrahemispheric topographical connectivity between specialized fronto-temporal (and parietal) ‘language-areas’ via white matter fibre bundles such as the arcuate, superior longitudinal or uncinate

Received April 13, 2010. Revised June 22, 2010. Accepted June 28, 2010

© The Author (2010). Published by Oxford University Press on behalf of the Guarantors of Brain. All rights reserved.
For Permissions, please email: journals.permissions@oxfordjournals.org

fasciculi, in the left (Catani *et al.*, 2005; Anwander *et al.*, 2007; Saur *et al.*, 2008; Xiang *et al.*, 2010) and/or the right hemisphere (Parker *et al.*, 2005; Powell *et al.*, 2006; Glasser and Rilling, 2008). It is only with the growing awareness of specific right-hemispheric language functions such as the processing of pragmatic (Jung-Beeman, 2005) and prosodic aspects (Gandour *et al.*, 2000; Friederici and Alter, 2004) that the importance of interhemispheric connectivity during language comprehension has come into the focus of scientific interest. Here, we report on the role of the corpus callosum in the coordination of sentence level syntactic and (linguistic) prosodic information.

Recent neurocognitive models of auditory language comprehension assume a differential relative specialization of the two hemispheres in the processing of linguistic features (Zatorre *et al.*, 2002; Poeppel, 2003; Scott and Johnsrude, 2003; Friederici and Alter, 2004). While they predominantly assign segmental processing of syntactic and lexical semantic information to left-hemispheric fronto-temporal brain areas (Friederici, 2002; Hagoort, 2005; Shalom and Poeppel, 2008), homologue areas in the right hemisphere are thought to have primacy for the processing of sentence level suprasegmental, prosodic information such as accentuation and boundary marking. For example, the processing of filtered or degraded speech merely carrying prosodic information has been related to a left ear advantage in dichotic listening studies (i.e. indicating processing in the right hemisphere; Blumstein and Cooper, 1974; Shipley-Brown *et al.*, 1988) and predominant right fronto-temporal brain activations in functional MRI experiments (Meyer *et al.*, 2002, 2004). Likewise, deficits in the processing of sentence level prosodic intonation have been described in patients with lesions in the right hemisphere (Weintraub *et al.*, 1981; Brådvik *et al.*, 1991), although other studies using natural speech stimuli reported impairment of prosodic processing in both patients with lesions in the left and right hemispheres (Heilman *et al.*, 1984; Bryan, 1989; Perkins *et al.*, 1996). This suggests that the actual lateralization of prosodic processing partly depends on task demands (Plante *et al.*, 2002; Gandour *et al.*, 2004; Kotz *et al.*, 2006) and on the presence of concurrent segmental information (Van Lancker and Sidtis, 1992; Baum and Pell, 1999; Friederici and Alter, 2004; Pannekamp *et al.*, 2005).

If there is a relative hemispheric division of labour between segmental (left hemisphere) and suprasegmental information processing (right hemisphere), the coordination of the lateralized speech streams requires smooth and rapid information exchange between the two hemispheres. This is likely to occur via the commissural fibres crossing through the corpus callosum. The corpus callosum is not a homogeneous structure but the fibre tracts are topographically ordered according to their cortical origin. Diffusion tensor imaging studies consistently locate projections between the orbital and frontal lobes through the anterior two-thirds of the corpus callosum (genu and truncus), whereas fibres connecting the temporal, parietal and occipital lobes cross through the posterior third of the corpus callosum (isthmus and splenium; Huang *et al.*, 2005; Hofer and Frahm, 2006; Zarei *et al.*, 2006; Dougherty *et al.*, 2007; Park *et al.*, 2008).

It is unclear though which part of the corpus callosum is particularly critical for the interhemispheric exchange of syntactic and

prosodic information, because both the syntax (Jost *et al.*, 1996; Caplan *et al.*, 1998; Dapretto and Bookheimer, 1999; Ni *et al.*, 2000) and the prosody processing stream (Meyer *et al.*, 2002, 2004) span temporal and frontal brain areas in the respective hemisphere connected via the posterior and anterior corpus callosum, respectively (for reviews, see Friederici, 2002; Friederici and Alter, 2004). A first answer to this question has been obtained in an event-related potential study testing the impact of prosodic intonation on syntactic processing in patients with corpus callosum lesions (Friederici *et al.*, 2007). In this experiment, patients with posterior corpus callosum lesions did not process violations of verb argument structure that could only be detected if preceding prosodic phrase structure information was processed. More precisely, these patients did not exhibit a N400-like effect for intransitive verbs that did not match the prosody-induced syntactic expectation to encounter a transitive verb. Conversely, patients with anterior corpus callosum lesions showed the N400-like effect (although modulated in morphology and topography). Altogether, these data suggest that prosodic information guides syntactic processing predominantly via the posterior corpus callosum with less strong involvement of the anterior corpus callosum.

The present study

Importantly, the prosody–syntax interface does not work unidirectionally. Not only does prosody influence the syntactic structuring and interpretation of spoken utterances as tested by Friederici *et al.* (2007) (for review, see Cutler *et al.*, 1997; Eckstein and Friederici, 2006), but also syntactic structure is highly predictive of prosodic phrasing, with syntactic and prosodic boundaries often (though not always) coinciding (Selkirk, 1986; Inkelas and Zec, 1996; Truckenbrodt, 1999). The current electroencephalography lesion study incorporated this idea and aimed to specify which portion of the corpus callosum crucially interfaces syntactic parsing effects on the processing of prosodic pitch contour. Therefore, the present paradigm used sentences with a highly predictable syntactic phrase structure allowing for strong predictions about the prosodic intonation of a forthcoming word (i.e. rising or falling pitch indicating sentence continuation or closure, respectively). Conditions containing a mismatch between the syntactically predicted and the actually perceived prosodic intonation have been shown to evoke an anteriorly distributed negativity in healthy students (Eckstein and Friederici, 2006).

The present study tested whether this index of syntax–prosody interaction is based in the corpus callosum communication between the left and right hemispheres, and whether syntactic information takes an anterior or posterior route via the corpus callosum to guide prosodic processing. Therefore, two patient groups were recruited according to the topographical organization of the corpus callosum (Huang *et al.*, 2005; Hofer and Frahm, 2006), one group with lesions in the anterior two-thirds of the corpus callosum and the other with lesions in the posterior third of the corpus callosum, as well as matched healthy controls. If the anterior negativity evoked by a mismatch between syntactic context and forthcoming prosodic intonation (i.e. indicating an impact of syntactic on prosodic processing) (Eckstein and Friederici, 2006)

is due to an interaction between the left and right hemispheres, then this effect should be absent in patients with lesions in the relevant portion of the corpus callosum. In other words, syntactic information processed in the left hemisphere should not influence prosodic processes in the right hemisphere. Based on the findings of Friederici *et al.* (2007), we hypothesized that lesions of posterior fibres connecting the temporal lobes involved in both prosodic and syntactic processing (Friederici, 2002; Friederici and Alter, 2004) would lead to an interruption of syntax–prosody interactions. Alternatively, given the strong involvement of the left inferior frontal gyrus and frontal operculum in syntactic processing (Friederici, 2002; Grodzinsky and Friederici, 2006; Vigneau *et al.*, 2006), syntactic guidance of prosodic processing may also be disrupted after lesion of anterior corpus callosum fibres connecting the orbital and frontal lobes.

Material and methods

Participants

Eleven patients were classified into two groups according to their lesion sites, involving either the anterior two-thirds of the corpus callosum ($n=5$, antCC group), or the posterior third of the corpus callosum ($n=6$, postCC group; see Table 1 and Fig. 1 for patient histories and lesion sites). Furthermore, 22 healthy controls, matched in handedness, gender and age were tested (age range: 23–71 years, mean: 50 years; 16 males, 6 females). Informed consent according to the Declaration of Helsinki (1964) was obtained from each participant prior to the experiment, which was approved by the local Ethical Committee.

To ensure that patients and controls did not differ in their peripheral hearing abilities, monaural pure tone thresholds were determined in all participants for frequencies of 170, 1000 and 4000 Hz, chosen to

cover the main frequency range of speech. Notably, 170 Hz was tested because the perception of prosodic manipulations in the stimulus material required the detection of F_0 shifts either exceeding or dropping below 170 Hz indicating sentence continuation or closure, respectively (for details, see Eckstein and Friederici, 2006). Audiograms for each ear and frequency were obtained in an adaptive up-down method using the test equipment of Bungert-Kahl *et al.* (2004) in a sound-proof cabin. The test values of each participant (average of left and right ears) were z-transformed based on the normative data provided by Bungert-Kahl *et al.* (2004) for different age groups. Missing normative data (e.g. for 170 Hz and ages >70 years) were interpolated using the cubic spline function in MATLAB 7.1. All experimental groups exhibited normal hearing thresholds in all tested frequencies [i.e. not exceeding 2 standard deviations (SD) of the standard normal distribution; Table 2], assuring that all groups were principally able to perceive the prosodic manipulations. Moreover, the hearing threshold did not differ between groups (all P -values > 0.132), ensuring that potential group differences in electrophysiological effects cannot be attributed to differential hearing abilities.

Stimulus material

The stimulus material employed has been used in a previous study with healthy students (Eckstein and Friederici, 2006). Participants listened to 288 German sentences starting with a proper name ('Steffen') and a verb ('sieht'/'sees') forming the matrix clause. The following subordinate clause was composed of a complementizer 'dass'/'that', a subject ('der Lehrer'/'the teacher'), a prepositional phrase consisting of a preposition fused with an article ('beim'/'during the') and a noun ('Tadel'/'reproval'), followed by the sentence final verb ('schmunzelt'/'smiles'; Table 3 and Supplementary Stimuli). Prosodic intonation of the noun at penultimate position was manipulated.

The critical noun (stem + suffix) was either marked for sentence continuation (congruent; rising pitch contour) or spoken with sentence final prosody (incongruent; falling pitch contour). The latter violates

Table 1 Individual patient histories

| Patient | Gender | Age | Hand | Onset | Aetiology | Callosal lesion | Additional lesions |
|---|--------|-----|------|-------|-----------|--|---|
| Anterior corpus callosum group (involving the anterior two-thirds of the corpus callosum) | | | | | | | |
| 104 | F | 69 | R | 9;9 | CMA | Rostrum, anterior c-body | Pontine, left basal ganglia lesion |
| 197 | F | 64 | R | 8;10 | IS/SAH | Rostrum, middle c-body | Left frontolateral ^a and temporopolar contusion, left basal ganglia lesion, left parietal necrosis |
| 286 | M | 66 | R | 8;4 | IS | Anterior c-body | CMA with pontine lacune |
| 521 | M | 51 | R | 6;5 | SAH | Anterior knee, left middle c-body, EVD | Right basal forebrain and temporopolar lesion |
| 724 | M | 25 | L | 5;4 | CH | Anterior c-body | Left thalamus, left internal capsule lesion |
| Posterior corpus callosum group (involving the posterior third of the corpus callosum) | | | | | | | |
| 126 | M | 75 | R | 12;0 | CH | Chronic ischemic posterior CC lesion | CMA, right thalamic lacune, right occipital bleeding |
| 339 | M | 59 | R | 7;11 | SAH | Posterior c-body (presplenial) | Post SAH, post EVD |
| 422 | M | 43 | R | 9;0 | TBI | Presplenial lesion | Minor atrophy of cerebellar vermis |
| 432 | F | 26 | R | 8;6 | CH | Posterior CC lesion | Embolized AVM, left posterior thalamus lesion |
| 675 | M | 24 | R | 5;2 | TBI | Small presplenial lesion | Left preinsular region, ^a left cerebral peduncle (midbrain) |
| 1499 | M | 53 | R | – | TBI | (pre)splenial lesion | CMA |

^a Patients were included despite lesions in the left frontal cortex known to be involved in syntactic processing (Friederici and Kotz, 2003; Vigneau *et al.*, 2006; Grodzinsky and Santi, 2008) as they showed a normal ELAN and P600 pattern (Fig. 3B and C) suggesting normal syntactic processing.

Handedness is as defined by the Edinburgh Handedness Inventory (Oldfield, 1971); Lesion onset is in years; months—no information available for Patient 1499.

AVM = arteriovenous malformation; CH = cerebral haemorrhage; CMA = cerebral microangiopathy; EVD = external ventricular drainage; F = female; IS = ischemic stroke; L = left handed; M = male; R = right handed; SAH = subarachnoid haemorrhage; TBI = traumatic brain injury.

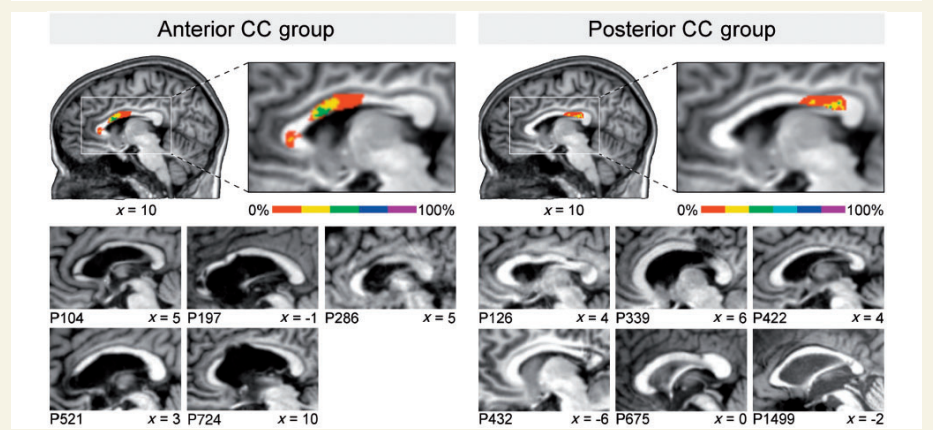


Figure 1 Lesion overlay and individual lesion sites of patients classified into the anterior and posterior corpus callosum (CC) lesion group. Coordinates refer to Talairach space (Talairach and Tournoux, 1988). The colour code indicates the percentage of patients with lesions in the given area (red = 1 patient, purple = all patients of the group).

Table 2 Between-group comparison of z-transformed hearing thresholds (mean ± SEM) at 170, 1000 and 4000 Hz

| | Controls | antCC | postCC | F(2,30) | P |
|---------|--------------|--------------|-------------|---------|--------|
| 170 Hz | −0.08 ± 0.23 | −0.83 ± 0.42 | 0.05 ± 0.46 | 1.13 | >0.335 |
| 1000 Hz | 0.10 ± 0.27 | −0.76 ± 0.19 | 0.29 ± 0.69 | 1.10 | >0.344 |
| 4000 Hz | −0.10 ± 0.35 | −0.49 ± 0.42 | 1.25 ± 0.70 | 2.16 | >0.132 |

Multivariate ANOVA with the fixed factor Group (healthy controls versus the antCC group versus the postCC group) computed on the z-transformed hearing thresholds at 170, 1000 and 4000 Hz (mean values of left and right ears).

prosodic expectations because the preceding syntactic context (e.g. ‘Steffen sieht, dass der Lehrer beim...’/‘Steffen sees that the teacher during the...’) predicts to be followed by at least two syntactic elements, a noun (required by the preposition ‘beim’/‘during the’) and a verb (required by the complementizer ‘dass’/‘that’). This anticipated syntactic structure allows for the prosodic prediction that the critical noun (‘Tadel’/‘reproval’) carries sentence continuation prosody. The manipulation of the noun’s prosodic intonation thus represents a violation of prosodic expectancies that are established through syntactic structure, allowing us to study the impact of syntactic phrase structure on prosodic processing. The electrophysiological marker of this syntax–prosody mismatch is typically an anterior negativity (Eckstein and Friederici, 2006).

To ensure that the predicted absence of the anterior negativity after corpus callosum lesion is due to a reduced interaction of syntactic with prosodic processing, it needs to be shown that participants process syntactic structure *per se*. Therefore, a syntactic error was included in half of the sentences, i.e. the word-category of the critical noun was manipulated at its suffix. All critical words were chosen such that their word stems (e.g. ‘Tade-’/‘reprov-’) were ambiguous with respect to word-category. In 50% of the sentences, a nominal suffix (e.g. ‘-l’/‘-al’) disambiguated the word towards being a noun

(‘Tade-l’/‘reprov-al’), which is highly expected in German after a preposition. In the other half of the sentences, the suffix (e.g. ‘-l’/‘-es’) indicated an inflected verb (‘tade-lt’/‘reprov-es’), which represents a word-category violation in a prepositional phrase. Typical electrophysiological markers evoked by these syntactic errors are an early left anterior negativity (ELAN) and a late posterior positivity (P600; Friederici et al., 1993; Hagoort et al., 1999; Hahne and Friederici, 1999; Kaan et al., 2000).

Taken together, the first part of the sentence establishing strong local syntactic and prosodic predictions was resolved into four experimental conditions: (i) congruent prosody + correct syntax (cc); (ii) incongruent prosody + correct syntax (pc); (iii) congruent prosody + incorrect syntax (cs); and (iv) incongruent prosody + incorrect syntax (ps) (audio samples of each condition are provided in the Supplementary material, which also includes an analysis of the syntactic processing at the suffix as a function of the prosodic information in the preceding word stem, i.e. informing us about the impact of prosody on syntax processing). In addition, two prosodically and syntactically correct filler conditions were included to avoid strategic effects that could emerge from the fact that verbal suffixes always coincided with a syntactic error. In these sentences, the critical word stem was presented with verbal suffix in a correct context (e.g. Filler 1: ‘Steffen sieht, dass der Lehrer mit Nachdruck tade-lt.’/‘Steffen sees that the teacher with emphasis reprov-es.’; Filler 2: ‘Steffen sieht, dass der Lehrer tade-lt.’/‘Steffen sees that the teacher reprov-es.’). Filler sentences did not enter the event-related potential data analysis. A detailed description of the cross-splicing procedure for stimulus construction and the acoustic properties of the material is provided by Eckstein and Friederici (2006).

Procedure

Participants sat in a comfortable chair in a sound-proof cabin while listening to the randomly presented 192 experimental and 96 filler sentences (48 sentences/condition). Stimuli were presented via loudspeakers at a comfortable volume using Experimental Run-Time

Table 3 Stimulus material

| Cond | Prosody | Syntax | Examples: 'Steffen sieht, dass der Lehrer beim...' ('Steffen sees that the teacher during the...') | | |
|------|-------------|-----------|--|--------------|-------------------------|
| cc | Congruent | Correct | '... Tade-l' | 'schmunzelt' | ('... reproval smiles') |
| cs | Congruent | Incorrect | '... tade-lt' | 'schmunzelt' | ('... reproves smiles') |
| pc | Incongruent | Correct | '... Tade-l' | 'schmunzelt' | ('... reproval smiles') |
| ps | Incongruent | Incorrect | '... tade-lt' | 'schmunzelt' | ('... reproves smiles') |

Literal English translations are given in parentheses. Further examples are provided as Supplementary material.

System (ERTS; BeriSoft Cooperation, Frankfurt am Main, Germany). Each trial started with a fixation star visible from 500 ms prior to sentence onset to 1500 ms after sentence off-set, followed by a 2000 ms response screen during which participants were asked to judge the grammaticality of the sentence by pressing a left or right button (key assignment was counterbalanced across participants). Each trial was followed by a 2500 ms interstimulus interval showing a blank screen. To familiarize participants with the procedure, they received training with 12 randomly presented trials (two sentences/condition). The experimental session was divided into four blocks and had a duration of ~60 min (training and breaks included).

Data acquisition and analysis

The electroencephalogram was recorded from 25 Ag/AgCl electrodes placed according to the extended international 10–20 system (Sharbrough *et al.*, 1991). The electrode positions were: FP1, FP2, F7, F8, F3, F4, FZ, FT7, FT8, FC3, FC4, T7, T8, C3, C4, CZ, CP5, CP6, P7, P8, P3, P4, PZ, O1, O2. Left mastoid (M1) served as reference; an additional electrode was placed on the right mastoid bone (M2) for off-line re-referencing. The ground electrode was located on the sternum. Horizontal and vertical electrooculograms were bipolarly recorded from electrodes placed on the outer canthus of each eye, as well as above and below the right eye. Impedances were kept below 5 k Ω . Signals were amplified with two synchronized PORTI-32/MREFA amplifiers (Twente Medical Systems International B.V., Enschede, The Netherlands) and digitized with a sampling rate of 250 Hz.

EEP 3.2 (ANT-software) was used to re-reference the data to linked mastoids. Further processing steps were accomplished using EEGLAB 5.03 (Delorme and Makeig, 2004) in MATLAB 7.7. Data were filtered using a 0.3 Hz highpass filter (fir, 2725 points, Blackman window), and strong muscle artefacts, electrode drifts or technical artefacts were manually rejected before entering the continuous data into an independent component analysis. The resulting component structure was used to reject eye movement and blink artefacts, muscle artefacts and slow drifts. Afterwards, the data were filtered with a 20 Hz lowpass filter (fir, 277 points, Blackman window), and cut into two different epochs. The first epoch of –100 to 1600 ms was time-locked to the onset of the critical word and ideally covered the prosodic violation in the word stem. The second epoch of –100 to 1200 ms was time-locked to the suffix of the critical word, i.e. the onset of the syntactic violations. Epochs were rejected whenever one or more electrodes exhibited voltages of $\pm 55 \mu\text{V}$, linear trends of $\pm 50 \mu\text{V}$ in a 400 ms gliding window or small blinks and drifts as identified by visual inspection. Non-rejected epochs were averaged separately for each condition. Only correctly answered trials were included. Averages were aligned to the respective –100 to 0 ms baseline. An average of 42.14 trials was included for each participant and each condition (mean \pm SD of healthy controls: 43.07 \pm 3.96, antCC: 39.85 \pm 4.44, postCC: 40.63 \pm 5.79; no significant differences between groups and conditions).

Effects of syntax on prosodic processing were analysed time-locked to the onset of the critical word (i.e. time-locked to the prosodic violation). Statistical analyses were carried out on the mean amplitudes of the prosodically congruent (cc and cs) and incongruent conditions (pc and ps) calculated in a time window from 200 to 500 ms after word onset (centred around the peak of the anterior negativity as identified by visual inspection) for six regions of interest: (i) left fronto-temporal (F7, FT7, T7); (ii) right fronto-temporal (F8, FT8, T8); (iii) left fronto-central (F3, FC3, C3); (iv) right fronto-central (F4, FC4, C4); (v) left posterior (P7, P3, O1); and (vi) right posterior (P8, P4, O2). Prosody-independent effects of syntactic violations were analysed time-locked to the onset of the suffix of the critical word. Statistical analyses were carried out on the mean amplitudes of syntactically correct (cc and pc) and incorrect trials (cs and ps) for the regions of interest described above, in two time windows: (i) from 250 to 350 ms (centred around the peak of the ELAN); and (ii) from 450 to 850 ms after suffix onset (centred around the peak of the P600). The Results in Supplementary materials contain an additional analysis of the modulation of the ELAN and P600 at the word suffix by the prosodic information in the preceding word stem. In healthy controls and (as a trend) in the antCC group, the ELAN lost its left lateralization, showing up bilaterally when the syntactic error was preceded by incongruent compared to congruent prosody. This indicates an immediate influence of phrasal prosody on early syntactic parsing stages. The patients in the postCC group showed no such ELAN topography shift, suggesting that prosodic information does not influence early steps of syntactic analysis after disconnection of posterior callosal fibres and supporting the notion that the posterior third of the corpus callosum is particularly relevant for interfacing prosodic and syntactic information during auditory language comprehension (Friederici *et al.*, 2007). No consistent results were found for the P600. Statistical details and a more extensive discussion are provided in the Results section of Supplementary materials. The analysis of the behavioural performance was restricted to the error rates. No reaction times were analysed due to the delayed response.

Results

Behavioural data

Participants showed an average performance of 94.59% correct (healthy controls: 97.51%, antCC group: 87.5%, postCC group: 89.76%), demonstrating that they were able to detect the syntactic violations. An ANOVA with the repeated measures factors Syntax (correct versus incorrect), Prosody (congruent versus incongruent) and the between-subjects factor Group (healthy controls versus antCC group versus postCC group) revealed a marginally better performance for prosodically congruent compared with

incongruent sentences (main effect of Prosody: $F(1,30)=4.08$, $P<0.053$), indicating that prosodic violations were processed although they were task irrelevant. No group differences (P -values >0.245) or other effects were found.

Impact of syntax on prosody

In healthy controls and patients in the antCC group, the prosodic violations elicited a negativity between 200 and 500 ms after word onset with a frontal scalp distribution (Fig. 2) as observed in previous experiments (Eckstein and Friederici, 2006). No such effect was found in the postCC group (Fig. 3A). ANOVAs with the within-subject factors Prosody (congruent versus incongruent), Region (fronto-temporal versus fronto-central versus posterior) and Hemisphere (left versus right) revealed a significant main effect of Prosody and a marginally significant interaction of Prosody \times Region (indicating a frontally distributed negativity) in healthy controls and the antCC group, but not in the postCC group (for statistical values see Table 4). An ANOVA with the additional between-subjects factor Group yielded a significant interaction of Prosody \times Group [$F(2,30)=4.05$, $P<0.028$, $\eta_p^2=0.213$] and a marginally significant interaction of Prosody \times Region \times Group [$F(4,60)=2.43$, $P<0.072$, $\eta_p^2=0.139$]. Subsequent ANOVAs comparing the postCC group versus healthy controls and the postCC group versus the antCC group revealed both a significant interaction of Prosody \times Group, whereas no such effect was found when comparing the antCC group versus healthy controls (for statistical details see Table 5). These data show that the processing of the syntax–prosody mismatch in the postCC group significantly deviated from the normal processing in healthy controls and the antCC group. This difference may be due to a missing functional interaction between syntactic cues and prosodic information. However, before drawing such a conclusion, it needs to be shown that the postCC group processed syntax normally, independent of prosodic information.

Prosody-independent syntactic processing

In all groups, the syntactic violation evoked an ELAN between 250 and 350 ms and a P600 between 450 and 850 ms after suffix onset (Figs 4, 3B and C), in line with previous studies (Friederici *et al.*, 1993; Hahne and Friederici, 1999; Eckstein and Friederici, 2006).

Early time window (250–350 ms)

ANOVAs with the factors Syntax (correct versus incorrect), Region and Hemisphere yielded an interaction of Syntax \times Region and/or Syntax \times Hemisphere in all groups; healthy controls showed an additional three-way interaction of Syntax \times Region \times Hemisphere (Table 4). To resolve this interaction, separate analyses of the variable Syntax were computed in each of the six regions of interest in healthy controls. A significant negativity (the ELAN) was observed only in the left and to a lesser degree in the right temporal region of interest (Table 6; P -values >0.07 in all other regions of interest). This effect was clearly left dominant as indicated by a significant interaction of Syntax \times Hemisphere in an

ANOVA with the factors Syntax and Hemisphere calculated for temporal regions of interest only [$F(1,21)=10.27$, $P<0.004$, $\eta_p^2=0.328$]. A similar pattern was observed in the antCC and postCC groups, both exhibiting a significant left-temporal negativity that was stronger than in right temporal regions of interest (Table 6; although the interactions of Syntax \times Hemisphere for temporal regions of interest failed to reach statistical significance; antCC group: $F(1,4)<1$, NS; postCC group: $F(1,5)=3.66$, $P>0.114$, $\eta_p^2=0.422$). An ANOVA including the factor Group did not reveal any group differences ($P>0.301$ for all interactions involving the factors Syntax and Group), indicating that early steps of syntactic parsing did not differ between the three experimental groups.

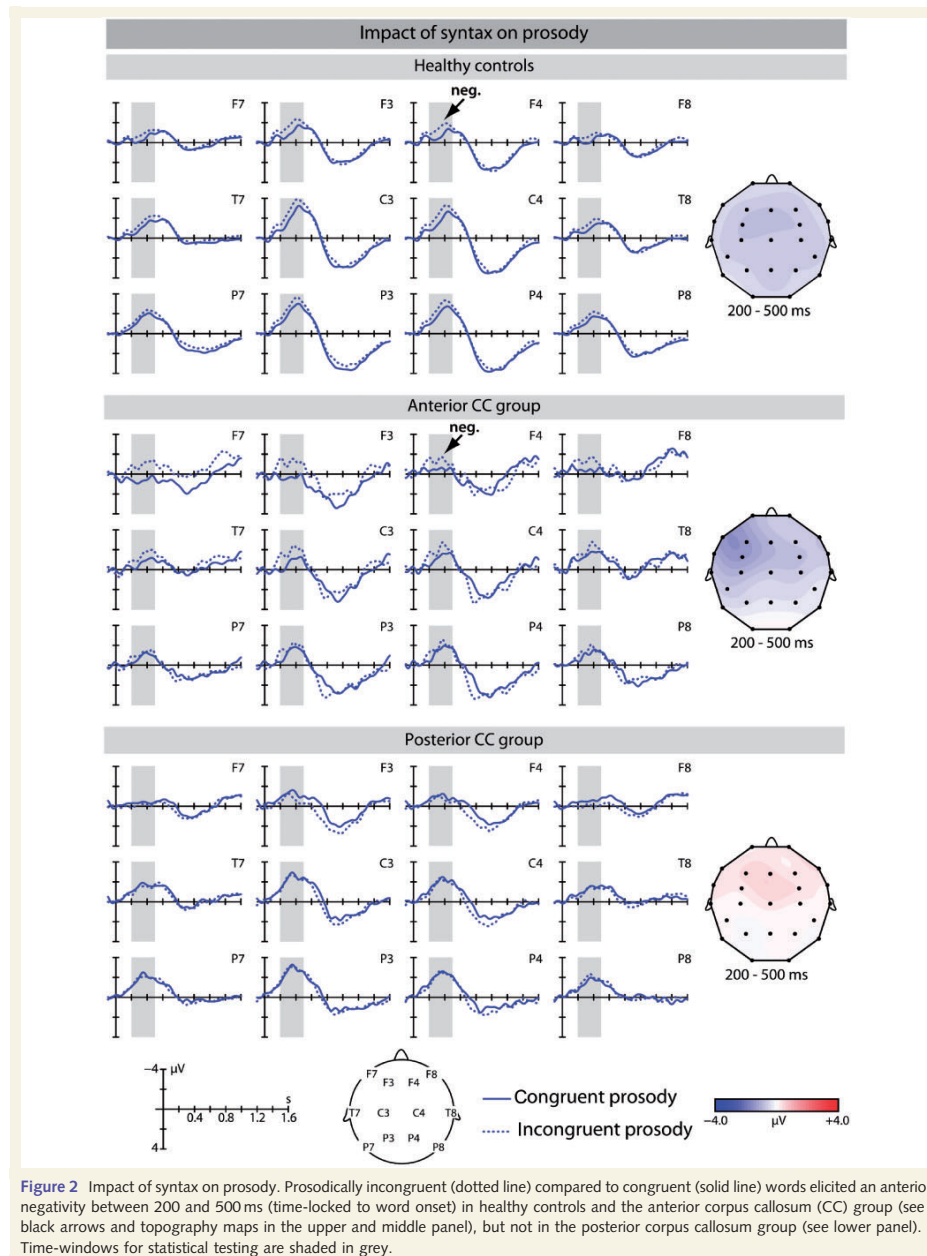
Late time window (450–850 ms)

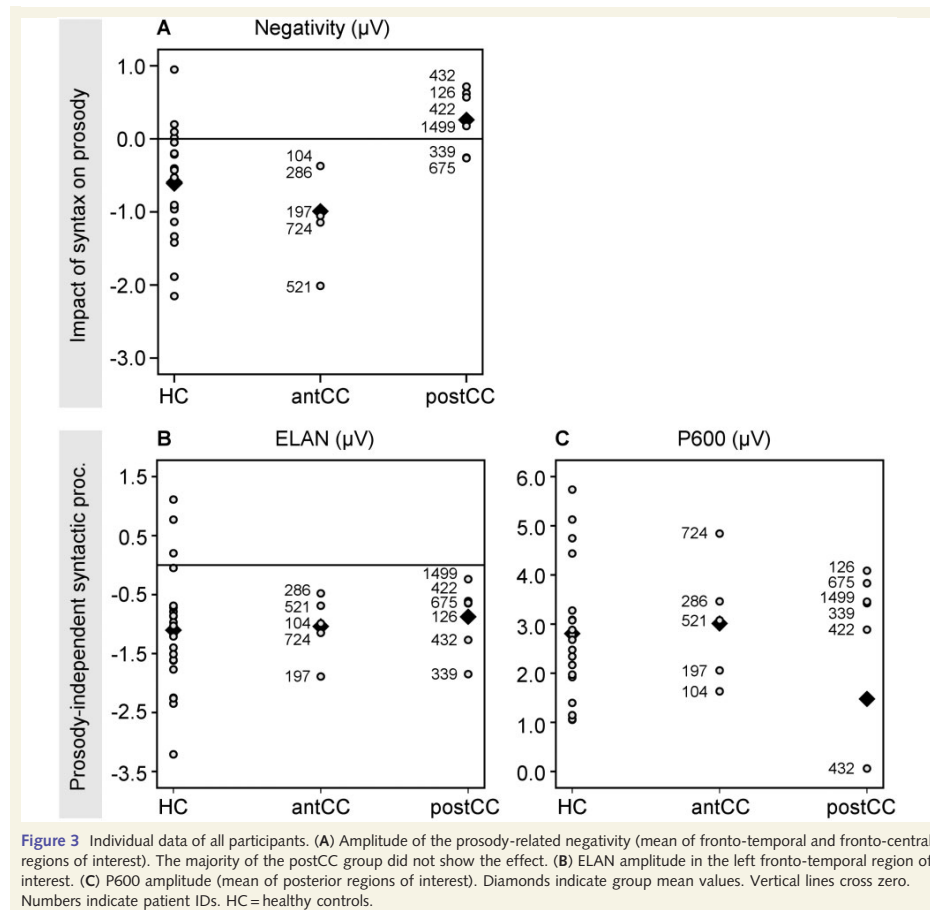
The statistical analyses showed a significant main effect of Syntax as well as an interaction of Syntax \times Region in all three experimental groups. Healthy controls exhibited additional interactions of Syntax \times Hemisphere, and Syntax \times Region \times Hemisphere (Table 4). Follow-up analyses in healthy controls showed that a syntactic positivity was significant in bilateral central and posterior regions of interest. A similar centro-parietal positivity was found in both patient groups (Table 6). No group differences were found ($P>0.154$ for all interactions involving the factors Syntax and Group), suggesting a comparable functioning of syntactic integration in all groups.

Discussion

The present study aimed to determine which portion of the corpus callosum is crucial for the interhemispheric information flow between the differently lateralized pathways processing syntax and prosody (Friederici and Alter, 2004). The combined data suggest that particularly the posterior third of the corpus callosum is critical for the dynamic interplay between the two streams during auditory language comprehension.

In healthy controls and the antCC (but not postCC) group, the prosodic violations elicited a negativity between 200 and 500 ms with a frontal scalp distribution, replicating the findings of earlier studies testing healthy students (Eckstein and Friederici, 2005, 2006). This negativity has been interpreted as a marker for the detection of a prosodic expectancy violation, i.e. a mismatch between the prosodic expectation of sentence continuation and the actual speech input signalling sentence closure. Notably, the build-up of such prosodic expectations crucially depends on the structural predictions made by the syntactic context that precedes the prosodic violation, suggesting a combined syntactic–prosodic origin of the negativity. This assumption is supported by the absence of the negativity in another study using the very same prosodic manipulation as applied in the present experiment (i.e. the penultimate word of a sentence presented with sentence final prosody; see stimulus pair 1 in Eckstein and Friederici, 2005). The mere difference between these and our stimuli was the degree of predictability of the upcoming syntactic (and prosodic) structure. While the context of our stimuli unmistakably predicted sentence closure not earlier than at the second element to follow,





the syntactic context in Eckstein and Friederici (2005) (e.g. 'Peter weiß, dass der Onkel...'/ 'Peter knows, that the uncle...') left open how many syntactic nodes were to follow, precluding as clear a prediction on prosodic phrasing. These combined findings lead us to conclude that the observed negative brain potential is a response to violations of local prosodic expectancies that are established through syntactic structure.

Along these lines, the fact that prosodic manipulations elicited this negativity in healthy controls and the antCC group indicates that both groups processed the syntactic context and used it to build up expectancies regarding upcoming prosodic boundaries, altogether suggesting an intact (interhemispheric) interaction of syntactic with prosodic information in these groups. The absence of this effect in the postCC group suggests that online syntactic

processes do not trigger the build-up of prosodic expectations after disconnection of posterior callosal fibres, qualifying the posterior third of the corpus callosum as the crucial interface between syntactic and prosodic information. This conclusion is based on two additional findings. First, the postCC group, like healthy controls and the antCC group, had normal hearing abilities in the speech frequency range, and particularly hearing thresholds around 170 Hz (i.e. the critical frequency for discriminating prosodically congruent and incongruent sentences in our study) were highly similar in all groups, suggesting equal preconditions for the processing of the prosodic manipulation. Moreover, as in the study of Friederici *et al.* (2007), the behavioural judgements of prosodically congruent and incongruent sentences did not differ between the postCC and antCC groups although their brain responses

Table 4 Results of the three-way ANOVAs testing prosodic and syntactic processing separately for each group

| | Healthy controls | | | | AntCC | | | | PostCC | | | |
|---|------------------|--------------|-------------------|--------------|-------|--------------|--------------|--------------|------------|--------------|--------------|--------------|
| | df | F | P | η_p^2 | df | F | P | η_p^2 | df | F | P | η_p^2 |
| ANOVA: Prosody \times Region \times Hemisphere (200–500 ms after word onset) | | | | | | | | | | | | |
| P | 1,21 | 19.03 | <0.0004 | 0.475 | 1,4 | 13.36 | 0.022 | 0.771 | 1,5 | <1 | NS | 0.065 |
| P \times R | 2,20 | 2.67 | 0.094 | 0.113 | 2,3 | 3.83 | 0.083 | 0.490 | 2,4 | <1 | NS | 0.121 |
| P \times H | 1,21 | <1 | NS | 0.000 | 1,4 | 1.52 | 0.285 | 0.275 | 1,5 | <1 | NS | 0.000 |
| P \times R \times H | 2,20 | <1 | NS | 0.029 | 2,3 | 1.68 | 0.246 | 0.295 | 2,4 | <1 | NS | 0.019 |
| ANOVA: Syntax \times Region \times Hemisphere (250–350 ms after suffix onset) | | | | | | | | | | | | |
| S | 1,21 | <1 | NS | 0.030 | 1,4 | <1 | NS | 0.024 | 1,5 | <1 | NS | 0.018 |
| S \times R | 2,20 | 11.31 | <0.0003 | 0.350 | 2,3 | 7.18 | 0.016 | 0.643 | 2,4 | 1.26 | 0.323 | 0.201 |
| S \times H | 1,21 | 10.51 | 0.004 | 0.333 | 1,4 | <1 | NS | 0.010 | 1,5 | 9.03 | 0.030 | 0.644 |
| S \times R \times H | 2,20 | 6.15 | 0.007 | 0.227 | 2,3 | <1 | NS | 0.054 | 2,4 | 2.07 | 0.195 | 0.289 |
| ANOVA: Syntax \times Region \times Hemisphere (450–850 ms after suffix onset) | | | | | | | | | | | | |
| S | 1,21 | 25.29 | <0.0001 | 0.546 | 1,4 | 32.14 | 0.005 | 0.889 | 1,5 | 19.07 | 0.007 | 0.792 |
| S \times R | 2,20 | 90.23 | <0.0001 | 0.811 | 2,3 | 9.33 | 0.008 | 0.700 | 2,4 | 7.75 | 0.009 | 0.607 |
| S \times H | 1,21 | 27.08 | <0.0001 | 0.563 | 1,4 | <1 | NS | 0.006 | 1,5 | 2.33 | 0.188 | 0.316 |
| S \times R \times H | 2,20 | 5.23 | 0.009 | 0.200 | 2,3 | <1 | NS | 0.088 | 2,4 | 1.12 | 0.365 | 0.180 |

Bold values indicate significant results. Partial eta squared $\eta_p^2 > 0.5$ = large effect size; $\eta_p^2 > 0.3$ = medium effect size; $\eta_p^2 > 0.1$ = small effect size (Bortz and Döring, 2003). H = hemisphere; NS = not significant; P = prosody; R = region; S = syntax. Bold values indicate significant results ($P < 0.05$) and large effect sizes.

Table 5 Results of the four-way ANOVAs testing group differences in prosodic processing

| | AntCC group versus healthy controls | | | | PostCC group versus healthy controls | | | | AntCC versus postCC groups | | | |
|---|-------------------------------------|------|-------|------------|--------------------------------------|-------------|--------------|------------|----------------------------|-------------|--------------|------------|
| | df | F | P | η_p^2 | df | F | P | η_p^2 | df | F | P | η_p^2 |
| ANOVA: Prosody \times Region \times Hemisphere \times Group (200–500 ms after word onset) | | | | | | | | | | | | |
| P \times G | 1,25 | <1 | NS | 0.012 | 1,26 | 6.47 | 0.017 | 0.200 | 1,9 | 6.77 | 0.029 | 0.429 |
| P \times R \times G | 2,24 | 2.43 | 0.110 | 0.089 | 2,25 | 1.73 | 0.194 | 0.062 | 2,8 | 3.98 | 0.047 | 0.307 |
| P \times H \times G | 1,25 | 3.09 | 0.091 | 0.109 | 1,26 | <1 | NS | 0.000 | 1,9 | 1.28 | 0.287 | 0.124 |
| P \times R \times H \times G | 2,24 | 1.57 | 0.218 | 0.059 | 2,25 | <1 | NS | 0.005 | 2,8 | 1.08 | 0.362 | 0.105 |

Bold values indicate significant results. Partial eta squared $\eta_p^2 > 0.5$ = large effect size; $\eta_p^2 > 0.3$ = medium effect size; $\eta_p^2 > 0.1$ = small effect size (Bortz and Döring, 2003). G = group; H = hemisphere; NS = not significant; P = prosody; R = region; S = syntax. Bold values indicate significant results ($P < 0.05$) and large effect sizes.

diverged, suggesting a comparable sensitivity to prosody alone but a differential interaction between prosody and syntax in both groups. Secondly, the postCC group, similar to healthy controls and the antCC group, exhibited normal prosody-independent syntactic processing, i.e. an intact early detection and late integration of word-category violations as indicated by the ELAN and P600, respectively (Friederici *et al.*, 1993; Hahne and Friederici, 1999; Lau *et al.*, 2006). Furthermore, the behavioural performance in detecting these syntactic errors did not differ between groups. Altogether, these combined findings suggest that the insensitivity of the postCC group towards the syntax–prosody mismatch is not due to deficient syntactic or prosodic processing *per se*, but relates to a disrupted transcallosal interaction of syntactic with prosodic information.

These data, along with the findings of Friederici *et al.* (2007), converge on the splenial and presplenial regions of the corpus callosum as the prosody–syntax interface. Notably, the combined results suggest a balanced bidirectional information exchange via this posterior route. The current data suggest a deficient transfer of syntactic information to the prosodic parser after posterior corpus callosum lesions, while the paradigm employed by

Friederici *et al.* (2007) revealed deficient transfer of prosodic information to the syntactic parser after similar brain damage (for partly converging evidence in the present data, see Results in Supplementary materials showing the absence of a prosody-induced topography shift of the syntactic ELAN in the postCC group). More generally, although both studies tested the prosody–syntax interplay by presenting the relevant prosodic and syntactic information with a temporal delay to determine the directionality of the interaction, it appears highly likely that both processing streams interact at any time of the parsing process reciprocally, dynamically and inseparably.

The relevance of the posterior corpus callosum as the crucial neural substrate for the transfer of auditory linguistic information (for the posterior corpus callosum's role in literacy see Dougherty *et al.*, 2007; Frye *et al.*, 2008; Carreiras *et al.*, 2009; Odegard *et al.*, 2009) is principally in line with previous behavioural studies showing that a reduced connectivity within the isthmus and/or splenium has most pronounced effects on the performance in verbal dichotic listening tasks (Sugishita *et al.*, 1995; Pollmann *et al.*, 2002; Bamiou *et al.*, 2007; Gadea *et al.*, 2009; Westerhausen *et al.*, 2009). Moreover, modern probabilistic maps

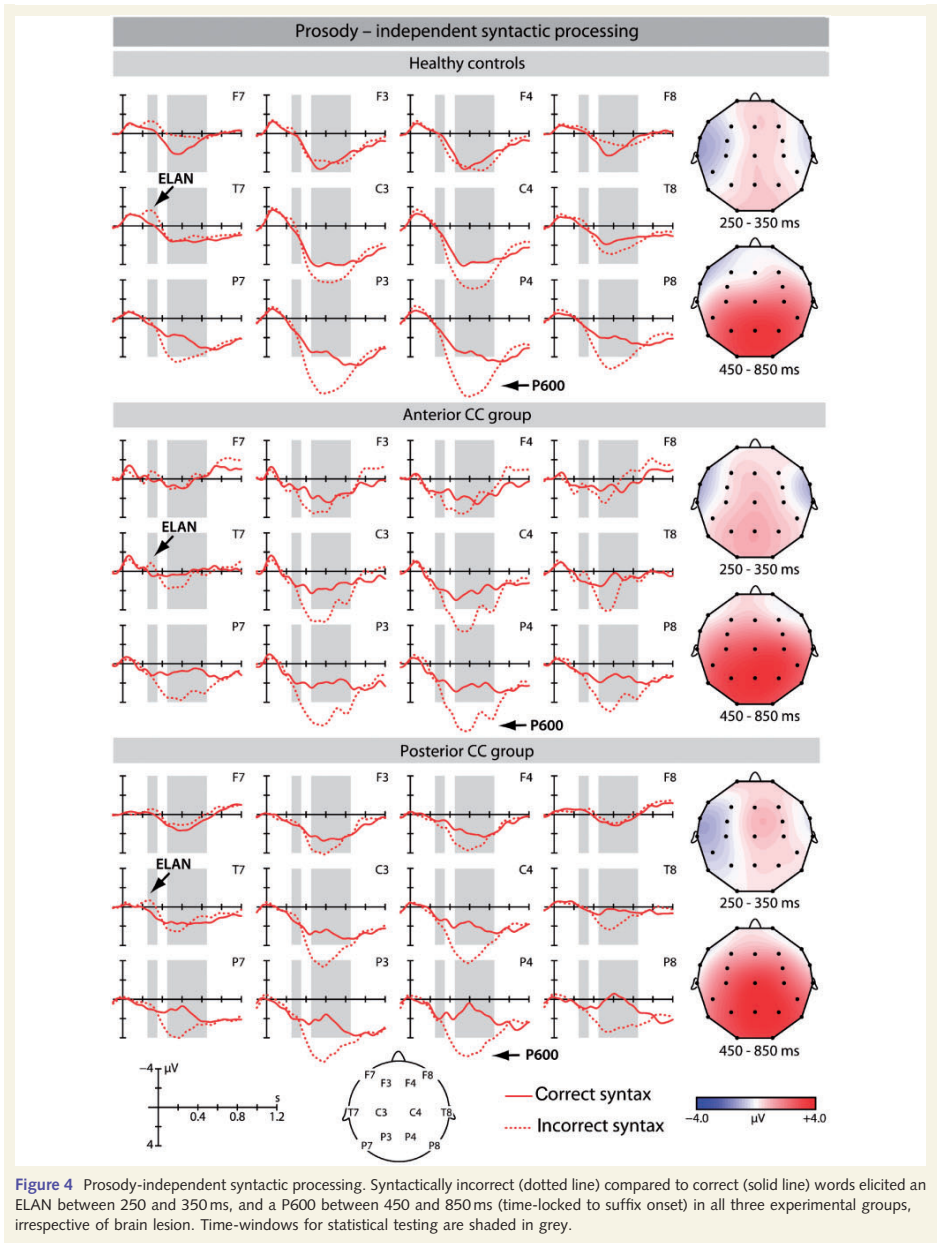


Table 6 Mean amplitude and significance of the ELAN and P600 in separate regions of interest

| | Healthy controls | | | | AntCC | | | | PostCC | | | |
|---|------------------|--------------|-------------------|--------------|--------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| | Δ (μV) | t(21) | P | d | Δ (μV) | t(4) | P | d | Δ (μV) | t(5) | P | d |
| ELAN: Syntax incorrect versus correct (250–350 ms after suffix onset) | | | | | | | | | | | | |
| LT | −1.10 | −5.10 | <0.0001 | 1.537 | −1.04 | −4.29 | 0.013 | 2.716 | −0.88 | −3.70 | 0.015 | 2.135 |
| RT | −0.49 | −2.24 | 0.037 | 0.676 | −0.78 | −1.50 | 0.210 | 0.947 | −0.19 | −0.75 | 0.488 | 0.432 |
| P600: Syntax incorrect versus correct (450–850 ms after suffix onset) | | | | | | | | | | | | |
| LT | −0.74 | −3.41 | 0.003 | 1.027 | +0.36 | 1.31 | 0.263 | 0.827 | −0.18 | −0.62 | 0.563 | 0.357 |
| RT | +0.23 | 0.98 | 0.338 | 0.296 | +0.55 | 2.58 | 0.063 | 1.631 | +0.40 | 0.88 | 0.422 | 0.505 |
| LC | +0.61 | 2.09 | 0.050 | 0.630 | +1.59 | 3.34 | 0.029 | 2.112 | +1.65 | 2.41 | 0.061 | 1.389 |
| RC | +1.29 | 3.59 | 0.002 | 1.084 | +1.56 | 2.71 | 0.054 | 1.712 | +1.99 | 2.68 | 0.044 | 1.547 |
| LP | +2.65 | 9.79 | <0.0001 | 2.953 | +3.04 | 5.08 | 0.007 | 3.210 | +2.62 | 3.99 | 0.011 | 2.301 |
| RP | +2.96 | 10.40 | <0.0001 | 3.134 | +2.98 | 5.33 | 0.006 | 3.371 | +3.29 | 5.54 | 0.003 | 3.200 |

T-tests for paired samples compared event-related potentials evoked by syntactically incorrect and correct words in the respective time windows. Bold values indicate significant results ($P < 0.05$) and large effect sizes. Cohen's $d > 0.8$ = large effect size; $d > 0.5$ = medium effect size; $d > 0.3$ = small effect size (Cohen, 1992). Regions of interest: LT = left fronto-temporal; RT = right fronto-temporal; LC = left fronto-central; RC = right fronto-central; LP = left posterior; RP = right posterior.

obtained with diffusion tensor imaging consistently located inter-hemispheric temporal connections, including auditory fibres, to the (anterior) splenium (Huang *et al.*, 2005; Hofer and Frahm, 2006; Zarei *et al.*, 2006; Park *et al.*, 2008; Westerhausen *et al.*, 2009), and other studies showed a degeneration particularly in splenial and presplenial portions of the corpus callosum as a consequence of temporal lesions (de Lacoste *et al.*, 1985) or temporal lobe epilepsy (Weber *et al.*, 2007). Therefore, the (pre)splenium appears to be the optimal route for the transfer of auditory linguistic information, particularly if processed in the temporal lobes like word-category (Friederici *et al.*, 2000, 2003; Rüschemeyer *et al.*, 2005; Brauer and Friederici, 2007; Hermann *et al.*, 2009) and prosodic intonation (Meyer *et al.*, 2002, 2004; Hesling *et al.*, 2005a,b; Humphries *et al.*, 2005; Ischebeck *et al.*, 2008).

Conclusion

Taken together, the present findings show that the lateralized prosody and syntax processing streams dynamically communicate via the posterior corpus callosum connecting auditory areas in the temporal lobes. Healthy controls and patients with anterior corpus callosum lesions showed an anterior negativity in response to violations of syntactically induced prosodic expectations, demonstrating an intact interaction of syntax and prosody. No such effect was found in patients with lesions in the posterior third of the corpus callosum, although they exhibited intact (prosody-independent) syntactic processing like healthy controls and the antCC group. In keeping with the dynamic dual pathway model (Friederici and Alter, 2004) and combined with earlier findings (Friederici *et al.*, 2007), these data argue for a reciprocal interplay between the speech processing pathways in the left and right hemispheres through the posterior corpus callosum, building the relevant brain basis for the timely coordination and smooth integration of syntactic and prosodic features during auditory speech comprehension.

Acknowledgements

The authors wish to thank Prof. Dr D.Y. von Cramon for kindly providing the patients, Anne-Kathrin Franz for patient recruitment, Heike Böthel for help with data acquisition, Dr Beate Sabisch for parts of the data analysis, Prof. R Rübsem for kindly providing the equipment for auditory testing, Dr Alexandra Ludwig for advice in auditory testing and Kerstin Flake for help with the figures. We also thank the editor and two anonymous reviewers for helpful comments.

Supplementary material

Supplementary material is available at *Brain* online.

References

Anwander A, Tittgemeyer M, von Cramon DY, Friederici AD, Knösche TR. Connectivity-based parcellation of Broca's area. *Cereb Cortex* 2007; 17: 816–25.

Bamiou DE, Sisodiya S, Musiek FE, Luxon LM. The role of the interhemispheric pathway in hearing. *Brain Res Rev* 2007; 56: 170–82.

Baum SR, Pell MD. The neural bases of prosody: Insights from lesion studies and neuroimaging. *Aphasiology* 1999; 8: 581–608.

Blumstein S, Cooper WE. Hemispheric processing of intonation contours. *Cortex* 1974; 10: 146–58.

Bortz J, Döring N. Forschungsmethoden und Evaluation: Für Human- und Sozialwissenschaftler. Berlin: Springer; 2003. p. 603–18.

Brauer J, Friederici AD. Functional neural networks of semantic and syntactic processes in the developing brain. *J Cogn Neurosci* 2007; 19: 1609–23.

Bryan K. Language prosody and the right hemisphere. *Aphasiology* 1989; 3: 285–99.

Brådvik B, Dravins C, Holtås S, Rosén I, Ryding E, Ingvar DH. Disturbances of speech prosody following right hemisphere infarcts. *Acta Neurol Scand* 1991; 84: 114–26.

Bungert-Kahl P, Biedermann F, Dörrscheidt GJ, von Cramon DY, Rübsem R. Psychoacoustic test tools for the detection of deficits in central auditory processing: Normative data. *Z Audiol* 2004; 43: 48–71.

- Caplan D, Alpert N, Waters G. Effects of syntactic structure and propositional number on patterns of regional cerebral blood flow. *J Cogn Neurosci* 1998; 10: 541–52.
- Carreiras M, Seghier ML, Baquero S, Estévez A, Lozano A, Devlin JT, et al. An anatomical signature for literacy. *Nature* 2009; 461: 983–6.
- Catani M, Jones DK, ffytche DH. Perisylvian language networks of the human brain. *Ann Neurol* 2005; 57: 8–16.
- Cohen J. A power primer. *Psychol Bull* 1992; 112: 155–9.
- Cutler A, Dahan D, van Donselaar W. Prosody in the comprehension of spoken language: a literature review. *Lang Speech* 1997; 40 (Pt 2): 141–201.
- Dapretto M, Bookheimer SY. Form and content: dissociating syntax and semantics in sentence comprehension. *Neuron* 1999; 24: 427–32.
- de Lacoste MC, Kirkpatrick JB, Ross ED. Topography of the human corpus callosum. *J Neuropathol Exp Neurol* 1985; 44: 578–91.
- Delorme A, Makeig S. EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Methods* 2004; 134: 9–21.
- Dougherty RF, Ben-Shachar M, Deutsch GK, Hernandez A, Fox GR, Wandell BA. Temporal-callosal pathway diffusivity predicts phonological skills in children. *Proc Natl Acad Sci USA* 2007; 104: 8556–61.
- Eckstein K, Friederici AD. Late interaction of syntactic and prosodic processes in sentence comprehension as revealed by ERPs. *Brain Res Cogn Brain Res* 2005; 25: 130–43.
- Eckstein K, Friederici AD. It's early: event-related potential evidence for initial interaction of syntax and prosody in speech comprehension. *J Cogn Neurosci* 2006; 18: 1696–711.
- Friederici AD. Towards a neural basis of auditory sentence processing. *Trends Cogn Sci* 2002; 6: 78–84.
- Friederici AD. Pathways to language: fibre tracts in the human brain. *Trends Cogn Sci* 2009; 13: 175–81.
- Friederici AD, Alter K. Lateralization of auditory language functions: a dynamic dual pathway model. *Brain Lang* 2004; 89: 267–76.
- Friederici AD, Kotz SA. The brain basis for syntactic processes: functional imaging and lesion studies. *NeuroImage* 2003; 20: 58–17.
- Friederici AD, Pfeifer E, Hahne A. Event-related brain potentials during natural speech processing: effects of semantic, morphological and syntactic violations. *Brain Res Cogn Brain Res* 1993; 1: 183–92.
- Friederici AD, Rüschemeyer SA, Hahne A, Fiebach CJ. The role of left inferior frontal and superior temporal cortex in sentence comprehension: localizing syntactic and semantic processes. *Cereb Cortex* 2003; 13: 170–7.
- Friederici AD, von Cramon DY, Kotz SA. Role of the corpus callosum in speech comprehension: interfacing syntax and prosody. *Neuron* 2007; 53: 135–45.
- Friederici AD, Wang Y, Herrmann CS, Maess B, Oertel U. Localization of early syntactic processes in frontal and temporal cortical areas: a magnetoencephalographic study. *Hum Brain Mapp* 2000; 11: 1–11.
- Frye RE, Hasan K, Xue L, Strickland D, Malmberg B, Liederman J, et al. Splenium microstructure is related to two dimensions of reading skill. *Neuroreport* 2008; 19: 1627–31.
- Gadea M, Marti-Bonmati L, Arana E, Espert R, Salvador A, Casanova B. Corpus callosum function in verbal dichotic listening: inferences from a longitudinal follow-up of relapsing-remitting multiple sclerosis patients. *Brain Lang* 2009; 110: 101–5.
- Gandour J, Tong Y, Wong D, Talavage T, Dziedzic M, Xu Y, et al. Hemispheric roles in the perception of speech prosody. *NeuroImage* 2004; 23: 344–57.
- Gandour J, Wong D, Hsieh L, Weinzaepfel B, Lancker DV, Hutchins GD. A crosslinguistic PET study of tone perception. *J Cogn Neurosci* 2000; 12: 207–22.
- Glasser MF, Rilling JK. DTI tractography of the human brain's language pathways. *Cereb Cortex* 2008; 18: 2471–82.
- Grodzinsky Y, Friederici AD. Neuroimaging of syntax and syntactic processing. *Curr Opin Neurobiol* 2006; 16: 240–6.
- Grodzinsky Y, Santi A. The battle for Broca's region. *Trends Cogn Sci* 2008; 12: 474–80.
- Hagoort P. On Broca, brain, and binding: A new framework. *Trends Cogn Sci* 2005; 9: 416–23.
- Hagoort P, Brown CM, Osterhout L. The neurocognition of syntactic processing. In: Brown CM, Hagoort P, editors. *The neurocognition of language*. New York: Oxford University Press; 1999. p. 273–316.
- Hahne A, Friederici AD. Electrophysiological evidence for two steps in syntactic analysis. Early automatic and late controlled processes. *J Cogn Neurosci* 1999; 11: 194–205.
- Heilman KM, Bowers D, Speedie L, Coslett HB. Comprehension of affective and nonaffective prosody. *Neurology* 1984; 34: 917–21.
- Herrmann B, Maess B, Hasting AS, Friederici AD. Localization of the syntactic mismatch negativity in the temporal cortex: an MEG study. *NeuroImage* 2009; 48: 590–600.
- Hesling I, Clément S, Bordesoules M, Allard M. Cerebral mechanisms of prosodic integration: evidence from connected speech. *NeuroImage* 2005a; 24: 937–47.
- Hesling I, Dilharreguy B, Clément S, Bordesoules M, Allard M. Cerebral mechanisms of prosodic sensory integration using low-frequency bands of connected speech. *Hum Brain Mapp* 2005b; 26: 157–69.
- Hickok G, Poeppel D. The cortical organization of speech processing. *Nat Rev Neurosci* 2007; 8: 393–402.
- Hofer S, Frahm J. Topography of the human corpus callosum revisited—comprehensive fibre tractography using diffusion tensor magnetic resonance imaging. *NeuroImage* 2006; 32: 989–94.
- Huang H, Zhang J, Jiang H, Wakana S, Poetscher L, Miller MI, et al. DTI tractography based parcellation of white matter: application to the mid-sagittal morphology of corpus callosum. *NeuroImage* 2005; 26: 195–205.
- Humphries C, Love T, Swinney D, Hickok G. Response of anterior temporal cortex to syntactic and prosodic manipulations during sentence processing. *Hum Brain Mapp* 2005; 26: 128–38.
- Inkelas S, Zec D. Syntax-phonology interface. In: Goldsmith JA, editor. *The handbook of phonological theory*. Blackwell Handbooks in Linguistics. Vol. 1. Oxford: Blackwell Publishers Ltd; 1996. p. 535–49.
- Ischebeck AK, Friederici AD, Alter K. Processing prosodic boundaries in natural and hummed speech: an fMRI study. *Cereb Cortex* 2008; 18: 541–52.
- Jung-Beeman M. Bilateral brain processes for comprehending natural language. *Trends Cogn Sci* 2005; 9: 512–8.
- Just MA, Carpenter PA, Keller TA, Eddy WF, Thulborn KR. Brain activation modulated by sentence comprehension. *Science* 1996; 274: 114–16.
- Kaan E, Harris A, Gibson E, Holcomb PJ. The P600 as an index of syntactic integration difficulty. *Lang Cognitive Proc* 2000; 15: 159–201.
- Kotz SA, Meyer M, Paulmann S. Lateralization of emotional prosody in the brain: an overview and synopsis on the impact of study design. *Prog Brain Res* 2006; 156: 285–94.
- Lau E, Stroud C, Plesch S, Phillips C. The role of structural prediction in rapid syntactic analysis. *Brain Lang* 2006; 98: 74–88.
- Meyer M, Alter K, Friederici AD, Lohmann G, von Cramon DY. fMRI reveals brain regions mediating slow prosodic modulations in spoken sentences. *Hum Brain Mapp* 2002; 17: 73–88.
- Meyer M, Steinhauer K, Alter K, Friederici AD, von Cramon DY. Brain activity varies with modulation of dynamic pitch variance in sentence melody. *Brain Lang* 2004; 89: 277–89.
- Ni W, Constable RT, Menci WE, Pugh KR, Fulbright RK, Shaywitz SE, et al. An event-related neuroimaging study distinguishing form and content in sentence processing. *J Cogn Neurosci* 2000; 12: 120–33.
- Odegard TN, Farris EA, Ring J, McColl R, Black J. Brain connectivity in non-reading impaired children and children diagnosed with developmental dyslexia. *Neuropsychologia* 2009; 47: 1972–7.
- Oldfield RC. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 1971; 9: 97–113.

- Pannekamp A, Toepel U, Alter K, Hahne A, Friederici AD. Prosody-driven sentence processing: an event-related brain potential study. *J Cogn Neurosci* 2005; 17: 407–21.
- Park HJ, Kim JJ, Lee SK, Seok JH, Chun J, Kim DI, et al. Corpus callosal connection mapping using cortical gray matter parcellation and DT-MRI. *Hum Brain Mapp* 2008; 29: 503–16.
- Parker GJM, Luzzi S, Alexander DC, Wheeler-Kingshott CAM, Ciccarelli O, Ralph MAL. Lateralization of ventral and dorsal auditory-language pathways in the human brain. *NeuroImage* 2005; 24: 656–66.
- Perkins JM, Baran JA, Gandour J. Hemispheric specialization in processing intonation contours. *Aphasiology* 1996; 10: 343–62.
- Plante E, Creusere M, Sabin C. Dissociating sentential prosody from sentence processing: activation interacts with task demands. *NeuroImage* 2002; 17: 401–10.
- Poeppel D. The analysis of speech in different temporal integration windows: Cerebral lateralization as 'asymmetric sampling in time'. *Speech Communication* 2003; 41: 245–55.
- Pollmann S, Maertens M, von Cramon DY, Lepsien J, Hugdahl K. Dichotic listening in patients with splenial and nonsplenial callosal lesions. *Neuropsychology* 2002; 16: 56–64.
- Powell HWR, Parker GJM, Alexander DC, Symms MR, Boulby PA, Wheeler-Kingshott CAM, et al. Hemispheric asymmetries in language-related pathways: a combined functional MRI and tractography study. *NeuroImage* 2006; 32: 388–99.
- Price CJ, Thierry G, Griffiths T. Speech-specific auditory processing: Where is it? *Trends Cogn Sci* 2005; 9: 271–6.
- Rüschmeyer S, Fiebach CJ, Kempe V, Friederici AD. Processing lexical semantic and syntactic information in first and second language: fMRI evidence from German and Russian. *Hum Brain Mapp* 2005; 25: 266–86.
- Saur D, Kreher BW, Schnell S, Kümmerer D, Kellmeyer P, Vry MS, et al. Ventral and dorsal pathways for language. *Proc Natl Acad Sci USA* 2008; 105: 18035–40.
- Scott SK, Johnsrude IS. The neuroanatomical and functional organization of speech perception. *Trends Neurosci* 2003; 26: 100–7.
- Selkirk E. On derived domains in sentence phonology. *Phonology* 1986; 3: 371–405.
- Shalom DB, Poeppel D. Functional anatomic models of language: Assembling the pieces. *Neuroscientist* 2008; 14: 119–27.
- Sharbrough F, Chatrian G, Lesser R, Lüders H, Nuwer M, Picton T. American Electroencephalographic Society guidelines for standard electrode position nomenclature. *J Clin Neurophysiol* 1991; 8: 200–2.
- Shipley-Brown F, Dingwall WO, Berlin CI, Yeni-Komshian G, Gordon-Salant S. Hemispheric processing of affective and linguistic intonation contours in normal subjects. *Brain Lang* 1988; 33: 16–26.
- Sugishita M, Otomo K, Yamazaki K, Shimizu H, Yoshioka M, Shinohara A. Dichotic listening in patients with partial section of the corpus callosum. *Brain* 1995; 118: 417–27.
- Talairach J, Tournoux P. Co-planar Stereotaxic Atlas of The Human Brain. 3-Dimensional proportional system: an approach to cerebral imaging. Stuttgart: Thieme; 1988.
- Truckenbrodt H. On the relation between syntactic phrases and phonological phrases. *Linguist Inq* 1999; 30: 219–55.
- Van Lancker D, Sidtis JJ. The identification of affective-prosodic stimuli by left- and right- hemisphere-damaged subjects: All errors are not created equal. *J Speech Hear Res* 1992; 35: 963–70.
- Vigneau M, Beaucousin V, Hervé PY, Duffau H, Crivello F, Houdé O, et al. Meta-analyzing left hemisphere language areas: Phonology, semantics, and sentence processing. *NeuroImage* 2006; 30: 1414–32.
- Weber B, Lüders E, Faber J, Richter S, Quesada CM, Urbach H, et al. Distinct regional atrophy in the corpus callosum of patients with temporal lobe epilepsy. *Brain* 2007; 130: 3149–54.
- Weintraub S, Mesulam MM, Kramer L. Disturbances in prosody. A right-hemisphere contribution to language. *Arch Neurol* 1981; 38: 742–744.
- Westerhausen R, Grüner R, Specht K, Hugdahl K. Functional relevance of interindividual differences in temporal lobe callosal pathways: a DTI tractography study. *Cereb Cortex* 2009; 19: 1322–9.
- Xiang HD, Fonteijn HM, Norris DG, Hagoort P. Topographical functional connectivity pattern in the perisylvian language networks. *Cereb Cortex* 2010; 20: 549–60.
- Zarei M, Johansen-Berg H, Smith S, Ciccarelli O, Thompson AJ, Matthews PM. Functional anatomy of interhemispheric cortical connections in the human brain. *J Anat* 2006; 209: 311–20.
- Zatorre RJ, Belin P, Penhune V. Structure and function of auditory cortex: music and speech. *Trends Cogn Sci* 2002; 6: 37–46.

Supplementary results – Impact of prosody on syntax

Materials and Methods

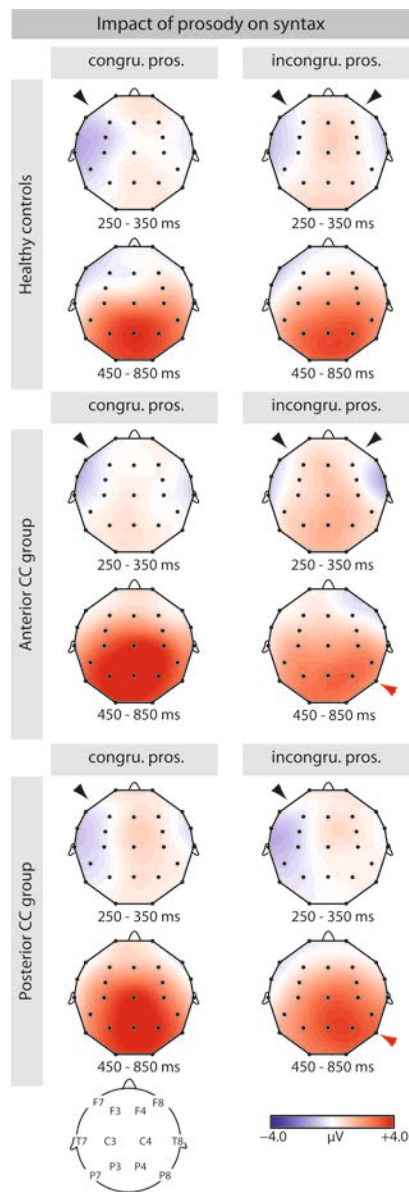
As stated in the main manuscript, half of the sentences contained a syntactic word-category error at the suffix of the critical word (*tade-It/reprov-es*) which should elicit an Early Left Anterior Negativity (ELAN) and P600 (Friederici et al., 1993; Hagoort et al., 1999; Hahne and Friederici, 1999; Kaan et al., 2000). Notably, this syntactic word-category information was always presented later than the prosodic information that was already discernable in the word stem (i.e. congruent rising or incongruent falling pitch contour; for a description of the acoustical properties of the material, see Eckstein and Friederici, 2006). Therefore, the paradigm allowed us to explore in how far the syntax-related ELAN and P600 evoked by the word-category violations are modulated by the preceding prosodic information, or in other words, in how far prosody influences syntax. Presenting these stimuli to 24 healthy students, Eckstein and Friederici (2006) observed that the (typically left-lateralized) ELAN was bilateral, and the P600 amplitude was overadditively enlarged when a syntactic error was preceded by a prosodic violation (for converging evidence, see Eckstein and Friederici, 2005). This prosody-induced topography shift of the ELAN and amplitude enhancement of the P600 may be viewed as additional indicators for a prosody-syntax interaction.

To test this influence of prosody on syntax, difference waves between syntactically incorrect and correct conditions were calculated separately for sentences with (ps minus pc) and without preceding prosodic incongruency (cs minus cc; for an analogous approach, see Eckstein and Friederici, 2006). Statistical analyses were computed on the mean amplitudes of these difference waves in the 250 to 350 ms (ELAN) and 450 to 850 ms time window (P600), in the above mentioned ROIs.

Results

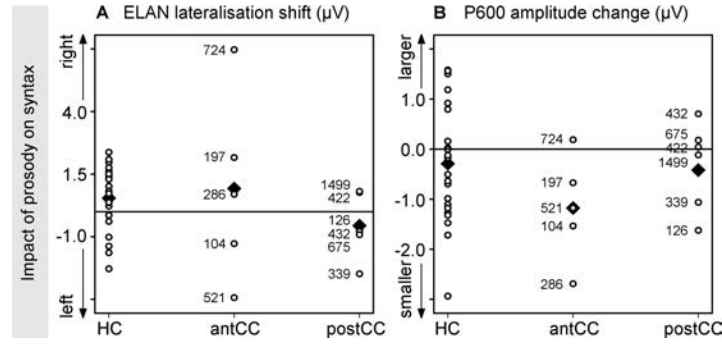
In HC and antCC (but not postCC) patients, syntactic violations elicited an apparently less strongly left lateralised ELAN (250 to 350 ms, time-locked to suffix onset) when preceding sentence prosody (at the word stem) was incongruent compared to when it was congruent

(Supplementary Figures 1 and 2 A). In antCC patients, the P600 (450 to 850 ms) appeared to be reduced in prosodically incongruent sentences, as will be explained below (Supplementary Figure 2 B).



Supplementary Figure 1. Impact of prosody on syntax.

Depicted are topography maps for the difference between syntactically incorrect and correct conditions separately for prosodically congruent and incongruent sentences in the ELAN (250-350 ms) and the P600 time window (450-850 ms). The ELAN was less strongly left lateralised and showed a more bilateral distribution after incongruent compared to congruent prosody in HC, as a trend also in antCC patients, but not in postCC patients (see black arrows in the upper topography maps for 250-350 ms). In HC, the P600 was not modulated by preceding prosodic information; in antCC patients the P600 amplitude was reduced after prosodic incongruities; and postCC patients showed a topography shift to the right after prosodic violations (see red arrows in the lower topography maps for 450-850 ms).



Supplementary Figure 2. Individual data of all participants. **A:** left (negative values) or right shift (positive values) of the ELAN lateralisation when the preceding prosody was incongruent (compared to congruent). While HC and antCC patients (with high variance) showed a right shift, postCC patients exhibited a trend of a left shift. **B:** increase (positive values) or decrease (negative values) of the P600 amplitude when preceding prosody was incongruent (compared to congruent). All groups showed a trend of an amplitude decrease. Diamonds indicate group mean values. Vertical lines cross zero. Numbers indicate patient ID's.

Early time window (250 to 350 ms). In HC, the ANOVA with the factors Congruency (preceding prosody congruent vs. incongruent), Region, and Hemisphere yielded a significant interaction of Congruency \times Hemisphere ($F(1,21) = 4.60, P < .045, \eta^2_p = .179$), indicating that the lateralisation of the ELAN was modulated by preceding prosodic information. Subsequent analyses showed a significant main effect of Hemisphere (i.e. a significant left-lateralisation of the ELAN) only for prosodically congruent sentences ($F(1,21) = 11.09, P < .004, \eta^2_p = .346$), not during incongruent conditions ($F(1,21) = 3.54, P > .074, \eta^2_p = .144$). Based on the findings of Eckstein and Friederici (2006), additional analyses were conducted specifically for the left and right temporal ROIs. In the left temporal ROI, a highly significant negativity (the ELAN) was found during prosodically congruent trials ($\Delta = -1.24 \mu V, F(1,21) = 23.17, P < .0001, \eta^2_p = .526$), that was slightly reduced in amplitude during incongruent conditions ($\Delta = -0.98 \mu V, F(1,21) = 17.08, P < .0006, \eta^2_p = .449$). In the right temporal ROI, a small and non-significant negativity was observed during congruent conditions ($\Delta = -0.35 \mu V, P > .255, \eta^2_p = .061$), that was slightly stronger and reached significance during incongruent trials ($\Delta = -0.63 \mu V, F(1,21) = 7.74, P < .012, \eta^2_p = .269$). Taken

together, the ELAN in HC lost its typical left-lateralisation due to the emergence of a right-temporal negativity when a prosodic violation preceded the word-category violation (see black arrows in the upper panel of Supplementary Figure 1).

A similar topography shift was observed in patients with antCC lesions (see black arrows in the middle panel of Supplementary Figure 1): In the left temporal ROI, the ELAN amplitude was larger in prosodically congruent ($\Delta = -1.23 \mu\text{V}$) compared to incongruent sentences ($\Delta = -0.53 \mu\text{V}$), whereas the reverse was found for right temporal regions (congruent: $\Delta = -0.87 \mu\text{V}$; incongruent: $\Delta = -1.09 \mu\text{V}$). Although this topography shift did not reach significance (no interaction of Congruency \times Hemisphere; $P > .778$, $\eta^2_p = .022$; perhaps due to the large interindividual variance; Supplementary Figure 2 A) the ERP pattern in antCC patients did not significantly differ from HC (no interaction involving Group in an ANOVA with the factors Congruency, Hemisphere and Group computed for temporal ROIs only; P 's $> .687$, $\eta^2_p < .007$).

Patients with postCC lesions exhibited a reverse pattern compared to the other groups (see black arrows in the lower panel of Supplementary Figure 1): In the left temporal ROI, a robust and significant negativity was found during prosodically congruent sentences ($\Delta = -0.83 \mu\text{V}$, $F(1,5) = 15.85$, $P < .011$, $\eta^2_p = .760$), that was virtually larger (not smaller) during incongruent trials ($\Delta = -0.94 \mu\text{V}$, $F(1,5) = 6.77$, $P < .049$, $\eta^2_p = .575$). In the right temporal ROI, a decrease (not increase) of the negativity's amplitude was found if preceding prosody was violated ($\Delta = +0.03 \mu\text{V}$, $F < 1$, n.s., $\eta^2_p = .002$) compared to correct ($\Delta = -0.41 \mu\text{V}$, $F(1,5) = 2.23$, $P > .195$, $\eta^2_p = .308$). This ERP pattern differed significantly from HC as indicated by a marginal Congruency \times Hemisphere \times Group interaction in the three-way ANOVA computed for temporal ROIs only ($F(1,26) = 3.52$, $P < .072$, $\eta^2_p = .119$). No difference was found between antCC and postCC patients, possibly due to the high variance in the antCC group ($P > .381$, $\eta^2_p = .086$). Altogether, these data suggest that particularly the brain potentials of postCC patients deviate from the processing in HC and antCC patients.

Late time window (450 to 850 ms). In HC, the P600 amplitude was not modulated by preceding prosodic incongruity (no interaction involving the factor Congruency; all P 's $> .262$, $\eta^2_p < .062$), contrary to previous findings (Eckstein and Friederici, 2005, 2006). Likewise, in antCC patients,

the three-way ANOVA did not reveal a significant influence of prosody on the P600, although exploratory analyses in each of the six ROIs yielded a marginally significantly reduced P600 amplitude after prosodic violations in posterior ROIs (left: $F(1,4) = 5.33$, $P < .083$, $\eta^2_p = .571$; right: $F(1,4) = 6.28$, $P < .067$, $\eta^2_p = .611$; see red arrow in the middle panel of Supplementary Figure 1). In postCC patients, the three-way ANOVA yielded a significant interaction of Congruency \times Hemisphere ($F(1,5) = 12.62$, $P < .016$, $\eta^2_p = .716$). Subsequent analyses showed that the P600 was more strongly right-lateralised during incongruent (left: $\Delta = +0.99 \mu V$; right: $\Delta = +1.98 \mu V$; $F(1,5) = 4.88$, $P < .079$, $\eta^2_p = .494$) than during congruent conditions (left: $\Delta = +1.69 \mu V$; right: $\Delta = +1.78 \mu V$; $F < 1$, n.s., $\eta^2_p = .023$; see red arrow in the lower panel of Supplementary Figure 1). The impact of prosody on syntactic processing did, however, not significantly differ between the three experimental groups (no interactions involving the factor Group in the four-way ANOVA; all P 's $> .123$, $\eta^2_p < .112$).

Discussion

The modulation of the syntactic ELAN as a function of preceding prosodic information in HC and antCC but not postCC patients mirrors the pattern of results reported in the main manuscript on the impact of syntax on prosody. In HC, the ELAN lost its left-lateralisation, showing up bilaterally when preceding sentence prosody was incongruent compared to when it was congruent. This prosody-induced topography shift of the ELAN replicates previous findings in healthy students (Eckstein and Friederici, 2006) and indicates an immediate influence of phrasal prosody on early syntactic parsing stages. As proposed by Eckstein and Friederici (2006), the changed scalp distribution of the ELAN most likely represents an additional recruitment of right-hemispheric resources in syntactic structure building after prosodic errors. The absence of such a topography shift in postCC patients suggests that prosodic information does not influence early steps of syntactic analysis after disconnection of posterior callosal fibers, supporting the notion that the posterior third of the CC is particularly relevant for interfacing prosodic and syntactic information during auditory language comprehension. AntCC patients showed a topography shift very similar to HC, although non-significant most likely due to the high variance of the data. It remains to be clarified whether this variability indicates at least a partial

transfer of prosodic-syntactic information through the anterior CC, as also suggested by the data of Friederici et al. (2007). This is conceivable because, apart from temporal generators, the ELAN has additional sources in inferior-frontal brain areas (Friederici et al., 2000; Knösche et al., 1999) communicating via the anterior CC (Hofer and Frahm, 2006; Huang et al., 2005; Park et al., 2008). Overall, these findings converge with the results of Friederici et al. (2007), adding weight to the assumption that prosodic information guides syntactic processing predominantly via the posterior CC.

The prosodic impact on late syntactic integration as indicated by the P600 (Friederici et al., 1993; Hagoort et al., 1999; Hahne and Friederici, 1999; Osterhout and Holcomb, 1992) requires a more extensive explanation. Contrary to the findings of Eckstein and Friederici (2005, 2006), the average P600 amplitude in HC was not enhanced when preceded by a prosodic incongruity. One reason for this finding may relate to the higher mean age of our participants compared to the previously tested students, and adult age differences in the use of prosody in syntactic parsing. It has been shown that elderly exploit prosody to a greater extent during syntactic parsing and have more difficulties in ignoring conflicting prosody than younger adults (Kjelgaard et al., 1999; Steinhauer et al., 2010; Titone et al., 2006; Wingfield et al., 1992) possibly due to age-related declines in some aspects of working memory, attentional and perceptual processes (Wingfield et al., 1988). Consequently, our HC may have been more distracted by the prosodic violations than younger participants, leading to allocation of less attention to the actual syntactic violation after a prosodic error. The weaker grammaticality judgments in prosodically incongruent compared to congruent sentences support this assumption. As the strength of the P600 depends on attention (Coulson et al., 1998; Gunter and Friederici, 1999), a reduction of the P600 amplitude due to attention distracting prosodic errors might have cancelled out the eventual enhancement of the P600 amplitude related to the prosody-syntax interaction. Along these lines, one may speculate that the significant reduction (instead of enhancement) of the P600 amplitude after prosodic violations in the antCC patients is due to their even stronger distractibility (compared to HC) possibly due to the lesion of callosal fibers normally connecting prefrontal areas involved in attentional processes (Cabeza and Nyberg, 2000) leading not only to a neutralisation but even reversal of the effect observed by Eckstein and Friederici (2005,

2006). Likewise, it remains unclear, whether the topography shift of the P600 to the right observed in postCC patients after prosodic violations actually indicates an interaction between prosody and syntax (e.g., via the anterior commissure also connecting the temporal lobes; Bamiou et al., 2007), or relates rather to age-related changes of the P600 scalp topography (Faustmann et al., 2007; Kemmer et al., 2004; Steinhauer et al., 2010). Altogether, the P600 amplitude might be confounded by the influence of general age-related attentional and working memory capacities on the integration of syntax and prosody. Therefore, further investigations are necessary before drawing final conclusions from these P600 data with respect to callosal transfer of prosodic and syntactic information.

References

- Cabeza R, Nyberg L. Imaging cognition II: An empirical review of 275 PET and fMRI studies. *J Cogn Neurosci* 2000; 12: 1–47.
- Coulson S, King J, Kutas M. Expect the unexpected: Event-related brain responses of morphosyntactic violations. *Language and Cognitive Processes* 1998; 13: 21–58.
- Faustmann A, Murdoch BE, Finnigan SP, Copland DA. Effects of advancing age on the processing of semantic anomalies in adults: evidence from event-related brain potentials. *Exp Aging Res* 2007; 33: 439–460.
- Gunter TC, Friederici AD. Concerning the automaticity of syntactic processing. *Psychophysiology* 1999; 36: 126–137.
- Kemmer L, Coulson S, Ochoa ED, Kutas M. Syntactic processing with aging: An event-related potential study. *Psychophysiology* 2004; 41: 372–384.
- Kjelgaard MM, Titone DA, Wingfield A. The influence of prosodic structure on the interpretation of temporary syntactic ambiguity by young and elderly listeners. *Experimental Aging Research* 1999; 25: 187–207.
- Knösche TR, Maess B, Friederici AD. Processing of syntactic information monitored by brain surface current density mapping based on MEG. *Brain Topogr* 1999; 12: 75–87.

Osterhout L, Holcomb PJ. Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language* 1992; 31: 785–806.

Steinhauer K, Abada SH, Pauker E, Itzhak I, Baum SR. Prosody-syntax interactions in aging: event-related potentials reveal dissociations between on-line and off-line measures. *Neurosci Lett* 2010; 472: 133–138.

Titone DA, Koh CK, Kjelgaard MM, Bruce S, Speer SR, Wingfield A. Age-related impairments in the revision of syntactic misanalyses: effects of prosody. *Lang Speech* 2006; 49: 75–99.

Wingfield A, Stine EA, Lahar CJ, Aberdeen JS. Does the capacity of working memory change with age? *Exp Aging Res* 1988; 14: 103–107.

Wingfield A, Wayland SC, Stine EA. Adult age differences in the use of prosody for syntactic parsing and recall of spoken sentences. *J Gerontol* 1992; 47: P350–P356.

Supplementary example stimuli

(literal English translations are given in parentheses)

Maria weiß, dass der Bauer beim ... [Maria knows that the farmer near the ...]

cc/pc: ... *Acke-r_[noun] rastet.* [... field_[noun] rests.]

cs/ps: ... *acke-rt rastet.* [... plows_[verb:3rd/sing] rests.]

Felix weiß, dass das Papier im ... [Felix knows that the paper in the ...]

cc/pc: ... *Filte-r_[noun] modert.* [... filter_[noun] moulders.]

cs/ps: ... *filte-rt_[verb:3rd/sing] modert.* [... filters_[verb:3rd/sing] moulders.]

Sonja weiß, dass der Gauner beim ... [Sonja knows that the chiseller in the ...]

cc/pc: ... *Hande-l_[noun] schwindelt.* [... deal_[noun] cheats.]

cs/ps: ... *hande-lt_[verb:3rd/sing] schwindelt.* [... deals_[verb:3rd/sing] cheats.]

Maren weiß, dass die Nonne trotz ... [Maren knows that the nun despite ...]

cc/pc: ... *Hunge-r_[noun] fastet.* [... hunger_[noun] fasts.]

cs/ps: ... *hunge-rt_[verb:3rd/sing] fastet.* [... starves_[verb:3rd/sing] fasts.]

Silke hört, dass der Sprecher im ... [Silke hears that the speaker in the ...]

cc/pc: ... *Jube-l*_[noun] *mitsingt*. [... jubilation_[noun] sings along.]

cs/ps: ... *jube-lt*_[verb:3rd/sing] *mitsingt*. [... jubilates_[verb:3rd/sing] sings along.]

Paula sieht, dass der Sportler zum ... [Paula sees that the sportsman to the ...]

cc/pc: ... *Kege-l*_[noun] *deutet*. [... bowling pin_[noun] points.]

cs/ps: ... *kege-lt*_[verb:3rd/sing] *deutet*. [... bowls_[verb:3rd/sing] points.]

Frederik sieht, dass der Priester vorm ... [Frederik sees that the priest in front of the ...]

cc/pc: ... *Opfe-r*_[noun] *wartet*. [... sacrifice_[noun] waits.]

cs/ps: ... *opfe-rt*_[verb:3rd/sing] *wartet*. [... sacrifices_[verb:3rd/sing] waits.]

Franziska sieht, dass der Camper trotz ... [Franziska sees that the camper despite ...]

cc/pc: ... *Padde-l*_[noun] *kentert*. [... paddle_[noun] capsizes.]

cs/ps: ... *padde-lt*_[verb:3rd/sing] *kentert*. [... paddles_[verb:3rd/sing] capsizes.]

Matthias weiß, dass der Jäger beim ... [Matthias knows that the hunter near the ...]

cc/pc: ... *Köde-r*_[noun] *lauert*. [... lure_[noun] lurks.]

cs/ps: ... *köde-rt*_[verb:3rd/sing] *lauert*. [... lures_[verb:3rd/sing] lurks.]

Manfred weiß, dass der Bäcker beim ... [Manfred knows that the baker with the ...]

cc/pc: ... *Zucke-r*_[noun] *knausert*. [... sugar_[noun] scrimps.]

cs/ps: ... *zucke-rt*_[verb:3rd/sing] *knausert*. [... sugars_[verb:3rd/sing] scrimps.]

Doreen weiß, dass das Mädchen im ... [Doreen knows that the girl in the ...]

cc/pc: ... *Schlumme-r*_[noun] *brabbelt*. [... nap_[noun] babbles.]

cs/ps: ... *schlumme-rt*_[verb:3rd/sing] *brabbelt*. [... naps_[verb:3rd/sing] babbles.]

Jutta weiß, dass der Schwimmer mit ... [Jutta knows that the swimmer with ...]

cc/pc: ... *Schnorche-l*_[noun] *startet*. [... snorkel_[noun] starts.]

cs/ps: ... *schnorche-lt*_[verb:3rd/sing] *startet*. [... snorkels_[verb:3rd/sing] starts.]

Sabine weiß, dass der Prüfling als ... [Sabine knows that the examinee as ...]

cc/pc: ... *Schuste-r*_[noun] *durchfällt*. [... cobbler_[noun] fails.]

cs/ps: ... *schuste-rt*_[verb:3rd/sing] *durchfällt*. [... cobbles_[verb:3rd/sing] fails.]

Henning sieht, dass der Kämpfer im ... [Henning sees that the warrior in the ...]

cc/pc: ... *Feue-r*_[noun] *umkommt*. [... fire_[noun] dies.]

- cs/ps: ... *feue-rt*_[verb:3rd/sing] *umkommt*. [... *fires*_[verb:3rd/sing] *dies*.]
- Steffen sieht, dass der Lehrer beim ...* [Steffen sees that the teacher during the ...]
- cc/pc: ... *Tade-l*_[noun] *schmunzelt*. [... *reproval*_[noun] *smiles*.]
- cs/ps: ... *tade-lt*_[verb:3rd/sing] *schmunzelt*. [... *reproves*_[verb:3rd/sing] *smiles*.]
- Britta weiß, dass der Forscher trotz ...* [Britta knows that the researcher despite ...]
- cc/pc: ... *Zweife-l*_[noun] *zusagt*. [... *doubt*_[noun] *accepts*.]
- cs/ps: ... *zweife-lt*_[verb:3rd/sing] *zusagt*. [... *doubts*_[verb:3rd/sing] *accepts*.]
- Daniela sieht, dass der Bursche mit ...* [Daniela sees that the boy with ...]
- cc/pc: ... *Bünde-l*_[noun] *wandert*. [... *bundle*_[noun] *hikes*.]
- cs/ps: ... *bünde-lt*_[verb:3rd/sing] *wandert*. [... *bundles*_[verb:3rd/sing] *hikes*.]
- Kevin sieht, dass der Lehrling beim ...* [Kevin sees that the trainee near the ...]
- cc/pc: ... *Male-r*_[noun] *faulenz*. [... *painter*_[noun] *lazes*.]
- cs/ps: ... *male-rt*_[verb:3rd/sing] *faulenz*. [... *paints*_[verb:3rd/sing] *lazes*.]
- Martin sieht, dass der Cowboy am ...* [Martin sees that the cowboy at the ...]
- cc/pc: ... *Satte-l*_[noun] *zupackt*. [... *saddle*_[noun] *grips*.]
- cs/ps: ... *satte-lt*_[verb:3rd/sing] *zupackt*. [... *saddles*_[verb:3rd/sing] *grips*.]
- Simon sieht, dass der Zocker zum ...* [Simon sees that the gambler to the ...]
- cc/pc: ... *Würfe-l*_[noun] *blinzelt*. [... *dice*_[noun] *squints*.]
- cs/ps: ... *würfe-lt*_[verb:3rd/sing] *blinzelt*. [... *dices*_[verb:3rd/sing] *squints*.]

3.2 Study V – Song melody and lyrics

Sammler, D., Baird, A., Valabrègue, R., Clément, S., Dupont, S., Belin, P., Samson, S. (2010). The relationship of lyrics and tunes in the processing of unfamiliar songs: An fMR-adaptation study. *The Journal of Neuroscience*, 30, 3572-3578.*

[Author contributions by D.S.: designed research with A.B., P.B., S.S., received materials from S.C. and S.D., performed research, analysed data with R.V., wrote the paper, incorporating edits by co-authors]

*Reprinted with permission from The Society for Neuroscience.

The Relationship of Lyrics and Tunes in the Processing of Unfamiliar Songs: A Functional Magnetic Resonance Adaptation Study

Daniela Sammler,^{1,2,3} Amee Baird,¹ Romain Valabrègue,^{3,4} Sylvain Clément,¹ Sophie Dupont,^{2,4} Pascal Belin,^{5,6} and Séverine Samson^{1,2}

¹Neuropsychologie et Cognition Auditive-JE2497, Université de Lille-Nord de France, 59653 Villeneuve d'Ascq, France, ²Hôpital de la Pitié-Salpêtrière, 75651 Paris, France, ³Centre de Neuroimagerie de Recherche, Paris, France, ⁴Centre de Recherche de l'Institut du Cerveau et de la Moëlle Épinrière, Université Pierre et Marie Curie–Unité Mixte de Recherche 7225 Centre National de la Recherche Scientifique–UMRS 975 INSERM, Paris, France, ⁵Centre for Cognitive Neuroimaging, Department of Psychology, University of Glasgow, Glasgow G12 8QB, United Kingdom, and ⁶Laboratories for Brain, Music, and Sound, Université de Montréal, Montréal, Québec H3C 3J7 and McGill University, Montréal, Québec H3A 2T5, Canada

The cognitive relationship between lyrics and tunes in song is currently under debate, with some researchers arguing that lyrics and tunes are represented as separate components, while others suggest that they are processed in integration. The present study addressed this issue by means of a functional magnetic resonance adaptation paradigm during passive listening to unfamiliar songs. The repetition and variation of lyrics and/or tunes in blocks of six songs was crossed in a 2×2 factorial design to induce selective adaptation for each component. Reductions of the hemodynamic response were observed along the superior temporal sulcus and gyrus (STS/STG) bilaterally. Within these regions, the left mid-STS showed an interaction of the adaptation effects for lyrics and tunes, suggesting an integrated processing of the two components at prelexical, phonemic processing levels. The degree of integration decayed toward more anterior regions of the left STS, where the lack of such an interaction and the stronger adaptation for lyrics than for tunes was suggestive of an independent processing of lyrics, perhaps resulting from the processing of meaning. Finally, evidence for an integrated representation of lyrics and tunes was found in the left dorsal precentral gyrus (PrCG), possibly relating to the build-up of a vocal code for singing in which musical and linguistic features of song are fused. Overall, these results demonstrate that lyrics and tunes are processed at varying degrees of integration (and separation) through the consecutive processing levels allocated along the posterior–anterior axis of the left STS and the left PrCG.

Introduction

Song is one of the richest formats of human communication, as it tightly binds verbal and musical information. A contemporary debate in music cognition research concerns the relationship between lyrics and tunes in the processing of song. Several lines of evidence suggest a separate processing of both components, as demonstrated by the better performance of nonfluent aphasics in producing the melody than the lyrics of songs (Hébert et al., 2003; Racette et al., 2006), the dissociation of lyrics and tunes in song memory after temporal lobe damage [Samson and Zatorre

(1991), their Experiment 2; Peretz (1996); Hébert and Peretz (2001)], or the differential brain signatures in healthy participants during listening to melodic or semantic errors in familiar songs (Besson et al., 1998). In contrast, other studies suggest an integrated processing of lyrics and tunes, as shown by the interaction between the perception of single pitches and vowels (Lidji et al., 2009), harmonic and phonemic information (Bigand et al., 2001), or lexical and semantic information (Poulin-Charronnat et al., 2005; see also Schön et al., 2005), as well as the failure of listeners to ignore the lyrics when required to recognize the tunes of songs, and vice versa [Serafine et al. (1984, 1986); Crowder et al. (1990); Samson and Zatorre (1991), their Experiment 1].

These divergent accounts are, however, not necessarily mutually exclusive. Rather, they may represent the extremes of a continuum with a more or less accentuated integration/dissociation at different stages of song perception, production, and memory. The present study examined the degree of integration (or separation) for song perception by means of a functional magnetic resonance (fMR)-adaptation paradigm. This method is based on the observation that the repetition of certain stimulus features reduces the activity in neuronal populations involved in representing these features (Grill-Spector, 2006; Krekelberg et al.,

Received May 26, 2009; revised Oct. 15, 2009; accepted Dec. 1, 2009.

This study was supported by a grant from "Agence Nationale pour la Recherche" of the French Ministry of Research (Project NT05-3_45987) to S.S. We sincerely acknowledge Bernard Bouchard, who constructed the stimulus material. We are grateful to Stéphane Lehericy, Eric Bardin, Eric Bertasi, and Kévin Nigaud of the Centre de Neuroimagerie de Recherche for great support during study preparation, fMRI data acquisition, and analysis. We also thank Aurélie Pimienta, Séverine Gilbert, and Amélie Coisne for help during data collection and Barbara Tillmann, Renée Béland, Emmanuel Bigand, Daniele Schön, and Hervé Platel for always fruitful discussion.

Correspondence should be addressed to either of the following at their present addresses: Prof. Séverine Samson, UFR de Psychologie, Université de Lille-Nord de France, BP 60 149, 59653 Villeneuve d'Ascq Cedex, France. E-mail: severine.samson@univ-lille3.fr; or Dr. Daniela Sammler, Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, 04103 Leipzig, Germany. E-mail: sammler@cps.mpg.de.

DOI:10.1523/JNEUROSCI.2751-09.2010

Copyright © 2010 the authors 0270-6474/10/303572-07\$15.00/0

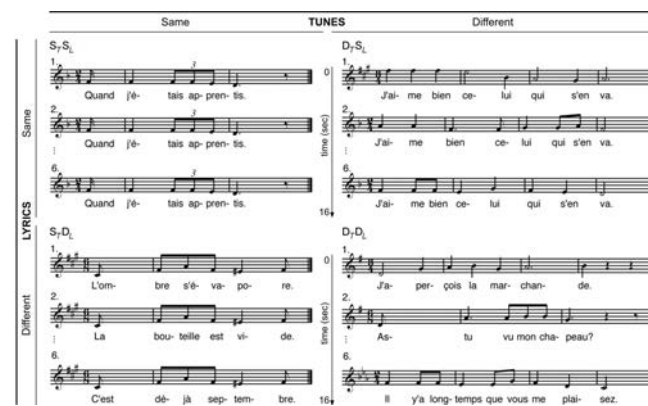


Figure 1. Experimental design. The repetition or variation of lyrics and/or tunes within blocks of six songs was crossed in a 2×2 factorial design.

2006). This response reduction, also referred to as repetition suppression or neural priming, might reflect the dynamic tuning of the perceptual apparatus and represent the neurophysiological basis of the implicit build-up of perceptual memory representations (Henson, 2003). This approach has been successfully used to study a variety of higher cognitive functions, such as the processing of numbers (Naccache and Dehaene, 2001), voices (Belin and Zatorre, 2003), or language (Dehaene-Lambertz et al., 2006).

We applied a variant of the adaptation paradigm to induce selective adaptation effects for lyrics and tunes during passive listening to unfamiliar songs. Blocks of six short songs (sung by different singers to rule out repetition effects for voice) were presented. The repetition or variation of lyrics and/or tunes within blocks was crossed in a 2×2 factorial design. We predicted that brain regions sensitive to the respective component (lyrics or tunes) would be less strongly activated in blocks in which that property was repeated compared to when it varied. In addition, we hypothesized that a significant interaction between the adaptation effects for the two components would be shown by any brain regions that integrate the processing of lyrics and tunes. The lack of such an interaction would specify brain regions that process lyrics and/or tunes independently, along a continuum between integration and separation.

Materials and Methods

Participants. The study was conducted with 12 healthy French native speakers (6 women, 6 men, mean age: 29 years, mean education: 15.58 years). All participants were right handed (mean laterality quotient: 82.64%) as assessed by the Edinburgh Handedness Inventory (Oldfield, 1971), and reported to have normal hearing. None of the participants was a professional musician or actively playing an instrument at the time of testing (mean years of musical training: 1.92 years). Written informed consent was obtained from each participant before the study, which was approved by the local ethics committee.

Materials. One hundred sixty-eight short unfamiliar songs with different (meaningful) lyrics and tunes were created by a professional composer based on a collection of 19th century French folk songs (Robine, 1994). Each song had an average of 7.65 notes and 5.61 words. Major (A, B, Eb, Bb, F, C, G, D, A, E, F#) and minor (bb, f, c, g, d, a, e, b, f#) mode and duple (2/4 or 4/4) and triple (3/4 or 6/8) time were balanced in the stimulus set.

All the songs were recorded by six trained singers (two sopranos, one alto, two tenors, and one bass; mean years of singing lessons: 5.3 years) in a sound studio, cut to 2500 ms, and normalized to -6 dB SPL using Adobe Audition 3 (Adobe Systems). Infrequent, slightly imprecisely sung pitches were adjusted using Celemony Melodyne Studio 3 (Celemony Software). Subsequently, 48 stimulus blocks were constructed consisting of six songs separated by 200 ms pauses resulting in a block duration of 16 s. To rule out potential adaptation to the singers' voices (Belin and Zatorre, 2003) or to simple pitch repetition, each song within a block was sung by another singer of varying age (range: 21–52 years) and sex (three men, three women), at an octave that best corresponded to the singer's voice (soprano, alto and tenor, bass). Consequently, pitch (i.e., octave) and voice-related parameters considerably varied within all 48 blocks, providing no basis for neuronal adaptation to singer's voice (in none of the four conditions described below) (Grill-Spector, 2006; Krekelberg et al., 2006). However, as a footnote it should be said that this manipulation does not completely exclude that

the changing voices may differentially interact with the adaptation for lyrics or tunes. Across blocks, each singer's voice occurred with equal probability at any of the six song positions.

There were four types of blocks corresponding to the four experimental conditions: (1) 12 blocks containing songs with the same tunes and same lyrics (S_7S_7), (2) 12 blocks with the same tunes but different lyrics (S_7D_7), (3) 12 blocks with different tunes but same lyrics (D_7S_7), and (4) 12 blocks with different tunes and different lyrics (D_7D_7) (Fig. 1; stimulus examples are available at www.jneurosci.org as supplemental material). There were no significant differences in word/note number, word/note length, word frequency according to LEXIQUE 2 (New et al., 2004), duple and triple time, major and minor modes, interval size, and number of contour reversals between conditions as revealed by a multivariate one-way ANOVA with the fixed factor condition (S_7S_7 vs S_7D_7 vs D_7S_7 vs D_7D_7) calculated for all these variables (p values >0.220) (see supplemental Table 1, available at www.jneurosci.org as supplemental material). To avoid adaptation to phonology, semantic content, or syntactic structure (Noppeney and Price, 2004), lyrics within S_7D_7 and D_7D_7 blocks did not rhyme, were semantically distant, and differed with respect to syntactic structure.

Procedure. Each participant was presented with one of four pseudo-randomizations of the 48 blocks. These were intermixed in a way that no more than two blocks of the same condition followed each other, and that transition probabilities between conditions were balanced. Interblock intervals were 10 ± 0.5 s to allow the hemodynamic response to return to baseline (Belin and Zatorre, 2003). Stimuli were presented using E-Prime 1.1 (Psychology Software Tools), and delivered binaurally through air pressure headphones (MR confon). The participants' task was to listen attentively with closed eyes, and to not hum or sing along with the melodies. After scanning, all participants rated on nine-point scales (1 = not at all, 9 = always) how attentively they had listened to the songs (mean: 7.75), and whether they had sung along overtly (mean = 0) or covertly (mean = 3.92) during the scan, confirming that they had followed the instructions. The duration of the experiment was ~ 30 min.

Scanning. Functional magnetic resonance imaging (fMRI) was performed on a 3T Siemens TRIO scanner (Siemens) at the Centre de Neuroimagerie de Recherche at the Salpêtrière Hospital in Paris. Before the functional scans, high-resolution T1-weighted images ($1 \times 1 \times 1$ mm voxel size) were acquired for anatomical coregistration using a magnetization-prepared rapid acquisition gradient echo (MPRAGE) sequence (TR = 2300 ms, TE = 4.18 ms). Subsequently, one series of 595 blood oxygenation level-dependent (BOLD) images was acquired using a single-shot echo-planar gradient-echo (EPI) pulse sequence (TR = 2120

ms, TE = 25 ms, the first six volumes were later discarded to allow for T1 saturation). Forty-four interleaved slices ($3 \times 3 \times 3$ mm voxel size, 10% interslice gap) perpendicular with respect to the hippocampal plane were collected with a head coil. The field of view was 192×192 mm with an in-plane resolution of 64×64 pixels and a flip angle of 90° . Scanner noise was continuous during the experiment, representing a constant auditory background.

Data analysis. fMRI data were analyzed using SPM5 (Wellcome Department of Imaging Neuroscience). Preprocessing of the functional data included spatial realignment, coregistration of the functional and anatomical data, spatial normalization into the MNI stereotactic space, and spatial smoothing using a 3D Gaussian kernel with 8 mm full-width at half-maximum (FWHM). Low-frequency drifts were eliminated using a temporal high-pass filter with a cutoff of 200 s. Statistical evaluation was performed using the general linear model (GLM). Four regressors were modeled (one for each of the four conditions) using boxcar functions convolved with a hemodynamic response function (HRF). In addition, estimated motion parameters were included as covariates of no interest to increase statistical sensitivity.

The combined brain activations of all four listening conditions were contrasted against baseline (all > baseline). Linear contrasts pertaining to the main effect of lyric repetition, i.e., $[(S_T D_L + D_T D_L) - (D_T S_L + S_T S_L)]$, the main effect of tune repetition, i.e., $[(D_T S_L + D_T D_L) - (S_T D_L + S_T S_L)]$, and the interactions of the factorial design were calculated, i.e., $[(S_T D_L + D_T S_L) - (D_T D_L + S_T S_L)]$ and $[(D_T D_L + S_T S_L) - (S_T D_L + D_T S_L)]$. To identify brain regions that showed stronger adaptation for lyrics than for tunes and vice versa, both main effects were contrasted, i.e., $[2 \times (D_T S_L - S_T D_L)]$ and $[2 \times (S_T D_L - D_T S_L)]$. To illustrate the influence of the repetition/variation of one component (lyrics or tunes) on the adaptation for the other, we also computed the four contrasts pertaining to the simple effects, i.e., $[D_T D_L - D_T S_L]$, $[S_T D_L - S_T S_L]$, $[D_T D_L - S_T D_L]$, and $[D_T S_L - S_T S_L]$.

For random effect group analyses, the individual contrast images were submitted to one-sample *t* tests. All SPMs were thresholded at $p < 0.001$, cluster extent $k \geq 5$ voxels. In a first step, only activations that survived the FDR correction ($p < 0.05$) were considered significant; in a second step, data were examined at a less conservative, uncorrected threshold ($p < 0.001$, $k \geq 5$). Analyses were conducted within a "song-sensitive mask" to increase signal detection (Friston et al., 1994). This mask was created at the group level using the all > baseline contrast and included only voxels for which passive listening to songs (collapsed across all four conditions) elicited significantly greater activation than baseline (thresholded at $p < 0.001$, $k \geq 5$, FDR corrected at $p < 0.05$, whole brain). The resulting mask (volume: 6895 voxels) spanned an auditory-motor network (see supplemental Fig. 1, Table 2, available at www.jneurosci.org as supplemental material). Peak activations were localized by an experienced neuroanatomist via visual inspection of the averaged high-resolution anatomical scan of all participants. Auditory activations included Heschl's gyrus (HG) and the superior temporal gyrus and sulcus (STG/STS) bilaterally, extending into the pars triangularis of the left inferior frontal gyrus (IFG, Brodmann area (BA) 45) and the left inferior temporal gyrus (ITG, BA 20). Motor acti-

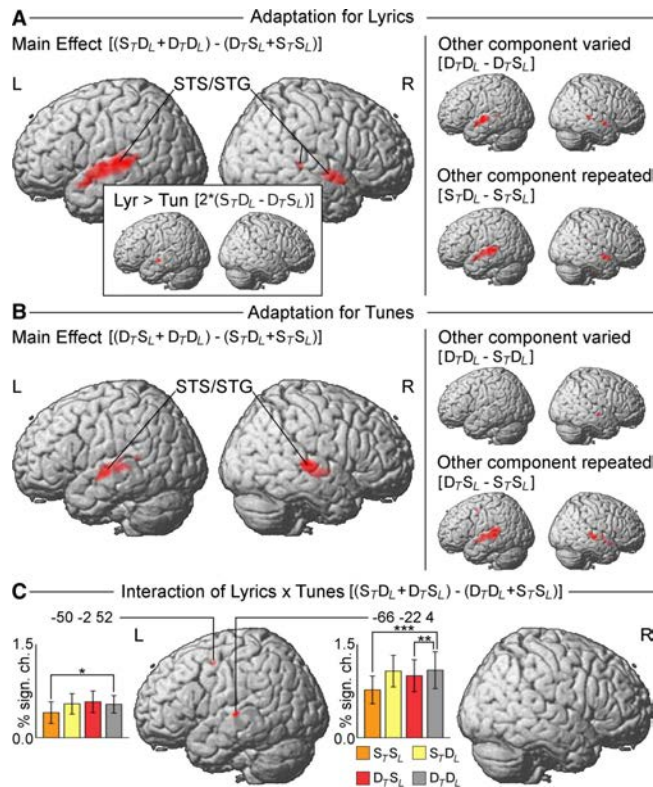


Figure 2. **A**, Adaptation effects for lyric repetition. Left, Main effect of the factorial design $[(S_T D_L + D_T D_L) - (D_T S_L + S_T S_L)]$. Right, Simple effects of lyric repetition when tunes varied $[D_T D_L - D_T S_L]$ (top), or when tunes were simultaneously repeated $[S_T D_L - S_T S_L]$ (bottom; see Results for details). The inset depicts stronger adaptation for the repetition of lyrics than of tunes $[2 \times (S_T D_L - D_T S_L)]$. **B**, Adaptation effects for tune repetition. Left, Main effect of the factorial design $[(D_T S_L + D_T D_L) - (S_T D_L + S_T S_L)]$. Right, Simple effects of tune repetition when lyrics varied $[D_T D_L - S_T D_L]$ (top), or when lyrics were simultaneously repeated $[D_T S_L - S_T S_L]$ (bottom). No brain region showed stronger adaptation for tunes than for lyrics $[2 \times (D_T S_L - S_T D_L)]$ (data not shown). **C**, Interaction of lyrics \times tunes $[(S_T D_L + D_T S_L) - (D_T D_L + S_T S_L)]$. Bar diagrams depict the percentage signal change of the peak voxels in the four conditions relative to baseline. Error bars indicate one SEM. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. For illustration, data are presented at a threshold of $p < 0.001$ (uncorrected, cluster size ≥ 5 voxels).

vations comprised the dorsal precentral gyrus (PrCG) and the cerebellum bilaterally, as well as parts of the right basal ganglia.

Results

Main effects

A main effect of lyric repetition was observed along the STG and STS with larger activations in the left (1147 voxels) than the right hemisphere (258 voxels) (Fig. 2A, left, and Table 1). These regions adapted to the repetition of lyrics even if the tunes varied (507 voxels) (Fig. 2A, top right), although the effect was spatially more than twice as extended if the tunes were simultaneously repeated (1063 voxels) (Fig. 2A, bottom right, and Table 1; for the BOLD percentage signal change of the peak voxels, see supplemental Fig. 2, available at www.jneurosci.org as supplemental material).

A main effect of tune repetition was found in similar areas along the STG/STS bilaterally (left: 362 voxels, right: 448 voxels)

Table 1. Main effects (top rows) and simple effects (middle and bottom rows) of lyric and tune repetition

| Gyrus or region | Adaptation for lyrics | | | | | Adaptation for tunes | | | | |
|---|---|-----|-----|-----|------|---|-----------|------------|-----------|-------------|
| | Size | x | y | z | Z | Size | x | y | z | Z |
| Main effects | $(S_T D_L + D_T D_L) - (D_T S_L + S_T S_L)$ | | | | | $(D_T S_L + D_T D_L) - (S_T D_L + S_T S_L)$ | | | | |
| Left hemisphere | | | | | | | | | | |
| STS/STG | 1147 | −66 | −26 | 4 | 5.14 | 342 | −54 | −16 | 0 | 5.71 |
| | | −50 | −30 | 4 | 4.60 | | −54 | −24 | 2 | 4.56 |
| | | −56 | 8 | −16 | 4.29 | | −66 | −12 | −6 | 4.29 |
| STS/STG | | | | | | 20 | −56 | −42 | 10 | 3.83 |
| Right hemisphere | | | | | | | | | | |
| STS/STG | 232 | 56 | 2 | −8 | 4.49 | 448 | 52 | −26 | 2 | 5.02 |
| | | 62 | −4 | −8 | 4.07 | | 60 | −16 | −4 | 4.42 |
| STS/STG | 26 | 68 | −32 | 0 | 3.52 | | 68 | −20 | 4 | 4.38 |
| Simple effects: variation of the other modality | $D_T D_L - D_T S_L$ | | | | | $D_T D_L - S_T D_L$ | | | | |
| Left hemisphere | | | | | | | | | | |
| STS/STG | 352 | −64 | −18 | −2 | 4.49 | | | | | |
| | | −56 | −10 | −6 | 4.47 | | | | | |
| | | −64 | −10 | 0 | 4.32 | | | | | |
| STS/STG | 38 | −46 | −38 | 4 | 4.00 | | | | | |
| STS/STG | 7 | −60 | −34 | 6 | 3.41 | | | | | |
| Right hemisphere | | | | | | | | | | |
| STS/STG | 36 | 70 | −32 | 0 | 4.57 | 11 | 58 | −14 | −4 | 3.46 |
| STS/STG | 74 | 58 | 0 | −12 | 3.66 | | | | | |
| Simple effects: repetition of the other modality | $S_T D_L - S_T S_L$ | | | | | $D_T S_L - S_T S_L$ | | | | |
| Left hemisphere | | | | | | | | | | |
| STS/STG | 857 | −64 | −30 | 4 | 5.03 | 975 | −58 | −32 | 0 | 5.38 |
| | | −48 | −32 | 4 | 4.83 | | −54 | −24 | 2 | 4.74 |
| | | −60 | −22 | 2 | 4.50 | | −66 | −36 | 14 | 4.44 |
| PrCG (BA 6) | | | | | | 39 | −42 | 0 | 46 | 3.74 |
| | | | | | | | −50 | −2 | 52 | 3.59 |
| Right hemisphere | | | | | | | | | | |
| STS/STG | 201 | 58 | −2 | −8 | 4.05 | 482 | 56 | 6 | −12 | 4.28 |
| STS/STG | 5 | 50 | −34 | 6 | 3.25 | | 60 | −24 | 0 | 4.20 |
| | | | | | | | 64 | −6 | 2 | 4.06 |

Plain type values. Thresholded at $p < 0.001$, cluster size ≥ 5 voxels, significant after FDR correction at $p < 0.05$; bold values, $p < 0.001$, cluster size ≥ 5 voxels, uncorrected. Brain atlas coordinates (MNI stereotactic space) are indicated in millimeters along left–right (x), anterior–posterior (y), and superior–inferior (z) axes.

(Fig. 2*B*, left, and Table 1). In FDR-corrected SPMs, no region adapted to the repetition of tunes if the lyrics varied, although a small effect was found in the right mid-STS at a less conservative, uncorrected threshold of $p < 0.001$, $k \geq 5$ (11 voxels) (Fig. 2*B*, top right). The adaptation effect emerged bilaterally (FDR corrected) if the lyrics were simultaneously repeated (1457 voxels, spatial extent enhanced by a factor of 132) (Fig. 2*B*, bottom right, and Table 1).

The direct comparison of the two main effects yielded no significant difference after FDR correction, probably due to the relatively low number of subjects. However, as several studies claim a dissociated processing of lyrics and tunes (e.g., Besson et al., 1998; Bonnel et al., 2001), possible differences between the adaptation for lyrics and tunes were tested directly at an uncorrected threshold ($p < 0.001$, $k \geq 5$). This test indicated a stronger adaptation effect for lyrics than for tunes in an anterior portion of the left STS ($x = -54$, $y = -12$, $z = -6$, cluster size = 35 voxels, $Z = 3.88$) (Fig. 2*A*, inset; for the BOLD percentage signal change of the peak voxel, see also supplemental Fig. 2, available at www.jneurosci.org as supplemental material). No brain regions showed stronger adaptation for tunes than for lyrics at this threshold.

Interaction

No voxel survived the FDR correction, but as previous studies reported an interaction between the processing of verbal and musical information in song (e.g., Bigand et al., 2001; Lidji et al., 2009; Schön et al., 2005) and guided by our hypothesis (see Introduction), a possible interaction of lyrics and tunes was tested

at a less conservative threshold ($p < 0.001$, $k \geq 5$ voxels, uncorrected). This analysis revealed an interaction of lyrics \times tunes [$(S_T D_L + D_T S_L) - (D_T D_L + S_T S_L)$] in the left mid-STS ($x = -66$, $y = -22$, $z = 4$, cluster extent = 16 voxels, $Z = 3.47$) and the left dorsal PrCG ($x = -50$, $y = -2$, $z = 52$, cluster extent = 5 voxels, $Z = 3.47$) (Fig. 2*C*). This indicates that in these regions, the combined repetition of lyrics and tunes ($S_T S_L$) induced significantly stronger adaptation (compared to $D_T D_L$) than the simple repetition of lyrics ($D_T S_L$) and tunes ($S_T D_L$) summed up, suggestive of an integrated processing of both components. No interactions were found in the right hemisphere and the reverse contrast [$(D_T D_L + S_T S_L) - (S_T D_L + D_T S_L)$]. Both clusters were distant to typical voice areas (Belin et al., 2000; Belin and Zatorre, 2003), indicating that the present effect was (as expected) not grounded on an interaction between lyrical/melodic and voice information. To further explore these effects, percentage signal change values were extracted from the peak voxels of each cluster in each participant using the MarsBaR SPM toolbox (<http://marsbar.sourceforge.net>). These values were subjected to *post hoc* paired-samples *t* tests, evaluating the adaptation effects when only lyrics ($D_T D_L$ vs $D_T S_L$), only tunes ($D_T D_L$ vs $S_T D_L$), or both components ($D_T D_L$ vs $S_T S_L$) were repeated (Fig. 2*C*, bar diagrams). In line with the interaction, the combined repetition of lyrics and tunes induced the strongest adaptation effects in both regions (left mid-STS: $t_{(11)} = 6.53$, $p < 0.001$; left PrCG: $t_{(11)} = 2.92$, $p < 0.015$). The adaptation effect for the simple repetition of lyrics was significant, but considerably weaker in the left mid-STS ($t_{(11)} = 3.43$, $p < 0.007$), and nonsignificant in the left PrCG ($p > 0.314$).

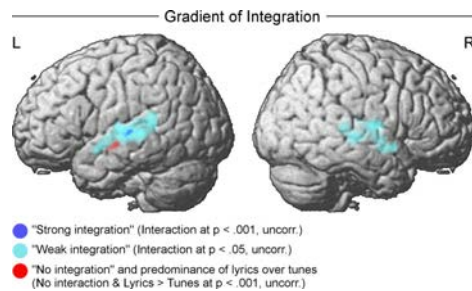


Figure 3. Posterior–anterior gradient of integration.

No cluster showed significant changes during the simple repetition of tunes (p values >0.717).

Gradient

To capture a possible gradient between integration and separation of lyrics and tunes, the interaction (taken as an index for integrated processing) was examined at different statistical thresholds ($p < 0.001$ and $p < 0.05$, uncorrected, $k \geq 5$ voxels). These clusters were compared with the regions that exhibited no interaction but a significantly stronger adaptation for lyrics than for tunes ($p < 0.001$, uncorrected, $k \geq 5$; no regions showed stronger adaptation for tunes than for lyrics; see above), indicating an independent (perhaps separate) processing of lyrics. Figure 3 illustrates that the interaction was confined to the left mid-STS at $p < 0.001$ (blue cluster), suggesting a relatively “strong integration” of both components. The interaction extended more anteriorly and posteriorly, and emerged also in the right STS/STG at a lowered threshold of $p < 0.05$ (cyan cluster), taken as a “weaker” form of integration. Anteroventral to this, the left STS showed no more interaction ($p > 0.05$), but a significantly stronger adaptation effect for the repetition of lyrics compared to tunes (red cluster) (see also Fig. 2A, inset), suggesting “no integration” and a predominance of lyrics over tunes in this region. Altogether, these findings appear to constitute a gradient from more to less integrated processing along the posterior–anterior axis of the left STS (Fig. 3, left).

Discussion

The present study demonstrates that lyrics and tunes of unfamiliar songs are processed at different degrees of integration along the axis of the superior temporal lobe and the left precentral gyrus (PrCG). This is consistent with the idea of a different weighting of integration (and separation) at different stages of the processing of unfamiliar songs.

Main adaptation effects were found along the superior temporal lobe bilaterally. These results are consistent with studies reporting activations of the STG/STS during listening to songs (Schön et al., 2005; Callan et al., 2006), and the processing of various aspects of language (Scott and Johnsrude, 2003; Vigneau et al., 2006) and music (Stewart et al., 2006). This suggests that the observed adaptation effects reflect the facilitated processing of the repeated lyrical and melodic information. Most importantly, the voice sensitivity of the STS (Belin et al., 2000; Belin and Zatorre, 2003) and pitch processing cannot account for the observed adaptation effects because singers’ voices and octave varied in all four conditions (see Materials).

The novel finding is that within these superior temporal regions, specifically in the left hemisphere, lyrics and tunes are

processed at varying degrees of integration, with some indication of an independent processing of lyrics in the left anterior STS. The left mid-STS, inferior to Heschl’s gyrus, showed an interaction of the adaptation effects for lyrics and tunes, indicating that the combined repetition of both components (S_1S_1) induced a significantly stronger response reduction (compared to D_1D_1) than the simple repetition of lyrics (D_1S_1) and tunes (S_1D_1) summed up. This overadditive effect demonstrates an integrated processing of both components within the left mid-STS. The interaction (and thus integration) decayed in regions anterior to this cluster. A more anteroventral portion of the left STS exhibited no more interaction, but a stronger adaptation for lyrics than tunes, suggesting a predominant processing of lyrics in this region (see below for a discussion why no region showed a predominance for tunes). Taking these findings together, the picture of a posterior–anterior gradient emerges along the axis of the left STS, from an integrated processing of lyrics and tunes in the mid-STS to the rather independent processing of lyrics in more anterior temporal regions.

This posterior–anterior gradient is reminiscent of the functional (Binder, 2000; Davis and Johnsrude, 2003; Liebenthal et al., 2005; Scott and Johnsrude, 2003) and temporal (Patterson et al., 2002; Kiebel et al., 2008; Overath et al., 2008) hierarchy of auditory (speech) perception in the superior temporal lobe. These models posit a rostral stream running from primary auditory areas to more lateral and anteroventral areas in the (left) STG and STS, comprising consecutive levels of processing that deal with increasingly abstract representations of the auditory information within growing temporal windows: spectrotemporal features in the millisecond range within the primary auditory cortices, prelexical phonemic information within the surrounding left mid-STG/STS (for an overview, see Obleser and Eisner, 2009), and sentential structure and meaning spanning several hundred milliseconds in more anterior temporal regions (Vandenberghe et al., 2002; Scott and Johnsrude, 2003; Crinion et al., 2006; Spitsyna et al., 2006).

Against this theoretical background, the localization of the lyrics–tunes interaction in the left mid-STS suggests an integration of musical and linguistic aspects of song during an intermediate, phonemic processing stage in this rostral auditory pathway (although the current study neither contains time course information, nor specifically manipulates acoustic, phonemic, or structural–semantic processing). No integration of lyrics and tunes was observed for early nonspecific sound analysis within primary auditory areas, although the present data do not exclude integration at this level. First, these pitch-sensitive regions were most likely blind to the repetition of the songs sung by different voices at different octaves, and second, their temporal integration window was probably too narrow to perceive the repetition of the 2.5 s songs (Kiebel et al., 2008). The localization of the interaction effect in the left mid-STS suggests that lyrics and tunes are particularly integrated at prelexical, phonemic processing levels (Obleser and Eisner, 2009). This observation is consistent with previous behavioral and EEG studies showing an interaction between the processing of melodic/harmonic information and nonsense syllables or vowels (Serafine et al., 1986; Crowder et al., 1990; Bigand et al., 2001; Lidji et al., 2009). Beyond that, the data suggest a separate processing of lyrics at subsequent levels of structural analysis and lexical–semantic representation or access in the left anterior STS. Note that this view would not contradict the ability of music to convey meaning (Koelsch et al., 2004) but propose a predominance and greater autonomy of linguistic (compared to musical) meaning in songs. In sum, it may be sug-

gested as a working hypothesis that the degree of integration of lyrics and tunes decreases as the processing of (unfamiliar) songs proceeds along the rostral auditory stream. Note that although the current study does not address song memory and production (precluding a direct comparison between our data and the majority of the prevailing studies), we speculate that also beyond auditory perceptual processing, the degree of integration/separation depends on the specific cognitive processes targeted by an experimental task (e.g., recognition vs recall or production of familiar vs unfamiliar songs), perhaps accounting for some of the conflicting results.

The profile of adaptation effects argues in favor of bidirectional connections between lyrics and tunes, as the adaptation for one component (lyrics or tunes) was modulated by the simultaneous repetition/variation of the other (Fig. 2*A,B*, right). However, it appears that the strength of these connections differs depending on their direction, in a way that tunes are tightly bound to lyrics, whereas the processing of lyrics exhibits a considerable autonomy [for converging behavioral data, see Serafine et al. (1984), Samson and Zatorre (1991), and Schön et al. (2005)]. Consistent with this notion, the left anterior STS showed stronger adaptation for lyrics than tunes (Fig. 2*A*, inset), whereas no reverse effects (tunes > lyrics) were found. It remains to be specified to what extent this imbalance of lyrics and tunes depends on the settings of the present experiment. Listeners may have paid particular attention to the lyrics (as they convey the message), probably boosting the adaptation effect (Chee and Tan, 2007). Correspondingly, deeper lexical–semantic processing (see above) may account for the more robust adaptation effects for lyrics. Alternatively, the predominance of lyrics might be due to the higher linguistic than musical expertise of the listeners (French native speakers, but musically untrained), consistent with the sensitivity of left STS activations to the expertise of listeners with the employed stimulus material (Leech et al., 2009). Future studies with trained musicians (i.e., balanced linguistic and musical expertise), focused listening to the melodies, and/or the use of nonsense lyrics could address these issues.

As a final footnote, the simple repetition of lyrics induced a bilateral response reduction with left hemisphere preponderance, whereas a small cluster in the right hemisphere tended to adapt to the simple repetition of tunes. This differential hemispheric weighting is consistent with prevailing models of a relative specialization of the left and right hemisphere for linguistic and musical stimulus features respectively, like temporal and spectral (Zatorre et al., 2002) or segmental and suprasegmental information (Friederici and Alter, 2004). Interestingly, lyrics and tunes appeared to be more strongly integrated in the left than in the right hemisphere. This might be due to the predominance of lyrics over tunes in the present study and, thus, a stronger involvement of the left hemisphere.

The interaction of adaptation effects in the left precentral gyrus (BA 6) also suggested an integrated processing of lyrics and tunes. The PrCG is the seat of primary motor and premotor areas, and its involvement in the present experiment may be associated either with (voluntary) internal singing or humming (Hickok et al., 2003; Callan et al., 2006), or with a more general (involuntary) coupling between the auditory and the motor system as proposed by models of auditory–motor integration in language (Scott and Johnsrude, 2003; Hickok and Poeppel, 2007) and music (Warren et al., 2005; Zatorre et al., 2007). These models posit a direct matching between the perception of an auditory signal like a speech sound or a piano tone and a stored (pre)motor code for its production. Along these lines, it may be speculated that the ad-

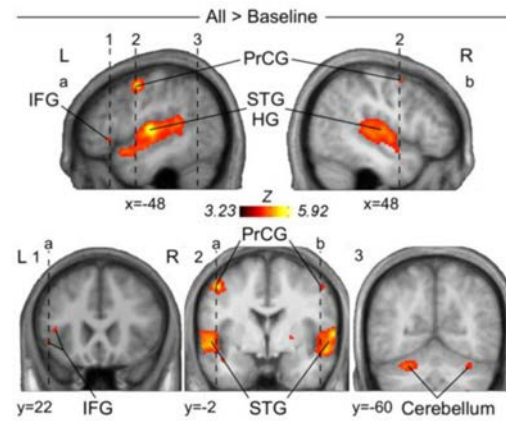
aptation of the neural activity in the left PrCG reflects either the increasing efficiency of subvocal rehearsal, i.e., vocal learning (Rauschecker et al., 2008), or the facilitated mirroring of articulatory gestures during passive listening. It appears that lyrical and melodic features must be integrated in a vocal code for singing as they are simultaneously articulated via the vocal tract.

To conclude, the present study is the first demonstration that lyrics and tunes of songs are processed at different degrees of integration (and separation) through the consecutive processing levels allocated along the posterior–anterior axis of the left superior temporal lobe and the left PrCG. While both components seem to be integrated at a prelexical, phonemic stage of the auditory analysis in the left mid-STS, and the preparation of a motor output in the left PrCG, lyrics may be processed independently at levels of structural and semantic integration in the left anterior STS. Overall, the findings demonstrate an anatomical and functional gradient of integration of lyrics and tunes during passive listening to unfamiliar songs.

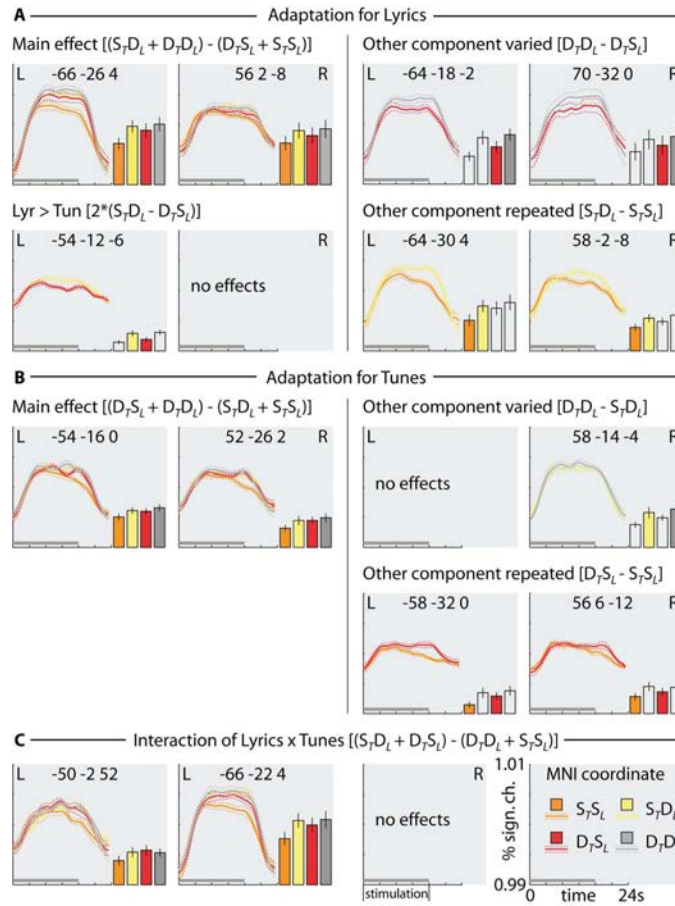
References

- Belin P, Zatorre RJ (2003) Adaptation to speaker's voice in right anterior temporal lobe. *Neuroreport* 14:2105–2109.
- Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B (2000) Voice-selective areas in human auditory cortex. *Nature* 403:309–312.
- Besson M, Ffytche F, Peretz I, Bonnel AM, Requin J (1998) Singing in the brain: independence of lyrics and tunes. *Psychol Sci* 9:494–498.
- Bigand E, Tillmann B, Poulin B, D'Adamo DA, Madurell F (2001) The effect of harmonic context on phoneme monitoring in vocal music. *Cognition* 81:B11–B20.
- Binder J (2000) The new neuroanatomy of speech perception. *Brain* 123:2371–2372.
- Bonnel AM, Ffytche F, Peretz I, Besson M (2001) Divided attention between lyrics and tunes of operatic songs: evidence for independent processing. *Percept Psychophys* 63:1201–1213.
- Callan DE, Tsytasarev V, Hanakawa T, Callan AM, Katsuhara M, Fukuyama H, Turner R (2006) Song and speech: brain regions involved with perception and covert production. *Neuroimage* 31:1327–1342.
- Chee MWL, Tan JC (2007) Inter-relationships between attention, activation, fMRI adaptation and long-term memory. *Neuroimage* 37:1487–1495.
- Crinion JT, Warburton EA, Lambon-Ralph MA, Howard D, Wise RJS (2006) Listening to narrative speech after aphasic stroke: the role of the left anterior temporal lobe. *Cereb Cortex* 16:1116–1125.
- Crowder RG, Serafine ML, Repp B (1990) Physical interaction and association by contiguity in memory for the words and melodies of songs. *Mem Cognit* 18:469–476.
- Davis MH, Johnsrude IS (2003) Hierarchical processing in spoken language comprehension. *J Neurosci* 23:3423–3431.
- Dehaene-Lambertz G, Dehaene S, Anton JL, Campagne A, Ciuciu P, Dehaene GP, Denghien I, Jobert A, LeBihan D, Sigman M, Pallier C, Poline JB (2006) Functional segregation of cortical language areas by sentence repetition. *Hum Brain Mapp* 27:360–371.
- Friederici AD, Alter K (2004) Lateralization of auditory language functions: A dynamic dual pathway model. *Brain Lang* 89:267–276.
- Friston KJ, Worsley RSJ, Frackowiak JC, Mazziotta JC, Evans AC (1994) Assessing the significance of focal activations using their spatial extent. *Hum Brain Mapp* 1:214–220.
- Grill-Spector K (2006) Selectivity of adaptation in single units: implications for fMRI experiments. *Neuron* 49:170–171.
- Hébert S, Peretz I (2001) Are text and tune of familiar songs separable by brain damage? *Brain Cogn* 46:169–175.
- Hébert S, Racette A, Gagnon L, Peretz I (2003) Revisiting the dissociation between singing and speaking in expressive aphasia. *Brain* 126:1838–1850.
- Henson RNA (2003) Neuroimaging studies of priming. *Prog Neurobiol* 70:53–81.
- Hickok G, Poeppel D (2007) The cortical organization of speech processing. *Nat Rev Neurosci* 8:393–402.
- Hickok G, Buchsbaum B, Humphries C, Muftuler T (2003) Auditory-

- motor interaction revealed by fMRI: speech, music, and working memory in area Spt. *J Cogn Neurosci* 15:673–682.
- Kiebel SJ, Daunizeau J, Friston KJ (2008) A hierarchy of time-scales and the brain. *PLoS Comput Biol* 4:e1000209.
- Koelsch S, Kasper E, Sammler D, Schulze K, Gunter T, Friederici AD (2004) Music, language and meaning: brain signatures of semantic processing. *Nat Neurosci* 7:302–307.
- Krekelberg B, Boynton GM, van Wezel RJA (2006) Adaptation: from single cells to BOLD signals. *Trends Neurosci* 29:250–256.
- Leech R, Holt LL, Devlin JT, Dick F (2009) Expertise with artificial non-speech sounds recruits speech-sensitive cortical regions. *J Neurosci* 29:5234–5239.
- Lidji P, Jolicoeur P, Moreau P, Kolinsky R, Peretz I (2009) Integrated pre-attentive processing of vowel and pitch: a mismatch negativity study. *Ann N Y Acad Sci* 1169:481–484.
- Liebenthal E, Binder JR, Spitzer SM, Possing ET, Medler DA (2005) Neural substrates of phonemic perception. *Cereb Cortex* 15:1621–1631.
- Naccache L, Dehaene S (2001) The priming method: imaging unconscious repetition priming reveals an abstract representation of number in the parietal lobes. *Cereb Cortex* 11:966–974.
- New B, Pallier C, Brysbaert M, Ferrand L (2004) Lexique 2: A new French lexical database. *Beh Res Meth Instr Comp* 36:516–524.
- Noppeney U, Price CJ (2004) An fMRI study of syntactic adaptation. *J Cogn Neurosci* 16:702–713.
- Obleser J, Eisner F (2009) Pre-lexical abstraction of speech in the auditory cortex. *Trends Cogn Sci* 13:14–19.
- Oldfield RC (1971) The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9:97–113.
- Overath T, Kumar S, von Kriegstein K, Griffiths TD (2008) Encoding of spectral correlation over time in auditory cortex. *J Neurosci* 28:13268–13273.
- Patterson RD, Uppenkamp S, Johnsrude IS, Griffiths TD (2002) The processing of temporal pitch and melody information in auditory cortex. *Neuron* 36:767–776.
- Peretz I (1996) Can we lose memories for music? A case of music agnosia in a nonmusician. *J Cogn Neurosci* 8:481–496.
- Poulin-Charronnat B, Bigand E, Madurell F, Peereman R (2005) Musical structure modulates semantic priming in vocal music. *Cognition* 94:B67–B78.
- Racette A, Bard C, Peretz I (2006) Making non-fluent aphasics speak: sing along! *Brain* 129:2571–2584.
- Rauschecker AM, Pringle A, Watkins KE (2008) Changes in neural activity associated with learning to articulate novel auditory pseudowords by covert repetition. *Hum Brain Mapp* 29:1231–1242.
- Robine M (1994) *Anthologie de la chanson française—des trouvères aux grands auteurs du XIXe siècle*. Paris: Albin Michel.
- Samson S, Zatorre RJ (1991) Recognition memory for text and melody of songs after unilateral temporal lobe lesion: evidence for dual encoding. *J Exp Psychol Learn Mem Cogn* 17:793–804.
- Schön D, Gordon RL, Besson M (2005) Musical and linguistic processing in song perception. *Ann N Y Acad Sci* 1060:71–81.
- Scott SK, Johnsrude IS (2003) The neuroanatomical and functional organization of speech perception. *Trends Neurosci* 26:100–107.
- Serafine ML, Crowder RG, Repp BH (1984) Integration of melody and text in memory for songs. *Cognition* 16:285–303.
- Serafine ML, Davidson J, Crowder RG, Repp B (1986) On the nature of melody-text integration in memory for songs. *J Mem Lang* 25:123–135.
- Spitsyna G, Warren JE, Scott SK, Turkheimer FE, Wise RJS (2006) Converging language streams in the human temporal lobe. *J Neurosci* 26:7328–7336.
- Stewart L, von Kriegstein K, Warren JD, Griffiths TD (2006) Music and the brain: disorders of musical listening. *Brain* 129:2533–2553.
- Vandenberghe R, Nobre AC, Price CJ (2002) The response of left temporal cortex to sentences. *J Cogn Neurosci* 14:550–560.
- Vigneau M, Beaucousin V, Hervé PY, Duffau H, Crivello F, Houdé O, Mazoyer B, Tzourio-Mazoyer N (2006) Metaanalyzing left hemisphere language areas: phonology, semantics, and sentence processing. *Neuroimage* 30:1414–1432.
- Warren JE, Wise RJS, Warren JD (2005) Sounds do-able: auditory-motor transformations and the posterior temporal plane. *Trends Neurosci* 28:636–643.
- Zatorre RJ, Belin P, Penhune VB (2002) Structure and function of auditory cortex: music and speech. *Trends Cogn Sci* 6:37–46.
- Zatorre RJ, Chen JL, Penhune VB (2007) When the brain plays music: auditory-motor interactions in music perception and production. *Nat Rev Neurosci* 8:547–558.

Supplementary Material – Sammler, Baird et al. (2010)

Supplemental Figure 1. Activations elicited during passive listening to songs (collapsed across all four conditions and contrasted against baseline) overlaid on the participants' average anatomical MR image (thresholded at $p < .001$, $k \geq 5$ voxels, FDR-corrected with $p < .05$, whole brain). PrCG: precentral gyrus, STG: superior temporal gyrus, HG: Heschl's gyrus, IFG: inferiorfrontal gyrus.



Supplemental Figure 2. Time course and average of the BOLD % signal change in the peak voxels of all effects depicted in Figure 2 in the article. Values were obtained using the MarsBaR SPM toolbox (<http://marsbar.sourceforge.net>). **(A)** Adaptation for Lyrics repetition. Left: main effect of the factorial design $[(S_T D_L + D_T D_L) - (D_T S_L + S_T S_L)]$ (top), and the stronger adaptation for the repetition of lyrics than of tunes $[2 * (S_T D_L - D_T S_L)]$ (bottom). Right: simple effects of lyrics repetition when tunes varied $[D_T D_L - D_T S_L]$ (top), or when tunes were simultaneously repeated $[S_T D_L - S_T S_L]$ (bottom). **(B)** Adaptation effects for Tunes repetition. Left: main effect of the factorial design $[(D_T S_L + D_T D_L) - (S_T D_L + S_T S_L)]$. Right: simple effects of tunes repetition when lyrics varied $[D_T D_L - S_T D_L]$ (top), or when lyrics were simultaneously repeated $[D_T S_L - S_T S_L]$ (bottom). **(C)** Interaction of Lyrics x Tunes $[(S_T D_L + D_T S_L) - (D_T D_L + S_T S_L)]$. Solid lines depict mean values, dotted lines and error bars correspond to one SEM. Uncolored bars in the bar diagrams are irrelevant for the respective contrast. The gray line beneath the x-axis indicates the time of stimulus presentation (i.e., one block of six songs).

Supplemental Table 1:

| Condition | S _T S _L | S _T D _L | D _T S _L | D _T D _L | <i>F</i> (3,164) | <i>p</i> |
|------------------|-------------------------------|-------------------------------|-------------------------------|-------------------------------|------------------|----------|
| Note number | 7.50 | 7.33 | 7.50 | 7.88 | 1.49 | .220 |
| Note length | 0.33 | 0.34 | 0.33 | 0.31 | 1.35 | .260 |
| Interval size | 1.64 | 1.84 | 1.86 | 1.79 | 0.53 | .659 |
| Contour reversal | 2.08 | 2.00 | 2.21 | 2.21 | 0.15 | .931 |
| Meter | ½ / ½ | ½ / ½ | ½ / ½ | ½ / ½ | - | - |
| Mode | ½ / ½ | ½ / ½ | ½ / ½ | ½ / ½ | - | - |
| Word number | 5.25 | 5.44 | 5.75 | 5.81 | 1.14 | .333 |
| Word length | 1.31 | 1.38 | 1.44 | 1.46 | 1.04 | .376 |
| Word frequency | 4723.59 | 4533.27 | 4701.61 | 4760.71 | 0.08 | .971 |

Supplemental Table 1. Stimulus properties that were balanced between conditions. Note length: seconds, interval size: semitones, number of contour reversals, number of pieces in duple/triple time, number of pieces in major/minor mode, word length: syllables, word frequency according to LEXIQUE 2 (New et al., 2004).

Supplemental Table 2:

| Gyrus or region | All > Baseline | | | | |
|-------------------------|----------------|-----|-----|-----|------|
| | size | x | y | z | Z |
| <i>Left hemisphere</i> | | | | | |
| HG (BA 41/42) / | 3391 | -50 | -14 | 4 | 5.87 |
| STG (BA 22) | | -56 | -20 | 6 | 5.60 |
| | | -42 | -28 | 10 | 5.15 |
| IFG, p. triang. (BA 45) | 18 | -42 | 24 | 8 | 4.09 |
| IFG, p. triang. (BA 45) | 13 | -46 | 24 | -2 | 3.33 |
| ITG (BA 20) | 12 | -60 | -44 | -14 | 3.84 |
| PrCG (BA 6) | 187 | -48 | -2 | 48 | 5.57 |
| <i>Right hemisphere</i> | | | | | |
| HG (BA 41/42) / | 2873 | 56 | -12 | -4 | 5.92 |
| STG (BA 22) | | 62 | -4 | 4 | 5.47 |
| | | 68 | -18 | 6 | 5.19 |
| STS/STG | 13 | 42 | 6 | -20 | 3.41 |
| PrCG (BA 6) | 14 | 50 | 2 | 52 | 3.40 |
| Caudate nucleus | 7 | 18 | 6 | 12 | 3.52 |
| Pallidum | 7 | 22 | -2 | 0 | 3.23 |
| <i>Left cerebellum</i> | | | | | |
| VI | 158 | -28 | -60 | -26 | 4.17 |
| VIIIB | 5 | -16 | -76 | -42 | 3.38 |
| <i>Right cerebellum</i> | | | | | |
| Crus 2 | 67 | 22 | -80 | -38 | 4.32 |
| VIIIA | 88 | 20 | -64 | -48 | 4.14 |
| | | 30 | -64 | -50 | 3.97 |
| VI | 18 | 18 | -68 | -28 | 3.66 |
| VI | 24 | 32 | -60 | -26 | 3.40 |

Supplemental Table 2. Activations during song perception (collapsed across all four conditions) compared to baseline. SPMs are thresholded at $p < .001$, $k \geq 5$ voxels and are significant after FDR-correction at $p < .05$. BA: Brodmann area, HG: Heschl's gyrus, IFG: inferior frontal gyrus, p. triang.: pars triangularis, ITG: inferior temporal gyrus, PrCG: precentral gyrus, STG: superior temporal gyrus, STS: superior temporal sulcus. Cerebellar localizations are based on Schmahmann et al. (1999).

3.3 Study VI – Song melody and lyrics

Alonso, I., **Sammler, D.**, Valabrègue, R., Dinkelacker, V., Dupont, S., Belin, P., Samson, S. (2014). Hippocampal sclerosis affects fMR-adaptation to lyrics and melodies in songs. *Frontiers in Human Neuroscience*, 8, Article 111.

[Author contributions by D.S.: designed research with P.B., S.S.; performed research with I.A., received materials from V.D. and S.D., analysed data with I.A., R.V.; co-wrote the paper with I.A., incorporating edits by co-authors]



Hippocampal sclerosis affects fMR-adaptation of lyrics and melodies in songs

Irene Alonso^{1,2,3,4}, Daniela Sammler⁵, Romain Valabrègue^{3,4}, Vera Dinkelacker^{2,4}, Sophie Dupont^{2,4}, Pascal Belin^{6,7,8} and Séverine Samson^{1,2*}

¹ Laboratoire de Neurosciences Fonctionnelles et Pathologies (EA 4559), Université Lille-Nord de France, Lille, France

² Epilepsy Unit, Hôpital de la Pitié-Salpêtrière, Paris, France

³ Centre de Neuroimagerie de Recherche, Groupe Hospitalier Pitié-Salpêtrière, Paris, France

⁴ Centre de Recherche de l'Institut du Cerveau et de la Moëlle Épinrière, UPMC – UMR 7225 CNRS – UMRS 975 INSERM, Paris, France

⁵ Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

⁶ Centre for Cognitive Neuroimaging, Department of Psychology, University of Glasgow, Glasgow, UK

⁷ Laboratories for Brain, Music and Sound, Université de Montréal and McGill University, Montreal, QC, Canada

⁸ Institut des Neurosciences de la Timone, UMR7289, CNRS-Université Aix Marseille, Marseille, France

Edited by:

Eckart Altenmüller, University of Music and Drama Hannover, Germany

Reviewed by:

Stephan Schuele, Northwestern University, USA

Kathrin Wagner, University Hospital Freiburg, Germany

*Correspondence:

Séverine Samson, Department of Psychology, University of Lille 3, BP 60 149, 59653 Villeneuve d'Ascq Cedex, France
e-mail: severine.samson@univ-lille3.fr

Songs constitute a natural combination of lyrics and melodies, but it is unclear whether and how these two song components are integrated during the emergence of a memory trace. Network theories of memory suggest a prominent role of the hippocampus, together with unimodal sensory areas, in the build-up of conjunctive representations. The present study tested the modulatory influence of the hippocampus on neural adaptation to songs in lateral temporal areas. Patients with unilateral hippocampal sclerosis and healthy matched controls were presented with blocks of short songs in which lyrics and/or melodies were varied or repeated in a crossed factorial design. Neural adaptation effects were taken as correlates of incidental emergent memory traces. We hypothesized that hippocampal lesions, particularly in the left hemisphere, would weaken adaptation effects, especially the integration of lyrics and melodies. Results revealed that lateral temporal lobe regions showed weaker adaptation to repeated lyrics as well as a reduced interaction of the adaptation effects for lyrics and melodies in patients with left hippocampal sclerosis. This suggests a deficient build-up of a sensory memory trace for lyrics and a reduced integration of lyrics with melodies, compared to healthy controls. Patients with right hippocampal sclerosis showed a similar profile of results although the effects did not reach significance in this population. We highlight the finding that the integrated representation of lyrics and melodies typically shown in healthy participants is likely tied to the integrity of the left medial temporal lobe. This novel finding provides the first neuroimaging evidence for the role of the hippocampus during repetitive exposure to lyrics and melodies and their integration into a song.

Keywords: neural adaptation, song, lyrics, hippocampal sclerosis, memory trace, conjunctive representation

INTRODUCTION

As humans, we learn and enjoy songs from a very early age on. Over the course of our lives, we hear and remember thousands of songs and, most of the time, we learn them implicitly without much effort especially after repeated presentations (as with hit songs on the radio). Songs naturally combine music and language into a unique acoustic signal. However, it remains unclear whether memory traces of lyrics and melodies are built separately or in integration. Indeed, evidence from healthy participants and brain-damaged patients diverge on this question. On the one hand, several behavioral studies in healthy participants support the tight association of lyrics and melodies during the creation of a song memory trace as shown by cueing effects of one element on the other during song recognition (Serafine et al., 1984, 1986; Crowder et al., 1990; Baur et al., 2000; Peretz et al., 2004; Peynircioglu et al., 2008; Johnson and Halpern, 2012). On the other hand, neuropsychological studies in patients with lesions in the medial or lateral temporal lobes reveal dissociated recognition impairments

for verbal and musical features of songs (Samson and Zatorre, 1991; Hébert and Peretz, 2001). These results suggest that the natural binding of lyrics and melodies into one unique song memory trace may be disrupted after brain damage. The present study seeks to find neural evidence for this hypothesis by investigating the effect of hippocampal damage on the emergence of integrated memory traces for lyrics and melodies during repeated exposure to songs.

Research over the last two decades testifies to a growing awareness that the hippocampus – beyond its classical role in explicit episodic memory (Scoville and Milner, 1957; Mishkin, 1982; Zola-Morgan and Squire, 1993) – plays a role in the implicit build-up of a memory trace (Chun and Phelps, 1999; Graham et al., 2010) and the bridging between perception and encoding (Bussey and Saksida, 2005; Baxter, 2009; Suzuki, 2009; Suzuki and Baxter, 2009; Olsen et al., 2012). According to the Emergent Memory Account (Graham et al., 2010) advancing a non-modular view of memory and perception, memory arises from a dynamic interaction

between the perceptual representations distributed across the whole brain and a key role of the medial temporal lobe. More specifically, the hippocampus is thought to form conjunctive representations of inputs from unimodal and polymodal sensory cortices and to continuously return the processed information to the sensory cortex via feedback connections (McClelland et al., 1995; Eichenbaum, 2000; Turk-Browne et al., 2006; Bast, 2007), thus constantly updating the current representations with new experiences. This cortico-hippocampal loop of flowing information guarantees the encoding of events and its storage (Eichenbaum, 2000). Note that this mechanism not only implies a shared, anatomically distributed cerebral network for both memory and perception, but also puts the medial temporal lobe into a cardinal position between perceptual processes (Lee et al., 2005; Lee, 2006; Lee and Rudebeck, 2010a) and memory (long-term as well as short-term and working memory: Zarahn, 2004; Axmacher et al., 2007; Lee and Rudebeck, 2010b; Rose et al., 2012). Crucially, the hippocampus' combined role in (i) memory formation and (ii) conjunction of sensory inputs (Sutherland and Rudy, 1989; Eichenbaum et al., 1994; Rudy and Sutherland, 1995; O'Reilly and Rudy, 2001; Winters, 2004; Cowell et al., 2006, 2010; Barense et al., 2007; Diana et al., 2007) makes it a potential key candidate for (i) the build-up of song memory traces, in which (ii) lyrics and melodies are integrated.

Although most of the studies on the hippocampus' role in memory formation and binding come from the visual domain (Davachi, 2006; Diana et al., 2007; Shimamura, 2010), we hypothesize that similar processes also apply to the auditory domain (Overath et al., 2007, 2008; Buchsbaum and D'Esposito, 2009), especially to songs. It is reasonable to assume that memory formation for lyrics and melodies happens through a cortico-hippocampal loop, and that the natural combination of a verbal and a melodic component into a single song percept and memory trace requires binding mechanisms as described above. Tentative support for this comes from lesion studies in patients with anterior temporal lobectomy for treatment of pharmaco-resistant epilepsy (Samson and Zatorre, 1991). Using explicit recognition memory tasks after presentation of short unfamiliar songs, these experiments revealed a clear deficit in recognition of sung and spoken lyrics after left temporal lobe resection, and impaired recognition of melodies (without text) after right temporal lobe resection. On top of that, the data suggest a lack of integration of lyrics and melodies in patients with unilateral left (but not those with right) temporal lobe lesions. While patients with right temporal lobe resections had deficits in melody recognition when the tune was sung with new words, i.e., showing that they had bound the melody to the original lyrics, no such conjunction was observed in left-hemisphere damaged patients. In fact, their recognition of lyrics was impaired irrespective of whether these were presented with (or without) old or new melodies, suggesting an independent processing of the two song components and an isolated deficit for lyrics.

While these results lend initial support for our hypothesis of hippocampal involvement in song memory formation, they leave two important questions open: first, in how far can these deficit patterns be attributed to hippocampal dysfunctions, and second, in how far may these results depend on the use of a recognition memory task? First, the resection always included anterior temporal

lobe structures beyond the hippocampus, making it difficult to pinpoint a specific hippocampal role. Furthermore, although the lesion description was based upon the surgeon's meticulous drawings, a precise assessment of how far the resection extended into the hippocampus was not possible at that time. Moreover, although recognition tasks certainly depend on successful encoding, they also involve aspects of memory retrieval making it difficult to disentangle these effects with behavioral data. The present study seeks to address the points by first, testing patients with circumscribed unilateral hippocampal sclerosis (i.e., prior to surgery without further macroscopic lesions) and precisely describing the extent of hippocampal damage by means of volumetric analyses. Second, the incidental build-up of a song memory trace was assessed unbeknownst to the participants by examining the dynamics of neural adaptation during natural passive listening as described below.

Numerous studies have investigated the neural correlates of song processing (Samson and Zatorre, 1991; Brown et al., 2004a,b; Schön et al., 2005; Callan et al., 2006; Suarez et al., 2010; Merrill et al., 2012; Saito et al., 2012; Tierney et al., 2012), however, rarely has any study touched upon the implicit emergence of song memory. Indirect evidence can be drawn from studies using the successive presentation of changed and unchanged song stimuli (Same vs. Different) (Schön et al., 2010) and neural adaptation paradigms (Sammler et al., 2010). Adaptation is "a reduction of neural activity following prolonged or repetitive exposure to identical or at least similar stimuli" (Dobbins et al., 2004; Ganel et al., 2006; Grill-Spector et al., 2006), similar to repetition priming (Old vs. New stimuli) (Krekelberg et al., 2006). Although typically described in studies on perception, it appears that neural adaptation may also be indicative of memory trace formation. In line with the Emergent Memory Account (Graham et al., 2010), neural adaptation may reflect the emergence of a memory trace within cortical areas of perceptual representation through implicit learning during repeated exposure. Given the role of the hippocampus in memory formation (Turk-Browne et al., 2006) and according to connectionist models of memory (Damasio, 1989; McClelland et al., 1995; Rolls, 1996; Fuster, 1997), it is reasonable to suggest that cortical adaptation effects are subject to top-down modulations driven by the hippocampus (Blondin and Lepage, 2005; Goh et al., 2007), including integration of lyrics and melodies through binding (for a review on binding, see Opitz, 2010).

Of particular relevance for our research question of how lyrics and melody are bound in a conjunctive song memory trace are those studies describing the cerebral substrates underlying the integration of verbal and melodic components of songs (Sammler et al., 2010; Schön et al., 2010). These studies, which consider songs to be more than the sum of lyrics and melodies, examined modulations of brain activity to investigate how the two components interact, and how their processing is lateralized. For instance, Schön et al. (2010, Exp. 2) presented pairs of sung words that could vary or repeat in terms of the verbal and/or the melody component in a same-different task. Their results showed interactive processing in the left and the right superior temporal gyrus (STG), suggesting an integrated processing of the two components in these areas. Sammler et al. (2010) adopted a similar approach, taking advantage of neural adaptation effects. In this study, healthy participants were presented with blocks of short songs in which

Table 1 | Demographic data.

| Group | N | Sex (males/females) | Mean age | Mean education | Musical expertise | Full scale IQ WAIS-R ^a | RAVLT forgetting% ^a |
|---------|----|---------------------|--------------|----------------|-------------------|-----------------------------------|--------------------------------|
| Control | 19 | 9/10 | 32.63 ± 7.90 | 14.53 ± 2.99 | 5.02 ± 3.68 | – | – |
| LTLE | 12 | 5/7 | 34.17 ± 8.71 | 12.25 ± 1.42 | 6.75 ± 4.46 | 92.66 ± 1.33 | 20.56 ± 22.99 |
| RTLE | 12 | 5/7 | 39.92 ± 1.23 | 12.67 ± 2.71 | 6.83 ± 5.45 | 95.43 ± 1.79 | 9.07 ± 10.76 |

^aMean for all except two RTLE patients due two missing data. RAVLT: Rey Auditory Verbal Learning Test.

repetition of lyrics and/or melodies was varied in a factorial design to induce selective adaptation to lyrics, melodies, or unified songs. Consistent with Schön et al. (2010), repeated lyrics or repeated tunes evoked adaptation effects in bilateral STG. Core areas of integration were found in the left middle superior temporal sulcus (STS) and the left premotor cortex (PMC). Based on the previously reported literature, we hypothesize that these adaptation effects and the integration of lyrics and melodies are likely mediated by the hippocampus through feedback connections to STG/STS and binding of verbal and melodic information.

To investigate the modulatory effect of the hippocampus on (i) the incidental emergence of a song memory trace and (ii) the integration of the verbal and melodic components of songs, we adopted the paradigm by Sammler et al. (2010) to test patients with unilateral left or right hippocampal sclerosis and healthy controls. We compared the patterns of adaptation produced by songs in which either the lyrics, or the melodies, or both were repeated. As demonstrated by diffusion-weighted imaging studies, patients with hippocampal sclerosis present disconnections between medial and lateral temporal lobe regions (Focke et al., 2008; Bettus et al., 2009; Diehl et al., 2010; Riley et al., 2010; Liao et al., 2011). Such lesions have the capacity to prevent the hippocampus from sending feedback predictions and from updating the sensory memory trace (as expected by default after repetitions) and thus weaken adaptation effects in general and integration of lyrics and melodies in particular. More precisely, following Samson and Zatorre (1991), we hypothesized reduced adaptation for lyrics after left and for melodies after right hippocampal sclerosis. Moreover, following previous studies showing binding deficits in patients with left anterior temporal lobe resections (Samson and Zatorre, 1991) and correlates of lyrics–melody integration mainly in the left hemisphere (Sammler et al., 2010), we hypothesized that left hippocampal lesions, in particular, would have a negative impact on integration of lyrics and melodies in songs.

MATERIALS AND METHODS

PARTICIPANTS

Twenty-four temporal lobe epilepsy patients with left ($n = 12$; LTLE) or right ($n = 12$; RTLE) hippocampal sclerosis participated in this study. They all presented with medically intractable epilepsy and were seen during pre-surgical evaluation at Pitié-Salpêtrière Hospital (Paris, France). All patients were right-handed according to the Edinburgh Handedness Inventory (Oldfield, 1971), except for one LTLE (−83.33) and one RTLE patient (−75). All patients had language lateralization to the left hemisphere except for the left-handed RTLE patient with bilateral language representation. Language lateralization was assessed by means of a verbal fluency test that is part of the standard functional magnetic resonance

imaging (fMRI) assessment prior to epilepsy surgery at the Pitié-Salpêtrière Hospital. In the scanner, patients are required to think as many words of a semantic category (e.g., tools) as possible. The number of activated left and right fronto-temporo-parietal voxels against baseline was used to calculate a standard language lateralization score (Lehéricy et al., 2000; Thivard et al., 2005). The control group consisted of 19 right-handed healthy participants including 12 subjects, who had already participated in a previous study (Sammler et al., 2010), and 7 new volunteers. All participants were French native speakers and reported to have normal hearing. Controls were carefully selected to match the patient groups in terms of age, mean years of education, and musical expertise (Ehrlé musical expertise questionnaire, unpublished). A verbal memory deficit was present in the LTLE as opposed to the RTLE patients, as assessed with the Rey Auditory Verbal Learning Test (RAVLT) (Rey, 1964; Sziklas and Jones-Gotman, 2008) in accordance with the usual neuropsychological profile of these patients. Demographic characteristics of the participants are summarized in Table 1. The sclerosis in either left or right hippocampus in the two patient groups was corroborated by a volumetric analysis using Freesurfer software (Fischl, 2012; Reuter et al., 2012) that attested an ipsilateral hippocampal volume reduction of an average of 24.51% in the LTLE and 29.71% in the RTLE group compared to healthy controls. Between-group comparisons confirmed the significance of these volume reductions in the atrophic hippocampus ($p < 0.05$). Volumes and percentage of reduction are summarized in Table 2 (for details on the volumetric analysis, see Data Analysis). The local ethics committee approved this study and informed consent was obtained from each participant.

MATERIALS

The material and the scanning protocol used here were previously published by Sammler et al. (2010). The stimulus set consisted of 48 blocks of 6 unfamiliar songs based on a collection of nineteenth century French folk songs (Robine, 1994). Each song within a block was sung by a different singer to avoid adaptation to the singer's voice (Belin and Zatorre, 2003), had a duration of 2.5 s and was followed by a 0.2 s pause. Repetition of lyrics and/or melodies within blocks was crossed in a 2×2 factorial design, forming four conditions. Songs within a block either had the same melodies and same lyrics ($S_M S_L$), the same melodies but different lyrics ($S_M D_L$), different melodies with same lyrics ($D_M S_L$), or different melodies and different lyrics ($D_M D_L$). Mode and tempo were balanced across the stimulus set, and each song had an average of 7.65 notes and 5.61 words. Songs in the four conditions did not differ with respect to length and number of word/note, word frequency, interval size, and number of contour reversals. In blocks where lyrics were varied, they did not rhyme, were semantically distant,

Table 2 | Medial temporal lobe (MTL) volumes (mm³).

| Region | Left MTL | | | | Right MTL | | | |
|-----------------------|----------------|----------------------------|----------------|----------------------------|----------------|----------------------------|----------------|----------------------------|
| | LTLE | | RTLE | | LTLE | | RTLE | |
| | Mean/ ± SD | Reduction (%) ^a | Mean/ ± SD | Reduction (%) ^a | Mean/ ± SD | Reduction (%) ^a | Mean/ ± SD | Reduction (%) ^a |
| Hippocampus | 2606.17/506.82 | 24.51 | 3467.95/245.68 | −0.46 | 3589.04/678.97 | −2.72 | 2455.90/414.39 | 29.71 |
| Entorhinal cortex | 1802.67/613.03 | 3.86 | 1847.25/218.17 | 1.48 | 1901.42/408.28 | 3.73 | 1802/246.24 | 8.76 |
| Parahippocampal gyrus | 2165.42/386.72 | 8.56 | 2304.17/222.16 | 2.70 | 2249.92/266.37 | 3.26 | 2026/268.47 | 12.89 |

^aPercentage of reduced volumes as compared to control group volumes.

and differed with respect to syntactic structure avoiding potential adaptation to phonology, semantic content, or syntactic structure (Noppeney and Price, 2004).

PROCEDURE

Participants were instructed to listen attentively with closed eyes while avoiding moving, humming, or singing along. No behavioral data were collected. Stimuli were presented using E-Prime 1.1 (Psychology Software Tools) and delivered binaurally through air pressure headphones (MR confon). Additionally, participants used earplugs to minimize noise interference. All blocks were presented in one of four pseudorandom orders, with a silent gap between blocks of 10 s (± 0.5 s) allowing the hemodynamic response to return to baseline (Belin and Zatorre, 2003). This resulted in a total duration of the experiment of around 30 min. Blocks of the same condition were not presented more than twice in a row. At the end of the experiment, all participants filled in a debriefing questionnaire with several nine-point scales (1 = not at all, 9 = always) in which they rated their attention during listening at 7.63 (Controls), 7.00 (LTLE), 7.57 (RTLE), and the amount of overt and/or covert singing during scanning at 0.00 and 2.89 (Controls), 0.47 and 2.71 (LTLE), and 0.21 and 2.14 (RTLE), showing that they had followed the instructions.

SCANNING

Functional magnetic resonance imaging was performed using a 3-T Siemens TRIO scanner (Siemens, Erlangen, Germany) at the *Centre de Neuroimagerie de Recherche* at the *Institut du Cerveau et de la Moëlle Épinière – ICM* (Groupe Hospitalier Pitié-Salpêtrière, Paris, France). Radiofrequency transmission was performed with a body coil and the signal was received with a 12-channel head coil. Before the functional scans, high-resolution T1-weighted images ($1 \times 1 \times 1$ mm³ voxel size) were collected for anatomical coregistration using a magnetization-prepared rapid acquisition gradient-echo (MPRAGE) sequence (TR = 2300 ms, TE = 4.18 ms). Subsequently, one series of 595 blood oxygenation level-dependent (BOLD) images was obtained using a single-shot echo-planar gradient-echo (EPI) pulse sequence (TR = 2120 ms, TE = 25 ms, the first six volumes were later discarded to allow for T1 saturation). Forty-four interleaved slices ($3 \text{ mm} \times 3 \text{ mm} \times 3 \text{ mm}$ voxel size, 10% interslice gap) perpendicular with respect to the hippocampal plane were collected. The field of view was 192×192 mm² with an in-plane resolution of

64×64 pixels and a flip angle of 90°. Scanner noise was continuous during the experiment representing a constant auditory background.

DATA ANALYSIS

The fMRI data were analyzed using SPM8 (Wellcome Trust Centre for Neuroimaging). Preprocessing included spatial realignment and reslicing and coregistration of the anatomical T1 to the mean functional data. The first level analysis was carried out in the native space. Four regressors were built for each experimental condition based on the general linear model (different melodies and different lyrics ($D_M D_L$); same melodies and different lyrics ($S_M D_L$); different melodies and same lyrics ($D_M S_L$) and same melodies and same lyrics ($S_M S_L$), and convolved with a hemodynamic response function (HRF). Movement parameters were included as regressors of no interest and serial correlations were modeled with an AR (1) process. A temporal high-pass filter with a cut-off of 200 s was used to eliminate low-frequency drifts. Six one-sample *t*-tests were computed for each participant: all conditions against silence to establish a “song-sensitive” mask, the main effects of adaptation to lyrics [$(D_M D_L + S_M D_L) - (D_M S_L + S_M S_L)$] and to melodies [$(D_M D_L + D_M S_L) - (S_M D_L + S_M S_L)$] to identify areas of general adaptation to the repetition of song components, as well as the interaction [$(D_M S_L + S_M D_L) - (D_M D_L + S_M S_L)$] to isolate areas of lyrics–melody integration. For the sake of completeness and consistency with the analysis of Sammler et al. (2010), we additionally compared both main effects to identify brain regions that showed an independent processing of either lyrics or melodies (i.e., stronger adaptation for lyrics than for melodies [$2 \times (S_M D_L)$] and vice versa [$2 \times (D_M S_L)$]).

Segmentation of the anatomical files was performed with the VBM8 toolbox (Ashburner and Friston, 2005) to form a normalized anatomical image and the DARTEL exported tissue types. A template with eight iterations was created in DARTEL (Ashburner, 2007) including all 43 subjects to improve anatomical accuracy in the normalization of the functional contrast images obtained in the first level. Contrast images were spatially smoothed using a three-dimensional Gaussian kernel with 8 mm full width at half maximum. For the second level, the DARTEL normalized contrast images were normalized to the Montreal Neurological Institute (MNI) space. The automatically generated mask from the first level analysis of each subject was also normalized with this procedure but without smoothing. Statistical analysis was confined

to a song-sensitive mask in gray matter to increase signal detection (Friston et al., 1994). To create this mask, a binary mask from the last iteration of the DARTTEL template thresholded at 0.3 was overlaid with active voxels in the “all conditions against silence” contrast at $p < 0.05$ (FWE correction for multiple comparisons), $k > 5$ for all 43 participants. All voxels that were involved in both were included into the explicit song-sensitive mask for statistics. This mask covered an auditory-motor network, including the temporal gyrus, the PMC, and the cerebellum. For random effects group analyses, the individual contrast images were submitted to one-sample t -tests, separately for healthy controls, LTLE and RTLE patients. Furthermore, two-sample t -tests were computed for all contrasts, comparing each patient group against controls. All SPMs were thresholded at $p < 0.001$ (uncorrected) with a minimum cluster extent of $k \geq 5$ voxels. Results will report the peak voxel p value and the number of voxels (k).

To assess the size of the hippocampal sclerosis and surrounding cortex, volumetric measures of hippocampal, entorhinal, and parahippocampal gyrus were obtained for all participants with the Freesurfer image analysis suite (Fischl, 2012; Reuter et al., 2012), which is documented and freely available for downloading online (<http://surfer.nmr.mgh.harvard.edu/>). Non-parametric tests (Kruskal–Wallis, SPSS 18.0) were used to compare these measures between the patient and controls groups. To control global differences, intracranial volume was included in the analysis as a covariate, which was not found to be significant. The percentage of reduction of each structure was calculated for each patient group in comparison to the control group and is reported in Table 2.

RESULTS

MAIN EFFECTS

A complete report of the results at threshold $p < 0.001$ (uncorrected) with a minimum cluster extent of $k \geq 5$ voxels can be seen in Table 3. All three groups of participants showed adaptation to lyrics in the left and right STG and STS that was however considerably more extended in Controls (2474 and 2423 voxels) than in LTLE (541 and 388 voxels) and RTLE patients (201 and 165 voxels). Between-group comparisons revealed significantly weaker adaptation effects in the LTLE but not in the RTLE as compared to Controls in the left STS (Figure 1A).

In all three groups, adaptation to melody was found in the left and right STG and STS, again more extended in Controls (2380 and 1830 voxels) than in LTLE (245 and 295 voxels) and RTLE patients (106 and 111 voxels), as well as in the cerebellum. The Control group showed, in addition, adaptation in the left PMC (52 voxels) that was not observed in patients (Figure 1B). However, between-group differences failed to reach significance.

INTERACTION EFFECTS

Interaction effects were calculated with the contrast $[(D_{MSL} + S_{MDL}) - (D_{MDL} + S_{MSL})]$ and were taken to represent an integrated processing of lyrics and melodies in songs. Only the control group showed interaction effects at $p < 0.001$ $k \geq 5$, which were located in the bilateral posterior STG/STS (left: 169 voxels and right: 323 voxels). No such effect was observed in LTLE and RTLE patients. To visualize areas that simply may not have passed our statistical criterion, we inspected the data at a very

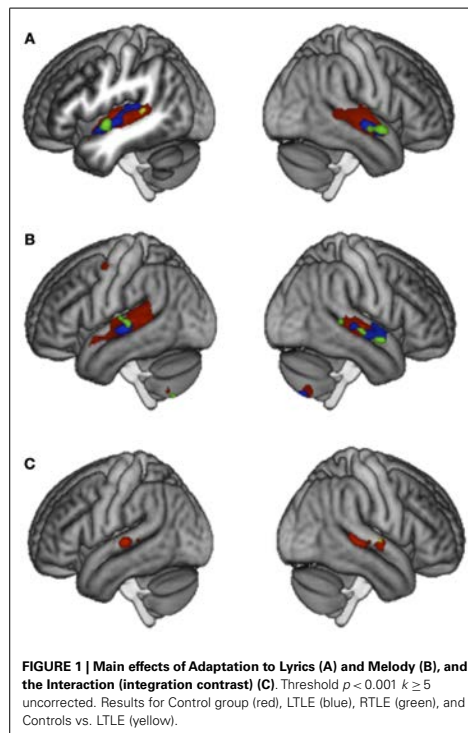
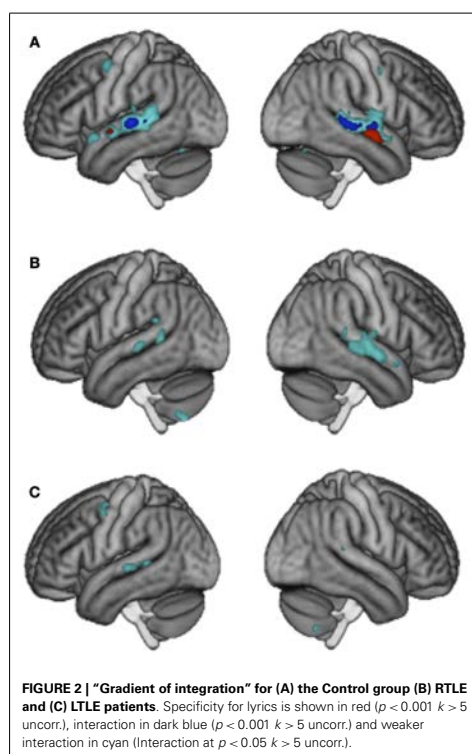


FIGURE 1 | Main effects of Adaptation to Lyrics (A) and Melody (B), and the Interaction (integration contrast) (C). Threshold $p < 0.001$ $k \geq 5$ uncorrected. Results for Control group (red), LTLE (blue), RTLE (green), and Controls vs. LTLE (yellow).

lenient level of $p < 0.05$ uncorrected ($k > 5$). Controls showed an extended region within the left (1936 voxels) and right (2176 voxels) STG/STS (Figure 2A). At this threshold, RTLE patients showed a pattern that was similar to Controls, but considerably less extended (554 and 1501 voxels) (Figure 2B). Interestingly, LTLE patients showed nearly no interaction in the temporal lobe at this very lenient threshold (238 and 35 voxels) (Figure 2C). Indeed, between-group comparisons revealed a significantly weaker interaction effect in the LTLE than the Control group in the right STG (Figure 1C) whereas the difference between the RTLE patients and Controls did not reach significance. Details on interaction effects are shown in Table 4.

INDEPENDENCE EFFECTS

Greater adaptation to lyrics as compared to melody was found bilaterally in the anterior region of the STG (23 and 196 voxels) in the control group, suggesting an independent processing of lyrics in this region. Greater adaptation to melody as compared to lyrics was obtained bilaterally in the cerebellum in RTLE patients. However, between-group differences failed to reach significance (Figure 2A). Details on independence effects are shown in Table 4.



DISCUSSION

The aim of the current study was to assess the modulatory effects of a unilateral hippocampal lesion on the incidental emergence of a song memory trace and the integration of lyrics and melodies into a conjunctive representation. To this end, neural adaptation to song repetition – as a proxy for song memory formation – was examined in patients with left or right hippocampal sclerosis in comparison to healthy controls using an fMR-adaptation paradigm. It was hypothesized that damage to the hippocampus may disrupt feedback connections to the lateral temporal lobe and thus preclude the establishment and update of a sensory memory trace. As a consequence, damage to the hippocampus may result in weaker neural adaptation in the STG. In particular, hippocampal lesions could hinder the integration of lyrics and melodies into a unified memory trace (Diana et al., 2007; Staesina and Davachi, 2009; Graham et al., 2010; Shimamura, 2010).

The main findings of this study were indeed that the neural adaptation to lyrics repetition as well as the integration of lyrics and melodies in songs (as reflected by the statistical interaction between adaptation effects for lyrics and melodies) was reduced in patients with left hippocampal sclerosis. More specifically, the

direct comparison of these patients with healthy control participants revealed a weaker adaptation to lyrics in the left STS and a weaker integration of lyrics and melodies in the right STG. If one accepts the notion that neural adaptation reflects the emergence of a memory trace (see Introduction), these results are in line with our hypotheses and previous work showing that left hippocampal damage may lead to weaker memory for lyrics (Samson and Zatorre, 1991) and may hinder the integration of lyrics and melodies into a unified memory representation (Samson and Zatorre, 1991; Sammler et al., 2010).

All three groups of participants showed adaptation to the repetition of lyrics or melodies in the bilateral STG and STS, but in both patient groups, these effects were markedly smaller in spatial extent when compared to healthy controls. Notably, patients with left (but not right) hippocampal sclerosis exhibited significantly decreased adaptation to lyrics in the left STS, which is known to play a role in phonemic processing and also known to be crucial for the perception of a sound as speech (Dehaene-Lambertz et al., 2005; Liebenthal, 2005; Möttönen et al., 2006; for a review on STS, see Hein and Knight, 2008). This finding is most likely tied to the role of the left medial temporal lobe in verbal processing (Meyer et al., 2005; Wagner et al., 2008; Greve et al., 2011) and may reflect the perturbed build-up of memory traces for lyrics (and verbal material in general) due to disrupted feedback connections between medial and lateral structures of the left temporal lobe (Eichenbaum, 2000). Such an interpretation could be supported by the verbal memory deficit documented in the LTLE patients of the present study (assessed with the RAVLT) and, although we did not collect behavioral data for this experiment, these results are also in agreement with the behavioral results of Samson and Zatorre (1991). That study showed that the recognition of sung lyrics after listening to unfamiliar songs was impaired in patients with left (but not right) medial temporal lobe lesions.

Although patients with right hippocampal sclerosis showed nominally reduced adaptation and integration effects, these did not significantly differ from those in healthy controls, suggesting rather normal song processing and lyrics-melody integration in these patients. While the latter is in line with previous behavioral data showing spared integration of lyrics and tunes after right anterior temporal lobe resection (Samson and Zatorre, 1991), our hypothesis on reduced adaptation to melodies was not confirmed. This may partly be due to the stimulus material used: even if melodies were repeated to induce adaptation, they differed in octave sung by sopranos, tenors, altos, and bass. Most likely, adaptation effects are not fully robust to transposition of melodies. Furthermore, adaptation to melodies was generally weaker than adaptation to lyrics, as attested by the results in healthy participants, possibly resulting in a floor effect. Our participants may have paid less attention to melodies than to lyrics (as the latter convey the message) leading to weak adaptation, given that a lack of attention reduces adaptation effects (Chee and Tan, 2007). Alternatively, several lines of evidence suggest that melodies may be processed more bilaterally than lyrics (Samson and Zatorre, 1992; Binder et al., 2000; Besson and Schön, 2003; Peretz and Coltheart, 2003; Schön et al., 2005; Patel, 2008; Koelsch, 2012), leading to less severe deficits in processing melodies than in verbal

Table 3 | Main effects of lyrics and melodies repetition for each group and comparison between Controls and LTLE.

| Group | Adaptation for lyrics | | | | Adaptation for melody | | | |
|------------------|-----------------------|---------|--------------|------|-----------------------|---------|---------------|------|
| | Area | Size(k) | x, y, z (mm) | Z | Area | Size(k) | x, y, z (mm) | Z |
| Control | Left hemisphere | | | | Left hemisphere | | | |
| | STG/STS | 2474 | −58, −6, −6 | 5.75 | STG/STS | 2380 | −64, −29, 3 | 5.12 |
| | | | −62, −17, 0 | 5.52 | | | −54, −39, 3 | 4.75 |
| | | | −48, −39, 6 | 5.04 | | | −66, −39, 12 | 4.68 |
| | Temporal pole | 11 | −51, 9, −18 | 3.86 | PrCG | 52 | −52, −5, 51 | 3.94 |
| | | | | | Cerebellum | 55 | −26, −62, −55 | 3.75 |
| | Right hemisphere | | | | Right hemisphere | | | |
| | STG/STS | 2423 | 62, −9, −6 | 5.59 | STG/STS | 1830 | 60, −17, −3 | 5.85 |
| | | | 60, −0, −13 | 4.76 | | | 46, −36, 2 | 4.26 |
| | | | 62, −24, 2 | 4.39 | | | 62, −0, −10 | 4.12 |
| LTLE | Cerebellum | 10 | 16, −80, −46 | 3.26 | Cerebellum | 148 | 34, −63, −58 | 3.91 |
| | | | | | Cerebellum | 27 | 30, −57, −27 | 3.59 |
| | Left hemisphere | | | | Left hemisphere | | | |
| | STG/STS | 541 | −58, −5, −9 | 4.60 | STG | 245 | −57, −21, −3 | 4.68 |
| | | | −56, −15, −3 | 4.05 | | | | |
| | | | −66, −33, 9 | 3.65 | | | | |
| | Right hemisphere | | | | Right hemisphere | | | |
| | STG/STS | 388 | 62, −9, −7 | 4.41 | STG/STS | 295 | 62, 2, −9 | 3.69 |
| | | | 62, 2, −13 | 3.85 | | | 57, −11, −9 | 3.67 |
| | | | | | | | 58, −5, −1 | 3.44 |
| RTLE | | | | | STG/STS | 92 | 54, −24, 0 | 3.69 |
| | | | | | Cerebellum | 134 | 26, −74, −60 | 3.67 |
| | Left hemisphere | | | | Left hemisphere | | | |
| | STG/STS | 201 | −63, −6, −6 | 4.26 | STG/STS | 106 | −66, −26, 2 | 3.55 |
| | | | | | | | −58, −18, 6 | 3.24 |
| | | | | | Cerebellum | 20 | −24, −66, −60 | 3.30 |
| | Right hemisphere | | | | Right hemisphere | | | |
| | Temporal pole | 165 | 62, 3, −10 | 4.13 | STG/STS | 61 | 62, −0, −12 | 3.95 |
| | STG/STS | | 63, −9, −10 | 3.29 | STG/STS | 50 | 58, −17, −4 | 3.65 |
| | | | | | STG/STS | 21 | 69, −36, 3 | 3.41 |
| Control vs. LTLE | Left hemisphere | | | | Left hemisphere | | | |
| | STS | 25 | −46, −39, 6 | 3.60 | | | | |

processing after unilateral temporal lobe damage. Further studies will be necessary to clarify this issue.

One novel finding is the main effect of melodies in the cerebellum in all groups (without group differences). Since activity in the cerebellum has been frequently reported in other studies using sung material (Parsons, 2001; Callan et al., 2007; Lebrun-Guillaud et al., 2008; Tillmann et al., 2008; Merrill et al., 2012), these effects may be linked to optimization of the fine sensory acquisition and internalization of input–output characteristics of stimuli, a process related to the creation of internal models of vocal articulation (Parsons, 2001; Callan et al., 2007; Stoodley and Schmahmann, 2009), that may function independently from the hippocampus.

As previously reported (Sammler et al., 2010), healthy participants presented maximum integration of lyrics and melodies in the posterior STS with a continuous decay of the lyrics–melodies integration along the posterior–anterior axis, toward regions of independent processing of lyrics in the anterior STG. These effects

were shown bilaterally in the present experiment, expanding the previously reported effect, which was restricted to the left hemisphere. This analysis illustrates a “gradient of integration” from more to less integrated processing. In line with the literature on music and language (Scott et al., 2000; Davis and Johnsruide, 2003; Scott and Johnsruide, 2003; Friederici, 2011; Gow, 2012), this gradient poses an integrative processing of songs at the prelexical and phonemic level in the mid-STS. Consequently, information can be transmitted both along an anterior pathway to the temporal pole for an independent analysis of the linguistic content, and along a posterior pathway to the left PMC for the integrated sensori-motor conversion of the stimuli. In other words, lyrics and melodies might split up in the ventral pathway for semantics and comprehension (Griffiths, 2001; Patterson et al., 2002; Hickok and Poeppel, 2007; Saur et al., 2008; Friederici, 2009, 2011; Hickok et al., 2011) but stay integrated in sensori-motor dorsal pathways (Kiebel et al., 2008; Loui et al., 2009).

Table 4 | Integration and independence for each group and between controls and LTLE.

| INTEGRATION | | | | |
|---------------------|------------------|----------|---------------|------|
| Group | Area | Size (k) | x, y, z (mm) | Z |
| Control | Left hemisphere | | | |
| | STG | 164 | −63, −23, 2 | 4.39 |
| | MTG | 5 | −63, −36, 3 | 3.18 |
| | Right hemisphere | | | |
| | STG | 235 | 56, −32, 0 | 3.98 |
| | STG | 82 | 51, −38, 5 | 3.65 |
| Control vs. LTLE | STG | 6 | 66, −11, −3 | 3.76 |
| | STG | 6 | 66, −20, 0 | 3.18 |
| | Right hemisphere | | | |
| | STG | 6 | 63, −9, 0 | 3.19 |
| LYRICS INDEPENDENCE | | | | |
| Control | Left hemisphere | | | |
| | STG | 15 | −63, −5, −6 | 3.29 |
| | STG | 8 | −46, −41, 8 | 3.32 |
| | Right hemisphere | | | |
| | STG | 196 | 63, −6, −15 | 3.79 |
| MELODY INDEPENDENCE | | | | |
| LTLE | Left hemisphere | | | |
| | Cerebellum | 10 | −18, −66, −61 | 3.58 |
| RTLE | Left hemisphere | | | |
| | Cerebellum | 19 | −12, −83, −46 | 3.77 |
| | Cerebellum | 6 | −24, −66, −61 | 3.21 |
| | Right hemisphere | | | |
| | Cerebellum | 18 | 14, −81, −46 | 3.62 |

Contrary to healthy participants, both patient groups showed very weak levels of lyrics–melody integration in the bilateral mid-STG/STS, and only after lowering the statistical threshold to $p < 0.05$ (uncorrected). This effect may reside on generally weaker adaptation effects in both patient groups. The spatial extent of this weak lyrics–melody interaction was particularly small in patients with left hippocampal sclerosis who also showed a significantly reduced interaction effect in the right STG as compared to controls. These tendencies suggest a partial (although not complete) disruption of integrated processing in clinical populations and indicate that the conjunctive representation of lyrics and melodies depends on intact medial temporal lobe structures, particularly in the left hemisphere. Overall, this finding is in line with previous studies in patients with anterior temporal lobe resection including parts of the hippocampus (Samson and Zatorre, 1991). These experiments showed a perturbed integration of verbal and melodic song components in patients with left (but not right) temporal lobe resections, i.e., a selective deficit in recognizing lyrics that was independent from recognition memory for melodies. It is worth to mention that in both the present and previous studies, the integration deficit may reside on a more general deficit to process lyrics, as supported by the weaker adaptation for lyrics and reduced performance in neuropsychological tests on verbal memory in our patients with left hippocampal sclerosis.

Taken together, adaptation to lyrics and integration of lyrics and melodies within songs appear to be less efficient in patients with left hippocampal damage as compared to healthy controls. We propose that these lesions may hinder the build-up of a sensory memory trace for lyrics (with rather preserved mechanisms for melodies), which in turn might be at the origin of the reduced integration of lyrics and melody. These combined effects could be attributed to hippocampal malfunction *per se* or to a more global disconnection of lateral temporal neocortical structures caused by repetitive seizures or epilepsy history (Yasuda et al., 2010; Besson et al., 2012), both of which can disrupt the hippocampal top-down modulatory influence on STG/STS. If this is the case, it is possible that adaptation could also be reduced for stimuli other than lyrics, melodies, or songs, demonstrating a more general adaptation and putative encoding deficit following disruption of cortico-hippocampal processing loops.

Interestingly, an independent analysis of the connectivity profiles in our patients showed asymmetries between the left and right hemispheric lesion groups: LTLE patients exhibited more extended and more strongly left-lateralized disconnections, as opposed to more discrete and bilateral connectivity deficits in RTLE (Besson et al., 2012). Such differences in connectivity profiles provide an additional explanation for the nominally stronger impairments in patients with left hippocampal sclerosis as compared to patients with right hippocampal sclerosis. In sum, the present data indicate that an imbalance in the left hippocampal-cortical system, due to hippocampal sclerosis and/or disrupted connectivity with STG/STS, affects the incidental emergence of a memory trace of verbal song components and precludes the build-up of a conjunctive representation that integrates lyrics and melodies.

CONCLUSION

To the best of our knowledge, this is the first study to investigate the processing of songs using fMRI in patients with unilateral hippocampal sclerosis. We showed that the adaptation to lyrics and the integration of lyrics and melodies was diminished in lateral temporal lobe regions in patients with left hippocampal sclerosis while a similar but non-significant result pattern was found in patients with right hippocampal sclerosis. These findings suggest the importance of hippocampal top-down modulations on the STG/STS during repetitive exposure to songs. We interpret the observed adaptation patterns to be a result of a disturbed connectivity in a hippocampal–cortical network, weakening the emergence of a memory trace for lyrics and the integrated processing of songs as a unified percept. Overall, these data provide a novel contribution by suggesting that the integration shown in healthy participants is tied to the integrity of the medial temporal lobe and its connections with the lateral temporal cortex.

ACKNOWLEDGMENTS

The authors are grateful to the CENIR team and Diana Omigie for their helpful assistance. Funding: the research leading to these results has received funding from an Early Stage Researcher fellowship to Irene Alonso by the European Community's Seventh Framework Programme under the Europe, Brain and Music (EBRAMUS) project – grant agreement n°238157 and by a grant

from “Agence Nationale pour la Recherche” of the French Ministry of research (project n° ANR-09-BLAN-0310-02) and a grant from the “Institut Universitaire de France” to Séverine Samson.

REFERENCES

- Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. *Neuroimage* 38, 95–113. doi:10.1016/j.neuroimage.2007.07.007
- Ashburner, J., and Friston, K. J. (2005). Unified segmentation. *Neuroimage* 26, 839–851. doi:10.1016/j.neuroimage.2005.02.018
- Axmacher, N., Mormann, F., Fernandez, G., Cohen, M. X., Elger, C. E., and Fell, J. (2007). Sustained neural activity patterns during working memory in the human medial temporal lobe. *J. Neurosci.* 27, 7807–7816. doi:10.1523/JNEUROSCI.0962-07.2007
- Barense, M. D., Gaffan, D., and Graham, K. S. (2007). The human medial temporal lobe processes online representations of complex objects. *Neuropsychologia* 45, 2963–2974. doi:10.1016/j.neuropsychologia.2007.05.023
- Bast, T. (2007). Toward an integrative perspective on hippocampal function: from the rapid encoding of experience to adaptive behavior. *Rev. Neurosci.* 18, 253–282. doi:10.1515/REVNEURO.2007.18.3-4.253
- Baur, B., Utner, I., Ilmberger, J., Fesl, G., and Mai, N. (2000). Music memory provides access to verbal knowledge in a patient with global amnesia. *Neurocase* 6, 415–421. doi:10.1080/13554790008402712
- Baxter, M. G. (2009). Involvement of medial temporal lobe structures in memory and perception. *Neuron* 61, 667–677. doi:10.1016/j.neuron.2009.02.007
- Belin, P., and Zatorre, R. (2003). Adaptation to speaker's voice in right anterior temporal lobe. *Brain Imaging* 14, 2105–2109. doi:10.1097/01.wnr.0000091689.94870.85
- Besson, M., and Schön, D. (2003). Comparison between language and music. *Ann. N. Y. Acad. Sci.* 930, 232–258. doi:10.1111/j.1749-6632.2001.tb05736.x
- Besson, P., Dinkelaar, V., Valabregue, R., Samson, S., Thivard, L., Baulac, M., et al. (2012). “Decreased connectivity of major cortical nodes is more pronounced and lateralized in left versus right mesial temporal sclerosis,” in *Poster Presented at 66th American Epilepsy Society Annual Meeting*, San Diego.
- Bettus, G., Guedj, E., Joyeux, F., Confort-Gouny, S., Soulier, E., Laguitton, V., et al. (2009). Decreased basal fMRI functional connectivity in epileptogenic networks and contralateral compensatory mechanisms. *Hum. Brain Mapp.* 30, 1580–1591. doi:10.1002/hbm.20625
- Binder, J. R., Frost, J. A., Hammeke, T., Bellgowan, P. S., Springer, J., Kaufman, J., et al. (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cereb. Cortex* 10, 512–528. doi:10.1093/cercor/10.5.512
- Blondin, F., and Lepage, M. (2005). Decrease and increase in brain activity during visual perceptual priming: an fMRI study on similar but perceptually different complex visual scenes. *Neuropsychologia* 43, 1887–1900. doi:10.1016/j.neuropsychologia.2005.03.021
- Brown, S., Martinez, M. J., Hodges, D. A., Fox, P. T., and Parsons, L. M. (2004a). The song system of the human brain. *Brain Res. Cogn. Brain Res.* 20, 363–375. doi:10.1016/j.cogbrainres.2004.03.016
- Brown, S., Martinez, M. J., and Parsons, L. M. (2004b). Passive music listening spontaneously engages limbic and paralimbic systems. *Neuroreport* 15, 2033–2037. doi:10.1097/00001756-200409150-00008
- Buchsbaum, B. R., and D'Esposito, M. (2009). Repetition suppression and reactivation in auditory-verbal short-term recognition memory. *Cereb. Cortex* 19, 1474–1485. doi:10.1093/cercor/bhn186
- Bussey, T. J., and Saksida, L. M. (2005). Object memory and perception in the medial temporal lobe: an alternative approach. *Curr. Opin. Neurobiol.* 15, 730–737. doi:10.1016/j.conb.2005.10.014
- Callan, D. E., Kawato, M., Parsons, L., and Turner, R. (2007). Speech and song: the role of the cerebellum. *Cerebellum* 6, 321–327. doi:10.1080/14734220601187733
- Callan, D. E., Tsytarev, V., Hanakawa, T., Callan, A. M., Katsuhara, M., Fukuyama, H., et al. (2006). Song and speech: brain regions involved with perception and covert production. *Neuroimage* 31, 1327–1342. doi:10.1016/j.neuroimage.2006.01.036
- Chee, M. W., and Tan, J. C. (2007). Inter-relationships between attention, activation, fMRI adaptation and long-term memory. *Neuroimage* 37, 1487–1495. doi:10.1016/j.neuroimage.2007.07.006
- Chun, M. M., and Phelps, E. A. (1999). Memory deficits for implicit contextual information in amnesic subjects with hippocampal damage. *Nat. Neurosci.* 2, 844–847. doi:10.1038/12222
- Cowell, R. A., Bussey, T. J., and Saksida, L. M. (2006). Why does brain damage impair memory? A connectionist model of object recognition memory in perirhinal cortex. *J. Neurosci.* 26, 12186–12197. doi:10.1523/JNEUROSCI.2818-06.2006
- Cowell, R. A., Bussey, T. J., and Saksida, L. M. (2010). Components of recognition memory: dissociable cognitive processes or just differences in representational complexity? *Hippocampus* 20, 1245–1262. doi:10.1002/hipo.20865
- Crowder, R. G., Serafine, M. L., and Repp, B. (1990). Physical interaction and association by contiguity in memory for the words and melodies of songs. *Mem. Cognit.* 18, 469–476. doi:10.3758/BF03198480
- Damasio, A. R. (1989). The brain binds entities and events by multiregional activation from convergence zones. *Neural Comput.* 1, 123–132. doi:10.1162/neco.1989.1.1.123
- Davachi, L. (2006). Item, context and relational episodic encoding in humans. *Curr. Opin. Neurobiol.* 16, 693–700. doi:10.1016/j.conb.2006.10.012
- Davis, M. H., and Johnsrude, I. S. (2003). Hierarchical processing in spoken language comprehension. *J. Neurosci.* 23, 3423–3431.
- Dehaene-Lambertz, G., Pallier, C., Serniclaes, W., Sprenger-Charolles, L., Jobert, A., and Dehaene, S. (2005). Neural correlates of switching from auditory to speech perception. *Neuroimage* 24, 21–33. doi:10.1016/j.neuroimage.2004.09.039
- Diana, R. A., Yonelinas, A. P., and Ranganath, C. (2007). Imaging recollection and familiarity in the medial temporal lobe: a three-component model. *Trends Cogn. Sci. (Regul. Ed.)* 11, 379–386. doi:10.1016/j.tics.2007.08.001
- Diehl, B., Tkach, J., Piao, Z., Ruggieri, P., LaPresto, E., Liu, P., et al. (2010). Diffusion tensor imaging in patients with focal epilepsy due to cortical dysplasia in the temporo-occipital region: electro-clinico-pathological correlations. *Epilepsy Res.* 90, 178–187. doi:10.1016/j.eplepsyres.2010.03.006
- Dobbins, I. G., Schnyer, D. M., Verfaellie, M., and Schacter, D. L. (2004). Cortical activity reductions during repetition priming can result from rapid response learning. *Nature* 428, 316–319. doi:10.1038/nature02400
- Eichenbaum, H. (2000). A cortical-hippocampal system for declarative memory. *Nat. Rev. Neurosci.* 1, 41–50. doi:10.1038/35036213
- Eichenbaum, H., Otto, T., and Cohen, N. J. (1994). Two functional components of the hippocampal memory system. *Behav. Brain Sci.* 17, 449–471. doi:10.1017/S0140525X00035391
- Fischl, B. (2012). FreeSurfer. *Neuroimage* 62, 774–781. doi:10.1016/j.neuroimage.2012.01.021
- Focke, N. K., Yogarajah, M., Bonelli, S. B., Bartlett, P. A., Symms, M. R., and Duncan, J. S. (2008). Voxel-based diffusion tensor imaging in patients with mesial temporal lobe epilepsy and hippocampal sclerosis. *Neuroimage* 40, 728–737. doi:10.1016/j.neuroimage.2007.12.031
- Friederici, A. D. (2009). Pathways to language: fiber tracts in the human brain. *Trends Cogn. Sci. (Regul. Ed.)* 13, 175–181. doi:10.1016/j.tics.2009.01.001
- Friederici, A. D. (2011). The brain basis of language processing: from structure to function. *Physiol. Rev.* 91, 1357–1392. doi:10.1152/physrev.00006.2011
- Friston, K., Jezzard, P., and Turner, R. (1994). Analysis of functional MRI time-series. *Hum. Brain Mapp.* 1, 153–171. doi:10.1002/hbm.460010207
- Fuster, J. M. (1997). Network memory. *Trends Neurosci.* 20, 451–459. doi:10.1016/S0166-2236(97)01128-4
- Ganel, T., Gonzalez, C. L., Valyear, K. F., Culham, J. C., Goodale, M. A., and Köhler, S. (2006). The relationship between fMRI adaptation and repetition priming. *Neuroimage* 32, 1432–1440. doi:10.1016/j.neuroimage.2006.05.039
- Goh, J. O. S., Chee, M. W. L., Tan, J., Venkatraman, V., Hebrank, A., Leshikar, E., et al. (2007). Age and culture modulate object processing and object-scene binding in the ventral visual areas. *Cogn. Affect. Behav. Neurosci.* 7, 44–52. doi:10.3758/CABN.7.1.44
- Gow, D. W. (2012). The cortical organization of lexical knowledge: a dual lexicon model of spoken language processing. *Brain Lang.* 121, 273–288. doi:10.1016/j.bandl.2012.03.005
- Graham, K. S., Barense, M. D., and Lee, A. C. (2010). Going beyond LTM in the MTL: a synthesis of neuropsychological and neuroimaging findings on the role of the medial temporal lobe in memory and perception. *Neuropsychologia* 48, 831–853. doi:10.1016/j.neuropsychologia.2010.01.001
- Greve, A., Evans, C. J., Graham, K. S., and Wilding, E. L. (2011). Functional specialisation in the hippocampus and perirhinal cortex during the encoding of verbal associations. *Neuropsychologia* 49, 2746–2754. doi:10.1016/j.neuropsychologia.2011.06.002
- Griffiths, T. D. (2001). The neural processing of complex sounds. *Ann. N. Y. Acad. Sci.* 930, 133–142. doi:10.1111/j.1749-6632.2001.tb05729.x

- Grill-Spector, K., Henson, R., and Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. *Trends Cogn. Sci. (Regul. Ed.)* 10, 14–23. doi:10.1016/j.tics.2005.11.006
- Hebert, S., and Peretz, I. (2001). Are text and tune of familiar songs separable by brain damage? *Brain Cogn.* 46, 169–175. doi:10.1016/S0278-2626(01)80058-0
- Hein, G., and Knight, R. T. (2008). Superior temporal sulcus-it's my area: or is it? *J. Cogn. Neurosci.* 20, 2125–2136. doi:10.1162/jocn.2008.20148
- Hickok, G., Houde, J., and Rong, F. (2011). Sensorimotor integration in speech processing: computational basis and neural organization. *Neuron* 69, 407–422. doi:10.1016/j.neuron.2011.01.019
- Hickok, G., and Poeppel, D. (2007). The cortical organization of speech processing. *Nat. Rev. Neurosci.* 8, 393–402. doi:10.1038/nrn2113
- Johnson, S. K., and Halpern, A. R. (2012). Semantic priming of familiar songs. *Mem. Cognit.* 40, 579–593. doi:10.3758/s13421-011-0175-z
- Kiebel, S. J., Daunizeau, J., and Friston, K. J. (2008). A hierarchy of time-scales and the brain. *PLoS Comput. Biol.* 4:e1000209. doi:10.1371/journal.pcbi.1000209
- Koelsch, S. (2012). *Brain and Music*. West Sussex: Wiley-Blackwell.
- Krekelberg, B., Boynton, G., and Vanwezel, R. (2006). Adaptation: from single cells to bold signals. *Trends Neurosci.* 29, 250–256. doi:10.1016/j.tins.2006.02.008
- Lebrun-Guillaud, G., Tillmann, B., and Justus, T. (2008). Perception of tonal and temporal structures in chord sequences by patients with cerebellar damage. *Music Percept.* 25, 271–283. doi:10.1525/mp.2008.25.4.271
- Lee, A. C. (2006). Differentiating the roles of the hippocampus and perirhinal cortex in processes beyond long-term declarative memory: a double dissociation in dementia. *J. Neurosci.* 26, 5198–5203. doi:10.1523/JNEUROSCI.3157-05.2006
- Lee, A. C., Bussey, T. J., Murray, E. A., Saksida, L. M., Epstein, R. A., Kapur, N., et al. (2005). Perceptual deficits in amnesia: challenging the medial temporal lobe 'mnemonic' view. *Neuropsychologia* 43, 1–11. doi:10.1016/j.neuropsychologia.2004.07.017
- Lee, A. C., and Rudebeck, S. R. (2010a). Human medial temporal lobe damage can disrupt the perception of single objects. *J. Neurosci.* 30, 6588–6594. doi:10.1523/JNEUROSCI.0116-10.2010
- Lee, A. C., and Rudebeck, S. R. (2010b). Investigating the interaction between spatial perception and working memory in the human medial temporal lobe. *J. Cogn. Neurosci.* 22, 2823–2835. doi:10.1162/jocn.2009.21396
- Lehericy, S., Cohen, L., Bazin, B., Samson, S., Giacomini, E., Rougetet, R., et al. (2000). Functional MR evaluation of temporal and frontal language dominance compared to the Wada test. *Neurology* 54, 1625–1633. doi:10.1212/WNL.54.8.1625
- Liao, W., Zhang, Z., Pan, Z., Mantini, D., Ding, J., Duan, X., et al. (2011). Default mode network abnormalities in mesial temporal lobe epilepsy: a study combining fMRI and DTI. *Hum. Brain Mapp.* 32, 883–895. doi:10.1002/hbm.21076
- Liebsenthal, E. (2005). Neural substrates of phonemic perception. *Cereb. Cortex* 15, 1621–1631. doi:10.1093/cercor/bhi040
- Loui, P., Wu, E. H., Wessel, D. L., and Knight, R. T. (2009). A generalized mechanism for perception of pitch patterns. *J. Neurosci.* 29, 454–459. doi:10.1523/JNEUROSCI.4503-08.2009
- McClelland, J. L., McNaughton, B. L., and O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychol. Rev.* 102, 419.
- Merrill, J., Sammler, D., Bangert, M., Goldhahn, D., Lohmann, G., Turner, R., et al. (2012). Perception of words and pitch patterns in song and speech. *Front. Psychol.* 3:76. doi:10.3389/fpsyg.2012.00076
- Meyer, P., Mecklinger, A., Grunwald, T., Fell, J., Elger, C. E., and Friederici, A. D. (2005). Language processing within the human medial temporal lobe. *Hippocampus* 15, 451–459. doi:10.1002/hipo.20070
- Mishkin, M. (1982). A memory system in the monkey. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 298, 85–95.
- Möttönen, R., Calvert, G. A., Jääskeläinen, I. P., Matthews, P. M., Thesen, T., Tuomainen, J., et al. (2006). Perceiving identical sounds as speech or non-speech modulates activity in the left posterior superior temporal sulcus. *Neuroimage* 30, 563–569. doi:10.1016/j.neuroimage.2005.10.002
- Noppeney, U., and Price, C. J. (2004). An fMRI study of syntactic adaptation. *J. Cogn. Neurosci.* 16, 702–713. doi:10.1162/089892904323057399
- Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113.
- Olsen, R. K., Moses, S. N., Riggs, L., and Ryan, J. D. (2012). The hippocampus supports multiple cognitive processes through relational binding and comparison. *Front. Hum. Neurosci.* 6:146. doi:10.3389/fnhum.2012.00146
- Opitz, B. (2010). Neural binding mechanisms in learning and memory. *Neurosci. Biobehav. Rev.* 34, 1036–1046. doi:10.1016/j.neubiorev.2009.11.001
- O'Reilly, R. C., and Rudy, J. W. (2001). Conjunctive representations in learning and memory: principles of cortical and hippocampal function. *Psychol. Rev.* 108, 311–345. doi:10.1037/0033-295X.108.2.311
- Overath, T., Cusack, R., Kumar, S., Kriegstein, K., von Warren, J. D., Grube, M., et al. (2007). An information theoretic characterisation of auditory encoding. *PLoS Biol.* 5:e288. doi:10.1371/journal.pbio.0050288
- Overath, T., Kumar, S., Kriegstein, K., and von Griffiths, T. D. (2008). Encoding of spectral correlation over time in auditory cortex. *J. Neurosci.* 28, 13268–13273. doi:10.1523/JNEUROSCI.4596-08.2008
- Parsons, L. M. (2001). Exploring the functional neuroanatomy of music performance, perception, and comprehension. *Ann. N. Y. Acad. Sci.* 930, 211–231. doi:10.1111/j.1749-6632.2001.tb05735.x
- Patel, A. (2008). *Music, Language, and the Brain*. New York: Oxford University Press.
- Patterson, R. D., Uppenkamp, S., Johnsrude, I. S., and Griffiths, T. D. (2002). The processing of temporal pitch and melody information in auditory cortex. *Neuron* 36, 767–776. doi:10.1016/S0896-6273(02)01060-7
- Peretz, I., and Coltheart, M. (2003). Modularity of music processing. *Nat. Neurosci.* 6, 688–691. doi:10.1038/nm1083
- Peretz, I., Gagnon, L., Hebert, S., and Macoir, J. (2004). Singing in the brain: insights from cognitive neuropsychology. *Music Percept.* 21, 373–390. doi:10.1525/mp.2004.21.3.373
- Peynircioglu, Z. F., Rabinovitz, B. E., and Thompson, J. L. (2008). Memory and metamemory for songs: the relative effectiveness of titles, lyrics, and melodies as Cues for each other. *Psychol. Music* 36, 47–61. doi:10.1177/0305735607079722
- Reuter, M., Schmansky, N. J., Rosas, H. D., and Fischl, B. (2012). Within-subject template estimation for unbiased longitudinal image analysis. *Neuroimage* 61, 1402–1418. doi:10.1016/j.neuroimage.2012.02.084
- Rey, A. (1964). *L'Examen Clinique en Psychologie*. Paris: Presses Universitaires de France.
- Riley, J. D., Franklin, D. L., Choi, V., Kim, R. C., Binder, D. K., Cramer, S. C., et al. (2010). Altered white matter integrity in temporal lobe epilepsy: association with cognitive and clinical profiles. *Epilepsia* 51, 536–545. doi:10.1111/j.1528-1167.2009.02508.x
- Robine, M. (1994). *Anthologie de la chanson française—des trouvères aux grands auteurs du XIXe siècle*. Paris: Albin Michel.
- Rolls, E. T. (1996). A theory of hippocampal function in memory. *Hippocampus* 6, 601–620.
- Rose, N. S., Olsen, R. K., Craik, F. I., and Rosenbaum, R. S. (2012). Working memory and amnesia: the role of stimulus novelty. *Neuropsychologia* 50, 11–18. doi:10.1016/j.neuropsychologia.2011.10.016
- Rudy, J. W., and Sutherland, R. J. (1995). Configural association theory and the hippocampal formation: an appraisal and reconfiguration. *Hippocampus* 5, 375–389.
- Saito, Y., Ishii, K., Sakuma, N., Kawasaki, K., Oda, K., and Mizusawa, H. (2012). Neural substrates for semantic memory of familiar songs: is there an interface between lyrics and melodies? *PLoS ONE* 7:e46354. doi:10.1371/journal.pone.0046354
- Sammler, D., Baird, A., Valabregue, R., Clement, S., Dupont, S., Belin, P., et al. (2010). The relationship of lyrics and tunes in the processing of unfamiliar songs: a functional magnetic resonance adaptation study. *J. Neurosci.* 30, 3572–3578. doi:10.1523/JNEUROSCI.2751-09.2010
- Samson, S., and Zatorre, R. J. (1991). Recognition memory for text and melody of songs after unilateral temporal lobe lesion: evidence for dual encoding. *J. Exp. Psychol. Learn. Mem. Cogn.* 17, 793–804.
- Samson, S., and Zatorre, R. J. (1992). Learning and retention of melodic and verbal information after unilateral temporal lobectomy. *Neuropsychologia* 30, 815–826.
- Saur, D., Kreher, B. W., Schnell, S., Kummerer, D., Kellmeyer, P., Vry, M.-S., et al. (2008). Ventral and dorsal pathways for language. *Proc. Natl. Acad. Sci. U.S.A.* 105, 18035–18040. doi:10.1073/pnas.0805234105
- Schön, D., Gordon, R., Campagne, A., Magne, C., Astésano, C., Anton, J., et al. (2010). Similar cerebral networks in language, music and song perception. *Neuroimage* 51, 450–461. doi:10.1016/j.neuroimage.2010.02.023

- Schön, D., Gordon, R. L., and Besson, M. (2005). Musical and linguistic processing in song perception. *Ann. N. Y. Acad. Sci.* 1060, 71–81. doi:10.1196/annals.1360.006
- Scott, S. K., Blank, C. C., Rosen, S., and Wise, R. J. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain* 123, 2400–2406. doi:10.1093/brain/123.12.2400
- Scott, S. K., and Johnsrude, I. S. (2003). The neuroanatomical and functional organization of speech perception. *Trends Neurosci.* 26, 100–107. doi:10.1016/S0166-2236(02)00037-1
- Scoville, W. B., and Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *J. Neurol. Neurosurg. Psychiatr.* 20, 11–21.
- Serafine, M., Crowder, R. G., and Repp, B. (1984). Integration of melody and text in memory for songs. *Cognition* 16, 285–303.
- Serafine, M., Davidson, J., Crowder, R. G., and Repp, B. (1986). On the nature of melody-text integration in memory for songs. *J. Mem. Lang.* 25, 123–135.
- Shimamura, A. P. (2010). Hierarchical relational binding in the medial temporal lobe: the strong get stronger. *Hippocampus* 20, 1206–1216. doi:10.1002/hipo.20856
- Staresina, B. P., and Davachi, L. (2009). Mind the gap: binding experiences across space and time in the human hippocampus. *Neuron* 63, 267–276. doi:10.1016/j.neuron.2009.06.024
- Stoodley, C., and Schmahmann, J. (2009). Functional topography in the human cerebellum: a meta-analysis of neuroimaging studies. *Neuroimage* 44, 489–501. doi:10.1016/j.neuroimage.2008.08.039
- Suarez, R. O., Golby, A., Whalen, S., Sato, S., Theodore, W. H., Kufita, C. V., et al. (2010). Contributions to singing ability by the posterior portion of the superior temporal gyrus of the non-language-dominant hemisphere: first evidence from subdural cortical stimulation, Wada testing, and fMRI. *Cortex* 46, 343–353. doi:10.1016/j.cortex.2009.04.010
- Sutherland, R. J., and Rudy, J. W. (1989). Configural association theory: the role of the hippocampal formation in learning, memory, and amnesia. *Psychobiology* 17, 129–144.
- Suzuki, W. A. (2009). Perception and the medial temporal lobe: evaluating the current evidence. *Neuron* 61, 657–666. doi:10.1016/j.neuron.2009.02.009
- Suzuki, W. A., and Baxter, M. G. (2009). Memory, perception, and the medial temporal lobe: a synthesis of opinions. *Neuron* 61, 678–679. doi:10.1016/j.neuron.2009.02.009
- Sziklas, V., and Jones-Gotman, M. (2008). RAVLT and nonverbal analog: French forms and clinical findings. *Can. J. Neurol. Sci.* 35, 323–330.
- Thivard, L., Hombrouck, J., Tézenas du Montcel, S., Delmaire, C., Cohen, L., Samson, S., et al. (2005). Productive and perceptive language reorganisation in temporal lobe epilepsy. *Neuroimage* 24, 841–851. doi:10.1016/j.neuroimage.2004.10.001
- Tierney, A., Dick, F., Deutsch, D., and Sereno, M. (2012). Speech versus song: multiple pitch-sensitive areas revealed by a naturally occurring musical illusion. *Cereb. Cortex* 23, 255–263. doi:10.1093/cercor/bhs003
- Tillmann, B., Justus, T., and Bigand, E. (2008). Cerebellar patients demonstrate preserved implicit knowledge of association strengths in musical sequences. *Brain Cogn.* 66, 161–167. doi:10.1016/j.bandc.2007.07.005
- Turk-Browne, N. B., Yi, D.-J., and Chun, M. M. (2006). Linking implicit and explicit memory: common encoding factors and shared representations. *Neuron* 49, 917–927. doi:10.1016/j.neuron.2006.01.030
- Wagner, K., Frings, L., Spreer, J., Buller, A., Everts, R., Halsband, U., et al. (2008). Differential effect of side of temporal lobe epilepsy on lateralization of hippocampal, temporolateral, and inferior frontal activation patterns during a verbal episodic memory task. *Epilepsy Behav.* 12, 382–387. doi:10.1016/j.yebeh.2007.11.003
- Winters, B. D. (2004). Double dissociation between the effects of peri-posterior rhinal cortex and hippocampal lesions on tests of object recognition and spatial memory: heterogeneity of function within the temporal lobe. *J. Neurosci.* 24, 5901–5908. doi:10.1523/JNEUROSCI.1346-04.2004
- Yasuda, C. L., Valise, C., Saúde, A. V., Pereira, A. R., Pereira, F. R., Ferreira Costa, A. L., et al. (2010). Dynamic changes in white and gray matter volume are associated with outcome of surgical treatment in temporal lobe epilepsy. *Neuroimage* 49, 71–79. doi:10.1016/j.neuroimage.2009.08.014
- Zarahn, E. (2004). Positive evidence against human hippocampal involvement in working memory maintenance of familiar stimuli. *Cereb. Cortex* 15, 303–316. doi:10.1093/cercor/bbh132
- Zola-Morgan, S., and Squire, L. R. (1993). Neuroanatomy of memory. *Annu. Rev. Neurosci.* 16, 547–563. doi:10.1146/annurev.ne.16.030193.002555

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 11 December 2013; paper pending published: 20 January 2014; accepted: 13 February 2014; published online: 27 February 2014.

Citation: Alonso I, Sammler D, Valabrègue R, Dinkelacker V, Dupont S, Belin P and Samson S (2014) Hippocampal sclerosis affects fMRI-adaptation of lyrics and melodies in songs. *Front. Hum. Neurosci.* 8:111. doi: 10.3389/fnhum.2014.00111

This article was submitted to the journal *Frontiers in Human Neuroscience*.

Copyright © 2014 Alonso, Sammler, Valabrègue, Dinkelacker, Dupont, Belin and Samson. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Chapter 4

Melody in (inter)action

4.1 Study VII – Speech prosody in (inter)action

Hellbernd, N., Sammler, D. (2016). Prosody conveys speaker's intentions: acoustic cues for speech act perception. *Journal of Memory and Language*, 88, 70-86.*

[Author contributions by D.S.: designed research with N.H.; supervised research and data analysis; co-wrote the paper with N.H.]

*Reprinted with permission from Elsevier.

Journal of Memory and Language 88 (2016) 70–86



Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Prosody conveys speaker's intentions: Acoustic cues for speech act perception



Nele Hellbernd*, Daniela Sammler

Otto Hahn Group "Neural Bases of Intonation in Speech", Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

ARTICLE INFO

Article history:
Received 7 May 2015
revision received 5 January 2016

Keywords:

Prosody
Intention
Speech acts
Acoustics
Pragmatics

ABSTRACT

Action-theoretic views of language posit that the recognition of others' intentions is key to successful interpersonal communication. Yet, speakers do not always code their intentions literally, raising the question of which mechanisms enable interlocutors to exchange communicative intents. The present study investigated whether and how prosody—the vocal tone—contributes to the identification of “unspoken” intentions. Single (non-)words were spoken with six intonations representing different speech acts—as carriers of communicative intentions. This corpus was acoustically analyzed (Experiment 1), and behaviorally evaluated in two experiments (Experiments 2 and 3). The combined results show characteristic prosodic feature configurations for different intentions that were reliably recognized by listeners. Interestingly, identification of intentions was not contingent on context (single words), lexical information (non-words), and recognition of the speaker's emotion (valence and arousal). Overall, the data demonstrate that speakers' intentions are represented in the prosodic signal which can, thus, determine the success of interpersonal communication.

© 2016 Elsevier Inc. All rights reserved.

Introduction

During conversations, humans regularly decode not only *what* is said but also *why* (Bühler, 1934; Grice, 1957; Wittgenstein, 1953). Depending on the latter, we may understand the same statement “It's hard to be punctual in the morning” as empathic concern, criticism, or simply as a matter of facts. Pragmatic theory posits that it is particularly the *why*—the communicative intention of the speaker—that drives the recipient's behavior and is the motive of communication. Yet, how intentions are (de)coded in interpersonal communication is still not fully

understood. Contemporary pragma-linguistic theories posit that listeners identify the speaker's goal via pragmatic inference (Wilson & Sperber, 2012), taking conversation context and “common ground” (Clark & Carlson, 1981; Levinson, 2013; Stalnaker, 2002; Tomasello, 2005; Wichmann, 2002) into account. Alternatively, other studies seek to identify extralinguistic cues that reveal a speaker's intention, such as facial expressions (Fridlund, 1994; Frith, 2009; Parkinson, 2005), properties of biological motion (Di Cesare, Di Dio, Marchi, & Rizzolatti, 2015), or gestures (Bucciarelli, Colle, & Bara, 2003; Enrici, Adenzato, Cappa, Bara, & Tettamanti, 2011). The present study will focus on speech prosody—the tone of the voice—and will weigh its potential to convey communicative intentions.

The question of how interlocutors decode the *why* of an utterance is grounded in *action-theories of language*. In the middle of the 20th century, scholars like Bühler (1934), Wittgenstein (1953), or Grice (1975) recognized that language is more than strings of symbols that are understood

* Corresponding author at: Otto Hahn Group “Neural Bases of Intonation in Speech”, Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstraße 1a, 04103 Leipzig, Germany. Fax: +49 341 9940 2204.

E-mail addresses: hellbernd@cbs.mpg.de (N. Hellbernd), sammler@cbs.mpg.de (D. Sammler).

<http://dx.doi.org/10.1016/j.jml.2016.01.001>

0749-596X/© 2016 Elsevier Inc. All rights reserved.

by retrieving their conventional, *coded* meaning. In their view, language is an *intentional action* and gains meaning through its employment. Utterances become instruments to influence the behavior of the interlocutor. The meaning of an utterance must be found in its underlying intention. It was Grice (1957) who particularly promoted the central role of intentions in communication. He advocated the idea that intentions drive speakers' behaviors (e.g., utterances) whose sole function is to have an effect on the addressee by virtue of having their intention recognized (cf. Levinson, 2006). Notably, the intention of the speaker—the *speaker meaning* in Grice's terms—not necessarily surfaces in the overt lexical content of the utterance, as shown in the example on punctuality above, but needs to be interpreted by the listener.

This idea later became central to speech act theory by Austin (1962) and Searle (1969) who considered utterances as actions—or *speech acts*—with specific interpersonal goals such as promising, apologizing, or warning. Like Grice, they claimed that speakers convey information on at least two levels: (1) the *propositional content* carrying the lexical meaning of *what* is said, and (2) the *illocutionary force* representing the action and speaker's intention—the *why*. As mentioned above, it is this second level—what the speaker is attempting to accomplish with a remark—that is thought to predominantly drive the interlocutor's (conversational) reaction. Notably, illocutionary force is often expressed implicitly (i.e. without the performative verb) or even indirectly, hence requiring some sort of inference on the part of the listener (Austin, 1962; Bach, 1994).

Interestingly, the notion of implicitness and indirectness conflicts with Grice's *cooperative principle* (1975), which describes principles for effective communication in conversation in four maxims. Following his *maxim of manner*, speakers ought to shape their utterances in ways that support the purpose of the conversation. Hence, speakers should produce unambiguous cues that make their intentions comprehensible to listeners. The fact that this seems often not to be the case but listeners still efficiently recognize the speaker's intent has fueled research on the cognitive and neural bases of comprehending communicative intentions. A great deal of work has focused on implicit speech acts, i.e. utterances that express the speaker's intention and illocutionary force without inclusion of the performative verb (e.g. "I will be there." expressing a promise without including the verb "promise"). These studies demonstrated the psychological reality of speech acts (Holtgraves, 2005), their automatic (Holtgraves, 2008a; Liu, 2011) and early recognition during conversation turns (Egorova, Pulvermüller, & Shtyrov, 2014; Egorova, Shtyrov, & Pulvermüller, 2013; Gisladdottir, Chwilla, & Levinson, 2015), and their importance for conversation memory (Holtgraves, 2008b). However, despite their importance for understanding human communication, these studies remain incomplete in one particular way: They often rely on written linguistic material and, thus, miss out on extralinguistic cues that are usually available during natural spoken conversations. These cues comprise signals expressed via additional communicative channels like eyes, face, body, or voice and may render the speaker's intention less implicit and indirect than typically thought.

The present study will focus on vocal acoustic cues, i.e., prosody, as one non-verbal channel in interpersonal conversation that may play an important role for speakers and listeners to express and recognize communicative intentions.

The term prosody refers to variations in pitch, loudness, timing, or voice quality over the course of an utterance (Warren, 1999) that can modify the communicative content of a message, both linguistically and paralinguistically (Bolinger, 1986). Linguistically, prosody has direct effects on the information structure of an utterance. It conveys, for example, semantic relationships (Cutler, Dahan, & van Donselaar, 1997; Wagner & Watson, 2010), disambiguates the syntactic constituent structure (Carlson, Frazier, & Clifton, 2009), and marks declarative vs. interrogative sentence mode (Sammler, Grosbras, Anwender, Bestelmeyer, & Belin, 2015; Schneider, Lintfert, Dogil, & Möbius, 2006; Srinivasan & Massaro, 2003). Paralinguistically, the "manner of saying" conveys additional information that goes beyond the linguistic content. Whether or not this includes intentions is a matter of debate (Bolinger, 1986) and will be topic of the present research.

Until now, most studies on paralinguistic prosody either focused on the speaker's emotion (Banse & Scherer, 1996; Bänziger & Scherer, 2005; Frick, 1985; Simon-Thomas, Keltner, Sauter, Sinicropi-Yao, & Abramson, 2009) or, more recently, on their attitude, for example, the politeness, confidence, or sincerity of the speaker (Jiang & Pell, 2015; Monetta, Cheang, & Pell, 2008; Rigoulot, Fish, & Pell, 2014) and often sought to determine links between the acoustics of the prosodic signal and the listeners' comprehension of the paralinguistic message. Although opinions diverge on whether prosody as such can convey meaning, i.e. without contextual information (see below) (Cutler, 1976; Wichmann, 2000, 2002), studies revealed distinct acoustic properties for the prosodic expression of different emotions (Banse & Scherer, 1996; Szameitat, Alter, Szameitat, Darwin, et al., 2009; Szameitat, Alter, Szameitat, Wildgruber, et al., 2009) and attitudes (Blanc & Dominey, 2003; Morlec, Bailly, & Aubergé, 2001; Uldall, 1960). Similarly, on the perception side, researchers showed that participants were able to identify the speaker's attitude (Morlec et al., 2001; Uldall, 1960) and emotion by prosodic differences alone, in verbal (Banse & Scherer, 1996; Morlec et al., 2001) and non-verbal utterances (Monetta et al., 2008; Sauter, Eisner, Calder, & Scott, 2010), in laughter (Szameitat, Alter, Szameitat, Darwin, et al., 2009; Szameitat, Alter, Szameitat, Wildgruber, et al., 2009), and to some extent even cross-culturally (Sauter, Eisner, Ekman, & Scott, 2010).

Compared to this active field of research, only little is known about the perceptual reality, relevance and effectiveness of prosodic cues in conveying *intentions*. We consider communicative intentions as the goals of interpersonal actions (e.g., language) that are meant to be recognized by the interlocutor and to influence her (conversational) reactions. This differentiates communicative intentions from basic emotions that do not necessarily need another person to be displayed, and attitudes that are not necessarily meant to purposefully influence conversation partners (Wichmann, 2000). Certainly, both

emotions and attitudes can be expressed for communicative purposes (Fridlund, 1994; Mead, 1934; Parkinson, 2005) and often take an effect on the listener by virtue of their “expressive function” (Bühler, 1934). Yet, their intended goal remains rather underspecified compared to the “specific intentions for specific turns” (cf. Holtgraves, 2008a) proposed by action-theoretic accounts of language, particularly by speech act theory (Austin, 1962; Searle, 1969).

To date, the role of prosody for the non-literal expression and recognition of different intentions still lacks detailed investigation, although several findings from developmental studies and psycholinguistics point to the relevance of extralinguistic vocal cues in intentional communication. For quite some time, studies on intonational development have been focusing on the emergence of illocutionary skills in infants, considering intonation patterns as primitive devices that preverbal infants use to express their communicative intentions (Dore, 1975). For example, 7- to 11-month-old babies were found to vocally distinguish between communicative and investigative (Papaeliou, Minadakis, & Cavouras, 2002) or emotional functions when babbling (Papaeliou & Trevarthen, 2006). This competence was proposed to regulate cooperative interactions with their parents as a prerequisite for language acquisition. Furthermore, infants’ intonations of babble at the end of their first year (Esteve-Gibert & Prieto, 2013), or words in their second year of life (e.g., Furrow, Podrouzek, & Moore, 1990; Marcos, 1987; Prieto, Estrella, Thorson, & Vanrell, 2012) were found to differ between simple speech acts such as complaining, requesting, or greeting. These combined findings were taken as evidence for a prosodic choice that prelinguistic infants make to communicate their intentions (illocutions) while their propositional (locutionary) abilities are still limited. One challenge that these studies have to face, though, is their dependency on adult, post hoc interpretations of infants’ vocal actions that are usually based on the context in which the vocalizations were produced. This bears the risk that raters—although experts (e.g., mothers or phoneticians)—might overestimate or misinterpret the children’s (true) motives or draw conclusions from cues other than prosody. Studies with adult speakers who can report on their intentions are necessary to corroborate the link between prosody and communicative intentions, and to show its persistence in adulthood.

The present study aimed to fill this gap by conceptualizing a speaker’s intention in terms of speech acts (Austin, 1962; Holtgraves, 2002; Searle, 1969) and investigating the role of prosody in decoding illocutionary force. Note that it was not our goal to describe the prosody of a complete set of speech acts or to investigate the reality of speech act theory. Rather, we aimed to demonstrate that—in identical utterances pronounced according to a limited set of intentions—speakers produce well-identifiable characteristic prosodic patterns, and that these patterns can be reliably recognized by listeners. This adds to the debate whether prosody can convey meaning on its own, i.e., may be conventionalized for different communicative concepts. Alternative views regard prosody as a contrastive marker that does not carry meaning by itself but signals the presence

of “unspoken” meaning by deviating from normal prosody, and hence, motivates listeners to infer the implied message by taking context information into account (Cutler & Isard, 1980; Levinson, 2013). Here, we tested the hypothesis that prosodic patterns as such can be sufficiently distinct, to a degree that listeners can recognize the broad communicative concept and intention in the prosodic speech signal. Therefore, our stimulus set comprised single words and non-words, i.e., tokens free of context and lexical meaning, that were pronounced with six different intonations representing the speech acts criticism, doubt, naming, suggestion, warning, and wish. In three experiments, we combined acoustic analyses of these speech signals with perceptual judgments of listeners (for a similar approach, see Banse & Scherer, 1996; Sauter, Eisner, Calder, et al., 2010). If prosody itself codes speakers’ intentions, different speakers should employ similar cue configurations when conveying the same intention, and participants should be able to recognize the intention without contextual information (i.e., in single words) and irrespective of whether the speech sound carries lexical meaning or not (i.e., in words and non-words).

One important consideration for our investigations of communicative intentions in prosody is the relation to emotional components in the speaker’s tone of voice. Although we advocated a conceptual differentiation of intentions and emotions above, we have to keep in mind that emotions (e.g., fear) might drive intentions (e.g., to warn the interlocutor). Hence, both may be intertwined in the production and perception of communicative utterances. In an attempt to show that the comprehension of intentions is more than the recognition of emotions or affect in the prosodic signal, we further assessed the valence and arousal of our speech stimuli according to dimensional models of affect (Remington, Fabrigar, & Visser, 2000; Wundt, 1896) (we will use the term emotion throughout the text to refer to these affect measures). These values were then used to correct the perceptual recognition of intentions for the contribution of emotion (see below).

The present study took three steps: We started with analyses of the acoustics of speakers’ vocal expressions of speech acts by means of discriminant analyses (Experiment 1). If speech acts as carriers of intentions are coded in characteristic prosodies (i.e. show some consistency of the prosodic pattern across speakers and across tokens within speakers), it should be possible to classify the different categories of speech acts based on their acoustic features alone, in words and non-words alike. Second, we tested whether listeners are able to identify the correct intention based on the prosodic pattern alone (Experiment 2) in a 6-alternative forced choice (6-AFC) categorization task and ratings of the stimuli on every speech act scale (e.g., “How much does it sound like criticism?”). If prosody conveys meaning in a partly conventionalized way, listeners should be able to classify the intentions despite lack of context (i.e., in single words) and irrespective of lexical meaning (i.e., similarly in words and non-words). Finally, we determined which acoustic parameters contribute most to the perception of the respective intention (Experiment 3). Therefore, we fed the acoustic parameters into

multiple regression analyses to predict the participants' ratings on each speech act scale. Furthermore, to control for a possible influence of emotion on intention recognition, the regression analyses were repeated once after valence and arousal ratings of the stimuli had been regressed out.

In summary, the present study sought to demonstrate that prosody carries information about the speaker's communicative intention by (i) identifying characteristic prosodic feature configurations of a set of speech acts that are (ii) reliably recognized by listeners, (iii) despite the lack of context information (single words) and semantic content (non-words) and the control for emotional processing of the stimuli.

Experiment 1 – acoustics

The goal of Experiment 1 was to investigate whether speakers use characteristic acoustic features to convey their intentions. If so, it should be possible to classify the speech stimuli into the corresponding speech act categories based on their acoustic features alone and irrespective of word meaning. Specifically, we focused on duration, intensity, pitch, and spectral features that have been analyzed in similar approaches in emotion research (e.g. Banse & Scherer, 1996; Blanc & Dominey, 2003; Sauter, Eisner, Calder, et al., 2010). In such studies, pitch cues were predominant when emotions were expressed verbally, compared to a stronger weighting of spectral features in non-verbal utterances, making it likely that pitch cues will play a major role in the present experiment.

For the current study, four speakers produced single-word stimuli with varying prosodies to express six different intentions, i.e., the speech acts criticism, doubt, naming, suggestion, warning, and wish. To obtain stimuli that are representative for typical language use, all speakers were non-actors, i.e., they relied on their intuition—not training in acting—to express the intention in a way that could be understood by an imaginary interlocutor. For high stimulus quality, all speakers were, however, familiar with sound recordings, i.e., working as voice coaches or speech scientists. This choice of professional speakers with only minimal training in acting is an attempt to face the criticism, first raised in emotion research, that actors' prosodic patterns may deviate from those used in everyday conversations (Jürgens, Hammerschmidt, & Fischer, 2011; see also General discussion). Apart from that, it should be mentioned that intentions are typically expressed more voluntarily than emotions and are, hence, less dependent on the spontaneity of the utterance. Altogether, the present stimuli were recorded such to grant generalizability of the results to natural language use.

Materials and methods

Ethics approval

The ethics committee of the University of Leipzig, Germany approved the present and all following experiments in this study.

Stimulus recordings

Four trained native German speakers (voice coaches, 2 female) were invited to record the German words “Bier” (beer) and “Bar” (bar) as well as the non-words “Diem” and “Dahm” (for examples, see Appendix D: [Supplementary material](#)). These (non-)words were intoned to express six different communicative intentions or speech acts: criticism, doubt, naming, suggestion, warning, and wish. The chosen speech acts were plausible for our stimulus words “beer” and “bar” and fit into the broader speech act categories as defined by Searle and Vanderveken (1985). To elicit the respective intentions in the speakers, they read short scenarios that described a situation in which they interacted with an interlocutor (see Appendix A). They were allowed to utter an initial sentence and to vocalize freely until they felt ready to articulate the intention shortened to the single essential word. This recording approach, instead of using natural speech recordings, was chosen to obtain clear portrayals of intentions in good sound quality. Recordings were conducted in a soundproof room with the microphone (Rode NT55) approximately 20 cm in front of the speaker and digitized at a 44.1 kHz sampling rate in a 16-bit mono format. The words and non-words were repeated several times to obtain eight variants per stimulus in good quality. The resulting stimulus set, thus, comprised 768 stimuli, with eight repetitions of four (non-) words expressed as six speech acts by four speakers.

Acoustic features

For investigating acoustic features of the speech acts, we obtained seven acoustic measures that are commonly used in experiments on human voice and speech stimuli (e.g., Banse & Scherer, 1996; Sauter, Eisner, Calder, et al., 2010). Using Praat software (Boersma & Weenink, 2014) we extracted the number of voiced frames as a measure of stimulus duration, mean intensity, harmonics-to-noise ratio (HNR), mean fundamental frequency (f0) as well as pitch rise, measured as the difference between offset and onset f0. Furthermore, we extracted the spectral center of gravity and the standard deviation of the spectrum. The mean acoustic characteristics as measured with Praat are presented in [Table C1](#) (Appendix C). Statistical analyses showed that speakers had used very similar acoustic cues to intone speech acts in words and non-words. *T*-tests for paired samples comparing the acoustics of words and non-words for each speech act category were largely non-significant. Only exception were HNR and spectral center of gravity that showed differences in some, but not all speech act categories (see [Table C2](#) in Appendix C for more details). These differences are, however, likely to be caused by the different consonants (“r” in words vs. “m” in non-words) rather than by differences in prosody.

Discriminant analyses

Discriminant analyses were performed for words and non-words separately, with the seven acoustic features as independent variables and the speech act category (criticism, doubt, etc.) as dependent variable. These analyses sought to identify linear functions of acoustic feature combinations that maximize differences between speech act categories. In other words, these analyses tested whether

Table 1

Results of cross-validated (jackknife) discriminant analysis for classification of speech acts from the acoustic features (in %). Correct classifications are shown in bold.

| Stimulus type | Classification | | | | | |
|------------------|----------------|-------------|------------|-------------|------------|-------------|
| | Criticism | Doubt | Naming | Suggestion | Warning | Wish |
| <i>Words</i> | | | | | | |
| Criticism | 76.6 | 4.7 | 0 | 10.9 | 7.8 | 0 |
| Doubt | 1.6 | 92.2 | 0 | 1.6 | 0 | 4.7 |
| Naming | 0 | 0 | 100 | 0 | 0 | 0 |
| Suggestion | 3.1 | 1.6 | 0 | 95.3 | 0 | 0 |
| Warning | 0 | 0 | 0 | 0 | 100 | 0 |
| Wish | 0 | 0 | 9.4 | 0 | 0 | 90.6 |
| <i>Non-words</i> | | | | | | |
| Criticism | 79.7 | 6.2 | 0 | 7.8 | 4.7 | 1.6 |
| Doubt | 7.8 | 84.4 | 3.1 | 0 | 0 | 4.7 |
| Naming | 0 | 0 | 100 | 0 | 0 | 0 |
| Suggestion | 1.6 | 0 | 0 | 98.4 | 0 | 0 |
| Warning | 0 | 0 | 0 | 0 | 100 | 0 |
| Wish | 0 | 0 | 4.7 | 0 | 0 | 95.3 |

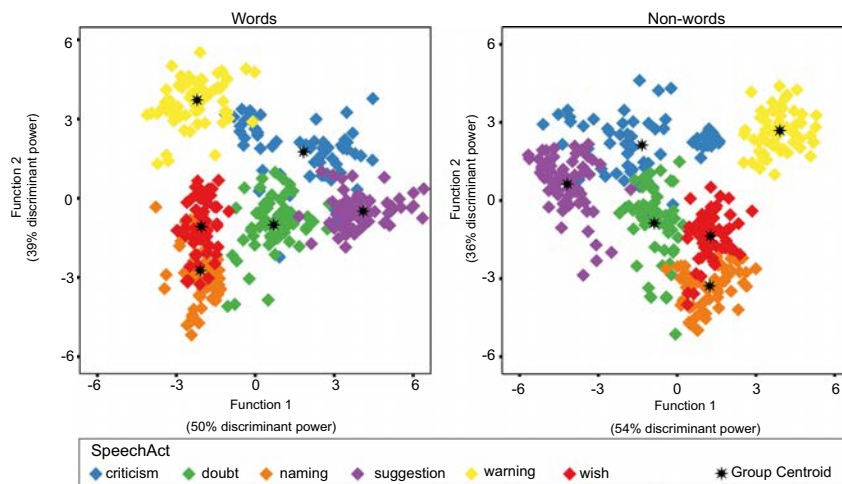


Fig. 1. Results of the discriminant analyses plotted along the first and second discriminant function. Data points correspond to stimuli of the six speech act categories.

the acoustic features alone have sufficient discriminant power to reliably group the stimuli that express the same intention. The discriminant analyses for both words and non-words were cross-validated with a jack-knife procedure, and the distribution of the results was validated with chi-square tests.

Results

The discriminant analyses classified the correct speech act category for 92% of all word stimuli and 93% of all non-word stimuli. These results are highly above chance-level (17%) as was tested with chi-square tests: $\chi^2(35) = 1717.6$, $p < .001$ for words and $\chi^2(35) = 1702.6$, $p < .001$ for non-words. Classification results of the discriminant analyses for the different types of speech acts are demonstrated as

confusion matrices in Table 1. The highest results were found for naming and warning (both 100% correct classification for words and non-words), while the lowest results were obtained for criticism in words (76.6% correct classification), and non-words (79.7%). Additional chi-square tests showed that the discriminant model classified our stimuli better than chance (chance-level: 17%) for every type of speech act ($\chi^2(5) > 151$, p 's $< .001$). Fig. 1 shows the classification of the different speech acts by the first two discriminant functions. The first function (x-axes) explained 49.6% of the variance for words and 54.2% for non-words and was mainly based on the acoustic measure of pitch rise (offset-onset f0). The second discriminant function (y-axes) had an additional discriminant power of 38.6% for words and 36.4% for non-words and was most related to the mean intensity and mean f0 of the stimulus (see Table C3,

Appendix C). Additionally, a third function (not depicted in Fig. 1 for reasons of clarity) explained 10.5% of variance for words and 7.7% for non-words and showed highest correlation with the duration of the stimuli. The last two discriminant functions from our analyses explained only minor effects (function 4: 0.9% for words and 1.2% for non-words, function 5: 0.4% for words and 0.5% for non-words) and were neglected from further investigations.

Discussion

The acoustic features of our stimulus set could be used to accurately classify the correct speech act, for words and non-words alike. This demonstrates the distinctiveness of the prosodic patterns that speakers deliberately applied to code their intentions in the tested speech act categories and the relative independence of prosody from lexical content. Furthermore, the high accuracy of the classification implies a reasonable consistency of the relevant prosodic cues across speakers and utterances, and may point to the existence of feature configurations that speakers consider conventional and appropriate for different communicative goals. For example, the warning stimuli were loudest and had the most arched pitch contour with a salient peak in the middle of the word as is appropriate for the urgent nature of a warning. In comparison, the naming stimuli showed the least salient acoustic features with low mean pitch, flat pitch contour, low intensity and little spectral variation in line with the neutral character of the expression. As expected, pitch rise and mean f_0 , together with mean intensity were the most influential acoustic features in these analyses, while spectral features had only weak discriminant power. In sum, the data show that speakers can use prosody as a channel of communication to convey their intentions. Note that we do not expect that speakers possess different prosodic patterns for all possible intentions or speech acts. Yet, we believe that speakers choose salient, distinguishable and probably culturally learned prosodic signatures to trigger cognitive processes in the addressee to infer the communicative intent of the speaker beyond the overt lexical meaning.

Experiment 2 – behavior

After finding consistent acoustic differences between prosodic speech act expressions, we were interested in participants' perception of the stimuli. We investigated whether participants would be able to identify the different intentions based on the prosodic information in a 6-alternative forced-choice (6-AFC) categorization task. Participants, further, judged the valence and arousal of every stimulus (Remington et al., 2000; Russell, 1980; Wundt, 1896) in the second half of the experiment, which allowed us to assess in how far speech acts may be classified based on their emotional tone.

Materials and methods

Participants

Ten participants were presented with the word stimuli (6 female, mean age \pm SD: 24.6 ± 4.9), ten other volunteers

performed the task with the non-word stimuli (4 females, mean age \pm SD: 24.9 ± 2.6). We tested separate groups of participants for words and non-words to avoid transfer of the semantic meaning to the non-word stimuli. All participants reported normal hearing ability, gave written informed consent and were paid 7€ per hour for their participation.

Design and procedure

In the first half of the experimental session, participants were asked to assign each stimulus to one of the six possible speech act categories (criticism, doubt, naming, suggestion, warning, or wish). After having read short definitions for the different speech acts (Appendix B), they heard each sound stimulus once via headphones and were instructed to press the keys 1–6 on a keyboard. The speech act labels with corresponding numbers were displayed on a computer screen throughout the experiment. No feedback for the correctness of the response and no time limits were given. The experiment was separated into four blocks—one for each speaker. Block order and stimulus order within each block were pseudo-randomized by preallocating the speech acts with balanced probabilities. Chi-square tests were performed to test for above-chance classification across all speech acts and within single speech act categories.

In the second half of the experimental session, participants were asked to evaluate the valence and arousal of each stimulus. Therefore, they listened to the same stimuli again, in the same order as before. After each sound, they saw two visual analogue scales on the screen, first for valence (positive/negative), then for arousal (calm/excited), and placed their ratings with a continuous slider. The scales showed the outermost pictures of the Self-Assessment Manikin (Bradley & Lang, 1994) at the margins. No time constraint was given for the answers. Friedman tests were calculated to examine differences in the affect ratings among the speech act categories.

Results

Speech act categorization

In the 6-AFC task, participants were able to identify the correct speech act category of our stimuli with high accuracy for words (mean \pm SD: $82 \pm 13\%$) and for non-words ($73 \pm 17\%$)—with no significant difference between the participant groups for words and non-words ($t(18) = 1.26$, $p = .22$). Chi-square tests showed that participants' classification of every speech act category was better than being predicted by chance (chance level: 17%), both for words ($\chi^2(5) > 1082$, p 's $< .001$) and non-words ($\chi^2(5) > 798$, p 's $< .001$).

Confusion matrices for words and non-words are presented in Table 2. As with the acoustic analyses, the identification of criticism was lowest among the six speech act categories. For words as well as non-words, participants misclassified criticism most often as being doubt, and to a lower extent as warning. Furthermore, common confusions of the non-word stimuli were found for suggestion taken as doubt, wish, or criticism. To some extent,

Table 2
Behavioral categorization of speech acts (in %). Correct categorizations are shown in bold.

| Stimulus type | Participants' responses | | | | | |
|------------------|-------------------------|-------------|-------------|-------------|-------------|-------------|
| | Criticism | Doubt | Naming | Suggestion | Warning | Wish |
| <i>Words</i> | | | | | | |
| Criticism | 62.0 | 24.0 | 0.2 | 4.9 | 6.9 | 2.0 |
| Doubt | 5.3 | 83.4 | 3.1 | 4.4 | 0.6 | 3.3 |
| Naming | 1.7 | 0.3 | 90.0 | 1.3 | 0.8 | 5.9 |
| Suggestion | 4.5 | 9.2 | 3.3 | 80.3 | 0.5 | 2.2 |
| Warning | 4.1 | 0 | 0.3 | 0.5 | 89.5 | 5.6 |
| Wish | 1.9 | 0.8 | 8.3 | 4.4 | 1.1 | 83.6 |
| <i>Non-words</i> | | | | | | |
| Criticism | 52.4 | 29.7 | 1.1 | 3.9 | 9.4 | 3.4 |
| Doubt | 2.7 | 82.6 | 2.2 | 8.5 | 0 | 4.1 |
| Naming | 17.7 | 1.9 | 74.1 | 2.3 | 0.3 | 3.8 |
| Suggestion | 9.7 | 12.8 | 3.3 | 63.7 | 0.5 | 10.0 |
| Warning | 3.8 | 0.5 | 0.0 | 0.2 | 94.2 | 1.4 |
| Wish | 5.3 | 1.6 | 8.9 | 14.2 | 0.6 | 69.4 |

participants also misclassified wish as suggestion, and naming as criticism.

Emotion ratings

For the perception of emotion, mean ratings for valence and arousal differed significantly between the speech act categories, for words (valence: $\chi^2(5) = 35$, arousal: $\chi^2(5) = 43$, p 's < .001) and non-words (valence: $\chi^2(5) = 44$, arousal: $\chi^2(5) = 45$, p 's < .001). The results were very similar for words and non-words in each speech act category (Fig. 2). On the valence scale, the speech acts warning and criticism were perceived most negatively, whereas wish and suggestion were associated with a more positive valence. Doubt and naming were rated neutrally with regard to valence. The perception of the speakers' arousal was very calm for naming, wish, and doubt, and very excited for warning and criticism. Suggestion stimuli were rated in the middle range for arousal.

Discussion

Participants were well able to identify the speaker's communicative intention from the prosody alone as indicated by the highly significant results in the 6-AFC categorization task. Importantly, participants were able to make use of the prosodic signal with minimal context descriptions and without lexical content (see below). These data show that prosody is a powerful communicative channel that is used by listeners to decode the "unspoken" meaning and intention of the speaker and that may determine their respective conversational reaction. Interestingly, criticism was identified least reliably and was specifically confused with doubt, in line with the similar acoustic features of these two speech acts (Fig. 1, Table 1). It is well conceivable that the acoustic similarity of criticism and doubt may amount from their conceptual similarity—a rather depreciative stance toward an inner or outer event—a fact that

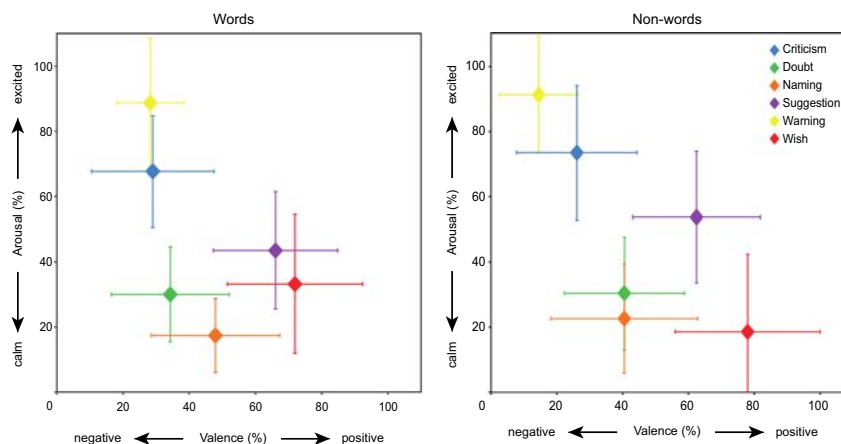


Fig. 2. Emotion ratings. Average scores of the valence and arousal ratings for each speech act category.

would further illustrate the intricate link between communicative intentions and prosody.

The valence and arousal ratings revealed distinctive affective properties of the different speech act categories, which were consistent across words and non-words (Fig. 2). This suggests that speaker's intentions may have emotional connotations that listeners are able to detect in the prosodic signal. A natural question that comes up is, how strongly the identification of intentions in prosody depends on emotion recognition and whether intentions can be recognized without taking emotion into account. We addressed this issue in Experiment 3.

Experiment 3 – behavior and acoustics

Experiment 1 and 2 demonstrated that prosodically coded intentions can be differentiated (i) physically based on characteristic acoustic feature configurations as well as (ii) perceptually in a 6-AFC task. What remains to be shown is, in how far the acoustic differences account for participants' ability to identify the speaker's intention. If participants use the prosody's acoustic information for intention understanding, it should be possible to predict listeners' perception from the acoustic measures and, further, to identify the different feature combinations that evoke specific speech act impressions. We addressed this question by feeding acoustic measures and typicality ratings for every speech act into a multiple regression analysis. Moreover, to assess the influence of emotion perception on intention recognition (see Experiment 2), we conducted an additional regression analysis in which valence and arousal ratings were regressed out.

Materials and methods

Participants

A new group of 20 healthy volunteers (10 females, mean age \pm SD: 24.8 ± 4.1 years) for the words and 20 participants for the non-words (10 females, mean age \pm SD: 24.6 ± 3.2 years) took part in a rating study. All participants reported normal hearing ability, gave written informed consent and were paid 7€ per hour for their participation.

Design and procedure

In this experiment, participants were asked to indicate to what extent each stimulus sounded like a given speech act category (criticism, doubt, naming, suggestion, warning, or wish). Compared to the 6-AFC categorization task, such speech act ratings provide a more refined and less strategy dependent measure for the participants' perception and allowed for the application of multiple regression analyses. In total, each stimulus was presented six times, once for every speech act scale, in separate blocks. Responses were given with a slider on a visual analogue scale from 0 to 100 ('intonation does not fit the intention at all' to 'intonation fits the intention very well'). Each block comprised the same 192 stimuli—four tokens of two (non-)words expressed as six speech acts by four speakers—that were chosen from the full stimulus pool of 384 stimuli. Stimuli and block order were again pseudo-

randomized. The timing of the experiment was self-paced and participants were able to take breaks between blocks. The results of the ratings were analyzed by repeated-measures ANOVAs with Greenhouse–Geisser correction for the factor SPEECH ACT for every speech act scale separately.

Multiple regressions

To elucidate which acoustic features guided the participants' ratings on the speech act scales, we performed linear multiple regression analyses. Specifically, we used the acoustic features as predictors (independent variable) for the subjective ratings of the 192 stimuli (dependent variable), separately for words and non-words. Acoustic features were the same as in Experiment 1 and were chosen such to include measures of duration, intensity, pitch, and spectrum while keeping multicollinearity low (variance inflation factor words: <3.526 , non-words: <4.561).

Furthermore, to demonstrate that intention perception is not merely determined by perceived emotional connotations in the speech signal, the emotion perception of the stimuli was regressed out in two steps: Firstly, separate regressions were calculated with the valence and arousal ratings as independent variables and the single speech act ratings as dependent variables. This way, we bound all the variance in the perceived speech act that could be explained by potentially perceived emotions. Thus, the residuals of these regressions should contain information about the participants' intention perception devoid of the perceived valence and arousal. Following this, new regressions were performed, now with the acoustic features as independent and the standardized residuals of the speech act ratings as dependent variables.

Results

Multiple regressions

The mean ratings of the stimuli according to the six different speech act scales are shown in Table 3. As can be seen in the diagonal, the highest ratings were obtained for the correct speech act category. This was confirmed by a significant main effect of SPEECH ACT in repeated-measures ANOVAs performed for every speech act scale separately (word stimuli: $F_s > 46.289$, $p_s < .001$; non-word stimuli: $F_s > 29.326$, $p_s < .001$). Post-hoc paired comparisons with Bonferroni correction showed that speech act stimuli were rated significantly higher on their corresponding scale than any other speech act category with $p_s < .03$. Altogether, the ratings replicate the findings in the 6-AFC categorization task, in that also this new group of participants was well able to recognize and evaluate the speech acts correctly.

To examine whether specific patterns of acoustic features can predict subjective evaluation of the different speech act stimuli, the ratings together with the acoustic measures for the single stimuli were entered into multiple regression analyses, separately for each speech act rating scale. These regressions yielded highly significant results on all scales (see Table 4 for detailed results). The variance explained by the regression models ranged from 11.6% for the wish ratings to 52.7% for the warning ratings of the

Table 3
Participants' ratings of speech acts (min = 0, max = 100). Ratings on corresponding speech act scale are shown in bold.

| Stimulus type | Speech act scale | | | | | |
|------------------|------------------|-------------|-------------|-------------|-------------|-------------|
| | Criticism | Doubt | Naming | Suggestion | Warning | Wish |
| <i>Words</i> | | | | | | |
| Criticism | 70.0 | 56.2 | 9.4 | 18.5 | 27.5 | 16.5 |
| Doubt | 45.7 | 82.5 | 13.7 | 16.7 | 10.7 | 9.5 |
| Naming | 12.2 | 13.2 | 85.3 | 10.5 | 9.3 | 17.4 |
| Suggestion | 15.7 | 28.0 | 17.7 | 77.3 | 7.7 | 26.9 |
| Warning | 22.8 | 18.6 | 8.2 | 16.0 | 86.0 | 29.9 |
| Wish | 6.5 | 10.1 | 21.3 | 16.3 | 6.0 | 80.4 |
| <i>Non-words</i> | | | | | | |
| Criticism | 63.3 | 45.5 | 11.3 | 21.3 | 31.3 | 13.3 |
| Doubt | 23.0 | 77.5 | 17.2 | 22.7 | 6.1 | 20.6 |
| Naming | 14.9 | 13.1 | 78.0 | 17.7 | 8.7 | 21.7 |
| Suggestion | 20.5 | 35.3 | 24.9 | 70.6 | 10.9 | 18.3 |
| Warning | 26.0 | 5.4 | 7.5 | 8.5 | 94.1 | 9.1 |
| Wish | 10.1 | 11.4 | 25.7 | 22.8 | 5.4 | 75.9 |

Table 4
Multiple regression analyses of acoustic features and speech act ratings (beta-weights).

| Acoustic parameter | Speech act ratings | | | | | |
|-------------------------------|--------------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| | Criticism | Doubt | Naming | Suggestion | Warning | Wish |
| <i>Words</i> | | | | | | |
| Voiced frames | 0.276*** | 0.348*** | −0.420*** | −0.227*** | 0.073*** | 0.083*** |
| Mean f0 | 0.371*** | 0.402*** | −0.326*** | −0.211*** | 0.385*** | −0.294*** |
| Offset-onset f0 | 0.224*** | 0.315*** | −0.277*** | 0.524*** | −0.387*** | −0.135*** |
| Mean intensity | −0.015 | −0.250*** | −0.266*** | 0.368*** | 0.343*** | 0.118*** |
| Mean HNR | −0.171*** | −0.106*** | 0.141*** | −0.008 | −0.127*** | 0.044 |
| Center of gravity | −0.083 | −0.106*** | 0.097*** | −0.059*** | −0.039*** | 0.042 |
| SD spectrum | −0.085*** | −0.082*** | 0.014 | 0.035*** | −0.107*** | 0.149*** |
| Adj R ² | 0.171*** | 0.273*** | 0.364*** | 0.349*** | 0.527*** | 0.116*** |
| Adj R ² (emo-corr) | 0.153*** | 0.285*** | 0.228*** | 0.305*** | 0.162*** | 0.175*** |
| <i>Non-words</i> | | | | | | |
| Voiced frames | 0.183*** | 0.237*** | −0.388*** | −0.184*** | 0.056*** | 0.141*** |
| Mean f0 | 0.121*** | 0.331*** | −0.376*** | −0.258*** | 0.289*** | −0.204*** |
| Offset-onset f0 | 0.259*** | 0.416*** | −0.206*** | 0.407*** | −0.387*** | −0.062*** |
| Mean intensity | 0.168*** | −0.315*** | −0.291*** | 0.250*** | 0.479*** | −0.100*** |
| Mean HNR | −0.052* | 0.033*** | 0.249*** | −0.093*** | −0.135*** | −0.056* |
| Center of gravity | 0.052* | −0.074*** | 0.199*** | −0.182*** | 0.015 | −0.077*** |
| SD spectrum | 0.105*** | 0.034*** | −0.033*** | −0.027*** | −0.145*** | 0.175*** |
| Adj R ² | 0.135*** | 0.269*** | 0.302*** | 0.260*** | 0.621*** | 0.176*** |
| Adj R ² (emo-corr) | 0.092*** | 0.305*** | 0.132*** | 0.159*** | 0.270*** | 0.142*** |

Beta weights and adjusted R^2 are depicted for multiple regressions using acoustic features as predictors and speech act ratings as dependent variables. Additionally, adjusted R^2 values are depicted after controlling for emotion perception (emo-corr). f0 = fundamental frequency; HNR = harmonics-to-noise ratio; SD = standard deviation; Adj = adjusted; emo-corr = overall performance of the multiple regressions after affective ratings had been regressed out.

* $p < .05$.

** $p < .01$.

*** $p < .001$.

words and from 13.5% for the criticism ratings to 62.1% for warning ratings of the non-words. The beta weights of the regression functions indicate the degree to which the acoustic parameters predicted the ratings. Put differently, high absolute values of the beta weights reflect the importance of the corresponding acoustic feature for the prediction of the regression model. Almost all acoustic features contributed significantly to the predictions of the speech act ratings (see Table 4). While the spectral features (center of gravity and standard deviation of the spectrum) as well as the HNR yielded very low beta values in general (all <0.2 , except for HNR in naming ratings in the non-words), the acoustic measures of pitch, amplitude, and

duration reached absolute beta values of up to 0.524, suggesting that these parameters are key features for the comprehension of the intentions.

Fig. 3 shows the beta weights of the main acoustic features and reveals specific patterns of acoustic parameters for the prediction of the different speech act ratings: While high ratings for criticism and doubt were mainly predicted for long stimuli with high mean pitch and a rising pitch contour (positive beta weights for voiced frames, mean f0 and offset-onset f0), high ratings for naming were associated with short and soft stimuli with low mean pitch and falling pitch contour (negative beta weights for these measures). Suggestion ratings relied on short stimuli with low

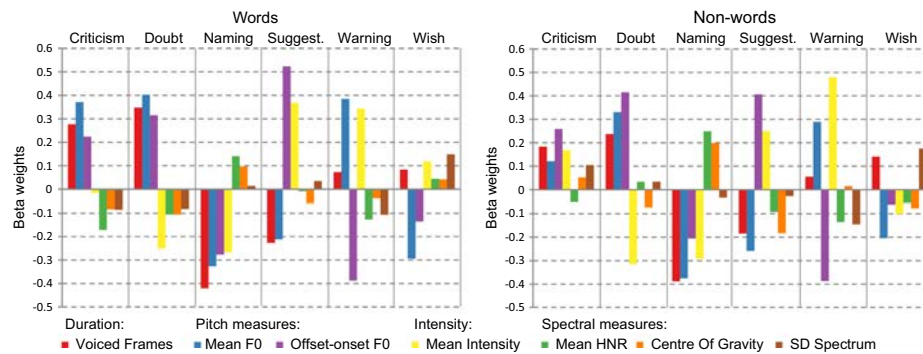


Fig. 3. Results of the multiple regressions for each speech act scale (columns). The bars represent the beta-weights for the seven acoustic features indicating how strongly they predicted the speech act rating.

mean pitch, but a strong pitch rise and high intensity. High ratings for warning were predicted, if the stimuli had a high mean pitch and intensity with a negative f0 offset-onset relation. Beta values in predictions for wish ratings showed the least clear pattern of acoustic information. For these stimuli, a low mean pitch was the most informative parameter. In total, the judgment of different speech act categories seems to be based on distinct acoustic patterns. Crucially, predictions of criticism and doubt ratings were based on similar patterns that only differed in the contribution of intensity. This is in line with the confusion of criticism and doubt in the behavioral categorization and ratings (Tables 2 and 3). Overall, the pitch-related features (mean f0 and offset-onset f0) had a high influence on the ratings in all speech act categories, qualifying them as the most important acoustic features in these analyses. A subset of speech acts was further influenced by amplitude and duration features, while spectral parameters only seemed to play a minor role. Notably, the results of the multiple regressions were very similar for the words and non-words, even though the analyzed data were not only based on a different stimulus set, but also on the ratings of two independent groups of participants. Therefore, these results validate the distinct acoustic patterns that shape the perception of different speech acts and intentions in single-word utterances.

Multiple regressions controlled for emotion

The second regression approach was performed to control for all variance that could be explained by the perception of valence and arousal. After an initial regression to predict speech act ratings as dependent variables from emotion ratings as independent variables, we conducted a second regression to explain residual speech act information by the acoustic measures. These regressions were still highly significant on all speech act scales (p 's < .001, Table 4) and explained variance in the range from 15.3% for criticism ratings to 30.5% for suggestion ratings of the words, and from 9.2% for criticism ratings to 30.5% for doubt ratings of the non-words. Compared to the original regressions, there was a noticeable decrease in explained

variance for naming and warning in words and non-words, which indicates that some of the variance could be explained by the emotion perception of these speech acts. Prediction of the other speech acts was virtually unchanged.

Discussion

The current experiment confirmed a link between the acoustics and perception of the speech act stimuli by multiple regression analyses in which distinct acoustic feature configurations significantly (but not fully) predicted the listeners' perception of the speech acts. The amount of variance explained (ranging between 12% and 62% depending on speech act type) was overall comparable to estimates found in previous studies on emotional prosody (Banse & Scherer, 1996; Sauter, Eisner, Calder, et al., 2010), validating our approach. In general, pitch features (mean pitch and offset-onset f0) were most influential for the perception of different intentions. Further important cues could be derived from intensity and duration measures, whereas spectral features contributed least to the intention predictions.

The amount of variance explained by the regressions was significant, but the values for some speech acts (e.g. 12% for wish in words or 13% for criticism in non-words) suggest that the acoustic features chosen for the analyses are not the only basis for intention perception. The inclusion of additional acoustic features might further increase the precision of the regression models. On the other hand, higher cognitive processes, such as social inference, may contribute to the recognition of the communicative intention (see below; Wichmann, 2000, 2002; Szameitat et al., 2010). Still, the fact that different acoustic patterns can explain the perception of different speech acts, generally leads to the assertion that prosody carries information about the speaker's intended meaning.

Emotions

Importantly, perception of intentions was not solely based on recognition of the speaker's emotion as shown

by the additional regression analyses, taking valence and arousal of stimuli into account. As mentioned in the introduction, we do not exclude that communicative intentions are partly based on the emotional state of the speaker. Indeed, listeners could classify the stimuli in terms of valence and arousal (Fig. 2; Experiment 2). Nevertheless, regression analyses still explained a significant amount of variance of the speech act ratings and most speech act predictions were virtually unchanged after these affective components had been regressed out (Table 4). Only warning and naming showed a considerable decrease in the prediction rate which might be explained by their extreme positions on the arousal scale (Fig. 2). On the other hand, participants might have first identified the speaker's intention and then assigned the corresponding valence and arousal because they were asked to do so in the experiment (Experiment 2). Overall, although emotional connotations may be important for the recognition of some speech acts, our results give no reason to assume a systematic influence of emotions on the recognition of communicative intentions.

Ratings vs. 6-AFC

Finally, it is of note that the ratings replicated the results of the 6-AFC task used in Experiment 2. Importantly, ratings are a more sensitive measure than forced-choice categorization tasks because they not only allow participants to reject predefined response categories but also to flexibly adjust their responses on every (visual-analog) speech act scale. The fact that both typicality ratings in Experiment 3 and 6-AFC judgments in Experiment 2 (conducted in separate participant groups) yielded very similar results demonstrates the robustness of our findings.

General discussion

Action-theoretic views of language (Austin, 1962; Bühler, 1934; Grice, 1957; Searle, 1969) propose that speakers' intentions are the main core and driver of interpersonal communication. Yet, speakers rarely express their intentions literally in the propositional content of an utterance, raising the question of how the speaker's meaning is transmitted from sender to receiver. Here, we conceptualized intentions in terms of speech acts and provide evidence that prosody serves as an extralinguistic channel to convey intentions non-verbally. Acoustically, speakers used distinct prosodic feature configurations for different speech acts. Behaviorally, listeners were well able to differentiate these intentions from voice tone alone, even when no semantic meaning (non-words) or situational context was available (single words). Further, a direct link between acoustics and perception was demonstrated, in that acoustic features reliably (although not fully) accounted for the listeners' perception of the stimulus—even when the emotional connotation of the stimuli was controlled for.

Notably, our results were consistent across all three experiments. For example, in all measures from acoustics to perception, warning was classified with highest and criticism with lowest accuracy. Moreover, in both the

stimulus-based discriminant analyses (Experiment 1) and the multiple regressions (Experiment 3), pitch rise, mean f_0 , as well as mean intensity and duration were the most and spectral features the least relevant cues for correct speech act recognition. This is consistent with a special role of pitch features observed in similar studies on verbal emotion (Banse & Scherer, 1996) and attitude recognition (Blanc & Dominey, 2003). Overall, the consistency of our results across experiments and participant groups lends strong support for the relevance of prosody in conveying communicative intentions.

Conventional prosodic expressions

Our results invite the assumption that speakers' intents are expressed in conventionalized prosodic forms. This view is supported (i) by the consistency of the prosodic patterns across four independent speakers for each of the six speech acts, and (ii) by the robustness of listeners' performance in identifying the expressed intentions, despite absence of contextual or semantic information. Arguably, prosodic patterns do not refer to communicative intentions as unambiguously as words refer to objects in the world. Rather we propose that they represent "communicative complexes" that connote a set of conceptually related pragmatic categories (e.g., speech acts), whose distributions of relevant acoustic cues partly overlap. This acoustic and conceptual overlap may account for the confusion of criticism and doubt in our experiments and predicts a rather loose labeling of speakers' intentions in open choice tasks, licensing our use of forced-choice task and typicality ratings (see below). Notably, our data suggest that the acoustic characteristics of these "complexes" are conventionalized to the extent that listeners can infer the relevant communicative concept by matching the perceived prosodic pattern with an internalized probabilistic distribution of acoustic cue configurations for different intentions.

Such a direct recognition of speakers' intent from prosody is reminiscent of previous work on written speech acts (Holtgraves, 2008a) suggesting that the default interpretation of illocutionary force can be based on generalized rather than particularized implicatures, i.e. can be directly understood without contextual information, similar to most idioms (e.g., to call it a day) or metaphors (e.g., He is a walking dictionary) (Glucksberg, 2003; Glucksberg, Gildea, & Bookin, 1982; Keysar, 1989). The relevance of context for the classification of speakers' prosodic intentions or attitudes has been a matter of debate for a while (Cutler, 1976; Wichmann, 2000, 2002). Some accounts posit that prosody mainly acts in a contrastive way, without conveying meaning by itself (e.g. Attardo, Eisterhold, Hay, & Poggi, 2003; Bryant & Fox Tree, 2005). By deviating from its "default", prosody is thought to motivate the listener to look for "unspoken" meanings in the utterance, i.e. to infer implicit speech actions from literal meaning by taking context information into account (Cutler & Isard, 1980; Levinson, 2013). However, as Wichmann (2002) rightly pointed out, this view requires knowledge about prosodic "defaults". We argue that this knowledge is best characterized as experience-dependent inventory of situationally distinct acoustic patterns that allows

listeners to recognize broad communicative concepts based on prosody. Such a distinguished role for prosody in intention transmission is supported by the fact that these communicative concepts could be conveyed despite absence of contextual information and without knowledge of the lexical content in non-words (Experiments 2 and 3).

Note that we do not claim that context plays no role at all. Very much like lexical and syntactic processing is not based on acoustics alone but varies with context (for example in case of homophones such as “meet” vs. “meat” or ambiguous word category as in “report”; for review, see Piantadosi, Tily, & Gibson, 2012), also the prosodic recognition of speakers’ intents can be shaped by context (Tanenhaus, Kurumada, & Brown, 2015). First, context predicts what interpretations are likely and may, thus, resolve perceptual ambiguity between overlapping distributions within the “communicative complex”, e.g., allowing listeners to better discriminate between doubt and criticism. Second, context provides a sample of the speaker’s prosodic “style” that allows listeners to flexibly adapt (even reverse) their prosodic interpretations accordingly (Tanenhaus et al., 2015). Altogether, we conclude that (paralinguistic) prosody is a signal that is able to convey a broad communicative concept on its own but becomes cognitively interlinked and specified with complementary contextual information, if available.

Prosody’s initial relevance for social communication

Overall, the transfer of intentions via prosody might be a capability that forms the initial, non-linguistic foundation of interpersonal communication (Bates, Camaioni, & Volterra, 1975; Dore, 1975) that becomes gradually complemented and refined—yet not erased—by growing verbal capacities, over the course of ontogeny and perhaps even phylogeny (Oller & Griebel, 2014). For example, primate calls have been found to signal the producer’s interactive stance intentionally (Schel, Townsend, Machanda, Zuberbühler, & Slocumbe, 2013) via distinct acoustic structures (Crockford & Boesch, 2003; Seyfarth & Cheney, 2014), even if they lack lexical (referential) meaning (Wheeler & Fischer, 2012). Developmentally, young infants start to produce acoustically distinct prosodic patterns in the middle of their first year of life that are initially used in communicative as opposed to self-centered emotional or exploratory contexts (Papaeliou & Trevarthen, 2006; Papaeliou et al., 2002), later express specific “primitive intents” (Esteve-Gibert & Prieto, 2013; Prieto et al., 2012) and endow pointing gestures with communicative goals (Grünloh & Liszkowski, 2015). Notably, interactive prosodic patterns emerge earlier than verbal skills and become meaningful communicative instruments, most likely because parents differentiate their responses based on the acoustics of the child’s vocalizations (cf. Lester et al., 1995; Oller & Griebel, 2014). The present data show that prosody continues to be indicative of speakers’ intents in adulthood, despite mature verbal skills. More than that, the data suggest that the use of prosodic cues evolves further beyond infancy to express more complex intentions than those infants would ever produce (e.g., criticism or doubt). Whether speakers resort more strongly to these

(early) prosodic building blocks of communication when verbal capacities may get lost or are nonexistent as in conditions of non-fluent aphasia (Barrett, Crucian, Raymer, & Heilman, 1999; Warren, Warren, Fox, & Warrington, 2003) or foreign languages is an interesting topic for future research.

Prosody in natural language use

Single-word utterances are part of our everyday life and humans start to use prosody to code for different pragmatic intentions in single words in early infancy (Dore, 1975; Prieto et al., 2012). Yet, compared to longer sentences with additional semantic information, the brevity of the present context-free stimuli may have led speakers to emphasize the relevant prosodic features. Listeners, in turn, may be more used to decode intentions in sentential contexts that often resolve ambiguities (even if ambiguities were mitigated by the 6-AFC task and typicality ratings in the present study). Future studies can help to generalize our results by using a wider set of recordings (as suggested by Banse & Scherer, 1996), for example, including sentence-level stimuli, more variable tokens (i.e. more words/sentences), more speech acts, and more speakers.

Apart from that, another point of discussion is in how far prosodies produced in the sound lab using fictional scenarios correspond to prosodies produced in natural conversations. Although a direct empirical investigation is still pending, there are several reasons that grant the ecological validity of our sound stimuli. First, cues for expressing intentions are typically produced voluntarily during an interaction. Therefore, they have a posed character by nature and may not suffer from artificial recording situations to the extent as emotions do (Jürgens, Grass, Drolet, & Fischer, 2015; Jürgens et al., 2011). Second, our speakers—although trained in producing clear and artifact-free speech—were non-actors. Hence, they relied on their everyday speech experience to express the intention in a way they would naturally do to be understood by an interlocutor. Last, studies on non-prosodic cues for speech acts (Bucciarelli et al., 2003; Reeder, 1980) and voluntary vocal expressions of social affect (Rilliard, Shochi, Martin, Erickson, & Auberger, 2009) suggest that cues for expressing intentions are not innate but culturally learned. On this assumption, the fact that our speakers and listeners used and understood the specific prosodic cues suggests that these cues must occur in natural conversations.

Future research on intentional prosody

An interesting question with regard to speaker’s intent in natural communication is, then, how prosodic cues are weighed and cognitively interlinked with other paralinguistic cues such as facial expressions. Notably, the latter have been shown to serve explicit interpersonal functions that reach beyond the inadvertent display of basic emotions (Ekman, 1992), for example when (voluntarily) communicating *social motives* (e.g., in case of compassion or empathy for pain) (Fridlund, 1994; Parkinson, 2005). Concerning audio–visual integration, recent motion-capture and neuroimaging studies revealed interactions between

linguistic/emotional prosody and facial expressions, in speakers (Cvejic, Kim, & Davis, 2012; Kitamura, Guellai, & Kim, 2014) and in listeners, respectively (Brück, Kreifelts, & Wildgruber, 2011; Watson et al., 2014). Yet, whether and how prosody and facial cues are fused in the transfer of speaker's meaning is currently not known and an interesting topic for future research.

Another point that deserves further examination is our observation that acoustic information predicted participant's speech act recognition successfully, yet not fully. This raises the interesting hypothesis that the comprehension of speaker's intentions from prosody relies on a weighted contribution of auditory-prosodic and other, socio-cognitive processes whose exact nature and ways of interaction still need to be clarified. On the socio-cognitive side, recent neuroimaging work lends initial evidence for inferential processes, i.e. involving theory of mind areas, during the comprehension of speech acts (Egorova, Shtyrov, & Pulvermüller, 2015; Egorova et al., 2014) and speaker meaning (Bašnáková, Weber, Petersson, van Berkum, & Hagoort, 2014; Jang et al., 2013), as well as motor system involvement during the processing of directive speech acts (Egorova et al., 2014) and indirect requests (van Ackeren, Casasanto, Bekkering, Hagoort, & Rueschemeyer, 2012). Yet, none of these studies involved prosody, leaving the fundamental question unresolved how prosody potentially interlinks with these socio-cognitive systems. Future neurocognitive investigations with the present stimuli may help to elucidate this question and are currently underway.

Conclusion

Speakers rarely code their intentions in the lexical content of an utterance. Yet, listeners easily recognize the speaker's communicative goals. The present study shows that conversationalists are able to use prosody as extralinguistic cue to specify communicative intentions—an early capacity that complements adults' mature verbal abilities. Interlocutors produce and understand prosodic cues independently of the semantic meaning, contextual information, and emotional coloring of the utterance. These results argue in favor of conventionalized acoustic feature configurations that connote communicative concepts, although their acoustic and conceptual distributions may partly overlap. The present study leads toward future research on the interaction between auditory-prosodic cues, conversation context, and socio-cognitive processes serving the transfer of speaker meaning as the foundation of successful interpersonal communication.

Acknowledgments

We thank Prof. Dr. J.D. Jescheniak and Prof. Dr. K. von Kriegstein for their helpful feedback on this project. The study was funded by the Otto Hahn award of the Max Planck Society to DS.

Appendix A

Situation descriptions presented to the speakers for speech act recordings (English translations). All speakers read the scenarios and were asked to place themselves in the interpersonal situation, for example, by uttering the example sentences (in bold), before uttering the single words (or non-words) with the corresponding prosody. Scenarios contained either words or non-words (in square brackets below).

Criticism

Your colleague Tom and you will present your first big job in an important meeting this afternoon. Therefore, you are extremely nervous and do not want to disappoint your boss. You sit at your desk and go through the presentation one more time. Suddenly, there is a knock on the door and Tom peeks into your office. He asks you whether you would like to join him for a beer [diem] in the bar [dahm], although he knows how important the upcoming meeting is. You think it would be best for him to prepare the joint talk and ask disapprovingly:

BIER [DIEM]: “(Are you serious? A) beer [diem], (now)?”
 BAR [DAHM]: “(Are you serious? The) bar [dahm], (now)?”

Doubt

You arrive home after a hard day at work and it is quite late. Your mobile rings when you have just hung your coat up. It is your friend, Eva, and she suggests having a beer [diem] at the bar [dahm]. You would actually like to meet her as you have not seen her for a long time, but are very tired and need to leave for work early the next day. You do not know whether this is a good idea. Therefore, you ask doubtfully:

BEER [DIEM]: “(A) beer [diem], (now)?”
 BAR [DAHM]: “(The) bar [dahm], (now)?”

Naming

Please say the words: beer/bar/diem/dahm with a neutral intonation, for example, as in the sentence:

“(I'm going to have a) beer [diem] (tonight).”
 “(I'm going to a) bar [dahm] (tonight).”

Suggestion

It is Thursday evening and you have almost finished your work. You achieved a lot today and are satisfied with your work. You really deserve to go to the bar [dahm] for a beer [diem] now. You think you can perhaps convince your

colleague, Anne, to join as you sometimes go to your favorite bar together after work. In pleasant anticipation of a nice evening, you peek into Anne's office and ask invitingly:

BEER [DIEM]: “(Are you up for a) beer [diem]?”
BAR [DAHM]: “(Do you want to go to a) bar [dahm]?”

Warning

You invited a friend to have a beer at your apartment. He talks excitedly about his last football match and vividly tries to imitate one of his maneuvers. He spins around wildly, back and forth, left and right, and you start getting worried about your furniture. He suddenly starts running and does not see your mini bar [dahm] where he put his glass of beer [diem]. You try to warn him:

BEER [DIEM]: “(Watch out, your) beer [diem]!”
BAR [DAHM]: “(Watch out, the) bar [dahm]!”

Wish

It is a hot summer day and you descend after a hard, but wonderful mountain hike. After all those kilometers and the great view, you are pleasantly exhausted, hungry, and thirsty—the hotel is almost within sight. You only have one thought on your mind: You would like a nice cool beer [diem] at the hotel bar [dahm] to make this day truly perfect. You say longingly:

BEER [DIEM]: “(Now for a) beer [diem]!”
BAR [DAHM]: “(Now to the) bar [dahm]!”

Appendix B

Definitions of speech acts presented to participants before the behavioral tests (English translations).

Criticism

The speaker, a friend of yours, is disapprovingly expressing criticism, for example, about one of your suggestions.

Doubt

The speaker is deliberately expressing doubt, for example about whether to accept a proposal you made.

Naming

The speaker is saying something for no specific purpose, for example, to name an object.

Suggestion

The speaker is invitingly suggesting something, for example, to undertake something together.

Warning

The speaker is warning you of a possible accident, for example, not to fall over an object.

Wish

The speaker is longingly expressing a wish for something, for example, a relaxing evening after a successful working day.

Appendix C

See Tables C1–C3.

Table C1

Mean acoustic features per speech act category.

| Speech act | Acoustic feature | | | | | | |
|------------------|-------------------------|--------------|----------------------|---------------------|---------------|---------------------------------|------------------|
| | Number of voiced frames | Mean f0 (Hz) | Offset-onset f0 (Hz) | Mean intensity (dB) | Mean HNR (dB) | Spectral center of gravity (Hz) | SD spectrum (Hz) |
| <i>Words</i> | | | | | | | |
| Criticism | 450.1 ± 61.1 | 230.7 ± 48.4 | 81.5 ± 91.6 | 65.1 ± 3.9 | 12.1 ± 3.2 | 712.1 ± 263.8 | 728.6 ± 243.4 |
| Doubt | 482.8 ± 68.7 | 188.9 ± 41.6 | 72.6 ± 30.9 | 57.2 ± 3.6 | 14.2 ± 3.4 | 511.4 ± 186.0 | 713.6 ± 204.7 |
| Naming | 341.1 ± 64.5 | 13.3 ± 42.6 | −51.6 ± 29.0 | 56.9 ± 4.2 | 13.3 ± 2.3 | 587.7 ± 248.1 | 554.4 ± 139.1 |
| Suggestion | 320.0 ± 56.4 | 206.1 ± 37.2 | 184.7 ± 55.2 | 63.3 ± 2.0 | 13.2 ± 2.7 | 617.1 ± 224.0 | 611.7 ± 136.0 |
| Warning | 428.4 ± 97.9 | 268.5 ± 49.4 | −122.7 ± 27.0 | 71.8 ± 2.3 | 13.9 ± 2.8 | 897.1 ± 209.3 | 762.2 ± 244.4 |
| Wish | 485.7 ± 62.9 | 148.6 ± 39.5 | −61.7 ± 16.4 | 59.1 ± 3.1 | 12.7 ± 2.1 | 591.1 ± 238.2 | 781.1 ± 319.6 |
| Average | 418.0 ± 95.3 | 196.9 ± 62.3 | 17.2 ± 115.5 | 62.2 ± 6.2 | 13.2 ± 2.9 | 652.8 ± 259.8 | 652.8 ± 237.0 |
| <i>Non-words</i> | | | | | | | |
| Criticism | 473.0 ± 81.1 | 250.1 ± 60.7 | 118.9 ± 112.6 | 64.8 ± 3.6 | 15.5 ± 4.7 | 645.5 ± 279.2 | 806.2 ± 227.4 |
| Doubt | 510.7 ± 57.5 | 185.8 ± 47.5 | 70.9 ± 30.8 | 57.0 ± 4.1 | 18.9 ± 4.1 | 386.8 ± 118.9 | 641.9 ± 230.0 |
| Naming | 427.3 ± 72.9 | 134.1 ± 41.7 | −56.8 ± 28.4 | 55.2 ± 4.2 | 17.2 ± 2.8 | 504.2 ± 235.0 | 597.5 ± 174.1 |
| Suggestion | 342.0 ± 50.8 | 217.4 ± 41.0 | 213.2 ± 49.5 | 62.2 ± 3.1 | 16.1 ± 3.6 | 455.0 ± 169.0 | 567.9 ± 154.4 |
| Warning | 474.2 ± 124.9 | 280.5 ± 46.7 | −125.3 ± 23.3 | 71.4 ± 1.6 | 18.1 ± 2.6 | 755.1 ± 288.8 | 713.7 ± 184.0 |
| Wish | 560.8 ± 87.1 | 142.3 ± 38.0 | −53.2 ± 19.0 | 57.9 ± 3.5 | 15.9 ± 3.9 | 494.5 ± 188.3 | 760.9 ± 205.4 |
| Average | 464.7 ± 106.8 | 201.7 ± 70.7 | 27.9 ± 128.7 | 61.4 ± 6.5 | 17.0 ± 3.9 | 540.2 ± 252.5 | 681.4 ± 214.6 |

Values depict mean ± SD. HNR = harmonics-to-noise ratio; SD = standard deviation. All values were extracted using PRAAT 5.3.01 (<http://www.praat.org>).

Table C2

Statistical comparison of acoustic features between words and non-words.

| Acoustic parameter | Speech act ratings | | | | | | | | | | | |
|--------------------|--------------------|------|----------------|-------------|----------------|-------------|--------------|-------------|----------------|-------------|--------------|-------------|
| | Criticism | | Doubt | | Naming | | Suggestion | | Warning | | Wish | |
| | t(6) | p | t(6) | p | t(6) | p | t(6) | p | t(6) | p | t(6) | p |
| Voiced frames | −0.453 | .666 | −1.376 | .218 | −1.831 | .117 | −0.710 | .504 | −0.512 | .627 | −1.679 | .144 |
| Mean f0 | −0.427 | .684 | −0.027 | .980 | 0.124 | .906 | −0.191 | .854 | −0.272 | .795 | 0.251 | .810 |
| Offset-onset f0 | −0.299 | .775 | 0.277 | .791 | 0.630 | .552 | −0.618 | .559 | −0.198 | .849 | −1.652 | .150 |
| Mean intensity | 0.159 | .879 | −0.191 | .855 | 0.605 | .567 | 0.837 | .435 | 0.105 | .920 | 0.903 | .401 |
| Mean HNR | −1.735 | .134 | − 7.098 | .000 | − 2.749 | .033 | −2.317 | .060 | − 2.659 | .038 | −2.309 | .060 |
| Center of gravity | 0.605 | .567 | 2.619 | .040 | 1.193 | .278 | 4.292 | .005 | 2.011 | .091 | 2.562 | .043 |
| SD spectrum | −0.962 | .373 | 0.611 | .564 | −0.415 | .693 | 0.689 | .516 | 0.443 | .673 | −0.008 | .994 |

Acoustic features of words and non-words were compared with paired t-tests for each speech act category (columns). Significant results ($p < .05$) are marked in bold. f0 = fundamental frequency; HNR = harmonics-to-noise ratio; SD = standard deviation.

Table C3

Results of the discriminant analyses (Experiment 1).

| | Words | Non-words |
|-----------------|-------|-----------|
| Function 1 | | |
| Offset-onset f0 | 0.881 | 0.836 |
| Function 2 | | |
| Mean intensity | 0.721 | 0.716 |
| Mean f0 | 0.467 | 0.538 |
| Function 3 | | |
| Voiced Frames | 0.768 | 0.708 |

Within-group correlations between acoustic measures and standardized canonical discriminant functions. Table includes values of the first three functions above a threshold of $r = 0.4$.

Appendix D. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.jml.2016.01.001>.

References

- Attardo, S., Eisterhold, J., Hay, J., & Poggi, I. (2003). Multimodal markers of irony and sarcasm. *Humor – International Journal of Humor Research*, 16, 243–260.
- Austin, J. L. (1962). *How to do things with words*. Cambridge: Harvard University Press.
- Bach, K. (1994). Conversational implicature. *Mind & Language*, 9, 124–162.
- Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, 70, 614–636.
- Bänziger, T., & Scherer, K. R. (2005). The role of intonation in emotional expressions. *Speech Communication*, 46, 252–267.
- Barrett, M. D., Crucian, G. P., Raymer, A. M., & Heilman, K. M. (1999). Spared comprehension of emotional prosody in a patient with global aphasia. *Neuropsychiatry, Neuropsychology, and Behavioral Neurology*, 12, 117–120.
- Bašnáková, J., Weber, K., Petersson, K. M., van Berkum, J., & Hagoort, P. (2014). Beyond the language given: The neural correlates of inferring speaker meaning. *Cerebral Cortex*, 24, 2572–2578.
- Bates, E., Camaioni, L., & Volterra, V. (1975). The acquisition of performatives prior to speech. *Merrill-Palmer Quarterly*, 21, 205–226.
- Blanc, J., & Dominey, P. (2003). Identification of prosodic attitudes by a temporal recurrent network. *Cognitive Brain Research*, 17, 693–699.
- Boersma, P., & Weenink, D. (2014). *Praat: Doing phonetics by computer [computer program]*. Version 5.3.80. <<http://www.praat.org/>> Retrieved 29.06.14.
- Bolinger, D. (1986). *Intonation and its parts: Melody in spoken English*. Stanford: Stanford University Press.
- Bradley, M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25, 49–59.

- Brück, C., Kreifelts, B., & Wildgruber, D. (2011). Emotional voices in context: A neurobiological model of multimodal affective information processing. *Physics of Life Reviews*, 8, 383–403.
- Bryant, G. A., & Fox Tree, J. E. (2005). Is there an ironic tone of voice? *Language and Speech*, 48, 257–277.
- Bucciarelli, M., Colle, L., & Bara, B. G. (2003). How children comprehend speech acts and communicative gestures. *Journal of Pragmatics*, 35, 207–241.
- Bühler, K. (1934). *Sprachtheorie: Die Darstellungsfunktion der Sprache*. Jena: Gustav Fischer.
- Carlson, K., Frazier, L., & Clifton, C. J. (2009). How prosody constrains comprehension: A limited effect of prosodic packaging. *Lingua*, 119, 1066–1082.
- Clark, H. H., & Carlson, T. B. (1981). Context for comprehension. In J. Ling & A. Baddeley (Eds.), *Attention and performance IX* (pp. 313–330). Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Crockford, C., & Boesch, C. (2003). Context-specific calls in wild chimpanzees, *Pan troglodytes verus*: Analysis of barks. *Animal Behaviour*, 66, 115–125.
- Cutler, A. (1976). *The context-dependence of “intonational meanings”*, pp. 104–115.
- Cutler, A., Dahan, D., & van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech*, 40, 141–201.
- Cutler, A., & Isard, S. D. (1980). The production of prosody. In B. Butterworth (Ed.), *Language production: Speech and talk* (Vol. 1, pp. 245–270). Academic Press.
- Cvejic, E., Kim, J., & Davis, C. (2012). Recognizing prosody across modalities, face areas and speakers: Examining perceivers' sensitivity to variable realizations of visual prosody. *Cognition*, 122, 442–453.
- Di Cesare, G., Di Dio, C., Marchi, M., & Rizzolatti, G. (2015). Expressing our internal states and understanding those of others. *Proceedings of the National Academy of Sciences*, 112, 10331–10335.
- Dore, J. (1975). Holophrases, speech acts and language universals. *Journal of Child Language*, 2, 21–40.
- Egorova, N., Pulvermüller, F., & Shtyrov, Y. (2014). Neural dynamics of speech act comprehension: An MEG study of naming and requesting. *Brain Topography*, 27, 375–392.
- Egorova, N., Shtyrov, Y., & Pulvermüller, F. (2013). Early and parallel processing of pragmatic and semantic information in speech acts: Neurophysiological evidence. *Frontiers in Human Neuroscience*, 7, 1–13.
- Egorova, N., Shtyrov, Y., & Pulvermüller, F. (2015). Brain basis of communicative actions in language. *NeuroImage*, 125, 857–867.
- Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion*.
- Enrici, I., Adenzato, M., Cappa, S., Bara, B. G., & Tettamanti, M. (2011). Intention processing in communication: A common brain network for language and gestures. *Journal of Cognitive Neuroscience*, 23, 2415–2431.
- Esteve-Gibert, N., & Prieto, P. (2013). Prosody signals the emergence of intentional communication in the first year of life: Evidence from Catalan-babbling infants. *Journal of Child Language*.
- Frick, R. W. (1985). Communicating emotion: The role of prosodic features. *Psychological Bulletin*, 97, 412–429.
- Fridlund, A. J. (1994). *Human facial expression: An evolutionary view*. Academic Press.
- Frith, C. (2009). Role of facial expressions in social interactions. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, 364, 3453–3458.

- Furrow, D., Podrouzek, W., & Moore, C. (1990). The acoustical analysis of children's use of prosody in assertive and directive contexts. *First Language*, 10, 37–49.
- Gisladottir, R. S., Chwilla, D. J., & Levinson, S. C. (2015). Conversation electrified: ERP correlates of speech act recognition in underspecified utterances. *PLoS ONE*, 10, 1–24.
- Glucksberg, S. (2003). The psycholinguistics of metaphor. *Trends in Cognitive Sciences*, 7, 92–96.
- Glucksberg, S., Gildea, P., & Bookin, H. B. (1982). On understanding nonliteral speech: Can people ignore metaphors? *Journal of Verbal Learning and Verbal Behavior*, 21, 85–98.
- Grice, H. P. (1957). Meaning. *The Philosophical Review*, 66, 377–388.
- Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics: 3. Speech acts* (pp. 41–58). New York: Academic Press.
- Grünloh, T., & Liszkowski, U. (2015). Prelinguistic vocalizations distinguish pointing acts. *Journal of Child Language*, 49, 1–48.
- Holtgraves, T. (2002). *Language as social action*. London: Lawrence Erlbaum Associates.
- Holtgraves, T. (2005). The production and perception of implicit performatives. *Journal of Pragmatics*, 37, 2024–2043.
- Holtgraves, T. (2008a). Automatic intention recognition in conversation processing. *Journal of Memory and Language*, 58, 627–645.
- Holtgraves, T. (2008b). Conversation, speech acts, and memory. *Memory & Cognition*, 36, 361–374.
- Jang, G., Yoon, S. A., Lee, S. E., Park, H., Kim, J., Ko, J. H., & Park, H. J. (2013). Everyday conversation requires cognitive inference: Neural bases of comprehending implicated meanings in conversations. *NeuroImage*, 81, 61–72.
- Jiang, X., & Pell, M. D. (2015). On how the brain decodes vocal cues about speaker confidence. *Cortex*, 66, 9–34.
- Jürgens, R., Grass, A., Drolet, M., & Fischer, J. (2015). Effect of acting experience on emotion expression and recognition in voice: Non-actors provide better stimuli than expected. *Journal of Nonverbal Behavior*, 195–214.
- Jürgens, R., Hammerschmidt, K., & Fischer, J. (2011). Authentic and play-acted vocal emotion expressions reveal acoustic differences. *Frontiers in Psychology*, 2, 1–11.
- Keyser, B. (1989). On the functional equivalence of literal and metaphorical interpretations in discourse. *Journal of Memory and Language*, 28, 375–385.
- Kitamura, C., Guellai, B., & Kim, J. (2014). Motherese by eye and ear: Infants perceive visual prosody in point-line displays of talking heads. *PLoS ONE*, 9, e111467.
- Lester, B. M., Boukydis, C. F., Zachariah Garcia-Coll, C. T., Peucker, M., McGrath, M. M., Vohr, B. R., ... Oh, W. (1995). Developmental outcome as a function of the goodness of fit between the infant's cry characteristics and the mother's perception of her infant's cry. *Pediatrics*, 95, 516–521.
- Levinson, S. C. (2006). On the “human interaction engine”. In N. J. Enfield & S. C. Levinson (Eds.), *Roots of human sociality: Culture, cognition, and interaction* (pp. 39–69). Oxford: Berg.
- Levinson, S. C. (2013). Action formation and ascription. In T. Stivers & J. Sidnell (Eds.), *The handbook of conversation analysis* (pp. 103–130). Wiley-Blackwell.
- Liu, S. (2011). An experimental study of the classification and recognition of Chinese speech acts. *Journal of Pragmatics*, 43, 1801–1817.
- Marcos, H. (1987). Communicative functions of pitch range and pitch direction in infants. *Journal of Child Language*, 14, 255.
- Mead, G. H. (1934). In C. W. Morris (Ed.), *Mind, self, and society: From the standpoint of a social behaviorist*. University of Chicago Press.
- Monetta, L., Cheang, H. S., & Pell, M. D. (2008). Understanding speaker attitudes from prosody by adults with Parkinson's disease. *Journal of Neuropsychology*, 2, 415–430.
- Morlec, Y., Bailly, G., & Aubergé, V. (2001). Generating prosodic attitudes in French: Data, model and evaluation. *Speech Communication*, 33, 357–371.
- Oller, D. K., & Griebel, U. (2014). On quantitative comparative research in communication and language evolution. *Biological Theory*, 9, 296–308.
- Papaeliou, C., Minadakis, G., & Cavouras, D. (2002). Acoustic patterns of infant vocalizations expressing emotions and communicative functions. *Journal of Speech, Language, and Hearing Research*, 45, 311–317.
- Papaeliou, C., & Trevarthen, C. (2006). Prelinguistic pitch patterns expressing “communication” and “apprehension”. *Journal of Child Language*, 33, 163–178.
- Parkinson, B. (2005). Do facial movements express emotions or communicate motives? *Personality and Social Psychology Review: An Official Journal of the Society for Personality and Social Psychology, Inc.*, 9, 278–311.
- Piantadosi, S. T., Tily, H., & Gibson, E. (2012). The communicative function of ambiguity in language. *Cognition*, 122, 280–291.
- Prieto, P., Estrella, A., Thorson, J., & Vanrell, M. D. M. (2012). Is prosodic development correlated with grammatical and lexical development? Evidence from emerging intonation in Catalan and Spanish. *Journal of Child Language*, 39, 221–257.
- Reeder, K. (1980). The emergence of illocutionary skills. *Journal of Child Language*, 7, 13–28.
- Remington, N. A., Fabrigar, L. R., & Visser, P. S. (2000). Reexamining the circumplex model of affect. *Journal of Personality and Social Psychology*, 79, 286–300.
- Rigoulot, S., Fish, K., & Pell, M. D. (2014). Neural correlates of inferring speaker sincerity from white lies: An event-related potential source localization study. *Brain Research*, 1565, 48–62.
- Rilliard, A., Shochi, T., Martin, J.-C., Erickson, D., & Auberger, V. (2009). Multimodal indices to Japanese and French prosodically expressed social affects. *Language and Speech*, 52, 223–243.
- Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*.
- Sammler, D., Grosbras, M.-H., Anwander, A., Bestelmeyer, P. E. G., & Belin, P. (2015). Dorsal and ventral pathways for prosody. *Current Biology*, 1–7.
- Sauter, D. A., Eisner, F., Calder, A. J., & Scott, S. K. (2010). Perceptual cues in nonverbal vocal expressions of emotion. *Quarterly Journal of Experimental Psychology*, 63, 2251–2272.
- Sauter, D. A., Eisner, F., Ekman, P., & Scott, S. K. (2010). Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proceedings of the National Academy of Sciences of the United States of America*, 107, 2408–2412.
- Schel, A. M., Townsend, S. W., Machanda, Z., Zuberbühler, K., & Slocumbe, K. E. (2013). Chimpanzee alarm call production meets key criteria for intentionality. *PLoS ONE*, 8.
- Schneider, K., Lintfert, B., Dogil, G., & Möbius, B. (2006). Phonetic grounding of prosodic categories. In *Methods in empirical prosody research* (pp. 335–362). Berlin: De Gruyter.
- Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge University Press.
- Searle, J. R., & Vanderveken, D. (1985). *Foundation of illocutionary logic*. Cambridge University Press.
- Seyfarth, R. M., & Cheney, D. L. (2014). The evolution of language from social cognition. *Current Opinion in Neurobiology*, 28, 5–9.
- Simon-Thomas, E. R., Keltner, D. J., Sauter, D., Sinicropi-Yao, L., & Abramson, A. (2009). The voice conveys specific emotions: Evidence from vocal burst displays. *Emotion (Washington, D.C.)*, 9, 838–846.
- Srinivasan, R. J., & Massaro, D. W. (2003). Perceiving prosody from the face and voice: Distinguishing statements from echoic questions in English. *Language and Speech*, 46, 1–22.
- Stalnaker, R. (2002). Common ground. *Linguistics and Philosophy*, 25, 701–721.
- Szameitat, D. P., Alter, K., Szameitat, A. J., Darwin, C. J., Wildgruber, D., Dietrich, S., & Sterr, A. (2009). Differentiation of emotions in laughter at the behavioral level. *Emotion (Washington, D.C.)*, 9, 397–405.
- Szameitat, D. P., Alter, K., Szameitat, A. J., Wildgruber, D., Sterr, A., & Darwin, C. J. (2009). Acoustic profiles of distinct emotional expressions in laughter. *The Journal of the Acoustical Society of America*, 126, 354–366.
- Szameitat, D. P., Kreifelts, B., Alter, K., Szameitat, A. J., Sterr, A., Grodd, W., & Wildgruber, D. (2010). It is not always tickling: Distinct cerebral responses during perception of different laughter types. *NeuroImage*, 53, 1264–1271.
- Tanenhaus, M. K., Kurumada, C., & Brown, M. (2015). Prosody and intention recognition. In L. Frazier & E. Gibson (Eds.), *Explicit and implicit prosody in sentence processing* (pp. 99–118). Springer.
- Tomasello, M. (2005). Intention reading and imitative learning. In S. Hurley & N. Chater (Eds.), *Perspectives on imitation: From neuroscience to social science* (pp. 133–148). MA: MIT Press.
- Uldall, E. (1960). Attitudinal meanings conveyed by intonation contours. *Language and Speech*, 3, 223–234.
- van Ackeren, M. J., Casasanto, D., Bekkering, H., Hagoort, P., & Rueschemeyer, S.-A. (2012). Pragmatics in action: Indirect requests engage theory of mind areas and the cortical motor network. *Journal of Cognitive Neuroscience*, 24, 2237–2247.
- Wagner, M., & Watson, D. G. (2010). Experimental and theoretical advances in prosody: A review. *Language and Cognitive Processes*, 25, 905–945.
- Warren, P. (1999). Prosody and language processing. In *Language processing* (pp. 155–188). Psychology Press Ltd.

- Warren, J. D., Warren, J. E., Fox, N. C., & Warrington, E. K. (2003). Nothing to say, something to sing: Primary progressive dynamic aphasia. *Neurocase*, 9, 140–155.
- Watson, R., Latinus, M., Noguchi, T., Garrod, O., Crabbe, F., & Belin, P. (2014). Crossmodal adaptation in right posterior superior temporal sulcus during face-voice emotional integration. *Journal of Neuroscience*, 34, 6813–6821.
- Wheeler, B. C., & Fischer, J. (2012). Functionally referential signals: A promising paradigm whose time has passed. *Evolutionary Anthropology: Issues, News, and Reviews*, 21, 195–205.
- Wichmann, A. (2000). The attitudinal effects of prosody, and how they relate to emotion. *Proceedings of the ISCA Workshop on Speech and Emotion*, Newcastle, pp. 143–148.
- Wichmann, A. (2002). Attitudinal intonation and the inferential process. *Proceedings of Speech Prosody Conference*, Aix-en-Provence.
- Wilson, D., & Sperber, D. (2012). *Meaning and relevance*. Cambridge University Press.
- Wittgenstein, L. (1953). *Philosophische Untersuchungen*. Suhrkamp.
- Wundt, W. (1896). *Grundriss der Psychologie*. Leipzig: Engelmann.

4.2 Study VIII – Musical harmony in (inter)action

Sammler, D.*, Novembre, G.*, Koelsch, S., Keller, P. E. (2013). Syntax in a pianist's hand: ERP signatures of “embodied” syntax processing in music. *Cortex*, 49, 1325-1339. (*equal contribution)*

[Author contributions by D.S.: designed research with G.N., P.E.K., performed research with G.N.; analysed data with G.N.; co-wrote the paper with G.N., incorporating edits by co-authors]

*Reprinted with permission from Elsevier.

CORTX 49 (2013) 1325–1339

Available online at www.sciencedirect.com

SciVerse ScienceDirect

Journal homepage: www.elsevier.com/locate/cortex

Research report

Syntax in a pianist's hand: ERP signatures of “embodied” syntax processing in music

Daniela Sammler^{a,*}, Giacomo Novembre^{b,1}, Stefan Koelsch^c and Peter E. Keller^b

^aDepartment of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

^bResearch Group “Music Cognition and Action”, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

^cCluster of Excellence “Languages of Emotion”, Free University Berlin, Berlin, Germany

ARTICLE INFO

Article history:

Received 11 November 2011

Reviewed 03 January 2012

Revised 2 April 2012

Accepted 13 June 2012

Action editor Stefano Cappa

Published online 29 June 2012

Keywords:

Music

Syntax

Action

Imitation

Embodiment

ABSTRACT

Syntactic operations in language and music are well established and known to be linked in cognitive and neuroanatomical terms. What remains a matter of debate is whether the notion of syntax also applies to human actions and how those may be linked to syntax in language and music. The present electroencephalography (EEG) study explored syntactic processes during the observation, motor programming, and execution of musical actions. Therefore, expert pianists watched and imitated silent videos of a hand playing 5-chord sequences in which the last chord was syntactically congruent or incongruent with the preceding harmonic context. 2-chord sequences that diluted the syntactic predictability of the last chord (by reducing the harmonic context) served as a control condition. We assumed that behavioural and event-related potential (ERP) effects (i.e., differences between congruent and incongruent trials) that were significantly stronger in the 5-chord compared to the 2-chord sequences are related to syntactic processing. According to this criterion, the present results show an influence of syntactic context on ERPs related to (i) action observation and (ii) the motor programming for action imitation, as well as (iii) participants' execution times and accuracy. In particular, the occurrence of electrophysiological indices of action inhibition and reprogramming when an incongruent chord had to be imitated implies that the pianist's motor system anticipated (and revoked) the congruent chord during action observation. Notably, this well-known anticipatory potential of the motor system seems to be strongly based upon the observer's music-syntactic knowledge, thus suggesting the “embodied” processing of musical syntax. The combined behavioural and electrophysiological data show that the notion of musical syntax not only applies to the auditory modality but transfers – in trained musicians – to a “grammar of musical action”.

© 2012 Elsevier Ltd. All rights reserved.

* Corresponding author. Max Planck Institute for Human Cognitive and Brain Sciences, 04103 Leipzig, Germany.

E-mail address: sammler@cbs.mpg.de (D. Sammler).

¹ These authors contributed equally to the study.

0010-9452/\$ – see front matter © 2012 Elsevier Ltd. All rights reserved.

<http://dx.doi.org/10.1016/j.cortex.2012.06.007>

1. Introduction

From a structural point of view, a linguistic sentence, a musical phrase and a goal-directed action share one key property: all are composed of discrete items (words, tones, motor acts) that are strung together according to specific rules (language-specific grammars, culture dependent tonal systems, motor constraints) to build-up meaning over the course of a sequence (Chomsky, 1957; Lashley, 1951; Lerdahl and Jackendoff, 1983). Modern comparative research further stresses analogies between these domains in terms of hierarchical organization and recursion, concepts that are particularly well established in language (Chomsky, 1957, 1995), and increasingly substantiated in harmonic structure in music (Katz and Pesetsky, 2011; Rohrmeier, 2011) as well as complex actions (Pastra and Aloimonos, 2012). We will refer to this shared property – i.e., the rule-based hierarchical and compositional ordering of discrete elements into sequences – as “syntax”. This term is clearly borrowed from (psycho)linguistics, a field that describes the organization of language, discusses the universals (Chomsky, 1986, 1995; Fitch, 2011; Moro, 2008) and essentials (Hauser et al., 2002) of the language faculty, and – most relevant to us – assumes a neural system that may be dedicated to the processing of syntax in natural languages (Moro et al., 2001; Musso et al., 2003; Pallier et al., 2011).

However, the intriguing parallels of “syntactic structure” in language, music, and perhaps action lead to the question whether this neural system and the cognitive operations necessary to isolate, process, and integrate syntactically organized elements are specific to language or may be shared between domains. The strongest evidence in favour of shared syntactic resources comes from investigations on speech and music perception (Koelsch, 2011; Patel, 2003). Here it has been shown that syntactic violations in the two domains elicit comparable electric brain potentials (Koelsch, 2005; Patel et al., 1998) and activate overlapping brain regions (Abrams et al., 2011; Sammler et al., 2009) including Broca’s area and its right hemisphere homotope (Maess et al., 2001), i.e., areas that have been typically associated with syntactic processing in language (Grodzinsky and Santi, 2008). Moreover, simultaneous presentations of syntactic errors in music and language evoke interference effects (Fedorenko et al., 2009; Koelsch et al., 2005; Slevc et al., 2009; Steinbeis and Koelsch, 2008), neurological patients show parallel syntactic deficits in both domains (Grodzinsky, 2000; Patel et al., 2008; Sammler et al., 2011), and syntactic capabilities in one domain are enhanced after training in the other domain (Jentschke and Koelsch, 2009; Jentschke et al., 2005; Marin, 2009). In other words, these combined findings gave rise to the idea that the brain’s dedicated syntax network (Friederici, 2011; Kaan and Swaab, 2002; Moro et al., 2001; Pallier et al., 2011) may be less language-specific than initially thought.

Since Lashley’s seminal article on the structural principles of goal-related actions (Lashley, 1951), it is a matter of debate whether the notion of syntax also applies to human actions. Most recent work has been dedicated to the formalization of the compositional (Guerra-Filho and Aloimonos, 2012; see also Zacks and Tversky, 2001) and generative organization of actions (Pastra and Aloimonos, 2012) in comparison to the

syntactic organization of language. Moreover, several studies have aimed at clarifying whether the cognitive processes (Allen et al., 2010; Greenfield, 1991) and underlying neural correlates (Farag et al., 2010; van Schie et al., 2006) that operate on compositional action structures are the same as the ones found in language and music. Similar parallels have also been discussed for visuo-spatial sequencing (Bahlmann et al., 2009; Tettamanti et al., 2009), logic (Monti et al., 2009) or arithmetic (Dehaene, 1997; Friedrich and Friederici, 2009; Nuñez-Peña and Honrubia-Serrano, 2004; Scheepers et al., 2011; although the rules of “syntactic” combination have to be explicitly taught in arithmetic, while they are implicitly acquired in language, music and simple actions, such as goal-related grasping).

Most authors adopt the view of a domain-general hierarchical syntax processor in the inferior frontal lobe (Fadiga et al., 2009; Fiebach and Schubotz, 2006; Gelfand and Bookheimer, 2003; Koelsch, 2011; Patel, 2003; Tettamanti and Weniger, 2006), although this is not yet unequivocally proven (Rogalsky et al., 2011). An alternative approach pertains to a polymodal sensorimotor theory of syntax, i.e., the involvement of action-perception circuits to mediate grammar processing in language (Pulvermüller and Fadiga, 2010; van Schie et al., 2006), music (Fadiga et al., 2009), and action (Clerget et al., 2009; Fazio et al., 2009). Although it is not clear how rule-based structures might be processed in sensorimotor areas alone (i.e., by means of a “mirror” mechanism without the recruitment of an extra parser that processes syntactic dependencies, see Tettamanti and Moro, 2012), it is possible that the motor system makes use of syntactic operations during the perception and production of sequences of acts forming goal-directed actions. Some evidence for such a syntax-action link can be inferred from models of incremental planning of serial actions such as speech or music (for a review, see Palmer and Pfordresher, 2003). These models suggest that the ongoing advance construction of motor programs during musical performance is governed by musical structure, e.g., melodic, harmonic or metrical relationships between tones and chords of a musical piece, whose statistical regularities have been acquired over the course of experience (Palmer and van de Sande, 1993, 1995; Restle, 1970).

In a recent behavioural study, Novembre and Keller (2011) explored the impact of syntactic knowledge on musical actions by means of an imitation paradigm. Expert pianists watched and imitated videos displaying one hand performing sequences of chords, including occasional chords that were harmonically, that is syntactically, incongruent with the preceding musical context (i.e., the events that precede the target chord and whose syntactic structure influences how the target chord is perceived). The experiment was run in the absence of sound. Results showed that imitation of chords was faster when they were embedded in a congruent (i.e., syntactically regular) context, suggesting that the harmonic rules implied by the observed actions induced strong expectancies that influenced action execution. Therefore, this study provided evidence in favour of syntactic structures regulating the progression of motor acts associated with producing music. The authors suggested that, as a result of musical training, the rules determining the position of chords within

chord sequences are internalized as a form of “embodied harmony”, i.e., that the motor system of skilled musicians makes use of syntactic rules in the perception and production of musical actions. On a more abstract level, this notion alludes to theories of embodied cognition that ground cognition in the bodily senses and mental simulation (Barsalou, 2008; Gibbs, 2006; Wilson, 2002) instead of segregating body and mind.

The present study set out to test further the hypothesis of “embodied” processing of harmony and zoomed into the neurophysiological correlates of syntactic operations during the observation and imitation of musical actions. Particularly, we aimed to reveal the time course and origin of the syntactic interference effects described by Novembre and Keller (2011), i.e., the influence of syntax on (i) the observation of musical performance, (ii) the translation of observed movements into a motor program, and (iii) the execution of the movements themselves. Therefore, electroencephalogram (EEG) and piano performance were recorded while skilled pianists watched and imitated the videos employed by Novembre and Keller (2011) displaying 5-chord sequences with and without syntactic violations. In order to control for differences between target chords other than syntactic congruity, such as visual appearance or motoric complexity (see *Methods*), we included an additional set of videos displaying 2-chord sequences. These videos kept the visual and motoric aspects of the target chords invariant, but diluted the music-syntactic predictability of the sequences by reducing harmonic context information (Fig. 1). Note that no sound was presented during the whole experiment, neither in the videos nor on the piano used by the participants in the imitation task.

We predicted that the harmonic rules – as implied by the observed and imitated movements – would induce motor expectations. Thus, the perception and imitation of the last chord should elicit distinct electrophysiological brain responses, imitation time and accuracy, depending on the chord’s

congruency with the preceding harmonic context. From what is known from auditory studies, the observed music-syntactic violations may evoke an early right anterior negativity (ERAN) – i.e., an electrophysiological marker of early musical structure building – and an N500 or P600 both reflecting later stages of syntactic integration (Koelsch, 2009; Koelsch et al., 2000; Patel et al., 1998), although these chord sequences have before never been studied in the context of action. Importantly, based on the observation that the sense of a tonal centre and corresponding music-syntactic expectancies usually gain strength over the course of a musical piece (Bigand and Parncutt, 1999; Koelsch et al., 2000; Leino et al., 2007), we assumed that the 5-chord sequences should induce stronger syntactic expectancies than the 2-chord sequences. Hence, any behavioural or event-related potential (ERP) effect related to the processing of syntax in musical actions should be stronger in the 5-chord than 2-chord sequences, i.e., evidenced by a statistical interaction of Congruency (congruent/incongruent) \times Context (5-chord/2-chord sequences). (The factor Context relates to the number of events that precede the target chords and whose syntactic structure influences how the target chords are perceived and imitated). This criterion, in combination with the excellent temporal resolution of the EEG, should allow us to specify the time course of interaction between perceptuo-motor and syntactic processes. Ultimately, the present paradigm should permit us to specify the neurophysiological signatures and computational underpinnings of putative “embodied” syntactic processes in action.

2. Methods

2.1. Participants

Twenty-seven right-handed pianists (nine males), aged 20–34 years [mean = 24.93, standard deviation (SD) = 3.55], were

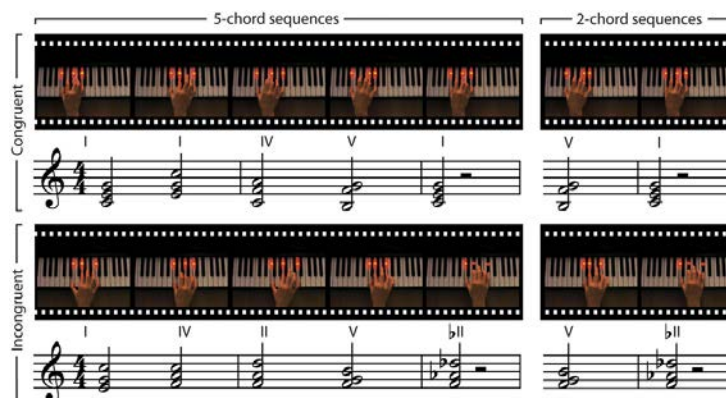


Fig. 1 – Experimental design. Participants watched and imitated silent videos showing a right hand playing chord sequences composed of five or two chords. Two-chord sequences were derived from the 5-chord sequences by deleting the first three chords. The final target chord of each sequence was either syntactically congruent (i.e., Tonic [I]) or incongruent (i.e., the major chord of the lowered second scale degree [bII]). Scores are only shown for illustration and were not presented to the pianists. The experiment was run in the absence of sound.

included in the analysis. Three additional participants were tested but excluded from data analysis because they were not able to perform the task. All pianists had a minimum of 14 years of formal training in classical music, the mean age at which piano studies commenced was 6.31 years ($SD = 1.52$), mean years of piano training was 16.96 years ($SD = 3.45$), and average weekly amount of practice was 7.70 h ($SD = 10.67$). All pianists were naïve with regard to the purpose of the study.

2.2. Stimuli

Stimuli were identical to a portion of those used by [Novembre and Keller \(2011\)](#). They consisted of silent videos showing a female pianist's right hand playing sequences of chords on a muted keyboard (Yamaha EZ200) equipped with red light-emitting diodes (LEDs). These LEDs were illuminated for the duration of each key press and made the identity of the pressed keys clear to the participant (cf. [Novembre and Keller, 2011](#)).

A total of 60 different chord sequences were used for this experiment: 30 were syntactically regular in the sense that they came to a conventional harmonic resolution (congruent condition; upper left panel in [Fig. 1](#)) and 30 were irregular in that they ended in an uncommon and unresolved harmony (incongruent condition; lower left panel in [Fig. 1](#)). For each condition, the chord sequences were in the key of C, D, or F major (10 sequences/key). All chords consisted of three piano keystrokes. The first chord was always the tonic of the given tonal context, and was followed either by a tonic, supertonic, or subdominant at the second position. Chords at the third position were the tonic, subdominant, supertonic or submediant. At the fourth position dominant seventh chords were presented in root position, or in first or third inversion. The chord at the fifth position was different between the two conditions: a tonic chord (congruent condition) or the major chord built on the lowered second scale degree (incongruent condition). Tonic chords were presented in root position, first, and second inversions. Incongruent chords were presented in both first and second inversions. In general, chord sequences had different 'melodic contours' (e.g., starting with the first, third, or fifth degree of the tonic chord) in the top voice.

It is important to note that the videos displayed non-manipulated biological movements (apart from the first chord; see below) – as recorded in natural piano playing – which were intended to maximally activate the observers' motor system ([Buccino et al., 2004](#); [Perani et al., 2001](#); [Stevens et al., 2000](#)). This implied, however, that the spatial trajectory performed by the model hand moving from the penultimate to the incongruent target chords was significantly longer (mean trajectory duration from movement onset to offset = 303 msec, $SD = 86$ msec) than when moving to the congruent targets [mean = 221 msec, $SD = 69$ msec; $t(58) = -4.07$, $p < .001$]. Moreover, other visual aspects such as movement fluency, finger configuration and number of black keys, along with motor task complexity and familiarity could not be kept entirely constant between congruent and incongruent target chords, necessitating an appropriate control condition to isolate syntax-related brain activity. Therefore, we included an additional set of 60 excerpt-videos displaying only the last two chords from the 5-chord sequences described above, i.e., 2-chord sequences (right panel of [Fig. 1](#)). Note that the control videos were truncated

versions of the original videos of the 5-chord sequences. As a result, the target chords of 5- and 2-chord sequences were physically identical (and thus also identical in terms of visual appearance, motoric complexity and familiarity), and merely differed in their syntactic predictability. In other words, the longer (5-chord) music-syntactic context should induce a stronger sense of tonality ([Bigand and Parncutt, 1999](#); [Koelsch et al., 2000](#); [Leino et al., 2007](#)) and thus stronger syntactic expectancies than the shorter (2-chord) sequences. Consequently, any behavioural or electrophysiological effect that is significantly stronger in 5- than 2-chord sequences – i.e., reflected in a statistical interaction of Congruency \times Context – should be clearly attributable to enhanced syntactic processing.

Nevertheless, it should be noted that the 2-chord sequences also contained a certain degree of syntactic information (as they constituted common 2-chord progressions in the Western tonal system). This implies (1) that we manipulated the amount of syntactic information rather than its presence or absence, and (2) that therefore the comparison of 5- and 2-chord sequences may cancel out some syntactic processing aspects.

Each video started with a stationary hand poised to press the three keys associated with the first chord for 3 sec, to give the participant enough time to match the initial position of his or her own hand with the position of the model hand in the video. After that, the model hand executed the chord progression with each chord lasting approximately 2 sec, leading to video durations of 13 sec and 7 sec for 5- and 2-chord sequences, respectively. Because data of interest were the brain responses to the perception and imitation of the last chord in each sequence, the presentation of this chord was time-locked to the video onset. This was done by decomposing each video into its constituent frames (of which there were 30/sec), extending or shortening the first chord, and thus moving the first frame in which the model hand pressed all three keys of the target chord (i.e., all three LEDs were on) to 11 sec (for the 5-chord sequences) or 5 sec (for the 2-chord sequences) after video onset (videos were edited using the software iMovie HD 6.0.3, Apple Computer, Inc.).

2.3. Procedure

Participants were asked to watch and simultaneously imitate the silent videos, which were presented on a computer monitor placed on a musical instrument digital interface (MIDI) piano (Yamaha Clavinova CLP150). They were instructed to imitate both the key presses and the fingerings as fast and correctly as possible with their right hand. Furthermore, they were asked to move as minimally as possible to avoid muscle artefacts in the EEG. Each trial started with a visual fixation cross presented for 500 msec.

Sixty 5-chord and sixty 2-chord sequences were presented separately in two blocks, which were repeated once in order to increase statistical power (resulting in 240 chord sequences in total). The order of the blocks alternated and was counter-balanced across participants (e.g., 5-chords, 2-chords, 5-chords, 2-chords). Trials within each block were randomized individually for each participant. To increase participants' familiarity with the stimuli and accuracy in the task, the experiment started with a short training session consisting of

a short 5-chord and 2-chord block in counterbalanced order, each comprising 20 sequences (10/condition, in the key of G major). To control for individual differences in task strategy, participants were asked to fill in a questionnaire at the end of the experiment. Specifically, they rated (from 1 to 9) to what extent they relied on auditory and/or motor imagery, and their theoretical knowledge of western harmony, in order to perform the task efficiently.

Presentation software (Version 14.2, Neurobehavioral Systems, Inc.) was used to control both stimulus presentation (i.e., videos) and response registration (i.e., keystrokes on the piano). A MIDI interface converted the MIDI key values received from the piano keyboard into a serial signal that was compatible with Presentation software. This permitted us to compute the times at which specific keys were struck in relation to event timing in the video. Additionally, a video camera (Sony, HDR-HC9E) placed above the piano recorded the performed fingering in the participant's hand from an aerial perspective.

2.4. EEG data acquisition

EEG was recorded from 61 Ag/AgCl electrodes mounted in an elastic cap according to the extended international 10–20 system (Sharbrough et al., 1991). The electrode positions were: FPZ, FP1, FP2, AFZ, AF3, AF4, AF7, AF8, FZ, F1, F2, F3, F4, F5, F6, F7, F8, FCZ, FC1, FC2, FC3, FC4, FC5, FC6, FT7, FT8, CZ, C1, C2, C3, C4, C5, C6, T7, T8, CPZ, CP1, CP2, CP3, CP4, CP5, CP6, TP7, TP8, PZ, P1, P2, P3, P4, P5, P6, P7, P8, POZ, PO3, PO4, PO7, PO8, OZ, O1, O2. Left mastoid (M1) served as reference; an additional electrode was placed on the right mastoid bone (M2) and the tip of the nose for off-line re-referencing. The ground electrode was located on the sternum. Horizontal and vertical electrooculograms were bipolarly recorded from electrodes placed on the outer canthus of each eye, as well as above and below the right eye. Impedances were kept below 5 k Ω . Signals were amplified with a 24 bit Brainvision QuickAmp 72 amplifier (Brain Products GmbH, Gilching, Germany) and digitized with a sampling rate of 500 Hz.

2.5. Behavioural data analysis

Errors and response times (RTs) for imitation of the target chord (i.e., the last chord) of each trial were analyzed in accordance with Novembre and Keller (2011). If both the last and the second-last chords had been correctly imitated in terms of the keys pressed and the fingering employed, then a trial was considered to be correct. Chords in which the keystrokes were not synchronous (i.e., when more than 150 msec intervened between the first and the last keystroke) were excluded from analysis (cf. Drost et al., 2005). Errors were counted if the target chord was incorrectly imitated in terms of the keys pressed, the fingering employed, or both. Errors were counted only if the previous chord (i.e., the second-last chord) had been correctly imitated in terms of both keys and fingering.

RTs were measured in correct trials by calculating the time elapsed between the presentation of the target chord (i.e., the frame in which the model hand struck all three target keys) and the participant's execution of the same chord (i.e., mean

of the three keystroke times composing the chord). RTs exceeding 3000 msec were not analyzed (cf. Drost et al., 2005). Statistical analyses were conducted on errors and RT data using separate two-way repeated measures analyses of variance (ANOVAs) with the variables Congruency (congruent/incongruent) and Context (5-chord/2-chord sequences).

2.6. EEG data analysis

EEP 3.2 (ANT-software) was used to re-reference the data to the algebraic mean of both mastoid leads. Further processing steps were done using EEGLAB 6.01 (Delorme and Makeig, 2004) in MATLAB 7.7. Data were filtered using a .3 Hz high-pass filter (fir, 5854 points, Blackman window), and strong muscle artifacts, electrode drifts or technical artifacts were manually rejected before entering the continuous data into an independent component analysis. The resulting component structure was used to reject eye movement and blink artifacts, muscle artifacts and slow drifts. Afterwards, the data were filtered with a 25 Hz lowpass filter (fir, 110 points, Blackman window), and cut into epochs ranging from –800 to 1000 msec relative to the target chord in the videos (i.e., the frame when all three target keys were pressed). Only correct trials (i.e., mean \pm SD long context: 39.02 \pm 10.18, short context: 49.61 \pm 6.61; according to the criteria in the behavioural data) were included in the ERP analysis. Trials were rejected whenever one or more electrodes exhibited voltages of ± 50 μ V. Altogether, this procedure allowed the complete elimination of movement artifacts caused by the imitation task, e.g., eye movements between screen and keyboard or tension of neck and shoulder muscles during playing. Non-rejected trials were averaged separately for each condition. Averages were aligned to a –800 to –300 msec baseline, i.e., to a time in the video during which the model hand rested on the keys of the penultimate chord, prior to the trajectory onset towards the target chord. An average of 39.17 trials was included for each participant and each condition (mean \pm SD long context: 35.67 \pm 11.16, short context: 42.67 \pm 10.02).

Effects of chord congruency and context length were analyzed time-locked to the target chord in the video, i.e., the point when the model hand struck the keys of the 5th chord in the 5-chord sequences and the 2nd chord in the 2-chord sequences. Statistical analyses were carried out on the mean amplitudes in each condition calculated for specific time windows (see Results) in nine regions of interest (ROIs): (i) left anterior (F3, F5, F7, FC3, FC5, FT7), (ii) left central (C3, C5, T7, CP3, CP5, TP7), (iii) left posterior (P3, P5, P7, PO3, PO7), (iv) middle anterior (F1, FZ, F2, FC1, FCZ, FC2), (v) middle central (C1, CZ, C2, CP1, CPZ, CP2), (vi) middle posterior (P1, PZ, P2, POZ), (vii) right anterior (F4, F6, F8, FC4, FC6, FT8), (viii) right central (C4, C6, T8, CP4, CP6, TP8), and (ix) right posterior (P4, P6, P8, PO4, PO8). Four time windows were defined separately in 5- and 2-chord sequences by visual inspection of the ERPs and topography plots according to the following criteria: assuming that different map topographies and polarities directly indicate different underlying generators, i.e., different cognitive processes (Michel et al., 2004), borders between time windows were set whenever the topography shifted or polarity of the effect flipped (for details, see Results). Note that this approach generated a different border between the first

and second time window in 5-chord (–80 msec) and 2-chord sequences (0 msec). This is most likely due to the better syntactic (and temporal) predictability of the 5- compared to the 2-chord sequences, possibly leading to an acceleration of cognitive processes and their related ERP components.

Statistical evaluation comprised a four-way ANOVA with the repeated measures factors Congruency (congruent/incongruent) \times Context (5-chord/2-chord sequences) \times AntPost (anterior/central/posterior) \times Laterality (left/middle/right). Whenever an interaction involving the factor Congruency was found, follow-up analyses were carried out by splitting up the factorial model.

3. Results

3.1. Behavioural data

Fig. 2A shows mean RTs for correctly produced target chords in each condition. A two-way ANOVA with the repeated measures factors Congruency (congruent/incongruent) and Context (5-chord/2-chord sequences) yielded a significant main effect of Congruency [$F(1,26) = 98.89, p < .001$] and a significant Congruency \times Context interaction [$F(1,26) = 13.98, p < .002$]. This indicates that imitation of congruent chords was overall faster than imitation of incongruent chords, and fastest when a congruent chord was embedded in a 5-chord than 2-chord sequence. Notably, *t*-tests for paired samples showed that the congruent chord in 5-chord sequences was executed significantly faster than in 2-chord sequences [$t(26) = -3.02, p < .007$], whereas no significant difference was found between incongruent chords across long and short contexts [$t(26) = .422, p > .676$]. This suggests that the extended harmonic context facilitated the execution of the congruent chord (rather than interfering with the execution of the incongruent chord). The main effect of Context was not significant [$F(1,26) = 2.09, p > .159$] demonstrating that imitation of the target chords across 5- and 2-chord sequences did not differ in terms of RT.

A similar trend was observed in the mean number of errors, as depicted in Fig. 2B. Less errors were committed during imitation of congruent compared to incongruent chords (main effect of Congruency [$F(1,26) = 23.17, p < .001$]) while errors did not differ between 5- and 2-chord sequences (no main effect of Context [$F(1,26) = 2.59, p > .119$]). Although particularly few errors were produced in the congruent condition in the 5-chord (compared to 2-chord) sequences, the Congruency \times Context interaction fell short of statistical significance [$F(1,26) = 1.56, p > .222$]. Consistently with what was observed for the RTs, *t*-tests for paired samples showed that significantly fewer errors were produced during imitation of congruent target chords in 5- compared to 2-chord sequences [$t(26) = -3.705, p < .002$], whereas no significant difference was found between incongruent chords between long and short contexts [$t(26) = -.251, p > .803$].

3.2. EEG data

In both the 5- and 2-chord sequences a four-phasic ERP pattern was found (Fig. 3), each phase will be described in turn.

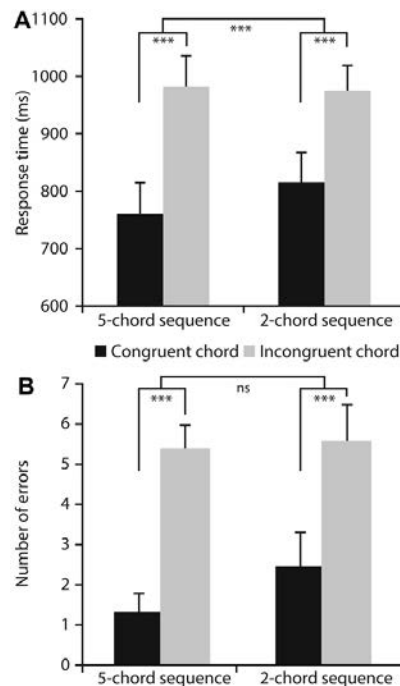
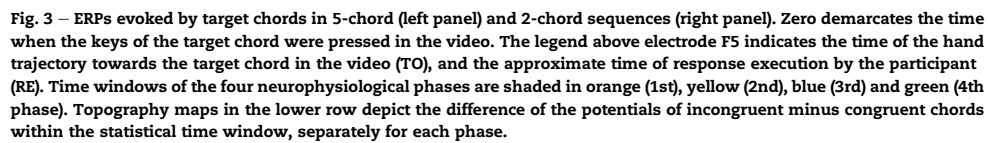


Fig. 2 – Behavioural data. (A) RTs time-locked to the key press in the video. (B) Number of errors. Error bars indicate one standard error of means.

In the 1st phase (shaded in orange in Fig. 3) prior to keystroke in the video, i.e., during the trajectory of the hand towards the target chord in the video, incongruent chords evoked a more positive potential than congruent target chords, in both the 5- and the 2-chord sequences. Yet, the positivity had a shorter duration in the 5- compared to 2-chord sequences: while it gave way to an anterior negativity around –80 msec in the 5-chord sequences, this happened only at 0 msec in the 2-chord sequences. This difference in timing may reflect a speeding-up of the 2nd phase anterior negativity (see below) due to higher predictability of the 5-chord sequences. To account for this difference, time windows for statistical testing were set to –300 to –80 msec in the 5-chord sequences and –300 to 0 msec in the 2-chord sequences. (–300 msec were chosen as onset because the trajectory towards the incongruent chords started on average at –300 msec; see Methods.) An ANOVA with the repeated measures factors Congruency (congruent/incongruent) \times Context (5-chord/2-chord sequences) \times AntPost (anterior/central/posterior) \times Laterality (left/middle/right) revealed a significant main effect of Congruency and an interaction of Congruency \times AntPost \times Laterality indicating a broadly distributed positivity irrespective of sequence length (for statistical values, see Table 1). A significant interaction of Congruency \times Context \times AntPost alluded to the more anterior



In the **2nd phase** (shaded in yellow in Fig. 3), incongruent compared to congruent chords evoked an anterior negativity between -80 and 150 msec in the 5-chord sequences and

| Effect | df | 1st phase: -300...-80 versus -300...0 msec ^a | | | 2nd phase: -80...150 versus 0...150 msec ^a | | | 3rd phase: 150...400 msec | | | 4th phase: 400...1000 msec | | |
|----------------|-------|---|-----------------|--------------|---|-----------------|------------|------------------------------|-----------------|--------------|-------------------------------|-----------------|------------|
| | | F | p-value | η_p^2 | F | p-value | η_p^2 | F | p-value | η_p^2 | F | p-value | η_p^2 |
| | | C | 1,26 | 10.77 | <.030 | .293 | <1 | >.532 | .015 | 21.41 | <.001 | .452 | 3.75 |
| C × Co | 1,26 | <1 | >.779 | .003 | 6.68 | <.016 | .204 | 7.08 | <.014 | .214 | <1 | >.863 | .001 |
| C × A | 2,52 | <1 | >.760 | .006 | 14.01 | <.001 | .350 | 1.96 | >.171 | .070 | <1 | >.765 | .005 |
| C × A × Co | 2,52 | 4.15 | <.049 | .138 | <1 | >.802 | .003 | 5.00 | <.032 | .161 | <1 | >.462 | .024 |
| C × L | 2,52 | 3.06 | >.059 | .105 | 11.40 | <.001 | .305 | 11.94 | <.001 | .315 | 2.79 | >.070 | .097 |
| C × L × Co | 2,52 | <1 | >.820 | .008 | 2.80 | >.073 | .097 | 1.40 | >.256 | .051 | 1.45 | >.244 | .053 |
| C × A × L | 4,104 | 3.50 | <.035 | .119 | 4.01 | <.013 | .134 | 9.84 | <.001 | .275 | <1 | >.414 | .036 |
| C × A × L × Co | 4,104 | <1 | >.485 | .031 | 1.04 | >.384 | .038 | <1 | >.454 | .033 | 1.08 | >.361 | .040 |

a Note that similar results were found when identical time windows were used for both 5- and 2-chord sequences, i.e., 1st phase -300...-80 msec and 2nd phase 0...150 msec.

between 0 and 150 msec in the 2-chord sequences. (150 msec was chosen as offset because of a remarkable posterior topography shift of the negativity in 5-chord sequences and a return to zero in 2-chord sequences at that time.) The four-way ANOVA showed significant interactions of Congruency \times AntPost, Congruency \times Laterality, and Congruency \times AntPost \times Laterality, demonstrating the middle-to-right frontal maximum of the negativity (Table 1). Follow-up ANOVAs with the factor Congruency computed for each ROI separately confirmed a predominantly middle-to-right anteriorly distributed negativity [middle anterior: $F(1,26) = 10.58$, $p < .004$, $\eta_p^2 = .289$; middle central: $F(1,26) = 4.86$, $p < .037$, $\eta_p^2 = .157$; right anterior: $F(1,26) = 8.36$, $p < .008$, $\eta_p^2 = .243$] accompanied by a left posterior positivity [$F(1,26) = 9.02$, $p < .006$, $\eta_p^2 = .258$; all other $ps > .158$] that most likely reflects the tail of the 1st phase positivity (see Fig. 3). The negativity was significantly greater in amplitude in the 5- compared to the 2-chord sequences as demonstrated by a significant interaction of Congruency \times Context across all electrodes (Table 1) as well as within single ROIs [middle anterior: $F(1,26) = 4.92$, $p < .036$, $\eta_p^2 = .159$; middle central: $F(1,26) = 7.57$, $p < .011$, $\eta_p^2 = .226$; middle posterior: $F(1,26) = 5.95$, $p < .022$, $\eta_p^2 = .186$; all other $ps > .061$].

To evaluate in how far the negativity in the 5-chord sequences may have been influenced by (conscious) auditory or motor imagery strategies or the application of music-theoretical knowledge, the mean amplitude of the difference wave (incongruent – congruent) in middle anterior, middle central, and right anterior ROIs was correlated with the ratings obtained in the debriefing. No significant relationships were found for auditory imagery ($r = .189$, $p > .344$, $R^2 = .036$) and motor imagery ($r = .157$, $p > .435$, $R^2 = .025$), whereas the negativity was reduced in amplitude with greater explicit reliance on music-theoretical knowledge ($r = .440$, $p < .023$, $R^2 = .193$), suggesting that the effect was not driven by the conscious identification of the music-syntactic incongruity.

In the 3rd phase (shaded in blue in Fig. 3), incongruent compared to congruent chords elicited a broadly distributed but posteriorly pronounced negativity in 5-chord sequences and an anteriorly distributed negativity in 2-chord sequences, both in the time range from 150 to 400 msec after keystrokes in the video. (400 msec was chosen as offset because the negativities in both contexts gave way to a positivity at that time; see below.) The four-way ANOVA revealed a significant main effect of Congruency, and significant interactions of Congruency \times Laterality and Congruency \times AntPost \times Laterality. Furthermore, interactions of Congruency \times Context and Congruency \times AntPost \times Context were observed (Table 1), indicating a stronger and more posteriorly distributed negativity in 5- compared to 2-chord sequences. Follow-up analyses with the factors Congruency \times Context in each ROI evidenced a significantly stronger negativity in 5- compared to 2-chord sequences at posterior electrodes [interaction of Congruency \times Context; left posterior: $F(1,26) = 11.49$, $p < .003$, $\eta_p^2 = .307$; middle posterior: $F(1,26) = 10.81$, $p < .003$, $\eta_p^2 = .294$; middle central: $F(1,26) = 5.67$, $p < .025$, $\eta_p^2 = .179$; right posterior: $F(1,26) = 9.26$, $p < .006$, $\eta_p^2 = .263$], whereas effects did not differ at anterior and central electrodes (no interaction of Congruency \times Context in the remaining ROIs; all $ps > .115$).

In the 4th phase (shaded in green in Fig. 3) between 400 and 1000 msec, incongruent chords evoked stronger positive

potentials than congruent chords similarly in both 5- and 2-chord sequences. The four-way ANOVA showed a marginally significant main effect of Congruency and interaction of Congruency \times Laterality (Table 1), suggestive of a stronger positivity in the left and right lateral compared to middle electrodes. No interactions were found between Congruency \times Context, demonstrating that the effects were similar in amplitude and topography in both 5- and 2-chord sequences.

4. Discussion

The present study explored the degree to which musical actions are governed by syntactic processes. Specifically, we aimed to examine the influence of syntax on different aspects of action such as the observation of another person's actions, as well as the programming and execution of one's own actions. To this end, expert piano players simultaneously watched and imitated videos of chord sequences in which the harmonic congruity of the last chord with the preceding syntactic context (congruent or incongruent) and the length of the context (5- or 2-chord sequences) were crossed in a 2×2 factorial design. The experiment was run in the total absence of sound. We defined behavioural and ERP effects that were significantly stronger in 5- compared to 2-chord sequences as related to syntactic processing, because a longer harmonic context establishes more specific syntactic expectancies in the listener (Koelsch et al., 2000; Leino et al., 2007; Tillmann et al., 2003). In other words, the syntactic regulation of motor acts should be reflected in an interaction of Congruency \times Context.

As will be discussed in detail below, the EEG data together with the replication of the behavioural findings reported by Novembre and Keller (2011) show that the observation and imitation of syntactically organized sequences of musical acts evokes motor expectancies that influence skilled pianists' imitation of musical actions. The EEG data extend this finding further by demonstrating that this link between musical syntax and action concerns intermediate processing stages of (i) syntactic analysis of the observed movements and (ii) motor programming for accurate imitation, whereas ERPs related to initial perceptual and late executive stages of the task were not (or only minimally) influenced by the syntactic predictability of the chord sequences. Particularly, electrophysiological indices of action inhibition and reprogramming imply that the observer's motor system anticipates forthcoming actions during imitation based upon his or her long-term music-syntactic knowledge, i.e., suggesting an "embodied" processing of musical harmony. The spatial neighbourhood and dense interconnection (Nieuwenhuys et al., 2008; pp. 841–887) of goal-related action programming in premotor cortex and the syntax-related properties of inferior frontal areas (including Broca's area) might provide a neuroanatomical basis for this interaction. Although it remains to be clarified whether the motor system is informed by an extra syntactic parser (Tettamanti and Moro, 2012) or acts as an independent syntax processor (Pulvermüller and Fadiga, 2010), the combined behavioural and neurophysiological data support the workings of syntax to reach beyond the auditory perception of music to include the action domain.

4.1. Behavioural data

The analysis of the RTs revealed that motor demands differed between congruent and incongruent target chords (main effect of Congruency) but were comparable across 5- and 2-chord sequences as demonstrated by the overall similar RTs in both contexts (no main effect of Context). Most importantly, the imitation of congruent chords was generally faster than imitation of incongruent chords, particularly when the target chord was embedded into a 5-chord sequence, thus replicating the results of [Novembre and Keller \(2011\)](#). This pattern (i.e., an interaction of Congruency \times Context) is entirely in line with our above described criterion for syntax effects in action. Notably, the data suggest that the long syntactic context led to the priming and facilitation of the congruent target chord (i.e., speedup and higher accuracy; possibly reflecting a subliminal modulation of the motor system), instead of processing costs for the incongruent chord (i.e., slowing and lower accuracy) (see also [Tillmann et al., 2003](#)). Overall, this pattern indicates that the harmonic rules implied in the observed action sequences induced strong expectancies in the pianists about forthcoming motor acts and influenced their imitation performance. The EEG data described next, particularly the 2nd and 3rd phases, lead us to argue that this behavioural effect is based on a syntax-driven anticipation of motor programs during action imitation.

4.2. EEG data 1st phase – perceptual processes

In the first phase, i.e., during the presentation of the hand moving towards the target chord in the videos, incongruous chords evoked a more positive potential than congruous chords in both 5- and 2-chord sequences, although with slightly different scalp topography (see below). The early onset of the effect around 300 msec before the hand in the video reached the keys suggests that this ERP component reflects sensory processes related to the perceptually different hand trajectory towards congruent and incongruent targets, i.e., different finger positions, hand shapes and movement onsets (see [Methods](#)). The more pronounced posterior distribution of the effect in 2- compared to 5-chord sequences may reflect the stronger involvement of visual cortical areas due to particular attention of the pianists to these visuo-spatial cues during early stages of musical context build-up (i.e., after the presentation of just one chord when the sense of tonality is still weak) in order to quickly and accurately imitate the observed musical acts. Interestingly, the effect was left-frontally distributed in 5-chord sequences, which raises the possibility of a left inferior frontal source. The left inferior frontal gyrus (IFG) and adjacent ventral premotor cortex (vPMC) have been frequently discussed as a domain-general “grammar processor” ([Fadiga et al., 2009](#); [Fiebach and Schubotz, 2006](#); [Gelfand and Bookheimer, 2003](#); [Koelsch, 2005](#); [Patel, 2003](#); [Tettamanti and Weniger, 2006](#)) involved in the structural sequencing of language ([Friederici, 2011](#); [Grodzinsky and Santi, 2008](#)), music ([Maess et al., 2001](#); [Sammler et al., 2011](#)), and action ([Clerget et al., 2009](#); [Fazio et al., 2009](#)). In this function, and once a clear tonality is established like in the 5-chord sequences, the IFG/vPMC might provide top-down predictions about upcoming chords that include form-based

estimates of the hand trajectory (such as hand shape and finger configurations), i.e., syntactically relevant visuo-motor cues in the movement sequences that are checked against perceptually and motorically salient elements in the video (for similar form-based syntactic estimations in auditory and visual language comprehension, see [Dikker et al., 2009](#); [Herrmann et al., 2009](#)). However, at this stage of research the possibility of top-down syntactic influence on the early perceptual processing of musical actions must remain an interesting hypothesis to test in future studies.

4.3. EEG data 2nd phase – mismatch detection and response conflict

In the second phase, incongruous target chords evoked a right anterior negativity that was significantly stronger and emerged slightly earlier in 5- compared to 2-chord sequences. This interaction of Congruency \times Context is consistent with our criterion indicating syntactic analysis of music performance. Although it remains to be clarified whether this brain response is specifically tied (i) to the detection of the syntactic violation, (ii) to the perception of the incongruous action as a performance error, or (iii) to cognitive control processes related to the participant's own response, as will be explained in detail below, we will argue that all three views demonstrate the impact of musical grammar on musical actions.

- (i) **Detection of the syntactic violation.** As pointed out earlier (see [Introduction](#)), the auditory presentation of harmonic expectancy violations (such as the ones employed in the current study) evoke an ERAN, an index of (early) music-syntactic processing mediated by the IFG and superior temporal gyrus ([Garza Villarreal et al., 2011](#); [Koelsch, 2009](#); [Sammler et al., 2009](#)). The observed 2nd phase negativity is reminiscent of the ERAN in terms of sensitivity to music-syntactic violations and context length ([Koelsch et al., 2000](#); [Leino et al., 2007](#)), right-anterior scalp topography and polarity inversion at mastoid leads, although the 2nd phase negativity peaked earlier than the ERAN. This acceleration of the effect is most likely due to the ability of skilled pianists to anticipate the congruous or incongruous action outcome in the videos based on the hand trajectory towards the target chord. Pianists may actually use subtle cues in finger configuration – i.e., similar to “coarticulatory” information in speech – to recognize the (in)congruity of the forthcoming chord prior to the actual keystrokes, accounting for the pre-zero onset of the 2nd phase negativity, i.e., an earlier peak than the ERAN in the auditory modality (for a similar action anticipation ability in high-performing athletes, see [Aglioti et al., 2008](#)). Altogether, the above mentioned parallels (despite different timing) may cast the 2nd phase negativity as an equivalent of the ERAN in the visuo-motor modality, and thus provide indirect evidence for modality-independent processing of syntactic irregularities in rule-based harmonic sequences. The idea of such an abstract processing mechanism is supported by experiments showing that reading of unexpected notes in musical scores ([Gunter et al., 2003](#); [Schön and Besson, 2002](#)) evokes early negativities similar to those elicited

when hearing such violations (James et al., 2008; Koelsch, 2005; Patel et al., 1998). Note that the absence of a significant correlation between the negativity's amplitude and the auditory imagery score obtained in the debriefing suggests that the effect is not driven by participants' strategic use of auditory images (Hasegawa et al., 2004; Haslinger et al., 2005; Hubbard, 2010) related to the visually presented stimuli. It more likely reflects the work of a polymodal musical syntax processor that operates on different expressions (i.e., auditory, visual or sensorimotor) of the same syntactic structure. Nevertheless, the possible co-occurrence of auditory images in the context of our motor task is an issue that deserves consideration and is more extensively discussed below.

- (ii) **Perception of a performance error.** In addition to modality-unspecific syntactic processes, the 2nd phase negativity might also reflect an error-related negativity (ERN), or error negativity (N_e), evoked if the incongruous actions in the videos were perceived as erroneous actions (although they were not erroneous *per se*, just unexpected). The ERN is evoked after self-generated errors (Falkenstein et al., 1990; Gehring et al., 1993; Herrojo Ruiz et al., 2009; Maidhof et al., 2010) as well as errors observed in another person (Miltner et al., 2004; van Schie et al., 2004) suggesting that the observer's own action control system internally simulates the required and perceived action (Iacoboni, 2005; Rizzolatti and Sinigaglia, 2010). The ERN is largest at fronto-central recording sites and is interpreted as the mismatch detection between the actual (i.e., incorrect) action compared to the required (i.e., correct) action (Falkenstein et al., 1990). Notably, the ERN amplitude depends on how well the representation of the required action is established (Falkenstein, 2004), and how strongly the dissimilarity between appropriate and actual response is perceived (Arbel and Donchin, 2011; Bernstein et al., 1995). This property of the ERN could account for its higher amplitude in our 5-chord sequences, which led to a stronger representation of the (required) congruous chord and a greater salience of the incongruous chord, than in 2-chord sequences (Bigand and Parncutt, 1999; Koelsch et al., 2000). Note that such an interpretation puts music-syntactic processes at the origin of a brain response evoked by the observation of an unexpected act. In other words, this finding would demonstrate that syntactic knowledge influences the way in which we perceive another person's action, possibly via simulation of this action in our own motor (syntactic) system (van Schie et al., 2004; Wilson and Knoblich, 2005).
- (iii) **Action control processes.** Beyond these relationships of the 2nd phase negativity to the observation of the incongruous chords in the videos (i.e., ERAN and observer ERN), this brain potential might also be related to the *participants' own response*. The imitation of the incongruent action sequences may have triggered cognitive control processes such as detection of response conflict and response inhibition to override the prepotent, syntax-driven impulse to produce a congruent sequence ending. In fact, a fronto-centrally distributed N2c or no-go N2 is usually elicited in response priming tasks whenever advance information is invalid (Kopp and Wessel, 2010;

Leuthold, 2004) and a planned response needs to be withheld (Bruin and Wijers, 2002; Falkenstein et al., 1999; Pfefferbaum et al., 1985). It has been suggested that these negativities reflect a control signal that is issued whenever response conflict is detected and is used to temporarily suppress the input to the motor execution system (Stürmer et al., 2002) to adjust or remedy ongoing but inappropriate actions (Kopp et al., 1996). Notably, the N2c amplitude (along with RT costs) has been shown to increase with stronger degree of processing conflict (Botvinick et al., 2001), a condition that is fulfilled especially in the 5-chord sequences. Most importantly, this interpretation does not only imply that the observed syntactically structured sequence of acts triggers an internal representation of the analogous motoric sequence (Rizzolatti and Sinigaglia, 2010). It also alludes to the future-oriented processing of action sequences proposed by incremental models of response preparation (Palmer and Pfordresher, 2003) and, most intriguingly, suggests the automatic advance programming of forthcoming actions (i.e., the congruent target chord) once they can be predicted from the syntactic context (Borroni et al., 2005; Kilner et al., 2004). In other words, the present data argue for an anticipated resonant response in the observer's motor system that does not immediately depend on the realization of the movement in the videos but on context-dependent predictions based on the long-term syntactic knowledge of the pianists.

Taken together, the 2nd phase negativity may be interpreted as an ERAN, an observer ERN, or an N2c/no-go N2 (or a superposition of them; for an overview, see Folstein and Van Petten, 2008), clearly calling for further studies (e.g., with passive observation instead of imitation). Note, importantly, that all three views, irrespective of functional interpretation, demonstrate the operation of musical grammar in the domain of action (observation or programming). On a more abstract level, this triad of processes potentially represents interrelated, syntax-based mechanisms that may play a role during joint musical performance, such as the syntactically guided and modality-unspecific moment-to-moment evaluation and anticipation of other players' musical actions, as well as the syntax-driven programming and flexible revocation of one's own motor acts in concert with other musicians' performances.

4.4. EEG data 3rd phase – response (re-)programming

In the third phase, incongruous (compared to congruous) chords in 5-chord sequences evoked a slightly right-lateralized posterior negativity that was not observed in 2-chord sequences and is therefore – in line with our definition of a Congruency \times Context interaction – most likely related to the syntactic regulation of the musical performance. More precisely, this effect may reflect mechanisms of movement reprogramming following the cancellation of the syntactically prepotent response, i.e., the programming of the incongruent chord in the face of the more dominant congruent chord (Mars et al., 2007). Response priming paradigms comparing the execution of an action after neutral,

valid, and invalid primes have consistently observed large RT costs along with a slightly right-dominant centroparietal negativity in invalid compared to valid and neutral precue conditions (Leuthold and Jentzsch, 2002; Vidal et al., 1995). Considering our 2-chord, congruent 5-chord, and incongruent 5-chord sequences as equivalents of the neutral, valid, and invalid conditions, respectively, this finding closely mirrors our results and thus supports the interpretation as response reprogramming. Importantly, the reprogramming of an action provides indirect support for the prior activation of an alternative motor program, i.e., the syntax-driven motoric anticipation of the congruent chord during action observation.

It should be noted that this interpretation does not necessarily exclude the simultaneous auditory anticipation of the congruent chord. In fact, a number of functional magnetic resonance imaging (fMRI) studies showed auditory activations during the observation (Hasegawa et al., 2004; Haslinger et al., 2005) or execution of mute piano performance (Bangert et al., 2006; Lahav et al., 2007; Zatorre et al., 2007), making simultaneous processing in auditory and motor areas likely in the context of the present task. However, we argue that the posterior scalp distribution of the effect is atypical for purely auditory-based potentials and rather compatible with a stronger involvement of the centroparietal motor (as opposed to auditory) system.

No significant centroparietal potential differences were found between congruent and incongruent endings in the 2-chord sequences, suggesting that their programming was less predictive, but rather reactive (i.e., without prior activation of the congruent motor program and thus no need of reprogramming in case of an incongruent chord). Instead, incongruent (compared to congruent) chords evoked a fronto-central negativity with slight right-hemispheric predominance. Note that a similar effect may have been also evoked by 5-chord sequences, and superposed by the prominent posterior negativity. As mentioned earlier (see *Methods*), 2-chord sequences – although syntactically less predictable than 5-chord sequences – were not syntax-free, as they contained very common chord pairs in Western tonal music. Therefore, this effect may reflect some delayed (in 2-chord sequences) or extended (in 5-chord sequences) music-syntactic processing of the visually perceived chord, akin to the negativities observed in the auditory (Koelsch, 2009; Koelsch et al., 2000) and visual modality (Gunter et al., 2003; Schön and Besson, 2002). Alternatively, given that incongruent target chords were equally higher in motoric (i.e., kinematic) complexity than congruent chords in both 5- and 2-chord sequences (see *Methods*), this ERP may be interpreted as a correlate of the greater effort during movement programming, possibly as one aspect of the contingent negative variation (CNV) of externally cued movements (Walter et al., 1964). Consistent with our results, the CNV is fronto-centrally distributed, is assumed to reflect (amongst others) preparatory motor activity, and exhibits larger amplitudes for motorically more complex and less familiar movements (Cui et al., 2000; Kranczioch et al., 2010) (for an overview, see Brunia, 2003).

4.5. EEG data 4th phase – response execution

In the 4th phase, incongruent chords evoked slightly stronger positive potentials than congruent chords at left and right

lateral electrodes in both 5- and 2-chord sequences. This effect may reflect syntactic reanalysis and repair processes (because both 5- and 2-chord sequences contained syntactic information; see above) that are commonly associated with a P300 (Janata, 1995), P600 (Patel et al., 1998), or late positive component (LPC; Besson and Faita, 1995), although those are usually centro-posteriorly distributed. Alternatively, the latency of the potential between 400 and 1000 msec corresponds to the latency of the participants' motor responses (Fig. 2), suggesting a link of the ERP with the execution of the final chords. Congruent chords required less cognitive and motor programming effort to be executed (e.g., due to a lower number of black keys; see *Methods*) and had significantly shorter RTs than incongruent chords (Fig. 2), possibly leading to an earlier onset of movement-related cortical potentials. These are negative shifts preceding movement onset, such as the slowly rising readiness potential (Bereitschaftspotential) in single self-paced movements (Kornhuber and Deecke, 1965; Shibasaki and Hallett, 2006) or a steady-state movement-related cortical negativity in repetitive movements (Gerloff et al., 1997). Notably, an earlier onset of the negative shift in the congruent conditions would show up as a relative positivity, as the one observed in the present experiment. Furthermore, complex movements (such as in piano playing) are known to recruit not only contra- but also ipsilateral motor areas (Haaland et al., 2004; Nishitani and Hari, 2000; Verstynen et al., 2005), consistent with the bilateral topography of the observed effect.

4.6. Is music-syntactic processing “embodied”?

The present data suggest an influence of music-syntactic knowledge on piano performance, as evidenced by longer RTs, higher error rates, and ERPs reflecting the cancellation (2nd phase) and reprogramming (3rd phase) of a prepotent motor response during imitation of syntactically incongruent chord sequences. In other words, the observed and imitated musical acts evoked an anticipatory response in the pianist's motor system that may be taken as an “embodied” correlate of music-syntactic processing.

Theories of embodiment ground cognitive processes, such as action perception (Wilson and Knoblich, 2005), language processing (Glenberg and Gallese, 2012; Glenberg and Kaschak, 2002; Pulvermüller and Fadiga, 2010), or social cognition (Gallese, 2007; Gallese and Sinigaglia, 2011; Goldman and de Vignemont, 2009), in a bodily format (for reviews, see Barsalou, 2008; Wilson, 2002). Recent attempts in the language domain to unravel bodily representations of syntactic processes in the sensorimotor (“mirror”) system (Pulvermüller and Fadiga, 2010) directly relate to our results, due to the structural and neurofunctional similarities of language and music (see *Introduction*). It should be noted though that our study – although it strongly suggests a role of the sensorimotor system in establishing a bodily image of forthcoming motor acts based on syntactic information – did not explicitly manipulate recursion and thus cannot clarify whether this predictive process involves merely linear chord transitions and overlearned syntactic patterns, or full-fledged recursive dependencies and the hierarchical depth structure of harmony (Katz and Pesetsky, 2011; Rohrmeier, 2011). Overall,

the cognitive instantiation of recursion in both music and action remains an issue that deserves further attention to build a clear theoretical framework on aspects of syntax that are specific to language or shared between domains.

It is conceivable that recursive hierarchy in action cannot be processed by the sensorimotor system alone (i.e., via a “mirror” mechanism) but requires reciprocal interactions with an external syntactic processor. This possibility would be in line with associative (as opposed to non-associative, “mirror”) kinds of embodied cognition theories in a way that a polymodal syntax processor informs and is informed by the sensorimotor system in two-way feedback loops, (back)translating syntactic processes from/into a bodily format (Tettamanti and Moro, 2012). The spatial neighbourhood and dense interconnection (Nieuwenhuys et al., 2008; pp. 841–887) of premotor cortex and putative syntactic properties of inferior frontal regions (including Broca’s area) make such an interaction anatomically plausible. Furthermore, deficits in processing or learning compositional action sequences after damage to left inferior frontal regions (Fazio et al., 2009) or transcranial magnetic stimulation over Broca’s area (Clerget et al., 2011, 2009) is compatible with this notion. The question which level of syntactic complexity the bodily senses themselves are able to parse remains an interesting topic for future research.

5. Conclusion

In sum, the present study shows that music-syntactic knowledge triggers the internal build-up of action expectancies in trained musicians when they imitate other players’ musical actions. In particular, the data suggest that the prediction of musical acts involves a resonant response in the observer’s motor system, i.e., an “embodied” application of the rules determining the position of chords within harmonic sequences (that in turn influences music performance). It is well-known that the motor system is endowed with the ability to predict abstract events (Schubotz, 2007) and to constantly set up anticipatory models of other persons’ actions (Borroni et al., 2005; Iacoboni et al., 2005; Kilner et al., 2004; van Schie et al., 2004) in order to predict other’s intentions ahead of their behavioural realization (for related monkey work, see Umiltà et al., 2001). The novel finding of the present study is that – in a musical context and despite the absence of auditory information – the motor system’s anticipatory potential is strongly based upon the observer’s long-term music-syntactic knowledge. The present study thus shows that the notion of syntax not only applies to music in the auditory modality but transfers – in trained musicians – to an embodied processing of a “grammar of musical action”.

Acknowledgements

First and foremost we thank all pianists who participated in the study. We further thank Katrin Cunitz, Maria Bader, Irene Nenoff, and Katharina Fath for their great help with data acquisition, Kerstin Träger for inviting the pianists, Jan Bergmann and Sven Gutekunst for technical support, Makiko

Sadakata for creating the video stimuli, and three anonymous reviewers for helpful comments.

REFERENCES

- Abrams DA, Bhatara A, Ryali S, Balaban E, Levitin DJ, and Menon V. Decoding temporal structure in music and speech relies on shared brain resources but elicits different fine-scale spatial patterns. *Cerebral Cortex*, 21(7): 1507–1518, 2011.
- Aglioti SM, Cesari P, Romani M, and Urgesi C. Action anticipation and motor resonance in elite basketball players. *Nature Neuroscience*, 11(9): 1109–1116, 2008.
- Allen K, Ibara S, Seymour A, Cordova N, and Botvinick M. Abstract structural representations of goal-directed behavior. *Psychological Science*, 21(10): 1518–1524, 2010.
- Arbel Y and Donchin E. How large the sin? A study of the event related potentials elicited by errors of varying magnitude. *Psychophysiology*, 48(12): 1611–1620, 2011.
- Bahlmann J, Schubotz RI, Mueller JL, Koester D, and Friederici AD. Neural circuits of hierarchical visuo-spatial sequence processing. *Brain Research*, 1298: 161–170, 2009.
- Bangert M, Peschel T, Schlaug G, Rotte M, Drescher D, Hinrichs H, et al. Shared networks for auditory and motor processing in professional pianists: Evidence from fMRI conjunction. *NeuroImage*, 30(3): 917–926, 2006.
- Barsalou LW. Grounded cognition. *Annual Review of Psychology*, 59: 617–645, 2008.
- Bernstein PS, Scheffers MK, and Coles MG. “Where did I go wrong?” A psychophysiological analysis of error detection. *Journal of Experimental Psychology: Human Perception and Performance*, 21(6): 1312–1322, 1995.
- Besson M and Faita F. An event-related (ERP) study of musical expectancy: Comparison of musicians with nonmusicians. *Journal of Experimental Psychology: Human Perception and Performance*, 21(6): 1278–1296, 1995.
- Bigand E and Parncutt R. Perceiving musical tension in long chord sequences. *Psychological Research*, 62(4): 237–254, 1999.
- Borroni P, Montagna M, Cerri G, and Baldissera F. Cyclic time course of motor excitability modulation during the observation of a cyclic hand movement. *Brain Research*, 1065(1–2): 115–124, 2005.
- Bortz JJ and Döring N. *Forschungsmethoden und Evaluation: Für Human- und Sozialwissenschaftler*. Berlin: Springer, 2003.
- Botvinick MM, Braver TS, Barch DM, Carter CS, and Cohen JD. Conflict monitoring and cognitive control. *Psychological Review*, 108(3): 624–652, 2001.
- Bruin KJ and Wijers AA. Inhibition, response mode, and stimulus probability: A comparative event-related potential study. *Clinical Neurophysiology*, 113(7): 1172–1182, 2002.
- Brunia CHM. CNV and SNP: Indices of anticipatory behavior. In Jahanshahi M and Hallett M (Eds), *The Bereitschaftspotential – Movement-related Cortical Potentials*. New York: Kluwer Academic/Plenum Publishers, 2003: 207–227.
- Buccino G, Lui F, Canessa N, Patteri I, Lagravinese G, Benuzzi F, et al. Neural circuits involved in the recognition of actions performed by nonconspecifics: An fMRI study. *Journal of Cognitive Neuroscience*, 16(1): 114–126, 2004.
- Chomsky AN. *Syntactic Structures*. Den Haag: Mouton, 1957.
- Chomsky AN. *Knowledge of Language: Its Nature, Origin and Use*. New York: Praeger, 1986.
- Chomsky AN. *The Minimalist Program*. Cambridge, MA: MIT Press, 1995.
- Clerget E, Poncin W, Fadiga L, and Olivier E. Role of Broca’s area in implicit motor skill learning: Evidence from continuous theta-burst magnetic stimulation. *Journal of Cognitive Neuroscience*, 24(1): 80–92, 2011.

- Clerget E, Winderickx A, Fadiga L, and Olivier E. Role of Broca's area in encoding sequential human actions: A virtual lesion study. *NeuroReport*, 20(16): 1496–1499, 2009.
- Cui RQ, Egkher A, Huter D, Lang W, Lindinger G, and Deecke L. High resolution spatiotemporal analysis of the contingent negative variation in simple or complex motor tasks and a non-motor task. *Clinical Neurophysiology*, 111(10): 1847–1859, 2000.
- Dehaene S. *The Number Sense – How the Mind Creates Mathematics*. New York: Oxford University Press, 1997.
- Delorme A and Makeig S. EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1): 9–21, 2004.
- Dikker S, Rabagliati H, and Pykkänen L. Sensitivity to syntax in visual cortex. *Cognition*, 110(3): 293–321, 2009.
- Drost UC, Rieger M, Brass M, Gunter TC, and Prinz W. When hearing turns into playing: Movement induction by auditory stimuli in pianists. *The Quarterly Journal of Experimental Psychology Section A*, 58(8): 1376–1389, 2005.
- Fadiga L, Craighero L, and D'Ausilio A. Broca's area in language, action, and music. *Annals of the New York Academy of Sciences*, 1169: 448–458, 2009.
- Falkenstein M. ERP correlates of erroneous performance. In Ullsperger M and Falkenstein M (Eds), *Errors, Conflicts, and the Brain. Current Opinions on Performance Monitoring*. Leipzig: Max Planck Institute for Human Cognitive and Brain Sciences, 2004: 5–14.
- Falkenstein M, Hohnsbein J, Hoormann J, and Blanke L. Effects of errors in choice reaction tasks on the ERP under focused and divided attention. In Brunia C, Gaillard A, and Kok A (Eds), *Psychophysiological Brain Research*. Tilburg: Tilburg University Press, 1990.
- Falkenstein M, Hoormann J, and Hohnsbein J. ERP components in Go/Nogo tasks and their relation to inhibition. *Acta Psychologica*, 101(2–3): 267–291, 1999.
- Farag C, Troiani V, Bonner M, Powers C, Avants B, Gee J, et al. Hierarchical organization of scripts: Converging evidence from fMRI and frontotemporal degeneration. *Cerebral Cortex*, 20(10): 2453–2463, 2010.
- Fazio P, Cantagallo A, Craighero L, D'Ausilio A, Roy AC, Pozzo T, et al. Encoding of human action in Broca's area. *Brain*, 132: 1980–1988, 2009.
- Fedorenko E, Patel A, Casasanto D, Winawer J, and Gibson E. Structural integration in language and music: Evidence for a shared system. *Memory & Cognition*, 37(1): 1–9, 2009.
- Fiebach CJ and Schubotz RI. Dynamic anticipatory processing of hierarchical sequential events: A common role for Broca's area and ventral premotor cortex across domains? *Cortex*, 42(4): 499–502, 2006.
- Fitch WT. Unity and diversity in human language. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1563): 376–388, 2011.
- Folstein JR and Van Petten C. Influence of cognitive control and mismatch on the N2 component of the ERP: A review. *Psychophysiology*, 45(1): 152–170, 2008.
- Friederici AD. The brain basis of language processing: From structure to function. *Physiological Reviews*, 91(4): 1357–1392, 2011.
- Friedrich R and Friederici AD. Mathematical logic in the human brain: Syntax. *Plos One*, 4(5): e5599, 2009.
- Gallese V. Before and below theory of mind: Embodied simulation and the neural correlates of social cognition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1480): 659–669, 2007.
- Gallese V and Sinigaglia C. What is so special about embodied simulation? *Trends in Cognitive Sciences*, 15(11): 512–518, 2011.
- Garza Villarreal EA, Brattico E, Leino S, Ostergaard L, and Vuust P. Distinct neural responses to chord violations: A multiple source analysis study. *Brain Research*, 1389: 103–114, 2011.
- Gehring WJ, Goss B, Coles MGH, Meyer DE, and Donchin E. A neural system for error-detection and compensation. *Psychological Science*, 4(6): 385–390, 1993.
- Gelfand JR and Bookheimer SY. Dissociating neural mechanisms of temporal sequencing and processing phonemes. *Neuron*, 38(5): 831–842, 2003.
- Gerloff C, Toro C, Uenishi N, Cohen LG, Leocani L, and Hallett M. Steady-state movement-related cortical potentials: A new approach to assessing cortical activity associated with fast repetitive finger movements. *Electroencephalography and Clinical Neurophysiology*, 102(2): 106–113, 1997.
- Gibbs RWJ. *Embodiment and Cognitive Science*. Cambridge: Cambridge University Press, 2006.
- Glenberg AM and Gallese V. Action-based language: A theory of language acquisition, comprehension, and production. *Cortex*, 48(7): 905–922, 2012.
- Glenberg AM and Kaschak MP. Grounding language in action. *Psychonomic Bulletin & Review*, 9(3): 558–565, 2002.
- Goldman A and de Vignemont F. Is social cognition embodied? *Trends in Cognitive Sciences*, 13(4): 154–159, 2009.
- Greenfield PM. Language, tools, and brain – The ontogeny and phylogeny of hierarchically organized sequential behavior. *Behavioral and Brain Sciences*, 14(4): 531–550, 1991.
- Grodzinsky Y. The neurology of syntax: Language use without Broca's area. *Behavioral and Brain Sciences*, 23(1): 1–71, 2000.
- Grodzinsky Y and Santi A. The battle for Broca's region. *Trends in Cognitive Sciences*, 12(12): 474–480, 2008.
- Guerra-Filho G and Aloimonos Y. The syntax of human actions and interactions. *Journal of Neurolinguistics*, 25(5): 500–514, 2012.
- Gunter TC, Schmidt BH, and Besson M. Let's face the music: A behavioral and electrophysiological exploration of score reading. *Psychophysiology*, 40(5): 742–751, 2003.
- Haaland KY, Elsinger CL, Mayer AR, Durgerian S, and Rao SM. Motor sequence complexity and performing hand produce differential patterns of hemispheric lateralization. *Journal of Cognitive Neuroscience*, 16(4): 621–636, 2004.
- Hasegawa T, Matsuki K, Ueno T, Maeda Y, Matsue Y, Konishi Y, et al. Learned audio–visual cross-modal associations in observed piano playing activate the left planum temporale. An fMRI study. *Cognitive Brain Research*, 20(3): 510–518, 2004.
- Haslinger B, Erhard P, Altenmüller E, Schroeder U, Boecker H, and Ceballos-Baumann AO. Transmodal sensorimotor networks during action observation in professional pianists. *Journal of Cognitive Neuroscience*, 17(2): 282–293, 2005.
- Hauser MD, Chomsky N, and Fitch WT. The faculty of language: What is it, who has it, and how did it evolve? *Science*, 298(5598): 1569–1579, 2002.
- Herrmann B, Maess B, Hasting AS, and Friederici AD. Localization of the syntactic mismatch negativity in the temporal cortex: An MEG study. *NeuroImage*, 48(3): 590–600, 2009.
- Herrero Ruiz M, Jabusch HC, and Altenmüller E. Detecting wrong notes in advance: Neuronal correlates of error monitoring in pianists. *Cerebral Cortex*, 19(11): 2625–2639, 2009.
- Hubbard TL. Auditory imagery: Empirical findings. *Psychological Bulletin*, 136(2): 302–329, 2010.
- Iacoboni M. Neural mechanisms of imitation. *Current Opinion in Neurobiology*, 15(6): 632–637, 2005.
- Iacoboni M, Molnar-Szakacs I, Gallese V, Buccino G, Mazziotta JC, and Rizzolatti G. Grasping the intentions of others with one's own mirror neuron system. *Plos Biology*, 3(3): e79, 2005.
- James CE, Britz J, Vuilleumier P, Hauert CA, and Michel CM. Early neuronal responses in right limbic structures mediate harmony incongruity processing in musical experts. *NeuroImage*, 42(4): 1597–1608, 2008.
- Janata P. ERP measures assay the degree of expectancy violation of harmonic contexts in music. *Journal of Cognitive Neuroscience*, 7(2): 153–164, 1995.

- Jentschke S and Koelsch S. Musical training modulates the development of syntax processing in children. *NeuroImage*, 47(2): 735–744, 2009.
- Jentschke S, Koelsch S, and Friederici AD. Investigating the relationship of music and language in children: Influences of musical training and language impairment. *Annals of the New York Academy of Sciences*, 1060: 231–242, 2005.
- Kaan E and Swaab TY. The brain circuitry of syntactic comprehension. *Trends in Cognitive Sciences*, 6(8): 350–356, 2002.
- Katz J and Pesetsky D. The Identity Thesis of Language and Music. *lingBuzz/000959*, 2011.
- Kilner JM, Vargas C, Duval S, Blakemore SJ, and Sirigu A. Motor activation prior to observation of a predicted movement. *Nature Neuroscience*, 7(12): 1299–1301, 2004.
- Koelsch S. Neural substrates of processing syntax and semantics in music. *Current Opinion in Neurobiology*, 15(2): 207–212, 2005.
- Koelsch S. Music-syntactic processing and auditory memory: Similarities and differences between ERAN and MMN. *Psychophysiology*, 46(1): 179–190, 2009.
- Koelsch S. Toward a neural basis of music perception – A review and updated model. *Frontiers in Psychology*, 2: 1–20, 2011.
- Koelsch S, Gunter T, Friederici AD, and Schröger E. Brain indices of music processing: “Nonmusicians” are musical. *Journal of Cognitive Neuroscience*, 12(3): 520–541, 2000.
- Koelsch S, Gunter TC, Wittfoth M, and Sammler D. Interaction between syntax processing in language and in music: An ERP study. *Journal of Cognitive Neuroscience*, 17(10): 1565–1577, 2005.
- Kopp B, Rist F, and Mattler U. N200 in the flanker task as a neurobehavioral tool for investigating executive control. *Psychophysiology*, 33(3): 282–294, 1996.
- Kopp B and Wessel K. Event-related brain potentials and cognitive processes related to perceptual-motor information transmission. *Cognitive, Affective, Behavioral Neuroscience*, 10(2): 316–327, 2010.
- Kornhuber HH and Deecke L. Hirnpotentialänderungen bei Willkürbewegungen und passiven Bewegungen des Menschen: Bereitschaftspotential und reafferente Potentiale. *Pflüger's Archiv für die gesamte Physiologie des Menschen und der Tiere*, 284(1): 1–17, 1965.
- Krancioch C, Mathews S, Dean P, and Sterr A. Task complexity differentially affects executed and imagined movement preparation: Evidence from movement-related potentials. *Plos One*, 5(2) 2010.
- Lahav A, Saltzman E, and Schlaug G. Auditory operations of the mirror neuron system: Action recognition network while listening to newly-acquired actions. *Neurology*, 68(12): A106, 2007.
- Lashley KS. The problem of serial order in behavior. In Jeffress LA (Ed), *Cerebral Mechanisms in Behavior*. New York: Wiley, 1951: 112–131.
- Leino S, Brattico E, Tervaniemi M, and Vuust P. Representation of harmony rules in the human brain: Further evidence from event-related potentials. *Brain Research*, 1142: 169–177, 2007.
- Lerdahl F and Jackendoff R. *A Generative Theory of Tonal Music*. Cambridge: MIT Press, 1983.
- Leuthold H. Action monitoring and control: In search of common principles. In Ullsperger M and Falkenstein M (Eds), *Errors, Conflicts, and the Brain. Current Opinions on Performance Monitoring*. Leipzig: Max Planck Institute for Human Cognitive and Brain Sciences, 2004: 70–76.
- Leuthold H and Jentsch I. Spatiotemporal source localisation reveals involvement of medial premotor areas in movement reprogramming. *Experimental Brain Research*, 144(2): 178–188, 2002.
- Maess B, Koelsch S, Gunter TC, and Friederici AD. Musical syntax is processed in Broca's area: An MEG study. *Nature Neuroscience*, 4(5): 540–545, 2001.
- Maidhof C, Vavatzanidis N, Prinz W, Rieger M, and Koelsch S. Processing expectancy violations during music performance and perception: An ERP study. *Journal of Cognitive Neuroscience*, 22(10): 2401–2413, 2010.
- Marin MM. Effects of early musical training on musical and linguistic syntactic abilities. *Annals of the New York Academy of Sciences*, 1169: 187–190, 2009.
- Mars RB, Piekema C, Coles MG, Hulstijn W, and Toni I. On the programming and reprogramming of actions. *Cerebral Cortex*, 17(12): 2972–2979, 2007.
- Michel CM, Murray MM, Lantz G, Gonzalez S, Spinelli L, and Grave de Peralta R. EEG source imaging. *Clinical Neurophysiology*, 115(10): 2195–2222, 2004.
- Miltner WHR, Brauer J, Hecht H, Trippe R, and Coles MGH. Parallel brain activity for self-generated and observed errors. In Ullsperger M and Falkenstein M (Eds), *Errors, Conflict, and the Brain. Current Opinions on Performance Monitoring*. Leipzig: MPI of Cognitive Neuroscience, 2004: 124–129.
- Monti MM, Parsons L, and Osherson DN. The boundaries of language and thought in deductive inference. *Proceedings of the National Academy of Sciences of the United States of America*, 106(30): 12554–12559, 2009.
- Moro A. *The Boundaries of Babel – The Brain and the Enigma of Impossible Languages*. Cambridge, MA: MIT Press, 2008.
- Moro A, Tettamanti M, Perani D, Donati C, Cappa SF, and Fazio F. Syntax and the brain: Disentangling grammar by selective anomalies. *NeuroImage*, 13(1): 110–118, 2001.
- Musso M, Moro A, Glauche V, Rijntjes M, Reichenbach J, Büchel C, et al. Broca's area and the language instinct. *Nature Neuroscience*, 6(7): 774–781, 2003.
- Nieuwenhuys R, Voogd J, and van Huijzen C. *The Human Central Nervous System*. Berlin: Springer, 2008.
- Nishitani N and Hari R. Temporal dynamics of cortical representation for action. *Proceedings of the National Academy of Sciences of the United States of America*, 97(2): 913–918, 2000.
- Novembre G and Keller PE. A grammar of action generates predictions in skilled musicians. *Consciousness and Cognition*, 20(4): 1232–1243, 2011.
- Núñez-Peña MI and Honrubia-Serrano ML. P600 related to rule violation in an arithmetic task. *Cognitive Brain Research*, 18(2): 130–141, 2004.
- Pallier C, Devauchelle AD, and Dehaene S. Cortical representation of the constituent structure of sentences. *Proceedings of the National Academy of Sciences of the United States of America*, 108(6): 2522–2527, 2011.
- Palmer C and Pfordresher PQ. Incremental planning in sequence production. *Psychological Review*, 110(4): 683–712, 2003.
- Palmer C and van de Sande C. Units of knowledge in music performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19(2): 457–470, 1993.
- Palmer C and van de Sande C. Range of planning in music performance. *Journal of Experimental Psychology: Human Perception and Performance*, 21(5): 947–962, 1995.
- Pastrak K and Aloimonos Y. The minimalist grammar of action. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1585): 103–117, 2012.
- Patel AD. Language, music, syntax and the brain. *Nature Neuroscience*, 6(7): 674–681, 2003.
- Patel AD, Gibson E, Ratner J, Besson M, and Holcomb PJ. Processing syntactic relations in language and music: An event-related potential study. *Journal of Cognitive Neuroscience*, 10(6): 717–733, 1998.
- Patel AD, Iversen JR, Wassenaar M, and Hagoort P. Musical syntax processing in agrammatic Broca's aphasia. *Aphasiology*, 22(7–8): 776–789, 2008.
- Perani D, Fazio F, Borghese NA, Tettamanti M, Ferrari S, Decety J, et al. Different brain correlates for watching real and virtual hand actions. *NeuroImage*, 14(3): 749–758, 2001.

- Pfefferbaum A, Ford JM, Weller BJ, and Kopell BS. ERPs to response production and inhibition. *Electroencephalography and Clinical Neurophysiology*, 60(5): 423–434, 1985.
- Pulvermüller F and Fadiga L. Active perception: Sensorimotor circuits as a cortical basis for language. *Nature Reviews Neuroscience*, 11(5): 351–360, 2010.
- Restle F. Theory of serial pattern learning: Structural trees. *Psychological Review*, 77(6): 481–495, 1970.
- Rizzolatti G and Sinigaglia C. The functional role of the parieto-frontal mirror circuit: Interpretations and misinterpretations. *Nature Reviews Neuroscience*, 11(4): 264–274, 2010.
- Rogalsky C, Rong F, Saberi K, and Hickok G. Functional anatomy of language and music perception: Temporal and structural factors investigated using functional magnetic resonance imaging. *Journal of Neuroscience*, 31(10): 3843–3852, 2011.
- Rohrmeier M. Towards a generative syntax of tonal harmony. *Journal of Mathematics and Music*, 5(1): 35–53, 2011.
- Sammler D, Koelsch S, Ball T, Brandt A, Elger CE, Friederici AD, et al. Overlap of musical and linguistic syntax processing: Intracranial ERP evidence. *Annals of the New York Academy of Sciences*, 1169: 494–498, 2009.
- Sammler D, Koelsch S, and Friederici AD. Are left fronto-temporal brain areas a prerequisite for normal music-syntactic processing? *Cortex*, 47(6): 659–673, 2011.
- Scheepers C, Sturt P, Martin CJ, Myachykov A, Teevan K, and Viskupova I. Structural priming across cognitive domains: From simple arithmetics to relative-clause attachment. *Psychological Science*, 22(10): 1319–1326, 2011.
- Schön D and Besson M. Processing pitch and duration in music reading: A RT–ERP study. *Neuropsychologia*, 40(7): 868–878, 2002.
- Schubotz RI. Prediction of external events with our motor system: Towards a new framework. *Trends in Cognitive Sciences*, 11(5): 211–218, 2007.
- Sharbrough F, Chatrian G-E, Lesser RP, Lüders H, Nuwer M, and Picton TW. American-electroencephalographic-society guidelines for standard electrode position nomenclature. *Journal of Clinical Neurophysiology*, 8(2): 200–202, 1991.
- Shibasaki H and Hallett M. What is the Bereitschaftspotential? *Clinical Neurophysiology*, 117(11): 2341–2356, 2006.
- Slevc LR, Rosenberg JC, and Patel AD. Making psycholinguistics musical: Self-paced reading time evidence for shared processing of linguistic and musical syntax. *Psychonomic Bulletin & Review*, 16(2): 374–381, 2009.
- Steinbeis N and Koelsch S. Shared neural resources between music and language indicate semantic processing of musical tension-resolution patterns. *Cerebral Cortex*, 18(5): 1169–1178, 2008.
- Stevens JA, Fonlupt P, Shiffrar M, and Decety J. New aspects of motion perception: Selective neural encoding of apparent human movements. *NeuroReport*, 11(1): 109–115, 2000.
- Stürmer B, Leuthold H, Soetens E, Schröter H, and Sommer W. Control over location-based response activation in the Simon task: Behavioral and electrophysiological evidence. *Journal of Experimental Psychology: Human Perception and Performance*, 28(6): 1345–1363, 2002.
- Tettamanti M and Moro A. Can syntax appear in a mirror (system)? *Cortex*, 48(7): 923–935, 2012.
- Tettamanti M, Rotondi I, Perani D, Scotti G, Fazio F, Cappa SF, et al. Syntax without language: Neurobiological evidence for cross-domain syntactic computations. *Cortex*, 45(7): 825–838, 2009.
- Tettamanti M and Weniger D. Broca's area: A supramodal hierarchical processor? *Cortex*, 42(4): 491–494, 2006.
- Tillmann B, Janata P, Birk J, and Bharucha JJ. The costs and benefits of tonal centers for chord processing. *Journal of Experimental Psychology: Human Perception and Performance*, 29(2): 470–482, 2003.
- Umliltà MA, Kohler E, Gallese V, Fogassi L, Fadiga L, Keysers C, et al. I know what you are doing: A neurophysiological study. *Neuron*, 31(1): 155–165, 2001.
- van Schie HT, Mars RB, Coles MGH, and Bekkering H. Modulation of activity in medial frontal and motor cortices during error observation. *Nature Neuroscience*, 7(5): 549–554, 2004.
- van Schie HT, Toni I, and Bekkering H. Comparable mechanisms for action and language: Neural systems behind intentions, goals, and means. *Cortex*, 42(4): 495–498, 2006.
- Verstynen T, Diedrichsen J, Albert N, Aparicio P, and Ivry RB. Ipsilateral motor cortex activity during unimanual hand movements relates to task complexity. *Journal of Neurophysiology*, 93(3): 1209–1222, 2005.
- Vidal F, Bonnet M, and Macar F. Programming the duration of a motor sequence: Role of the primary and supplementary motor areas in man. *Experimental Brain Research*, 106(2): 339–350, 1995.
- Walter WG, Winter AL, Cooper R, McCallum WC, and Aldridge VJ. Contingent negative variation – Electric sign of sensorimotor association and expectancy in human brain. *Nature*, 203(494): 380–384, 1964.
- Wilson M. Six views of embodied cognition. *Psychonomic Bulletin and Review*, 9(4): 625–636, 2002.
- Wilson M and Knoblich G. The case of motor involvement in perceiving conspecifics. *Psychological Bulletin*, 131(3): 460–473, 2005.
- Zacks JM and Tversky B. Event structure in perception and conception. *Psychological Bulletin*, 127(1): 3–21, 2001.
- Zatorre RJ, Chen JL, and Penhune VB. When the brain plays music: Auditory–motor interactions in music perception and production. *Nature Reviews Neuroscience*, 8(7): 547–558, 2007.

Part III

Appendix

Zusammenfassung

Einleitung

Melodie und Tonfall spielen eine Schlüsselrolle in Gesellschaft und Kultur. Sie sind ohne Zweifel Grundlage für musikästhetischen Genuss (Huron, 2006; Patel, 2008) und erleichtern das Sprachverstehen (Cole, 2015; Cutler et al., 1997), indem sie z.B. den akustischen Strom musikalischer und sprachlicher Information sinngebend strukturieren. Zum anderen sind Melodie und Tonfall ganz ohne Worte entscheidend im zwischenmenschlichen Umgang, für das Gefühl sozialer Nähe und Zusammengehörigkeit und das Erkennen innerer Einstellungen zueinander. Intuitiv erlauben (sprach)melodische Elemente eine Verständigung über interpersonelle Belange jenseits von Worten (Cross & Woodruff, 2009; Malloch & Trevarthen, 2009b), bilden eine ganz ursprüngliche Form der Kommunikation, die onto- und vielleicht sogar phylogentischer Grundstein (Darwin, 1871; Fitch, 2010; Mithen, 2005) für die Formung sozio-affektiv und linguistisch kompetenter Menschen ist.

Die vorliegende Arbeit beleuchtet die *soziale* und *linguistische* Relevanz von Melodie und Tonfall in Sprache und Musik in der Kommunikation zwischen Erwachsenen und eruiert deren neuronale Grundlagen. Dabei verfolgt die Arbeit ein dreistufiges Forschungsprogramm (Abbildung 4.1), das bestehendes Wissen in mehrfacher Hinsicht erweitert: Nach einer einführenden Diskussion zu strukturellen Gemeinsamkeiten und Unterschieden von Melodie in Sprache und Musik zeigt die Arbeit, **(A)** dass Melodie in beiden Domänen in komplexen *Netzwerken* räumlich getrennter, aber eng miteinander verknüpfter Hirnareale verarbeitet wird, **(B)** dass diese Netzwerke mit (nicht-melodischen) linguistischen und sozio-kognitiven neuronalen Netzwerken *interagieren*, um **(C)** den Gebrauch von Melodie und Tonfall *innerhalb und zwischen* (Inter)Akteuren zu steuern.

Ausgangspunkt 1 der Arbeit ist das Bewusstsein, dass weder Sprache (Friederici, 2011; Poeppel, 2014) noch Musik (Koelsch, 2011a; Peretz & Coltheart, 2003) *monolithische* Fähigkeiten sind, sondern eine Komposition aus vielen kognitiven Subroutinen. Diese Subroutinen erfordern eine ganze Reihe an spezialisierten Hirnregionen, die über Faserbündel der weißen Substanz fortlaufend miteinander kommunizieren und so Informationseinheiten zwischen aufeinanderfolgenden Verarbeitungsstufen übermitteln. Moderne neurobiologische Modelle der Sprachverarbeitung haben diesen Netzwerkgedanken aufgenommen und es gilt heute als unbestritten, dass z.B. das kortikale Sprachsystem der linken Hemisphäre, das u.a. syntaktische und semantische Anteile der Sprache verarbeitet, anatomisch und funktionell unterscheidbare **dorsale und ventrale Pfade** umfasst (Bornkessel-Schlesewsky et al., 2015; Friederici, 2011; Hickok & Poeppel, 2007; Rauschecker & Scott, 2009). Kapitel 2 („Neuronale Netzwerke für Melodie“) überträgt die Idee multipler Pfade auf die Verarbeitung von Melodie. Es präsentiert ein *rechtshemisphärisches* Zwei-Pfad-Modell für die Wahrnehmung von Prosodie, die Sprachmelodie (**Studien I & II**), und zieht Parallelen zu Netzwerken der Harmoniewahrnehmung und -produktion in Musik (**Studie III**). Danach fokussiert Kapitel 3 („Melodie und Fakten“) auf die Interaktion zwischen melodischen und verbalen Verarbeitungspfaden. Es präsentiert Evidenz für den Informationsaustausch zwischen rechtshemisphärischen prosodischen und linkshemisphärischen syntaktischen Prozessen in Sprache (**Studie IV**) und untersucht die Verknüpfung von Melodie und Text bei der Wahrnehmung gesungener Lieder (**Studien V & VI**).



Abbildung 4.1: Dreistufiges Forschungsprogramm zur Untersuchung der neuronalen Grundlagen von Intonation in Sprache und Musik. Der Begriff Melodie bezeichnet hier übergreifend *Tonverläufe über die Zeit*, einschließlich kontinuierlicher Tonkonturen in Sprachprosodie sowie Kompositionen aus Einzeltönen und Akkorden in Musik.

Ausgangspunkt 2 der Arbeit ist die Tatsache, dass sowohl Sprache als auch Musik üblicherweise *in Gesellschaft anderer* auftreten, wobei Gesprächspartner oder Musiker

flexibel miteinander interagieren, ihr Verhalten aufeinander abstimmen und wechselseitig lenken. In den sozialen Neurowissenschaften wird davon ausgegangen, dass erfolgreiche zwischenmenschliche Interaktion die Fähigkeit erfordert, **Handlungsziele** und **soziale Intentionen** des Gegenübers zu erkennen, wobei zwei mögliche Mechanismen zum Tragen kommen: Motorsimulation und *Mentalizing* (einen Überblick geben Van Overwalle & Baetens, 2009). Kapitel 4 („Melodie in (Inter)Aktion“) überträgt diese Konzepte auf die Nutzung von Melodie und Tonfall in Sprache und Musik. Es prüft, inwieweit Sprecher soziale Intentionen über Prosodie ausdrücken und damit die Reaktion des Gesprächspartners beeinflussen (**Studie VII**), inwiefern Musiker beim Musizieren die Handlungen ihrer Mitmusiker simulieren und damit die Qualität des Spiels erhöhen (**Studie VIII**), und welche Rolle kulturell verankerte Konventionen hierbei spielen.

Die vorliegende Arbeit geht diesen Fragen in acht empirischen Studien nach, in denen sie die neuronalen Grundlagen von Melodie und Tonfall in Sprache und Musik, in Perzeption und Produktion, mit jeweils angemessenen psychoakustisch-behavioralen, elektrophysiologischen (z.B. Elektroenzephalographie, EEG), bildgebenden (z.B. funktionelle Magnetresonanztomographie, fMRT) und nichtinvasiven Stimulationsverfahren (z.B. transkranielle Magnetstimulation, TMS) bei gesunden Erwachsenen, Patienten mit neurologischen Störungen und professionellen Musikern untersucht.

Experimentelle Befunde und Diskussion

Neuronale Netzwerke für Melodie

Studien I-III eruierten die für Melodiewahrnehmung relevanten Hirnregionen und deren anatomische und funktionelle Vernetzung. Zuerst zeigte **Studie Ia** (Sammler et al., 2015) mittels funktioneller und diffusionsgewichteter MRT, dass sowohl **dorsale** als auch **ventrale Pfade** in der *rechten* Hemisphäre für **prosodische Konturwahrnehmung** relevant sind. Versuchspersonen wurden gebeten, einsilbige Wörter hinsichtlich zweier orthogonalisierter Dimensionen zu beurteilen: deren Prosodie, die zwischen Aussage und Frage (fallendem und steigendem Tonverlauf) variierte, oder deren wortinitiales Phonem, das zwischen /b/ und /p/ variierte. Hirnregionen, die während der Prosodieaufgabe stärker aktiv waren als in der Phonemaufgabe, wurden als prosodierelevant klassifiziert. Probabilistische Faser-Traktographie zwischen diesen Arealen zeigte die anatomischen Verbindungen in diesem Netzwerk auf. Dies waren der Fasciculus longitudinalis medius

(MdLF) im ventralen Pfad, der den posterioren (pSTS) mit dem anterioren Sulcus temporalis superior (aSTS) verbindet, sowie zwei parallel verlaufende Anteile des Fasciculus arcuatus / longitudinalis superior (AF/SLF) im dorsalen Pfad, die den pSTS zum einen mit dorso-posterioren Anteilen des Gyrus frontalis inferior (IFG) und zum anderen mit dem prämotorischen Kortex (PMC) verbinden (siehe Abbildung 4.2). Insgesamt erinnert diese Zwei-Pfad-Struktur des prosodischen Netzwerks stark an die dorsalen/ventralen Pfade der linken Hemisphäre, die für Sprachverarbeitung bekannt (Bornkessel-Schlesewsky et al., 2015; Friederici, 2011; Hickok & Poeppel, 2007; Rauschecker & Scott, 2009; Saur et al., 2008), bei Prosodie aber rechthemisphärisch gewichtet sind.

Besonders hervorzuheben ist die klare Beteiligung des rechten *dorsalen* Pfades, dessen Relevanz für Sprache bis heute häufig in Frage gestellt wird (Hickok, 2012; Hickok & Poeppel, 2007). Dessen *kausale* Beteiligung an der Verarbeitung von (linguistischer) Sprachprosodie war deshalb Forschungsgegenstand der Studien Ib & II. In **Studie Ib** (Sammler et al., 2015) führten 15 Minuten inhibitorischer TMS des rechten PMC im dorsalen Pfad (1Hz, offline) zu einer signifikanten Verschlechterung der Prosodiewahrnehmung, verglichen mit (i) Phonemwahrnehmung, (ii) Sham-Stimulation und (iii) Stimulation des linken PMC. In **Studie II** (Sammler et al., 2018) zeigte ein Patient mit ödembedingten Gewebeauffälligkeiten des rechten dorsalen Pfades (inkl. AF/SLF) eine signifikante postoperative Verbesserung seiner Prosodiewahrnehmung, verglichen mit (i) seiner präoperativen Leistung, (ii) seiner Leistung in einer nicht-prosodischen verbalen Kontrollaufgabe und (iii) der Leistung von zehn gesunden Kontrollprobanden. Zusammengefasst liefern beide Studien Argumente für eine *kausale* Rolle des rechten dorsalen Pfades in der Wahrnehmung vokal-prosodischer Information, insbesondere von Tonverläufen.

Studie III (Bianco et al., 2016) zeigte schließlich, dass melodische Information in **Musik** ebenfalls entlang multipler Pfade verarbeitet wird. Inspiriert durch die vermutete Arbeitsteilung zwischen ventralen („Verstehen“) und dorsalen Pfaden („Produzieren“) in Modellen der Sprachverarbeitung (Hickok & Poeppel, 2007; Rauschecker & Scott, 2009; Saur et al., 2008), wurde mittels fMRT die Hirnaktivität von Pianisten registriert, während sie harmonisch reguläre und irreguläre Akkordsequenzen *hörten* (ohne zu spielen) oder *spielten* (ohne Ton). In Kombination mit funktionellen Konnektivitätsanalysen in Ruhe-Messungen (*resting-state* MRT) bestätigten die Daten die Beteiligung rechtshemisphärischer ventraler und dorsaler Pfade in der Perzeption und Produktion harmonischer Zusammenhänge in Musik. Temporal-auditorische Regionen, die während der *Wahrnehmung* aktiviert waren, interagierten mit den Anteilen des IFG (BA44/45), in denen Capsula extrema (EmC) und Fasciculus uncinatus (UF; beides ven-

trale Verbindungen) sowie AF/SLF (dorsale Verbindung) enden. Parietal-motorische Regionen, die während der *Produktion* aktiv waren, zeigten funktionelle Kopplung mit dorsalen Anteilen des IFC (BA44/6), die ausschließlich Endpunkte des AF/SLF sind. Innerhalb dieser Zwei-Pfad-Struktur sei an dieser Stelle besonders auf die klare Beteiligung des rechten *ventralen* Pfades hingewiesen. Bisherige Studien fokussierten nahezu ausschließlich auf den rechten dorsalen Pfad in Musik (Chen et al., 2015; Loui et al., 2009, 2011) und unterschätzten möglicherweise die Komplexität des Netzwerkes.

Zusammengenommen demonstrieren Studien I-III, dass die Arbeitsteilung multipler Pfade ein übergreifendes Merkmal neurokognitiver Organisation zu sein scheint, das gleichermaßen auf Melodie/Prosodie, Musik und Sprache zutrifft. Hervorzuheben ist, dass die drei Studien gemeinschaftlich mit der Vorstellung brechen, dass der rechte dorsale Pfad an Sprachwahrnehmung nicht beteiligt ist und die Wahrnehmung harmonischer Zusammenhänge in Musik ausschließlich dorsal verläuft. Stattdessen kristallisierte sich sowohl in Sprache als auch Musik eine Zwei-Pfad-Struktur (mit relativer Rechtsdominanz) für den Gebrauch von Melodie und Tonfall heraus.

Melodie und Fakten

Kapitel 3 thematisierte die Interaktion von Melodie und Tonfall mit *verbaler* Information. Zuerst untersuchte **Studie IV** (Sammner, Kotz, et al., 2010) die neuronale Integration von Prosodie und Syntax in **Sprache**. Fragestellung war, wie prosodische und syntaktische Phrasenstruktur, die üblicherweise im Satz aufeinander ausgerichtet sind, trotz ihrer Verarbeitung in getrennten Hemisphären (Friederici & Alter, 2004) miteinander verknüpft werden (Cole, 2015; Cutler et al., 1997). Im Fokus der Studie stand das Corpus callosum (CC; siehe Abbildung 4.2) als wahrscheinlichster Kommunikationskanal zwischen den lateralisierten Verarbeitungspfaden (Friederici et al., 2007), der hier mittels EEG bei neurologischen Patienten geprüft wurde. Patienten mit Läsionen im *posterioren* CC, d.h. unterbrochenem Informationsfluss zwischen den beiden Temporallappen (Hofer & Frahm, 2006), zeigten im EEG keine anteriore Negativierung (~200ms) bei Diskrepanzen zwischen prosodischer und syntaktischer Phrasenstruktur in auditorisch präsentierten Sätzen. Dieser elektrophysiologische Marker für Prosodie-Syntax-Integration (Eckstein & Friederici, 2006) wurde dahingegen sehr wohl bei gesunden Kontrollprobanden und auch bei Patienten mit Läsionen im *anterioren* CC, d.h. trotz unterbrochenem Informationsfluss zwischen den beiden Frontallappen, beobachtet. Insgesamt sprechen diese

Daten demnach für eine Integration prosodischer Kontur und syntaktischer Information bereits auf *frühen* auditorischen Verarbeitungsstufen im Temporallappen, die insbesondere durch das posteriore CC gewährleistet wird, dessen Läsion zu einer Entkopplung beider Informationsströme führte.

Studien V & VI stellten sich der Forschungsfrage aus Perspektive der **Musik** und widmeten sich der bis heute bestehenden Kontroverse, wie Melodie und Text in *Liedern* neuronal repräsentiert werden: als separate Komponenten (z.B. Samson & Zatorre, 1991, Experiment 2) oder als integriertes Ganzes (z.B. Samson & Zatorre, 1991, Experiment 1). Mittels fMRT konnten beide Studien zeigen, dass sich diese scheinbar gegensätzlichen Positionen gegenseitig nicht ausschließen, sondern dass die Segregation oder Integration beider Komponenten (i) von der kognitiven Verarbeitungsstufe (**Studie V**; Sammler, Baird, et al., 2010) und (ii) von der neuronalen Unversehrtheit der Hörer abhängt (**Studie VI**; Alonso et al., 2014). Beide Studien nutzten ein neu entwickeltes Adaptationsparadigma (*repetition-suppression*; siehe z.B. Grill-Spector et al., 2006) und zeigten supra-additive Adaptationseffekte für Melodie und Text im (bilateralen) mittleren STS und PMC, die entlang des ventralen Pfades hin zum anterioren STS ab zunahmen, zugunsten einer stärkeren Adaptation für Text allein. Eine Interpretation dieser Befunde ist, dass auf frühen prälexikalen, auditorischen Verarbeitungsstufen ein integriertes Perzept aus Melodie und Text geformt wird, das sowohl entlang des ventralen als auch dorsalen Pfades weitergeleitet wird: ventral zum aSTS für eine zunehmend unabhängige Analyse des linguistischen Gehalts (Bedeutung und Struktur; Friederici, 2011; Hickok & Poeppel, 2007; Scott et al., 2000) und dorsal zum PMC zur Übersetzung in ein integriertes Motorprogramm für subvokales (Mit)Singen (Zarate, 2013).

Aufbauend auf diesen Ergebnissen bei gesunden Probanden, zeigte **Studie VI** schließlich, dass die frühe auditorische Integration von Melodie und Text im ventralen Pfad von der Unversehrtheit des Hippocampus abhängt. Patienten mit unilateraler Sklerose des linken (oder rechten) Hippocampus zeigten schwächere kortikale Adaptation für Text und schwächere Integration von Melodie und Text als gesunde Probanden. Dem Hippocampus wird im steten Austausch mit kortikalen Regionen eine Schlüsselrolle in der Verknüpfung (*binding*) sensorischer Information zu einer integrierten Repräsentation zugeschrieben (Opitz, 2010). Hippocampale Läsionen (Samson & Zatorre, 1991) und/oder die daraus resultierende Unterbrechung der kortiko-hippocampalen Verarbeitungsschleifen könnten demnach unwillkürlich zu einer Entkopplung melodischer und verbaler Information führen und mögliche Erklärung für widersprüchliche Befunde in der Literatur liefern.

Zusammengenommen verdeutlichen Studien IV-VI, dass die Verknüpfung melodischer und verbaler Information während der Wahrnehmung von Sprache und Gesang eine komplexe Integrationsleistung ist, die sowohl einen flexiblen intra- und inter-hemisphärischen als auch kortiko-subkortikalen Informationsaustausch innerhalb und zwischen neuronalen Netzwerken erfordert. Trotz unterschiedlicher experimenteller Ansätze scheint sich zwischen den Studien Konsens über die neuronale Integration auf vergleichsweise frühen, auditorischen Verarbeitungsstufen im mittleren/posterioren Temporallappen abzuzeichnen.

Melodie in (Inter)Aktion

Nach Studien zur Frage, wie Melodie und Tonfall mit verbal-linguistischer Information interagieren, betrachtete Kapitel 4 schließlich die kommunikative Wirkung von *Melodie allein*, jenseits verbaler Botschaften. **Studie VII** (Hellbernd & Sammler, 2016) zog Inspiration aus Theorien der Pragmatik, die davon ausgehen, dass in sprachlicher Kommunikation nicht unbedingt zählt, *was* gesagt wird, sondern *warum* etwas gesagt wird (Bühler, 1934; Grice, 1957). In Abhängigkeit von letzterem kann dieselbe Äußerung „Es ist schwer, morgens pünktlich zu sein.“ als Kritik, Mitgefühl oder einfache Feststellung verstanden werden. Es ist hauptsächlich das *warum* – die kommunikative Intention des Sprechers – die die Reaktion des Gegenübers bestimmt und das wahre Motiv von Kommunikation ausmacht (Austin, 1962; Grice, 1957; Searle, 1969). Jedoch drücken Sprecher ihre Absicht häufig nicht offen im Wortlaut aus (Holtgraves, 2005). Ziel von Studie VII war zu zeigen, dass konventionelle prosodische Muster diese „unausgesprochenen“ sozialen Intentionen in der alltäglichen Kommunikation zwischen Erwachsenen vermitteln. Drei Experimente belegen, dass akustisch-prosodische Information in Einwortäußerungen ausreicht, um sowohl statistisch (Diskriminanzanalysen und multiple Regressionen) als auch behavioral sechs verschiedene Intentionen reliabel zu klassifizieren, ganz ohne situativen Kontext, über Sprecher hinweg und unabhängig von Wortbedeutung und emotionaler Färbung des Tonfalls. Diese Daten zeigen, dass Prosodie *selbst* kommunikative Intentionen jenseits des Wortlauts vermittelt und somit den Verlauf und die Qualität einer Konversation beeinflussen kann.

An dieser Stelle soll erwähnt werden, dass es sich in einer Folgestudie (Hellbernd & Sammler, 2018) bereits abzeichnet, dass erwartungsgemäß auditorisch-prosodische Netzwerke (wie in Kapitel 2 untersucht) an der Entschlüsselung der kommunikativen In-

tention beteiligt sind. Darüber hinaus interagieren diese Netzwerke mit sozio-kognitiven Regionen, wie z. B. der Amygdala und *Mentalizing*-Arealen (Frith & Frith, 2006), womit das Erkennen vokal kodierter sozialer Intentionen ähnlichen Mechanismen zu unterliegen scheint wie nonverbale Interaktion.

Abschließend wandte sich **Studie VIII** (Sammler et al., 2013) dem Erkennen von Intentionen und Handlungszielen in **Musik** zu, als Grundvoraussetzung für die Virtuosität, mit der Musiker in Ensembles interagieren. Die Studie zog Inspiration aus der Hypothese, dass erfolgreiche zwischenmenschliche Interaktion die stete *Antizipation* der Handlungsziele eines Gegenübers erfordert (Pickering & Garrod, 2013; Sebanz & Knoblich, 2009). Diese Antizipation beruht auf internen Vorwärtsmodellen, die es dem Beobachter erlauben, die Bewegungen des Partners zu **simulieren** (Wolpert et al., 2003), das Ziel der Handlung vorherzusehen und die eigene (Re)Aktion entsprechend zu planen oder anzupassen. Es ist bekannt, dass Musiker diese Vorwärtsmodelle nutzen, um vorherzusehen, *wann* ein Handlungsziel erreicht sein wird (einen Überblick gibt Keller, 2012). Studie VIII zeigte mit Hilfe eines Imitationsparadigmas (Novembre & Keller, 2011) und EEG-Messungen bei Pianisten, dass Musiker darüber hinaus gleichermaßen antizipieren, *was* der Partner als nächstes spielen wird. Das ist umso bemerkenswerter, da es sich hier um *ungeübte* Stücke handelte, in denen sich die Musiker lediglich auf melodische Konventionen, die musikalische „Syntax“ des Stückes, verlassen konnten. Studie VIII identifizierte somit Melodie und deren kulturell verankerten konventionellen Aufbau als ein Gerüst, das beide Interakteure als gemeinsame mentale Basis nutzen können, um die Handlung des jeweiligen Gegenübers zu antizipieren, ihr Verhalten wechselseitig zu koordinieren und in einer gemeinsamen musikalischen Linie zu harmonisieren.

Insgesamt sind beide Studien – trotz ihrer unterschiedlichen Blickwinkel auf abstrakte Tonkonturen (Studie VII) oder konkrete Körperbewegungen (Studie VIII) – Beispiele für die soziale Relevanz von melodischen Konventionen, die in beiden Fällen Rückschlüsse auf die Intentionen und Handlungsziele eines Interaktionspartners erlaubten. Das vermutlich kulturell erworbene Wissen, in welcher Situation Melodie und Tonfall wie und warum angewandt werden, wurde von Hörern und Sprechern/Musikern gleichermaßen konsultiert und als wesentlicher Grundstein für die Koordination gleichzeitiger oder gegenseitiger Handlungen genutzt. Zukünftige Untersuchungen in „echten“ interaktiven Situationen sind erstrebenswert (Novembre et al., 2016).

Diskussion

Alles in allem umreißt die Arbeit einen Forschungsrahmen, der den Gebrauch von Intonation in Sprache und Musik in ausgedehnten neuronalen Netzwerken (der rechten Hemisphäre) verortet, die ihrerseits mit subkortikalen, linguistischen und sozio-kognitiven Systemen interagieren, um den sprachgebunden und sozialen Umgang zwischen Personen zu dirigieren. Abbildung 4.2 fasst die wesentlichsten Befunde schematisch zusammen (rechte Seite) und stellt sie bestehenden Sprachmodellen gegenüber (linke Seite).

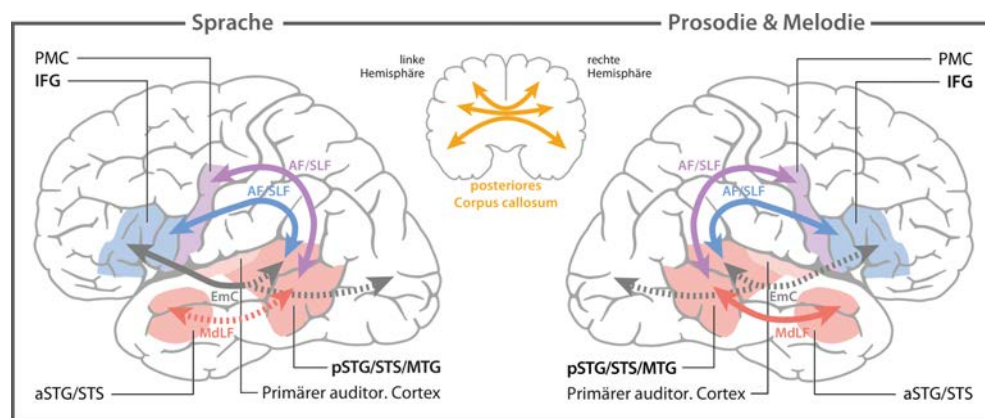


Abbildung 4.2: Schema der Verarbeitungspfade in linker und rechter Hemisphäre. Linkshemisphärische dorsale und ventrale Pfade für Sprache sind in Anlehnung an Friederici (2011) abgebildet: temporo-(prä)motorische Verbindungen für audio-motorische Übertragung von Silben und Phonemen (violett), temporo-frontale Verbindungen für hierarchische syntaktische Strukturbildung (blau), ventrale Verbindungen für semantische Prozesse (rot) und lokale syntaktische Strukturbildung (grau). Rechtshemisphärische Pfade für Prosodie und Melodie entsprechen den Ergebnissen der Studien I-III, V & VI der vorliegenden Arbeit: temporo-(prä)motorische Verbindungen für den zeitgenauen Nachvollzug von Tonverläufen und Schärfung auditorische Prozesse durch Motorsimulation (violett), temporo-frontale Verbindungen für musikalische und prosodische Strukturbildung (blau), ventrale Verbindungen für lokale musikalische Strukturbildung (grau) und Formung auditorischer Gestalten, die bedeutungstragend sein können (rot). Das posteriore Corpus callosum ist als interhemisphärische Schnittstelle zwischen prosodischen und nicht-prosodischen Pfaden abgebildet (Studie IV). AF: Fasciculus arcuatus, SLF: Fasciculus longitudinalis superior, EmC: Capsula extrema, MdLF: Fasciculus longitudinalis medialis, IFG: Gyrus frontalis inferior, STG/STS: Cylus/Sulcus temporalis superior, MTG: Cylus temporalis medius, PMC: prämotorischer Kortex, a: anterior, p: posterior.

Es ist nicht von der Hand zu weisen, dass die Wahrnehmung von Melodie sowohl in Sprache (Studien I & II) als auch Musik (Studien III, V & VI) auf **multiple Verarbeitungspfade mit relativer Rechtsdominanz** zurückgeht, die nahezu spiegelbildlich

zu bekannten linkshemisphärischen Sprachnetzwerken organisiert zu sein scheinen (Friederici, 2011; Hickok & Poeppel, 2007; Rauschecker & Scott, 2009). Zum heutigen Zeitpunkt lässt sich über die spezifische Funktion dieser Pfade nur spekulieren. Als Arbeitshypothese für Folgestudien sollen jedoch den korrespondierenden links- und rechtshemisphärischen Pfaden vorerst analoge Funktionen zugeschrieben werden, die sie entsprechend für verbale(n) Inhalt und Struktur bzw. Melodie und Tonfall ausüben. Die relative (aber keineswegs absolute) Rechtsdominanz melodischer Prozesse in Sprache und Musik geht konform mit bestehenden Modellen hemisphärischer Spezialisierung, die eine rechtshemisphärische Präferenz für spektrale Information (Zatorre, 2002) und vergleichsweise langsame Signalanteile (wie Tonkontur) vorschlagen, die sich über längere Zeitfenster erstrecken (Giraud et al., 2007; Poeppel, 2003).

Ventraler Pfad I – Bedeutung. Der ventrale Pfad entlang des rechten Temporallappens (rot in Abbildung 4.2) könnte durch sukzessive Integration und Abstraktion der akustischen Parameter des melodischen Signals (z.B. Norman-Haignere et al., 2015; Patterson et al., 2002) die Formung einer kategorisierbaren, auditorischen „Gestalt“ fördern (Schirmer & Kotz, 2006), der schließlich eine kommunikative Bedeutung zugeordnet werden kann. Grundvoraussetzung für einen derartigen Mechanismus ist eine einigermaßen konsistente Kopplung zwischen prosodischer Form (z.B. steigende Tonkontur) und kommunikativer Funktion (→ Frage), über Sprecher hinweg. Diese relative Konsistenz wurde in Studie VII (z.B. Hellbernd & Sammler, 2016), sogar für abstrakte zwischenmenschliche Konzepte wie Kritik und Zweifel, aufgezeigt. Demnach ist eine „norm-basierte“ Kodierung prosodischer Bedeutung durchaus denkbar, bei der sowohl Sprecher als auch Hörer die prosodische „Gestalt“ mit einem erfahrungsbasierten Inventar stilisierter akustischer Bilder abgleichen, um die kommunikative Bedeutung der Kontur zu (v)ermitteln. Die neuronale Realität eines derartigen Mechanismus, seine sehr wahrscheinlich probabilistische Natur und die unweigerliche Interaktion mit höheren Inferenzprozessen im situativen Kontext sind Themen für zukünftige Forschung.

Ventraler Pfad II – Einfache Struktur. In Musik kann Bedeutung – über die sozio-affektiven Konnotationen von Prosodie hinaus – auch über Strukturverhältnisse zwischen Tönen und Akkorden vermittelt werden (Koelsch, 2011b; Meyer, 1956). Diese „intra-musikalische“ Art der Bedeutung und die zugrundeliegende „syntaktische“ Strukturbildung mag Grund für die offensichtliche Beteiligung des ventralen fronto-temporalen Pfades (grau in Abbildung 4.2) beim Hören der Akkordsequenzen in Studie III gewesen sein (Bianco et al., 2016), der (bisher) keine Entsprechung in Prosodie fand. Grund hierfür mag die Nutzung von Einwortäußerungen in Studie I sein. In Studie II

wurde der ventrale Pfad nicht thematisiert. Unabhängig davon ist dieses Ergebnis in der Hinsicht bemerkenswert, als dass es bisherige implizite Annahmen, dass musikalische Strukturbildung *ausschließlich* Aufgabe dorsaler Pfade sei, revidiert (z.B. Loui et al., 2009, 2011; Peretz, 2016) und ein komplexeres Netzwerk als bisher angenommen propagiert (siehe unten).

Dorsaler Pfad I – Motorsimulation. Die Wahrnehmung von Melodie mag in beiden Domänen den dorsalen auditorisch-prämotorischen Pfad (violett in Abbildung 4.2) als geeignetes Werkzeug für den zeitgenauen *Nachvollzug* von Tonkonturen heranziehen (ähnlich der Nutzung auditorischen Feedbacks beim Sprechen oder Singen; Hickok, 2012; Houde & Chang, 2015; Zarate, 2013), oder auch zur *Schärfung* der Melodiewahrnehmung durch Simulation von Vokalisationsgesten beanspruchen (Rauschecker, 2011) (siehe auch A. M. Liberman & Mattingly, 1985). Während ersteres in Studien V & VI (Alonso et al., 2014; Sammler, Baird, et al., 2010) beim passiven Hören von Gesang der Fall gewesen sein mag, wurde letzteres besonders durch Studie Ib (Sammler et al., 2015) bestätigt. Eine Schlüsselfrage für Folgestudien wird sein, ob der PMC unwillkürlich oder strategisch (d.h. in Abhängigkeit von einer expliziten Aufgabe) zur Melodiewahrnehmung beiträgt.

Dorsaler Pfad II – Strukturbildung. Dorsale Verbindungen zwischen temporalen und inferior-frontalen Arealen (blau in Abbildung 4.2) könnten höhere Analysen der musikalischen oder sprachmelodischen Struktur leisten. Auf dem Gebiet der Musikkognition ist dieser Gedanke nicht neu (siehe oben; Loui et al., 2009, 2011) (vgl. jedoch Chen et al., 2015) und findet Unterstützung in Studie III (Bianco et al., 2016), die in Anlehnung an Sprachmodelle (Friederici, 2011) eine Zwei-Pfad-Struktur für melodische Strukturwahrnehmung vorschlägt. Kürzlich wurde begonnen, dem dorsalen Pfad auch prosodische Strukturverarbeitung zuzusprechen (Bornkessel-Schlesewsky & Schlesewsky, 2013), inkl. Segmentierung prosodischer Elemente und deren Kombination in komplexere syntaktische Einheiten. Die Idee ist mit den Ergebnissen von Studie II kompatibel (jedoch nicht endgültig verifizierbar; Sammler et al., 2018) und erfordert Folgestudien. Anzumerken bleibt, dass prosodische Segmentierung sehr wahrscheinlich nicht nur von prosodischen Elementen abhängt, sondern von syntaktischer Information profitiert (Cole, 2015; Cutler et al., 1997), was die Frage der Interaktion zwischen Pfaden aufwirft.

Interaktion zwischen Pfaden. Die transkallosale Interaktion der lateralisierten Verarbeitungspfade (Studie IV; Sammler, Kotz, et al., 2010) ist ein plausibler, jedoch sicherlich

keineswegs der einzige Weg der Informationsintegration. Zukünftige multimodale Studien werden spezifizieren müssen wo, wann und wie reziproke Interaktionen (i) zwischen dorsalen und ventralen Pfaden innerhalb der Hemisphäre (siehe Studie V und Cloutman, 2013), (ii) zwischen melodischen und *linguistischen* (Sammler, Baird, et al., 2010; Sammler, Kotz, et al., 2010) sowie *sozio-kognitiven* Netzwerken (Hellbernd & Sammler, 2018) auftreten. Möglicherweise erweisen sich hierbei subkortikale Strukturen als Schlüssel, wie die *hippokampale* Beteiligung bei der Bindung von Melodie und Text in Liedern bereits andeutet (Studie VI; Alonso et al., 2014).

Alles in allem bestehen die neuronalen Grundlagen von Intonation in Sprache und Musik aus einem flexiblen „Geben und Nehmen“ innerhalb und zwischen ausgedehnten Netzwerken. Obwohl weitere Forschungsarbeit sehr wohl notwendig ist, um die vorgeschlagene Arbeitsteilung – zwischen Pfaden und zwischen Hemisphären – zu prüfen und zu etablieren, betont das skizzierte Modell bereits heute die gemeinsame Beteiligung multipler, sich funktionell *ergänzender* und z.T. *redundanter* Verarbeitungspfade. Diese Mehrgleisigkeit mag eine hohe Anpassungsfähigkeit des Systems gewährleisten, wodurch es den vielfältigen Anforderungen unterschiedlicher Hörsituationen und individueller Sprechergewohnheiten im Alltag optimal gewachsen scheint.

Sozio-kommunikative Rolle von Melodie. Obwohl die Wichtigkeit *neuronaler Vernetzung* nicht genug betont werden kann, wäre jedes Modell unvollständig, das Melodie im *zwischenmenschlichen Umgang* miteinander vernachlässigt. Besonders Studien VII & VIII (Hellbernd & Sammler, 2016; Sammler et al., 2013) haben gezeigt, dass melodische Konventionen ein Gerüst für interpersonelle Koordination und Kooperation bieten – sowohl in Sprache als auch Musik. Regeln und Gebräuche, wie Töne kombiniert und Tonkonturen geformt werden, sind aktiven und rezeptiven Interakteuren bekannt und ermöglichen gegenseitiges Verständnis. Inwiefern melodische Gebräuche allein kulturell erlernt sind oder auf einer Reihe von universellen Grundmotive beruhen (z.B. die Präferenz für Konsonanz), ist eine interessante Frage für weiterführende kulturvergleichende Forschung.

Zusammengenommen besitzt Melodie in Sprache und Musik beziehungsstiftende Werte, die Interakteure zwischenmenschliche Begegnungen gestalten lassen. Diese Beobachtung hat tiefgreifende Folgen für neuronale Modelle von Sprache und Musik, die nun vor der Herausforderung stehen, die *sozialen* Mechanismen von Sprach- und Musik-*Gebrauch* zu bestimmen, über die Dekodierung des expliziten Stimulus hinaus. Die Fähigkeiten zur Handlungssimulation (Pickering & Garrod, 2013; Sebanz & Knoblich,

2009) und Zuschreibung mentaler Zustände (Frith & Frith, 2006) könnten hier Schlüssel sein.

Schlusswort. Die vorliegende Arbeit zeigt, dass wir erst zu verstehen beginnen, mit welcher komplexer neuronaler Maschinerie uns Mutter Natur ausgestattet hat, um die Melodien unserer Umwelt für vielfältige linguistische und zwischenmenschliche Belange zu nutzen. Die Zukunft muss zeigen, inwiefern der *Gebrauch* von Melodie Teil unseres evolutionsbiologischen Erbes ist (Cross & Woodruff, 2009; Darwin, 1871; Fitch, 2010; Mithen, 2005) oder kultureller Formung unterliegt. Es wird jedoch bereits heute deutlich, dass der bio- und soziokulturelle Status von Melodie einen ganzheitlichen Blick auf umfangreiche auditorisch-melodische, linguistische und sozio-kognitive neuronale Systeme und deren Interaktion erfordert, um der Vielfalt gerecht zu werden, mit der Melodie gesellschaftliche Harmonie und intellektuelles Erleben beeinflusst.

Bibliography

- Alonso, I., Davachi, L., Valabrègue, R., Lambrecq, V., Dupont, S., & Samson, S. (2016). Neural correlates of binding lyrics and melodies for the encoding of new songs. *NeuroImage*, 127, 333–345.
- Alonso, I., Sammler, D., Valabrègue, R., Dinkelacker, V., Dupont, S., Belin, P., & Samson, S. (2014). Hippocampal sclerosis affects fMR-adaptation of lyrics and melodies in songs. *Frontiers in Human Neuroscience*, 8, 111.
- Austin, J. L. (1962). *How to do things with words* (J. O. Urmson & M. Sbisà, Eds.). Oxford: Clarendon Press.
- Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, 70(3), 614–36.
- Bänziger, T., & Scherer, K. R. (2005). The role of intonation in emotional expressions. *Speech Communication*, 46(3), 252–267.
- Bašnáková, J., Weber, K., Petersson, K. M., van Berkum, J., & Hagoort, P. (2013). Beyond the language given: The neural correlates of inferring speaker meaning. *Cerebral Cortex*, 24(10), 2572–2578.
- Bekkering, H., de Bruijn, E. R. A., Cuijpers, R. H., Newman-Norlund, R., van Schie, H. T., & Meulenbroek, R. (2009). Joint action: Neurocognitive mechanisms supporting human interaction. *Topics in Cognitive Science*, 1(2), 340–352.
- Belin, P., & Zatorre, R. J. (2003). Adaptation to speaker's voice in right anterior temporal lobe. *Neuroreport*, 14(16), 2105–2109.
- Besson, M., Faïta, F., Peretz, I., Bonnel, A.-M., & Requin, J. (1998). Singing in the brain: Independence of lyrics and tunes. *Psychological Science*, 9(6), 494–498.
- Besson, M., Schön, D., Moreno, S., Santos, A., & Magne, C. (2007). Influence of musical expertise and musical training on pitch processing in music and language. *Restorative Neurology and Neuroscience*, 25(3-4), 399–410.
- Besson, P., Dinkelacker, V., Valabrègue, R., Thivard, L., Leclerc, X., Baulac, M., ... Dupont, S. (2014). Structural connectivity differences in left and right temporal lobe epilepsy.

- NeuroImage*, 100, 135–144.
- Bianco, R., Novembre, G., Keller, P. E., Kim, S.-G., Scharf, F., Friederici, A. D., ... Sammler, D. (2016). Neural networks for harmonic structure in music perception and action. *NeuroImage*, 142, 454–464.
- Bigand, E., Tillmann, B., Poulin, B., D'Adamo, D. A., & Madurell, F. (2001). The effect of harmonic context on phoneme monitoring in vocal music. *Cognition*, 81(1), B11–20.
- Bögels, S., Barr, D. J., Garrod, S., & Kessler, K. (2015). Conversational interaction in the scanner: Mentalizing during language processing as revealed by MEG. *Cerebral Cortex*, 25(9), 3219–3234.
- Bolinger, D. (1986). *Intonation and its parts - Melody in spoken English*. Stanford: Stanford University Press.
- Bornkessel-Schlesewsky, I., & Schlewsky, M. (2013). Reconciling time, space and function: A new dorsal-ventral stream model of sentence comprehension. *Brain and Language*, 125(1), 60–76.
- Bornkessel-Schlesewsky, I., Schlewsky, M., Small, S. L., & Rauschecker, J. P. (2015). Neurobiological roots of language in primate audition: Common computational properties. *Trends in Cognitive Sciences*, 19(3), 142–50.
- Brandt, A., Gebrian, M., & Slevc, L. R. (2012). Music and early language acquisition. *Frontiers in Psychology*, 3, 327.
- Brown, S., Ngan, E., & Liotti, M. (2008). A larynx area in the human motor cortex. *Cerebral Cortex*, 18(4), 837–845.
- Brück, C., Kreifelts, B., & Wildgruber, D. (2011). Emotional voices in context: A neurobiological model of multimodal affective information processing. *Physics of Life Reviews*, 8(4), 383–403.
- Bryant, G. A., & Barrett, H. C. (2007). Recognizing intentions in infant-directed speech: Evidence for universals. *Psychological Science*, 18(8), 746–751.
- Bryant, G. A., & Barrett, H. C. (2008). Vocal emotion recognition across disparate cultures. *Journal of Cognition and Culture*, 8(1), 135–148.
- Bühler, K. (1934). *Sprachtheorie: Die Darstellungsfunktion der Sprache* (3rd ed.). Jena: Gustav Fischer.
- Buxó-Lugo, A., & Watson, D. G. (2016). Evidence for the influence of syntax on prosodic parsing. *Journal of Memory and Language*, 90, 1–13.
- Canessa, N., Alemanno, F., Riva, F., Zani, A., Proverbio, A. M., Mannara, N., ... Ferguson, M. (2012). The neural bases of social intention understanding: The role of interaction goals. *PLoS ONE*, 7(7), e42347.

- Chen, J. L., Kumar, S., Williamson, V. J., Scholz, J., Griffiths, T. D., & Stewart, L. (2015). Detection of the arcuate fasciculus in congenital amusia depends on the tractography algorithm. *Frontiers in Psychology*, 6, 1–11.
- Chen, J. L., Rae, C., & Watkins, K. E. (2012). Learning to play a melody: An fMRI study examining the formation of auditory-motor associations. *NeuroImage*, 59(2), 1200–1208.
- Chevillet, M., Riesenhuber, M., & Rauschecker, J. P. (2011). Functional correlates of the anterolateral processing hierarchy in human auditory cortex. *The Journal of Neuroscience*, 31(25), 9345–9352.
- Chun, M. M., & Phelps, E. A. (1999). Memory deficits for implicit contextual information in amnesic subjects with hippocampal damage. *Nature Neuroscience*, 2(9), 844–847.
- Cloutman, L. L. (2013). Interaction between dorsal and ventral processing streams: Where, when and how? *Brain and Language*, 127(2), 251–63.
- Cole, J. (2015). Prosody in context: A review. *Language, Cognition and Neuroscience*, 30(1-2), 1–31.
- Collins, A., & Quillian, M. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning & Verbal Behavior*, 8(2), 240–247.
- Cross, I. (2012). Cognitive science and the cultural nature of music. *Topics in Cognitive Science*, 4(4), 668–677.
- Cross, I., & Woodruff, G. E. (2009). Music as a communicative medium. In R. Botha & C. Knight (Eds.), *The prehistory of language* (pp. 113–144). Oxford: Oxford University Press.
- Crowder, R. G., Serafine, M. L., & Repp, B. (1990). Physical interaction and association by contiguity in memory for the words and melodies of songs. *Memory & Cognition*, 18(5), 469–76.
- Cruttenden, A. (1994). Phonetic and prosodic aspects of baby talk. In C. Gallaway & B. Richards (Eds.), *Input and interaction in language acquisition* (pp. 135–152). Cambridge: Cambridge University Press.
- Cutler, A., Dahan, D., & Van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech*, 40, 141–201.
- Darwin, C. (1871). *The descent of man, and selection in relation to sex*. London: John Murray.
- D'Ausilio, A., Bufalari, I., Salmas, P., Busan, P., & Fadiga, L. (2011). Vocal pitch discrimination in the motor system. *Brain and Language*, 118(1-2), 9–14.
- D'Ausilio, A., Novembre, G., Fadiga, L., & Keller, P. E. (2015). What can music tell us about social interaction? *Trends in Cognitive Sciences*, 19(3), 111–114.

- Davachi, L. (2006). Item, context and relational episodic encoding in humans. *Current Opinion in Neurobiology*, 16(6), 693–700.
- David, O., Maess, B., Eckstein, K., & Friederici, A. D. (2011). Dynamic causal modeling of subcortical connectivity of language. *The Journal of Neuroscience*, 31(7), 2712–7.
- de Pijper, J. R., & Sanderman, A. A. (1994). On the perceptual strength of prosodic boundaries and its relation to suprasegmental cues. *The Journal of the Acoustical Society of America*, 96(4), 2037–2047.
- Deutsch, D., Henthorn, T., & Lapidis, R. (2011). Illusory transformation from speech to song. *The Journal of the Acoustical Society of America*, 129(4), 2245–2252.
- Dissanayake, E. (1992). *Homo aestheticus: Where art comes from and why*. New York: Free Press.
- Dowling, W. J. (1978). Scale and contour: Two components of a theory of memory for melodies. *Psychological Review*, 85, 341–354.
- Dowling, W. J. (1982). Melodic information processing and its development. In D. Deutsch (Ed.), *The psychology of music* (pp. 413–429). New York: Academic Press.
- Eckstein, K., & Friederici, A. D. (2006). It's early: Event-related potential evidence for initial interaction of syntax and prosody in speech comprehension. *Journal of Cognitive Neuroscience*, 18(10), 1696–1711.
- Egorova, N., Pulvermüller, F., & Shtyrov, Y. (2014). Neural dynamics of speech act comprehension: An MEG study of naming and requesting. *Brain Topography*, 27(3), 375–392.
- Egorova, N., Shtyrov, Y., & Pulvermüller, F. (2016). Brain basis of communicative actions in language. *NeuroImage*, 125, 857–67.
- Eichenbaum, H. (2000). A cortical-hippocampal system for declarative memory. *Nature Reviews Neuroscience*, 1(1), 41–50.
- Engel, A., Hijmans, B. S., Cerliani, L., Bangert, M., Nanetti, L., Keller, P. E., & Keysers, C. (2014). Inter-individual differences in audio-motor learning of piano melodies and white matter fiber tract architecture. *Human Brain Mapping*, 35(5), 2483–2497.
- Esteve-Gibert, N., & Prieto, P. (2013). Prosody signals the emergence of intentional communication in the first year of life: evidence from Catalan-babbling infants. *Journal of Child Language*, 40(5), 919–944.
- Ethofer, T., Anders, S., Erb, M., Herbert, C., Wiethoff, S., Kissler, J., ... Wildgruber, D. (2006). Cerebral pathways in processing of affective prosody: A dynamic causal modeling study. *NeuroImage*, 30(2), 580–587.
- Ethofer, T., Bretscher, J., Gschwind, M., Kreifelts, B., Wildgruber, D., & Vuilleumier, P.

- (2012). Emotional voice areas: Anatomic location, functional properties, and structural connections revealed by combined fMRI/DTI. *Cerebral Cortex*, 22(1), 191–200.
- Falk, D. (2004). Prelinguistic evolution in early hominins: Whence motherese? *Behavioral and Brain Sciences*, 27(4), 491–503.
- Falk, S. (2011). Melodic versus intonational coding of communicative functions: A comparison of tonal contours in infant-directed song and speech. *Psychomusicology: Music, Mind & Brain*, 21(1-2), 54–68.
- Falk, S., Rathcke, T., & Dalla Bella, S. (2014). When speech sounds like music. *Journal of Experimental Psychology: Human Perception and Performance*, 40(4), 1491–1506.
- Fernald, A. (1989). Intonation and communicative intent in mothers' speech to infants: Is the melody the message? *Child Development*, 60(6), 1497–510.
- Fernald, A., & Kuhl, P. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development*, 10(3), 279–293.
- Fitch, W. T. (2006). The biology and evolution of music: A comparative perspective. *Cognition*, 100(1), 173–215.
- Fitch, W. T. (2010). *The evolution of language*. Cambridge: Cambridge University Press.
- Frazier, L., Carlson, K., & Clifton Jr., C. (2006). Prosodic phrasing is central to language comprehension. *Trends in Cognitive Sciences*, 10(6), 244–249.
- Frick, R. W. (1985). Communicating emotion - The role of prosodic features. *Psychological Bulletin*, 97(3), 412–429.
- Friederici, A. D. (2011). The brain basis of language processing: From structure to function. *Physiological Reviews*, 91(4), 1357–1392.
- Friederici, A. D. (2012). The cortical language circuit: From auditory perception to sentence comprehension. *Trends in Cognitive Sciences*, 16(5), 262–268.
- Friederici, A. D., & Alter, K. (2004). Lateralization of auditory language functions: A dynamic dual pathway model. *Brain and Language*, 89(2), 267–276.
- Friederici, A. D., von Cramon, D. Y., & Kotz, S. A. (2007). Role of the corpus callosum in speech comprehension: Interfacing syntax and prosody. *Neuron*, 53(1), 135–145.
- Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. *Neuron*, 50(4), 531–534.
- Fritz, T., Jentschke, S., Gosselin, N., Sammler, D., Peretz, I., Turner, R., . . . Koelsch, S. (2009). Universal recognition of three basic emotions in music. *Current Biology*, 19(7), 573–576.
- Frühholz, S., & Grandjean, D. (2012). Towards a fronto-temporal neural network for the decoding of angry vocal expressions. *NeuroImage*, 62(3), 1658–1666.
- Frühholz, S., Trost, W., & Grandjean, D. (2014). The role of the medial temporal limbic

- system in processing emotions in voice and music. *Progress in Neurobiology*, 123, 1–17.
- Gabrielsson, A., & Juslin, P. N. (2003). Emotional expression in music. In *Handbook of affective sciences* (pp. 503–534). New York: Oxford University Press.
- Galantucci, S., Tartaglia, M. C., Wilson, S. M., Henry, M. L., Filippi, M., Agosta, F., . . . Gorno-Tempini, M. L. (2011). White matter damage in primary progressive aphasia: A diffusion tensor tractography study. *Brain*, 134, 3011–3029.
- Gandour, J. T., & Krishnan, A. (2015). Processing tone languages. In G. Hickok & S. L. Small (Eds.), *Neurobiology of language* (pp. 1095–1107). Amsterdam: Elsevier.
- Gierhan, S. M. E. (2013). Connections for auditory language in the human brain. *Brain and Language*, 127, 205–221.
- Giraud, A. L., Kleinschmidt, A., Poeppel, D., Lund, T. E., Frackowiak, R. S., & Laufs, H. (2007). Endogenous cortical rhythms determine cerebral specialization for speech perception and production. *Neuron*, 56(6), 1127–1134.
- Giraud, A. L., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience*, 15(4), 511–517.
- Glasser, M. F., & Rilling, J. K. (2008). DTI tractography of the human brain's language pathways. *Cerebral Cortex*, 18(11), 2471–2482.
- Gordon, R. L., Schön, D., Magne, C., Astésano, C., & Besson, M. (2010). Words and melody are intertwined in perception of sung words: EEG and behavioral evidence. *PLoS ONE*, 5(3), e9889.
- Graf Estes, K., & Hurley, K. (2013). Infant-directed prosody helps infants map sounds to meanings. *Infancy*, 18(5), 797–824.
- Graham, K. S., Barense, M. D., & Lee, A. C. (2010). Going beyond LTM in the MTL: A synthesis of neuropsychological and neuroimaging findings on the role of the medial temporal lobe in memory and perception. *Neuropsychologia*, 48(4), 831–853.
- Grice, H. P. (1957). Meaning. *Philosophical Review*, 66, 377–388.
- Griffiths, J. D., Marslen-Wilson, W. D., Stamatakis, E. A., & Tyler, L. K. (2013). Functional organization of the neural language system: Dorsal and ventral pathways are critical for syntax. *Cerebral Cortex*, 23(1), 139–47.
- Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: Neural models of stimulus-specific effects. *Trends in Cognitive Sciences*, 10(1), 14–23.
- Grossmann, T., Oberecker, R., Koch, S. P., & Friederici, A. D. (2010). The developmental origins of voice processing in the human brain. *Neuron*, 65(6), 852–8.
- Guenther, F. H., & Vladusich, T. (2012). A neural theory of speech acquisition and pro-

- duction. *Journal of Neurolinguistics*, 25(5), 408–422.
- Hagoort, P. (2009). Reflections on the neurobiology of syntax. In D. Bickerton & E. Szathmáry (Eds.), *Biological foundations and origin of syntax* (pp. 279–296). Cambridge, MA: MIT Press.
- Halwani, G. F., Loui, P., Rüber, T., & Schlaug, G. (2011). Effects of practice and experience on the arcuate fasciculus: Comparing singers, instrumentalists, and non-musicians. *Frontiers in Psychology*, 2, 156.
- Haspelmath, M., Dryer, M. S., Gil, D., & Comrie, B. (Eds.). (2005). *The World Atlas of Language Structures*. Oxford: Oxford University Press.
- Hébert, S., & Peretz, I. (2001). Are text and tune of familiar songs separable by brain damage? *Brain and Cognition*, 46(1-2), 169–75.
- Hébert, S., Racette, A., Gagnon, L., & Peretz, I. (2003). Revisiting the dissociation between singing and speaking in expressive aphasia. *Brain*, 126(8), 1838–1850.
- Hellbernd, N., & Sammler, D. (2016). Prosody conveys speaker's intentions: Acoustic cues for speech act perception. *Journal of Memory and Language*, 88, 70–86.
- Hellbernd, N., & Sammler, D. (2018). Neural bases of social communicative intentions in speech. *Social Cognitive and Affective Neuroscience*, 13(6), 604–615.
- Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience*, 13(2), 135–145.
- Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8(5), 393–402.
- Hickok, G., & Rogalsky, C. (2011). What Does Broca's Area Activation to Sentences Reflect? *Journal of Cognitive Neuroscience*, 23(10), 2629–2631.
- Hofer, S., & Frahm, J. (2006). Topography of the human corpus callosum revisited - Comprehensive fiber tractography using diffusion tensor magnetic resonance imaging. *NeuroImage*, 32(3), 989–994.
- Holtgraves, T. (2005). The production and perception of implicit performatives. *Journal of Pragmatics*, 37(12), 2024–2043.
- Holtgraves, T. (2008). Automatic intention recognition in conversation processing. *Journal of Memory and Language*, 58(3), 627–645.
- Houde, J. F., & Chang, E. F. (2015). The cortical computations underlying feedback control in vocal production. *Current Opinion in Neurobiology*, 33, 174–181.
- Huang, H., Zhang, J., Jiang, H., Wakana, S., Poetscher, L., Miller, M. I., ... Mori, S. (2005). DTI tractography based parcellation of white matter: Application to the mid-sagittal morphology of corpus callosum. *NeuroImage*, 26(1), 195–205.

- Huron, D. (2006). *Sweet anticipation: Music and the psychology of expectation*. Cambridge, MA: MIT Press.
- Hutchins, S., Gosselin, N., & Peretz, I. (2010). Identification of changes along a continuum of speech intonation is impaired in congenital amusia. *Frontiers in Psychology*, 1, 236.
- Hyde, K. L., Zatorre, R. J., & Peretz, I. (2011). Functional MRI evidence of an abnormal neural network for pitch processing in congenital amusia. *Cerebral Cortex*, 21(2), 292–299.
- Jusczyk, P. W., & Krumhansl, C. L. (1993). Pitch and rhythmic patterns affecting infants' sensitivity to musical phrase structure. *Journal of Experimental Psychology: Human Perception and Performance*, 19(3), 627–40.
- Juslin, P. N. (2013). What does music express? Basic emotions and beyond. *Frontiers in Psychology*, 4, 596.
- Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, 129(5), 770–814.
- Kawahara, H. (2006). STRAIGHT, exploitation of the other aspect of vocoder: Perceptually isomorphic decomposition of speech sounds. *Acoustical Science and Technology*, 27(6), 349–353.
- Keller, P. E. (2012). Mental imagery in music performance: Underlying mechanisms and potential benefits. *Annals of the New York Academy of Sciences*, 1252(1), 206–213.
- Kiebel, S. J., Daunizeau, J., & Friston, K. J. (2008). A hierarchy of time-scales and the brain. *PLoS Computational Biology*, 4(11), e1000209.
- Kilner, J. M. (2011). More than one pathway to action understanding. *Trends in Cognitive Sciences*, 15(8), 352–357.
- Knoblich, G., Butterfill, S., & Sebanz, N. (2011). Psychological research on joint action: Theory and data. In B. Ross (Ed.), *The psychology of learning and motivation* (pp. 59–101). Burlington: Academic Press.
- Knösche, T. R., Neuhaus, C., Haueisen, J., Alter, K., Maess, B., Witte, O. W., & Friederici, A. D. (2005). Perception of phrase structure in music. *Human Brain Mapping*, 24(4), 259–273.
- Koechlin, E., & Jubault, T. (2006). Broca's area and the hierarchical organization of human behavior. *Neuron*, 50(6), 963–974.
- Koelsch, S. (2011a). Toward a neural basis of music perception - A review and updated model. *Frontiers in Psychology*, 2, 1–20.

- Koelsch, S. (2011b). Towards a neural basis of processing musical semantics. *Physics of Life Reviews*, 8(2), 89–105.
- Koelsch, S. (2014). Brain correlates of music-evoked emotions. *Nature Reviews Neuroscience*, 15(3), 170–180.
- Koelsch, S., Skouras, S., & Jentschke, S. (2013). Neural correlates of emotional personality: a structural and functional magnetic resonance imaging study. *Plos One*, 8(11), e77196.
- Kotz, S. A., Meyer, M., & Paulmann, S. (2006). Lateralization of emotional prosody in the brain: An overview and synopsis on the impact of study design. *Progress in Brain Research*, 156, 285–294.
- Kotz, S. A., & Paulmann, S. (2011). Emotion, language, and the brain. *Language and Linguistics Compass*, 5(3), 108–125.
- Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Reviews Neuroscience*, 11(8), 599–605.
- Kravitz, D. J., Saleem, K. S., Baker, C. I., & Mishkin, M. (2011). A new neural framework for visuospatial processing. *Nature Reviews Neuroscience*, 12(4), 217–230.
- Kreiner, H., & Eviatar, Z. (2014). The missing link in the embodiment of syntax: Prosody. *Brain and Language*, 137, 91–102.
- Krekelberg, B., Boynton, G. M., & van Wezel, R. J. A. (2006). Adaptation: From single cells to BOLD signals. *Trends in Neurosciences*, 29(5), 250–256.
- Krumhansl, C. L. (1979). The psychological representation of musical pitch in a tonal context. *Cognitive Psychology*, 11, 445–463.
- Kümmerer, D., Hartwigsen, G., Kellmeyer, P., Glauche, V., Mader, I., Klöppel, S., ... Saur, D. (2013). Damage to ventral and dorsal language pathways in acute aphasia. *Brain*, 136(2), 619–29.
- Ladd, D. R. (1988). Declination "reset" and the hierarchical organization of utterances. *The Journal of the Acoustical Society of America*, 84(2), 530–544.
- Latinus, M., McAleer, P., Bestelmeyer, P. E., & Belin, P. (2013). Norm-based coding of voice identity in human auditory cortex. *Current Biology*, 23(12), 1075–1080.
- Laukka, P., Elfenbein, H. A., Söder, N., Nordström, H., Althoff, J., Chui, W., ... Thingujam, N. S. (2013). Cross-cultural decoding of positive and negative non-linguistic emotion vocalizations. *Frontiers in Psychology*, 4, 353.
- Leaver, A. M., & Rauschecker, J. P. (2010). Cortical representation of natural complex sounds: Effects of acoustic features and auditory object category. *The Journal of Neuroscience*, 30(22), 7604–7612.

- Leitman, D. I., Wolf, D. H., Ragland, J. D., Laukka, P., Loughhead, J., Valdez, J. N., ... Gur, R. C. (2010). "It's not what you say, but how you say it": A reciprocal temporofrontal network for affective prosody. *Frontiers in Human Neuroscience*, 4, 19.
- Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge: MIT Press.
- Leuthold, H. (2004). Action monitoring and control: in search of common principles. In M. Ullsperger & M. Falkenstein (Eds.), *Errors, conflicts, and the brain. current opinions on performance monitoring* (pp. 70–76). Leipzig: Max Planck Institute for Human Cognitive and Brain Sciences.
- Lévesque, Y., & Schön, D. (2015). Modulation of the motor cortex during singing-voice perception. *Neuropsychologia*, 70, 58–63.
- Levinson, S. C. (2006). On the human "interactional engine". In N. J. Enfield & S. C. Levinson (Eds.), *Roots of human sociality: Culture, cognition and interaction* (pp. 39–69). New York: Berg.
- Levinson, S. C., & Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology*, 6, 731.
- Lieberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21(1), 1–36.
- Lieberman, M., & Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry*, 8(2), 249–336.
- Lidji, P., Jolicœur, P., Moreau, P., Kolinsky, R., & Peretz, I. (2009). Integrated preattentive processing of vowel and pitch: A mismatch negativity study. *Annals of the New York Academy of Sciences*, 1169(1), 481–484.
- Liégeois-Chauvel, C., Peretz, I., Babai, M., Laguitton, V., & Chauvel, P. (1998). Contribution of different cortical areas in the temporal lobes to music processing. *Brain*, 121, 1853–1867.
- Liu, F., Patel, A. D., Fourcin, A., & Stewart, L. (2010). Intonation processing in congenital amusia: Discrimination, identification and imitation. *Brain*, 133, 1682–1693.
- Loui, P., Alsop, D., & Schlaug, G. (2009). Tone deafness: A new disconnection syndrome? *The Journal of Neuroscience*, 29(33), 10215–10220.
- Loui, P., Li, H. C., & Schlaug, G. (2011). White matter integrity in right hemisphere predicts pitch-related grammar learning. *NeuroImage*, 55(2), 500–507.
- Magne, C., Schön, D., & Besson, M. (2006). Musician children detect pitch violations in both music and language better than nonmusician children: Behavioral and electrophysiological approaches. *Journal of Cognitive Neuroscience*, 18(2), 199–211.

- Malloch, S., & Trevarthen, C. (Eds.). (2009a). *Communicative Musicality: Exploring the Basis of Human Companionship*. Oxford, UK: Oxford University Press.
- Malloch, S., & Trevarthen, C. (2009b). Musicality: Communicating the vitality and interests of life. In S. Malloch & C. Trevarthen (Eds.), *Communicative musicality: Exploring the basis of human companionship* (pp. 1–16). Oxford: Oxford University Press.
- Mampe, B., Friederici, A. D., Christophe, A., & Wermke, K. (2009). Newborns' cry melody is shaped by their native language. *Current Biology*, 19(23), 1994–1997.
- Männel, C., & Friederici, A. D. (2013). Accentuate or repeat? Brain signatures of developmental periods in infant word recognition. *Cortex*, 49(10), 2788–2798.
- McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, 102(3), 419–457.
- McDermott, J. H., Schultz, A. F., Undurraga, E. A., & Godoy, R. A. (2016). Indifference to dissonance in native Amazonians reveals cultural variation in music perception. *Nature*, 535(7613), 547–550.
- Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, 29(2), 143–178.
- Merrill, J., Bangert, M., Sammler, D., & Friederici, A. D. (2016). Classifying song and speech: Effects of focal temporal lesions and musical disorder. *Neurocase*.
- Mesulam, M. M. (1990). Large-scale neurocognitive networks and distributed processing for attention, language, and memory. *Annals of Neurology*, 28(5), 597–613.
- Meyer, L. (1956). *Emotion and meaning in music*. Chicago: University of Chicago Press.
- Mithen, S. (2005). *The singing Neanderthals - The origins of music, language, mind and body*. London: Weidenfeld & Nicolson.
- Moreno, S., & Besson, M. (2006). Musical training and language-related brain electrical activity in children. *Psychophysiology*, 43(3), 287–291.
- Musso, M., Weiller, C., Horn, A., Glauche, V., Umarova, R., Hennig, J., ... Rijntjes, M. (2015). A single dual-stream framework for syntactic computations in music and language. *NeuroImage*, 117, 267–283.
- Norman-Haignere, S., Kanwisher, N. G., & McDermott, J. H. (2015). Distinct cortical pathways for music and speech revealed by hypothesis-free voxel decomposition. *Neuron*, 88(6), 1281–1296.
- Novembre, G., & Keller, P. E. (2011). A grammar of action generates predictions in skilled musicians. *Consciousness and Cognition*, 20(4), 1232–1243.

- Novembre, G., Sammler, D., & Keller, P. E. (2016). Neural alpha oscillations index the balance between self-other integration and segregation in real-time joint action. *Neuropsychologia*, 89, 414–425.
- Obleser, J., & Eisner, F. (2009). Pre-lexical abstraction of speech in the auditory cortex. *Trends in Cognitive Sciences*, 13(1), 14–19.
- Oller, D. K., & Griebel, U. (2014). On quantitative comparative research in communication and language evolution. *Biological Theory*, 9(3), 296–308.
- Opitz, B. (2010). Neural binding mechanisms in learning and memory. *Neuroscience & Biobehavioral Reviews*, 34(7), 1036–1046.
- Papoušek, H., & Papoušek, M. (1987). Intuitive parenting: A dialectic counterpart to the infant's integrative competence. In J. D. Osofsky (Ed.), *Handbook of infant development* (2nd ed., pp. 669–720). New York: Wiley.
- Papoušek, M., Bornstein, M. H., Nuzzo, C., Papoušek, H., & Symmes, D. (1990). Infant responses to prototypical melodic contours in parental speech. *Infant Behavior and Development*, 13(4), 539–545.
- Patel, A. D. (2008). *Music, language, and the brain*. New York: Oxford University Press.
- Patel, A. D., Foxton, J. M., & Griffiths, T. D. (2005). Musically tone-deaf individuals have difficulty discriminating intonation contours extracted from speech. *Brain and Cognition*, 59(3), 310–313.
- Patel, A. D., Peretz, I., Tramo, M., & Labreque, R. (1998). Processing prosodic and musical patterns: A neuropsychological investigation. *Brain and Language*, 61(1), 123–144.
- Patterson, R. D., Uppenkamp, S., Johnsrude, I. S., & Griffiths, T. D. (2002). The processing of temporal pitch and melody information in auditory cortex. *Neuron*, 36(4), 767–776.
- Pell, M. D., Paulmann, S., Dara, C., Alasser, A., & Kotz, S. A. (2009). Factors in the recognition of vocally expressed emotions: A comparison of four languages. *Journal of Phonetics*, 37(4), 417–435.
- Perani, D., Saccuman, M. C., Scifo, P., Anwander, A., Spada, D., Baldoli, C., ... Friederici, A. D. (2011). Neural language networks at birth. *Proceedings of the National Academy of Sciences of the United States of America*, 108(38), 16056–16061.
- Peretz, I. (1990). Processing of local and global musical information by unilateral brain-damaged patients. *Brain*, 113(4), 1185–1205.
- Peretz, I. (1996). Can we lose memory for music? A case of music agnosia in a nonmusician. *Journal of Cognitive Neuroscience*, 8(6), 481–496.
- Peretz, I. (2016). Neurobiology of congenital amusia. *Trends in Cognitive Sciences*, 20(11),

857–867.

Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nature Neuroscience*, 6(7), 688–691.

Peretz, I., Gagnon, L., Hébert, S., & Macoir, J. (2004). Singing in the brain: Insights from cognitive neuropsychology. *Music Perception*, 21(3), 373–390.

Peretz, I., & Hyde, K. L. (2003). What is specific to music processing? Insights from congenital amusia. *Trends in Cognitive Sciences*, 7(8), 362–367.

Petkov, C. I., Logothetis, N. K., & Obleser, J. (2009). Where are the human speech and voice regions, and do other animals have anything like them? *Neuroscientist*, 15(5), 419–429.

Pichon, S., & Kell, C. A. (2013). Affective and sensorimotor components of emotional prosody generation. *The Journal of Neuroscience*, 33(4), 1640–1650.

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, 36(4), 329–347.

Plack, C. J., Oxenham, A. J., & Fay, R. R. (Eds.). (2005). *Pitch: Neural coding and perception*. New York: Springer.

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as 'asymmetric sampling in time'. *Speech Communication*, 41(1), 245–255.

Poeppel, D. (2014). The neuroanatomic and neurophysiological infrastructure for speech and language. *Current Opinion in Neurobiology*, 28, 142–149.

Poulin-Charronnat, B., Bigand, E., Madurell, F., & Peereman, R. (2005). Musical structure modulates semantic priming in vocal music. *Cognition*, 94(3), B67–B78.

Prieto, P., Estrella, A., Thorson, J., & del Mar Vanrell, M. (2012). Is prosodic development correlated with grammatical and lexical development? Evidence from emerging intonation in Catalan and Spanish. *Journal of Child Language*, 39(2), 221–257.

Racette, A., Bard, C., & Peretz, I. (2006). Making non-fluent aphasics speak: Sing along! *Brain*, 129(10), 2571–2584.

Rauschecker, J. P. (2011). An expanded role for the dorsal auditory pathway in sensorimotor control and integration. *Hearing Research*, 271(1-2), 16–25.

Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: Non-human primates illuminate human speech processing. *Nature Neuroscience*, 12(6), 718–724.

Riemann, H. (1877/1971). *Musikalische Syntaxis: Grundriss einer harmonischen Satzbildungslehre*. Niederwalluf: Sändig.

- Rilliard, A., Shochi, T., Martin, J. C., Erickson, D., & Aubergé, V. (2009). Multimodal indices to Japanese and French prosodically expressed social affects. *Language and Speech*, 52(2-3), 223–243.
- Sammler, D., Baird, A., Valabrègue, R., Clément, S., Dupont, S., Belin, P., & Samson, S. (2010). The relationship of lyrics and tunes in the processing of unfamiliar songs: A functional magnetic resonance adaptation study. *The Journal of Neuroscience*, 30(10), 3572–3578.
- Sammler, D., Cunitz, K., Gierhan, S. M. E., Anwander, A., Adermann, J., Meixensberger, J., & Friederici, A. D. (2018). White matter pathways for prosodic structure building: A case study. *Brain and Language*, 183, 1–10.
- Sammler, D., Grosbras, M.-H., Anwander, A., Bestelmeyer, P. E. G., & Belin, P. (2015). Dorsal and ventral pathways for prosody. *Current Biology*, 25(23), 3079–3085.
- Sammler, D., Kotz, S. A., Eckstein, K., Ott, D. V. M., & Friederici, A. D. (2010). Prosody meets syntax: The role of the corpus callosum. *Brain*, 133(9), 2643–2655.
- Sammler, D., Novembre, G., Koelsch, S., & Keller, P. E. (2013). Syntax in a pianist's hand: ERP signatures of embodied syntax processing in music. *Cortex*, 49(5), 1325–1339.
- Samson, S., & Zatorre, R. J. (1991). Recognition memory for text and melody of songs after unilateral temporal lobe lesion: Evidence for dual encoding. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 17(4), 793–804.
- Saur, D., Kreher, B. W., Schnell, S., Kümmerer, D., Kellmeyer, P., Vry, M. S., ... Weiller, C. (2008). Ventral and dorsal pathways for language. *Proceedings of the National Academy of Sciences of the United States of America*, 105(46), 18035–18040.
- Sauter, D. A., Eisner, F., Calder, A. J., & Scott, S. K. (2010). Perceptual cues in nonverbal vocal expressions of emotion. *The Quarterly Journal of Experimental Psychology*, 63(11), 2251–2272.
- Schaafsma, S. M., Pfaff, D. W., Spunt, R. P., & Adolphs, R. (2015). Deconstructing and reconstructing theory of mind. *Trends in Cognitive Sciences*, 19(2), 65–72.
- Schirmer, A., & Kotz, S. A. (2006). Beyond the right hemisphere: Brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences*, 10(1), 24–30.
- Schneider, K., Lintfert, B., Dogil, G., & Möbius, B. (2006). Phonetic grounding of prosodic categories. In S. Sudhoff et al. (Eds.), *Methods in empirical prosody research* (Vol. 3, pp. 335–361). Berlin: Walter de Gruyter.
- Schomers, M. R., & Pulvermüller, F. (2016). Is the sensorimotor cortex relevant for speech perception and understanding? An integrative review. *Frontiers in Human Neuroscience*, 10, 435.

- Schön, D., Magne, C., & Besson, M. (2004). The music of speech: Music training facilitates pitch processing in both music and language. *Psychophysiology*, 41(3), 341–349.
- Schuppert, M., Münte, T. F., Wieringa, B. M., & Altenmüller, E. (2000). Receptive amusia: Evidence for cross-hemispheric neural networks underlying music processing strategies. *Brain*, 123, 546–559.
- Scott, S. K., Blank, C. C., Rosen, S., & Wise, R. J. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain*, 123, 2400–2406.
- Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge: Cambridge University Press.
- Sebanz, N., & Knoblich, G. (2009). Prediction in joint action: What, when, and where. *Topics in Cognitive Science*, 1(2), 353–367.
- Serafine, M. L. (1984). Integration of melody and text in memory for songs. *Cognition*, 16(3), 285–303.
- Serafine, M. L., Davidson, J., Crowder, R. G., & Repp, B. H. (1986). On the nature of melody-text integration in memory for songs. *Journal of Memory and Language*, 25(2), 123–135.
- Sievers, B., Polansky, L., Casey, M., & Wheatley, T. (2013). Music and movement share a dynamic structure that supports universal expressions of emotion. *Proceedings of the National Academy of Sciences of the United States of America*, 110(1), 70–75.
- Skipper, J. I., Devlin, J. T., & Lametti, D. R. (2017). The hearing ear is always found close to the speaking tongue: Review of the role of the motor system in speech perception. *Brain and Language*, 164, 77–105.
- Smith, E., Shoben, E., & Rips, L. (1974). Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review*, 81(3), 214–241.
- Spitsyna, G., Warren, J. E., Scott, S. K., Turkheimer, F. E., & Wise, R. J. S. (2006). Converging language streams in the human temporal lobe. *The Journal of Neuroscience*, 26(28), 7328–7336.
- Srinivasan, R. J., & Massaro, D. W. (2003). Perceiving prosody from the face and voice: Distinguishing statements from echoic questions in English. *Language and Speech*, 46(1), 1–22.
- Stefanics, G., Háden, G. P., Sziller, I., Balázs, L., Beke, A., & Winkler, I. (2009). Newborn infants process pitch intervals. *Clinical Neurophysiology*, 120(2), 304–308.
- Steinbeis, N., & Koelsch, S. (2009). Understanding the intentions behind man-made products elicits neural activity in areas dedicated to mental state attribution. *Cerebral Cortex*, 19(3), 619–623.

- Steinhauer, K., Alter, K., & Friederici, A. D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience*, 2(2), 191–196.
- Steinmann, S., & Mulert, C. (2012). Functional relevance of interhemispheric fiber tracts in speech processing. *Journal of Neurolinguistics*, 25(1), 1–12.
- Stewart, L., Overath, T., Warren, J. D., Foxton, J. M., & Griffiths, T. D. (2008). fMRI evidence for a cortical hierarchy of pitch pattern processing. *Plos One*, 3(1), e1470.
- Stolk, A., Verhagen, L., Schoffelen, J. M., Oostenveld, R., Blokpoel, M., Hagoort, P., . . . Toni, I. (2013). Neural mechanisms of communicative innovation. *Proceedings of the National Academy of Sciences of the United States of America*, 110(36), 14574–14579.
- Streeter, L. A. (1978). Acoustic determinants of phrase boundary perception. *The Journal of the Acoustical Society of America*, 64(6), 1582–1592.
- Swain, J. (1995). The concept of musical syntax. *The Musical Quarterly*, 79(2), 281–308.
- Tanenhaus, M. K., Kurumada, C., & Brown, M. (2015). Prosody and intention recognition. In L. Frazier & E. Gibson (Eds.), *Explicit and implicit prosody in sentence processing* (pp. 99–118). Cham: Springer.
- Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy*, 7(1), 53–71.
- Tierney, A., Dick, F., Deutsch, D., & Sereno, M. (2013). Speech versus song: Multiple pitch-sensitive areas revealed by a naturally occurring musical illusion. *Cerebral Cortex*, 23(2), 249–54.
- Tillmann, B., L  v  que, Y., Forloni, L., Albouy, P., & Caclin, A. (2016). Impaired short-term memory for pitch in congenital amusia. *Brain Research*, 1640, 251–263.
- Trainor, L. J., & Trehub, S. E. (1992). A comparison of infants' and adults' sensitivity to Western musical structure. *Journal of Experimental Psychology: Human Perception and Performance*, 18(2), 394–402.
- Trehub, S. E., & Trainor, L. J. (1998). Singing to infants: Lullabies and play songs. *Advances in Infancy Research*, 12, 43–77.
- Trehub, S. E., Unyk, A. M., & Trainor, L. J. (1993). Adults identify infant-directed music across cultures. *Infant Behavior and Development*, 16(2), 193–211.
- Ueno, T., Saito, S., Rogers, T. T., & Lambon Ralph, M. A. (2011). Lichtheim 2: Synthesizing aphasia and the neural basis of language in a neurocomputational model of the dual dorsal-ventral language pathways. *Neuron*, 72(2), 385–396.
- Van Overwalle, F., & Baetens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: A meta-analysis. *NeuroImage*, 48(3), 564–584.
- Vandenberghe, R., Nobre, A. C., & Price, C. J. (2002). The response of left temporal cortex

- to sentences. *Journal of Cognitive Neuroscience*, 14(4), 550–560.
- van Schie, H. T., Mars, R. B., Coles, M. G. H., & Bekkering, H. (2004). Modulation of activity in medial frontal and motor cortices during error observation. *Nature Neuroscience*, 7(5), 549–554.
- Wagner, M., & Watson, D. G. (2010). Experimental and theoretical advances in prosody: A review. *Language and Cognitive Processes*, 25(7-9), 905–945.
- Warren, P. (1999). Prosody and language processing. In S. Garrod & Pickering, M. J. (Eds.), *Language processing* (pp. 155–188). Hove: Psychology Press.
- Weiller, C., Bormann, T., Saur, D., Musso, M., & Rijntjes, M. (2011). How the ventral pathway got lost - And what its recovery might mean. *Brain and Language*, 118(1-2), 29–39.
- Wermke, K., & Mende, W. (2009). Musical elements in human infants' cries: In the beginning is the melody. *Musicae Scientiae*, 13, 151–175.
- Wichmann, A. (2002). Attitudinal intonation and the inferential process. In B. Bel & I. Marlien (Eds.), *Speech prosody* (pp. 11–15).
- Wildgruber, D., Ackermann, H., Kreifelts, B., & Ethofer, T. (2006). Cerebral processing of linguistic and emotional prosody: fMRI studies. *Progress in brain research*, 156, 249–268.
- Wildgruber, D., Ethofer, T., Grandjean, D., & Kreifelts, B. (2009). A cerebral network model of speech prosody comprehension. *International Journal of Speech-Language Pathology*, 11(4), 277–281.
- Wilson, M., & Knoblich, G. (2005). The case for motor involvement in perceiving conspecifics. *Psychological Bulletin*, 131(3), 460–473.
- Wilson, S. M., Galantucci, S., Tartaglia, M. C., Rising, K., Patterson, D. K., Henry, M. L., ... Gorno-Tempini, M. L. (2011). Syntactic processing depends on dorsal language tracts. *Neuron*, 72(2), 397–403.
- Witteman, J., van Ijzendoorn, M. H., van de Velde, D., van Heuven, V. J., & Schiller, N. O. (2011). The nature of hemispheric specialization for linguistic and emotional prosodic perception: A meta-analysis of the lesion literature. *Neuropsychologia*, 49(13), 3722–3738.
- Wittgenstein, L. (1953). *Philosophische Untersuchungen* (G. E. M. Anscombe & R. Rhees, Eds.). Oxford: Blackwell.
- Wolpert, D. M. (1997). Computational approaches to motor control. *Trends in Cognitive Sciences*, 1(6), 209–216.
- Wolpert, D. M., Doya, K., & Kawato, M. (2003). A unifying computational framework for

- motor control and social interaction. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 358(1431), 593–602.
- Zarate, J. M. (2013). The neural control of singing. *Frontiers in Human Neuroscience*, 7, 237.
- Zatorre, R. J. (2002). How does the brain process complex sound? Current controversies. *Journal of Neurology Neurosurgery and Psychiatry*, 72(6), 825.
- Zatorre, R. J., Chen, J. L., & Penhune, V. B. (2007). When the brain plays music: Auditory-motor interactions in music perception and production. *Nature Reviews Neuroscience*, 8(7), 547–558.
- Zatorre, R. J., & Krumhansl, C. L. (2002). Mental models and musical minds. *Science*, 298(5601), 2138–2139.

List of figures

| | | |
|-----|--|-----|
| 1.1 | Research agenda | 13 |
| 1.2 | Schema of processing streams in the left and right hemisphere | 25 |
| 4.1 | Forschungsprogramm | 166 |
| 4.2 | Schema der Verarbeitungspfade in linker und rechter Hemisphäre | 173 |

Curriculum Vitae

Daniela Sammler

* 9. September 1978, Meerane, Deutschland

Forschungsschwerpunkte auf zwischenmenschlicher Kommunikation

- **Wahrnehmung** und Verstehen sprachmelodischer und musikalischer Signale
- Handlungskontrolle bei der **Produktion** von Sprachmelodie und Musik
- **Psychophysik**, experimentelle **Elektrophysiologie** und **Neuroimaging**

Derzeitig Stelle

| | |
|--------------|--|
| seit 07/2013 | Forschungsgruppenleiterin (W2-Stelle, befristet) , Otto-Hahn-Gruppe „Neuronale Grundlagen von Intonation in Sprache und Musik“, Max-Planck-Institut für Kognitions- und Neurowissenschaften, Leipzig, Deutschland |
|--------------|--|

Ausbildung und akademischer Werdegang

| | |
|-------------------|---|
| 07/1997 | Abitur (1,0 mit Auszeichnung) am Pestalozzi-Gymnasium Meerane, Deutschland |
| 10/1997 – 03/2004 | Diplomstudium Psychologie (1,1 mit Auszeichnung) an der Universität Leipzig – „Musik und Emotion: Elektrophysiologische Korrelate der Verarbeitung angenehmer und unangenehmer Musikstücke“ |
| 09/1999 – 07/2000 | Licence de Psychologie (12.-Platzierte des Jahrgangs) an der Université Louis Pasteur, Strasbourg, Frankreich |
| 04/2004 – 03/2008 | Promotion (summa cum laude) an der Universität Leipzig & MPI für Kognitions- und Neurowissenschaften, Leipzig, Deutschland – “The neuroanatomical overlap of syntax processing in music and language: Evidence from lesion and intracranial ERP studies” |
| 04/2008 – 03/2009 | Postdoktorandin an der Université Nord-de-France, Lille und am Hôpital de la Pitié-Salpêtrière, Paris, Frankreich |
| 04/2009 – 09/2011 | Postdoktorandin in der Abteilung Neuropsychologie, MPI für Kognitions- und Neurowissenschaften, Leipzig, Deutschland |
| 10/2011 – 06/2013 | Gastwissenschaftlerin am Centre for Cognitive Neuroimaging (CCNi), University of Glasgow, Schottland, Großbritannien |
| 07/2014 – 10/2014 | Gastwissenschaftlerin am MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Australien |

Preise und Auszeichnungen

| | |
|------|--|
| 2010 | Otto-Hahn-Medaille der Max-Planck-Gesellschaft für eine ausgezeichnete Dissertation, 5000 € (http://www.mpg.de/86170/otto_hahn_gruppen) |
|------|--|

| | |
|------|--|
| 2010 | Otto-Hahn-Award der Max-Planck-Gesellschaft, Finanzierung meines 2-jährigen Forschungsaufenthaltes als Gastwissenschaftlerin an der University of Glasgow + vollständige Finanzierung meiner derzeitigen Otto-Hahn-Gruppe |
| 2010 | Posterpreis , 16th Annual Meeting of the Organization of Human Brain Mapping |
| 2012 | Goldmedaille für herausragende Leistungen in der Denkmalpflege in Europa , 10. Europäische Messe für Denkmalpflege, Leipzig, Germany |

Drittmittel

| | |
|-----------|---|
| 2014 | „Nonverbale Kommunikation in Sprache und Musik“ – Stipendium <ul style="list-style-type: none"> • Antragsteller: Dr. Peter E. Keller & Dr. Daniela Sammler • Gefördert durch: MARCS Institute, Western Sydney University, Australien • Summe: 9.000\$ (~6.200€) |
| 2013-2018 | „Neuronale Grundlagen von Intonation in Sprache und Musik“ – Otto-Hahn-Award <ul style="list-style-type: none"> • Antragsteller: Dr. Daniela Sammler • Gefördert durch: Max-Planck-Gesellschaft • Summe: 600.000€ |
| 1999-2000 | Licence de Psychologie – ERASMUS Stipendium <ul style="list-style-type: none"> • Antragsteller: Daniela Sammler • Gefördert durch: ERASMUS • Summe: 1.200€ |

Akademische Selbstverwaltung

| | |
|-------------|--|
| seit 2016 | Mitglied des Ethikrates der Max-Planck-Gesellschaft |
| seit 2015 | Mitglied verschiedener Berufungskommissionen der Max-Planck-Gesellschaft (Berufung von Max-Planck-DirektorInnen und ForschungsgruppenleiterInnen) |
| seit 2013 | Gewählte wissenschaftliche Mitarbeitervertreterin im Wissenschaftlichen Rat der Max-Planck-Gesellschaft |
| seit 2013 | Mentorin für Nachwuchswissenschaftlerinnen im Minerva-FemmeNet der Max-Planck-Gesellschaft (http://www.minerva-femmenet.mpg.de) |
| seit 2013 | Associate Faculty Member, International Max Planck Research School (IMPRS) NeuroCom |
| 2010 – 2011 | Modulkoordinatorin „Biologische Psychologie“ an der Martin-Luther-Universität Halle-Wittenberg, Halle / Saale, Deutschland |

Kongresse, Workshops, und Symposien

| | |
|---------|---|
| 08/2015 | Sammler, D. (2015). The Melodic Mind. Workshop. Max-Planck-Institut für Kognitions- und Neurowissenschaften, Leipzig, Deutschland. |
|---------|---|

10/2013

Sammler, D., Anwender, A., Belin, P. (2013). CCNI workshop on Diffusion-weighted imaging. Workshop. Centre for Cognitive Neuroimaging (CCNi) & Institute of Neuroscience and Psychology (INP), University of Glasgow, Großbritannien.

Gutachtertätigkeiten

Gutachterin für internationale Fördermittelgeber

European Research Council (ERC), Portuguese Foundation for Science and Technology (FCT), National Science and Engineering Council of Canada (NSERC)

Redakteurin für

Frontiers in Auditory Cognitive Neuroscience

Gutachterin für

Behavior Research Methods, BMC Neuroscience, Brain, Brain and Language, Brain Research, Brain Structure and Function, Brain Topography, Cerebral Cortex, Cognition, Cognitive Neurodynamics, Cognitive Neuroscience, Cortex, Frontiers in Auditory Cognitive Neuroscience, Human Brain Mapping, International Journal of Psychophysiology, Journal of Cognitive Neuroscience, Journal of Neurophysiology, Journal of Neuroscience, Music Perception, Nature Communications, NeuroImage, Neuropsychologia, Philosophical Transactions of the Royal Society B, PloS ONE, Psychophysiology, Royal Society Open Science, Social Cognitive and Affective Neuroscience

Mitgliedschaften

Deutsche Gesellschaft für Psychologie (DGPs)
Deutsche Gesellschaft für Sprachwissenschaft (DGfS)
Society for the Neurobiology of Language (SNL)
Organization for Human Brain Mapping (OHBM)

Methoden

Psychophysische und psychoakustische Verhaltenstestung
Elektro- und Magnetoenzephalographie (E/MEG)
Funktionelle und strukturelle Magnetresonanztomographie (MRT)
Subdurale Elektrokortikographie (ECoG)
Transkranielle Magnet- und Gleichstromstimulation (TMS/tDCS)
Arbeit mit gesunden Erwachsenen und neurologischen Patienten

Nationale und internationale Kooperationspartner

| | |
|-------------------|---|
| Prof. P. Belin | Institut des Neurosciences de La Timone, Université Aix-Marseille, Frankreich |
| Prof. T. G. Bever | University of Arizona, Tucson, Arizona, USA |
| Prof. C. Davis | Western Sydney University, Sydney, Australien |

| | |
|------------------------|---|
| Prof. C. E. Elger | Klinik für Epileptologie, Universitätsklinikum Bonn, Deutschland |
| Prof. A. D. Friederici | Abteilung Neuropsychologie, MPI-CBS Leipzig, Deutschland |
| Prof. P. E. Keller | Western Sydney University, Sydney, Australien |
| Prof. S. A. Kotz | Maastricht University, Maastricht, Niederlande |
| Prof. J. Meixensberger | Klinik und Poliklinik für Neurochirurgie, Universitätsklinikum Leipzig, Deutschland |
| Dr. G. Novembre | University College London, Großbritannien |
| Prof. S. Samson | Hôpital de la Pitié-Salpêtrière, Paris, Université Lille Nord-de-France, Frankreich |
| Prof. Schulze-Bonhage | Freiburg Epilepsiezentrum, Universitätsklinikum Freiburg, Freiburg, Deutschland |
| Prof. A. Villringer | Abteilung Neurologie, MPI-CBS, Leipzig, Deutschland |

Weiterbildungen zur Lehre

| | |
|---------|---|
| 2015/16 | Sächsisches Hochschuldidaktikzertifikat (200 Arbeitseinheiten [AE], Abschluss 2017) |
| Modul 1 | „Hochschuldidaktische Grundlagen“ (60 AE) |
| Modul 2 | „Aktivierende Lehrmethoden“ (16 AE) |
| | „Abschlussarbeiten wirksam betreuen“ (16 AE) |
| | „Bewerten, Benoten, Prüfen“ (16 AE) |
| | „Beleg- und Diplomarbeiten schnell und objektiv bewerten“ (4 AE) |
| | „Zielkonsistente Prüfungsfragen konstruieren“ (4 AE) |
| | „Wirksame Beratungen“ (5AE) |

Seminare und Vorlesungen

Universität Leipzig (mit Professor Jörg Jescheniak)

| | |
|---------|--|
| 2016 | Kognitive Psychologie II – Denken und Sprache (Seminar, B.Sc. Psychologie) |
| 2015/16 | Einführung in die Psychologie (Seminar, B.Sc. Psychologie) |
| 2015 | Kognitive Psychologie II – Denken und Sprache (Seminar, B.Sc. Psychologie) |
| 2014/15 | Einführung in die Psychologie (Seminar, B.Sc. Psychologie) |
| 2014/15 | Empiriepraktikum – Posner-Paradigma (Praktische Übung, B.Sc. Psychologie) |
| 2014 | Kognitive Psychologie II – Denken und Sprache (Seminar, B.Sc. Psychologie) |
| 2013/14 | Einführung in die Psychologie (Seminar, B.Sc. Psychologie) |
| 2006 | Wahrnehmung (Seminar, Nebenfach Psychologie, mit Dr. Kirsten Volz) |
| 2005/06 | Biologische Psychologie (Vorlesung, Nebenfach Psychologie, mit Dr. Stefan Koelsch) |

International Max Planck Research School (IMPRS), MPI für Kognitions- und Neurowissenschaften

| | |
|------|--|
| 2016 | Sprache und Musik (Vorlesung, Promotionsstudium) |
|------|--|

Martin-Luther-Universität Halle-Wittenberg

| | |
|-------------|--|
| 2010 – 2011 | Modulkoordinatorin Biologische Psychologie (B.Sc. Psychologie) |
|-------------|--|

| | |
|---------|---|
| 2010/11 | Biologische Psychologie (Vorlesung, B.Sc. Psychologie) |
| 2010/11 | Biologische Psychologie (Übung, B.Sc. Psychologie) |
| 2010 | Biologische Psychologie – Musik & Gehirn (Seminar, B.Sc. Psychologie) |
| 2009/10 | Biologische Psychologie (Vorlesung, B.Sc. Psychologie, mit Dr. Roman Liepelt) |
| 2009/10 | Biologische Psychologie (Übung, B.Sc. Psychologie, mit Dr. Roman Liepelt) |

Supervision, Co-Supervision, Mentoring

Supervision von Doktoranden

| | |
|-------|---|
| 2017- | Maria Dotzer, [Temporal dynamics of prosody perception] |
| 2016- | Pei-Ju Chien, [Cross-cultural understanding of tone and intonation in speech] |
| 2013- | Nele Hellbernd, [The neural bases of prosody in speech acts] |
| 2012- | Roberta Bianco, [Principles of action planning in music production] |

Co-Supervision von Doktoranden

| | |
|-----------|---|
| 2010- | Maria Felber, [Prosody and music] |
| 2010-2016 | Eleanor Harding, [Rhythm-syntax interactions in music and speech] |
| 2010-2015 | Irene Alonso Fernández, [Hippocampal binding of lyrics and melodies to songs] |
| 2009-2012 | Julia Merrill (geb. Groh), [Speech and song perception] |

Diplom- und Master-Studenten

| | |
|------|---|
| 2016 | Katharina Roscher, M.Sc. Psychologie, Universität Regensburg [Die Rolle von motorischem und syntaktischem Priming beim Klavierspielen] |
| 2016 | Jasper Pöhls, M.Sc. Psychologie, Martin-Luther-Universität Halle-Wittenberg [Syntax in musical actions – tDCS evidence] |
| 2016 | Felix Haiduk, M.Sc. Psychologie, Universität Leipzig [The singer's formant] |
| 2016 | Natalie Kohler, M.Sc. in Music, Mind, and Brain, Goldsmiths University of London [Interaction between melodic pitch and linguistic prosodic expectancy] |
| 2015 | Nico Benz, M.A. Linguistik, Universität Leipzig [Reanalyse bei der Verarbeitung hierarchischer Strukturen] |
| 2011 | Katharina Fath, Diplom in Psychologie, Eberhard-Karls-Universität Tübingen [Interaktion von Musik und Handlung] |
| 2011 | Katrin Cunitz, Diplom in Psychologie, Universität Bremen [Die Rolle des Fasciculus Arcuatus beim Sprachverstehen] |
| 2010 | Eleanor Harding, M.Sc. Clinical Linguistics, Universität Potsdam [Neural correlates of music and action] |
| 2004 | Lena Nowicki, Diplom in Biologie, Universität Leipzig [Emotionale Wahrnehmung von Geräuschen] |

Bachelor-Studenten

| | |
|------|---|
| 2016 | Jannis Steininger, B.Sc. Psychologie, Universität Leipzig [Prosodische Äußerung von Intentionen – ein Vergleich zwischen Schauspielern und Nicht-Schauspielern] |
| 2016 | Elias Matin, B.Sc. Psychologie, Universität Leipzig [Prosodiewahrnehmung im prämotorischen Kortex – Etablierung eines neuen Paradigmas für TMS] |
| 2011 | Tobias Ay, B.Sc. Psychologie, Martin-Luther-Universität Halle-Wittenberg [Meaning in music] |

Praktikanten

| | |
|------|---|
| 2016 | Natalie Kohler (Goldsmiths, University of London, Großbritannien) |
| 2016 | Jannis Starzinski (Universität Leipzig) |
| 2016 | Jannis Steininger (Universität Leipzig) |
| 2015 | Elias Matin (Universität Leipzig) |
| 2014 | Corinna Eidner (Christliches Gymnasium Jena) |
| 2014 | Jessica Jiesi Zhao (Wilhelm-Ostwald-Gymnasium, Leipzig) |
| 2014 | İpek Pınar Bekâr (Ankara University, Türkei) |
| 2013 | Florian Scharf (Universität Leipzig) |
| 2012 | Roberta Bianco (Universität Mailand, Italien) |
| 2011 | Sophie-Marie Raeder (McGill University, Montréal, Kanada) |
| 2011 | Katharina Fath (Eberhard-Karls-Universität Tübingen) |
| 2010 | Irene Alonso Fernández (Universität Oviedo, Spanien) |
| 2010 | Tobias Ay (Martin-Luther-Universität Halle-Wittenberg) |
| 2010 | Johanna Reichert (Julius-Maximilian-Universität Würzburg) |
| 2009 | Eleanor Harding (Universität Potsdam) |
| 2009 | Katrin Cunitz (Universität Bremen) |

Studentische Hilfskräfte

| | |
|-----------|---|
| 2016- | Annika Just (Universität Leipzig) |
| 2015-2016 | Jasper Pöhls (Martin-Luther-Universität Halle-Wittenberg) |
| 2014-2015 | Florian Scharf (Universität Leipzig) |
| 2011-2012 | Anne Wagner (Universität Leipzig) |
| 2009-2011 | Katrin Cunitz (Universität Bremen) |

Publications

†part of this thesis

‡indicates student/trainee author

Books and chapters

1. **Sammler, D.** (2017). Neurophysiologische Aspekte des Singens. In: M. Fuchs (Ed.). *Die Stimme im pädagogischen Alltag*. (pp. 187-196). Berlin: Logos-Verlag.
2. **Sammler, D.** (2009). *The neuroanatomical overlap of syntax processing in music and language: Evidence from lesion and intracranial ERP studies (MPI Series in Human Cognitive and Brain Sciences, Vol. 108)*. Leipzig: Max Planck Institute for Human Cognitive and Brain Sciences.

Five most important publications (peer reviewed)

3. †‡Hellbernd, N., **Sammler, D.** (2016). Prosody conveys speaker's intentions: acoustic cues for speech act perception. *Journal of Memory and Language*, 88, 70-86.
4. †**Sammler, D.**, Grosbras, M.-H., Anwender, A., Bestelmeyer, P. E. G., Belin, P. (2015). Dorsal and ventral pathways for prosody. *Current Biology*, 25, 3079-3085.
5. †**Sammler, D.**, Kotz, S. A., Eckstein, K., Ott, D. V. M., Friederici, A. D. (2010). Prosody meets syntax: the role of the corpus callosum. *Brain*, 133, 2643-2655.
6. **Sammler, D.**, Koelsch, S., Ball, T., Brandt, A., Grigutsch, M., Huppertz, H.-J., Knösche, T. R., Wellmer, J., Widman, G., Elger, C. E., Friederici, A. D., Schulze-

- Bonhage, A. (2013). Co-localizing linguistic and musical syntax with intracranial EEG. *NeuroImage*, 64, 134-146.
7. †**Sammler, D.**, Novembre, G., Koelsch, S., Keller, P. E. (2013). Syntax in a pianist's hand: ERP signatures of "embodied" syntax processing in music. *Cortex*, 49, 1325-1339.

Publications in peer reviewed journals (continued)

8. †Merrill, J., Bangert, M., **Sammler, D.**, Friederici, A. D. (2016). Classifying song and speech: effects of focal temporal lesions and musical disorder. *Neurocase*, 22, 496-504.
9. †Bianco, R., Novembre, G., Keller, P. E., Kim, S.-G., †Scharf, F., Friederici, A. D., Villringer, A., **Sammler, D.** (2016). Neural networks for musical syntax in perception and action. *NeuroImage*, 142, 454-464.
10. Novembre, G., **Sammler, D.**, Keller, P. E. (2016). Self-other entrainment and co-representation are linked via alpha oscillations in musical joint action. *Neuropsychologia*, 89, 414-425.
11. †Bianco, R., Novembre, G., Keller, P. E., †Scharf, F., Friederici, A. D., Villringer, A., **Sammler, D.** (2016). Syntax in action has priority over movement selection in piano playing: An ERP study. *Journal of Cognitive Neuroscience*, 28, 41-56.
12. Maffongelli, L., Bartoli, E., **Sammler, D.**, Koelsch, S., Campus, C., Olivier, E., Fadiga, L., D'Ausilio, A. (2015). Distinct brain signatures of content and structure violation during action sequence observation. *Neuropsychologia*, 75, 30-39.
13. Besson P., Dinkelacker V., Valabrègue R., Thivard L., Leclerc, X., Baulac M., **Sammler D.**, Colliot O., Lehericy, S., Samson, S., Dupont S. (2014). Structural connectivity differences in left and right temporal lobe epilepsy. *NeuroImage*, 100, 135-144.
14. †Alonso, I., **Sammler, D.**, Valabrègue, R., Dinkelacker, V., Dupont, S., Belin, P., Samson, S. (2014). Hippocampal sclerosis affects fMR-adaptation of lyrics and melodies in songs. *Frontiers in Human Neuroscience*, 8, Article 111.

15. †Merrill, J., **Sammler, D.**, Bangert, M., Goldhahn, D., Lohmann, G., Turner, R., Friederici, A. D. (2012). Perception of words and pitch patterns in song and speech. *Frontiers in Psychology*, 3, Article 76.
16. **Sammler, D.**, Koelsch, S., Friederici, A. D. (2011). Are left fronto-temporal “language areas” a prerequisite for normal music-syntactic processing? *Cortex*, 47, 659-673.
17. **Sammler, D.**, Baird, A., Valabrègue, R., Clément, S., Dupont, S., Belin, P., Samson, S. (2010). The relationship of lyrics and tunes in the processing of unfamiliar songs: An fMR adaptation study. *The Journal of Neuroscience*, 30, 3572-3578.
18. **Sammler, D.**, Koelsch, S., Ball, T., Brandt, A., Elger, C. E., Friederici, A. D., Huppertz, H.-J., Knösche, T. R., Wellmer, J., Widman, G., Schulze-Bonhage, A. (2009). Overlap of musical and linguistic syntax processing: Intracranial ERP evidence. *Annals of the New York Academy of Sciences*, 1169, 494-498.
19. Fritz, T., Jentschke, S., Gosselin, N., **Sammler, D.**, Peretz, I., Turner, R., Friederici, A. D., Koelsch, S. (2009). Universal recognition of three basic emotions in music. *Current Biology*, 19, 573-576.
20. Koelsch, S., Schulze, K., **Sammler, D.**, Fritz, T., Mueller, K., Gruber, O. (2009). Functional architecture of verbal and tonal memory: An fMRI study. *Human Brain Mapping*, 30, 859-873.
21. Koelsch, S., **Sammler, D.**, Jentschke, S., Siebel, W. A. (2008). EEG correlates of moderate intermittent explosive disorder. *Clinical Neurophysiology*, 119, 151-162.
22. Koelsch, S., **Sammler, D.** (2008). Cognitive components of regularity processing in the auditory domain. *PLoS one*, 3, e2650.
23. Koelsch, S., Remppis, A., **Sammler, D.**, Jentschke, S., Mietchen, D., Fritz, T., Bonnemeier, H., Siebel, W. A. (2007). A Cardiac signature of emotionality. *European Journal of Neuroscience*, 26, 3328-3338.
24. Koelsch, S., Jentschke, S., **Sammler, D.**, Mietchen, D. (2007). Untangling syntactic and sensory processing: An ERP study of music perception. *Psychophysiology*, 44, 476-490.
25. **Sammler, D.**, Grigutsch, M., Fritz, T., Koelsch, S. (2007). Music and Emotion: Electrophysiological correlates of the processing of pleasant and unpleasant music. *Psychophysiology*, 44, 293-304.

26. Koelsch, S., Heinke, W., **Sammler, D.**, Olthoff, D. (2006). Auditory processing during deep propofol sedation and recovery from unconsciousness. *Clinical Neurophysiology*, 117, 1746-1759.
27. Koelsch, S., Gunter, T. C., Wittfoth, M., **Sammler, D.** (2005). Interaction between syntax processing in language and in music: An ERP study. *Journal of Cognitive Neuroscience*, 17, 1565-1577.
28. Heinke, W., Kenntner, R., Gunter, T. C., **Sammler, D.**, Olthoff, D., Koelsch, S. (2004). Sequential effects of increasing propofol sedation on frontal and temporal cortices as indexed by auditory event-related potentials. *Anesthesiology*, 100, 617-625.
29. Koelsch, S., Kasper, E., **Sammler, D.**, Schulze, K., Gunter, T., Friederici, A. D. (2004). Music, language and meaning: Brain signatures of semantic processing. *Nature Neuroscience*, 7, 302-307.
30. Scheid, R., Voltz, R., Guthke, T., **Sammler, D.**, von Cramon, D. Y. (2003). Neuropsychiatric findings in anti-MA2-positive limbic encephalitis. *Neurology*, 61, 1159-1160.
31. Koelsch, S., Gunter, T. C., Schröger, E., Tervaniemi, M., **Sammler, D.**, Friederici, A. D. (2001). Differentiating ERAN and MMN: An ERP study. *Neuroreport*, 12, 1385-1389.

Publications without peer review

32. **Sammler, D.** (2016). Musik- und Sprachwahrnehmung bei Epilepsiepatienten. *Zeitschrift für Epileptologie*, 29, 16-20.
33. **Sammler, D.** (2014). Prosodie – Die Musik der Sprache. *LogoTHEMA*, 11, 9-13.
34. **Sammler, D.** (2011). Amusie und Sprachstörungen. *Sprachheilarbeit*, 56, 194-197.

Publications currently under review

35. †Bianco, R., Novembre, G., Keller, P. E., Friederici, A. D., Villringer, A., **Sammler, D.** (under review). Musical genre-dependent behavioural and EEG signatures of action planning: A comparison between classical and jazz pianists

36. ‡**Sammler, D.**, †Cunitz, K., Gierhan, S. M. E., Anwander, A., Adermann, J., Meixensberger, J., Friederici, A. D. (under review). White matter pathways for prosodic structure building: A case study.

Selbständigkeitserklärung

Hiermit erkläre ich, dass ich die vorliegende Habilitationsschrift selbständig verfasst, keine anderen als die angegebenen Quellen und Hilfsmittel benutzt und wörtlich oder inhaltlich übernommene Stellen als solche gekennzeichnet habe.

Dr. Daniela Sammler

Leipzig, den 24. Februar 2017

MPI Series in Human Cognitive and Brain Sciences:

- 1 Anja Hahne
Charakteristika syntaktischer und semantischer Prozesse bei der auditiven Sprachverarbeitung: Evidenz aus ereigniskorrelierten Potentialstudien
- 2 Ricarda Schubotz
Erinnern kurzer Zeitdauern: Behaviorale und neurophysiologische Korrelate einer Arbeitsgedächtnisfunktion
- 3 Volker Bosch
Das Halten von Information im Arbeitsgedächtnis: Dissoziationen langsamer corticaler Potentiale
- 4 Jorge Jovicich
An investigation of the use of Gradient- and Spin-Echo (GRASE) imaging for functional MRI of the human brain
- 5 Rosemary C. Dymond
Spatial Specificity and Temporal Accuracy in Functional Magnetic Resonance Investigations
- 6 Stefan Zysset
Eine experimentalspsychologische Studie zu Gedächtnisabrufprozessen unter Verwendung der funktionellen Magnetresonanztomographie
- 7 Ulrich Hartmann
Ein mechanisches Finite-Elemente-Modell des menschlichen Kopfes
- 8 Bertram Opitz
Funktionelle Neuroanatomie der Verarbeitung einfacher und komplexer akustischer Reize: Integration haemodynamischer und elektrophysiologischer Maße
- 9 Gisela Müller-Plath
Formale Modellierung visueller Suchstrategien mit Anwendungen bei der Lokalisation von Hirnfunktionen und in der Diagnostik von Aufmerksamkeitsstörungen
- 10 Thomas Jacobsen
Characteristics of processing morphological structural and inherent case in language comprehension
- 11 Stefan Kölsch
*Brain and Music
A contribution to the investigation of central auditory processing with a new electrophysiological approach*
- 12 Stefan Frisch
Verb-Argument-Struktur, Kasus und thematische Interpretation beim Sprachverstehen
- 13 Markus Ullsperger
The role of retrieval inhibition in directed forgetting – an event-related brain potential analysis
- 14 Martin Koch
Measurement of the Self-Diffusion Tensor of Water in the Human Brain
- 15 Axel Hutt
Methoden zur Untersuchung der Dynamik raumzeitlicher Signale
- 16 Frithjof Kruggel
Detektion und Quantifizierung von Hirnaktivität mit der funktionellen Magnetresonanztomographie
- 17 Anja Dove
Lokalisierung an internen Kontrollprozessen beteiligter Hirngebiete mithilfe des Aufgabenwechselparadigmas und der ereigniskorrelierten funktionellen Magnetresonanztomographie
- 18 Karsten Steinhauer
Hirnphysiologische Korrelate prosodischer Satzverarbeitung bei gesprochener und geschriebener Sprache
- 19 Silke Urban
Verbinformationen im Satzverstehen
- 20 Katja Werheid
Implizites Sequenzlernen bei Morbus Parkinson
- 21 Doreen Nessler
Is it Memory or Illusion? Electrophysiological Characteristics of True and False Recognition
- 22 Christoph Herrmann
Die Bedeutung von 40-Hz-Oszillationen für kognitive Prozesse
- 23 Christian Fiebach
*Working Memory and Syntax during Sentence Processing.
A neurocognitive investigation with event-related brain potentials and functional magnetic resonance imaging*
- 24 Grit Hein
Lokalisation von Doppelaufgabendefiziten bei gesunden älteren Personen und neurologischen Patienten
- 25 Monica de Filippis
Die visuelle Verarbeitung unbeachteter Wörter. Ein elektrophysiologischer Ansatz
- 26 Ulrich Müller
Die catecholaminerge Modulation präfrontaler kognitiver Funktionen beim Menschen
- 27 Kristina Uhl
Kontrollfunktion des Arbeitsgedächtnisses über interferierende Information
- 28 Ina Bornkessel
The Argument Dependency Model: A Neurocognitive Approach to Incremental Interpretation
- 29 Sonja Lattner
Neurophysiologische Untersuchungen zur auditorischen Verarbeitung von Stimminformationen
- 30 Christin Grünewald
Die Rolle motorischer Schemata bei der Objektrepräsentation: Untersuchungen mit funktioneller Magnetresonanztomographie
- 31 Annett Schirmer
Emotional Speech Perception: Electrophysiological Insights into the Processing of Emotional Prosody and Word Valence in Men and Women
- 32 André J. Szameitat
Die Funktionalität des lateral-präfrontalen Cortex für die Verarbeitung von Doppelaufgaben
- 33 Susanne Wagner
Verbales Arbeitsgedächtnis und die Verarbeitung ambiger Wörter in Wort- und Satzkontexten
- 34 Sophie Manthey
Hirn und Handlung: Untersuchung der Handlungsrepräsentation im ventralen prämotorischen Cortex mit Hilfe der funktionellen Magnet-Resonanz-Tomographie
- 35 Stefan Heim
Towards a Common Neural Network Model of Language Production and Comprehension: fMRI Evidence for the Processing of Phonological and Syntactic Information in Single Words
- 36 Claudia Friedrich
Prosody and spoken word recognition: Behavioral and ERP correlates
- 37 Ulrike Lex
Sprachlateralisierung bei Rechts- und Linkshändern mit funktioneller Magnetresonanztomographie

- 38 Thomas Arnold
Computergestützte Befundung klinischer Elektroenzephalogramme
- 39 Carsten H. Wolters
Influence of Tissue Conductivity Inhomogeneity and Anisotropy on EEG/MEG based Source Localization in the Human Brain
- 40 Ansgar Hantsch
Fisch oder Karpfen? Lexikale Aktivierung von Benennungsalternativen bei der Objektbenennung
- 41 Peggy Bungert
*Zentralnervöse Verarbeitung akustischer Informationen
Signalidentifikation, Signallateralisation und zeitgebundene Informationsverarbeitung bei Patienten mit erworbenen Hirnschädigungen*
- 42 Daniel Senkowski
Neuronal correlates of selective attention: An investigation of electrophysiological brain responses in the EEG and MEG
- 43 Gert Wollny
Analysis of Changes in Temporal Series of Medical Images
- 51 Markus Ullsperger & Michael Falkenstein
Errors, Conflicts, and the Brain Current Opinions on Performance Monitoring
- 44 Angelika Wolf
Sprachverstehen mit Cochlea-Implantat: EKP-Studien mit postlingual ertaubten erwachsenen CI-Trägern
- 45 Kirsten G. Volz
Brain correlates of uncertain decisions: Types and degrees of uncertainty
- 46 Hagen Huttner
Magnetresonanztomographische Untersuchungen über die anatomische Variabilität des Frontallappens des menschlichen Großhirns
- 47 Dirk Köster
Morphology and Spoken Word Comprehension: Electrophysiological Investigations of Internal Compound Structure
- 48 Claudia A. Hruska
Einflüsse kontextueller und prosodischer Informationen in der auditorischen Satzverarbeitung: Untersuchungen mit ereigniskorrelierten Hirnpotentialen
- 49 Hannes Ruge
Eine Analyse des raum-zeitlichen Musters neuronaler Aktivierung im Aufgabenwechselparadigma zur Untersuchung handlungssteuernder Prozesse
- 50 Ricarda I. Schubotz
Human premotor cortex: Beyond motor performance
- 51 Clemens von Zerssen
Bewusstes Erinnern und falsches Wiedererkennen: Eine funktionelle MRT Studie neuroanatomischer Gedächtniskorrelate
- 52 Christiane Weber
*Rhythm is gonna get you.
Electrophysiological markers of rhythmic processing in infants with and without risk for Specific Language Impairment (SLI)*
- 53 Marc Schönwiesner
Functional Mapping of Basic Acoustic Parameters in the Human Central Auditory System
- 54 Katja Fiehler
Temporospatial characteristics of error correction
- 55 Britta Stolterfoht
Processing Word Order Variations and Ellipses: The Interplay of Syntax and Information Structure during Sentence Comprehension
- 56 Claudia Danielmeier
Neuronale Grundlagen der Interferenz zwischen Handlung und visueller Wahrnehmung
- 57 Margret Hund-Georgiadis
Die Organisation von Sprache und ihre Reorganisation bei ausgewählten, neurologischen Erkrankungen gemessen mit funktioneller Magnetresonanztomographie – Einflüsse von Händigkeit, Läsion, Performanz und Perfusion
- 58 Jutta L. Mueller
Mechanisms of auditory sentence comprehension in first and second language: An electrophysiological miniature grammar study
- 59 Franziska Biedermann
Auditorische Diskriminationsleistungen nach unilateralen Läsionen im Di- und Telenzephalon
- 60 Shirley-Ann Rüschmeyer
The Processing of Lexical Semantic and Syntactic Information in Spoken Sentences: Neuroimaging and Behavioral Studies of Native and Non-Native Speakers
- 61 Kerstin Leuckefeld
The Development of Argument Processing Mechanisms in German. An Electrophysiological Investigation with School-Aged Children and Adults
- 62 Axel Christian Kühn
Bestimmung der Lateralisierung von Sprachprozessen unter besondere Berücksichtigung des temporalen Cortex, gemessen mit fMRT
- 63 Ann Pannekamp
Prosodische Informationsverarbeitung bei normalsprachlichem und deviantem Satzmaterial: Untersuchungen mit ereigniskorrelierten Hirnpotentialen
- 64 Jan Derrfuß
Functional specialization in the lateral frontal cortex: The role of the inferior frontal junction in cognitive control
- 65 Andrea Mona Philipp
The cognitive representation of tasks – Exploring the role of response modalities using the task-switching paradigm
- 66 Ulrike Toepel
Contrastive Topic and Focus Information in Discourse – Prosodic Realisation and Electrophysiological Brain Correlates
- 67 Karsten Müller
Die Anwendung von Spektral- und Waveletanalyse zur Untersuchung der Dynamik von BOLD-Zeitreihen verschiedener Hirnareale
- 68 Sonja A. Kotz
The role of the basal ganglia in auditory language processing: Evidence from ERP lesion studies and functional neuroimaging
- 69 Sonja Rossi
The role of proficiency in syntactic second language processing: Evidence from event-related brain potentials in German and Italian
- 70 Birte U. Forstmann
Behavioral and neural correlates of endogenous control processes in task switching
- 71 Silke Paulmann
Electrophysiological Evidence on the Processing of Emotional Prosody: Insights from Healthy and Patient Populations
- 72 Matthias L. Schroeter
Enlightening the Brain – Optical Imaging in Cognitive Neuroscience
- 73 Julia Reinholz
Interhemispheric interaction in object- and word-related visual areas
- 74 Evelyn C. Ferstl
The Functional Neuroanatomy of Text Comprehension
- 75 Miriam Gade
Aufgabeneinhibition als Mechanismus der Konfliktreduktion zwischen Aufgabenrepräsentationen

- 76 Juliane Hofmann
Phonological, Morphological, and Semantic Aspects of Grammatical Gender Processing in German
- 77 Petra Augurzyk
Attaching Relative Clauses in German – The Role of Implicit and Explicit Prosody in Sentence Processing
- 78 Uta Wolfensteller
Habituelle und arbiträre sensorimotorische Verknüpfungen im lateralen prämotorischen Kortex des Menschen
- 79 Päivi Sivonen
Event-related brain activation in speech perception: From sensory to cognitive processes
- 80 Yun Nan
Music phrase structure perception: the neural basis, the effects of acculturation and musical training
- 81 Katrin Schulze
Neural Correlates of Working Memory for Verbal and Tonal Stimuli in Nonmusicians and Musicians With and Without Absolute Pitch
- 82 Korinna Eckstein
Interaktion von Syntax und Prosodie beim Sprachverstehen: Untersuchungen anhand ereigniskorrelierter Hirnpotentiale
- 83 Florian Th. Siebörger
Funktionelle Neuroanatomie des Textverstehens: Kohärenzbildung bei Witzen und anderen ungewöhnlichen Texten
- 84 Diana Böttger
Aktivität im Gamma-Frequenzbereich des EEG: Einfluss demographischer Faktoren und kognitiver Korrelate
- 85 Jörg Bahlmann
Neural correlates of the processing of linear and hierarchical artificial grammar rules: Electrophysiological and neuroimaging studies
- 86 Jan Zwickel
Specific Interference Effects Between Temporally Overlapping Action and Perception
- 87 Markus Ullsperger
Functional Neuroanatomy of Performance Monitoring: fMRI, ERP, and Patient Studies
- 88 Susanne Dietrich
Vom Brüllen zum Wort – MRT-Studien zur kognitiven Verarbeitung emotionaler Vokalisationen
- 89 Maren Schmidt-Kassow
What's Beat got to do with it? The Influence of Meter on Syntactic Processing: ERP Evidence from Healthy and Patient populations
- 90 Monika Lück
Die Verarbeitung morphologisch komplexer Wörter bei Kindern im Schulalter: Neurophysiologische Korrelate der Entwicklung
- 91 Diana P. Szameitat
Perzeption und akustische Eigenschaften von Emotionen in menschlichem Lachen
- 92 Beate Sabisch
Mechanisms of auditory sentence comprehension in children with specific language impairment and children with developmental dyslexia: A neurophysiological investigation
- 93 Regine Oberecker
Grammatikverarbeitung im Kindesalter: EKP-Studien zum auditorischen Satzverstehen
- 94 Şükriü Banş Demiral
Incremental Argument Interpretation in Turkish Sentence Comprehension
- 95 Henning Holle
The Comprehension of Co-Speech Iconic Gestures: Behavioral, Electrophysiological and Neuroimaging Studies
- 96 Marcel Braß
Das inferior frontale Kreuzungsareal und seine Rolle bei der kognitiven Kontrolle unseres Verhaltens
- 97 Anna S. Hasting
Syntax in a blink: Early and automatic processing of syntactic rules as revealed by event-related brain potentials
- 98 Sebastian Jentschke
Neural Correlates of Processing Syntax in Music and Language – Influences of Development, Musical Training and Language Impairment
- 99 Amelie Mahlstedt
The Acquisition of Case marking Information as a Cue to Argument Interpretation in German: An Electrophysiological Investigation with Pre-school Children
- 100 Nikolaus Steinbeis
Investigating the meaning of music using EEG and fMRI
- 101 Tilmann A. Klein
Learning from errors: Genetic evidence for a central role of dopamine in human performance monitoring
- 102 Franziska Maria Korb
Die funktionelle Spezialisierung des lateralen präfrontalen Cortex: Untersuchungen mittels funktioneller Magnetresonanztomographie
- 103 Sonja Fleischhauer
Neuronale Verarbeitung emotionaler Prosodie und Syntax: die Rolle des verbalen Arbeitsgedächtnisses
- 104 Friederike Sophie Haupt
The component mapping problem: An investigation of grammatical function reanalysis in differing experimental contexts using eventrelated brain potentials
- 105 Jens Brauer
Functional development and structural maturation in the brain's neural network underlying language comprehension
- 106 Philipp Kanske
Exploring executive attention in emotion: ERP and fMRI evidence
- 107 Julia Grieser Painter
Music, meaning, and a semantic space for musical sounds
- 108 Daniela Sammler
The Neuroanatomical Overlap of Syntax Processing in Music and Language - Evidence from Lesion and Intracranial ERP Studies
- 109 Norbert Zmyj
Selective Imitation in One-Year-Olds: How a Model's Characteristics Influence Imitation
- 110 Thomas Fritz
Emotion investigated with music of variable valence – neurophysiology and cultural influence
- 111 Stefanie Regel
The comprehension of figurative language: Electrophysiological evidence on the processing of irony
- 112 Miriam Beisert
Transformation Rules in Tool Use
- 113 Veronika Krieghoff
Neural correlates of Intentional Actions
- 114 Andreja Bubić
Violation of expectations in sequence processing

- 115 Claudia Männel
Prosodic processing during language acquisition: Electrophysiological studies on intonational phrase processing
- 116 Konstanze Albrecht
Brain correlates of cognitive processes underlying intertemporal choice for self and other
- 117 Katrin Sakreida
Nicht-motorische Funktionen des prämotorischen Kortex: Patientenstudien und funktionelle Bildgebung
- 118 Susann Wolff
The interplay of free word order and pro-drop in incremental sentence processing: Neurophysiological evidence from Japanese
- 119 Tim Raettig
The Cortical Infrastructure of Language Processing: Evidence from Functional and Anatomical Neuroimaging
- 120 Maria Golde
Premotor cortex contributions to abstract and action-related relational processing
- 121 Daniel S. Margulies
Resting-State Functional Connectivity fMRI: A new approach for assessing functional neuroanatomy in humans with applications to neuroanatomical, developmental and clinical questions
- 122 Franziska Süß
The interplay between attention and syntactic processes in the adult and developing brain: ERP evidences
- 123 Stefan Bode
From stimuli to motor responses: Decoding rules and decision mechanisms in the human brain
- 124 Christiane Diefenbach
Interactions between sentence comprehension and concurrent action: The role of movement effects and timing
- 125 Moritz M. Daum
Mechanismen der frühkindlichen Entwicklung des Handlungsverständnisses
- 126 Jürgen Dukart
Contribution of FDG-PET and MRI to improve Understanding, Detection and Differentiation of Dementia
- 127 Kamal Kumar Choudhary
Incremental Argument Interpretation in a Split Ergative Language: Neurophysiological Evidence from Hindi
- 128 Peggy Sparenberg
Filling the Gap: Temporal and Motor Aspects of the Mental Simulation of Occluded Actions
- 129 Luming Wang
The Influence of Animacy and Context on Word Order Processing: Neurophysiological Evidence from Mandarin Chinese
- 130 Barbara Ettrich
Beeinträchtigung frontomedianer Funktionen bei Schädel-Hirn-Trauma
- 131 Sandra Dietrich
Coordination of Unimanual Continuous Movements with External Events
- 132 R. Muralikrishnan
An Electrophysiological Investigation Of Tamil Dative-Subject Constructions
- 133 Christian Obermeier
Exploring the significance of task, timing and background noise on gesture-speech integration
- 134 Björn Herrmann
Grammar and perception: Dissociation of early auditory processes in the brain
- 135 Eugenia Solano-Castiella
In vivo anatomical segmentation of the human amygdala and parcellation of emotional processing
- 136 Marco Taubert
Plastizität im sensorimotorischen System – Leminduzierte Veränderungen in der Struktur und Funktion des menschlichen Gehirns
- 137 Patricia Garrido Vázquez
Emotion Processing in Parkinson's Disease: The Role of Motor Symptom Asymmetry
- 138 Michael Schwartze
Adaptation to temporal structure
- 139 Christine S. Schipke
Processing Mechanisms of Argument Structure and Case-marking in Child Development: Neural Correlates and Behavioral Evidence
- 140 Sarah Jessen
Emotion Perception in the Multisensory Brain
- 141 Jane Neumann
Beyond activation detection: Advancing computational techniques for the analysis of functional MRI data
- 142 Franziska Knolle
Knowing what's next: The role of the cerebellum in generating predictions
- 143 Michael Skeide
Syntax and semantics networks in the developing brain
- 144 Sarah M. E. Gierhan
Brain networks for language: Anatomy and functional roles of neural pathways supporting language comprehension and repetition
- 145 Lars Meyer
The Working Memory of Argument-Verb Dependencies: Spatiotemporal Brain Dynamics during Sentence Processing
- 146 Benjamin Stahl
Treatment of Non-Fluent Aphasia through Melody, Rhythm and Formulaic Language
- 147 Kathrin Rothermich
The rhythm's gonna get you: ERP and fMRI evidence on the interaction of metric and semantic processing
- 148 Julia Merrill
Song and Speech Perception – Evidence from fMRI, Lesion Studies and Musical Disorder
- 149 Klaus-Martin Krönke
Learning by Doing? Gesture-Based Word-Learning and its Neural Correlates in Healthy Volunteers and Patients with Residual Aphasia
- 150 Lisa Joana Knoll
When the hedgehog kisses the frog: A functional and structural investigation of syntactic processing in the developing brain
- 151 Nadine Diersch
Action prediction in the aging mind
- 152 Thomas Dolk
A Referential Coding Account for the Social Simon Effect
- 153 Mareike Bacha-Trams
Neurotransmitter receptor distribution in Broca's area and the posterior superior temporal gyrus
- 154 Andrea Michaela Walter
The role of goal representations in action control

- 155 Anne Keitel
Action perception in development: The role of experience
- 156 Iris Nikola Knierim
Rules don't come easy: Investigating feedback-based learning of phonotactic rules in language.
- 157 Jan Schreiber
Plausibility Tracking: A method to evaluate anatomical connectivity and microstructural properties along fiber pathways
- 158 Katja Macher
Die Beteiligung des Cerebellums am verbalen Arbeitsgedächtnis
- 159 Julia Erb
The neural dynamics of perceptual adaptation to degraded speech
- 160 Philipp Kanske
Neural bases of emotional processing in affective disorders
- 161 David Moreno-Dominguez
Whole-brain cortical parcellation: A hierarchical method based on dMRI tractography
- 162 Maria Christine van der Steen
Temporal adaptation and anticipation mechanisms in sensorimotor synchronization
- 163 Antje Strauß
Neural oscillatory dynamics of spoken word recognition
- 164 Jonas Obleser
The brain dynamics of comprehending degraded speech
- 165 Corinna E. Bonhage
Memory and Prediction in Sentence Processing
- S 2 Tania Singer, Bethany E. Kok, Boris Bornemann, Matthias Bolz, and Christina A. Bochow
*The Resource Project
Background, Design, Samples, and Measurements*
- 166 Anna Wilsch
Neural oscillations in auditory working memory
- 167 Dominique Goltz
Sustained Spatial Attention in Touch: Underlying Brain Areas and Their Interaction
- 168 Juliane Dinse
A Model-Based Cortical Parcellation Scheme for High-Resolution 7 Tesla MRI Data
- 169 Gesa Schaadt
Visual, Auditory, and Visual-Auditory Speech Processing in School Children with Writing Difficulties
- 170 Laura Verga
Learning together or learning alone: Investigating the role of social interaction in second language word learning
- 171 Eva Maria Quinque
Brain, mood and cognition in hypothyroidism
- 172 Malte Wöstmann
Neural dynamics of selective attention to speech in noise
- 173 Charles-Étienne Benoit
Music-based gait rehabilitation in Parkinson's disease
- 174 Anja Fengler
How the Brain Attunes to Sentence Processing Relating Behavior, Structure, and Function
- 175 Emiliano Zaccarella
Breaking Down Complexity: The Neural Basis of the Syntactic Merge Mechanism in the Human Brain
- S 2 Tania Singer, Bethany E. Kok, Boris Bornemann, Matthias Bolz, and Christina A. Bochow
*2nd Edition The Resource Project
Background, Design, Samples, and Measurements*
- 176 Manja Attig
Handlungsverständnis in den ersten Lebensjahren: retrospektive und prospektive Verarbeitung
- 177 Andrea Reiter
*Out of control behaviors?
Investigating mechanisms of behavioral control in alcohol addiction, binge eating disorder, and associated risk factors*
- 178 Anna Strotseva-Feinschmidt
The processing of complex syntax in early childhood
- 179 Smadar Ovadia-Caro
Plasticity following stroke: the recovery of functional networks as measured by resting-state functional connectivity
- 180 Indra Kraft
Predicting developmental dyslexia at a preliterate age by combining behavioral assessment with structural MRI
- 181 Sabine Frenzel
*How actors become attractors
A neurocognitive investigation of linguistic actorhood*
- 182 Anja Dietrich
Food craving regulation in the brain: the role of weight status and associated personality aspects
- 183 Haakon G. Engen
On the Endogenous Generation of Emotion
- 184 Seung-Goo Kim
Myeloarchitecture and Intrinsic Functional Connectivity of Auditory Cortex in Musicians with Absolute Pitch
- 185 Yaqiong Xiao
Resting-state functional connectivity in the brain and its relation to language development in preschool children
- 186 Sofie Louise Valk
*The Structure of the Social Brain:
Dissociating socio-affective and socio-cognitive networks through the study of individual differences, brain plasticity, and disease models*
- 187 Douglas Weinbrenner
*Abstract pointing
ERP and behavioral evidence for its role in reference tracking*
- 188 Elisabeth Kaminski
Augmenting dynamic balance performance by transcranial direct current stimulation

- 189 Claudia Barth
Exploring structural and functional brain dynamics across the menstrual cycle
- 190 Eleanor Elizabeth Harding
Neurocognitive entrainment to meter influences syntactic comprehension in music and language: An individual-differences approach
- 191 Maike Hoff
*Motorische Plastizität über die Lebensspanne
Untersuchungen zur Reduktion altersbedingter feinmotorischer Defizite durch motorisches Lernen und nicht-invasiver Hirnstimulation*
- 192 Viola Rjosk
*Augmenting Motor Performance with Mirror Visual Feedback (MVF):
Underlying Mechanisms and Neural Correlates*
- 193 Charlotte Grosse Wiesmann
*The Emergence of Theory of Mind
Cognitive and Neural Basis of False Belief Understanding in Preschool Age*
- 194 Michael A. Skeide
The brain basis of developmental dyslexia