Julia Merrill: Song and Speech Perception: Evidence from fMRI, Lesion Studies and Musical Disorder. Leipzig: Max Planck Institute for Human Cognitive and Brain Sciences, 2013 (MPI Series in Human Cognitive and Brain Sciences; 148)

Song and Speech Perception - Evidence from fMRI, Lesion Studies and Musical Disorder

Impressum

Max Planck Institute for Human Cognitive and Brain Sciences, 2013

Druck: Sächsisches Druck- und Verlagshaus Direct World, Dresden
Titelbild: © Julia Merrill, 2013

ISBN 978-3-941504-32-5

Song and Speech Perception - Evidence from fMRI, Lesion Studies and Musical Disorder

Der Humanwissenschaftlichen Fakultät
der Universität Potsdam
eingereichte

DISSERTATION
zur Erlangung des akademischen Grades
doctor philosophiae
Dr. phil.

vorgelegt
von Julia Merrill
geboren am 18. September 1981 in Kassel

Dekan:
Prof. Dr. Frank Mayer
Gutachter:
Prof. Dr. Angela D. Friederici
Prof. Dr. Stefan Kölsch

Tag der Verteidigung: Potsdam, den 08.04.2013

Danksagung

Eine Dissertation verfasst sich nicht (von) allein. Unterstützung unterschiedlichster Art erfuhr ich von Kollegen, Freunden und Familie, denen an dieser Stelle herzlichst gedankt sein soll.

An erster Stelle gilt mein Dank Frau Prof. Dr. Angela D. Friederici für die Unterstützung und die besondere Möglichkeit an diesem herausragenden Institut die Arbeit für meine Promotion durchführen zu dürfen. An zweiter Stelle danke ich Prof. Dr. Stefan Kölsch für die Begutachtung dieser Dissertation.

Während meiner Doktorandenzeit standen mir vor allem zwei Personen zur Seite, ohne die diese Arbeit nicht die wäre, die sie ist: Dr. Marc Bangert und Dr. Daniela Sammler. Es ist unbezahlbar euch als Kollegen und Freunde zu haben.

Methodisch bereichert wurde diese Arbeit durch die Mitwirkung von Prof. Dr. Robert Turner, Dirk Goldhahn und Carsten Bogler.

Besonders beim Feinschliff der Dissertation halfen mir Dr. Michael Schwartze und Dr. Mathias Scharinger. Vielen Dank für euren Einsatz – vom kritischen Gegenlesen bis hin zur Vorbereitung der Disputation.

Die alltäglichen Höhen und Tiefen des Doktorandendaseins konnte ich immer mit meinen Kollegen teilen: Besonderer Dank geht an meine Bürokollegin und Freundin Lisa J. Knoll und an Corinna Bonhage, Dr. Patricia Garrido-Vasques, Dr. Sarah Gierhan, Eleanor Harding, Dr. Sarah Jessen, Prof. Dr. Peter Keller, Dana Marinos, Dr. Anna Mestres-Misse, Dr. Christian Obermeier, Dr. Jonas Obleser und Marie Uhlig.

Für eine gelungene Promotion muss auch privat genug Rückhalt sein. Den bekam ich von meinem Mann Michael, der sich gerne rückversicherte: “Will this dissertation ever be finished?!”, während er Korrektur las. Ganz besonders danken möchte ich meinen Eltern, die meinen langen Bildungsweg gelassen verfolgten (und vielfältig unterstützten) und sich sicher fragten, ob dies mal ein (gutes) Ende nehmen würde. Zu guter Letzt ist meine Tochter Leonie zu nennen, die ihre ersten Lebensmonate damit verbrachte, mir geduldig beim Schreiben zuzusehen.

Contents

Preface

The interrelationship of music and language has been of scholarly interest for years. Singing is where both domains meet - song and speech are both vocal, bear linguistic and lexical content and carry pitch contour to convey meaning. Despite sharing similarities, apparently there are equally salient differences between song and speech; otherwise it would not be so easy to instantaneously differentiate whether someone is singing or speaking. In this context, the general research question of the present dissertation is: How does our brain code for differences in song and speech? This question will be tackled in more detail by systematically utilizing approaches covering (i) functional imaging in healthy adults, (ii) systemic perception alterations in lesion patients, and (iii) considerations along a theoretical framework of musical disorder.

Given the aforementioned degree of feature-similarity in song and speech, it is no surprise that the brain activity associated with the perception of song and speech demonstrated some overlap. Previous research, however, provides evidence for hemispheric specialization, reflecting the traditional view of music processing being predominantly lateralized to the right and language lateralized to the left hemisphere (Riecker et al., 2000; Jeffries et al., 2003; Callan et al., 2006). The specifics of a putative left/right separation remain unresolved as both sung and spoken sentences (1) express meaning through words thus bearing linguistic information, (2) contain melodic information and (3) have underlying rhythm patterns. Phenomenologically, song relies on melodies with discrete pitch relations and typically shows discrete rhythmic onsets at integer multiples of the underlying metric beat or its subdivisions. Speech, by contrast, does not show these discrete relations, neither in pitch pattern (which shows continuous transition or ‘gliding’ pitch) nor in the periodicity of rhythmic/metric timing. However, spectral and temporal features are known to be processed differently, with spectral information being processed in the right and temporal information in the left hemisphere (e. g., Zatorre et al., 2002). The possibility that different spectral and temporal aspects of song and speech lead to different lateralization patterns calls for an experiment that carefully dissects these aspects in order to draw a conclusive picture on the neural distinction of song and speech perception.

Under normal conditions it is easy to differentiate between song and speech. Yet, when the auditory perception is distorted, discrimination performance may show a notable drop. Distortion of the sensory percept can be a result of one of two changes: (1) modification of the acoustical input, which can be manipulated systematically in order to artificially induce a performance drop in discrimination tasks; (2) impairments of the perception system itself, as a result of brain lesions or congenital amusia (i. e. a music perception disability). Both options will be exploited in the current study. Introducing stimuli that are manipulated in ways to bridge the continuum of acoustic parameters between song and speech (being perceived as 'halfway between') in a forced choice paradigm allows investigating to what extent people show a bias in classifying those stimuli as song or speech. By testing patients with focal lesions, the specific role of the temporal lobe in speech and song perception (for which imaging studies suggested overlap) may be investigated. Moreover, individuals with congenital amusia can help to understand the influence of music processing on song and speech discrimination.

This dissertation sets out to investigate song and speech perception on the level of their underlying constituents, words and pitch patterns (Experiment 1) and the special role of the temporal lobe and music processing abilities on the discrimination of song and speech (Experiment 2).

The first part of this dissertation provides the theoretical, methodological and empirical background framework. Chapter 1 will give an overview of the functional neuroanatomy of language and speech processing, specifying the diverse sub-processes in sentence comprehension, based on a leading model. An emphasis will be on word and pitch processing in speech, i. e. intelligibility, speech-specificity and prosody. Chapter 2 will outline two models of music perception. One describes music processing reflecting the complexity of language perception and is in close correspondence with the model of sentence comprehension, the other proposes distinct modules for processing language and music. Furthermore, the mechanisms underlying pitch processing in music will be addressed in more detail, and the peculiarities of song (as it combines word and music processing) will be touched upon briefly. Chapter 3 reviews the literature on similarities and differences of music and language, with a focus on words, pitch and rhythm properties being the main shared features of both domains. The specific roles of the temporal and the parietal lobe in music and speech processing will be reviewed. Chapter 4 will discuss song and speech at the phenomenological and neurophysiological level, and will elaborate on the research question of this dissertation along with Chapter 5, where musical disabilities, congenital and post-trauma, will be described. Part one of the current dissertation will conclude providing methodological information (Chapter 6) about the techniques used for the investigations, such as functional Magnetic Resonance Imaging (fMRI), and will critically discuss the lesion approach.

After outlining the research questions in Chapter 7, the second part of this dissertation presents two main experiments. In Experiment 1, the neural underpinnings of song and speech were investigated using fMRI. A general introduction in Chapter 8 is followed by the description of a behavioral pretest (Chapter 9, Experiment 1A) to evaluate the stimuli and the task used in the main experiment. The main experiment – an fMRI study – was analyzed with two different goals: Along with the experimental approach of previous studies, Experiment 1B Part 1 (Chapter 10) compares sung and spoken stimuli without taking underlying constituents into account. Experiment 1B Part 2 (Chapter 11) compares sung and spoken stimuli on different levels: words and pitch patterns. In the second Experiment (Chapter 12) patients with focal lesions in the temporal lobe and individuals with amusia were tested on their song and speech classification abilities. Chapter 13 will discuss the findings reported in the previous chapters with respect to the literature.

Part I

Theoretical and Empirical Background

Chapter 1

Language Processing

As far as we know, the human ability to use language is the most distinguished feature between us and any other being. Once learned, the ease with which we use language masks the inherent complexity therein. In this chapter, based on a recent neurocognitive model on sentence comprehension, the current idea of how and where the different features of language are processed will be described: acoustical, phonological, syntactic, semantic and prosodic information. Then, more detailed insights in speech processing will be given as the current dissertation focuses on speech processing rather than language processing. Speech is defined as the motor-output of language, i. e. the articulation and the resulting quality, pitch, resonance and intensity of the voice. The following sections will describe the processing of segmental information (i. e. consonants and vowels combined to meaningful units, such as words) and suprasegmental (i. e. prosodic) information, with special emphasis on the involvement of the temporal lobe. This chapter will give an overview on how and where sentence processing takes place and will further describe speech processing on the segmental and suprasegmental level, i. e. the processing of speech-sounds as vowels and consonants (spoken words) and linguistic prosodic information.

1.1 Functional Neurocognitive Architecture

Since the classical Broca-Wernicke model (Broca, 1861; Wernicke, 1874; Lichtheim, 1885; Geschwind, 1970), which was based on aphasic symptoms, functional magnetic resonance imaging (fMRI) research has advanced the understanding of language and speech processing in the brain. A large number of studies focused on the identification of brain areas supporting different aspects of language comprehension in general and speech comprehension in particular and lead to a variety of functional anatomic models. A segregation of the functions of the reported brain areas is a complex issue, as the observed involvement and interplay of regions may depend on the focus of the research, as a result of testing a specific model.

The complexity of language caused researchers to narrow down and focus on specific linguistic levels of description, such as prelexical phonemic processing, semantic processing (of spoken words and sentences), syntactic processing, as well as prosodic processing on sentence level. They reported a mainly left-lateralized fronto-temporo-parietal neural network for speech perception, including parts of the inferior frontal gyrus (IFG), the superior temporal gyrus (STG), middle temporal gyrus (MTG), inferior parietal and angular gyrus (for reviews and meta-analyses see e. g., Bookheimer, 2002; Friederici, 2002, 2011, 2012; Vigneau et al., 2006; Hickok & Poeppel, 2007; C. J. Price, 2010).

Sentence comprehension will be discussed in more detail including a leading model on language comprehension which was recently updated (Friederici, 2011, 2012; see Figure 1.1). The model describes which brain regions support particular language functions, such as phonetic, syntactic, sentence-level semantic and prosodic processes – based on evidence from electroencephalography (EEG), fMRI and connectivity studies.

Speech processing starts with the acoustic-phonetic analysis in the auditory cortex bilaterally with specialized temporal (left) and spectral (right) processing in the hemispheres (e. g., Obleser et al., 2008; Zatorre et al., 2002). While Heschl's gyrus (HG) reacts to any kind of sound, the planum temporale (PT) seems to be involved in sound categorization processes. It has been suggested to act as a 'computational hub' (Griffiths & Warren, 2002) that segregates auditory objects and gates that information to higher-order cortical areas. From the primary auditory cortex (PAC) the information goes to more posterior (PT and posterior STG, pSTG) and more anterior regions (planum polare, PP, and anterior STG) in the temporal lobe.

Activation in the anterior superior temporal sulcus (STS) was found for intelligible speech, while the posterior STS responded to sounds that contain some phonetic information, i. e. meaningful sounds, such as consonants and vowels, that were acoustically masked, and therefore not necessarily intelligible (Shannon et al., 1995; Scott et al., 2000; the latter study is reported in detail below). Speech selectivity was observed ventrally (Specht et al., 2009; Vaden et al., 2009) and posteriorly (Specht et al., 2009) as well as anteriorly to HG (Leff et al., 2009). A left lateralization in the posterior and ventral STS was suggested when white noise was gradually morphed into speech sounds compared to the transformation into morphed music sounds (Specht et al., 2009). C. J. Price (2010) suggested multiple prelexical processing pathways, depending on whether the experimental conditions were more or less demanding and allowed for expectancies from prior experience.

From the anterior STG (aSTG) extending in frontal regions, the aSTG and the left frontal operculum were suggested to form a network for initial local structure building

(Friederici, 2011). In an assumed left-lateralized temporo-frontal network, semantic and syntactic relations are processed in parallel. More precisely, the semantic network involves the middle and posterior STG and MTG and also BA 45 (and BA 47); the syntactic network (for complex sentence structures) involves the posterior STG/STS and BA 44. Syntactic and semantic integration processes are assumed to involve the STG/STS and the basal ganglia (BG; Friederici, 2011).

As far as this study is concerned, the involvement of the anterior temporal lobe (TL), as suggested by Friederici (2011), will be discussed in more detail. The lateral anterior TL was found for sentences over word lists (i. e. syntax vs. non-syntax; Friederici et al., 2000; Humphries et al., 2006; Mazoyer et al., 1993), and meaningless pseudoword sentences over meaningless pseudoword lists (Friederici et al., 2000; Humphries et al., 2006), which lead to the conclusion that whenever syntactic structure has to be processed (in sentences versus word lists), the anterior STG is involved (Friederici, 2011). Note that the evidence for syntactic structure processing in anterior TL is stronger than for semantic processing. The above mentioned studies, for example, did not reveal activation for semantic processing in the comparison of real-word stimuli and pseudoword-stimuli (Friederici et al., 2000; Humphries et al., 2006).

Furthermore, syntactic ambiguity was reported to engage the inferior parietal lobe and the MTG anterior to HG (Tyler & Marslen-Wilson, 2008), semantic sentence ambiguity was found to activate the left posterior temporal cortex including STS, MTG and inferior temporal gyrus (Rodd et al., 2005). Both also activate the left IFG (Tyler & Marslen-Wilson, 2008).

The left IFG (specifically Broca's area, encompassing BA 44 and BA 45) has always been considered to play an important role in language comprehension, but its exact functions are still a matter of debate (for a review see Friederici, 2011). Evidence from various studies though reported the frontal operculum for syntax processing, suggesting BA 44 as being a most critical region for processing syntactic complexity (Friederici, 2011). Studies also suggest the role of BA 45 in semantic processing (e. g., Thompson-Schill et al., 1997; Friederici, 2012), or assume a gradient of phonological-syntactic-semantic processing along the posterior-anterior axis of the IFG (Hagoort, 2005). Regarding lateralization, syntactic processes showed a clear left-lateralization of the temporo-frontal network while semantic processes seem to be less lateralized.

In contrast, suprasegmental prosodic information was found to recruit a temporo-frontal network in the right hemisphere (superior temporal and fronto-opercular cortices; Meyer et al., 2002), which is in close interaction with the left hemisphere through the posterior portion of the corpus callosum (CC; Friederici et al., 2007; Sammler, Kotz, et al., 2010), connecting the temporal structures of both hemispheres (for more details about prosodic processing, see below).

Of special interest for the current thesis is the processing of segmental information, such as phones and phonemes, and suprasegmental information, such as prosody. Therefore the following paragraphs focus on these specific components in speech.

1.1.1 Processing of Words in Speech

Language uses units on several levels that can be combined in many different ways. Segments, such as phones (i. e. articulated speech sounds) and phonemes (i. e. meaningful segments, such as consonants and vowels) are combined to form words. This phonetic and phonological aspect of speech, the segmental information (suprasegmental information, such as prosody, is discussed in the following section), is the topic of this section. In comparison to other sound systems, such as environmental sounds or instrumental music, words as well as larger units, such as sentences might be speech-specific, and the question remains whether speech-typical factors can be demarcated from others and whether there are speech-selective areas in the brain.

In a study on the intelligibility of speech sounds, regions processing phonetic information and intelligible speech were investigated (Scott et al., 2000). Intelligible speech was contrasted with so called rotated speech, which has very similar temporal and spectral pattern to ordinary speech, but does not contain phonetic features (Blesser, 1972), i. e. functional acoustic properties that carry meaningful information, and is not intelligible as the acoustic information is masked or degraded by spectral rotation. Intelligible speech was also presented as so called noise-vocoded speech (Shannon et al., 1995), in which fine spectral details are removed and thus hampers the extraction of phonetically relevant properties, consequently affecting intelligibility. Furthermore, a combination of both, so called rotated noise vocoded speech, does not convey pitch and phonetic information. A region ventrolaterally to HG responded to all types of stimuli, phonetic information was processed in pSTS and ventrolateral STS, and intelligible speech followed an anterolateral stream from PAC with responses from mid left STS and anterior STS (see also e. g., Davis & Johnsrude, 2003; Liebenthal et al., 2005; Zekveld et al., 2006; Obleser et al., 2007). In the same vein, Obleser et al. (2006) found the aSTG activated for vowel sounds over non-speech (band-passed noise) stimuli.

A recent meta-analysis by F. Samson et al. (2011) confirmed this pattern by comparing intelligible speech to unintelligible spectro-temporally complex sounds and reported speech-related activity in lateral non-primary superior temporal regions, specifically in posterior STG, and anterior and middle STS. They reported engaged areas beyond auditory cortex in left IFG and prefrontal cortex and proposed an expanded hierarchical model of

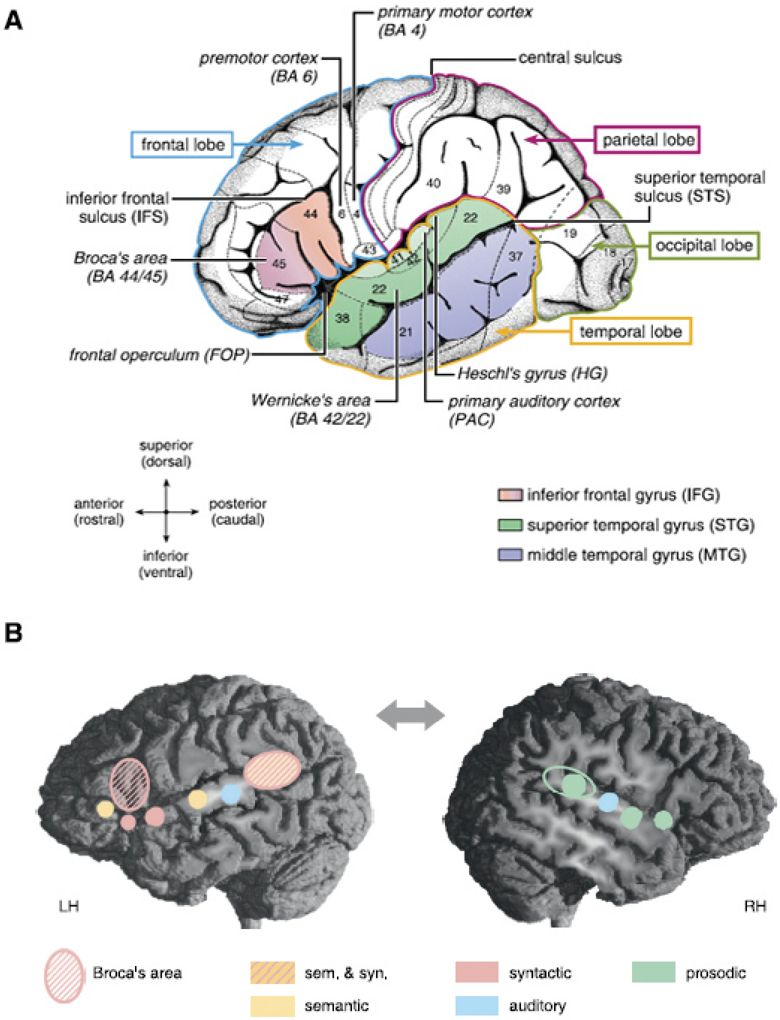

Figure 1.1: (A) Anatomical and cytoarchitectonic details of the left hemisphere. Major language relevant areas are color coded such as the superior temporal gyrus, middle temporal gyrus and inferior temporal gyrus, along with the Brodmann areas (BA). (B) Brain basis of auditory sentence comprehension (both taken from Friederici 2011). While syntactic, semantic and syntactic-semantic integration processes are supported by networks in the left hemisphere, the right hemisphere supports prosodic information processing. Syntax (in the left) and prosody (in the right hemisphere) interact via the posterior portion of the corpus callosum. For further explanation see text and Friederici (2011).

speech processing: originating from PAC and extending in non-auditory regions, such as motor, premotor and prefrontal regions (see Figure 1.2).

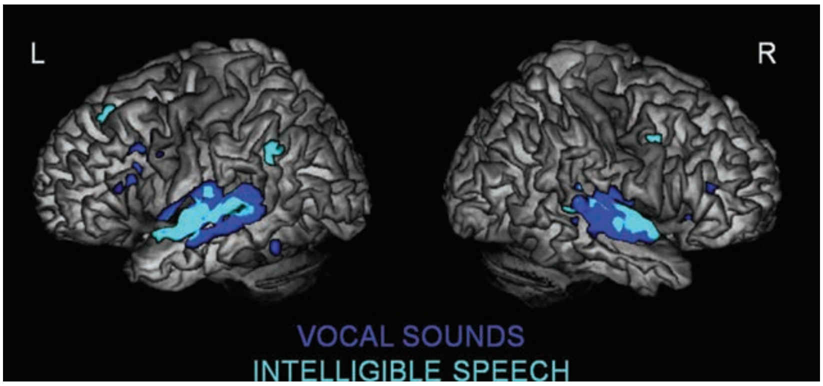

Figure 1.2: Activation Likelihood Estimation maps showing clusters of activity related to vocal sounds (dark blue) and intelligible speech (pale blue; taken from Samson et al. (2011).

By contrast, C. J. Price et al. (2005) argued that there are "no micro-anatomical structures dedicated to speech in the human brain" (C. J. Price et al., 2005, p. 275) – the regions were shared with the processing of other, non-verbal stimuli. They base their claim on a meta-analytic comparison of speech (i. e. sentences and words) with different kinds of stimuli such as meaningful sounds or music melody and pitch chroma. As they found overlapping areas for music and speech, this topic is further discussed in Chapter 3.

To further explore the involvement of the temporal lobe in speech and language processing for the current thesis, a quick glance at lesion studies is useful. Patients with lesions in the posterior part of the STG suffer from Wernicke's aphasia (Wernicke's area in pSTG; Eggert, 1977), exhibiting fluent speech, but displaying impaired repetition and impaired auditory comprehension (i. e. word repetition, non-word repetition, digit span and phonological discrimination). Only recently it has been shown that the impairment is not only restricted to the acoustic-phonological analysis (associated with pSTG) but was also found for semantic cognition (associated with pMTG and angular gyrus), suggesting a dual deficit in Wernicke's aphasia (Robson et al., 2012).

To conclude, an auditory hierarchical organization has been suggested, with primary auditory areas on the superior temporal plane responding rather indistinctively to all sounds, and more anterior, lateral and ventral association areas in the STG and STS showing sensitivity to spectro-temporal complexity and linguistic intelligibility (Binder et al., 2000; Scott et al., 2000; Narain et al., 2003; Poeppel, 2003; Giraud et al., 2004). Different aspects of speech and language processing could be defined in the temporal lobe, but the demarcation of speech to other meaningful sounds is still a matter of debate. In the following chapters of this thesis, the demarcation to music will be discussed.

1.1.2 Processing of Pitch in Speech

In speech, the closest analog to melodic contour is the trajectory of fundamental frequency (F0) over time, which is commonly called intonation. Intonation is a basic part of the organization and perception of spoken language and contributes to marking the boundaries of structural units (intonational phrase boundaries) distinguishing pragmatic categories of utterances (e. g., statement, question, command), and to signaling focus (Lehiste & Meltzer, 1973; Bolinger, 1989). These linguistic prosodic cues can help disambiguate syntactically ambiguous sentences through structural segmentation of speech (Magne et al., 2003; Steinhauer et al., 1999). Note that language conveys a second type of prosodic information - emotional prosody, which indicates a speaker's emotional state or emphasized content on an emotional basis. Considered as an extralinguistic cue it is not germane to this thesis. The third type of linguistically relevant intonation is lexical tone (e. g., in Mandarin), which is interesting to look at in combination with musical tone and will be discussed in turn.

Neurocognitively, it appears that the right hemisphere shows a specialization for suprasegmental prosodic information whereas the left hemisphere shows a relative preference for processing segmental information (e. g., Friederici & Alter, 2004). A model on cerebral lateralization of speech concludes that (1) there is a strong bilateral sensitivity to speech perception and (2) that the right superior temporal lobes support the processing of slowly changing spectral information, namely the pitch contour in spoken utterances (Poeppel, 2003). Imaging studies suggest that the right hemisphere is responsible for the processing of suprasegmental linguistic prosody (Meyer et al., 2002, 2004; for a review see Friederici & Alter, 2004; Friederici, 2011).

Evidence from lesion studies revealed cases of specific deficits in speech prosody, i. e. 'aprosodia'. Aprosodia can occur due to damage to the right or left hemisphere (Ross, 1981; Cancelliere & Kertesz, 1990; Blonder et al., 1991), however, one has to be aware of whether linguistic or suprasegmental emotional prosodies are being probed, which was also be shown to be processed in the right hemisphere (Schirmer & Kotz, 2006). Nicholson et al. (2003) report a case of right-hemisphere stroke with preserved segmental speech perception, but failure to discriminate pitch or rhythm patterns in linguistic or musical stimuli as well as prosodic perception tasks (discriminating statements from questions).

It seems that the lateralization of prosody is not as fixed as other linguistic features such as e. g., syntax, and can be modulated by task demands (Plante et al., 2002; for emotional prosody: Kotz et al., 2003) and the function of prosodic pitch in language, such as tone languages. Native tone language speakers processed tonal pitch in left-lateralized areas when associated with semantic meaning, but English native speakers, being unaware of the semantic content, processed tonal pitch in right-lateralized areas (Gandour et al., 2004).

To conclude, the processing of prosodic information at the sentence level - i. e. F0 contour – recruits the superior temporal cortex and the opercular cortex in the right hemisphere (Meyer et al., 2002, 2004). The lateralization can be modulated by task demands – and the function of the pitch patterns, e. g., being music or speech related. The following chapters focus on pitch in music and compare pitch patterns in speech and song leading the discussion along.

1.2 Summary

This chapter has tried to summarize some recent opinions on the organization of language in the brain with special regard to the functional neuroanatomy of speech comprehension. When language is perceived auditorily, the temporal lobe plays an important role (e. g., Friederici, 2011; Hickok & Poeppel, 2007; Vigneau et al., 2006). Overall, activation for sentence processing was reported in the auditory cortex, posterior MTG and pSTG, the angular gyrus, the anterior TL and the IFG (e. g., Friederici, 2002, 2011; Friederici et al., 2010; C. J. Price, 2010). Starting with the spectro-temporal analysis of the incoming signal (speech and non-speech sounds) in the HG and surrounding areas (PT, PP), the signal is projected into different directions while further linguistic features (such as phonological, syntactic and semantic aspects) are processed and analyzed. The STG/STS is suggested to support phonemic information (for a review see, Hickok & Poeppel, 2007) and phonological processing (Binder et al., 2000; Indefrey & Levelt, 2004; Liebenthal et al., 2005; Scott et al., 2000), while the right STG/STS, is specialized in prosodic information (Friederici & Alter, 2004; Meyer et al., 2002, 2004). It still remains unclear whether there are areas that only respond to speech – regardless of the linguistic level of processing (syntax, semantic, prosody). Hence, in the context of the current thesis, it is of special interest how different features, like words and pitch patterns are processed in a different, but related field: music. The next chapters will focus on areas involved in music processing, also in comparison to speech.

Chapter 2

Music Processing

Processing music involves a plethora of mental processes, such as analyzing spectral and temporal information in an acoustic signal, processing structure and meaning, involving memories and emotion etc. Investigating these features especially in comparison with language helps to understand general neural mechanisms beyond single domains.

In this chapter two prevailing models of music processing (Peretz & Coltheart, 2003; Koelsch & Siebel, 2005; Koelsch, 2011) will be discussed. Brain regions that are relevant for music processing will be discussed subsequently in more detail. This will be done with special regard to how and where lyrics in songs and pitch patterns (melodies) are processed and when the temporal lobe comes into play.

2.1 Neurocognitive Models of Music Processing

Only a few models of music processing have been proposed so far (Peretz & Coltheart, 2003; Koelsch & Siebel, 2005; Brown et al., 2006; note that Brown et al., 2006 compared music and language processing and is therefore part of the next chapter). In the following paragraphs, two models will be discussed. The first is based on evidence from lesion studies which lead to musical processing deficits (Peretz & Coltheart, 2003), the second is based on neuroimaging studies (mainly electroencephalography, EEG, but also fMRI) with healthy participants (Koelsch & Siebel, 2005; Koelsch, 2011). Evidence from EEG studies allowed for a model that considers the temporal component, in contrast to the first model.

2.1.1 A Model Based on Brain Lesions and Musical Disorder

The music perception model by Peretz and Coltheart (2003; see Figure 2.1) relies on the idea of Fodorian modules (Fodor, 1983), the concept of modularity of cognitive processes. The authors argue for the existence of a music processing module – “a distinct mental module with its own procedures and knowledge bases that are associated with dedicated and separate neural substrates" (Peretz & Coltheart, 2003, p. 688). The model divides music processing into different neuropsychological components, based on evidence from brain damaged patients with impairments or preservations of specific musical abilities. It also comprises language as a different module, which is processed in parallel to music. Within the music module, pitch and rhythm are claimed being independent subsystems (Ayotte et al., 2000; Di Pietro et al., 2004; Liégeois-Chauvel et al., 1998; Peretz, 1990; Peretz & Kolinsky, 1993; Piccirilli et al., 2000; Vignolo, 2003), processed also in parallel and comprising a number of encapsulated module components. The pitch subsystem contains three modules: contour analysis, interval analysis and tonal encoding; the rhythm subsystem contains two modules: rhythm and meter analysis.

The following processing steps are suggested: After the acoustic input has underwent an acoustic analysis (note that no temporal order is concretely provided by the authors), all modules would receive the input, but only the ones that were specialized for the respective information would respond. Then both, the temporal and the pitch subsystem, would send their information to the musical lexicon (which stores a representation of all heard musical phrases) or the emotion expression analysis. These steps follow the vocal production plan, which is reported in detail in turn.

Interestingly, some of the music processing components are claimed to be music specific, as individuals with a specific impairment provide evidence for it. The pitch organization is claimed to be music specific as individuals with impaired pitch but preserved rhythm (and language etc.) processing were found. The music-specificity of rhythm had originally been unknown, but has recently been evidenced by a single case study under participation of the same author. They reported for the first time an individual with beat production and perception impairment in the musical context only. Regarding the model, a neurobiological origin for 'time' distinct from pitch in music processing was stated (Phillips-Silver et al., 2011). Note that the participants' speech rhythm perception has not yet been tested. Note that the emotion expression analysis is also yet not known as music specific.

Of special interest for the present dissertation is that the model proposes two distinct processing modules: one for music and one for speech. For example, if the acoustical input is a song that contains music and lyrics, it is assumed that the lyric information of the song (the output of the early acoustic analysis) is sent to the parallel working language processing system. This assumption comes from evidence that song and speech are processed differently, e. g., preserved singing abilities in non-fluent aphasic patients (Hebert et al., 2003; Racette et al., 2006) or "speech arrest but not song arrest" revealed in a transcranial magnetic stimulation (TMS) study (Stewart et al., 2001). In the next step, the 'acoustic-to-phonological conversion' forwards the information to the phonological lexicon. If the goal is to speak the lyrics, no interaction with the musical processing system is necessary - but if the goal is to sing the song, the information of the phonological and the musical lexicon will be integrated and planned for the vocal production ('vocal plan formation').

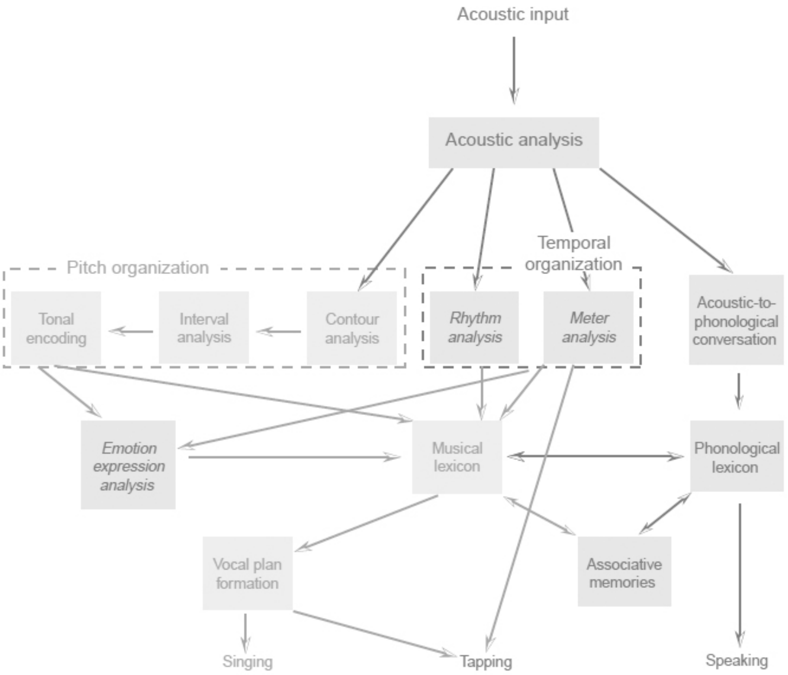

Figure 2.1: A cognitive model of music processing developed by Peretz and Coltheart (2003). The model is based on patient data with circumscribed brain lesions. It divides music processing into different neuropsychological components (grey boxes) and their communication pathways (arrows). Brain damage may disrupt either a component or a pathway. It also comprises language as a different module, which is processed in parallel to music.

To mention more processes that are only involved in music processing, the specificity of music processing is furthermore described in Peretz and Hyde (2003).

2.1.2 A Model Based on Healthy Individuals

While the model by Peretz and Coltheart (2003) is based on lesion studies, the model by Koelsch and Siebel (2005; and the updated version by Koelsch, 2011) is based mainly on studies with healthy participants and accounts for several processing steps in music perception and their neural correlates, also specifying the temporal order of music processing (see Figure 2.2).

The auditory input of music passes through different processing stages and engages a multitude of cognitive functions, comprising the basic acoustic analysis, memory and emotions; it even affects the autonomic nervous system, the hormonal and immune system.

After the feature extraction in the auditory brainstem (the superior olivary complex and the inferior colliculus), where both spectral and timing information are processed (e. g., Langner & Schreiner, 1988), and the thalamus, the acoustic information reaches the auditory cortex. Here the feature extraction of basic information takes place, as pitch height, pitch chroma, timbre, intensity and roughness (e. g., Griffiths et al., 1999; Tramo et al., 2002; Warren & Griffiths, 2003) - the acoustic information is transformed into a percept (e. g., frequency information into pitch height or chroma). While for most of these features further evidence regarding the exact localization and functional integration is needed, evidence for spectral and temporal processing is available; bilateral auditory cortices responded, but with a left hemisphere predominance for rapidly changing acoustic (temporal) information and the right for spectral information (Zatorre & Belin, 2001; Zatorre et al., 2002; Jamison et al., 2006; Hyde et al., 2008; Perani et al., 2010). These observations gave further indication for a hemispheric classification for speech to the left hemisphere (higher temporal resolution needed) and music to the right (higher spectral resolution needed). After the feature extraction, the acoustic information is stored in auditory sensory memory in which BA 6, 44 and 45 are involved, possibly “due to their role in attention processes, sequencing, and working memory processes” (Koelsch, 2011, p. 4). In music perception, grouping (*Gestalt* formation) has been demonstrated for timbre and loudness, proximity of pitch or onset time and continuation of pitch (see Bregman et al., 1990). After the process of grouping, a more fine-grained analysis of chords, melodies and time intervals follows. Pitch relations in melodies and chords are processed successively in a melody or simultaneously in a chord. The interval representation was found to activate the anterior and posterior supratemporal cortex bilaterally (Peretz & Zatorre, 2005; Liégeois-Chauvel et al., 1998; Patterson et al., 2002), complimentary, the planum polare was found to be sensitive to pitch intervals and sound sequences (Patterson et al., 2002; Warren & Griffiths, 2003; Zatorre et al., 1994). On the other hand, the contour representation as part of the *Gestalt* formation, was found posterior to Heschl’s gyrus in the right hemisphere, in line with the finding that the planum temporale was found to be crucial for auditory scene analysis and stream segregation (i. e. the identification of auditory objects; Griffiths & Warren, 2002).

In the model by Koelsch and Siebel (2005), Koelsch (2011), structure building processes of harmony, meter, rhythm and timbre follow next. Besides the anterior STG (Koelsch, Fritz, et al., 2005; Koelsch et al., 2002) and the ventrolateral premotor cortex (Janata et al., 2002; Parsons, 2001; Koelsch et al., 2002; Koelsch, Gunter, et al., 2005), music-syntactic processing activates BA 44 bilaterally, but with a right-hemispheric weighting. The activation of BA 44 was stronger in response to irregular and therefore unexpected chords, as the harmonic relation was less familiar than for more regular occurring chords (Koelsch, Fritz, et al., 2005). Furthermore, meaning and emotion are assumed to interact with all of these processes and take place during all stages of music perception. Music can activate representations of meaningful concepts and – as speech – can transfer meaning (Koelsch et al., 2004; Steinbeis & Koelsch, 2008b).

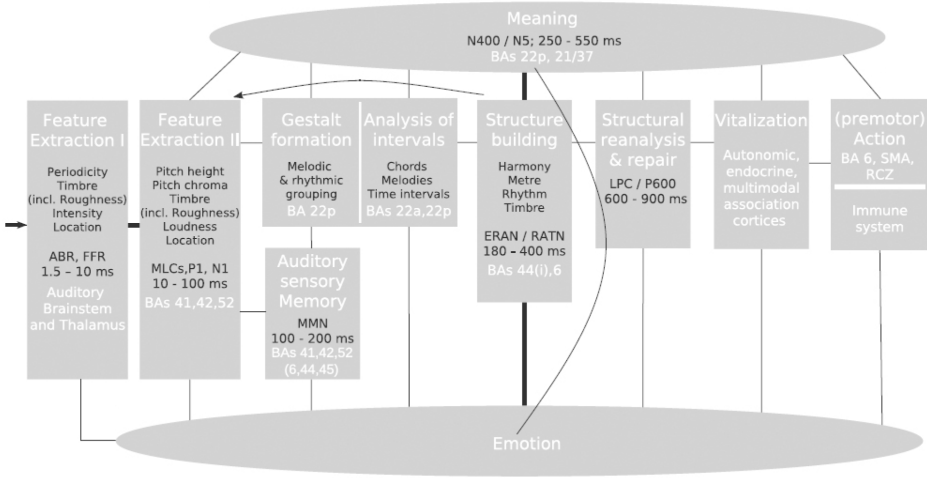

Figure 2.2: A neurocognitive model of music processing developed by (Koelsch & Siebel, 2005, updated by and taken from Koelsch, 2011). The model is based mainly on data of healthy subjects and provides information about the time course (based on EEG/MEG data) and the neurocognitive correlates (based on fMRI/MEG data) of single processing steps.

To conclude, the models represent two prevailing views on music processing. Based on these, the neuroanatomical underpinnings of a most crucial feature of music will be discussed in more detail – pitch; and a special case that transports linguistic information into music – lyrics in song. As rhythm is also an important feature it will be discussed together with speech in Chapter 3. The following sections will discuss the question: What is specific for pitch, melody and linguistic processing in music and when does the temporal lobe come into play?

2.2 Functional Neurocognitive Architecture

2.2.1 Processing of Words in Music

Most research on music has been conducted with instrumental music. But there is one field where linguistic information finds its way into music: song. In the following paragraphs, research on the relationship between tunes and lyrics will be discussed. Peretz and Coltheart (2003) proposed that the lyrics of song are processed independently and in parallel to the musical features (i. e. pitch and rhythm) in the 'language processing system'. Evidence comes from (1) preserved singing abilities in non-fluent aphasic patients (Hebert et al., 2003; Racette et al., 2006), (2) the dissociation of lyrics and tunes in song memory that have been shown after temporal lobe damage (S. Samson & Zatorre, 1991; Peretz, 1996; Hebert & Peretz, 2001), or (3) in healthy participants where different brain signatures were involved when listening to familiar songs that showed melodic or semantic incongruities (Besson et al., 1998).

In contrast, interactions between the processing of verbal and musical information in song were reported (Bigand et al., 2001; Lidji et al., 2010; Schön et al., 2005; R. L. Gordon et al., 2010; Sammler, Baird, et al., 2010). Sammler, Baird, et al. (2010) suggested an integrated processing of lyrics and tunes. In an adaptation paradigm healthy participants listened to four different sets of stimuli: songs with the same tunes and same lyrics, with the same tunes but different lyrics, with different tunes but same lyrics and with different tunes and different lyrics. Adaptation effects for lyric repetition could be found along the STG and STS with a larger cluster extension in the left hemisphere, adaptation effects for tune repetition were observed in similar areas along the STG and STS with a larger activation in the right hemisphere. Therefore, lyrics and tunes were found to interact in parts of the STG/STS. Furthermore, the authors suggest a 'gradient of integration' from more to less integrated processing along the posterior to anterior axis of the left STS, demonstrating in a similar way the hierarchical order of auditory processing in the STG (e. g., Patterson et al., 2002; Scott & Johnsrude, 2003). A strong integration was observed in the left mid-STS and a weaker integration extended more anteriorly and posteriorly in the left and also the right STG/STS. No integration (no significant interaction) was observed in the more anteroventral part of the STS and a predominance of lyrics over tunes (see Figure 2.3). As song combines musical and linguistic features and they therefore suggestively share the vocal representation in the brain (e. g., production and sub-vocal rehearsal), the left dorsal precentral gyrus (BA 6) was found as an area for integration of lyrics and tunes.

Studies on musical imagery also showed a different lateralization depending on whether imagery was carried out with or without lyrics, in which the secondary auditory cortex was reliably found (Halpern & Zatorre, 1999; Zatorre et al., 1996; Kraemer et al., 2005; Yoo et al., 2001). Bilateral activation was found when familiar songs with lyrics were used (Zatorre et al., 1996) and right sided activation when instrumental music was used (Halpern et al., 2004). The absence of lyrics was suggested to cause the shift to the right and underlines the crucial role of the right auditory cortex in pitch processing (Zatorre et al., 2002).

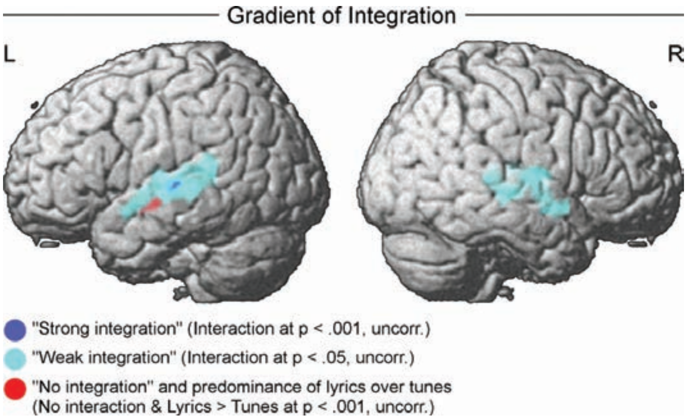

Figure 2.3: Posterior-anterior gradient of integration of lyrics and tunes (taken from Sammler, Baird, et al., 2010). The left mid-STS showed a strong integration and in the STS anteroventral to this, no integration, and a predominance of lyrics over tunes.

Other studies compared sung and spoken sentences directly to show word processing in speech and music. Before going into more detail, it should be noted that these approaches are infected by additional activations due to melodic and rhythmic processes. A clear activation pattern for word processing in song is only possible if underlying parameters such as melody and rhythm are eliminated first. Such an approach was done by Schön et al. (2010), who compared three-syllabic sung words with a vocalized stimulus (same melody sung on syllable) to isolate brain activations specific to words in sung stimuli. For the words in song (Schön et al., 2010 referred to it as ‘speech’ or ‘language processing’) a pronounced left-hemispheric involvement in the STG, STS and MTG was found, also activations in the left IFG (BA 9, 45 und 47) and the precentral gyrus (BA 6, more ventral compared to Sammler, Baird, et al., 2010, as a region of separation, not integration). Referring to the function of the remaining brain areas activated by this contrast, Schön et al. (2010) suggested that these areas reflect the phonological and semantic processing of speech/language. The study did not provide a comparable contrast of spoken words compared to vocalized prosody. The current dissertation will give further insights on brain areas involved in word processing in song and speech (see Chapter 3).

To conclude, lyrics and tunes were suggested to be processed in integration to some degree but to diverge at some point. Lyrics and tunes (or words and melodies) share most probably the secondary auditory cortex and posterior and anterior parts of the STG and STS. Divergence was found for lyrics in an anteroventral part of the STS, the MTG and the left IFG - areas involved in language processing (see Chapter 1). Overall, a left-hemispheric lateralization for lyrics in song was suggested.

2.2.2 Processing of Pitch in Music

Pitch is considered the fundamental component of music (throughout different cultures; McDermott & Hauser, 2005). Pitch can be defined as “that property of sound that enables it to be ordered on a scale going from low to high” (Acoustical Society of America Standard Acoustical Terminology, cf. Randel, 1978). Pitch is closely related to frequency, but frequency is an acoustic event whereas pitch is a percept - sound wave oscillations mapped tonotopically in the cochlea and the auditory cortex.

Pitch presented (1) in a row builds up a melody (patterns of pitch over time), (2) simultaneously builds up a chord and (3) simultaneously in a row (two or more melodies simultaneously) builds up harmonies. These sequential and simultaneous pitch structures follow rules according to e. g., Western tonal music using tonal keys – the frame that defines which notes within the octave are allowed to be combined (cf. Lerdahl & Jackendoff, 1983).

To assign these different processes to brain areas is still at an early stage – especially for tonal analysis. Pitch analysis in the brain starts with analyzing the physical stimulus properties, which is common to all perceived sounds, and takes place in sub- and cortical structures of the auditory ascending pathway: periodicity (Langner, 1992), temporal regularity (Griffiths et al., 1998; Gutschalk et al., 2002) and frequency spectrum, which is represented bilaterally in a tonotopic map (running from low to high) that mirrors the pitch map in the cochlear (Formisano et al., 2003; Bendor & Wang, 2005; Pantev et al., 1989).

Exceeding the (pitch) processing in the primary auditory cortex (PAC), the information enters surrounding areas. Studying pitch strength has shown that varying the temporal structure (Gutschalk et al., 2004; Patterson et al., 2002) and the spectrum (Penagos et al., 2004) involves the secondary auditory cortex, rather than the primary auditory cortex. That region in anterolateral Heschl’s gyrus (HG) was found in many fMRI studies, e. g., subserving the processing of fixed pitch (Patterson et al., 2002), pitch direction (Johnsrude et al., 2000), pitch sequences (Warren & Griffiths, 2003) and pitch salience (Penagos et al., 2004). Together with converging evidence from primates (Bendor & Wang, 2005), these results suggest the existence of a ‘pitch center’ (Griffiths, 2003) in lateral HG that represents the pitch of individual notes. This ‘pitch center’ seems to represent pitch as a percept, rather than the physical properties of the stimulus - this is a critical point about pitch, as the relationship between pitch as a percept and its physical attributes is still under debate (Plack et al., 2005).

The divergence from PAC happens with increasing stimulus complexity. In line with results from fMRI studies, a hierarchy of pitch processing seems to exist. While noise is processed in HG and fixed pitches (vs. noise) in secondary auditory cortex, pitch sequences (variable vs. monotonous pitch) exceed into anterior and posterior STG with different hemispheric weighting (Griffiths et al., 1998; Patterson et al., 2002; see Figure 2.4). Posterior

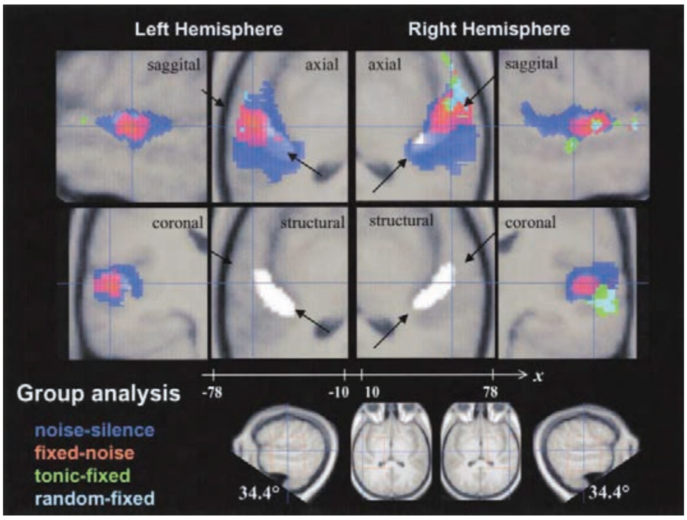

Figure 2.4: Pitch processing in superior temporal cortex (taken from Patterson et al., 2002). Averaged maps of Heschl's Gyrus of nine participants are shown in white, changes in brain activity (BOLD response) are shown in color. Noise versus silence (blue) activate primary and secondary auditory cortex, planum polare and planum temporale. Fixed pitch versus noise stimuli in a similar spectrum (green) activate secondary auditory cortex. Diatonic (green) and random (cyan) melody versus fixed pitch in a monotonic sequence activate lateral Heschl's gyrus, posterior superior temporal gyrus and planum polare.

regions of secondary auditory cortex process pitch height and anterior regions process pitch chroma or pitch class (e. g., all the notes called 'C' on a piano keyboard; Warren & Griffiths, 2003; see also Tervaniemi & Huotilainen, 2003 for converging evidence from the event-related potential, ERP). Interestingly, no significant difference was found between random tone sequences and tone melodies. Maybe, the rule-based system of tonal music is not applicable at this point – and in this contrast. When rule-based structures are tested in music processing, frontal lobe areas come into play. Experiments with fulfilled and unfulfilled expectations in chord sequences (Koelsch, Gunter, et al., 2005) and orchestral music compared with scrambled music (Levitin & Menon, 2003) revealed activations in frontal areas (e. g., BA 44). This is overlapping with syntax processing in speech (e. g., Friederici, 2011), but with a different hemispheric involvement: for speech on the left, for musical syntax bilaterally with a right-hemispheric weighting.

Many findings have shown that musical pitch processing preferentially involves right auditory cortical structures. For example, studies of brain-lesioned patients have shown that the right auditory cortex is critical for melody discrimination (Milner, 1962), perception of missing fundamental pitch (Zatorre, 1988), direction of pitch change (Johnsrude et al., 2000) and in using melodic contextual cues in pitch judgments (Warrier & Zatorre, 2004).

Consistent evidence comes from neuroimaging studies of normal subjects, showing that right secondary auditory regions are central in various aspects of musical pitch processing, such as in melodic processing (Patterson et al., 2002; Zatorre et al., 1994), imagery for tunes (Halpern & Zatorre, 1999), pitch judgments within melodies (Zatorre et al., 1994), discrimination of pitch and duration in short patterns (Griffiths et al., 1999), reproduction of tonal rhythmic patterns (Penhune et al., 1998), timbre judgments in dichotic stimuli (Hugdahl et al., 1999) and detection of deviant chords (Tervaniemi et al., 2000).

The perception of melody in terms of its contour has revealed different results regarding lateralization. With ‘contour’, the ups and downs of a melody are described. The contour can change on a global level, defining the ups and downs, and on a local level, defining the precise distance between pitches. Based on the mean task performance of lesion patients testing the detection of the global and the local contour, the results indicated global structure processing in the right and local structure processing in the left hemisphere (Peretz, 1990; Liégeois-Chauvel et al., 1998). This result was challenged by an fMRI study with healthy participants, showing a bilateral posterior STS (pSTS) involvement for the local contour (interval differences with preserved direction) and a left pSTS involvement for the global contour (violation in interval direction; Stewart et al., 2008). By revisiting the aforementioned patient studies, Stewart et al. (2008) showed that the results revealed variance on single level and the proposed lateralization can therefore only account for some of the patients. Furthermore, the study by Stewart et al. (2008) is further evidence for a cortical hierarchy of melody processing. Contrasting the local with the global structure revealed the right planum temporale (PT) and the pSTS for local structure, suggesting that “the global structure of a pitch sequence acts as a ‘framework’ on which the local detail is subsequently hung” (Stewart et al., 2008, p. 1).

A recent study by Lee et al. (2011) investigated melodic contour processing with a multivariate (MVPA; see Chapter 6.1.1) approach. Comparison of diatonic ascending and descending melodic sequences (5 notes per sequence) revealed discriminating brain regions in right mid-STS, left inferior parietal cortex (IPC) and anterior cingulate gyrus (ACC). These results underline the importance of the right mid-STS in melodic contour processing. Further, the left IPC has also been found in other studies using musical stimuli (e. g., Janata et al., 2002; Foster & Zatorre, 2010; Klein & Zatorre, 2011), suggesting that the IPC is somewhat important in melodic processing (for further discussions see Chapter 3.6).

It is challenging to take all these results together and form a full picture of musical melody processing in the brain. Stewart et al. (2006) meta-analyzed case studies with acquired amusia and summarized the functional architecture of (1) pitch interval, (2) pitch pattern (melodies) and (3) tonal structure in the musically deficient brain. Results are listed here shortly, for a detailed explanation see Stewart et al. (2006). (1) For impaired pitch difference detection, subcortical structures, the ascending auditory pathway or PAC were mentioned; for impaired pitch direction discrimination, the right lateral HG. This is in line with the aforementioned studies, as the frequency spectrum is represented in PAC and PT (e. g., Formisano et al., 2003; Penagos et al., 2004) and the lateral HG, representing a ‘pitch center’ (Griffiths, 2003). (2) Analysis of pitch patterns was associated with lesions posterior to HG (PT and parieto-temporal junction) and in anterior STG; discrimination of pitch patterns was found after right but not left temporal lobectomy, in line with a suggested pitch processing hierarchy in the STG/STS (e. g., Patterson et al., 2002; Stewart et al., 2008). (3) Tonal analysis was disturbed with lesions in right lateral HG, PT, parieto-temporal junction, insula, anterior STG and frontal operculum (see Figure 2.5), in line with a more right lateralized processing of musical pitch (e. g., Patterson et al., 2002; Zatorre et al., 2002) and the involvement of higher-order areas, such as frontal areas, in rule-based music processing (e. g., Koelsch, Gunter, et al., 2005).

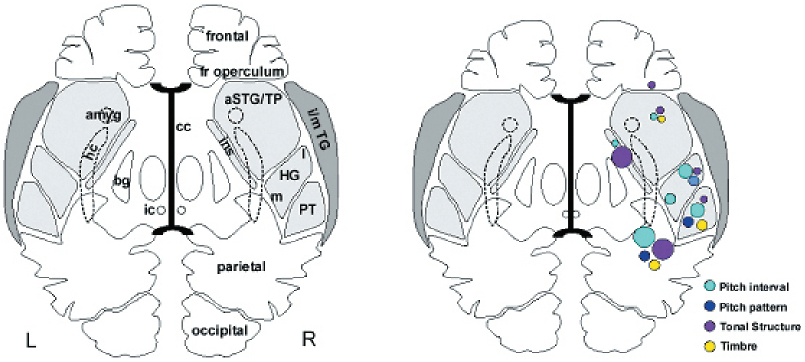

Figure 2.5: Critical brain substrates for musical listening disorders across studies (taken from Stewart et al., 2006). The schematic axial brain section includes all key anatomical areas involved in music listening, the superior temporal plane (light grey) and the middle/inferior temporal gyrus (dark grey, in exploded view) are color coded. Labels are included on the left graphic. The right graphic shows critical brain substrates for pitch processing: pitch interval (light blue), pitch patterns (dark blue), tonal structure (purple) and timbre (yellow). amyg = amygdala; aSTG = anterior superior temporal gyrus; bg = basal ganglia; cc = corpus callosum (in black); fr = frontal; hc = hippocampal; HG = Heschl’s gyrus; ic = inferior colliculi; i = inferior; ins = insula; l = lateral; m = medial; thal = thalamus; PT = planum temporale; TG = temporal gyrus. For further information see Stewart et al. (2006).

2.3 Summary

Music engages the brain on many levels. Subcortical and cortical structures are involved in processing the different features comprising music such as pitch characteristics, melody, harmony, rhythm, meter, structure, emotions, semantics and even more. How the brain treats pitch patterns (melodies) and how the temporal lobe comes into play, is of special interest for the current dissertation.

The right secondary auditory cortex is central in various aspects of musical pitch processing, suggestive of a ‘pitch center’ in lateral HG (Griffiths, 2003). Pitch patterns, i. e. melodies, are distributed along the STG and STS, following a pitch processing hierarchy along the anterior STG with a right-hemispheric predominance (Patterson et al., 2002).

Furthermore, an ongoing debate is whether pitch and rhythm are independent processes or intertwined. A cognitive model on music processing claimed that the temporal organization was processed independently from pitch (see model by Peretz & Coltheart, 2003, Chapter 2.1.1). Levitin and Tirovolas (2009) stated that a prevailing view was that pitch, rhythm, and loudness are processed separately and then come together later (where ‘later’ in neural processing time may be 25-50 ms later) to give us the impression of a fully realized musical object or phrase. Moreover, a double dissociation of music and language has been claimed (Peretz & Coltheart, 2003), but also integrated processes for lyrics and melodies were found in the anterior and posterior STG/STS, with a strong integration in left mid-STS (Sammler, Baird, et al., 2010).

To conclude, music consists of spectral and temporal properties, forming melody, rhythm and (in song) lyrics/text - this is congruent with speech processing. Therefore, comparing both domains is quite obvious. The next chapter will compare the features of both domains and lay out the state-of-the-art research.

Chapter 3

Music and Language

When music was recognized as a unique tool to investigate the brain's functions, it was obvious to compare it with language – to gain closer insights of one through the other.

In how far they are homologues or even constitute separate domains is still a matter of debate. In the previous chapter, two models on music processing were presented. Peretz and Coltheart (2003) claimed for separate modules of music and language processing. Koelsch (2011) demonstrated similarities between music and language on all levels referring to the language comprehension model by Friederici (2002; the updated model of 2011/2012 is described in Chapter 1) which is in close correspondence with Koelsch's music processing model.

3.1 A Double-Dissociation of Music and Language

Peretz and Coltheart (2003) claimed for parallel modules of music and language processing evidenced by special, mostly individual cases which show exclusive impairments in either the music or the language domain: aphasic patients without amusic symptoms (i. e. impairment of music perception) and amusics without aphasic symptoms.

'Aphasia without amusia' was reported in professional (classical) musicians (for an overview see Sergent, 1993), of which Maurice Ravel (1875–1937; Amaducci et al., 2002; Henson, 1988) is probably the most popular one. He suffered from a progressive cerebral disease of uncertain etiology with an assumed prominent involvement of the left hemisphere. Aphasia, apraxia, agraphia, and alexia and the loss of the ability to compose were appendant - but he was still able to recognize and remember his own work up to a few weeks before his death. Another case is the Russian composer Vissarion Shebalin (1902–1963; Luria et al., 1965) who had two strokes. In Shebalin's case a postmortem examination of the brain revealed impairments in left temporal and inferior parietal regions. Suffering from a ‘severe sensory aphasia’, he accomplished some of his most important compositions. The organist Jean Langlais (1907–1991) was blind from the age of two, but counts as one of the most successful French composers. After cerebral hemorrhage in the left temporo-parietal region, he became aphasic, alexic and agraphic but kept composing and improvising with no loss in the ability to read musical notation (in braille; Signoret et al., 1987). Two more cases can be mentioned, which are less well reported and both composers died shortly after the initial illness. Benjamin Britten (1913–1976), a British composer, had a stroke from which he suffered briefly from aphasia, which did not seem to influence his musical skills (Henson, 1988). George Gershwin (1898–1937), had a glioblastoma in the right temporal lobe. The course of illness was only brief with no obvious signs of cerebral dysfunction (Carp, 1977). He died after an operation at the age of 38.

To summarize, with a (probably) left-hemispheric impairment, all cases were reported to be still in possession of their musical abilities. The cases with the best foundations were Shebalin and Langlais; in Ravel’s case however, a postmortem brain examination was denied and only Langlais was tested systematically on his abilities, which makes those cases more anecdotal than beneficial in a scientific context. Furthermore, all reported cases were professional musicians, which questions that aphasia without amusia exists in the general population. This would be of special interest, as significant functional and anatomical differences between musicians and non-musicians have been shown (e. g., Münte et al., 2002; Gaser & Schlaug, 2003). A generalization from these cases, leading to different modules of music and language processing, is therefore questionable.

On the other hand cases of ‘amusia without aphasia’, which showed musical impairments without language comprehension deficits, have been reported more frequently and were tested systematically. Even without brain injuries, musical disabilities have been described, in so called congenital amusia (for details see Chapter 5). Reported cases (amusics) showed impaired pitch, melodic and sometimes also rhythmic processing. Most of the studies showed preserved speech processing abilities, which lead Peretz and Coltheart (2003) to the conclusion for a double-dissociation of music and language. Although, some studies showed impaired speech processing, for example word discrimination problems in mandarin amusics (Liu, Jiang, et al., 2012) or discrimination problems with questions and statements (Liu, Jiang, et al., 2012; Patel et al., 1998). In a very recent study though, Peretz et al. (2012) proved with a statistical learning test that amusics learned new words, but failed to learn musical material.

To conclude, amusia without aphasia is more defined than aphasia without amusia – those special cases however (as lesions and congenital impairments) are also influenced by normal interindividual differences (between amusics this was the case in Patel et al., 1998) and the question remains if these cases are strong enough to be taken as evidence for a double-dissociation of music and language. Besides these special individual cases, a variety of studies with also healthy participants were conducted to disentangle the relationship of music and language – and showed considerable commonalities.

3.2 A Model of Music and Language Processing

The discussion about a double-dissociation of music and language is challenged by a variety of studies with mainly healthy participants that showed similarities and overlap between both domains (see below). A model of music and language processing (Brown et al., 2006), which can be seen as a compromise in the discussion about the double-dissociation as it lays out shared, parallel and distinct features between music and language, will now be described, supplemented by findings from different studies.

Shared features. Shared resources and overlapping activations are suggested for primary auditory processing (acoustic input) in the primary and secondary auditory cortex (BA 41/42), and vocal motor-somatosensory processing (motor output/vocalization) in the primary motor cortex (BA 4; Brown et al., 2006).

Parallel features. Patel et al. (1998) already suggested similar brain mechanisms for music and speech as they share the use of structured pitch and rhythmic patterns to create meaning. Music and speech have the same underlying parameters of sound elements: pitch (contour and interval), duration (temporal features, as rhythm and meter), articulation, timbre, loudness and spatial location (see also Levitin & Tirovolas, 2009). Out of these, a limited number of discrete elements (e. g., phonemes and notes) are formed. These discrete elements form rule-based melodic and rhythmic phrases, unfolding over time. Phrases are generated following a set of syntactic rules that define how the discrete elements are arranged (rule-based sequencing). These sequential, combinatorial arrangements help the parser to decode the auditory stream (of both domains) and to resolve ambiguity in a sentence or a musical piece.

Irregularities in music and language were shown to be detected in the same way: processing of syntactically unexpected words or chords revealed strong overlap of music and language (for language: Friederici et al., 1993; Hahne & Friederici, 1999; for music: Koelsch et al., 2000; Koelsch & Siebel, 2005). This can be explained by the fact that both domains rely on structures and principles which we have acquired already in early childhood (e. g., McMullen & Saffran, 2004).

To conclude, the combinatory generativity of complex sound sequences (phonology) in music and language is suggested to be represented in parallel and partially overlapping in homologous brain regions of the STG/STS (BA 22; sensory center) and the IFG (i. e. BA 44/45; motor center). Therefore, music and language show parallels at the level of syntax and pitch contour (Brown et al., 2006, see also Swain, 1997; Patel, 2003a).

Distinct features. Brown et al. (2006) suggested divergences between language and music at the level of meaning (or semantics; note that also *parallels* in meaning were suggested, see Koelsch et al., 2004; Steinbeis & Koelsch, 2008a, 2008b) and a domain-specificity for distinctive, information bearing functions is suggested to be in temporal areas, such as middle and inferior temporal gyrus as well as the temporal pole (BA 20, 21, 38) and temporo-parietal regions (BA 39, 40). Also distinct between music and language are the discretization of pitch and time (see also Fitch, 2006). As reported in the literature (e. g., Patel et al., 2008), music employs isometric rhythms and discrete pitch, and language utilizes words (except the special case of song). The following paragraphs will focus on these differences between music and language in word, pitch and rhythm processing.

3.3 Word Processing in Music and Speech

Words find their way into music in songs and play a significant role in the opera and musicals. While music is possible without words (instrumental music), speech is not; despite phrases that are meant to mimic someone else's statement, produced on a hummed or vowel sound, but which only work in a very clear context. Sung and spoken utterances express meaning through words and thus share the phonology, phonotactics, syntax and semantics of the communicated language (Brown et al., 2006). Moreover, as both stimuli are vocal, the production processes of song and speech share the same vocal tract. But the use of the 'articulators' such as lips, tongue, lower jaw, pharynx and larynx, which determine the resonant characteristics of the vocal tract (Sundberg, 1996), differ during the production process of both. While during song production the articulators remain more stable (with a usually lowered larynx and a widened pharynx; Sundberg, 1996), during speech processing they move more continuously (co-articulation processes). Furthermore, as song has to be transported over a longer distance, articulation is usually more precise than during speaking. Most typical, the vowels remain in a more stable position for a longer time than during speech, as the vowels carry the sound in song (for an overview see Seidner & Wendler, 1978). As a last important sound characteristic specific to song, the so called 'Singer's Formant' (Sundberg, 1970) in professional singing has to be mentioned. As it is important for the singing voice to compete with the orchestral sound, this special formant is a strong resonance (around 2500-3000 Hz) which is added to the fundamental frequency and the vowel formants, which allows the voice to sound more powerful and brilliant and thus to project outwards into a concert hall.

Neurocognitive studies unraveling the pure ‘word’ processing *between* music (i. e. song) and speech have not been reported yet, i. e. statistically comparing sung and spoken sentences or words, corrected for their pitch and rhythm information.

Two fMRI studies contrasted a sung word or phrase with a vocalized or hummed stimulus with the same melody, to unravel vocalization processes (Özdemir et al., 2006) and perception (Schön et al., 2010). Özdemir et al. (2006) revealed in a production task (repeating two-syllabic words or phrases with one second per syllable) a bihemispheric network for ‘intoned’ (= sung) words, with activations in the pre- and postcentral gyrus, mid-STG, STS and the IFG bilaterally. Schön et al. (2010) revealed in a perception study with a same-different task a stronger left-hemispheric network, with the STG/STS and middle temporal gyrus (MTG) responding bilaterally and the precentral gyrus and the IFG (BA 45, 47, 9) in the left hemisphere (the study by Schön et al., 2010 is reported in more detail in Chapter 4).

Özdemir et al. (2006) suggested (not tested) similarities between the sung words vs. hummed melody and the speaking vs. silence contrasts, claiming that the areas involved in the singing output were the same as for speaking. Schön et al. (2010) assumed that the IFG involvement represents linguistic processing in song. Both studies were missing of a contrast speech vs. prosody to unravel regions involved in word processing in speech and to compare words in song and speech.

It is of note that comparisons between sung and spoken utterances were reported (e. g., Callan et al., 2006; for a review of the literature see Chapter 4), but they do not reflect pure word processing in both, as pitch patterns and rhythm differences in both stimuli confound the linguistic processing.

Differences reported between song and speech are usually interpreted as reflecting the ‘musical’ aspect (mostly referring to the pitch domain) in song – note that a speech over song contrast has not yet revealed significant results. Because of the lack in research regarding linguistic processing between song and speech, it has been investigated in the current dissertation (Chapter 11).

3.4 Pitch Processing in Music and Speech

When investigating the acoustical features of the pitch patterns (pitch contour, melodies) in music and speech, a fundamental difference remains in the gliding, continuously changing pitch in speech and the opposed discrete pitch in music; and while musical melodies are built around a stable set of pitch intervals (according to Western tonal music), spoken prosody is not (e. g., Patel et al., 2008).

Regarding speech prosody and musical melody, Zatorre and Baum (2012) suggested two mechanisms for processing pitch information: one focuses on contour processing, a more coarse-grained processing and another focuses on a more fine-grained, accurate pitch encoding. While the contour processing may overlap across music and language, a more fine-grained pitch encoding may be specific to music. In line with this, contour processing has been suggested to form a framework on which the ‘local’ information (i. e. interval or fine-grained processing) is subsequently hung (Stewart et al., 2008; see also Chapter 2.2.2). Contour in spoken prosody indicates for example a question or statement, in music the end or the continuation of a musical phrase.

Therefore, music processing might require both processing mechanisms contour and interval processing, and spoken prosody requires only contour processing. In music, the ups and downs of a melody as well as the discrete relations between pitches have to be processed in order to analyze musical scales, depending on rules of Western tonal music. As spoken prosody is not relying on a discrete and stable set of pitch intervals, only contour processing is necessary in order to interpret the signal. One may suggest that the discrete relations of musical intervals require fine-grained processing and should therefore be processed differently than prosody.

Neuroimaging studies though have mostly shown overlap for musical melody and prosody. EEG studies revealed that musical training can enhance pitch perception in linguistic material (e. g., Besson et al., 2007; Magne et al., 2006; Moreno & Besson, 2006; Schön et al., 2004; Chobert et al., 2011). In fMRI studies, prosodic information and melodic contour were both found to engage the right STG (in separate studies, for prosody: e. g., Friederici & Alter, 2004; Friederici, 2011; Meyer et al., 2002, 2004; for melodic contour: e. g., Ayotte et al., 2000; Liégeois-Chauvel et al., 1998; Peretz, 1990; Schuppert et al., 2000).

In the above mentioned EEG studies however, the musical training enhanced both contour and interval perception, but if the linguistic material only required contour processing this might explain the overlap. Keeping in line with that, the assumed overlap in the fMRI studies might be due to shared contour processing. This needs further investigations.

In a meta-analytic review, speech was compared with other sound systems (C. J. Price et al., 2005). The anterior STS bilaterally and the (left) posterior STS were suggested as shared areas of speech and music. The authors suggest, as musical melodies as well as prosodic contours need to be integrated over longer time frames, the overlap in anterior STS was most likely due to pitch pattern processing in music and language: melody and prosody. It is of note that the analysis did not reveal a significant lateralization of speech and music. Interestingly, this finding was corroborated by a recent meta-analysis by Schirmer et al. (2012): While speech showed a relatively small lateralization effect to the left, music did not reveal a significant lateralization at all, suggesting equal hemispheric mechanisms for music processing.

On the basis of contour and interval processing, lateralization of prosody and melody might also have to be reconsidered. As discussed in Chapter 2.2.2, studies have reported divergent evidence regarding the lateralization of interval and contour processing. Overall, the contour analysis of melody and prosody should reveal the same neural correlates, but they should diverge at some point, revealing brain areas representing the interval processing required in musical melody. Candidate structures to observe are the anterior and posterior STG/STS and the IPC (e. g., Stewart et al., 2008; Lee et al., 2011).

To conclude, musical melody and prosody differ acoustically in their pitch relations (glide vs. discrete) that should rely on different processing mechanisms in the brain. So far, similarities for both have been assumed and overlap has been shown. This might rely on the shared contour processing mechanism and studies were not able to tease out differences regarding interval processing, i. e. musical melody requires presumably a fine-grained, accurate pitch encoding to analyze the discrete pitch relations. Studies on the gliding and discrete tone relations are required to unravel these detailed differences in which temporal and parietal areas are of special interest.

Evidence from Lesions and Musical Disorder. As already mentioned in the beginning of this chapter, patient data form the basis of a proposed double-dissociation of music and language. Almost all of the presented patients exhibited musical impairments without language impairments – so called amusics (for more details and single case studies, see Chapter 5). Amusics have been shown to exhibit pitch and melody processing deficits. As speech contains prosody, a discussion is pending on the amusics' prosodic perception, compared to musical melody, and if the lesion site can give further insights on involved brain areas.

Musical impairments and prosodic impairments have been reported after right-hemispheric lesions (for music: Liégeois-Chauvel et al., 1998; for prosody: Pell & Baum, 1997b; for both: Nicholson et al., 2002, 2003; Patel et al., 1998), proposing shared mechanisms for prosody and music. Patel et al. (1998) for example reported a case with impaired musical contour and speech intonation processing with extended lesions in both temporal lobes and the right frontal lobe. An overlap of pitch processing in speech and music was suggested in the STG bilaterally (e. g., Patel et al., 1998).

On the contrary, a patient was reported who developed amusia without *aprosodia* after a right frontal lobe resection for intractable seizures (McChesney-Atkins et al., 2003). This case suggests that the right hemisphere has different sites for musical ability distinct from the centers regarding prosody.

Furthermore, authors that argue in favor of a strong dissociation between music and language skills presented cases with amusia that only exhibited impaired processing in the music domain. Amusia can occur after brain damage (e. g., Peretz et al., 1994; Peretz & Belleville, 1997; Hebert & Peretz, 2001; Steinke et al., 2001), but can also be congenital (i. e. without brain damage; e. g., Ayotte et al., 2002). Therefore, these speacial cases have problems with the processing of the discrete pitch in music and showed even an *advantage* for gliding pitch in speech (e. g., Liu, Xub, et al., 2012; see also Chapter 5).

To conclude, cognitively, amusics exhibited a dissociation between prosody and musical melody; neuroanatomically, the literature is diverse, as the lesion sites are extended and mixed results were reported. Studies with healthy participants on the other hand suggested overlap between musical melody and the prosodic aspect of speech, mainly in the STG/STS with a more right-hemispheric involvement.

In a direct comparison though – in the same study with comparable conditions –, the acoustic differences of the pitch patterns of music and speech, i. e. discrete and gliding pitch, should reveal differences in brain activity between prosody and musical melody. This approach was taken in the current dissertation, using sung and spoken stimuli (Chapter 11).

3.5 Rhythm Processing in Music and Speech

Rhythm is known to be an important feature of both, music and speech. This section will discuss similarities and differences on the acoustical and neuroanatomical level. In addition, an example taken from the stimulus set of the current dissertation will illustrate how speech and musical rhythm can be aligned in song.

Patel (2008) defines rhythm as "the systematic patterning of sound in terms of timing, accent, and grouping" (p. 96). A commonality was found on the level of grouping. Grouping in both domains means that smaller units group together to form larger rhythmic chunks or phrases (Wightman et al., 1992). In speech, words are not perceived as being equally separated, neither are tones in music. Tones and words are grouped together and organized and perceived as phrases. Even some of the acoustic cues indicating the boundaries between phrases are similar, such as pitch drop and durational lengthening (Todd, 1985). The pitch drop, for example, has traditionally been used as a main indicator for the end of a musical piece and also for the end of a spoken statement (P. J. Price et al., 1991; Narmour, 1990); this acoustical cue has also been used in the current dissertation for both, the sung and the spoken phrases (for the sheet music of the stimuli see Appendix).

That these phrase boundaries are perceived equally in speech and music has been shown in neuroimaging studies. An EEG (electroencephalography) component that reflects the perception of phrase boundaries – the closure positive shift (CPS) – was found in both domains (for language: Steinhauer et al., 1999; for prosody: Steinhauer & Friederici, 2001; Pannekamp et al., 2005; for music: Knösche et al., 2005).

While grouping is common to both, periodicity has been claimed to be a core difference between rhythm in music and speech (Patel, 2003b; Patel et al., 2008). Music is perceived as more regular than speech and leads people to synchronize their movements, e. g., dancing or just tapping their foot or moving their head. This is typically not caused by speech. Note that there are special cases demonstrating overlap in music and speech, as rap music or poetry, where speech actually does induce a beat. As these two types of creative expression are even possible, the question arises how ‘irregular’ speech actually is.

Classically, languages such as German and English were classified as ‘stressed-timed languages’, in which stressed syllables are perceived as prominent (in contrast to ‘syllable-timed languages’ as French and Italian). It has been claimed that they are characterized by (roughly) equal temporal intervals between stresses (Pike, 1945), following the isochrony hypothesis. However, pure measurements of speech did not reveal evidence for periodicity, i. e. regular intervals between stressed syllables, and the isochrony hypothesis has been abandoned in the 1980’s by e. g., Levelt (1989; for further references see Bertinetto, 1989 and Patel, 2008). Lehiste (1977) made the interesting suggestion that periodicity was maybe stronger in perception than production. Therefore, the alternation of stressed and unstressed syllables may function as a temporal and structural cue in speech comprehension which allows predicting when the next stressed syllable is likely to occur (Large & Kolen, 1994; Mattys, 1997). For communication, regularity in speech rhythm was shown to be important for the facilitation of semantic sentence processing by regular meter (see below; Rothermich et al., 2012) and by the interaction of rhythm and higher level linguistic faculties, such as syntax (Schmidt-Kassow & Kotz, 2009).

Music also shows alternations of stressed and unstressed events (e. g., tones) which follow a regularly timed beat (or an isochronous pulse) and are grouped in multiples of two or three in terms of how many beats constitute a basic unit. In contrast to rhythm, this regular occurrence of beats is defined as meter (cf. Lerdahl & Jackendoff, 1983). Therefore, temporal periodicity in musical meter is much stricter than in speech and the periodicities of music allow meter to serve as a strong mental framework for music perception. For example, syncopated rhythms in music, where prominences are on an off-beat, are usually perceived as metric. As stresses in speech are not as regular as in music, a phenomenon as syncopation is not allowed in speech.

To conclude, rhythm differences in music and speech concern mainly the periodicity (Patel, 2008), i. e. the metric conception. As pointed out by Patel (2008) “there is no evidence that speech has a regular beat, or has meter in the sense of multiple periodicities”

		Je-	de	Nacht	um	die-	se	**Zeit**		kommt	der	Mond	her-	vor.
speech	trochaic	1	2	1	2	1	2	**1**		1	2	1	2	1
song	4/4 time	1	2	3	4	1	2	3	**4**	1	2	3	4	1

Table 3.1: Schematic example of a spoken and sung realization of the same sentence in a metrical grid, taken from the current stimulus material. The trochaic pattern is disrupted by the combination of two one-syllabic, stressed words, 'Zeit kommt'. In speech no time interval in between needs to be preserved as opposed to song, where the verbal and the musical accent needs to be aligned.

(p. 194). But, linguistic metrical grids are not abstract periodic mental patterns (like musical metrical grids) but are simply maps of heard prominences, full of temporal irregularities. Brown and Weishaar (2010) described the differences in terms of a 'metric conception' for song as opposed to a 'heterometric conception' for speech. For the current dissertation it is of interest how speech and musical rhythm come together. In song, in a process called 'textsetting' (Halle & Lerdahl, 1993), words have to be set to metrical music without violating the pattern of stronger and weaker syllables in speech (in stress-timed languages). This alignment facilitates the comprehension of lyrics and enhances musical beat tracking by synchronizing neural activity with strong syllables (R. L. Gordon et al., 2011).

An example from the current thesis (see Table 3.1) is a sentence with mostly a trochaic rhythm (two syllables following each other, first strong, second weak). In the middle of the sentence the one-syllabic, stressed word 'Zeit' is followed by another stressed word 'kommt'. In speech, these two stressed syllables can follow each other without preserving a distinct time interval in between. In music however, the word 'kommt' should occur on the next strong beat, otherwise it would sound unnatural. Therefore, the word 'Zeit' needs either to be lengthened or a pause needs to fill the gap to align the verbal and the musical accent. This example demonstrates briefly the metric conception (as suggested by Brown & Weishaar, 2010) in music and the heterometric conception in speech and why music is perceived as more regular than speech.

Functional Neuroanatomy. After the initial analysis of the incoming sound for periodicity (Langner, 1992) and temporal regularity (Griffiths et al., 1998; Gutschalk et al., 2002) in the auditory system, rhythm perception and production engages regions in the cerebellum and basal ganglia (BG; Ivry & Keele, 1989; Janata & Grafton, 2003; Schwartze et al., 2011; Kotz et al., 2009; Kotz & Schwartze, 2010), the thalamus, as well as motor areas such as the premotor cortex (PMC) and supplementary motor area (SMA; Halsband et al., 1993). Investigations of meter and rhythm processing report activations in the posterior IFG and the STG (e. g., Aleman et al., 2005; Koelsch & Siebel, 2005; Platel et al., 1997; Riecker et al., 2002; Tillmann & Bharucha, 2002; Zaehle et al., 2004; Geiser et al., 2008; Liégeois-Chauvel et al., 1998; Chen et al., 2008). Activation for pure naïve passive listening to musical rhythm is reported by Chen et al. (2008) who found activations in bilateral planum temporale (PT) and midline SMA, bilateral mid-PMC, and bilateral cerebellar lobule IV. The most important finding was that motor regions were even activated during passive listening. Another interesting study in this context was conducted by Grahn and Brett (2007) who compared regular and irregular rhythms. Specific to rhythms in which accents arise at regular intervals (which give the feeling of a simple beat), activation in the BG, pre-SMA/SMA, and the anterior STG (aSTG) bilaterally were found.

Internal timekeeper. In a behavioral experiment Grahn and Brett (2007) showed that metrically simple rhythms were more accurately reproduced by the participants, speaking for the idea of an internal timekeeper in the brain (McAuley & Miller, 2007). The 'timer' might therefore be beat-based and metrically regular. Such a timer in adults may preferentially support the processing of small-integer ratio temporal intervals typical of Western music (cf. Ivry & Hazeltine, 1995; Poeppel, 1997). Moreover, it is also known that the auditory system continuously searches for regularities in the acoustic signal and, once detected, allows the listener to form predictions about upcoming events (Winkler et al., 2009). To conclude, if stimuli induce a stronger perception of a beat they are easier to recall (Grahn & Brett, 2007) and activate the BG, the cerebellum and auditory as well as motor regions - maybe forming an internal timekeeper in the brain.

Taken together, song and speech share rhythm as an underlying feature, which show similarities at the level of grouping and differences at the level of periodicity. The brain seems to have an advantage in processing regular (musical) rhythms, speaking for an internal timekeeper that involves the BG, SMA and cerebellum, PMC, PT and aSTG bilaterally.

3.6 The Role of the Temporal Lobe

The auditory cortices and adjacent association areas in the superior temporal lobes are crucial for the analysis of linguistic material, pitch and pitch patterns and rhythm as could be seen in the above review of the literature. The importance of the TL in music and speech processing has already been reported in detail. Within this section the crucial areas in the TL and their functions, with importance for the current dissertation, are summed: the role of the secondary auditory cortex (AC), the anterior and posterior STG in processing pitch sequences (belonging to both domains) and speech.

While all incoming sound is first analyzed in the primary auditory cortex (PAC), the secondary AC was suggested to constitute a 'pitch center' in the brain as all sorts of pitch processing could be observed here (e. g., fixed pitch, pitch direction, sequences). A pitch processing hierarchy (e. g., Patterson et al., 2002) was suggested with increasing anterior and posterior STG involvement in pitch complexity, e. g., random and tonal pitch sequences.

These effects engage more the right hemisphere than the left, as evidenced by lesion studies. For example, the discrimination of pitch patterns was found after right but not left temporal lobectomy (Liégeois-Chauvel et al., 1998). The analysis of pitch patterns was impaired by lesions posterior to Heschl's Gyrus (HG; PT and parieto-temporal junction) and in aSTG. Furthermore, the discrimination of unfamiliar melodies that differed in one tone has been shown in unilateral (S. Samson & Zatorre, 1988; Zatorre et al., 1995) and bilateral anterior TL lesions (Satoh et al., 2005). Also in anterior TL, processing rhythm with a regular beat and grouping were found to activate the PT and aSTG (Chen et al., 2008; Grahn & Brett, 2007).

Regarding the posterior STG, activations were reported when contrasting the passive listening to melodies vs. noise (e. g., Zatorre et al., 1994). Patients with unilateral cortectomy of the posterior STG (sparing HG) showed pronounced deficits in the processing of sequential pitch variations, demonstrating the critical role of this region in melody perception (e. g., Liégeois-Chauvel et al., 1998).

Speech. First, acoustic phonological processes take place in the PAC and the PT. The information is then forwarded to the STG/STS and distributed to the anterior and posterior part. A region anterolateral to HG was found to differentiate between speech and non-speech sounds and even further, the aSTG was found to react to intelligible speech (e. g., Scott et al., 2000). The posterior STG and PT were involved in processing basic acoustic characteristics. From the aSTG extending in frontal regions, the aSTG and the left frontal operculum were suggested to form a network for initial local structure building. Using a left-lateralized temporo-frontal network, semantic and syntactic relations are processed in parallel. To be exact, the semantic network involves the middle and posterior STG and MTG and also BA 45 and BA 47; the syntactic network (complex sentence structures) involves the anterior and posterior STG/STS and BA 44. Syntactic and semantic integration processes are assumed to involve the STG/STS and the BG. Regarding lateralization, sentence comprehension involves a left-lateralized temporo-frontal network for syntax and a less strongly lateralized network for semantic processing.

Areas in the temporal lobe were suggested to be selectively sensitive to speech sounds. A meta-analysis suggested speech-related activity in lateral non-primary superior temporal regions, specifically in posterior STG and anterior and middle STS (F. Samson et al., 2011). A left lateralization in the posterior and ventral STS for morphed speech sounds compared to morphed music sounds was suggested (Specht et al., 2009).

C. J. Price et al. (2005) on the other hand proposed that there were no speech-selective areas in the brain – speech would share these areas with other sound systems. Overlap between music and language was found in posterior and aSTG with no significant lateralization effect, and specifically the region in aSTG was interpreted to react to both, musical melody and prosody, as the aSTG is involved in integration of information over longer time frames (C. J. Price et al., 2005).

In a recent meta-analysis however, claims for differences between sound classes and the combinatorial nature of the anterior pathway in the STG could not be verified, e. g. vocalized stimuli against music, and spoken words against environmental sounds revealed greater activation in the STG, but did not demarcate an anterior pathway. Also, acoustic complexity, i. e. contrasting phonemes/syllables and tones, could not be shown to activate aSTG more than the other (Schirmer et al., 2012).

Therefore, disentangling music- and speech-related processes in the temporal lobe is only possible in contrast, i. e. one region is more activated by music than speech, but does not stand for exclusiveness, as other sounds (e. g., environmental sounds, voice) might share this region with music or speech.

3.7 The Role of the Parietal Lobe

The parietal lobe is known as an association area, subserving numerous functions, such as attention, orientation, spatial location, spatial and non-spatial working memory, mental imagery and rotation and task-switching (for an overview see Cabeza & Nyberg, 2000; Culham & Kanwisher, 2001). The following paragraphs will lay out a new parcellation of the inferior parietal cortex (IPC), its role in verbal and tonal working memory and attention, and a novel approach to melodic pitch processing.

The IPC is defined in 7 cytoarchitectonic areas (Caspers et al., 2006, 2008) which are grouped in 3 regions (Caspers et al., 2012; see Figure 3.1C). The intraparietal sulcus (IPS) constitutes the dorsal border of the IPC, separating the IPC and the superior parietal cortex (SPC). In the anterior middle IPS three areas are defined: the human intraparietal area 1 (hIP1) and 2 (Choi et al., 2006) and the hIP3 (Scheperjans et al., 2008). They are located in the depths of the IPS, with hIP1 and 2 at the lateral bank of the IPS and hIP3 at the medial wall of the IPS, at the boarder of hIP2 (see Figure 3.1B).

A most influential model on working memory (WM) hypothesizes the existence of a central executive, which controls the incoming information and its distribution to two subsystems: the phonological loop, which is responsible for storing verbal material, and the visuospatial sketchpad, which is responsible for the integration of visual input, spatial information and object properties (i. e. color and size; for an overview see Baddeley, 2012). Which function of this model is associated with which area is still under debate. Classically, the IPS is involved in spatial WM, while the supramarginal gyrus (SMG, ventral part of the IPC) is involved in phonological store (Baddeley, 2003). In a recent meta-analysis the IPC was found to be related to shifting attention within WM (Nee et al., 2012). A more attentional role of the anterior end of the IPS as opposed to a role in storage of information (which was found in posterior IPS) has been suggested in visual WM (Trapp & Lepsien, 2012). The IPS has also been suggested as a multimodal or ‘amodal’ region of WM capacity as it was involved in WM maintenance regardless of the modality of the stimuli (Cowan et al., 2011); this notion is confirmed by studies on the macaque brain (for a review see Grefkes & Fink, 2005).

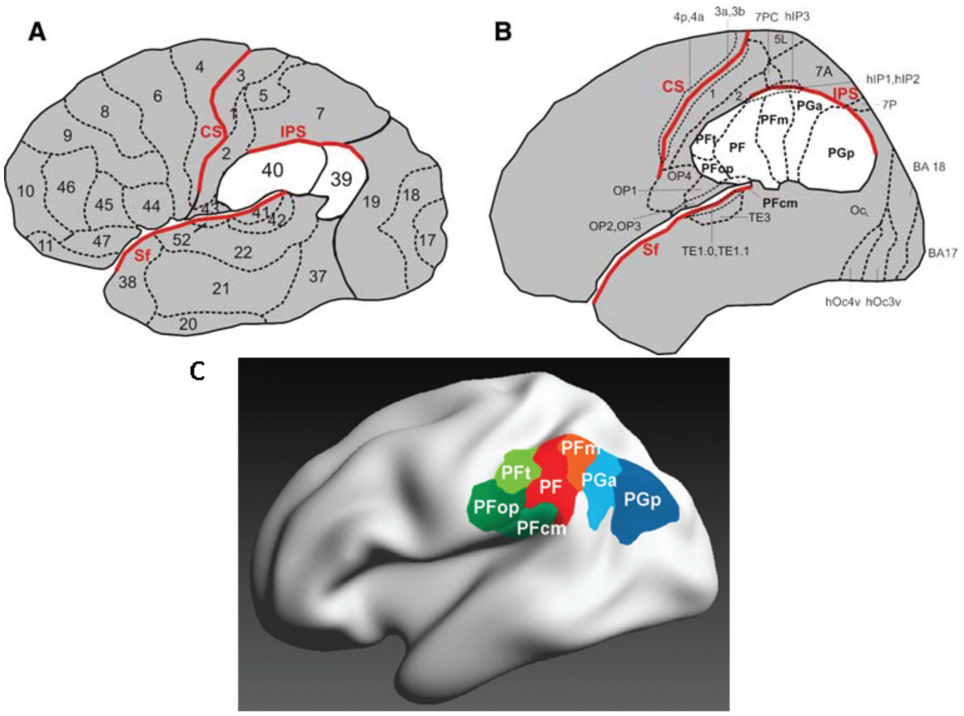

Figure 3.1: A: Schematic drawing of a lateral view of Brodmann's brain map. B: Schematic drawing of the intraparietal cortex, areas provided by the parcellation by Caspers et al., 2006, 2008 and of surrounding areas as published (e. g. hIP1-2: Choi et al., 2006; hIP3: Scheperjans et al., 2008). C: The new parcellation by Caspers et al., 2012, including the 7 areas and the 3-region model. Areas are color coded corresponding to the receptor-based cluster segregation: rostral cluster (areas PFt, PFop and PFcm): shaded green; middle cluster (areas PF and PFm): shaded red; caudal cluster (areas PGa and PGp): shaded blue. Pictures A and B taken from Caspers et al., 2008; C from Caspers et al., 2012.

Another recent meta-analysis underlines this notion by showing no significant difference and even a conjunction in the IPS between verbal and non-verbal material (Rottschy et al., 2012). In this meta-analysis working memory 'core' areas were observed and anatomically labeled by reference to probabilistic cytoarchitectonic maps of the human brain: bilaterally in dorsal BA 44 (extending into the premotor cortex), anterior insula, (pre-) SMA, and IPS (left areas hIP1-3, right hIP3) and lateral prefrontal cortex (LPFC), located anterior to BA 45 in the right, slightly overlapping with BA 45 in the left hemisphere (Rottschy et al., 2012).

Two related studies investigated the verbal and tonal WM (Koelsch et al., 2009) in musicians and non-musicians (Schulze et al., 2011). Verbal and tonal WM showed overlap in recruiting the typical WM network, including frontal and parietal areas. Activation peaks were reported in BA 40 (IPC, SMG) and BA 7 (SPC). Converting the reported coordinates and mapping them on the cytoarchitectonic map (Caspers et al., 2006, 2008, 2012) revealed the following: Verbal and tonal WM activated area PFt in the IPC (BA 40), but only the verbal WM extended in more dorsal areas (PFm and hIP3 in the anterior IPS). Both areas were significantly more strongly activated for the verbal than the tonal WM. This implies that the tonal WM might not extend in the anterior IPS.

This observation is interesting in light of recent studies which have reported activation for musical pitch in areas hIP2 and PFm (Foster & Zatorre, 2010; Klein & Zatorre, 2011), in which the above authors tried to exclude WM demands as a reason for the activation. To further explore the involvement of the IPS in musical pitch tasks (also in comparison with phoneme tasks), relevant studies are reported in detail in the following.

The role of the IPS in music or melody or pitch processing was already discussed in a study by Zatorre et al. in 1994 in which the pitches of the first and last notes of melodies were compared with listening to unfamiliar tonal melodies. In that study they could not rule out the task influence but speculated "given the widely acknowledged role of inferior parietal regions in spatial processing that a recoding of pitch information may be taking place during the performance of this task" (Zatorre et al., 1994, p. 1916). A row of subsequent studies found IPS involvement in musical tasks, suggesting that the IPS may play a role in dealing with the frequency relationship between stimuli. IPS involvement was found in discrimination tasks for large pitch shifts (Rinne et al., 2007), in auditory feedback monitoring for vocal regulation following pitch-shifts (Zarate & Zatorre, 2008) and furthermore suggesting an interaction with the right posterior STS in order to extract the directionality of a pitch-shift (Zarate et al., 2010). Getting closer to the exact role of the IPS in musical pitch processing, two recent studies examined relative pitch in melodies and chords while trying to control for other influential factors, such as task demands and cognitive load. As these two studies are in close correspondence to the current work, they will be described in more detail.

In an fMRI study Foster and Zatorre (2010) investigated the neural substrates of relative pitch. Participants were required to perform a same/different discrimination task on simple and transposed melodies, together with two control tasks, a rhythm discrimination and a phoneme discrimination task. Shared activation for all tasks was found in auditory areas and the ventrolateral frontal cortex (VLFC) and distinct activation for the transposed melody task in the IPS. As all tasks shared cognitive demands, such as auditory perception, working memory and a discrimination judgment, the authors argued against the IPS involvement due to cognitive load per se by showing that the BOLD signal was not related to e. g., the number of items used in the melody task. Furthermore, Foster and Zatorre tested for task difficulty (between the two melody tasks) and ruled out differences in attentional demands or effort-related factors. They even demonstrated a correlation between a (higher) IPS activation and (better) behavioral performance. As the anterior IPS was only activated in the transposed melody task, the authors suggest that the IPS reflects relative pitch encoding which is a fundamental element in music perception. This interpretation will be discussed in more detail together with the result from another study, which will now be described.

Klein and Zatorre (2011) investigated categorical perception in music, using stimuli that crossed the minor/major boundary compared to stimuli that changed absolute frequency (the whole chord changed in frequency, not only one tone within the cord). The authors used comparable paradigms (habituation and adaptation) as previous studies on the categorical perception of phonemes (Liebenthal et al., 2005; Joanisse et al., 2007). While phoneme tasks showed activation in left STS, Klein and Zatorre found for the music task activation in the right STS and additional activation in the IPS bilaterally. Neither Liebenthal et al. nor Joanisse et al. found the IPS in the phoneme studies. Interestingly, the above mentioned study by Foster and Zatorre (2010) did not reveal IPS activation for the phoneme task either. Klein and Zatorre suggested that this together with the findings by Foster and Zatorre point to a role of the IPS in differentiating interval types in stimuli. As Klein and Zatorre compared two similar oddball stimuli (in an adaptation paradigm) it can be assumed that those most probably did not differ in cognitive load; however, Klein and Zatorre did not discuss this possibility, as the focus of the study was on the involvement of the right STS and the IPS activity was not a result they had predicted.

Taken together, the IPS might become a region on the list of predictable areas involved in musical pitch processing. It is interesting that the IPS was found in studies using musical pitch tasks, but not in comparable non-musical (e. g., phoneme) tasks. While one most probably can never be sure in how far a different cognitive demand (e. g., working memory, monitoring, difficulty, attention) influences the activation of the IPS, the evidence that points to a role of the IPS in encoding musical pitch in some way, is too strong to be negated.

3.8 Summary

The possibility for a double-dissociation of music and language processing was discussed and individuals with impairments in one dimension but not the other were reported as evidence. Amusics revealed problems with processing the discrete pitch relations in music but showed an advantage for processing the gliding pitch in speech (Liu, Xub, et al., 2012). On the contrary, evidence from studies with healthy participants revealed overlap between music and language (e. g., by comparing leading music and language processing models, such as Koelsch, 2011; Friederici, 2011).

Music and language share undoubtedly the primary auditory analysis, processing periodicity, spectral and temporal information. Overlap between both was suggested in the use of structured pitch and rhythm patterns (e. g., on the level of grouping). Distinct features are the discretization of pitch and time. A fundamental difference remains in the gliding, continuously changing pitch in speech and the opposed discrete pitch in music. It has been suggested that both, musical melody and prosody require a more coarse-grained processing to analyze the contour, but only musical melody requires a more fine-grained, accurate pitch encoding to analyze interval relations, assuming different neural correlates for melody and prosody (Zatorre & Baum, 2012). However, musical melody and prosody were suggested to be processed in overlap in the (anterior) STG (e. g., Patel et al., 1998; C. J. Price et al., 2005) in the right hemisphere (e. g., Nicholson et al., 2003; Patel et al., 1998; Meyer et al., 2002, 2004; Liégeois-Chauvel et al., 1998). Regarding the temporal domain, music and speech differ in their metric conception, with music being periodic (following an isochronous pulse), while a more heterogenous conception for speech rhythm was suggested (Brown & Weishaar, 2010).

Furthermore, music and language share the linguistic dimension - words in music occur in song. As no direct comparison of the word processing in song and in speech has yet been reported, results from related studies suggest that sung and spoken words (being linguistic features) should be processed in overlap. It is of note that song and speech differ in the articulation of (mainly) vowels, which remain longer in a stable articulatory position in song, maybe causing differences in the neural representation. Thus, it is of value to compare words in speech and lyrics in song to test the relationship between music and language.

The next chapter will engross the thoughts about the relationship between music and language through focusing on their closest approximations – song and speech.

Chapter 4

Song and Speech

The separability of song and speech is in focus of this chapter. Therefore some evolutionary and developmental considerations, as well as evidence from production and perception studies, will be laid out trying to draw a more complete picture of both features.

As stated, our use of language might be the most distinguished feature between us and any other being – not so with singing. Song is agreed to be present in many non-human species, as for example birds, whales and gibbons – overall the ability to sing are supposed to have about 5400 species, and even some are able to learn and even compose new songs and also to improvise (for an overview see Wallin et al., 2000; for birdsong see e. g., Marler, 1970). This fact that singing is not unique to the human species, as language, one might assume that singing as a form of communication was used before speaking, but still, it is not known when the human species started singing and speaking (Altenmüller & Grossbach, 2002).

In children's development singing and speaking emerge in parallel – both without special vocal training or musical or language tutoring (for an overview see Welch, 2005). Children sing spontaneously around the age of one and by the age of 5 they have developed a state of singing ability that will remain in this state, if not further trained, by adulthood (Dowling, 1999). Thus, comparing singing and speaking means comparing two phenomena that are deeply rooted in humans. Singing seems to be as natural as speaking for humans and therefore it is not surprising that singing can persist, while the ability to speak is impaired as reported in patients with severe expressive aphasia and their preserved ability to sing familiar songs although they are widely unable to speak (e. g., Amaducci et al., 2002; Hebert et al., 2003; Warren & Griffiths, 2003; Yamadori et al., 1977; Straube et al., 2008).

These considerations are interesting in the light of most imaging studies that show a stronger and wider network involved in song than in speech – it is not clear yet why this is the case. Including data and results from related subjects (e. g., song and speech in production) is therefore necessary to help understand the functions of song and speech in the brain.

4.1 Similarities and Differences between Song and Speech

As song is part of the music domain, and speech is part of the language domain, similarities and differences between music and language pertain to song and speech as well. The following paragraphs will therefore summarize some of the afore mentioned features of music and language (see Chapter 3) and describe how they directly apply to song and speech.

Acoustically, song as well as speech consist of sequential acoustic patterns which show orderly variations of pitch (intonation), stress (duration and loudness) and rhythm of elements. At this level of comparison of basic acoustic parameters, music and the prosodic aspect of speech are in close correspondence. Furthermore, both music and speech show specific and relatively fixed developmental time courses (Trehub, 2001) and all known human societies make use of both speech and music, regardless of technological sophistication.

According to the voice-scientific view our speaking and singing behaviors exploit the same voice as their main instrument (Sundberg, 1996; Thurman & Welch, 2000). The same physiological mechanisms (including the throat, the vocal folds, and the larynx) generate both speaking and singing behaviors (Sundberg, 1996, 2001). Similarities between the prosodic aspects of language and certain aspects of music have been noted for centuries (e. g., Steele, 1775; Bolinger, 1989). Such notions make an implicit or explicit link between intonational and rhythmic aspects of language and the melodic and rhythmic dimensions of music. Only later on in history, speaking and singing behaviors began to be regarded as two different sets of vocal behavior (Welch 2005).

While sequential acoustic patterns constituting song and speech are typically studied by different research communities, their relationship has long interested scholars from both fields. For example, linguists have borrowed musicological concepts in building prosodic theories (Liberman, 1975; Selkirk, 1984), and musicologists have used tools from linguistic theory to describe musical structure (Lerdahl & Jackendoff, 1983). Analyses of acoustical structure revealed that music reflects patterns of durational contrast between successive vowels in spoken sentences, as well as patterns of pitch interval variability in speech (Patel et al., 2006). Descending pitch contours and syllables or notes of long duration typically mark ends of phrases in speech (P. J. Price et al., 1991) and in music (Narmour, 1990).

Acoustic differences in song and speech can be investigated by displaying the acoustic event in a spectrogram. A spectrogram displays the frequency structure of a signal in temporal succession. Time is plotted on the x-axis and frequency on the y-axis. The spectral amplitude is plotted using a gray-scale. Higher amplitudes are darker and represent broad energy bands that correspond to the formants. A formant is a pronounced acoustic resonance and gives the vowel its characteristic timbre. The first two formants (lowest black energy bands in the spectrogram) are important to forming the vowels, but formants 3 and 4 are also influenced by the vowels but to a lesser degree (for an overview see, Lindholm & Sundberg, 2007). Spectrograms taken from the current stimulus material (see Figure 4.1) show the spectral formation of a sentence part spoken (left) and sung (right). Two visually striking differences in the spectrographic pattern can be observed: Firstly, the f0 frequency (indicated by the thin black line) reveals a more stable pitch height of the vowels in song than speech and secondly, song exhibits higher energy in the spectral band than speech, indicated by darker amplitudes overall but especially in the third and fourth formant (above the added white line). These formants are still present in speech but to a much lesser degree. To conclude, it can be stated from the spectrogram that song has a richer spectral characteristic than speech.

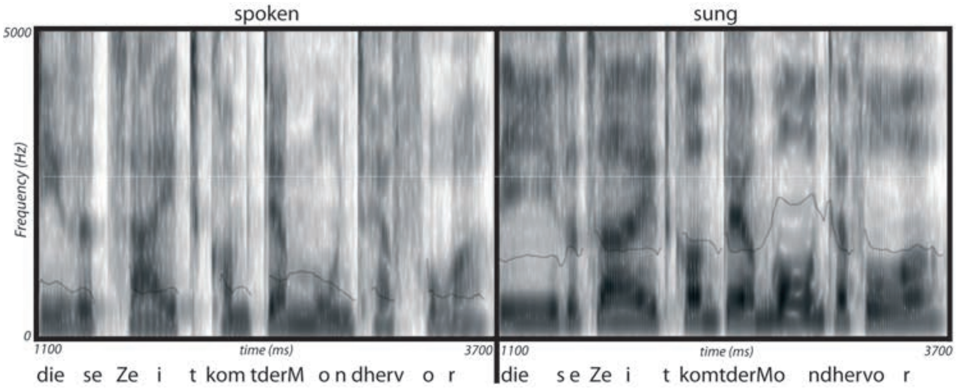

Figure 4.1: Spectrograms of stimulus set #5 from the current stimulus material with the text part 'diese Zeit kommt der Mond hervor', speech (left) and song (right). The approximate text distribution is given below the graph. Time is plotted on the x-axis, frequency on the y-axis. The spectral amplitude is plotted using a gray scale. Higher amplitudes are darker and represent broad energy bands that correspond to the formants. The thin black line indicates the F0 frequency. The thin white line was included to mark the boundary between the second and third formant.

Yet, singing as a feature of musical behavior and prosody as a feature of speech exhibit pronounced differences, on the phenomenological as well as the neurophysiological level: At the phenomenological level, song is 'quantized', i. e. (1) its pitches usually have discrete relations at n$\cdot \sqrt[12]{2}$ (semitones in Western music) and (2) its rhythm typically shows discrete onsets at integer multiples of the underlying metric beat or its subdivisions. Moreover, the pitch variations important for most types of musical processing are relatively small. For example, melodies typically use pitch changes of the order of 1/12 to 1/6 of an octave, and the subdivisions of the octave across cultures tend to be in this range (Vos & Troost, 1989). By contrast, speech typically does not have this strict ‘quantized’ quality, neither in pitch contour (which shows continuous transitions) nor in rhythmic/metric timing (except the special case of recited poetry). Pitch variations used for intonation contours in speech tend to be greater than half an octave, and specific interval relationships are not critical (Patel et al., 1998).

At the neurophysiological level, contradicting views about overlap and independent processing of language and music are held, which has already been discussed in detail in the previous chapters (Chapter 2 and 3). This probably reflects different aspects of language and music processing focused on in different studies – and moreover, it is a historical fact that the relation between the processing of crucial patterns – as melody and rhythm – in language and music has rarely been addressed empirically.

The following paragraphs will therefore review the literature on song and speech processing and comment on critical points regarding stimulus material and experimental design.

4.2 Song and Speech in Production

In the light of the similarities and differences listed above, the question remains as to what extent do song and speech share neural mechanisms. The apparent overlap between acoustical features of music and prosody and the hypothesis that this might be reflected in an ‘overlap’ in the underlying neural substrate, is further supported by some behavioral evidence, e. g., the notion that music lessons promote sensitivity to speech prosody (Thompson et al., 2003). Thus, musical training, by refining the frequency-processing network, facilitates the detection of pitch changes not only in music, but in language as well (Schön et al., 2004).

A remarkable number of clinical studies (Amaducci et al., 2002; Hebert et al., 2003; Yamadori et al., 1977; Straube et al., 2008; Pell & Baum, 1997a, 1997b; Liégeois-Chauvel et al., 1998; Nicholson et al., 2003; Patel et al., 1998) and quite a number of imaging studies on speech and song production (Perry et al., 1999; Formby et al., 1989; Wildgruber et al., 1996; Gunji et al., 2007; Jeffries et al., 2003; Riecker et al., 2000; Özdemir et al., 2006; Brown et al., 2004, 2006) suggest a neural network involving the superior temporal gyrus and opercular and rolandic cortices as major parts. All of these studies have a clear focus on the ability to speak and the ability to sing, i. e. these all mainly address the active production process only. Studies investigating the perception of song and speech, surprisingly, are sparse.

Despite the methodological problems and constraints involved with actual motor output and overt vocalization during 3D brain image acquisition, surprisingly, most studies so far have investigated singing production rather than song vs. speech perception (often using active singing, speaking, and humming as conditions). Most of the studies propose a different lateralization for singing and speaking according to the classical model of a language/left and music/right distribution. On the level of activation the results are rather diverse, most probably as a result of the different conditions used for the experiment. The paradigms using actual melodies rather than one single pitch (Perry et al., 1999), can be categorized in (1) rote singing of familiar melodies (Formby et al., 1989; Wildgruber et al., 1996; Gunji et al., 2007; Jeffries et al., 2003); (2) repetition of melodies (Riecker et al., 2000; Özdemir et al., 2006); or (3) invention of novel melodies (Brown et al., 2004, 2006). It is of note that most of the studies used well-known songs and melodies for the singing condition and highly automated word strings, such as the name of the months, or speaking the lyrics of the song as speaking condition.

Most of the studies proposed similar networks for singing and speaking, but with a different distribution: A lateralization for speech to the left and song to the right was found (Riecker et al., 2000; Jeffries et al., 2003; Callan et al., 2006) and suggested but not statistically compared (Perry et al., 1999; Brown et al., 2006); furthermore, a bilateral involvement for singing has been suggested (Brown et al., 2004; Özdemir et al., 2006; Gunji et al., 2007). In the following, activations found in the studies are reported.

One of the earliest positron emission tomography (PET) studies of *singing* suggested a similar network for speaking and singing (motor related areas and auditory areas, as well as the inferior parietal cortex, IPC) with a opposite lateralization for song (even in contrast to complex pitch perception) in right auditory areas - although that study evaluated only overt singing of a single pitch (Perry et al., 1999).

Riecker et al. (2000) used fMRI to study the production of melody, but in the absence of words. They showed a double-dissociation for insular contributions to the task; the left insula being more active for speech articulation; the right insula being more active for overt singing.

In a PET study by Jeffries et al. (2003), subjects spoke or sang the words to a familiar song. The otherwise rarely reported contrast speaking vs. singing revealed activation in 'classical perisylvian language areas' in the left hemisphere (e. g., pSTG, supramarginal gyrus (SMG) and frontal operculum). Singing, in contrast, activated areas in the right hemisphere more strongly (e. g., aSTG, STS, aMTG, insula, medial and dorsolateral prefrontal cortices). The authors claim for 'word' processing (not controlled for melody processing) during singing in the right hemisphere, which would underline the effect of fluency in speech therapy (e. g., Melodic Intonation Therapy, MIT, see below).

As mentioned above, overt vocalization is detrimental to MR and MEG acquisition procedures. One way to circumvent this issue is to examine covert rather than overt singing. Wildgruber et al. (1996) reported significant lateralization for speaking/left and singing/right (focusing on the primary motor cortex) but did not compare both conditions statistically.

Studies investigating the perceptual approach on song versus speech, surprisingly, are sparse. One study compared covert production and passive perception of singing and speaking, respectively, the song lyrics of highly familiar (Japanese) songs (Callan et al., 2006, using fMRI). Activations common to the conditions were found in aSTG, planum polare (PP), and cerebellum (this study is discussed in detail in the following section). For this specific set of tasks involving familiar song melodies and familiar song lyrics, a hemispheric advantage for speech could be demonstrated in the left hemisphere – for production and perception; and for song in the right hemisphere – for perception only, maybe also demonstrating a bihemispheric network for singing (not assumed by the authors).

While earlier studies have investigated how singing and speaking tasks are lateralized to different hemispheres of this functional network (Jeffries et al., 2003; Riecker et al., 2000; Wildgruber et al., 1996), a more recent study has discovered a shared network for sensorimotor processes in both singing and speaking, with a more bihemispheric network for singing (Özdemir et al., 2006). This is also suggested by Brown et al. (2004).

In PET studies (Brown et al., 2004, 2006) exploiting the human capabilities to invent melodies (such as harmonizing with a presented melody, ‘listen and respond’ and vocal improvisations), a functional network was recruited including primary and secondary AC, Broca’s area, supplementary motor area (SMA), insula, cerebellum, and basal ganglia (BG). Additionally, improvisation and creation of novel melodies and sentences activated BA 38, which is interpreted as being related to an even higher level of musical processing than BA 22.

It is of note that most studies did not reveal significant speaking more than singing contrasts. Nevertheless, most reported studies propose the left hemisphere to be engaged in speech production (and perception). Singing was found in the right hemisphere, but some studies lead to the suggestion of a more bihemispheric network for singing - but still with a right-hemispheric weighting.

This circumstance would explain why preserved singing abilities are helpful to recover speaking abilities after lesions in the left hemisphere. A speech therapy called ‘Melodic Intonation Therapy’ (Albert et al., 1973; Sparks et al., 1974) is really successful (according to Norton et al., 2009) when treating patients with left-hemispheric frontal lesions (in Broca’s area). The idea behind the MIT is that through enhancing right frontal areas (homologue to the impaired left-hemispheric areas) with musical or intonational stimuli, they would compensate the left-hemispheric’s function – based on the suggestion that speaking would predominantly depend on the left hemisphere and singing would be a function of the right hemisphere (Jeffries et al., 2003; S. Samson & Zatorre, 1991, 1992; Smith, 1966; H. W. Gordon & Bogen, 1974; Sparks & Deck, 1994; Cadalbert et al., 1994). This account could explain why aphasics with left-hemispheric lesions and a preserved right hemisphere are able to sing (e. g., Amaducci et al., 2002; Smith, 1966; Warren & Griffiths, 2003; Yamadori et al., 1977).

It is of note that the MIT stands in contrast to a double-dissociation of music and language as the lyrics in song belong to the language domain and were therefore claimed to be processed in parallel to music and not intertwined (see Chapter 3 for details). Evidence from patients and fMRI demonstrated that singing did not help in speech recovery (Hebert et al., 2003; Peretz et al., 2004; Racette et al., 2006; Straube et al., 2008; Stahl et al., 2011) and the brain correlates after the treatment showed for example a deactivation in the right hemisphere (Belin et al., 1996).

To conclude, despite being confronted with the typical problems of lesion studies, a successful treatment with MIT underlines the power of song and might explain the stronger representation of song in the brain than speech and also a bihemispheric network underlying song. These two main considerations – strength and lateralization – are also a main issue in song and speech perception studies, which will be discussed in the next section.

4.3 Song and Speech in Perception

As stated, only three studies involved passive listening tasks in their study (Saito et al., 2006; Callan et al., 2006), only one reported a perception study (i. e. without production task; Schön et al., 2010). The results and paradigms will now be discussed.

Saito et al. (2006) reported the activation for singing and speaking separately for a passive listening task. Listening to singing activated bilateral HG, bilateral STS, bilateral PT, left PP, right dorsal premotor cortex (PMC) and right SMA; listening to speaking activated as singing bilateral HG and STS, but only left PT and right PP. Unfortunately, as both conditions are not statistically contrasted, no conclusion can be drawn on differences between song and speech in perception.

Callan et al. (2006) compared listening to well-known Japanese songs with the same spoken lyrics. Song over speech revealed activation in aSTG, PT, and PMC bilaterally, also left orbitofrontal cortex. Especially the aSTG was found to be involved in complex pitch perception, so the authors suggest that it likely represents processing of the melodic aspects of the sung stimuli. A significant laterality effect for the right aSTG was found suggesting a right lateralized specialization for melodic processing. The reverse contrast only showed a lateralized activity to the left hemisphere on a lower threshold, consistent with a left hemisphere advantage for speech processing. Interesting for the current thesis is also the comparison of both covert production conditions. The contrast for speech production over singing production revealed a large extent of differential activity, which is unexpected as the speech over song perception contrast did not reveal any significant results. This circumstance is most probably due to the added difficulty of producing only the lyrics of well-known songs – without its associated melody. For the thesis at hand, these results lead to the following conclusions: Firstly, if contrasting sung lyrics and spoken words, unfamiliar songs should be used, as the brain response to spoken lyrics will most probably be confounded by an inner rehearsal of the associated melody. Secondly, if interested in the pure perception of song and speech, randomly mixing production with perception tasks will also most probably confound the pure listening condition with e. g., motor responses in anticipation of the next production trial. Thirdly, different task demands might occur between conditions due to the less familiar spoken production of the traditionally sung lyrics.

Schön et al. (2010) investigated the domain specificity of linguistic and musical processing (note that only the first part of the study, Experiment 1, is reported here because of the relatedness to the current dissertation). Participants listened to pairs of spoken three-syllabic words, sung words and ‘vocalise’ (i. e. singing on syllable) while performing a same-different task. Results revealed bilateral involvement of MTG, STG, IFG and middle frontal gyrus (MFG) while listening to the conditions. Activations related to music processing were investigated by contrasting song versus speech and showed a larger network in the right hemisphere. Activations related to speech processing were investigated by contrasting speech over song (which did not yield any suprathreshold clusters) and song over vocalise which showed predominance in the left hemisphere (e. g., STG/STS, MTG, IFG, i. e. BA 45, 47, 9), possibly reflecting phonological and semantic processing. However, the contrast of speech versus vocalise did not show any suprathreshold clusters, and the reverse contrast showed only motor-related regions. The authors suggest that these differences are possibly due to a somewhat greater difficulty of the task with vocalise than speech. From these results the authors conclude that the left hemisphere is specifically involved in speech processing and that the right hemisphere is specifically involved in music processing.

To summarize, song and speech stimuli (text/lyrics) in contrast with musical vocalized stimuli (only melody on syllable) were used to unravel linguistic and music processing in sung and spoken stimuli. The approach taken was partly successful, as only some of the contrasts revealed significant results. This might be caused by the following circumstances: Firstly, different task demands between conditions were assumed to have influenced the perception of the stimuli. Secondly, a prosodic condition is missing, i. e. a vocalized stimulus for the spoken stimuli. Thirdly, the contrast of song over speech was supposed to reflect music related processes – this opens the questions which music related processes were meant: for example, the discrete pitch information or the periodic beat/rhythm? Since spectral and temporal information was found to be processed in different hemispheres (Zatorre et al., 2002) this attempt is problematic as full sung and spoken stimuli contain both, melodic and rhythmic information and furthermore, the influence of the words, which are realized differently in song and speech (see Chapter 3), is also not separable. For fully separating out these processes between song and speech, an extended paradigm is necessary and a study focusing on passive listening, with low task demands – this was part of the current thesis and will be discussed in Chapter 11.

4.4 Summary

Song and speech are a conglomerate of similar spectral, temporal and linguistic features and therefore hard to discriminate on a neural basis. Lesion studies showed that one can be preserved while the other is impaired and were taken to prove a double-dissociation between music and language, i. e. lyrics are processed independently of melody (e. g., Peretz & Coltheart, 2003). Still, most neuroimaging studies have failed to show activation for speech over song, and conjunction analyses showed a significant overlap between both (e. g., Callan et al., 2006). But also, perception studies of song and speech in contrast are rare; conditions are differentially influenced by task demands, such as production or discrimination tasks. Furthermore different statistical analyses than the traditional univariate analysis should be used additionally to get closer insights in differential pattern of the brain response, e. g., multivariate pattern analysis.

To conclude, when one carefully looks at the composition of the stimulus materials used in neuroimaging studies (familiar or new songs, melodies with or without lyrics, etc.), a general pattern in the more recent studies seems to be a bilateral activation of temporal areas for both sung and spoken stimuli, with a right hemisphere advantage for musical pitch patterns, melody of song, and prosodic/intonational aspects of speech, and a left hemisphere advantage for rhythm, segmental, and linguistic aspects of both speech and song. To disentangle these acoustic features and the neural underpinnings of song and speech, a different paradigm needs to be created which takes linguistic, melodic and rhythmic differences of both into account.

Chapter 5

Deficits in Music Processing

5.1 Congenital Amusia

Congenital amusia is a lifelong impairment of music perception, in spite of normal audiograms, education, cognitive ability and memory (Ayotte et al., 2002), which affects ~4% of the population (Kalmus & Fry, 1980; Nan et al., 2010), and it seems to be not treatable – not even in the developing brain (e. g., through daily music listening; Goulet et al., 2012). Commonly known as ‘tone deafness’ (first noted by Allen, 1878, and called ‘note-deafness’) or ‘musical agnosia’ (e. g., Ayotte et al., 2000), later termed amusia (Peretz et al., 2001), it was first characterized in detail in 2002 by Ayotte et al. A significant amount of research has been done since then, although with some conflicting results.

In the beginning, the disorder appeared specific to the musical domain (pitch, rhythm and meter; Ayotte et al., 2002; Peretz et al., 2002) while perception of common environmental sounds and human voices was unimpaired. Ayotte et al. (2002) described impairments in music memory and recognition, also impairments in singing and the ability to tap in time to music. But it was observed that the temporal domain in many subjects with amusia (amusics hereafter) was unaffected, so it was suggested that selective abnormalities of pitch perception might underlie deficits in music perception (Peretz et al., 2002). This was confirmed later when amusics showed a deficit in the detection of pitch changes within a sequence of monotonic tones, but not in analogous temporal patterns (Hyde & Peretz, 2004). In further studies, when a standardized test for detecting amusia was used, there were always participants who performed normally on the temporal subtests (rhythm and meter) while failing in the pitch tasks. Only recently Phillips-Silver et al. (2011) reported the first case of an individual whose only impairment was beat production and perception in a musical context. Many further studies focused on the pitch processing deficits, using amusic participants with English, French and also tone languages as their native language. They found impaired discrimination of fine-grained pitch changes and detection of incorrect pitches in melodies (e. g., Foxton et al., 2004; Hyde & Peretz, 2004; Jiang et al., 2011; Jones et al., 2009; Peretz et al., 2002), deficits in processing contours of pitch sequences (Dalla Bella et al., 2009; Foxton et al., 2004; Jiang et al., 2010), judging/discriminating pitch direction (Foxton et al., 2004; Liu et al., 2010; Loui et al., 2008; Patel et al., 2008) and impairments of pitch memory (e. g., Tillmann et al., 2009). The pitch detection and discrimination problem of many amusics is up to and including a semitone (Foxton et al., 2004; Hyde & Peretz, 2004; Liu et al., 2010). This became also significant in musical singing; amusical singing was impaired as compared with normal performance (Ayotte et al., 2002). Moreover, amusics showed short-term memory deficits for pitch (Tillmann et al., 2009; Williamson & Stewart, 2010; for music: Ayotte et al., 2000) but have yet to show any long-term memory impairment (Jones et al., 2009).

Furthermore, the question arose, whether or not amusics are just consciously unaware of pitch changes while the implicit knowledge is intact. Imaging studies have revealed unconscious processing of pitch changes and melodic anomalies (e. g., Hyde et al., 2011; Moreau et al., 2009). With EEG Peretz et al. (2009) revealed brain activity for mistuned notes in melodies while behavioral performance was at chance level, reflecting implicit processing of musical pitch information. In detail, a brain response to quarter-tone pitch differences could be shown (indicated by an early negative event-related potential) but a lack of awareness for these fine-grained pitch changes could be shown by the absence of a late positive component which was observed in the healthy controls. In further studies, amusics showed near-normal performance in production and imitation tasks, where conscious processing of pitch information was not required (identification of pitch direction: Loui et al., 2008; lexical tone: Nan et al., 2010; speech intonation: Liu et al., 2010; note, a single case study of Tillmann et al., 2007, is reported below, see 'Acquired Amusia'). Additionally, while performing as well as controls in statistical learning tasks, amusics were less confident about their performance. The authors suggest that amusics may not be contingent on explicit knowledge formation or level of awareness to the degree shown in typical individuals (Omigie et al., 2012). Moreover, extending amusia research into musical structure processing, Tillmann et al. (2012) reported a deficit for chord processing, but also showed that amusics have internalized sophisticated syntactic-like functions of chords in the Western tonal musical system. As stated, from the beginning of amusia research, the disorder appeared specific to the musical domain.

To demarcate musical processing from other environmental sounds, studies on speech processing (speech intonation contours) in amusia have been conducted (Ayotte et al., 2002; Jiang et al., 2010; Liu et al., 2010; Patel et al., 2005, 2008; Foxton et al., 2004; Loui et al., 2008). Spoken language also relies on melodic contour perception and up until now it is not entirely clear in how far amusics' intonation processing is also impaired. Some early work suggested that amusics had no problems with the perception of speech intonation (Ayotte et al., 2002), while Patel et al. (2008) showed that about 30 % of (brain-damaged, not congenital) amusics from independent studies have difficulty discriminating a statement from a question on the basis of a final pitch fall or rise. Further studies demonstrated subtle problems with linguistic tone and intonation processing in amusics (Hutchins et al., 2010; Jiang et al., 2010; Liu et al., 2010; Nan et al., 2010; Patel et al., 2008). Unfortunately, some of the studies reported mixed results, for example, when discriminating speech intonation in natural speech and in complex tone analogs, i. e. converting intonation patterns to tone analogs through extracting the median F0 of each syllable (for details see Patel et al., 1998; Liu et al., 2010). In Ayotte et al. (2002), Jiang et al. (2010), Patel et al. (2005) amusics showed better performance for speech intonation in natural speech and opposingly in Liu et al. (2010) and Patel et al. (2008) for the tone analogs; no difference between both was found by Liu, Xub, et al. (2012).

Evidence from a production study suggested that inaccurate pitch imitation was selectively impaired in the music domain without being affected in speech (Dalla Bella et al., 2011). In general, amusical or poor pitch singing is implicated by a particular inaccuracy in producing pitch: for example familiar songs or intervals are typically sung quite far from the target. Ayotte et al. (2002) tested singing performance in congenital amusics which was rated to be more impaired on the pitch than on the temporal dimension, but which was not rated very high either. Interestingly, also manifested in production, congenital amusics showed poor-pitch singing with intact rhythm (Dalla Bella et al., 2009; Tremblay-Champoux et al., 2010).

The latest studies on 'amusical' pitch perception (Liu, Xub, et al., 2012; Liu, Jiang, et al., 2012) contrasted speech and non-speech sounds and different pitch patterns in (Mandarin speaking) amusics to examine what the 'melodic contour deafness' (Patel et al., 2008) in congenital amusia is associated with. Interestingly, they showed that amusics have a gliding pitch stimulus advantage in relation to discrete pitch stimuli, which could explain why amusics have often showed greater difficulty with musical pitch perception (discrete pitch processing needed) than speech perception (gliding pitch processing needed; Liu, Xub, et al., 2012). Furthermore, they examined the mechanisms of speech processing (Liu, Jiang, et al., 2012) where they confirmed the afore mentioned results and extended them insofar as amusics exhibit word discrimination, but not word identification problems. They stated that if tasks involve multiple acoustic cues to communicative meaning they were not impacted by amusia. Very recently though (Peretz et al., 2012) proved with statistical learning that amusics can learn new words, but fail to learn music material.

Extending behavioral perception studies, brain morphometric and connectivity studies were conducted. Congenital amusia cannot be explained by brain damage, but brain abnormalities were found. Amusics exhibited a reduction in white matter concentration in the right inferior frontal gyrus (IFG; Hyde et al., 2006) and an increase in grey matter in the same area plus the right auditory cortex (Hyde et al., 2007), as well as reduced functional connectivity between both areas (Hyde et al., 2011). These abnormalities might keep amusics from acquiring normal musical competence as the communication in a right fronto-temporal auditory pathway is disturbed. Furthermore, the right IFG showed an abnormal deactivation in amusics while listening to pure-tone, melodic-like sequences (Hyde et al., 2011). Interestingly, brain activity in the left and right auditory cortices increased as a function of increasing pitch distance in amusics, likewise their controls. As already shown with EEG (Peretz et al., 2009), this study confirms that the auditory cortices of amusics respond normally to pitch (and that the disorder might originate in higher cognitive functions). Furthermore, evidence from diffusion tensor tractography studies (Loui et al., 2009; Loui & Schlaug, 2009) revealed that musically tone-deaf people have a reduced connectivity in the arcuate fasciculus which connects the superior temporal gyrus with the premotor cortex – the sound perception to production pathway.

Besides other musical production deficits, like tapping along with music, singing in amusics is impaired, specifically more regarding pitch than temporal aspects (Ayotte et al., 2002). Recognizing a song sung by an amusic is in most cases only possible due to the lyrics, which are usually articulated correctly in time. Their inaccurate production of pitch becomes obvious when asked to produce familiar songs or imitate single pitches or intervals, as they sing quite far from the target (Ayotte et al., 2002; Tremblay-Champoux et al., 2010; Dalla Bella et al., 2011, 2012). Disturbed perception processes, motor control, deficient sensorimotor mapping and memory have been reported as causes for poor-pitch singing in the general population and amusics (Pfordresher et al., 2010; Hutchins & Peretz, 2011; Dalla Bella et al., 2011, 2012).

5.2 Acquired Amusia

In the last part the focus was on congenital amusia, but some already reported evidence for impaired music perception came from patients who suffered from brain lesions and exhibited music processing deficits, so called ‘acquired amusia’. The following paragraphs focus on the lesion sites and the resulting music processing deficits reported in different studies. This is of special interest for the current dissertation as patients with temporal lobe lesions and acquired amusia participated in the study (see Chapter 12); further interest was on implicit testing of pitch processing.

In 1994, Peretz et al. published the first two cases with acquired amusia – C. N. and G. L. both showed (besides deficits in other domains) impaired melodic but not rhythmic processing. Two years later, Peretz (1996; also Patel et al., 1998) reported that patient C. N. had recovered almost all musical abilities, despite singing from memory. In 1997 (Peretz and Belleville), patient I. R. was found and for the next ten years she became the most studied amusic case. Even with a brain damage that had been 20 years prior, she failed in most music tests (e. g., memory and discrimination tasks) while basic auditory functions and language abilities were still intact. I. R. displayed extended bilateral lesions: in the temporal lobe (among others, on the left: STG, HG, PT, MTG, SMG, insula; on the right: aSTG near the pole – sparing HG and PT – and anterior insula), in the frontal lobe (among others, precentral gyrus, IFG, lateral orbito-frontal gyrus, MFG) and in the left anterior inferior parietal cortex (IPC). Interestingly, common to all (G. L., C. N. and I. R.) were lesions in the temporal lobe bilaterally, including anterior STG, which already points to a crucial role of the (anterior) temporal lobe in music processing. Some more single cases have been reported since then: H. V. (Griffiths et al., 1997) showed unilateral lesions on the right hemisphere (for details on all lesion sites see Table 5.1) and was not able to sing from memory. N. R. and R. C. (Ayotte et al., 2000) showed also unilateral lesions on the right; they both showed deficits in all musical tasks (Montreal Battery of Evaluation of Amusia, MBEA; Peretz et al., 2003). N. N. (Satoh et al., 2005) had lesions in the anterior portion of the temporal lobes bilaterally and showed impairments in the recognition and discrimination of familiar melodies and the discrimination of unfamiliar phrases and isolated chords. R. K. (Schuppert et al., 2000), whose lesions were unfortunately not described in detail (they tested 20 patients who suffered from unilateral focal cerebrovascular cortical lesions in the frontal, temporal or parietal region) and showed strong impairments in melody and rhythm processing and performed poorly in pitch as well as in metre discrimination tasks (test was similar to the MBEA).

As depicted in Table 5.1, the STG was damaged in all patients, the anterior STG was damaged in all, except one case; half showed bilateral, half showed right sided aSTG lesions. The posterior STG (the PT in most cases) was damaged in some cases, accordingly, less prominent in causing music disabilities. Also interestingly, the HG was only reported to be damaged in some cases, two right sided and two left sided lesions.

It is striking that the reported amusic patients exhibited extended rather than focal lesions. In all cases (N = 8, including R. K.) the temporal lobe was damaged, and no case of acquired amusia without a temporal lobe damage has been reported so far. This is in line with pitch and melody processing studies (Griffiths et al., 1998; Patterson et al., 2002; Zatorre, 1998, see also Chapter 2.2.2), which all show involvement of the upper bank of the STG (HG and posterior and anterior distributions of the auditory input into the temporal

study	Peretz 1994	e. g., Peretz 1997; Patel 1998	Peretz 1994, 1996; Patel, 1998	Griffiths 1997	Ayotte 2000		Satoh 2005
patient	G. L.	I. R.	C. N.	H. V.	N. R.	R. C.	N. N.
STG	R/L	R/L	R/L	R	R	R	R/L
aSTG	R/L	R	R/L		R	R	R/L
pSTG/PT		L		R	R	R	
HG	L	L			R	R	
MTG	L	L			R		R/L
insula	R/L	R/L			R	R	L
IPC	L	L		R	R		
frontal	R/L	R/L			R		

Table 5.1: Lesion overview of single cases with acquired amusia after brain damage. R = right hemisphere; L = left hemisphere; STG = superior temporal gyrus; aSTG = anterior STG; pSTG = posterior STG; PT = planum temporale; HG = Heschl's Gyrus; MTG = middle temporal gyrus; IPC = inferior parietal cortex.

lobe). Furthermore, Peretz et al. (2009) showed near-normal pitch processing in congenital amusics (implicit processing of pitch) with EEG and suggested that the (congenital) amusic brain cannot make contact with musical pitch knowledge along the auditory-frontal neural pathway. In the amusic patients this pathway is interrupted by a damaged STG. Also, damage in the parietal lobe was reported in 5 (out of 8) cases and frontal lobe damage in 4 (out of 8) cases. But as these lesions did not occur in separation (without temporal lobe damage) their involvement in music processing cannot be demonstrated by this set of patients. Evidence from acquired and congenital amusia and pitch/melody processing studies demonstrate the crucial role of the anterior temporal lobe (i. e. the STG and STS) in music processing.

Some last comments on the here reported patients need to be made. Patients were chosen according to their reported musical deficits. In all cases a test similar the MBEA (or an earlier version of it) was used, so that musical deficits were comparable and they all could be classified as amusics along the definition of Peretz et al. (2003). For a meta-analysis with a large patient group which exhibited impairments in single musical features, the reader is referred to Stewart et al. (2006). For a large scale patient study where patients were chosen with regard to their lesion site and musical deficits, the reader is referred to Liégeois-Chauvel et al. (1998).

Implicit music processing. Of interest for the current thesis is that amusics (acquired and congenital) obviously fail in explicit musical tasks, but if tested implicitly they showed normal pitch processing abilities. This issue has already been discussed (see above in 'Congenital Amusia') and will be extended by Tillmann et al. (2007) who tested patient I. R. on implicit music processing using chord sequences and sung syllables. A harmonic priming effect was observed for the amusic patient I. R.: Phoneme (Exp. 1) or timbre (Exp. 2) identification was faster when the target was on a related tonic chord than when it was on a less related subdominant chord. Interesting for this dissertation is that with sung stimuli (sung syllables were used for the phoneme task) an implicit music processing could be seen. Experiment 3 and 4 required explicit completion judgments. I. R.'s musical disorder was confirmed as her judgments of related and less related chord sequences (Exp. 3) did not differ. That the language domain was unaffected was shown by the sentence completion judgments where I. R.'s performance was equal to the controls. While Experiment 3 confirmed I. R.'s musical disorder in an explicit task, the implicit tasks of Experiments 1 and 2 lead to the conclusion that I. R. still has some knowledge of Western tonal music that influences the musical perception at an implicit level.

Prosody and music. Recently, only congenital amusics have been tested on language compared to music processing; one patient study with mixed results will be mentioned here. Patient I. R. together with patient C. N. (who showed bilateral TL lesions, see above) assessed linguistic and musical discrimination tasks to explore prosodic and musical patterns (Patel et al., 1998); while I. R. exhibited difficulties, C. N. showed a good performance. As in both the level of performance was statistically similar, the authors suggest shared neural resources for prosody and music.

Song and Speech. Song and speech in production is also of interest, as resulting from particular lesions, the production abilities or disabilities can help making assumptions on the lateralization of song and speech. The notion that lesions in the left hemisphere lead to speech impairment with preserved singing ability, should reversely mean that lesions in the right hemisphere should lead to impaired singing and preserved speaking abilities. Among some older cases only a few have been reported with lesions in the right hemisphere that showed impaired singing abilities (Bautista & Ciampetti, 2003; Confavreux et al., 1992; McFarland & Fortin, 1982; Murayama et al., 2004; Russell & Golfinos, 2003; Terao et al., 2006). Among the more recent acquired amusic cases, the focus was on impaired music perception, not production, and therefore singing abilities were only tested in two of the patients; one with impaired singing after unilateral right hemispheric lesion (Griffiths et al., 1997) and one with bilateral anterior STG lesion (Peretz et al., 1994; Peretz, 1996). Taken together, poor singing mostly occurs in combination with music perception disabilities and lesions in the right hemisphere.

Closing this section, some considerations on acquired amusics should be taken. In amusia studies, where a group of similar lesion patients have been observed, most of them showed impairments in one or two features of music processing (many with additional aphasic and other symptoms), but only a few patients were impaired on the whole amusic spectrum, e. g., Peretz (1990) tested 20 lesion patients and presented two amusic cases (Peretz et al., 1994; presumably from the pool in 1990). Schuppert et al. (2000) discovered one amusic individual, and Ayotte et al. (2000) two amusics (both out of 20). Lesion sites and their functional impact are not replicated across patients, furthermore lesions are mostly not focal (see also Chapter 6.2).

5.3 Summary

To conclude, cases of congenital amusia serve as a unique opportunity to study cognitive processes underlying music perception. It has been shown that music processing deficits can occur regarding the pitch domain (melodic contour and interval) with and without affecting the temporal domain (rhythm and meter). Deficits have also been shown regarding music memory – e. g., singing and recognizing familiar tunes were impaired. An ongoing debate is how domain specific this disorder is. Therefore it has been of interest to separate music perception from other processes such as – most prominent – speech perception. This discourse is of special interest for the current work, mainly the question if amusics' pitch processing deficit is fully limited to the musical domain, or if it also affects prosodic processing. It has been suggested that amusics have a speech perception advantage, indicated by a better processing of gliding than discrete pitch. Only very subtle impairments in speech perception have been shown, e. g., in a word discrimination task (in a tone language), suggesting that intact speech perception is dependent on multiple acoustic cues.

Moreover it has been shown that amusics can actually process pitch changes, even smaller than a semitone, if tested implicitly, reflecting perception without awareness. This implies the necessity of a most careful approach for all amusia studies involving melody tasks, as the behavioral task itself may easily introduce response artifacts based on subjective coping strategies.

Furthermore, the combined results of different single case studies showed that a damaged anterior STG leads (except one case) to music processing deficits (acquired amusia), mainly in the melody/pitch domain. In the experimental part of the current thesis (Chapter 12), more single cases will be described to widen the congenital and acquired amusia spectrum.

Chapter 6

Methods

6.1 Functional Magnetic Resonance Imaging (fMRI)

The clinical MRI uses the nuclei of hydrogen atoms for the imaging process. Hydrogen atoms consist of a nucleus with only one proton which spins around its axis (magnetic dipole). The orientation of this axis is without a magnetic field random - but in a static magnetic field it aligns with it either parallel or anti-parallel. The spins react with an evasive movement (precession). This movement/spin has a characteristic frequency which is called Larmor frequency. It is dependent on the gyromagnetic ratio of the protons and the strength of the magnetic field (B_0) and can be expressed by the following equation: $\omega = \gamma B_0$.

The spins precess in the magnetic field and start to align longitudinally with it. This resulting stable spin system can be distorted by a radio frequency (RF) pulse which leads to an altered spin precession and a change in orientation: from the z-axis to the x-y-plane. The longitudinal alignment is turned into a transversal alignment which is reversed when the RF pulse is turned off. The tip to turn over the hydrogen atoms causes a measurable signal by inducing a current in a coil (the Free Induction Decay, FID). The return to the longitudinal magnetization is called T1 relaxation and as at the same time the precession in the x-y-plane dephases and the transversal magnetization becomes less and less, this decay is called T2 relaxation. To get an anatomical image of the human brain the T1 relaxation time is measured at different coordinates, creating an image with different brightness values.

While T1 measures the structure in the human brain, T2 is the basis of functional MR imaging. T2/T2* is additionally influenced by deoxygenated hemoglobin (among other factors), which means the intensity of blood oxygenation in a specific area can be measured - the BOLD response, the blood oxygenation level dependent response. Changes in blood supply in a certain area in the brain can therefore be measured and is associated with an increase in energy and therefore cognitive demand. This process needs time and thus, fMRI is not adequate for a high temporal resolution. The BOLD response has three typical phases: an initial dip at 1 to 2 seconds post-stimulus, an overshoot and an post-stimulus undershoot. The maximal BOLD reaction is observed in the overshoot-phase which has a time-to-peak at about 5 seconds post-stimulus, a full-width at half-maximum of about 4 seconds and a spatial distribution of 3–5 mm (Norris, 2006).

It is of note that the exact relationship between hemodynamic activity, neuronal activity and the MR signal is still not fully understood. It has been assumed that the vascular response is caused by local synaptic activity and not by neuronal sparking, which means that the BOLD signal is linked to the incoming but not (as much) to the neuronal output signal (e. g. Logothetis et al., 2001; Logothetis, 2008). Nonetheless, fMRI has been applied in a broad spectrum of research issues in cognitive neuroscience with the common sense that the BOLD response is more an indirect marker of neuronal activity. Detailed insights in fMRI methods are provided by e. g., Buxton (2002); Jezzard et al. (2001); Papanicolaou (1998).

6.1.1 Multivariate Pattern Analysis (MVPA)

fMRI data have traditionally been analyzed using univariate analyses (UVA), which compare mean activity of extended brain regions associated with different experimental conditions. Pattern classification methods however are sensitive to any changes in patterns of two conditions - including the activation mean as well as changes in patterns where mean activity does not vary. While in UVA only “the tip of the iceberg” (Formisano et al., 2008) of shared activity over all participants becomes visible, pattern classification uses fine-grained spatial activation patterns of individual participant’s data. This approach has been shown to be sensitive to small effects not detectable with conventional UVA (Formisano et al., 2008; Haynes & Rees, 2005, 2006) because information in brain regions might not be reflected in differences in the average signal. In neuroimaging the multivariate pattern analysis (MVPA) has recently gained much attention with different approaches for the analysis of fMRI activation patterns (Haynes & Rees, 2006; Kriegeskorte & Bandettini, 2007; Haxby et al., 2001; Norman et al., 2006).

The approach used in the current study is summarized in the following (see also the fMRI-Methods section, Chapter 11.1). After preprocessing steps, the feature selection takes place, which means defining a set of voxels that constitutes the patterns. Activation measures are then typically used as feature values (build a feature vector) for the classification analysis. It is possible to use activation patterns from all voxels for classification, or a subset of those could be used as features. In the current study a searchlight (Kriegeskorte et al., 2006) approach was chosen which constructs a sphere with a defined radius around each voxel. From each of these 3-dimensional clusters the classification accuracy can be achieved and be used for further analyses. In the current work, the data set was divided in a training set and a test set. The pattern vectors of the training data set were used to train a classifier. A classifier learns to distinguish between activation patterns relating to different experimental conditions. In the current case a linear support vector machine (SVM; C.-C. Chang & Lin, 2011) classifier was used, which estimates a decision hyperplane (a decision boundary) to separate the patterns of experimental conditions. This trained model is then used on an independent data set to classify the conditions.

Via cross-validation a better estimation of the classification accuracy can be obtained and over-fitting is controlled for. By dividing the data into independent sub-sets each one is used for testing the decision boundary on generalization. The decoding accuracy is then estimated by the percentage of correct classifications from all cross-validation steps (Haynes & Rees, 2006). The averaged decoding accuracy across all cross-validation steps is statistically tested against chance level. The exact parameters for the multivariate pattern analysis in the current thesis are provided in Chapter 11.1.

6.2 The Lesion Method

Leborgne and Lelong became famous postmortem through to their severe symptoms of aphasia - both were patients of Pierre Paul Broca, who found in both the same area to be damaged: Broca's area (Broca, 1861). From the 1860's on the essence of the lesion method is the establishment of a correlation between cognitive function and brain anatomy by investigating changes in behavioral performance after brain damage (Damasio & Damasio, 1989). This research tradition was challenged by imaging techniques as fMRI and PET, which are non-invasive imaging methods, measuring brain activations in healthy participants. Together with novel imaging techniques, Damasio and Damasio (1989) argue that lesion studies are still a unique opportunity to study the functions of the human brain. The advantages and limitations of the lesion method are laid out in the following paragraphs.

Firstly, naturally occurring brain lesions are rarely restricted to one structure; they are dictated by pathological processes, i. e. vascular anatomy. Secondly, a damaged area is also rarely restricted to one cognitive module; that means that co-existing deficits need to be taken into account when assigning and interpreting particular functions. Thirdly, reorganization processes change the pattern of deficits over time. This does not mean that acute lesions are the best to study; acute data are mostly confounded by a temporally disconnection between intact areas (diaschisis). When chronically lesions however lead to a consistent deficiency in behavior, this might be taken as evidence for the involvement of that area in a certain process. Furthermore, reorganization processes can also explain an intact behavior regardless a lesion; the deficiency of one area may have been compensated by another. Fourthly, the lesion method is based on the locality assumption, which ignores that the majority of functions is processed in distributed neural networks (Farah, 1994). Lesions occur within anatomical and functional networks: Anatomically, a lesion encompasses also white matter and disrupts the fiber tracts between communicating areas; functionally, due to a lesion a processing step cannot be accomplished and disrupts the functional network to fulfill a task.

Taken all these limitations together, in a most careful selection process of the patients some of these challenges can be faced: A lesion site description with MRI is necessary and eventually DTI (diffusion tensor imaging; to detect fiber tract anomalies) and a careful behavioral testing to detect functional deficits. Also, group studies rather than single case studies are able to create a more reliable picture of correlations between areas and their function by testing patients with similar as well as different damaged areas.

Furthermore, most of the critical points mentioned above can be overcome with imaging techniques applied in healthy participants. The spatial (fMRI and PET, MEG) and temporal (EEG, MEG) resolution is higher than in lesions studies. But, imaging techniques measure activation instead of disruption, which is not an indicator whether a certain structure is actually necessary to do the task. An activation of a certain area might not be due to its direct role in information processing, but depends on the control condition or might be co-activated due to its neural connection with eloquent brain areas, e. g. the homotope area in the contralateral hemisphere.

To conclude, both approaches, the lesion method and imaging techniques, build a useful combination of methods to investigate the correlation between brain areas and their function. Taking care of the choice of patients by precisely imaging the extent of the lesion and by a careful behavioral testing to examine co-existing functional deficits, the lesion method is especially in the context of fMRI investigations a valid approach in neuroscientific research.

Part II

Experiments

Chapter 7

Research Questions

The preceding overview has given some indications where further research is required: on the neural correlates of song and speech perception rather than production with focus on linguistic and melodic processing. Furthermore, the specific role of the temporal lobe needs to be addressed, and the influence of music processing deficits will provide further insights into song and speech perception.

Why song and speech research? In general, it helps to understand how sound is represented in the brain. It can be utilized to see how the brain treats periodicity, timbre, pitch characteristics, contour and interval (discrete or glide), melodic and rhythmic grouping, structure building processes, vocalization and even meaning and emotions. In language vs. music approaches, usually spoken sentences and instrumental music are used, which are acoustically two different stimuli – using song and speech stimuli provides vocal stimuli which are acoustically and spectrally in close correspondence (e. g., no timbre differences).

Song and speech are a conglomerate of spectral, temporal and linguistic features, and production and perception studies have tried to discriminate both on a neural basis. Only two studies investigated the perception of sung and spoken material: Schön et al. (2010); Callan et al. (2006); the latter in combination with a production task. Both studies failed to show a speech over song contrast in perception, and only one study tried to investigate song and speech separated by their underlying constituents words and pitch patterns (influenced by differential task demands; Schön et al., 2010). This is of special interest as these are the main components which song and speech share but also differ in.

What is missing? A perception study with low task demands of song and speech that investigates the underlying parameters of song and speech and their neural correlates. To disentangle these acoustic features and the neural underpinnings of song and speech, a different paradigm needs to be created which takes into account the linguistic, melodic and rhythmic differences between the two.

A general pattern in the more recent studies seems to be a bilateral activation of temporal areas for both sung and spoken stimuli, with a right hemisphere advantage for musical pitch patterns, melody of song, and prosodic aspects of speech, and a left hemisphere advantage for rhythm, segmental, and linguistic aspects of both speech and song. This notion was approached with Experiment 1.

In particular, the involvement of the temporal lobe is not yet clear – overlap of song and speech was suggested with the aSTG being more strongly involved in song than in speech (Callan et al., 2006). Studies on pitch and melodic processing for example proposed a hierarchical processing in the STG with more anterior parts of this region involved in pitch complexity processing (Patterson et al., 2002; Griffiths et al., 1998). This would suggest that song and speech are treated differently in the (anterior) temporal lobes as song might be more dependent on pitch than speech.

The influence on pitch and music processing on song and speech perception can be investigated by testing amusic individuals, who show music processing deficits. To resolve the activation pattern in the STG and the temporal lobe in general, patients with focal lesions in the temporal lobe can be behaviorally tested on their song and speech perception. These influences were approached in Experiment 2.

In summary, the current dissertation provides:

I. a study of the neural correlates of song and speech with focus on perception
and low task demands (Exp. 1)

II. sung and spoken stimuli in a hierarchical paradigm which provides
   a. separated investigations of words and pitch patterns
   (controlled for rhythm; Exp. 1B, Part 2)
   b. currently missing speech over song contrasts (Exp. 1B, Part 2)
   c. and the comparison of sung and spoken pitch patterns (Exp. 1B, Part 1)

III. an approach to investigate the involvement of the temporal lobe in song and speech
processing
   a. in an fMRI study with healthy participants (Exp. 1B)
   b. in a behavioral testing on patients with lesions in the temporal lobe (Exp. 2)

IV. insights on how music processing deficits alter song and speech perception (Exp. 2).

Chapter 8

Experiment 1 – fMRI Study – Perception of Song and Speech

The following empirical part consists of two main studies, of which one has already been published and appears unchanged in this chapter (Introduction), Chapter 9 (Materials and Methods) and Chapter 10 (Data Analysis, Results, Discussion and Conclusion): Merrill, J., Sammler, D., Bangert, M., Goldhahn, D., Lohmann, G., Turner, R. and Friederici, A. D. (2012): Perception of words and pitch patterns in song and speech. *Frontiers in Psychology, 3*(76), 1-13.

A behavioral pretest for the published fMRI study was included in this dissertation in Chapter 9 (Experiment 1A). Additional analyses of the fMRI data for the current thesis are described in Chapter 10 (Experiment 1B Part 1). These analyses catch up with previous studies (e. g., Callan et al., 2006; Schön et al., 2010) and report the 'direct comparisons', i. e. song and speech in comparison, regardless of the different underlying parameters, before in Chapter 11 (Experiment 1B Part 2) a new approach on song and speech data with regard to the perception of words and pitch patterns in both is described.

8.1 Introduction[1]

Nobody would ever confuse a dialog and an aria in an opera such as Mozart's 'The Magic Flute', just as everybody would be able to tell whether the lyrics of the national anthem were spoken or sung. What makes the difference between song and speech, and how do our brains code for it?

Song and speech are multi-faceted stimuli which are similar and at the same time different in many features. For example, both sung and spoken utterances express meaning

---

[1] This part corresponds to Merrill et al. (2012).

through words and thus share the phonology, phonotactics, syntax and semantics of the communicated language (Brown et al., 2006). However, words in sung and spoken language exhibit important differences in fine-grained acoustic aspects: Articulation of the same words is often more precise and vowel duration considerably longer in sung compared to spoken language (Seidner & Wendler, 1978). Furthermore, the formant structure of the vowels is often modified by singing style and technique, as for example reflected in a Singer's Formant in professional singing (Sundberg, 1970).

Both song and speech have an underlying melody or pitch pattern, but these vary in some detail. Song melody depends on the rule-based (syntactic) arrangement of 11 discrete pitches per octave into scales as described by music theory (cf. Lerdahl & Jackendoff, 1983). The melody underlying a spoken utterance is called prosody and may indicate a speaker's emotional state (emotional prosody), determine the category of sentences such as question or statement and aid language comprehension in terms of accentuation and boundary marking (linguistic prosody). In contrast to a sung melody, a natural spoken utterance carries a pattern of gliding, not discrete, pitches that are not related to scales but vary rather continuously (for an overview see Patel et al., 2008).

Altogether, song and speech, although similar in many aspects, differ in a number of acoustic parameters that our brains may capture and analyze to determine whether a stimulus is sung or spoken. The present study sets out to explore the neurocognitive architecture underlying the perception of song and speech at the level of their underlying constituents – words and pitch patterns.

Previous fMRI studies on the neural correlates of singing and speaking focused predominantly on differences between song and speech production (overt, covert or imagined; Gunji et al., 2007; Jeffries et al., 2003; Özdemir et al., 2006; Riecker et al., 2000; Wildgruber et al., 1996) or compared production with perception (Callan et al., 2006; Saito et al., 2006) whereas pure perception studies are rare (Sammler, Baird, et al., 2010; Schön et al., 2010). Two main experimental approaches have been used in this field: either syllable singing of folksongs or known instrumental music was contrasted with the recitation of highly automated word strings (e. g., names of the months; Riecker et al., 2000; Wildgruber et al., 1996), or well-known sung folksongs were contrasted with the spoken lyrics of the same song (Callan et al., 2006; Gunji et al., 2007; Jeffries et al., 2003; Saito et al., 2006).

Despite their above mentioned methodological diversity, most of the production as well as perception studies report a general lateralization effect for speech to the left and for song to the right hemisphere. For example, Callan et al. (2006) compared listening to sung (SNG) and spoken (SPK) versions of well-known Japanese songs and found significantly stronger activation of the right anterior superior temporal gyrus (STG) for SNG and a more strongly left-lateralized activity pattern for SPK. These findings led the authors to suggest that the right or left lateralization could act as a neural determiner for melody or speech processing, respectively. Schön et al. (2010) extended this view by suggesting that within song, linguistic (i. e. words) and musical (i. e. pitch) parameters show a differential hemispheric specialization. Their participants listened to pairs of spoken words, sung words, and 'vocalise' (i. e. singing on a syllable) while performing a same/different task. Brain activation patterns related to the processing of musical aspects of song isolated by contrasting the sung versus spoken words showed more extended activations in the right temporal lobe, whereas the processing of linguistic aspects (such as phonology, syntax, and semantics) determined by contrasting song versus vocalise showed a predominance in the left temporal lobe.

Thus, both production and perception data seem to suggest a predominant role of the right hemisphere in the processing of song due to pronounced musical features of the stimulus and a stronger left hemisphere involvement in speech due to focused linguistic processing. Notably, the most recent studies (Callan et al., 2006; Schön et al., 2010) allude to the possibility that different aspects of spoken and sung language lead to different lateralization patterns, calling for an experiment that carefully dissects these aspects in order to draw a conclusive picture on the neural distinction of song and speech perception.

Due to a restricted number and the particular choice of experimental conditions, previous studies (Callan et al., 2006; Schön et al., 2010) did not allow for fully separating out the influence of words, pitch patterns or other (uncontrolled) acoustic parameters on the differential coding for sung and spoken language in the brain.

Particularly, when it comes to the comparison of pitch patterns between song and speech, it must be taken into account that the melodies in song and speech (most obvious when they are produced on sentence level) do not only differ in their pitch contour, but have also different underlying rhythm patterns. Rhythm differences in song and speech concern mainly the periodicity, i. e. the metric conception. Meter describes the grouping of beats and their accentuation. Temporal periodicity in musical meter is much stricter than in speech and the regular periodicities of music allow meter to serve as a mental framework for sound perception. As pointed out by Patel et al. (2008, p. 194) "there is no evidence that speech has a regular beat, or has meter in the sense of multiple periodicities". Brown and Weishaar (2010) described the differences in terms of a 'metric conception' for song as opposed to a 'heterometric conception' for speech.

Consequently, the influence of the differential rhythm patterns must be parceled out (for example by adding a respective control condition) in order to draw firm conclusions on melody and prosody processing – which has not been done so far. This is also of specific relevance because the left and right hemispheres are known to have a relative preference for temporal (rhythm) and spectral (pitch) information, respectively (Jamison et al., 2006; Obleser et al., 2008; Zatorre & Belin, 2001).

Furthermore, the methodological approaches of the reported fMRI studies were limited to univariate analyses (UVA), which mostly subtract two conditions and provide information about which extended brain regions have a greater mean magnitude of activation for one stimulus relative to another. This activation based method relies on the assumption that a functional region extends over a number of voxels and usually applies spatial smoothing to increase statistical power.

Recent methodological developments in neuroimaging have brought up multivariate pattern analysis (MVPA; Haxby et al., 2001; Norman et al., 2006) which does not only take into account activation differences in single voxels, but analyses the information present in multiple voxels. In addition to regions that react more strongly to one condition than another, as in UVA, MVPA can thus also identify brain areas in which a fine spatial pattern of activation of several voxels discriminates between experimental conditions (Kriegeskorte et al., 2006). Notably, this allows identifying the differential involvement of the same brain area in two conditions that would be cancelled out in conventional univariate subtraction methods (Okada et al., 2010).

UVA and MVPA approaches complement each other in that weak extended activation differences will be boosted by the spatial smoothing employed by the UVA, whereas the MVPA will highlight non-directional differential activation patterns between two conditions. Consequently, the combination of the two methods should define neural networks in a more complete way than each of these methods alone. Note that a considerable overlap of the UVA and MVPA results is not unusual given that the similarity or difference of activation patterns is partly also determined by their spatial average activity level (for studies that explicitly isolate and compare multivariate and univariate contributions to functional brain mapping see Abrams et al., 2011; Kriegeskorte et al., 2006; Okada et al., 2010).

The present study used UVA as well as MVPA in a hierarchical paradigm to isolate the neural correlates of the word- and pitch-based discrimination between song and speech, corrected for the rhythmic differences mentioned above. Song and speech stimuli were constructed such to contain first all the three features (words, pitch and rhythm) of a full sung and spoken sentence, second only the pitch and rhythm patterns, and third, as a control for pitch processing, only the rhythm (see Figure 8.1). To assure maximal comparability, these three levels were derived from one another, spoken and sung material was kept parallel, task demands were kept as minimal as possible, and the study focused purely on perception. The hierarchical structure of the paradigm allowed us to (i) subtract each level from the above one to obtain brain areas only involved in word (first minus second level) and pitch (second minus third level) processing in either song and speech and (ii) compare these activation patterns.

We hypothesized first that words (or text and lyrics) in both song and speech may recruit left frontal and temporal regions where lexical semantics and syntax are processed (for a review see Bookheimer, 2002; Friederici, 2002, 2011). Second, the neural activation of prosody in speech and melody in song may be driven by its acoustic, pitch-related properties that are known to evoke a relative predominance of right-hemispheric involvement (Jamison et al., 2006; Obleser et al., 2008; Zatorre & Belin, 2001). Furthermore, we expected differences with respect to gliding and discrete pitches to be reflected in particular brain signatures.

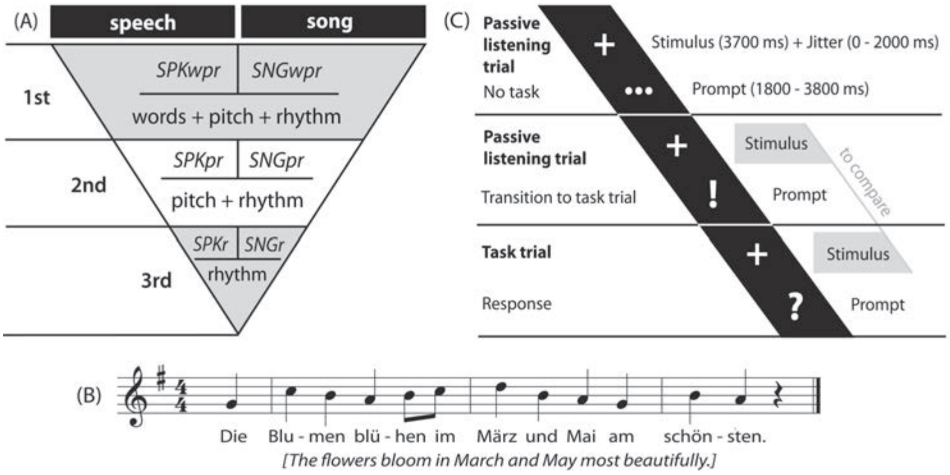

Figure 8.1: (A) Experimental design. Six conditions in a subtractive hierarchy on three levels: first level: SPKwpr and SNGwpr (containing words, pitch pattern and rhythm), second level: SPKpr and SNGpr (containing pitch pattern and rhythm), third level: SPKr and SNGr (containing rhythm). (B) Stimulus example. (C) Timeline of passive listening trial and task trial.

Chapter 9

Experiment 1A – Behavioral Study

The behavioral study served mainly as a pretest for the subsequent fMRI study. The goal was to (1) test if the task was feasible, (2) test the task with background scanner noise and (3) conduct a valence and arousal testing of the stimuli. Furthermore, the results of the pretesting provided a first glimpse at the participants' ability to correctly associate pitch and rhythm patterns to the songs and spoken sentences that they were derived from, and additionally whether or not participants' listening habits and experiences had an influence on their rating accuracy.

To do so, hierarchically organized sung and spoken stimuli were presented, carrying less and less information with each level; one feature was cancelled out from each respective level: from a full sentence to the pitch-rhythm pattern (minus text) to the pure rhythm (minus pitch contour; for details see below and Figure 8.1). The task was to indicate whether two stimuli had the same source, or in other words, if they derived from the same sentence or song respectively (see Table 9.1). Rating responses (accuracy) and response time were recorded and analyzed. A hypothesis for the rating accuracy and the response time was that if a stimulus contains more information it is easier to remember and therefore the rating accuracy should be higher and the response time shorter. Moving down the stimulus hierarchy this should mainly affect the pairs with a rhythm component as this is the most abstract stimulus.

9.1 Methods

Participants. Forty-two healthy German native speakers (18 male, mean age 22.8 years, SD: 2.8 years) participated in the study. None of the participants were professional musicians, nor had learned to play a musical instrument for more than four years.

	1st	2nd
SNG match	SNGpr SNGr	SNGwpr
	SNGwpr	SNGpr SNGr
SPK match	SPKpr SPKr	SPKwpr
	SPKwpr	SPKpr SPKr
SNG non match	SNGpr SNGr	SPKwpr
	SPKwpr	SNGpr SNGr
SPK non match	SPKpr SPKr	SNGwpr
	SNGwpr	SPKpr SPKr

Table 9.1: Possible stimulus pairs within one melody/sentence. SNG = song; SPK = speech;, wpr = stimulus contains words, pitch, rhythm; pr = stimulus contains pitch, rhythm; r = stimulus contains rhythm.

Materials. The paradigm consisted of 6 conditions, arranged in a subtractive hierarchy: spoken (SPKwpr) and sung sentences (SNGwpr) containing words, pitch patterns and rhythm; hummed speech prosody (SPKpr) and song melody (SNGpr) containing only pitch patterns and rhythm, as well as the speech or musical rhythm (SPKr and SNGr; see Figure 8.1). In this manner, 6 different sets of stimuli were created.

The sentences for the '-wpr' stimuli were 6 different statements with a constant number of twelve syllables across all conditions. The 6 melodies for the sung (SNG) stimuli were composed according to the rules of western tonal music, in related major and minor keys, duple and triple meters, and with and without upbeat depending on the sentences. The lyric/tone relation was mostly syllabic. For the rhythm (-r) conditions, a hummed tone (G3) was recorded and cut to 170 ms with a 20 ms fade in and out. Sequences of hummed tones were created by placing the tone onset on the vowel onsets of each syllable according to the original sung and spoken material using Adobe Audition 3 (Adobe Systems; for more details see Chapter 10).

Procedure. Participants were required to indicate via button press whether a pair of two stimuli was derived from the same original sentence or song. One of the stimuli (e. g., SNGwpr) was always taken from a different hierarchical level than the other (e. g., SNGr).

A melody (second level) or rhythm (third level) stimulus could only build a pair with a full sentence or song (first level). No combinations of -pr and -r stimuli were possible. Also, stimuli swapped positions (e. g., SNGwpr/SNGpr and SNGpr/SNGwpr) and a combination did not match if the stimuli did not derive from the same origin (e. g., SPKwpr/SNGpr; see Table 9.1). As stated, 6 different sets of stimuli were created. Of significant importance for the task was that pairs were only built out of the stimuli from one set. For example, no sung melody from song #1 was allowed to follow song #2 and the spoken rhythm of set #3 could not follow sung melody #1 and so on. Therefore, it was not a test for e. g., song memory.

Within each of the six stimulus categories, sixteen stimulus pairs were possible, which lead to 96 stimulus pairs in total. Each participant perceived all stimuli in one of 42 pseudo-random orders to ensure the task rules and avoid identical stimuli following each other.

Participants sat in front of a monitor that prompted ‘y’ and ‘n’ which stood for the buttons on the ‘QWERTZ’-keyboard they pressed to give their ‘yes’ or ‘no’ answer for matching the stimulus pairs. As soon as they had made their decision they could press a button; it was not necessary to wait until the stimulus was fully played. The next stimulus only started after a button was pressed. There was a one-second delay between the two stimuli of a pair. Prior to the experiment participants were instructed via audio introduction, which played examples and gave the answer after four seconds, regardless if the pairs matched or not. After that, participants had to rate 10 samples (taken from the original stimulus set). Those test ratings were evaluated immediately to assure the participants' understanding of the task. None of them showed an obvious misunderstanding. Participants had not been informed about the type of stimuli, for example song or speech, melody or rhythm; sounds were only called ‘Klangbeispiele’ (= sound examples).

Data analysis. A within-subject analysis of variance (ANOVA) with factors match (matching vs. non-matching stimulus pairs), level (-pr/-wpr vs. -r/-wpr stimulus pairs) and order (1st vs. 2nd position) and paired-sample t-tests were performed on all stimulus pairs using SPSS Statistics 17.0 to analyze differences in rating accuracy (RA) and response time (RT). A further ANOVA with factors Domain (SNG vs. SPK), Level (-pr/-wpr vs. -r/-wpr) and Order (1st vs. 2nd position) and paired-sample t-tests were performed only on the matching pairs to evaluate differences between song and speech stimuli.

Response bias. Taking a response bias into account, a d-prime analysis was conducted, according to the signal detection theory. The sensitivity index d’ (‘d-prime’) represents the sensitivity of the subjects in distinguishing between the two given stimuli. The criterion c represents the tendency of subjects to either answer ‘yes’ or ‘no’ more frequently than the other option, respectively. The calculation of d’ and c depends on the Hit, Miss, False

Alarm (FA) and Correct Rejection (CR) rate. Intuitively, the best participant maximizes the Hit rate (and thus minimizes the Miss rate) and minimizes the False Alarm rate (and thus maximizes the Correct Rejection rate); thus the larger the difference between Hits and FA, the better the subject's sensitivity. The statistic d' is a measure of this difference; it is the distance between the z-transforms of Hits and False Alarms: $d' = z(H) - z(FA)$. If Hits and FA are equal, $d' = 0$, the participant made a random decision. The effective limit (with a Hit rate of 99 % and a False Alarm rate of 1 %) of the highest possible d' (greatest sensitivity) is 4.65. Typical values are up to 2.0, and 69 % correct for both 'yes' and 'no' trials corresponds to a d' of 1.0 (Macmillan & Creelman, 2009).

A response bias is measured as the inclination of the participant to say 'yes' (or 'no'). The bias measure c is a function of Hits and FA: $c = -0{,}5 * (z(H) - z(F))$. The value $c < 0$ indicates a tendency toward the response 'yes' and $c > 0$ toward 'no'.

Task with scanner noise. For the fMRI experiment (Chapter 10) it needed to be tested, (1) whether the task was feasible with the scanner noise and (2) if participants reacted emotionally to the stimuli and (3) if they perceived them as more negative or positive. To ensure that the task could be properly performed in the scanner as well as outside the scanner without scanner noise, a smaller sample set was presented with a recording of the scanner noise in the background. Each stimulus pair was presented 13–16 times across participants.

Valence and arousal rating. Each stimulus of the first level (sung and spoken stimuli with text) was tested for valence and arousal. A low level of arousal (2–3) and a neutral to positive valence (5) were desirable for the fMRI experiment. This was tested without scanner noise.

Self assessment. After the rating, participants were asked to judge on a 9 point scale (1 easy, 9 difficult) how difficult they had found the pair-matching task. The goal was to find out if their response behavior would reflect their assessment. Additionally, participants were asked to describe the heard stimuli and to name them. As participants were not informed beforehand about the hierarchical structure of the stimuli, these questions were targeted at finding out how participants associated the connection between sung and spoken sentences and their underlying constituents, melody and rhythm.

Listening habits. Furthermore, an additional questionnaire was conducted with the goal of looking for correlations between the rating results and participants' individual listening habits and experiences. It asked for hours of music listening per week, hours of listening to audio books per month and concert and theater visits per year. The hypothesis was that maybe participants who had an advantage in rating song stimulus pairs would be those who went to concerts or listened to music more often. An advantage for rating spoken stimulus pairs would speak to a greater propensity to listening to audio books or a larger interest in going to the theater.

9.2 Results

Task with scanner noise. The goal of this pretest was to prove that the task could be performed equally with and without scanner noise. There was no significant difference in the performance in both presentation forms ($p > 0.6$), which means no distraction due to the noise and no frequency overlap between the stimuli and the noise could be observed.

Valence and arousal rating. Participants did not react emotionally to the stimuli and rated most of them as neutral. As the text and the melodies had a folk song character, two of them were rated more positively (see Table 9.2) than the others. A more positive character of a stimulus was not a criterion for excluding it, as long as participants did not have to reveal any kind of emotional expression during the scanning session.

	wpr 1	wpr 2	wpr 3	wpr 4	wpr 5	wpr 6
arousal	2.72	2.74	2.91	2.70	3.40	2.81
valence	4.81	5.30	4.79	5.93	7.16	5.72

Table 9.2: Valence and arousal rating of full stimuli (-wpr). Means of a 9 point scale rating (1 low, 9 high). Stimuli were rated less arousing and emotionally more neutral to positive.

RA and RT for matching and non-matching stimulus pairs. The within-subject ANOVA on all stimulus pairs for the rating accuracy (RA) revealed significant main effects for the factors Match and Level, indicating a better performance for matching vs. non-matching pairs (see also ‘Response bias’) and -pr vs. -r pairs. Significant interactions were found for Level × Order and Match × Level × Order (see Figure 9.1, Table 9.3 for details).

The within-subject ANOVA for the response time (RT) revealed significant main effects for Level and Order, indicating faster responses for -pr than -r pairs, and -pr or -r stimuli in the second over first place. A significant interaction for Level × Order was found (see Table 9.3 for details).

In detail, the main effect for Level showed that the accuracy was higher and RTs were shorter for -pr pairs than for -r pairs (see Figure 9.2). The response was about 1100 ms faster. That means that in the melodic comparison (mean 3410 ms) the decision was made while

ANOVA		**contrasts**	**RA**		**RT**	
(**all stimulus pairs**)			F(1,41)	p-value	F(1,41)	p-value
	Match	match > non-match	48.53	< .000		
ME	Level	pr > r	164.216	< .000	149.55	< .000
	Order	2nd > 1st			27.034	< .000
IA	Level × Order		8.173	< .007	15.55	< .000
	Match × Level × Order		5.482	< .024		
***t*-tests**	**for IA**		t(41)	p-value	t(41)	p-value
	Level × Order	pr 1st = pr 2nd				
		r 2nd > r 1st	-1.956	> .057	-5.588	< .000
		pr 1st > r 1st	13.182	< .000	-8.307	< .000
		pr 2nd > r 2nd	9.586	< .000	-12.381	< .000

Table 9.3: Results for rating accuracy (RA) and response time (RT) of the ANOVA for all stimulus pairs and additional two-sample t-tests. ME = main effect; IA = interaction. pr = pitch-rhythm (melody) comparison; r = rhythm comparison. SNG = song, SPK = speech; 1st = stimulus occurred in the first position of a pair, 2nd = stimulus occurred in the second position of a pair.

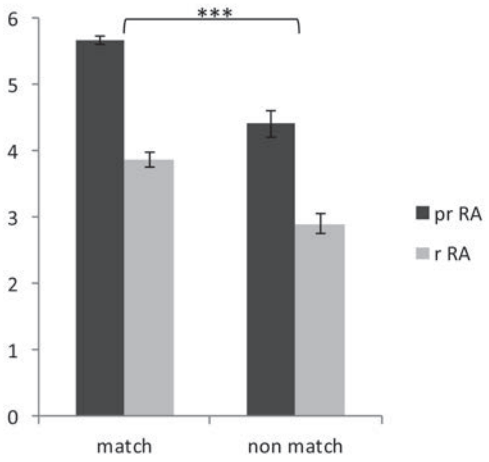

Figure 9.1: Results for the rating accuracy (RA; number of correct rated pairs out of 6) for matching and non-matching pairs.

the stimulus was still running, but in the rhythm comparison (mean 4520 ms) participants pressed the button 82 ms after stimulus offset. The melodic comparison decision was made faster than the rhythm comparison. This is in close correspondence to the result that the melodic comparison was easier to make, reflected in a higher accuracy.

A main effect for Order revealed significant results for RT and reflected a difference whether a -pr/-r stimulus occurred in the first or second position. Participants were faster with their response when the -pr or -r stimulus occurred in the second position, i. e. after a -wpr stimulus (see Figure 9.3).

To interpret the Level × Order interaction, additional t-tests were conducted. -r stimuli were rated more accurately and 55.4 ms faster when they were presented after a -wpr stimulus. Furthermore, the t-tests underline the result that -pr stimuli were rated more accurately and faster than -r stimuli in both positions (see Figure 9.3).

The Match × Level × Order interaction is most likely driven by a low RA and RT for rhythm stimulus pairs. Participants' RA and RT were best in SNGpr stimuli occurring in the second position.

RA and RT for matching stimulus pairs. The within-subject ANOVA on matching stimulus pairs only for the rating accuracy (RA) revealed significant main effects for the factors Domain and Level, indicating a higher RA and a lower RT for sung than for spoken stimuli, and a significant interaction for Domain × Level, which is most likely driven by the significantly low performance of the SPKr stimuli (see Figure 9.4 and Table 9.4 for details).

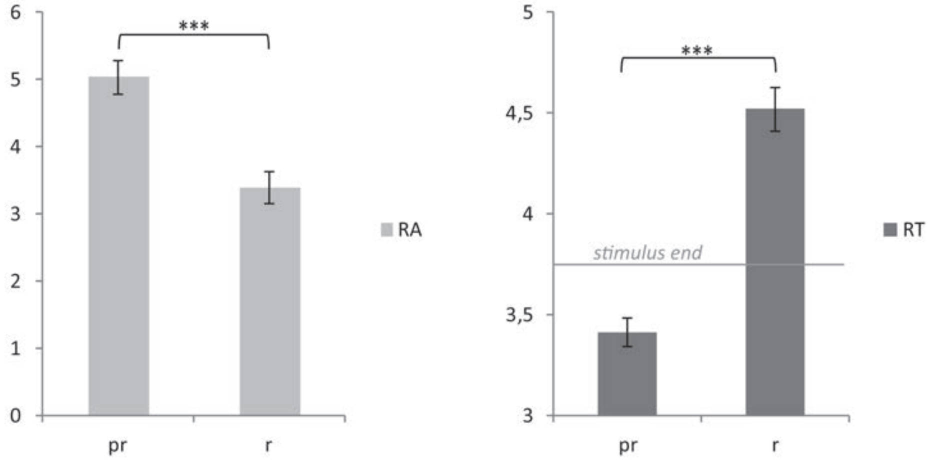

Figure 9.2: Results for the rating accuracy (RA; number of correct rated pairs out of 6) and response time (RT; in seconds) of -pr an -r pairs.

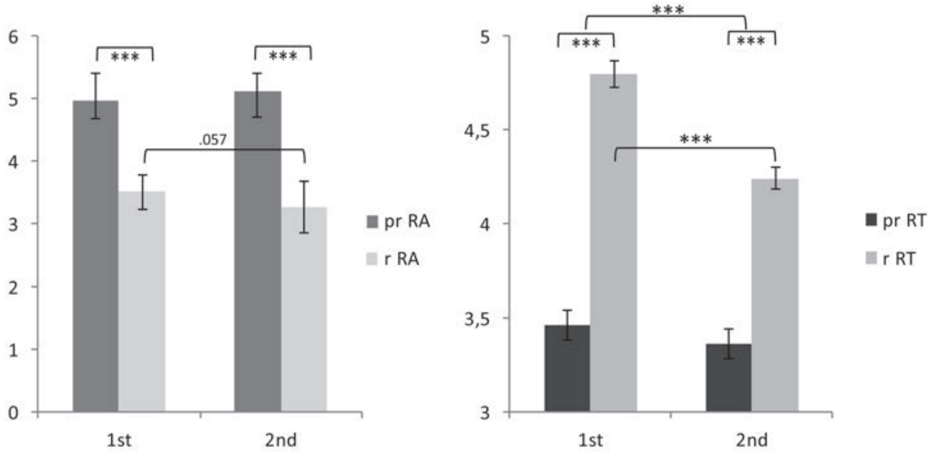

Figure 9.3: Results for the rating accuracy (RA; number of correct rated pairs out of 6) and response time (RT; in seconds) for stimuli occurred in the 1st or 2nd position.

ANOVA (matching stimulus pairs)		**contrasts**	**RA**		**RT**	
			$F(1,41)$	p-value	$F(1,41)$	p-value
ME	Domain	SNG > SPK	16.428	< .000	18.356	< .000
	Level	pr > r	261.797	< .000	139.544	< .000
	Order	2nd > 1st			15.181	< .000
IA	Domain × Level		5.778	< .021	6.213	< .017
	Level × Order				10.636	< .002

***t*-tests**	**for IA**		$t(41)$	p-value	$t(41)$	p-value
	Domain × Level	SNGpr > SPKpr	3.059	< .004	-4.901	< .000
		SNGr > SPKr	4.233	< .000		
		SNGpr > SNGr	8.321	< .000	-10.303	< .000
		SPKpr > SPKr	11.784	< .000	-8.988	< .000

Table 9.4: Results for rating accuracy (RA) and response time (RT) of the ANOVA for matching stimulus pairs only and additional two-sample t-tests. ME = main effect; IA = interaction. pr = pitch-rhythm (melody) comparison; r = rhythm comparison. SNG = song, SPK = speech; 1st = stimulus occurred in the first position of a pair, 2nd = stimulus occurred in the second position of a pair.

The within-subject ANOVA for the response time (RT) revealed significant main effects for Domain, Level and Order and a significant interaction for Level × Order (see Figure 9.4 and Table 9.4 for details). For the explanation of the remaining main effects and interactions, see RA and RT for matching and non-matching pairs, which showed similar results.

Correlation of RA and RT. There was a significant correlation (Spearman's rho, $r = -,697$, $p < 0.01$) between the rating response behavior and the response time; the less correct the answers were, the more time participants needed to answer and the more correct the answers were, the less time they needed.

Response bias. Participants showed a response bias toward 'yes', shown with the bias measure c, a function of Hits and False Alarms. Positive c-values represent answers toward 'no', negative c-values toward 'yes'. Only two participants had a c slightly above 0 (c = 0.03; 0.19), nine participants under 0 (c > -0.4), all the others were c < -0.4 (min. c = -1.09), which represents a strong response bias toward 'yes'.

Sensitivity and chance level. While participants scored better than predicted by chance in all -pr pairs, in the SNGwpr/SNGr, SNGr/SNGwpr and SPKwpr/SPKr pairs (one-sample t-test (0.5): $t(41) > 3.206$, $p < .003$; for details, see Table 9.6), they scored at chance for all

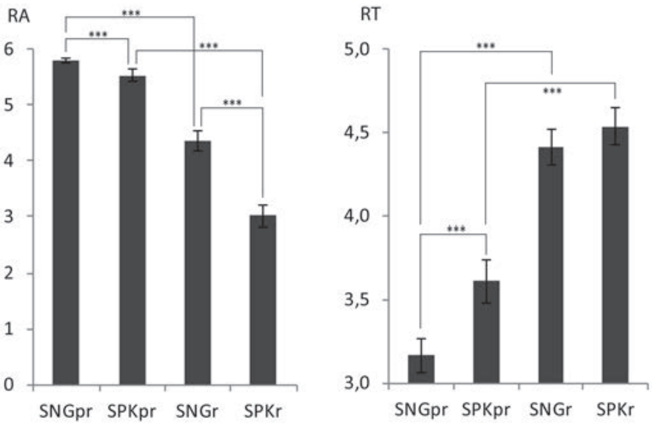

Figure 9.4: Results for the rating accuracy (RA; number of correct rated pairs out of 6) and response time (RT; in seconds) for only matching stimulus pairs.

the incorrect rhythm pairs and the SPKr/SPKwpr ($p > .110$). Participants' sensitivity for the -pr pairs was much better (d' > 2.01) than for the -r pairs (d' < 0.56), which indicates a rating accuracy under 69 % for the rhythm pairs. Only SPKr in the first position had a sensitivity of d' = 0 which is equal to a random decision (see Figure 9.5 and Table 9.5).

Stimulus Pair	Hit	FA	z(H)	z(F)	d'
SNGpr 1st	244	57	1.86	-0.75	2.61
SNGpr 2nd	243	68	1.80	-0.61	2.42
SPKpr 1st	192	24	0.71	-1.31	2.02
SPKpr 2nd	170	15	0.45	-1.56	2.01
SNGr 1st	186	136	0.64	0.1	0.54
SNGr 2nd	180	126	0.57	-1.4E-16	0.57
SPKr 1st	108	113	-0.18	-0.13	-0.06
SPKr 2nd	138	106	0.12	-0.20	0.32

Table 9.5: d-prime analysis for the stimulus pairs. Hits (max. 252) and False Alarms and their z-transformation. d' = 2 indicates a very good sensitivity, d' = 1 indicates a good sensitivity and is equal to 69 % rating accuracy, d' = 0 indicates a random decision.

Over all, participants had a good sensitivity (mean d' = 1.17), which indicates a rating accuracy higher than 69 %. 11 participants had a sensitivity lower than d' = 1 and two participants had a sensitivity lower than d' = 0, which indicates random decisions (see Figure 9.6), which was mainly driven by -r stimulus pairs.

		SNG				**SPK**				
		match		non-match		match		non-match		
		pr/r 1st	pr/r 2nd	pr/r 1st	pr/r 2nd	pr/r 1st	pr/r 2nd	pr/r 1st	pr/r 2nd	**mean**
	RT(sec)	3.09	3.24	3.43	3.41	3.66	3.57	3.26	3.66	3.41
pr	RA (%)	96.83	96.43	76.19	67.46	90.48	94.05	77.38	73.02	83.98
	Chance (*p*-value)	< .000	< .000	< .000	< .003	< .000	< .000	< .000	< .000	
	RT(sec)	4.18	4.65	4.25	4.95	4.28	4.79	4.25	4.79	4.52
r	RA (%)	73.81	71.43	42.86	54.76	55.16	57.94	46.03	50	56.5
	Chance (*p*-value)	< .000	< .000	**> .110**	**> .262**	**> .181**	**> .065**	**> .359**	**1.000**	

Table 9.6: Means of response time (RT in seconds), rating accuracy (RA in %) and chance level for -pr (melody) and -r (rhythm) combinations over all participants. The left columns show all possible SNGwpr combinations, the right columns show all possible SPKwpr combinations. The combinations are divided in matching and non-matching pairs and the position in which the -pr or -r stimulus in combination with the -wpr occurred: in the first (before a wpr stimulus) or second position (after a -wpr stimulus). p-values for combinations at chance are bold. SNG = song, SPK = speech.

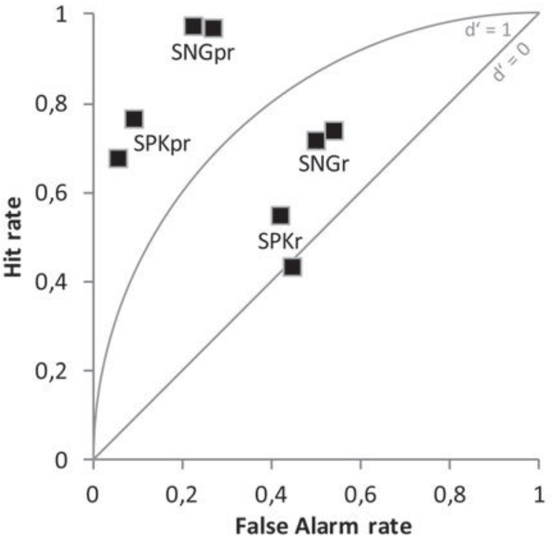

Figure 9.5: d-prime analysis for the stimulus pairs. The two dots per condition represent the two positions (1st and 2nd) in which a -pr and -r stimulus could occur. d' = 1 is equal to 69 % rating accuracy. d' = 0 is a random decision.

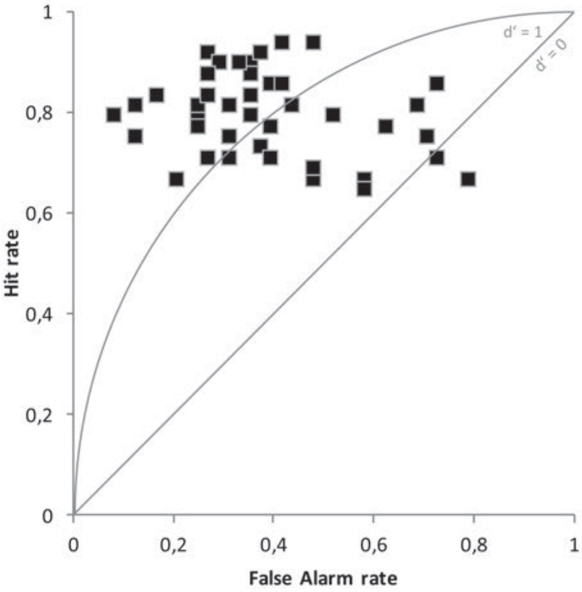

Figure 9.6: d-prime for each participant. 11 participants had a sensitivity lower than d' = 1, which is lower than 69 % rating accuracy and 2 participants had a sensitivity lower than d' = 0, which means they made random decisions.

Self assessment. A correlation between rating accuracy and self-assessment was found for the combinations SNGpr/SNGwpr, SNGr/SNGwpr, SPKpr/SPKwpr, SPKr/SPKwpr. When a full sentence (sung or spoken) followed a matching melody, participants rated the combination as easy and their performance was good. When a full sentence (sung or spoken) followed a matching rhythm, participants rated the combination as difficult and their performance was not as good.

Listening habits. No correlation was found between the response behavior and the participants' listening habits and experiences.

9.3 Discussion

The results show that (1) the stimuli were suited for the scanner, (2) participants had problems matching the rhythm stimulus pairs which (3) confirmed the plan to integrate the task in the scanning session only to ensure participants' attention and not for analyzing purposes; furthermore to use the rhythm condition only as a control condition and focus on melody and word processing in song and speech.

In the following, the rating results will be discussed in more detail. To summarize the findings, firstly, matching sung stimulus pairs (SNGwpr/SNGpr and SNGwpr/SNGr, and reverse) provoked a higher and faster rating performance than matching spoken stimulus pairs. Secondly, rating pitch-rhythm (melody) stimulus pairs revealed the highest performance (96 %), in contrast to rhythm stimuli, where performance was in some cases at chance. Thirdly, participants were faster in matching pairs with the -pr or -r stimulus in the second position.

From these combined results can be assumed that working memory plays a major role in matching these stimulus pairs; the less information a stimulus contains (the more abstract a stimulus is) and the less regular it is, the less likely it is to be remembered and therefore it is harder to compare with any other stimulus. (1) Sung stimuli are more regular than spoken stimuli and (2) rhythm stimuli are more abstract than -pr stimuli. Comparing regular and irregular rhythms, Grahn and Brett (2007) used a production task (tapping heard rhythms back) and showed that the metric simple (regular) rhythms are remembered correctly significantly more often than non-metric (irregular) rhythms. A regular grouping in the metric simple rhythms would lead to a stronger perception of a beat than in the metric complex rhythms. As discussed in Chapter 3.5 periodicity is the main difference between musical rhythm and speech rhythm. Regarding this experiment, a pair consisting of a full stimulus (-wpr) and melody (-pr) seemed to be the easiest combination to match (highest performance, 96 %). Both stimuli are easy to remember and therefore easier to match. Pairs that contain a full stimulus (-wpr) and rhythm (-r) were much harder to match because a rhythm stimulus contained less information than the -pr stimulus and was also more artificial, even though it was vocal, too. Miller (1956) stated that the number of items that can be held in working memory is limited to 7 ± 2 chunks. As the rhythm stimuli consisted of twelve monotonous tones (no pitch changes) they were much harder to remember than the -pr stimuli. Also, chunking (Miller, 1956) a sung stimulus with regular beats and discrete pitch changes is easier than spoken stimuli with a gliding, non-discrete pitch contour and irregular rhythm. Participants were also faster in matching pairs with the -wpr stimulus in the first position, most likely because it was the easiest to remember as it contained the most information and was a concrete stimulus.

The performance on the rhythm pairs was at chance for most of the rhythm conditions, but participants performed at chance mainly for the incorrect rhythm pairs. This can be explained by the participants' significant bias to rate pairs as matching (to respond 'yes'). Also evidently, participants scored significantly high for the correct song rhythm pairs (71.43 and 73.81 %). Furthermore, as shown by the d-prime analysis, participants' rating of SPKr pairs was random. While participants showed a very good sensitivity for all -pr pairs, the sensitivity for the -r pairs was low, especially for the SPKr pairs, which indicated a rating with random decisions.

As the results can be explained with working memory load, it can only be assumed how participants associated the connection between sung and spoken sentences and their underlying constituents – melody and rhythm. In a debriefing, participants were asked how they would describe and call the different stimuli – they had not been told about the type of stimuli beforehand (i. e. melody, rhythm, song or speech). They chose the word 'melody' to describe the -pr stimuli, but only a few participants actually named the rhythm stimuli 'rhythm'. Much more common were descriptions like 'beep tones'. Participants could not name the rhythm condition and therefore did not express a clear association for rhythm underlying song and speech. That could be explained by their lack of musical expertise as participants were not musically trained. For example, hardly anyone would match the hummed melody of the national anthem with the *'ABC'*-song, but contrasted with the rhythm (and even harder with a prosodic rhythm), a naïve, not musically trained person would have trouble matching them correctly, as he is not trained to memorize those abstract stimuli.

Or maybe participants did not perceive the (speech) rhythm as 'rhythmical', which means, the participants' expectation of what a 'rhythm' is, was different. Most likely, they expect a 'rhythm' that has a regular beat with pronounced metric accents, which is obvious in music; a speech rhythm does not exhibit such a regular beat. Maybe on these grounds participants did not term the rhythm stimuli 'rhythm'.

The debriefing questionnaire leads to the following conclusions. No correlation was found between the response behavior and the participants' listening habits and experiences. All participants were non-musicians and had not played an instrument for more than four years. The task required only a basic understanding of music and in the case that someone had not been able to perform above chance in the -pr pairs, a test for musical disorder would have been appropriate. Despite none ever having learned an instrument for very long, participants self-reported listening to music from 3 to 80 hours a week. The hypothesis that there could be a connection between the task performance and the interest in listening to music or spoken text (i. e. audio books) did not become significant. For example, participants who listened to music a lot did not show an advantage in matching the song pairs and so on.

Furthermore, the participants' self assessment reflected their response behavior. If a -wpr followed a -pr stimulus it was obviously easy to decide whether they matched or not, but if a -wpr followed a rhythm stimulus it was hard to decide whether they matched or not. That means in general, participants were conscious about their response behavior, especially about the discrepancy in matching -pr/-wpr pairs and -r/-wpr pairs.

9.4 Conclusion

It could be shown that the stimulus material was suited for the scanner. The poor performance on rhythm stimulus pairs confirmed the plan to use the task as an attention test only during the scanning session and exclude the task trials from the analysis, and to implement the rhythm condition as a control condition to extract pitch pattern perception in song and speech (see following fMRI study).

The participants' ability to correctly associate underlying pitch- and rhythm patterns in song and speech, respectively, could be shown for the pitch patterns, seen in the high performance, but not for the rhythms, which revealed a low performance.

Whether or not this has to do with a lack of understanding of rhythm as an underlying constituent of song and speech could not be shown, but can be considered unlikely. This observation was overshadowed by a high working memory load that lead most likely to the poor performance on rhythm stimulus pairs.

Chapter 10

Experiment 1B – fMRI Study Part 1 – Direct Comparisons

As previous fMRI studies on the neural correlates of song and speech focused predominantly on the direct comparison of (full) sung and spoken stimuli, i. e. stimuli containing linguistic information, pitch patterns and rhythm at the same time (Gunji et al., 2007; Jeffries et al., 2003; Özdemir et al., 2006; Riecker et al., 2000; Wildgruber et al., 1996; Callan et al., 2006; Saito et al., 2006; Schön et al., 2010), the following part reports the direct contrasts of sung and spoken sentences (first level) and melodies (second level; see Figure 8.1A) to broaden the spectrum of fMRI data on the perception of song and speech.

To summarize previous findings, both production and perception data from previous studies seem to suggest a predominant role of the right hemisphere in the processing of song due to pronounced musical features of the stimulus and a stronger left hemisphere involvement in speech due to focused linguistic processing. In particular, perception studies (Callan et al., 2006; Schön et al., 2010) found for the contrast song over speech (SNG > SPK) activations in the superior temporal gyrus and sulcus (STG/STS; BA 21, 22) bilaterally, Callan et al. more specifically in the anterior STG (aSTG) bilaterally. Both showed a right lateralized activation in the STG, Callan et al. by a lateralization test, Schön et al. by comparing the width of activation, which is not a statistical test. Furthermore, Callan et al. found activations in the rostral and dorsal part of the lateral premotor cortex (PMC) bilaterally and left orbitofrontal cortex and Schön et al. in the thalamus. Regarding the reverse contrast (SPK > SNG), neither Callan et al. nor Schön et al. found significant differential activity. While in Schön et al. no explanation was provided, Callan et al. reported under-threshold activation as a tendency for interpretation (for a detailed description of the studies see Chapter 4).

Taken together, the aSTG bilaterally seems to play a major role in song perception, as well as the PMC. A general lateralization for song to the right and speech to the left was suggested, but was not fully supported by statistical tests.

The current analysis will get back to this issue and broaden the song and speech perception research by three further insights: Firstly, by more data on the SNG vs. SPK contrasts, secondly, by new contrasts of SPKpr vs. SNGpr as none of the mentioned studies compared hummed spoken and hummed sung stimuli to evaluate differences in the underlying pitch-patterns of song and speech, and thirdly, by analyzing all contrasts for the first time with two different approaches, with univariate and multivariate analyses.

10.1 Methods[1]

10.1.1 Participants

Twenty-one healthy German native speakers (14 male, mean age 24.2 years, SD: 2.4 years) participated in the study. None of the participants were professional musicians, nor had learned to play a musical instrument for more than two years. All control participants reported to have normal hearing. Informed consent according to the Declaration of Helsinki was obtained from each participant prior to the experiment which was approved by the local Ethical Committee.

10.1.2 Materials

The paradigm consisted of 6 conditions (with 36 stimuli each) arranged in a subtractive hierarchy: spoken (SPKwpr) and sung sentences (SNGwpr) containing words, pitch patterns and rhythm; hummed speech prosody (SPKpr) and song melody (SNGpr) containing only pitch patterns and rhythm, as well as the speech or musical rhythm (SPKr and SNGr; see Figure 8.1A; sample stimuli will be provided on request).

The sentences for the 'wpr' stimuli were 6 different statements, with a constant number of twelve syllables across all conditions. The actual text content (lyrics) was carefully selected in order to be (a) semantically plausible in both, song and propositional speech (it is obviously not plausible to sing about taking the trash out) and (b) both the regular and irregular stress patterns were rhythmically compatible with the underlying melody (a stressed or prominent point in the melody never coincided with an unstressed word or syllable; see Figure 8.1B).

The 6 melodies for the sung (SNG) stimuli were composed according to the rules of Western tonal music, in related major and minor keys, duple and triple meters, and with and

---

[1] This part corresponds to Merrill et al. (2012).

without upbeat depending on the sentences. The lyric/tone relation was mostly syllabic. The melodies had to be highly distinguishable in key, rhythm and meter to make the task feasible (see below).

Melodies and lyrics were both unfamiliar to avoid activations due to long-term memory processes, automatic linguistic (lyric) priming, and task cueing. Spoken, sung (wpr) and hummed (pr and r) stimuli were recorded by a female trained voice who was instructed to avoid the Singer's Formant and ornaments like vibrato in the sung stimuli, to speak the spoken stimuli with emotionally neutral prosody and not to stress them rhythmically in order to keep them as natural as possible.

For the rhythm (r) conditions, a hummed tone (G3) was recorded and cut to 170 ms with 20 ms fade in and out. Sequences of hummed tones were created by setting the tone onset on the vowel onsets of each syllable according to the original sung and spoken material using Adobe Audition 3 (Adobe Systems). To control the hummed stimuli (pr and r) to be exactly equal in time and pitch as the spoken and sung sentences (wpr), they were adjusted using Celemony Melodyne Studio X (Celemony Software). All stimuli were cut to 3700 ms, normalized and compressed using Adobe Audition 3 (Adobe Systems).

10.1.3 Procedure

Across the experiment, each of the 36 stimuli was presented 6 times in a pseudo-random order (see below), interleaved with 20 baseline conditions (no sound played) and 36 task trials (requiring a response), resulting in 272 stimulus presentations in total. In an effort to avoid adaptation effects, exactly the same stimuli, stimuli with the same melody/text, or stimuli from the same level (wpr, pr, r) were not allowed to follow each other in the pseudo-randomized stimulus list.

The duration of the experiment was 34 minutes. For stimulus presentation and recording of behavioral responses, the software Presentation 13.0 (Neurobehavioral Systems, Inc., San Francisco, CA) was used.

The participants were instructed to passively listen to the sounds, without being informed about the kind of stimuli, like song or speech, melody or rhythm. To assure the participants' attention, 36 task trials required a same/different judgment with the stimulus of the preceding trial. The stimulus of the task trial (e. g., SNGwpr) was always taken from a different hierarchical level than the preceding stimulus (e. g., SNGr) and participants were required to indicate via button press whether the two stimuli were derived from the same original sentence or song. Prior to the experiment, participants received a short training to assure quick and accurate responses.

The timeline of a single passive listening trial (for sounds and silence) is depicted in Figure 8.1C: The duration of a passive listening trial was 7500 ms, during which the presentation of the stimulus (3700 ms; prompted by '+') with a jittered onset delay of 0, 500, 1000, 1500 or 2000 ms was followed either by '...' or '!' shown for the remaining trial duration between 1800 and 3800 ms. The three dots ('...') indicated that no task would follow. The exclamation mark ('!') informed the listeners that instead, a task trial would follow, i. e. that they had to compare the next stimulus with the stimulus they had just heard.

The timeline of a task trial was analogous to a passive listening trial except for the last prompt, a '?' indicating the time to respond via button press (see Figure 8.1C). Trials were presented in a fast event-related design. Task trials did not enter data analysis.

Scanning. Functional magnetic resonance imaging (fMRI) was performed on a 3T Siemens TrioTim scanner (Erlangen, Germany) at the Max Planck Institute for Human Cognitive and Brain Sciences in Leipzig. In an anatomical T1-weighted 2D-image (TR 1300 ms, TE 7.4 ms, flip angle 90°) 36 transversal slices were acquired. During the following functional scan one series of 816 BOLD images was continuously acquired using a gradient echo-planar imaging sequence (TR 2500 ms, TE 30 ms, flip angle 90°, matrix 64×64). 36 interleaved axial slices (3×3×3 mm voxel size, 1 mm interslice gap) were collected to cover the whole brain and the cerebellum. We made sure that participants were well able to hear the stimuli in the scanner.[2]

10.1.4 Data Analysis

Univariate Analysis. FMRI data were analyzed using SPM 8 (Welcome Department of Imaging Neuroscience). Images were realigned, unwarped using a fieldmap scan, spatially normalized into the MNI stereotactic space, and smoothed using a 4 mm FWHM Gaussian kernel. Low-frequency drifts were removed using a temporal high-pass filter with a cutoff of 128 s.

A general linear model using 4 regressors of interest (one for each of the four conditions, without the rhythm control condition) was estimated in each participant. Regressors were modeled using a boxcar function convolved with a hemodynamic response function to create predictor variables for analysis. For random effect group analyses, the individual SPM's were submitted to one-sample t-tests.

For direct comparisons of sung and spoken sentences (first level; see Figure 8.1A), regardless pitch- and rhythm-specificity in song and speech, SPKwpr-SNGwpr and SNGwpr-SPKwpr were contrasted. For direct comparisons of prosodic and musical pitch patterns (second level), regardless rhythm-specificity in song and speech, SPKpr-SNGpr and SNGpr-SPKpr were contrasted.

---

[2] End of quotation from the published manuscript.

Multivariate Pattern Analysis. The multivariate pattern analysis (MVPA) was carried out using SPM8 (Welcome Department of Imaging Neuroscience) and PyMVPA 0.4 (Hanke et al., 2009). Images were motion corrected before a temporal high-pass filter with a cut-off of 128 s was applied to remove low-frequency drifts. At this point no spatial smoothing and no normalization into MNI stereotactic space were performed to preserve the fine spatial activity patterns. Next, a contrast of interest was chosen. These contrasts included the same as with the UVA. MVPA was performed using a linear support vector machine (libsvm C-SVC, C.-C. Chang & Lin, 2011). For every task trial of the conditions, one image was selected as input for MVPA. To accommodate hemodynamic response, an image 7 seconds after stimulus onset was acquired by linear interpolation of the fMRI time series. Data were divided into 5 subsets each containing 7 images per condition to allow for cross validation. Each subset was independently z-scored relative to baseline condition. We used a searchlight approach (Kriegeskorte et al., 2006) with a radius of 8 mm to map brain regions which were differentially activated during both conditions of interest. This resulted in accuracy maps of the whole brain. The resulting images were spatially normalized into the MNI stereotactic space, and smoothed using a 4 mm FWHM Gaussian kernel.

Accuracy maps of all subjects were then submitted to a second-level group analysis comparing the mean accuracy for each voxel to chance level (50 %) by means of one-sample t-tests and thresholded at p (cluster-size corrected) < 0.05 in combination with p (voxel-level uncorrected) < 0.001. Localization of brain areas was done with reference to the Juelich Histological Atlas, Harvard-Oxford (Sub)Cortical Structural Atlas and activity within the cerebellum was determined with reference to the atlas of Schmahmann et al. (2000).

10.2 Results

10.2.1 Direct Comparison of Sung and Spoken Sentences (first level)

UVA. The univariate analysis revealed significant activation clusters for the SNGwpr > SPKwpr contrast in the bilateral aSTG and the right PMC (see Table 10.1 and Figure 10.1, left column, top row for details) and for the reverse contrast SPKwpr > SNGwpr in the left inferior parietal lobule in area PGp adjacent to the occipital cortex (see Table 10.1 and Figure 10.1, left column, middle row).

MVPA. Brain regions that distinguish between spoken and sung sentences were found in the STG/STS bilaterally, left posterior cingulum, right PMC (extending into the somatosensory cortex) and supplementary motor area (SMA; see Figure 10.1, left column, bottom row).

10.2.2 Direct Comparison of Sung and Spoken Melodies (second level)

UVA. The univariate analysis revealed significant activation clusters for the SPKpr > SNGpr contrast in left posterior STS/middle temporal gyrus (pSTS/pMTG), while the reverse contrast did not reveal any suprathreshold clusters (see Table 10.1 and Figure 10.1, right column, top row).

MVPA. Brain regions that distinguish between pitch patterns in speech and song were found in the STG/STS bilaterally and in the right anterior intraparietal sulcus (IPS; see Figure 10.1, right column, bottom row).

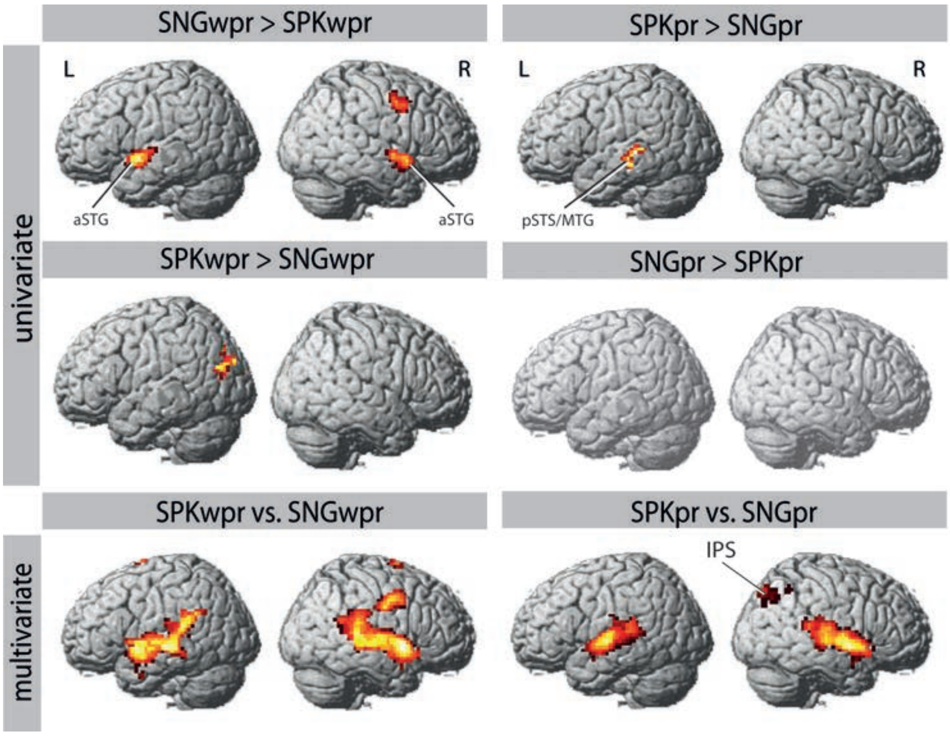

Figure 10.1: Brain regions that distinguish between song and speech on sentence level (first level) and pitch-rhythm pattern level (second level) in univariate and multivariate analyses (p (cluster-size corrected) < 0.5 in combination with p (uncorrected) < 0.001). aSTG = anterior superior temporal gyrus; pSTS/MTG = posterior superior temporal sulcus/middle temporal gyrus.

10.3 Discussion

The two most striking results are the activation in the aSTG bilaterally for SNGwpr > SPKwpr which most likely represents the music relatedness of song and is in

SONG vs. SPEECH (direct contrasts)

Region	BA	hem	cluster extent	MNI coordinates			Z value	cluster p(cor)
				x	y	z		
1st level (wpr)								
SNGwpr > SPKwpr (UVA)								
aSTG (Planum polare)	22	L	164	-51	2	-5	5.63	.000
aSTG (Temporal pole)	22	R	143	51	8	-8	4.95	.000
Premotor cortex	6		74	54	-1	46	5.3	.004
SPKwpr > SNGwpr (UVA)								
Inferior parietal lobule	(PGp)	L	73	-33	-88	28	4.71	.005
SPKwpr vs. SNGwpr (MVPA)								
STG/STS	22	L	1205	-57	-28	-5	5.19	.000
Posterior cingulum			76	-6	-19	37	3.95	.026
STG/STS	22	R	1151	57	8	1	5.35	.000
Primary somatosensory cortex	3b		195	51	-10	34	4.85	.000
Supplementary motor area	6		129	9	-1	76	4.74	.001
2nd level (pr)								
SPKpr > SNGpr (UVA)								
pSTS/MTG	22	L	48	-66	-31	4	3.99	.049
SPKpr vs. SNGpr (MVPA)								
STG/STS	22	L	703	-51	-13	-2	6.27	.000
STG/STS	22	R	999	57	-7	1	6.06	.000
Inferior parietal lobule	(PGp)		92	36	-79	46	3.95	.012
Anterior intraparietal sulcus	(hIp3)			36	-55	46	3.73	

Table 10.1: Brain areas involved in the processing of sung and spoken sentences, and prosodic and melodic pitch patterns. All p(cluster-size corrected) < 0.05 in combination with p(uncorrected) < 0.001)

line with previous studies, and in the IPS for SPKpr vs. SNGpr. As this contrast is presented for the first time, the role of the IPS in differentiating between prosodic and musical stimuli needs to be discussed.

10.3.1 Comparison of Song and Speech

Anterior Superior Temporal Gyrus. The bilateral aSTG activation for song over speech is in line with results from previous studies (Callan et al., 2006; Schön et al., 2010; also in production: Jeffries et al., 2003; see above). The assumption was that the aSTG represents (1) the music related processes and (2) the melodic aspects of song – associated with complex pitch perception in the aSTG (referring to Scott et al., 2000).

With 'music relatedness' the authors most likely refer to the rule-based dependency of song in Western tonal music, of which the 'melodic aspect' is just one specification. The aSTG was suggested to represent melodic processes in song. This statement cannot be taken as the only possibility as not only the melodic aspects of song and speech were compared. Both stimuli contained words and rhythmic aspects at the same time. Therefore, speculations of other influences can be made.

Firstly, the aSTG could also represent the regularity of song compared to speech which includes the regularity of pitch patterns (discrete pitch) and the regularity of the rhythm (a regular timed beat). The aSTG was found to be involved in syntactic processes in music (Koelsch, Fritz, et al., 2005; Koelsch et al., 2002) – but also in speech (e. g., Friederici, 2011). Hypothetically, this could mean that structure building processes are more engaged in song perception than in speech when the structure of the text, the pitch pattern and the rhythm pattern need to be processed.

Secondly, why should the melody in music be more complex than prosody in speech and therefore involve the aSTG stronger than speech? In a number of melody and pitch perception studies, the (mostly right) aSTG has been found to be involved – e. g., in patients with lesions in the right anterior TL the melodic processing was disturbed (Liégeois-Chauvel et al., 1998). In the temporal lobe (with a right-hemispheric weighting in more anterior regions) a hierarchy of melody processing has been suggested (Patterson et al., 2002; Griffiths et al., 1998). Random and tonal melodies activated anterior regions to Heschl's Gyrus more than fixed pitch - but with no significant difference between random and tonal melodies. As both stimuli contained discrete pitch, no statement regarding song and speech, where discrete vs. gliding pitch is processed, can be made, therefore the assumption that discrete pitch is more complex than gliding pitch cannot be held.

To speculate on another possibility, under the assumption that the aSTG is more involved in processing the complexity of sounds (e. g., complex pitch), it might be that the aSTG reflects the more complex spectral sound structure of song compared to speech. In song, vowels are produced in a more stable way and therefore the spectro-temporal complexity of sounds might be more demanding than in speech. The aSTG might be a core region in classifying not music (i. e. instrumental music) from speech but rather song from speech due to the combination of articulation and pitch (see Chapter 4).

An explanation for the activation of the STG bilaterally for song was that the involvement of the left hemisphere might be due to vocal lyrical singing (Callan et al., 2006) – as opposed to results from other studies that used non-vocal stimuli – and therefore that the context of language causes a shift to the left hemisphere (Wong et al., 2004). Furthermore, it is likely that the basic acoustic features of song and speech are so close that they are processed in overlap in the temporal lobe but yet somehow differently, as suggested by the MVPA results which show a discriminating pattern between both, but without a clear lateralization. Lateralization might be more likely to be found in higher-order areas outside the STG, like the IFG, the IPS, the MTG or others.

To conclude, the aSTG bilaterally is definitely a core region in song perception, but it is not clear yet due to which component of song. An investigation with separated components (words, pitch and rhythm) in an fMRI study, together with studies on lesions in the aSTG might shed light on this issue.

PMC and SMA. The direct contrast of SNGwpr > SPKwpr revealed activations in BA 6. Considering song having more regular stimulus features than speech, the activation in the PMC could represent regularity. Chen et al. (2008) for example found motor areas to be involved in music and therefore regular rhythm perception. Moreover, the PMC was associated with auditory-to-articulatory mapping (PMC; Brown et al., 2008; Hickok et al., 2003; Kleber et al., 2010; Wilson et al., 2004) and could represent sub-vocal rehearsal. These and more motor-regions were also found in the following analysis of the fMRI data which are discussed in detail in Chapter 11.3.

Area PGp. In the current study, the SPK > SNG contrast revealed activation in PGp, the caudal part of the inferior parietal cortex (IPC), which is located at the border of the visual cortex (see Figure 3.1). Studies on the organization of the IPC based on receptor architectonics showed that area PGp might provide a transition to adjoining visual cortex, evidenced by a similar fingerprint to that of ventral extrastriate visual cortex. Area PGp might be classified as not being a typical parietal cortex but as a higher visual area, linking occipital and parietal cortex (Caspers et al., 2012). It is not clear why visual areas are activated in this contrast, maybe due to visual imagery of the stimuli (Kleber et al., 2007; Foster & Zatorre, 2010). Activations of visual areas were also found in previous studies

(e. g., Perry et al., 1999; Kleber et al., 2007) and in the following analysis (see Chapter 11).

A note on the SPK > SNG contrasts: As the SPK > SNG contrast in the studies by Schön et al. (2010) and Callan et al. (2006) and the current study did not reveal noteworthy significant differences, one could also state that song activates a stronger network overall compared to speech, as it is known that the STG also plays a role in language processing (e. g., Friederici, 2002; Friederici & Alter, 2004; Friederici, 2011). A reason for the stronger network of song could be that it is the acoustically richer stimulus than speech and also the less familiar stimulus compared to neutral speech, which represents the acoustical event of everyday life.

10.3.2 Comparison of Prosody and Melody

Intraparietal Sulcus. The IPS is part of what has classically been considered the 'dorsal stream' (Culham & Kanwisher, 2001). It is anatomically and functionally connected with the dorsolateral prefrontal cortex, playing a role in working memory tasks requiring monitoring (Champod & Petrides, 2007). A fronto-parietal network has also been discussed in relation to selective attention and feature integration in the visual domain (Donner et al., 2002; Shafritz et al., 2002; Wei et al., 2011), but also for spatial transformations such as visual mental rotation (for a meta-analysis see Zacks, 2008). However, the function of the IPS is not restricted to the visual domain – and modality independent processing within the IPS was suggested by research on macaque brains and corresponding human analogue functional imaging data (for a review see Grefkes & Fink, 2005). A link between visual rotation and auditory transformation was proposed by Zatorre et al. (2010), demonstrated in a melody imagery task. Ruling out a visuospatial strategy to perform the task, they proposed that transformation of a mental image can take place in both, visual and auditory transformations: a visual arrangement has to be computed regarding its spatial position, an auditory regarding its temporal position, as tones in a melody are presented over time.

It was suggested that the IPS plays a role in dealing with the frequency relationship between stimuli, as has for example been shown in tasks involving pitch-shifts (Rinne et al., 2007; Zarate & Zatorre, 2008; Zarate et al., 2010). Two recent studies examined relative pitch in melodies and chords while trying to control for other influential factors, such as task demands and cognitive load. One found distinct activation for transposed melodies (compared to simple melodies) in the IPS (Foster & Zatorre, 2010), the other investigated categorical perception in music, using minor and major chords, which also revealed activation in the IPS (Klein & Zatorre, 2011; for details see Chapter 3.1). The authors suggest that the IPS might reflect relative pitch encoding which is a fundamental element in music perception.

So far, in the light of the current approach, the reported function of the IPS in the literature can be described with the computation of the relationship of elements (as evidenced by Champod & Petrides, 2007; Foster & Zatorre, 2010; Shafritz et al., 2002; Zatorre et al., 2010; Wei et al., 2011), in music the relative pitch relations, and to integrate their neural representation (Alexander et al., 2005; Cusack, 2005; Donner et al., 2002; Hill & Miller, 2010). The role of the IPS in musical pitch processing needs further investigation with passive listening studies to control for task demands and working memory load. In the current dissertation task demands were kept low by a small number of task trials (36 out of 272 trials in total) and were not subjected to the analysis. This will be discussed in detail in Chapter 11.5, together with the results from the following analysis of the fMRI data.

To conclude, the current analysis revealed the IPS to differentiate between prosody and musical melody. As this was non-directional information (MVPA), it can only be assumed from the above discussed studies that the IPS might code for the musical pitch information in the hummed musical melody. Musical pitch information might be easier to monitor and to pay attention to due to its discreteness – a feature that prosody is lacking. To be more concrete, the relationship between discrete tones represented over time might be better to detect and to monitor than the non-discrete pitch glides in speech. Evidence for that came from the behavioral pretest that revealed the highest performance with the lowest standard error for the musical melody pairs (i. e. comparing full sentences with the underlying melody) over all conditions. This is in line with a correlation between a high IPS activation and a good behavioral performance reported by Foster and Zatorre (2010).

Posterior STS/MTG. A region encompassing the left pSTS and MTG (MNI coordinates: -66, -31, 4) was found to be activated for hummed prosody over hummed melody (SPKpr > SNGpr). Similar activations were found for contrasting speech-stimuli with musical stimuli: detecting consonant-vowel pairs during dichotic listening with musical instruments (simple tones; Hugdahl et al., 1999; Talairach coordinates: -61, -26, 8) and attention to consonant-vowels versus attention to musical instruments (Hugdahl et al., 2000; Tal.: -59, -32, 7), for word identification and meaning (Jernigan et al., 1998; Tal.: -60, -30, 0), and also for morphed speech sounds compared to morphed music sounds (Specht et al., 2009; Tal.: -57, -36, 0). In a recent meta-analysis the left MTG was found in the contrast human vocalizations > music (Schirmer et al., 2012; Tal.: -52, -32, 0). Concluding, this area in the posterior STS seems to be specific for speech sounds. Furthermore, the MTG was also suggested to play a specific role in language processing. Generally it has been suggested as a region supporting the sound-to-meaning mapping, i. e. widely distributed lexical-semantic representations and concepts (e. g., Binder et al., 1997, 2000; Hickok & Poeppel, 2007; Rissman et al., 2003), and specified by lesion studies that show that damage to the posterior MTG causes severe deficits of word-level comprehension (Bates et al., 2003; Dronkers et al., 2004). Transferring this to the current finding, the MTG activation could represent the participants' attempt to assign meaning to the hummed prosodic stimuli. In neutrally spoken stimuli the words transfer the content and create the meaning (semantics). As the stimuli in the current study did not contain words, it can be assumed that participants tried to add words in order to make out the meaning in the prosodic stimulus. This is more likely for the hummed prosodic stimulus than for the sung melody as music is independent from words (i. e. instrumental music). It is likely that participants tried to figure out what has been said and therefore the activation in the posterior MTG was found.

A note on the SNGpr > SPKpr contrasts: The reverse contrasts SNGpr > SPKpr did not reveal significant results. Again, this might be the case, because a hummed song melody is the more familiar sound and therefore does not activate the brain as much as a less familiar hummed prosody. As the same areas found in these contrasts were also found in the following analysis, they are discussed together in the next chapter.

10.4 Limitations of the Current Approach

As stated in the beginning of this chapter, the approach on different neural patterns underlying song and speech processing by comparing stimuli directly is limited and results are open for speculations. Some examples are discussed in the following paragraph to underline this thesis.

The described activations in the bilateral aSTG for song over speech were suggested to reflect music related processes, i. e. the processing of the melodic aspects of song (see also Callan et al., 2006). This might be true, but in this contrast, not only the melodic aspects of song and speech were compared, but also linguistic and timing aspects (as words/lyrics and rhythm) which affected the results. Therefore, one cannot be sure if the aSTG only reflects pitch processing in music. As another example, the reverse contrast (SPK > SNG) has not yet revealed meaningful results as the sung stimulus is the acoustically richer stimulus; and speech most likely 'disappears' in contrast to song. But if song and speech stimuli are divided into their components (i. e. words, pitch and rhythm patterns), stimuli become acoustically less complex and the influence of other acoustic cues becomes smaller. This should lead to a more specific interpretation of engaged brain areas involved in song and speech processing. As an example from the following analysis (Chapter 11), the comparison of spoken words and sung words (corrected for pitch and rhythm information) revealed that the spoken words engaged the left IFG while sung words did not – a results for SPK > SNG.

10.5 Conclusion

In the current analysis, stimuli from the same level were contrasted directly: sung and spoken sentences, as well as hummed prosody versus musical melody. Taken together, the results from previous studies and the current analysis constitute two core areas in song and speech perception: the aSTG bilaterally and the IPS. While the aSTG has suggested itself to be crucial in song processing, the IPS is associated with one or the other, but it is not yet clear which. Although, while the IPS is most probably involved in differentiating song and speech on the pitch pattern level, the exact role of the aSTG in song perception has to be clarified by investigating the constituting parameters of song separately – using the subtractive hierarchy of the stimuli. This approach is presented in the next chapter.

Chapter 11

Experiment 1B – fMRI Study Part 2 – Perception of Words and Pitch Patterns

Direct contrasts of song and speech have been reported in previous studies and the last chapter. The critique on this approach was that those contrasts did not allow for fully separating out the influence of words and pitch patterns on the differential coding for song and speech in the brain. The stimuli differed in their acoustic features and therefore participants were required to pay attention to more than one dimension (i. e. linguistic and melodic dimension in one stimulus). The following analysis of the fMRI data focused on a hierarchical paradigm to isolate the neural correlates of the word- and pitch-based discrimination between song and speech.

11.1 Data Analysis[1]

Univariate Analysis. FMRI data were analyzed using SPM8 (Welcome Department of Imaging Neuroscience). Images were realigned, unwarped using a fieldmap scan, spatially normalized into the MNI stereotactic space, and smoothed using a 6 mm FWHM Gaussian kernel. Low-frequency drifts were removed using a temporal high-pass filter with a cutoff of 128 s.

A general linear model using 6 regressors of interest (one for each of the six conditions) was estimated in each participant. Regressors were modeled using a boxcar function convolved with a hemodynamic response function to create predictor variables for analysis.

---

[1] This part corresponds to Merrill et al. (2012).

The no-stimulus (silent) trials served as an implicit baseline. Contrasts of all 6 conditions against the baseline were then submitted to a second-level within-subject analysis of variance. Specific contrasts were assessed to identify brain areas involved in word and pitch processing in spoken and sung stimuli in the human brain.

For word processing, the activations for the hummed stimuli were subtracted from the full spoken and sung stimuli separately for song and speech (SPKwpr-SPKpr and SNGwpr-SNGpr). To obtain differences in word processing between song and speech, these results were compared, i. e. [(SPKwpr-SPKpr)-(SNGwpr-SNGpr)] and [(SNGwpr-SNGpr)-(SPKwpr-SPKpr)].

To identify brain areas involved in the pure pitch processing in song and speech, the activation for the rhythm condition was subtracted from the pitch-rhythm condition (SPKpr-SPKr and SNGpr-SNGr) and compared, i. e. [(SPKpr-SPKr)-(SNGpr-SNGr)] and [(SNGpr-SNGr)-(SPKpr-SPKr)].

To identify brain areas that are commonly activated by the different parameters of speech and song, additional conjunction analyses were conducted for words, i. e. [(SPKwpr-SPKpr) ∩ (SNGwpr-SNGpr)] as well as pitch patterns, i. e. [(SPKpr-SPKr) ∩ (SNGpr-SNGr)] using the principle of the minimum statistic compared to the conjunction null (Nichols et al., 2005).

Multivariate Pattern Analysis. The multivariate pattern analysis (MVPA) was carried out using SPM8 (Welcome Department of Imaging Neuroscience) and PyMVPA 0.4 (Hanke et al., 2009). Images were motion corrected before a temporal high-pass filter with a cut-off of 128 s was applied to remove low-frequency drifts. At this point no spatial smoothing and no normalization into MNI stereotactic space were performed to preserve the fine spatial activity patterns. Next, a contrast of interest was chosen. These contrasts included the same as with the UVA. MVPA was performed using a linear support vector machine (libsvm C-SVC, Chih-Chung Chang and Chih-Jen Lin). For every task trial of the conditions, one image was selected as input for MVPA. To accommodate hemodynamic response, an image 7 seconds after stimulus onset was acquired by linear interpolation of the fMRI time series. Data were divided into 5 subsets each containing 7 images per condition to allow for cross validation. Each subset was independently z-scored relative to baseline condition. We used a searchlight approach (Kriegeskorte et al., 2006) with a radius of 6 mm to map brain regions which were differentially activated during both conditions of interest. This resulted in accuracy maps of the whole brain. The resulting images were spatially normalized into the MNI stereotactic space, and smoothed using a 6 mm FWHM Gaussian kernel. Accuracy maps of all subjects were then submitted to a second-level group analysis comparing the mean accuracy for each voxel to chance level (50%) by means of one-sample *t*-tests.

In general, analyzing multivariate data is still a methodological quest, specifically regarding the best way of performing group statistics. t-tests on accuracy maps are common practice (Bode et al., 2012; Bogler et al., 2011; Haxby et al., 2001; Kahnt et al., 2010; Tusche et al., 2010) although accuracies are not necessarily normally distributed. Nonparametric tests and especially permutation tests have better theoretical justification, but remain computationally less feasible.

All reported group SPM statistics for the univariate and the multivariate analyses were thresholded at $p(\text{cluster-size corrected}) < 0.05$ in combination with $p(\text{voxel-level uncorrected}) < 0.001$. The extent of activation is indicated by the number of suprathreshold voxels per cluster. Localization of brain areas was done with reference to the Juelich Histological Atlas, Harvard-Oxford (Sub)Cortical Structural Atlas and activity within the cerebellum was determined with reference to the atlas of Schmahmann et al. (2000).

Region of interest analysis. To test for the lateralization of effects and specify differences between song and speech in the IFG and IPS, regions of interest (ROIs) were defined. According to the main activation peaks found in the whole-brain analysis, ROIs for left and right BA 47 were taken from the Brodmann Map using the template implemented in MRIcron (http://www.mccauslandcenter.sc.edu/mricro/mricron/). ROIs for the left and right IPS (hIP3) were taken from the SPM-implemented anatomy toolbox (Eickhoff et al., 2005). Contrast values from the uni- (beta values) and multivariate (accuracy values) analyses were extracted for each participant in each ROI by means of MarsBar (http://marsbar.sourceforge.net). Within-subject analyses of variance (ANOVA) and paired-sample t-tests were performed for each ROI using PASW Statistics 18.0. Normal distribution of the accuracies was verified in all ROIs using Kolmogorov-Smirnov tests (p's > 0.643).

11.2 Results

11.2.1 Words in Song and Speech

UVA. The contrasts of spoken words over prosodic pitch-rhythm patterns (SPKwpr-SPKpr) and sung words over musical pitch-rhythm patterns (SNGwpr-SNGpr) showed similar activated core regions (with more extended cluster activations for the sung stimuli) in the superior temporal gyrus/sulcus (STG/STS) bilaterally and for the SNGwpr-SNGpr additionally in left medial geniculate body (see Table 11.1 and Figure 11.1, top row for details). The overlap of these activations was nearly complete as evidenced by a conjunction analysis and no significant differences in the direct comparison of both contrasts, i. e.

[(SPKwpr-SPKpr)-(SNGwpr-SNGpr)] and [(SNGwpr-SNGpr)-(SPKwpr-SPKpr)].

MVPA. The MVPA revealed brain regions that distinguish significantly between words and pitch-rhythm patterns for both song (SNGwpr vs. SNGpr) and speech (SPKwpr vs. SPKpr) in the STG/STS and premotor cortex bilaterally (extending into the motor and somatosensory cortex; see Table 11.1 for details). For speech, in the SPKwpr vs. SPKpr contrast, additional information patterns were found in the supplementary motor area (SMA), the cerebellum, the pars orbitalis of the left IFG (BA 47), the right superior parietal lobule (BA 7) and the visual cortex (BA 17). For song, the SNGwpr vs. SNGpr contrast showed additional peaks in the pars orbitalis of the right IFG (BA 47) and the adjacent frontal operculum (see Figure 11.1, bottom row). Interestingly, the results were suggestive of a different lateralization of IFG involvement in spoken and sung words. To further explore this observation, accuracy values were extracted from anatomically defined regions of interest (ROI) in the left and right BA 47 (see Methods) and subjected to an ANOVA for repeated measures with the factors Hemisphere (left/right) and Modality (speech/song). This analysis showed a significant interaction of Hemisphere × Modality [$F(1,20) = 5.049$, $p < 0.036$], indicating that the left and right BA 47 were differentially involved in discriminating words from pitch in song and speech. Subsequent t-tests for paired samples revealed that in song, right BA 47 showed predominance over left BA 47 [$t(20) = -2.485$, $p < 0.022$], whereas the nominally opposite lateralization in speech fell short of significance ($p > 0.05$). Moreover, left BA 47 showed predominance for word-pitch discrimination in speech compared to song [$t(20) = 2.453$, $p < 0.023$] (see bar graphs in Figure 11.1).

11.2.2 Pitch Patterns in Song and Speech

UVA. Activation for processing pitch information was revealed in the contrast of prosodic pitch-rhythm patterns vs. prosodic rhythm patterns (SPKpr-SPKr) for speech and in the contrast musical pitch-rhythm patterns vs. musical rhythm patterns (SNGpr-SNGr) for song (Table 11.2 and Figure 11.2, top row). Note that these contrasts allow for investigating pitch in song and speech corrected for differential rhythm patterns. Both showed activations in the STG/STS bilaterally and in the premotor cortex bilaterally. For speech, the prosodic pitch patterns (SPKpr-SPKr) showed further activations in the pars orbitalis of the left IFG (BA 47) and the SMA.

The musical pitch patterns (SNGpr-SNGr) showed further activations in the pars orbitalis of the right IFG (BA 47), the cerebellum bilaterally, the left anterior cingulate cortex (ACC), the left lateral occipital cortex, the midline of the visual cortex, the right caudate nucleus, as well as a cluster in the parietal lobe with peaks in the left precuneus and the anterior intraparietal sulcus (IPS; see Table 11.2 and Figure 11.2, top row).

WORDS

Region	BA	hem	cluster extent	MNI coordinates			Z value	cluster p(cor)
				x	y	z		
Speech								
SPKwpr > SPKpr (UVA)								
STG/STS	22	L	1124	-36	-31	10	Inf	.000
STG/STS	22	R	757	42	-25	10	Inf	.000
SPKwpr vs. SPKpr (MVPA)								
STG/STS	22	L	3624	-66	-16	4	7.65	.000
Premotor cortex	6			-45	-4	49	4.66	
Cerebellum Crus I			158	-45	-67	-20	5.71	.000
Cerebellum VI lobule				-27	-58	-23	4.33	
IFG	47		146	-42	26	-11	4.18	.000
Visual cortex V1	17		87	-15	-109	4	4.13	.010
STG/STS	22	R	3581	66	-7	1	7.37	.000
Primary motor cortex	4a			54	-7	43	5.45	
Supplementary motor area	6		352	0	2	61	4.95	.000
Superior parietal lobule	7		115	3	-70	25	3.94	.002
Song								
SNGwpr > SNGpr (UVA)								
STG/STS	22	L	1486	-60	-10	-2	Inf	.000
Thalamus (Medial geniculate body)			194	-12	-28	-2	5.62	.000
STG/STS	22	R	1112	42	-25	10	Inf	.000
SNGwpr vs. SNGpr (MVPA)								
STG/STS	22	L	2663	-57	-13	4	7.67	.000
Premotor cortex	6			-45	-10	49	4.34	
STG/STS	22	R	2486	51	-7	7	7.82	.000
IFG	47			45	20	-11	4.37	
Frontal operculum				36	23	4	4.00	
Primary somatosensory cortex	1		164	57	-7	40	5.47	.000
Conjunction SPKwpr-SPKpr ∩ SNGwpr-SNGpr								
STG/STS (Planum temporale)	22	L	1065	-36	-31	10	Inf	.000
STG/STS (Planum temporale)	22	R	700	42	-25	10	Inf	.000

Table 11.1: Brain areas involved in the processing of words in song and speech. All p(cluster-size corrected) < 0.05 in combination with p(uncorrected) < 0.001.

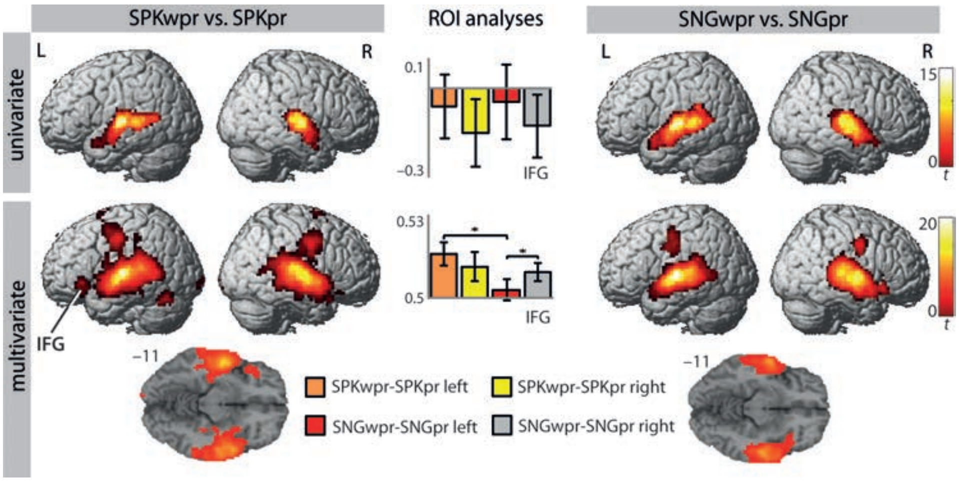

Figure 11.1: Brain regions that distinguish between words and pitch-rhythm patterns in song and speech (first vs. second level: SPKwpr vs. SPKpr and SNGwpr vs. SNGpr) (p(cluster-size corrected) < 0.05 in combination with p(uncorrected) < 0.001). Bar graphs depict beta values (UVA) and accuracy values (MVPA) of the shown contrasts extracted from left and right BA 47. Significant differences between conditions are indicated by an asterisk (p < 0.05). Color scales on the right indicate t-values for each row. IFG: inferior frontal gyrus.*

A conjunction analysis of both contrasts showed shared bilateral activations in the STG/STS (planum polare) and in the premotor cortex bilaterally. Despite the differential involvement of IFG, cerebellum and IPS listed above, these differences between pitch-related processes in song and speech fell short of statistical significance in the whole-brain analysis.

Again, the results were suggestive of a differential lateralization of IFG activity during pitch processing in speech and song. Therefore, an ANOVA with the repeated-measures factors Hemisphere (left/right) and Modality (speech/song) as well as t-tests for paired samples (comparing the hemispheres within each modality) were conducted on the beta values of the contrast images extracted from ROIs in the left and right BA 47 (see Methods). This analysis showed a significant interaction of Hemisphere × Modality [$F(1,20) = 5.185$, $p < 0.034$], indicating that the left and right BA 47 were differentially involved in the processing of pitch patterns in speech and song. Subsequent t-tests showed that while left BA 47 was more strongly involved during spoken pitch processing than right BA 47 [$t(20) = 2.837$, $p < 0.01$], no such lateralization was found for sung pitch [$t(20)$, $p > 0.9$]. Furthermore, involvement of right BA 47 was marginally stronger during pitch processing in song compared to speech [$t(20)$ = -2.032, $p < 0.056$], whereas no such difference was found for left BA 47.

Considering the growing evidence that the IPS is involved in the processing of pitch in music (Foster & Zatorre, 2010; Klein & Zatorre, 2011; Zatorre et al., 2009, 1994) and as the IPS was only activated in the sung pitch contrast (SNGpr-SNGr) and not in the spoken pitch contrast (SPKpr-SPKr), an additional ROI analysis was performed to further explore differences in sung pitch and spoken pitch. Therefore, contrast values were extracted from anatomically defined ROIs in the left and right IPS (see Methods) and subjected to an ANOVA for repeated measures with the factors Hemisphere (left/right) and Modality (speech/song). This analysis showed a significant main effect of Modality [$F(1,20) = 5.565$, $p < 0.029$] and no significant interaction of Hemisphere × Modality [$F(1,20) = 1.421$, $p > 0.3$], indicating that both, the left and the right IPS, were more strongly activated by sung than spoken pitch patterns.

MVPA. The MVPA revealed brain regions that distinguish between pitch-rhythm patterns and rhythm patterns for both song and speech in the STG/STS bilaterally, bilateral premotor cortex (extending into motor and somatosensory cortex) and SMA. For the SPKpr vs. SPKr comparison a peak in the left IFG (BA 45) was found (see Figure 11.2, bottom row). For SNGpr vs. SNGr additional clusters were found in the left anterior cingulate gyrus and left anterior IPS. Converging with the UVA results, the ROI analysis on the extracted contrast values revealed that the bilateral IPS was more involved in processing pitch relations in song than in speech, as shown by a significant main effect of Modality [$F(1,20) = 7.471$, $p < 0.013$] and no significant interaction of Hemisphere × Modality [$F(1,20) = 0.456$, $p > 0.5$].

11.2.3 Word and Pitch Processing in Vocal Stimuli

To further explore whether there are brain regions that show stronger activation for words than for pitch patterns and vice versa, irrespective of whether presented as song or speech, two additional contrasts were defined (wpr-pr and pr-r) and compared (see Table 11.3 and Figure 11.3). The comparison of word and pitch processing [(wpr-pr)-(pr-r)] showed a stronger activation for words in the planum temporale bilaterally, and the left insula. The reverse comparison [(pr-r)-(wpr-pr)] showed activations for pitch in the planum polare of the STG bilaterally, the pars orbitalis of the right IFG (BA 47), the right premotor cortex, right SMA, left cerebellum, the left caudate and putamen, and the left parietal operculum.

11.3 Discussion

The goal of the present study was to clarify how the human brain responds to different parameters in song and speech, and to what extent the neural discrimination relies on phonological and vocalization differences in spoken and sung words and discrete and gliding pitches in speech prosody and song melody. Based on UVA and MVPA of the functional brain activity three main results were obtained: Firstly, song and speech recruited a largely

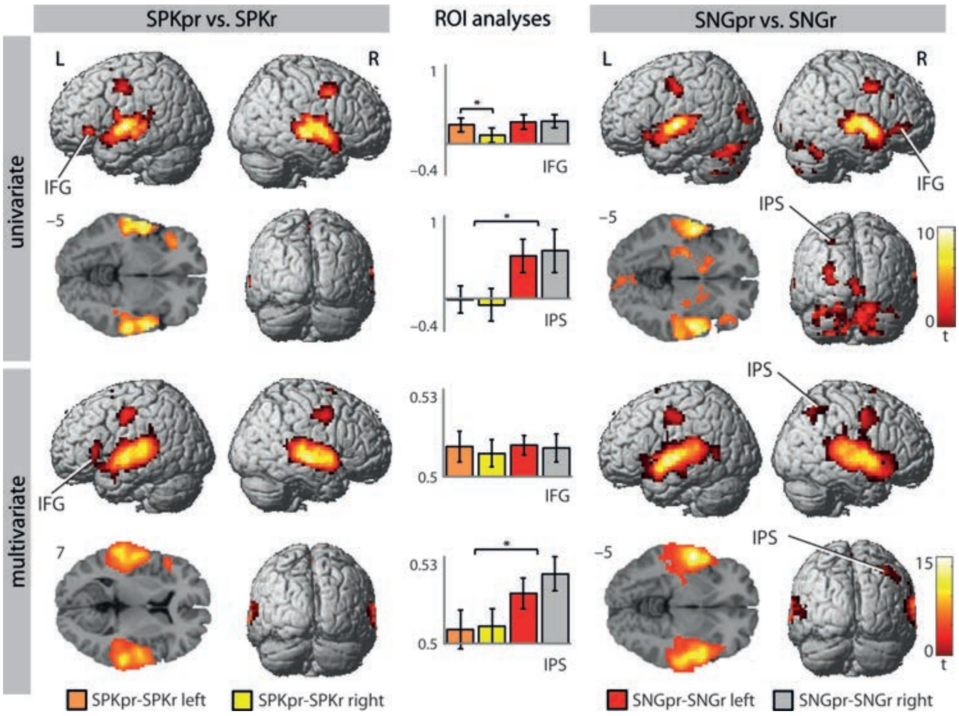

Figure 11.2: Brain regions that distinguish between pitch-rhythm patterns and rhythm in song and speech (second vs. third level: SPKpr vs. SPKr and SNGpr vs. SNGr) (p(cluster-size corrected) < 0.05 in combination with p(uncorrected) < 0.001). Bar graphs depict beta values (UVA) and accuracy values (MVPA) of the shown contrasts extracted from left and right BA 47 and the IPS. Significant results of the ROI analysis are indicated by an asterisk (p < 0.05). Color scales on the right indicate t-values for each row. IFG: inferior frontal gyrus, IPS: intraparietal sulcus.*

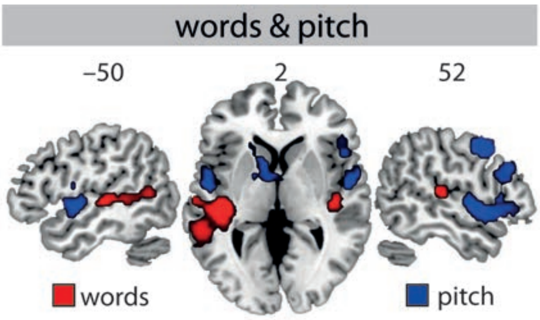

Figure 11.3: Comparison of word and pitch processing in vocal stimuli. Words-pitch (red) [(wpr-pr)-(pr-r)], pitch-words (blue) [(pr-r)-(wpr-pr)] (p(cluster-size corrected) < 0.05 in combination with p(uncorrected) < 0.001).

PITCH

Region	BA	hem	cluster	MNI coordinates			Z	cluster
			extent	x	y	z	value	p(cor)
Speech								
SPKpr > SPKr (UVA)								
STG/STS	22	L	802	-54	-4	-2	Inf	.000
Premotor cortex	6		99	-54	-7	49	5.23	.007
IFG	47		86	-36	32	-5	4,74	.014
STG/STS	22	R	993	60	2	-5	Inf	.000
Premotor cortex	6		101	54	2	43	7.03	.007
Supplementary motor area	6		60	3	5	64	4.45	.054
SPKpr vs. SPKr (MVPA)								
STG/STS	22	L	1664	-57	-10	-2	6.75	.000
IFG	45			-48	23	7	4.06	
Primary somatosensory cortex	2		181	-51	-19	46	4.82	.000
STG/STS	22	R	1512	63	-19	-5	6.93	.000
Premotor cortex	6		152	54	2	46	4.99	.000
Supplementary motor area	6		75	0	2	67	4.20	.022
Song								
SNGpr > SNGr (UVA)								
STG/STS	22	L	866	-54	-4	-2	Inf	.000
Anterior cingulate cortex	24		171	-3	-7	43	5.19	.000
Premotor cortex	6		110	-51	-7	52	4.64	.004
Lat. occip. cortex (sup. division)	18		105	-27	-82	19	4.16	.006
Parietal Lobe. Precuneus. WM			88	-21	-43	37	4.54	.013
Anterior intraparietal sulcus	(hIP3)			-24	-61	49	3.40	
Cerebellum VI lobule			1451	-27	-61	-26	6.50	.000
Cerebellum VI lobule		R		18	-70	-26	5.53	
STG/STS	22		1690	54	-7	1	Inf	.000
Caudate nucleus				18	8	10	4.75	
IFG	47			48	26	-5	4.09	
Visual cortex V1	17		113	3	-88	-2	3.86	.004
Premotor cortex	6		90	54	2	43	6.55	.011
SNGpr vs. SNGr (MVPA)								
STG/STS	22	L	2223	-57	-4	-5	7.26	.000
Primary motor cortex	4a		219	-48	-10	46	4.62	.000
Anterior cingulate cortex	24		152	-3	-10	40	4.69	.000
STG/STS	22	R	2622	57	-4	1	6.93	.000
Premotor cortex	6			54	-4	40	4.37	
Anterior intraparietal sulcus	(hIP3)		129	33	-49	52	4.27	.001
Supplementary motor area	6		82	0	2	61	3.94	.013
Conjunction SPKpr-SPKr ∩ SNGpr-SNGr								
STG/STS (Planum polare)	22	L	571	-54	-4	-2	Inf	.000
Premotor cortex (BA 6)	6		76	-51	-7	52	7.66	.023
STG/STS (Planum polare)	22	R	713	60	2	-2	7.66	.000
Premotor cortex (BA 6)	6		77	54	2	43	6.55	.022

Table 11.2: Brain areas involved in the processing of pitch patterns in song and speech. All p(cluster-size corrected) < 0.05 in combination with p(uncorrected) < 0.001).

WORDS and PITCH

Region	BA	hem	cluster extent	MNI coordinates			Z value	cluster p(cor)
				x	y	z		
words > pitch								
STG (Planum temporale)	41	L	435	-39	-34	10	6.92	.000
Insula Ig2				-39	-22	4	6.49	
STG (Planum temporale)	41			-45	-31	4	5.54	
STG (Planum temporale)	41	R	148	42	-25	10	6.98	.001
pitch > words								
STG (Planum polare)	22	L	84	-51	-4	-2	5.24	.015
Parietal operculum OP4				-51	-1	13	3.47	
Cerebellum VI lobule			95	-30	-61	-26	4.97	.009
Caudate nucleus			128	-18	-7	19	4.12	.002
Putamen				-21	-1	13	4.11	
Premotor cortex	6	R	83	54	5	43	6.56	.016
Supplementary motor area	6		237	6	5	64	6.07	.000
STG (Planum polare)	22		368	51	5	-5	5.36	.000
IFG	47			48	26	-5	4.64	

Table 11.3: Brain areas involved in the processing of words and pitch in vocal stimuli. All p(cluster-size corrected) < 0.05 in combination with p(uncorrected) < 0.001).

overlapping bilateral temporo-frontal network in which the superior temporal gyrus and the premotor cortex were found to code for differences between words and pitch independent of song and speech. Secondly, the left IFG coded for spoken words and showed dominance over the right IFG for pitch in speech, whereas an opposite lateralization was found for pitch in song. Thirdly, the IPS responded more strongly to discrete pitch relations in song compared to pitch in speech.

We will discuss the neuroanatomical findings and their functional significance in more detail below.

Inferior Frontal Gyrus. The IFG was involved with a differential hemispheric preponderance depending on whether words or melodies were presented in song or speech. The results suggest that the left IFG shows relative predominance in differentiating words and melodies in speech (compared to song) whereas the right IFG (compared to the left) shows predominance in discriminating words from melodies in song. (This effect was found in the MVPA only, demonstrating the higher sensitivity of MVPA to the differential fine-scale coding of information.) The left IFG involvement in speech most likely reflects the focused processing of segmental linguistic information, such as lexical semantics and syntax (for a review see Bookheimer, 2002; Friederici, 2002), to decode the message of the heard sentence. The right IFG involvement in song might be due to the specific way sung words are vocalized – as for example characterized by a lengthening of vowels. The right hemisphere is known to process auditory information at broader time scales than the left hemisphere (Boemio et al., 2005; Giraud et al., 2004; Poeppel et al., 2004). This may be a possible reason why the right IFG showed specific sensitivity to sung words. Alternatively, due to the non-directional nature of MVPA results, the right frontal involvement may also reflect the predominant processing of pitch in song. Although our right IFG result stands in apparent contrast to the left IFG activations observed in an UVA for sung words over vocalize by Schön et al. (2010) this discrepancy may be due to the different analysis method and stimulus material employed. Single words when they are sung as in Schön et al. (2010) may draw more attention to segmental information (e. g., meaning) and thus lead to a stronger left-hemispheric involvement than sung sentences (as used in the present study).

The processing of prosodic pitch patterns involved the left IFG (more than the right IFG), whereas melodic pitch patterns activated the right IFG (more than prosodic pitch patterns). The right IFG activation in melody processing is in line with previous results in music (Zatorre et al., 1994; Koelsch & Siebel, 2005; Schmithorst, 2005; Tillmann et al., 2006). Furthermore, this result along with the overall stronger involvement of the right IFG in pitch compared to word processing (Figure 11.3), is in keeping with the preference of the right hemisphere for processing spectral (as opposed to temporal) stimulus properties

The left-hemispheric predominance for prosodic pitch is most likely driven by the language-relatedness of the stimuli, superseding the right-hemispheric competence of processing spectral information. The lateralization of prosodic processing has been a matter of debate with evidence from functional neuroimaging for both, a left (Gandour et al., 2003, 2000; Hsieh et al., 2001; Klein & Zatorre, 2011), or a right hemisphere predominance (Gandour et al., 2003; Meyer et al., 2002, 2004; Plante et al., 2002; Wildgruber et al., 2002). Recent views suggest that the lateralization can be modulated by the function of pitch in language and task demands (Gandour et al., 2004; Kotz et al., 2003; Plante et al., 2002). For example, Gandour et al. (2004) found that pitch in tone languages was processed in left-lateralized areas when associated with semantic meaning (in native tone language speakers) and right-lateralized areas when analyzed by lower-level acoustic/auditory processes (in English speakers that were unaware of the semantic content).

Furthermore, Kotz et al. (2003) found that randomly switching between prosodic (i. e. filtered) and normal speech in an event-related paradigm led to an overall left-hemispheric predominance for processing emotional prosody, which might be due to the carry-over of a ‘speech mode’ of auditory processing to filtered speech triggered by the normal speech trials. In line with these findings, our participants may have associated the prosodic pitch patterns with normal speech in order to do the task, leading to an involvement of language-related area in the left IFG. On a more abstract level, the combined results on speech prosody and musical melody suggest that the lateralization of pitch patterns in the brain may be determined by their function (speech- or song-related) and not their form (being pitch modulations in both speech and song; Friederici, 2011).

Intraparietal Sulcus. The left and right IPS were found to play a significant role in processing musical pitch rather than prosodic pitch. The IPS has been discussed with respect to a number of functions. It is known to be specialized in spatial processing integrating visual, tactile, auditory, and/or motor processing (for a review see Grefkes & Fink, 2005). It also seems to be involved in non-spatial operations, such as manipulating working memory contents and maintaining or controlling attention (Husain & Nachev, 2007).

Related to the present study, the role of the IPS in pitch processing has attracted increasing attention. In an early study, Zatorre et al. (1994) found a bilateral activation in the inferior parietal lobe for a pitch judgment task (pitch processing) and suggested that a recoding of pitch information might be taking place during the performance of that task. More recent studies extended this interpretation, claiming that the IPS would be involved in a more general processing of pitch intervals and the transformation of auditory information. This idea is supported by the findings of Zatorre and colleagues showing an IPS involvement in the mental reversal of imagined melodies (Zatorre et al., 2009), the encoding of relative pitch by comparing transposed with simple melodies (Foster & Zatorre, 2010), as well as the categorical perception of major and minor chords (Klein & Zatorre, 2011).

While these results suggest that the IPS involvement for pitch patterns in song reflects the processing of different interval types or relative pitch per se, it remains to be explained why no similar activation was found in speech (i. e. comparing prosody against its underlying rhythm). It could be argued that the IPS is particularly involved in the processing of discrete pitches and fixed intervals typical in song, and not when perceiving gliding pitches and continuous pitch shifts as in speech. Indeed, to the best of our knowledge, no study on prosodic processing has ever reported IPS activations, eventually highlighting the IPS as one brain area that discriminates between discrete and gliding pitch as a core difference between song and speech (Fitch, 2006; Patel et al., 2008). Further evidence for this hypothesis needs to be collected in future studies.

Superior Temporal Cortex. The temporal lobe exhibited significant overlap between the processing of song and speech, at all different stimulus levels. Interestingly, however, words and pitch (irrespective of whether presented as speech or song) showed a different activation pattern in the temporal lobe. Beyond the antero-lateral STG that was jointly activated by words and pitch, activation for words extended additionally ventrally and posteriorly relative to Heschl's gyrus, and activation for pitch patterns spread medially and anteriorly.

These results are in line with processing streams for pitch described in the literature. For example, Patterson et al. (2002) described a hierarchy of pitch processing in the temporal lobe. As the processing of auditory sounds proceeded from no pitch (noise) via fixed pitch towards melody, the center of activity moved antero-laterally away from primary auditory cortex, reflecting the representation of increasingly complex pitch patterns, such as the ones employed in the present study.

Likewise, posterior temporal brain areas, in particular the planum temporale (PT), have been specifically described in the fine-grained analysis of spectro-temporally complex stimuli (Griffiths & Warren, 2002; F. Samson et al., 2011; Schönwiesner & Zatorre, 2008; Warren et al., 2005) and phonological processing in human speech (E. F. Chang et al., 2010). Accordingly, the fact that the PT in our study (location confirmed according to Westbury et al., 1999) showed stronger activation in the contrast of words over pitch for both song and speech may be due to a greater spectro-temporal complexity of the 'word'-stimulus (as grounded in e. g., the fast changing variety of high-band formants in the speech sounds) than the hummed 'pitch' stimulus.

(Pre)motor Areas. A number of brain areas that are classically associated with motor control, i. e. BA 2, 4, 6, SMA, ACC, caudate nucleus and putamen consistently showed activation in our study. This is in line with previous work showing that premotor and motor areas are not only activated in vocal production, but also in passive perception (Callan et al., 2006; Saito et al., 2006; Sammler, Baird, et al., 2010; Schön et al., 2010), the discrimination of acoustic stimuli (Brown & Martinez, 2007; Zatorre et al., 1992), processes for sub-vocal rehearsal and low-level vocal motor control (ACC; Perry et al., 1999), vocal imagery (SMA; Halpern & Zatorre, 1999), or more generally auditory-to-articulatory mapping (PMC; Brown et al., 2008; Hickok et al., 2003; Kleber et al., 2010; Wilson et al., 2004). Indeed, our participants reported that they had tried to speak or sing along with the stimuli in their head and, thus, most likely recruited a subset of the above mentioned processes.

In keeping with this, the precentral activation observed in the present study is close to the larynx-phonation area (LPA) identified by Brown et al. (2008) that is thought to mediate both vocalization and audition.

Other areas. *Cerebellum.* We also found effects in the cerebellum, another area associated with motor control (for an overview see Stoodley & Schmahmann, 2009). Apart from that, the discrimination between spoken words and prosodic pitch patterns (left crus I/VI lobe) as well as musical pitch patterns and musical rhythm (bilaterally, widely distributed, peaks in

VI lobule) in the cerebellum fits with its multiple roles in language task (bilateral lobe VI; Stoodley & Schmahmann, 2009), sensory auditory processing (especially the left lateral crus I; Petacchi et al., 2005) and motor articulation and perception and the instantiation of internal models of vocal tract articulation (VI lobe; for an overview see Callan et al., 2007).

Visual Cortex/Occipital Lobe. Activations observed in the visual cortex (BA 17, 18) seemed to be connected with processing pitch or melodic information. Previous findings support this idea, as similar regions were activated during pitch processing (Zatorre et al., 1992), listening to melodies (Foster & Zatorre, 2010; Zatorre et al., 1994), and singing production (Kleber et al., 2007; Perry et al., 1999). Note that visual prompts did not seem to be responsible, as in Perry et al. (1999) for example participants had their eyes closed, and in the current study participants followed the same visual prompts in all conditions. Following Perry et al. (1999) and Foster and Zatorre (2010), activation might be due to a mental visual imagery.

11.4 Conclusion

In summary, the subtractive hierarchy used in the study provided a further step in uncovering brain areas involved in the perception of song and speech. Apart from a considerable overlap of song- and speech-related brain areas, the IFG and IPS were identified as candidate structures involved in discriminating words and pitch patterns in song and speech. While the left IFG coded for spoken words and showed predominance over the right IFG in pitch processing in speech, the right IFG showed predominance over the left for pitch processing in song.

Furthermore, the IPS was qualified as a core area for the processing of musical (i. e. discrete) pitches and intervals as opposed to gliding pitch in speech.

Overall, the data show that subtle differences in stimulus characteristics between speech and song can be dissected and are reflected in differential brain activity, on top of a considerable overlap.[2]

11.5 Summary of Part 1 and Part 2

fMRI data on sung and spoken stimuli were analyzed in the first step with focus on the direct comparison of (1) sung and spoken sentences and (2) prosody and musical melody. Previous findings were able to be replicated by showing that sung sentences over spoken sentences activate the aSTG bilaterally. As previous studies lacked of a prosodic condition,

---

[2] End of quotation from the published manuscript.

the comparison of both melodies revealed differential patterns (MVPA) in the STG/STS and the IPS. As stated, the compared stimuli contained more than one parameter (words, pitch or rhythm), hence the information pattern found in the brain could not be defined precisely. Therefore, a second analysis used the hierarchical stimulus paradigm to demonstrate differences and similarities between song and speech on the level of word and pitch processing.

Taken together, the temporal lobe revealed overlap in song and speech (overlapping activity was found with the UVA, a discriminating pattern with the MVPA), but the direct contrasts revealed that song activates a particular region more than speech: the bilateral aSTG. Investigating the pitch and word distribution in the temporal lobe revealed that words extended posteriorly to HG while pitch extended anteriorly to HG - this could lead to the assumption that song processing is reflected in the aSTG due to its specific pitch pattern compared to speech. Maybe the aSTG reacts to the discrete pitch in music more than to the gliding pitch in speech, which needs to be clarified in further studies.

Separating song and speech into their underlying components revealed another core area: the IFG. The right IFG coded for sung words (and less strongly lateralized for musical pitch pattern) and the left IFG for prosody (and less strongly lateralized for spoken words). This rather unexpected distribution of words in song and prosody in speech lead to a general lateralization for speech to the left (only the STG/STS and the PMC was involved bilaterally) and for song to a more right-hemispheric involvement, although further areas such as the IPS and the cerebellum were activated bilaterally.

The IPS was suggested to code for the discrete pitch relations in song rather than the gliding pitch in speech, which defines the IPS as a core area in song perception. Taking the results from both analyses together, this statement needs to be discussed in more detail; especially due to its role in working memory tasks (see also Chapter 3.7). Two arguments can be held against an involvement of the IPS due to working memory load: Firstly, in the current study, task trials were unlikely to have driven the IPS activation (e. g., were excluded from the analysis, see below) and secondly, there is supporting evidence from other studies also arguing against this interpretation (e. g., Klein & Zatorre, 2011; Foster & Zatorre, 2010; discussed in Chapter 3.7).

In the current study, task trials (36 out of 272 trials in total) were excluded and only the ‘passive’ listening trials were subjected to the analysis, therefore the cognitive load during passive listening was probably low. Quantitatively, stimuli were matched in duration and number of elements, hence the quantity of items to maintain and monitor should have been equal indicating that working memory load per se might not have affected IPS level of activity. Still, as the task occurred randomly, participants expected a task trial all the time. As previously mentioned, a passive listening trial was played after which participants were told if a task trial followed or not. The task was to compare the stimulus just heard with the following one, which was excluded from the analysis. Therefore, participants might have paid selective attention to certain information, parameters or features of a stimulus in order to solve the randomly possible upcoming task. It is of note that this was intended as the task included the decision whether two stimuli belonged to the same category - song or speech. As participants might have paid attention to certain aspects of the stimuli to solve the task, the activation in the IPS could have reflected the difficulty between task trials including different stimuli. The behavioral study tested for differences in task difficulties between pairs. Firstly, pairs including a melody (melody-pair) were significantly easier and faster to match than pairs including a rhythm stimulus (rhythm-pair), but the IPS was active for musical melody over musical rhythm, therefore the IPS is not active due to the task difficulty as the rhythm-pairs were significantly more difficult to match. Secondly, the behavioral study revealed this exact same result for speech, i. e. prosody-pairs were faster and easier to match than rhythm-pairs, but the IPS was not found to be significant. Furthermore, musical melody was compared with two different conditions and the IPS was active in both contrasts: melody vs. prosody (i. e. discrete vs. gliding pitch) and melody vs. rhythm (i. e. discrete vs. no pitch), showing that the IPS was a consistent finding for discrete pitch in song even in contrast with different conditions.

An interesting side observation is that the tonal working memory (WM) in non-musicians engages (along with further areas) area PFt (rostral and ventral in the IPC; Koelsch et al., 2009; Schulze et al., 2011; see also Chapter 3.7). The peak activation observed in the current dissertation was in hIP3, which is the dorsal boarder of the IPC. Peak activations in studies investigating relative pitch (Foster & Zatorre, 2010; Klein & Zatorre, 2011) were observed in similar regions, in hIP2 (right next to hIP3) and PFm (medial in the IPC). Furthermore, the tonal WM engages the IPC more in the left hemisphere, while in the reported musical pitch-studies, the IPC/IPS was involved bilaterally (in some cases with a right-hemispheric weighting). Overall, musical pitch-tasks involve medial and dorsal parts of the IPC, including the IPS, but the tonal WM engages rostral and ventral areas. Furthermore, areas PFt and PFm were recently considered as parts of different groups in the IPC, based on receptor architectonics (Caspers et al., 2012; see Figure 3.1C): area PFt in a rostroventral group and area PFm in an intermediate group, and therefore maybe serving different functions. This speculation needs further investigations to exactly label and associate areas and function within the tonal WM network in comparison with the musical pitch processing in the reported studies.

Taken together, activation in the anterior IPS (hIP3) can generally be accounted for by working memory demands or attentional processes. While attention plays a role in a magnitude of cognitive tasks (Corbetta & Shulman, 2002) and is therefore challenging to tackle, working memory demands can be partly controlled when comparing stimuli with equal task demands – or no task, which should be investigated in future studies. In conclusion, in the current dissertation, the IPS codes for the musical pitch patterns in song – irrespective of whether the activation was driven by WM demands or attention, as no IPS involvement was observed for prosodic pitch.

To conclude, while speech mainly involves areas in the left hemisphere, song activates a bihemispheric network. While mainly the IFG and the IPS seem to discriminate between both, the involvement of the temporal lobe is not clear yet. Despite a more focal activation of the anterior STG bilaterally, the involvement of the temporal lobe overall remains unclear. A study with patients who suffer from lesions in the temporal lobe, mainly the STG, should shed light on this issue. In the following chapter a group of focal temporal lobe lesion patients as well as a group of participants with musical disorder were tested on their song and speech perception to test how depended a discrimination between both is on the temporal lobe and undisturbed music processing.

Chapter 12

Experiment 2 – Discriminating Song and Speech: A Behavioral Approach on Focal Lesions and Musical Disorder

Case studies on congenital and acquired amusia have shown a variety of music processing deficits which can be summarized in spectral and temporal processing deficits. As these two features are not unique to music, amusics have been tested on the speech domain as well. With an advantage for gliding pitch perception (Patel et al., 2008; Liu, Xub, et al., 2012), amusics showed only subtle speech perception deficits. Their pitch perception deficit seems to be most noticeable regarding the discrete pitch in music, ranging up to a semitone. A unique approach on processing deficits between music and language is to compare the vocal equivalent between language and music: speech and song. It allows comparing the specific melodic and rhythmic pattern between speech and music, while keeping text/lyrics constant.

If this leads to a behavioral experiment, it needs to be taken into account that amusics have a pitch awareness problem, and therefore that the behavioral task itself may easily introduce response biases based on subjective coping strategies. If tested with an implicit task or with imaging techniques, congenital amusics did unconsciously react to even quarter tones; reflected in early event-related potentials (EEG; Peretz et al., 2009) and a normal reaction of the auditory cortices (Hyde et al., 2011).

A damage of the temporal lobe can also lead to music processing deficits (acquired amusia), which has been shown in a variety of case studies (e. g., Peretz, 1990; Peretz et al., 1994; Peretz, 1996; Liégeois-Chauvel et al., 1998; Ayotte et al., 2000; Schuppert et al., 2000). This can be explained by the findings that the auditory cortices and adjacent association areas in the superior temporal lobes are crucial for the analysis of music – and speech (for details see Chapter 3.6).

If due to a lesion in the temporal lobe or a predisposition to congenital amusia, these essential processes, such as pitch and music processing are disturbed, the question arises, whether the prosody of speech sounds as opposed to closely matched song sounds can still be correctly assigned. Song comprises the music-specific components such as discrete pitch and regular rhythm and meter. Speech on the other hand exhibits gliding pitch and more irregular rhythmic patterns. As both are vocal stimuli, both can be performed with text, and therefore these features (regularity vs. irregularity in rhythm and meter and discrete vs. gliding pitch) are prominent acoustic cues to differentiate both. Usually, under normal hearing conditions, it is very easy to tell song from speech, but disturbed pitch processing (e. g., through lesions in the STG) should influence the perception and the classification.

Thus, the present study spotlights perception of prosody and melody in patients with lesions in the temporal lobe (TL) as well as congenital amusics (amusics) and specifically strives to investigate if their impairment (lesion and music processing deficit) will prevent a successful discrimination of both. As stated, under normal hearing conditions it is very easy to tell song from speech. In addition to such 'very easy' stimulus conditions, for the current dissertation additional stimuli were introduced that were manipulated in ways to bridge the continuum of acoustic parameters between song and speech. These additional hybrid stimuli (referred to as Ambiguous Stimuli (AMB) in the following) were cross-validated by a large participant sample in a pilot rating study in order to select ambiguous material that would be perceived as being 'halfway between' song and speech by a majority of the population. This particular 50:50 ambiguity will become important for another rationale of the study. More specifically, an interesting approach is to exploit the defining feature of the aforementioned ambiguous sounds, namely, that they are validated to equally likely evoke the responses 'song' and 'speech', respectively. Combining the ambiguous stimulus subset with a forced-choice paradigm (requiring participants to cognitively choose to perceive each of these stimuli as sung or spoken, respectively) allows investigating if patients show a bias in classifying those stimuli as song or speech.

Thus, the questions investigated with the present study are, will people with lesions in the temporal lobe and individuals with congenital amusia (1) perceptually demarcate melodic singing from prosodic speaking and (2) classify stimuli that exhibit ambiguous pitch contours into the speech or song category?

In theory, lesion patients might not be able to discriminate a gliding pitch contour (prosody) from a discrete pitch contour (sung melody), due to a very fundamental part in the brain, responsible for the computation of spectro-temporal information, being damaged. Amusics might not be able to *explicitly* classify them as song or speech. Regarding the ambiguous stimuli, healthy controls should not be able to make a decision whether they are sung or spoken. Because of the forced-choice paradigm, they should rate the same stimulus, which was repeated five times, as song or speech in an almost balanced manner. While controls may show indecisive ratings on the ambiguous stimuli, lesion patients and amusics might exhibit a bias in one or the other direction.

12.1 Methods

All together, the present study investigated song and speech perception in three groups, temporal lobe lesion patients and amusics, and one matched group of healthy controls (N = 14).

12.1.1 Participants

Temporal lobe lesion patients – lesion sites. Patients were chosen according to their lesion site and not according to functional deficits. The group comprised nine patients (4 female) with focal lesions in the left (N = 5) or right (N = 4) temporal lobe, encompassing the anterior and/or posterior superior temporal gyrus and sulcus (STG/STS) and the anterior temporal lobe (ATL; except for S. H. and A. P., whose lesions were restricted to inferior temporal gyrus, ITG, and middle temporal gyrus, MTG; see Table 12.1 and Figure 12.1). The lesions had different etiologies: ischemic stroke (N = 6), cerebral hemorrhage (N = 2) or herpes encephalitis (N = 1; see Table 12.1). Eight patients were right handed, one was left handed, according to the Edinburgh Handedness Inventory (Oldfield, 1971). They had a mean age of 55.22 years (age range: 46–67 years) and had suffered their lesion on average 4 years and 6 months before they participated in the present study (range: 1 year 9 months – 8 years 10 months). Education varied from 10 to 12 years (M = 10.89; see Table 12.2 for detailed information).

Amusics. We tested five volunteers (2 female, mean age 58 years, range 49–67, see Table 12.2) who could be classified as having pitch processing deficits based on their performance on the first three melody subtests of the Montreal Battery of Evaluation of Amusia (MBEA; Peretz et al., 2003; see below).

Matched Controls. For each TL patient and amusic (N = 14), one healthy, musically untrained control subject matched in gender, age (± 5 years), handedness (evaluated by the Edinburgh Handedness Inventory (Oldfield, 1971) and school education was recruited for the study (see Table 12.3).

Lesion sites	**Left temporal lobe lesion patients**					**Right temporal lobe lesion patients**			
	H. P.	S. E.	S. H.	A. P.	R. K.	P. B.	G. Z.	R. M.	T. M.
Etiology	IS	HE	IS	CH	IS	IS	IS	IS	CH
Lesion onset (yrs.;mon.)	7;10	6;4	4;6	2;5	4;7	5;6	1;10	1;9	8;10
TL lesions (BA)	21,22	20,21,22, 38	20,21	20,21	21,22,38	21,22,38	21,22,42	21,22,38, 41,42	22,41,42
Additional lesions	39,40	28,34,35, 36, insula, amygdala				11,47, insula		insula	39,40

Table 12.1: Description of the lesions for each individual patient. Etiology: IS = ischemic stroke, CH = cerebral hemorrhage, HE = herpes encephalitis. The average time between lesion onset and the experimental session is indicated in years;months. Lesions are listed in Brodmann Areas (BA). The extent of the lesion was determined by MRI scans.

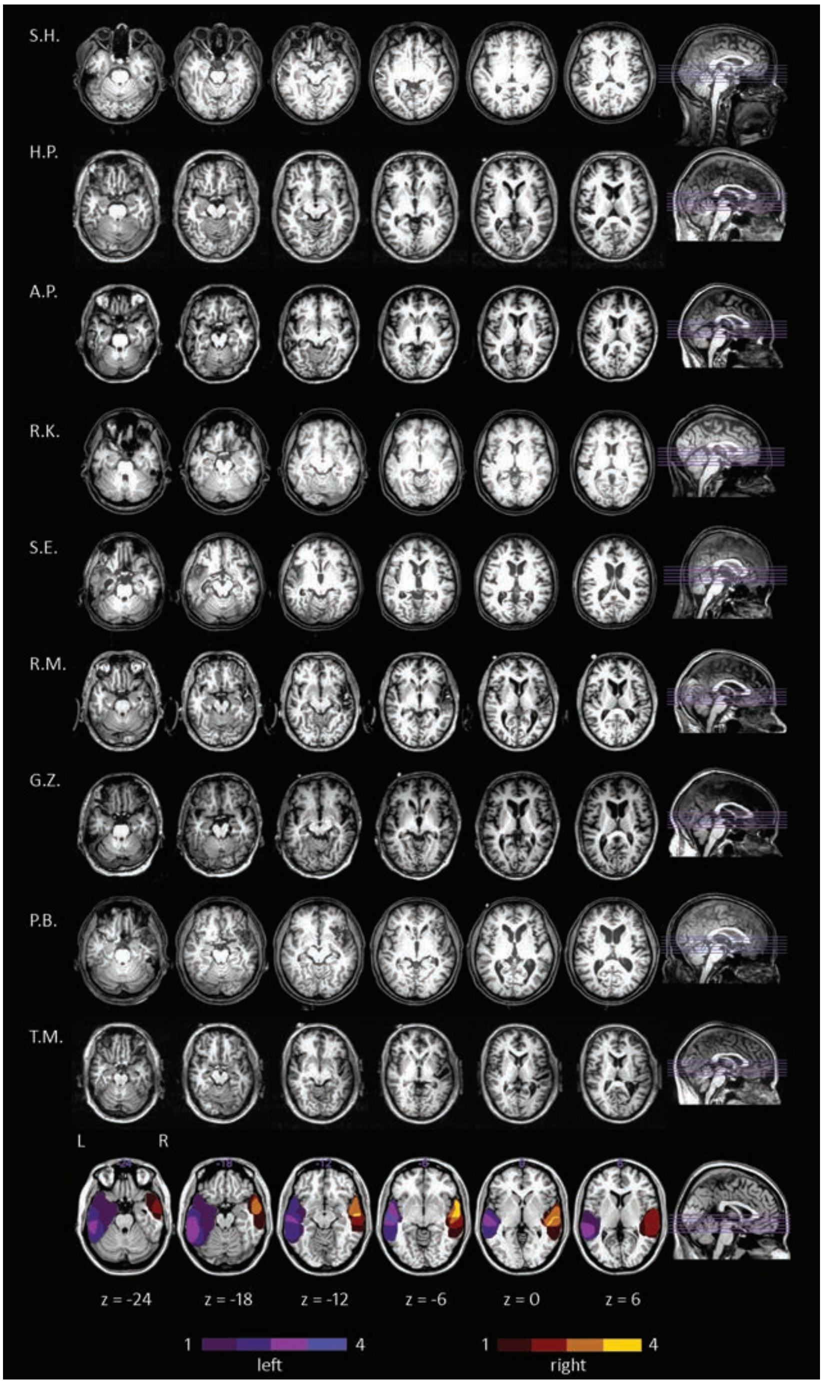

Figure 12.1: MRI scans of the temporal lobe lesion group. The top five rows depict the left temporal lobe lesion patients, the next four depict the right temporal lobe lesions patients. Bottom row: overlay of the temporal lobe lesions of all patients in MNI space. Lesions are color coded; left sided TL lesions in blue-purple, right sided TL lesions in red-yellow.

	Temporal lobe lesion patients									**Amusics**				
	P. B.	G. Z.	R. M.	T. M.	R. K.	A. P.	H. P.	S. E.	S. H.	D. E.	R. K.	R. S.	T. F.	R. K.
Age	67	67	57	50	46	49	59	51	51	67	49	67	57	51
Gender	M	M	M	M	F	F	M	F	F	M	F	M	M	F
Handedness	L	R	R	R	R	R	R	R	R	R	A	A	R	R
School education	12	12	10	10	10	10	12	12	10	10	10	12	12	10
Token Test	0	0	0	11	7	0	1	1	0	0	0	0	0	0
MMSE	29	29	29	24	24	25	25	25	29	27	30	29	27	29
Digit span	95/85	28/78	88/67	**0/2**	**1/2**	98/52	**13/2**	53/34	98/34	28/27	95/52	98/85	98/80	95/34
MBEA Scale	20	22	16	21	24	28	23	18	30	23	20	18	18	16
MBEA Contour	22	18	19	25	16	23	22	17	28	21	23	21	18	17
MBEA Interval	14	14	21	24	12	18	24	18	27	15	20	19	16	16
MBEA Total	**56**	**54**	**56**	70	**52**	69	69	**53**	85	**59**	**63**	**58**	**52**	**49**
Sing-Speak Rating	.65	.92	.94	.70	.55	.71	.71	.48	.88	.94	.94	.81	.67	.83

Table 12.2: Personal and neuropsychological data of the patients. Gender: F = female, M = male. Handedness is indicated according to the Edinburgh Handedness Inventory (Oldfield, 1971). R = right handed, A = ambidextrous. School education is indicated in years. The severity of language comprehension deficits is indicated by the number of mistakes in the Token Test: no/very mild disorder (0–6); mild (7–21); medium (22–40); severe (> 40). Mini-Mental State Examination (MMSE): mild cognitive impairment (< 26); moderate (< 18); severe (< 10). Verbal STM and WM were tested with the digit span test forward and backward. Results (age corrected) are indicated as percentages (digit span forward/backward). Values below 16 (bold font) indicate a deficient memory performance. The scores achieved in the MBEA are depicted as number of correct responses. In the MBEA Total, bold font indicates scores under the cut-off of 65 points (Peretz et al., 2003; Liu et al., 2010, 2012), i. e. marks a deficient performance.

	Controls W. H.	D. H.	V. D.	M. R.	I. S.	C. G.	R. K.	P. M.	C. V.	A. K.	G. D.	R. D.	B. W.	S. P.
Age	65	70	60	51	46	44	59	51	52	65	47	69	61	50
Gender	M	M	M	M	F	F	M	F	F	M	F	M	M	F
Handedness	L	R	R	R	R	R	R	R	R	R	R	R	R	R
School education	12	12	10	10	10	10	12	12	10	10	10	12	12	10
Token Test	0	0	0	0	2	1	0	0	0	2	0	0	0	5
MMSE	27	27	29	30	30	29	29	29	29	28	28	27	23	22
Digit span	53/27	76/93	67/38	100/71	53/**2**	75/**12**	48/**5**	53/52	53/71	76/**13**	69/**1**	76/53	67/67	53/52
MBEA Scale	26	28	24	30	28	28	26	29	26	26	29	26	29	28
MBEA Contour	26	26	24	26	24	23	22	25	24	23	27	24	29	24
MBEA Interval	24	28	21	23	26	23	23	25	23	23	26	22	23	27
MBEA Total	76	82	69	79	78	74	71	79	73	72	82	72	81	79

Table 12.3: Personal and neuropsychological data of the healthy controls. Participants are arranged in the same order as the patients. Gender: F = female, M = male. Handedness is indicated according to the Edinburgh Handedness Inventory (Oldfield, 1971). R = right handed. School education is indicated in years. The severity of language comprehension deficits is indicated by the number of mistakes in the Token Test: no/very mild disorder (0–6). Mini-Mental State Examination (MMSE): mild cognitive impairment (< 26). Verbal STM and WM were tested with the digit span test forward and backward. Results (age corrected) are indicated as percentages (digit span forward/backward). Values below 16 (bold font) indicate a deficient memory performance. The scores achieved in the MBEA are depicted as number of correct responses. In the MBEA Total, scores above the cut-off of 65 points (Peretz et al., 2003; Liu et al., 2010, 2012) indicate unimpaired music processing.

Hearing Abilities

TL, amusics and controls. All participants had normal hearing which was tested with the HTTS Audiometry by SAX GmbH (http://www.sax-gmbh.de/htts/httsmain.htm). HTTS is a program for performing a hearing test (audiometry) on a multimedia PC. It tests both ears independently, uses a logarithmic frequency scale and pitches are presented in a randomized order.

Language Comprehension

Language comprehension deficits were assessed by the Token Test, a subtest of the Aachen Aphasia Test (Huber et al., 1993, see Table 12.2 and Table 12.3), to make sure that all participants were able to understand the instructions.

TL. Patients showed mostly very mild, only one TL patient mild, or no language comprehension deficits at the time of testing.
Amusics. None of the participants showed language comprehension deficits.
Controls. Controls showed only very mild or no language comprehension deficits.

Cognitive Abilities

To evaluate the participants' mental state, the Mini-Mental State Examination (MMSE; Folstein et al., 1975) was conducted. Short term (STM) and working memory (WM) capacities were assessed with a verbal digit span forward (STM) and backward (WM) in all patients (part of the Wechsler Memory Scale, WMS; Wechsler, 1987). Raw data were age corrected and converted into the percentile equivalents.

TL. All participants were included in the study, as no participant performed lower than 24 out of 30 (<26: mild cognitive impairment). Two patients showed deficits in the STM and the WM task (R. K. and T. M.; see Table 12.2) and one a deficit in the WM task (H. P.). Note that there was no correlation between working and short-term memory capacities and the MBEA results ($p > .441$).

Amusics. No participant scored under 27 points in the MMSE which indicate no cognitive impairment. Also, none showed STM or WM deficits (see Table 12.2).

Controls. In the MMSE two candidates showed a mild cognitive impairment (B. W. and S. P.; see Table 12.3). STM was intact in all participants, even though three showed deficits in the WM task (I. S., R. K., G. D.). To compare the memory performance between the patients and the controls, two-sample *t*-tests were calculated. Neither STM nor WM differed significantly from the performance of the TL patients and the amusics ($p > 0.3$; see Table 12.4). Accordingly, differences in behavioral performance cannot be attributed to a different STM or WM capacity of patients and controls.

Music Perception

The Montreal Battery of Evaluation of Amusia (MBEA; Peretz et al., 2003) was carried out to detect music processing deficits in the participants. The test is based on the model by Peretz and Coltheart (2003; described in Chapter 2.1.1) and is targeted at musically untrained adult individuals.

According to the model, musical input must be processed along the melodic and the temporal dimension. As both dimensions contain various modules, the battery consists of 6 musical tests: three to assess pitch processing: (1) scale, (2) contour, (3) interval; two tests to assess temporal processing: (4) rhythm, (5) meter; and a last test to assess memory components, (6) incidentally learning new melodies. During the pitch subtests, candidates listen to pairs of short melodies and indicate whether a pair is the same or different. Consisting of 30 melody pairs per test, 15 pairs are identical, 15 pairs include a second melody in which one tone has been changed. These violations are different in each of the pitch subtests. In the scale test, one tone is modified to be out of scale. While the first test retains the original melodic contour, the second test violates the contour through changing the pitch direction of the surrounding intervals, while maintaining the original key. In the third, the interval test, one interval has been changed while maintaining contour (the pitch direction) and scale.

As the focus of this study was on the pitch differences in song and speech while the rhythm in the stimulus material was kept strictly constant over the conditions, only the three pitch-related subtests of the MBEA were performed. Subtests on rhythm, meter and incidentally learning novel melodies were not of interest for the purpose of this study. The authors of the MBEA provide test results of 160 normal participants of variable age and education. To evaluate the MBEA data of the three pitch-based subtests, a composite score was calculated using 65 as a cut-off score (Peretz et al., 2003; Liu et al., 2010; Liu, Xub, et al., 2012). This is the sum of the three first subtests, based on the test results of 160 normal participants of variable age and education serving as a reference sample; scale subtest (1) cut-off 22 / mean 27 / SD 2.3; contour subtest (2) cut-off 22 / mean 27 / SD 2.2; interval subtest (3) cut-off 21 / mean 26 / SD 2.4 (Peretz et al., 2003).

TL. Five TL patients (two with left- and three with right-hemispheric lesions) scored under the 65 cut-off score and could be classified as having a musical disability regarding pitch perception. Four patients scored normally, above the 65 cut-off score (see Figure 12.2 and Table 12.2).

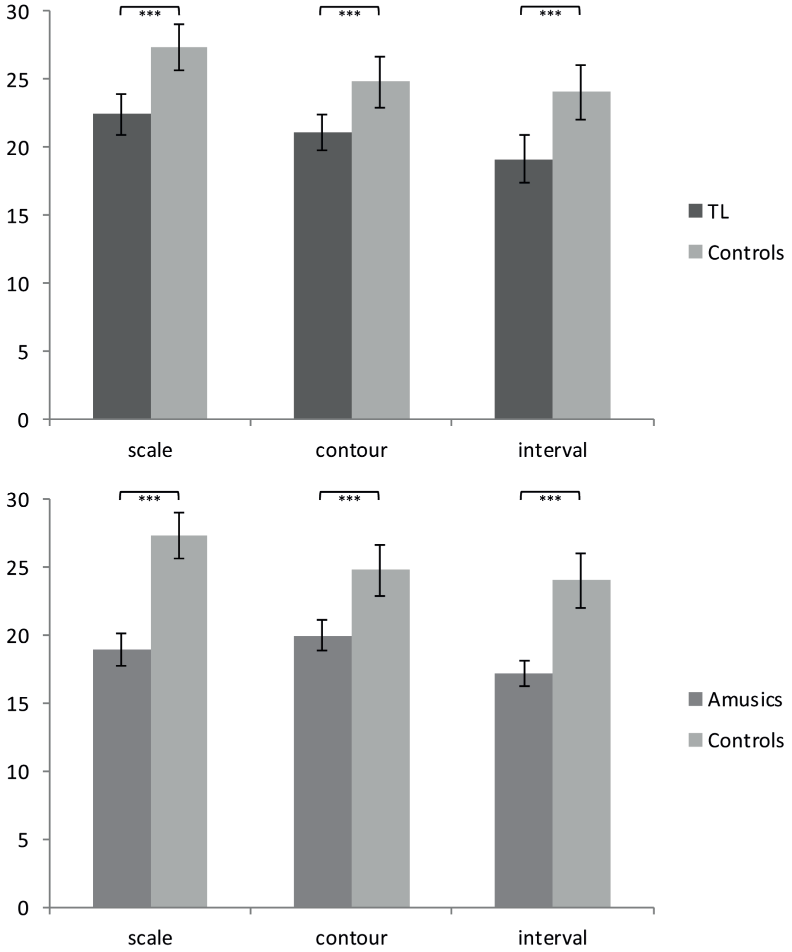

Figure 12.2: Results for the three pitch subtests of the MBEA for all groups. Highest possible score was 30. TL lesion patients and amusics performed significantly different than the controls in all three subtests.

Amusics. All five participants scored below the 65 cut-off score in the first three subtests and therefore could be classified as having pitch amusia (see Figure 12.2).

Controls. In all MBEA subtests, the control group performed above the 65 cut-off score, indicating unimpaired music processing. In order to statistically compare the performance of patients and controls, two-sample t-tests were calculated for each subtest. These analyses showed that the TL patient group scored significantly lower in all three pitch subtests of the MBEA than the controls. Participants in the amusic group were classified as having pitch amusia with the first three subtests of the MBEA, and as expected, they scored significantly lower in all three pitch subtests of the MBEA than the controls (see Figure 12.2 and Table 12.3 and 12.4).

12.1.2 Stimulus Material and Procedure

Stimuli. The general aim of the present study was to focus on the typical pitch patterns of music and speech, respectively, while keeping other putatively distinguishing parameters constant across conditions, namely, temporal (rhythm, meter, and timing) and spectral features (like a Singer's Formant as normally present in trained operatic singing), and accentuation (stress). Both sung and spoken stimulus types should contain linguistic text (lyrics; no hummed conditions). In order to avoid familiarity and transfer effects (like obligatory melody imagery when listening to a spoken line taken from song lyrics), no lyrics or melodies from well-known songs were used.

Sentences, Melodies and Prosody. Ten simple German sentences were used for stimulus recordings. They featured identical grammatical structure and were semantically straight-forward everyday life statements. Emotional valence of the sentences was obtained in a previous rating study (Kotz & Paulmann, 2007) and only neutral sentences were used. The number of syllables was held constant in all sentences. In order to make rhythm and meter of the sung stimuli similar to the linguistic stress pattern of the spoken stimuli, and vice versa, sentence structure was devised such that the accents of the pattern when spoken resembled the downbeats of a ternary meter when sung (see Figure 12.3E). For the sung conditions, four different melodies were generated along the interval transition probabilities typical of Western tonal music (Dowling, 1968; see Figure 12.3). They consisted of (A) a monotonous pitch contour, (B) a melody that mimics a neutral spoken prosody, (C) a typical classical music cadence, and (D) a melody consisting of fifths.

Voice Recordings, Stimulus Evaluation and Selection for the Study.[1] Two native speakers of German (one male, one female, both trained in professional voice as speakers as well as singers) produced all sentences. Sentences were recorded with a PC soundcard and digitized at 16-bit/44.1 kHz sampling rate. The singers were instructed to produce a wide variety of vocal timbres, rhythm variations, and intonations throughout the recording session, for sung (ranging from pop to operatic style) as well as for spoken sentences (ranging

---

[1] The stimuli and the results from the rating study were provided by Marc Bangert.
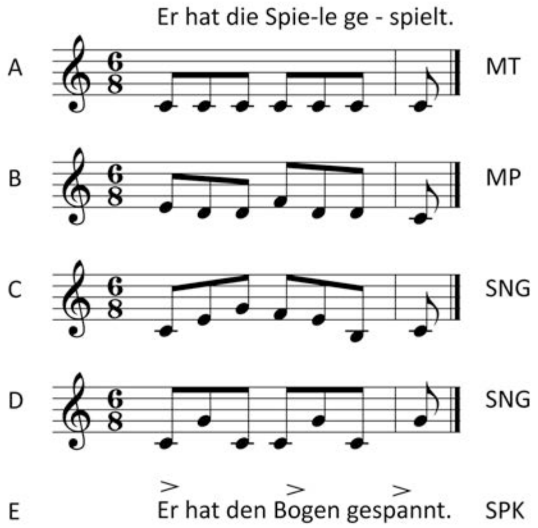
Figure 12.3: Stimulus material. Four different melodies were created and consisted of (A) a monotonous pitch contour = MT, (B) a melody that mimics a neutral spoken prosody = MP, (C) a typical classical music cadence, and (D) a melody consisting of fifths, and (E) neutrally spoken sentences. Rhythm and meter were ternary for both, sung (SNG) and spoken (SPK) stimuli. Therefore, rhythm and meter differences cannot be attributed for differences in song and speech perception.

from everyday pronunciation to highly articulate recitation). This approach was chosen in order to generate subsets of stimuli, which could be clearly distinguished as speech or song, respectively, but also to obtain a fair number of perceptually ambiguous vocalizations. All recordings were cut into individual sentences using custom-developed Matlab scripts; the cuts were then digitally normalized and individually re-sampled to be of 3-second duration exactly.

The resulting stimulus database consisted of 674 full sentences, from which 40 stimuli were selected for the experiment in an extensive rating evaluation (pilot study, not reported in detail): An observer group of 62 participants rated the entire stimulus pool on a 10-point visual-analog scale where 0 was Speech, 9 was Song, and in-between a number of intermediate steps for the more ambiguous or unclear vocalizations. Based on the total of 29617 ratings, we consecutively selected the 10 different sentences with the lowest average rating for the ‘speech’ condition (SPK), the 10 sentences with the highest ratings for the ‘song’ condition, and the 20 sentences closest to the center of the scale for the ambiguous condition (AMB). A standard deviation < 1.5 on the rating scale was used as additional criterion (this was to make sure a center-scale average was due to the fact that most participants actually agree on the ambiguity of the stimulus, rather than just having a group average effect when half rate it song and the other half rate it speech). Two different sets of 10 stimuli for the AMB stimuli were chosen. The stimuli were taken out of the middle of the rating pool which consisted of stimuli based on melodies A and B (see Figure 12.3). This lead to two different kinds of ambiguous stimulus sets: one based on melody A with a monotonous (MT), and one based on melody B with a mimicked prosodic pitch contour ('mimicked prosody'; MP). Stimuli exhibiting a marked Singer's Formant (peak ~3000 Hz) were excluded from the selection.

Paradigm and Procedure. The stimulus compilation outlined above yielded four conditions – Speech (SPK), Ambiguous (AMB) – divided into Monotonous (MT) and Mimicked prosody (MP) – and Song (SNG). Within each condition, 10 stimuli were presented five times in a randomized order, totaling 200 stimulus presentations during the experiment. Participants were seated in front of a computer. For presentation and recording of responses, a custom-developed flash animation was used. The acoustical stimuli were presented binaurally via headphones. The presentation level of the stimuli was adjusted to a comfortable level. After each stimulus presentation, participants were asked to decide whether the stimulus was 'song' or 'speech' by a forced-2-choice button decision. The respective prompts were presented as visual cues ('0' for speech, '1' for song).

Of interest for the statistical analysis was (1) if participants were able to correctly respond to clear song and spoken stimuli and (2) if they showed a bias to rate for ambiguous stimuli in a distinctive way (more song or more speech). While controls may show indecisive ratings, lesion patients and amusics might exhibit a bias in one or the other direction. All statistics were done with SPSS Statistics 17.0.

12.2 Results

12.2.1 Temporal Lobe Lesion Patients

Clear sung and spoken stimuli. All participants (controls, TL patients and amusics) were able to correctly classify between the clear sung and spoken stimuli, indicated by 99.8 % correct responses.

Ambiguous stimuli. Due to the forced-choice nature of the paradigm, the AMB stimuli revealed individual response biases of the participants, which in the control group were almost balanced as eight participants rated toward speaking (ranging from 0.21 to 0.48) and six toward singing (0.63 to 0.9). A one-sample t-test with a test value of 0.5 revealed no tendency in the controls' rating ($p > .845$).

Most of the TL patients showed a response bias toward singing (one-sample t-test (0.5): $t(8) = 4.242$, $p < .003$) , i. e. rated the ambiguous stimuli more likely as song than speech

(0.48 to 0.94). Only two of them were indecisive in rating for song or speech (S. E., 0.48 and R. K., 0.55). The difference between the rating of TL and controls became significant as evaluated by a two-sample t-test [$t(21) = 2.713$, $p < .013$]. Also, both groups performed differently for the MT [$t(21) = 2.546$, $p < .019$] and MP [$t(21) = 2.258$, $p < .035$] stimuli (see Figure 12.4 and Table 12.4).

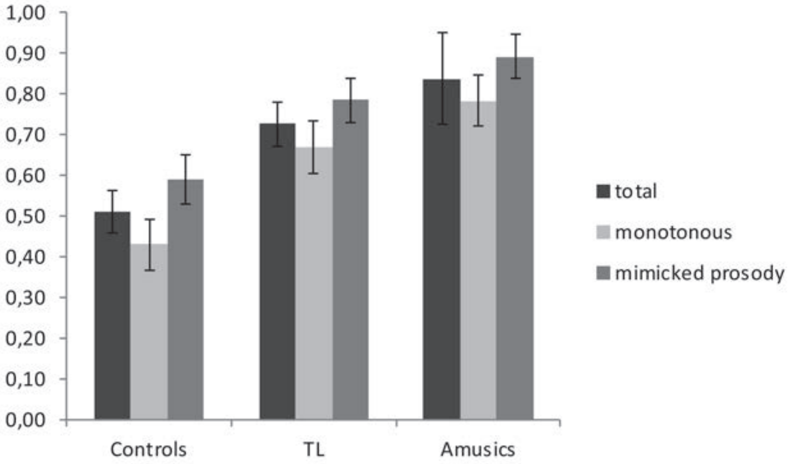

Figure 12.4: Results for the AMB stimuli of the Sing-Speak Rating for all groups. Ratings above 0.5 were considered as toward singing, ratings under 0.5 toward speaking and indecisive ratings occurred around 0.5.

Effects within group. In the control group, a significant difference between the monotonous and mimicked prosody stimuli [$t(13) = -2.8$, $p < .015$] could be found. The MT stimuli were rated on average with 0.43 and the MP with 0.59, which means that the MT stimuli were more rated toward speaking. Within the TL group this difference did not become significant ($p > .064$) but a tendency for the same effect could be seen.

Sing-Speak Rating and MBEA. Significant correlations between the Sing-Speak Rating and the MBEA were found between the first MBEA subtest, the 'scale' subtest, and the AMB stimuli (Pearson's $r = -.377$, $p < .048$). The higher participants scored in the MBEA scale test, i. e. the better their musical perception, the less they were biased toward singing (see Figure 12.5). Neither the MBEA contour ($p > .273$) nor the MBEA interval subtest ($p > .318$) correlated with the Sing-Speak Rating.

Left and right TL. The difference between the left (N = 5) and right (N = 4) TL lesion patients was assessed by a non-parametric Mann-Whitney-U test which revealed no signifi-

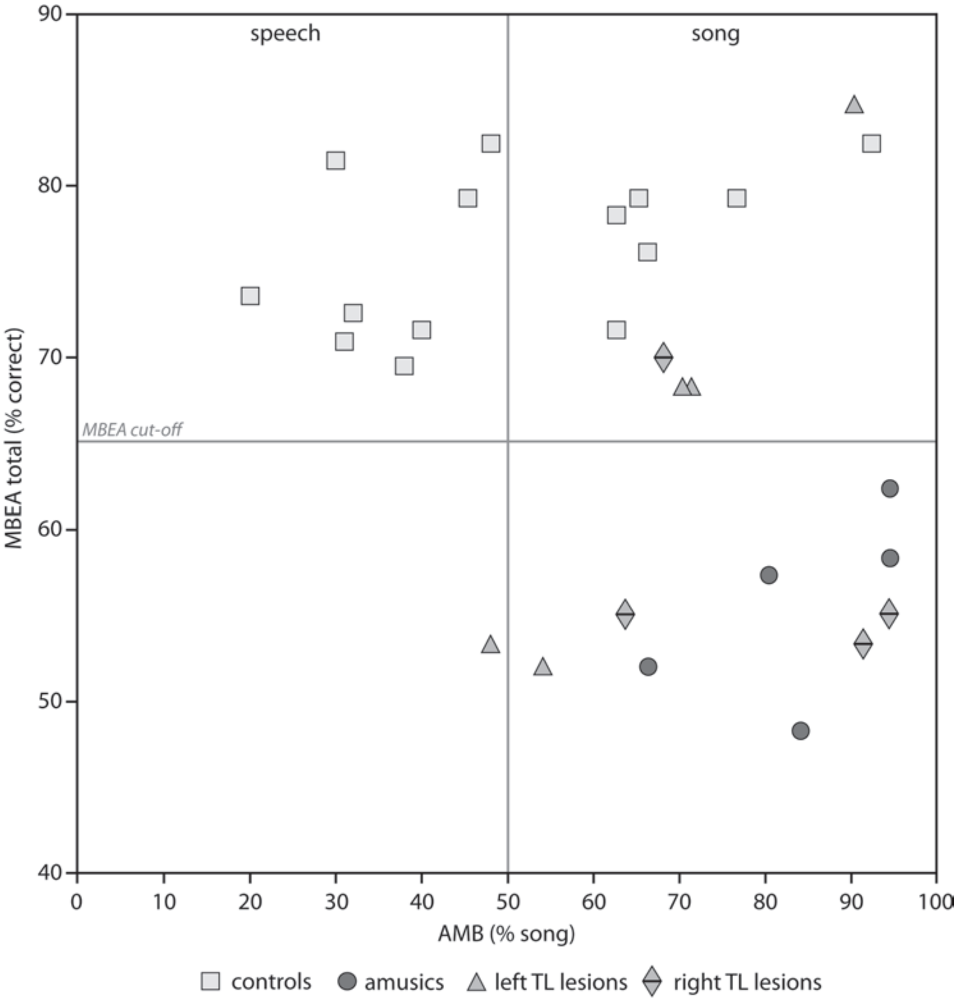

Figure 12.5: Performance of all participants of the MBEA total and the Sing-Speak Rating of ambiguous (AMB) stimuli. Squares = controls; triangle = left temporal lobe lesion patients; diamond = right temporal lobe lesion patients; circles = amusics. Number of AMB stimuli in percentage, rated as song. The MBEA cut-off was at 65 % and a lower performance classified participants as having pitch amusia. Note that four TL lesion patients did not exhibit pitch amusia.

cant difference in the Sing-Speak Rating between both groups ($p > .325$). Interestingly, the two highest 'singing' ratings came from two right lesion patients (G. Z. 0.92 and R. M. 0.94) and the two lowest, in this case 'indecisive', ratings came from two left lesion patients (S. E. 0.48 and R. K. 0.55). As all four could be classified as having pitch amusia, it is interesting to mention that none of them rated toward speaking.

12.2.2 Amusics

Ambiguous stimuli. The pitch amusia group exhibited an unequivocal bias to classifying the AMB stimuli as singing (one-sample t-test (0.5): $t(4) = 6.768$, $p < .002$). None of these participants displayed a tendency to rate for speaking (ranging from 0.67 to 0.94).

The difference between the AMB stimuli ratings of amusics and controls became significant as evaluated by a two-sample t-test [$t(18) = 3{,}421$, $p < .003$]. Also, both groups rated the MT [$t(18) = 3.151$, $p < .006$] and MP [$t(17) = 2.874$, $p < .011$] stimuli differently (see Figure 12.4 and Table 12.4).

Effects within group. A difference between MT and MP stimuli did not become significant within the amusic group ($p > .151$).

Between TL patients and amusics. Amusics and TL did not differ in their Sing-Speak Rating ($p > .196$; see Table 12.4 for details).

Test	**TL vs. Controls**		**Amusics vs. Contr.**		**TL vs. Amusics**	
	$t(21)$	p-value	$t(17)$	p-value	$t(12)$	p-value
Sing-Speak Rating Total	2.713	< .013	3.421	< .003	-1.368	> .196
Monotonous (MT)	2.546	< .019	3.151	< .006	-1.164	> .267
Mimicked prosody (MP)	2.258	< .035	2.874	< .011	-1.274	> .227
Digit span forward	-1.107	> .281	1.512	> .149		
Digit span backward	-.016	> .987	.935	> .363		
MBEA Scale	-3.78	< .001	-8.122	< .000		
MBEA Contour	-3.068	< .006	-4.522	< .000		
MBEA Interval	-3.248	< .004	-6.488	< .000		
MBEA Total	-4.083	< .001	-8.190	< .000		

Table 12.4: Results of the two-sample t-tests comparing the cognitive abilities, music perception and Sing-Speak Rating results between patients and controls. The digit span forward and backward assess STM and WM functions. Note that the performance of patients and controls did not differ in these basic tests. TL patients and amusics performed significantly lower in all MBEA subtests than the controls. This pattern remains the same with Bonferroni correction.

12.3 Discussion

SNG and SPK. Despite the lesions and the pitch processing deficits, none of the participants showed difficulties at all in classifying sung and spoken sentences. As the performance was nearly perfect, the assumption is that the test had a ceiling effect. Probably, the sung and spoken sentences used in the current study were too distinct and therefore the rating did not tease out the TLs' and amusics' actual processing problems. As an alternative explanation, a unilateral lesion (as in the tested patients) might not have been sufficient to disrupt the discrimination of song and speech, as this function most likely involves the temporal lobes bilaterally (i. e. prosody and pitch processing; e. g., Zatorre & Belin, 2001; Zatorre & Samson, 1991; Zatorre et al., 2002; Meyer et al., 2002; Tzourio et al., 1997) and the contralesional temporal lobe could have taken over to classify song and speech. Moreover, the clarity of the song and speech stimuli might have allowed a classification at even lower levels of the auditory processing hierarchy. Before an auditory stimulus reaches the (primary) auditory cortex and the STG, it is processed in the nuclei of the auditory brainstem and the thalamus, where various temporal (e. g., Griffiths et al., 2001) and spectral (e. g., Musacchia et al., 2007) properties of the signal are analyzed. Furthermore, in the assumption of song being the stimulus that arouses more than neutral speech, emotional cues play a role. Koelsch and Siebel (2005) stated that emotional processing can take place during all stages of music perception, even already at the level of the thalamus which is directly linked to major components of the emotion circuit (e. g., the amygdala and the medial orbitofrontal cortex).

To conclude, neither TL patients nor amusics had difficulties in classifying song and speech stimuli, which can be accounted for by the distinct and unambiguous stimuli. Therefore, the ambiguous stimulus set gave further insights into processing problems of both groups.

Ambiguous Stimuli. Surprisingly, if presented with the categorical decision that the stimuli are either spoken or sung, respectively, music processing deficits appear to lead subjects to classify unclear and unusual vocal stimuli as being sung. Apparently, their response behavior does not reflect their actual pitch processing mechanisms, but rather a cognitive strategy possibly based on an over-compensatory coping bias - with normal speech constituting the perceptual null hypothesis in everyday life. After all, speech is our daily communicative basis, while listening to song is a special event.

Within the control group, the whole rating spectrum of AMB stimuli was covered (0.21–0.94) which underlined the interindividual differences in rating AMB stimuli toward song or speech. Still, controls agreed on rating both AMB stimulus sets differently: the monotonous stimuli were more rated toward speaking and the mimicked prosody stimuli with a tendency toward singing. The difference between both stimulus sets was the pitch contour (words, rhythm and stress/beat were the same). This leads to the conclusion that less melody in a stimulus leads healthy participant to rate toward speaking and even a mimicked prosodic pitch contour leads toward singing. This, again, reflects every day experience: Speech can occur in a more monotonous manner, but song shows typically wider pitch variations.

This effect could not be seen in both patient groups. Both patient groups rated all of the AMB stimuli more or less toward singing (none below 0.48). Note that even the stimuli without pitch contour (monotonous) were rated as singing. One could assume that they would rate the AMB stimuli as speaking as their spectral/pitch processing malfunctions in explicit tasks (e. g., Ayotte et al., 2002; Peretz et al., 2002) and therefore, the specificity of song, the discrete pitch contour, might not serve as a feature for them to recognize the stimuli as song.

This finding suggests that pitch processing deficits lead participants to overcompensate their ratings toward singing. They might strongly rely on a template of proper speech and rate anything that does not sound like speech – including the less familiar ambiguous stimuli – as singing. Under this viewpoint, the rating response behavior of the TL patients and amusics reflects the use of a specific strategy that is indirectly related to their actual pitch processing deficit. Notably, this also implies the necessity of a most careful approach for all amusia studies involving melody tasks, as the behavioral task itself may easily introduce response biases based on subjective coping strategies.

MBEA and Sing-Speak Rating. The lower participants scored in the MBEA scale test, the more they rated toward singing. The scale subtest especially detects severe problems in music perception, as the violations in this task are extremely obvious – the changed pitch sounds out of tune. Participants who scored low in this task were obviously unable to detect significant changes in a music piece. The results indicate that rating for singing stands in strong connection with music processing deficits.

Interestingly, four of the TLs did not show a music processing deficit, but rated for song, too. This could be explained by interindividual differences in perceiving AMB stimuli, as has been shown in the control group, which covered the whole range from singing to speaking. The TL patients – in contrast to amusics – showed a similar pattern to the controls in at least having some indecisive ratings. Also, three of the four patients scored near the cut-off point (69 and 70) which underlines the described connection between music processing deficits and perceiving AMB stimuli as song. Further investigations are necessary on a wider population range of patients on the temporal lobe lesion spectrum.

Brain Structures Involved in Music and Song Processing. The temporal lobe lesion group encompassed four patients who did not show musical impairments (non-amusics) and five patients who could be classified as having pitch-amusia (amusics). Therefore it is worth contrasting the lesion profiles, as the difference in their patterns of neural damage could help suggest the involvement or lack of specific neural regions in music processing (with regard to pitch and melody). Three observations are conspicuous: (1) left as well as right TL lesion patients showed acquired pitch-amusia (left: 2 out of 5, right: 3 out of 4), (2) all five amusics had lesions in the anterior STG (aSTG), and four of the five had a lesion in BA 38, the temporal pole (two right and two left lesion patients), which had none of the non-amusics; (3) two of the four non-amusics had lesions in ITG and MTG, not in the STG (BA 22) as all the other patients, and could not be classified as having pitch amusia.

Note that an advantage of the current study is that the acquired amusics showed rather focal lesions, in contrast to other reported lesion studies where patients showed extensive lesions in the TL, extending in other e. g., frontal areas (for details, see Chapter 5.2). The observations in the current study could lead to the assumption that the STG bilaterally and the temporal pole bilaterally are crucial for pitch processing in music, which is in line with the literature. Regions in anterior and posterior STG (from Heschl's gyrus; HG) were found to be involved in the analysis of simple pitch sequences - instead of single pitches (e. g., Griffiths et al., 1998; Krumhansl, 2004; Patterson et al., 2002; Zatorre, 1988). The importance of the anterior temporal lobe (ATL) in music processing has also been shown in other lesion studies (Ayotte et al., 2000; Peretz, 1990; Liégeois-Chauvel et al., 1998; Schuppert et al., 2000). Music processing deficits have been shown after unilateral (S. Samson & Zatorre, 1988; Zatorre, 1985) and bilateral ATL lesions (Satoh et al., 2005; for details on music processing and the TL see Chapter 3; for details on acquired amusia see Chapter 5.2). BA 38 was also found to be involved in music-related processes (but not for speech or language; Brown et al., 2004, 2006; Schön et al., 2010), furthermore in the improvisation and creation of novel melodies, which was interpreted as being related to an even higher level of musical processing than BA 22 (Brown et al., 2006), corresponding to Griffiths et al. (1998) who proposed BA 38 (temporal pole) to be involved in processing complex musical patterns.

Most important for the current examination is a very recent finding by Tierney et al. (2012) who found that the aSTG bilaterally was activated when speech was perceived as song. Participants in an fMRI study were presented with naturally occurring 'boundary cases' between song and speech (taken from audiobooks) where a spoken phrase sounded as if it was sung when isolated and repeated. They conclude that song processing needs an increased demand on pitch processing and audio-motor integration. Together with the results from the current study, the aSTG seems not only to be important in music perception but to be especially important for song perception. Further evidence for the important role of the aSTG in song processing came from song and speech perception studies (Callan et al., 2006; Schön et al., 2010 and the current fMRI study, see Chapter 10 and 11).

Regarding the left hemisphere, Ayotte et al. (2000) demonstrated that left lesion patients after middle cerebral artery infarct showed impaired musical long-term memory.

Unfortunately, controversial findings in lesion studies regarding the involvement of the ATL in music/melody processing have been reported. S. Samson and Zatorre (1988), for example, showed melodic processing deficits when (1) the right ATL is damaged, and also when (2) the left ATL and the HG are damaged at the same time, but (3) not the left ATL only. Also, the right ATL has been shown to have little impact on music discrimination and memory (Liégeois-Chauvel et al., 1998). Ayotte et al. (2000) on the other hand reported cases with music processing deficits and right temporal pole damage. One could argue that the HG plays a crucial role and it undoubtedly does, but a lesion here is not mandatory to exhibit music processing deficits, as some of the individual cases did show impairments with intact HG (see Table 12.5). As a further explanation, the above mentioned findings rely on MRI scans from the years of 1988 to 2000, when the resolution of the images was very low. Further research with higher resolution images might help solving the divergence.

In Chapter 5.2, specific single cases of acquired amusia, previously reported in the literature, were discussed and it could be shown that the aSTG was damaged in all but one case. Extending this observation by adding the above mentioned cases, the whole picture of 12 patients with acquired amusia shows only one case with musical deficits and without damage in the aSTG (H. V., Griffiths et al., 1997; see Figure 12.1 and Table 12.5, current case studies are added). Despite patient G. Z., four patients from the current study displayed an additional lesion in the temporal pole. To conclude, the current study provides further evidence for the importance in musical melody and song processing of (1) the bilateral STG and aSTG (BA 22) and (2) the temporal pole (BA 38) - here demonstrated unilateral in right and also the left hemisphere (see Figure 12.6).

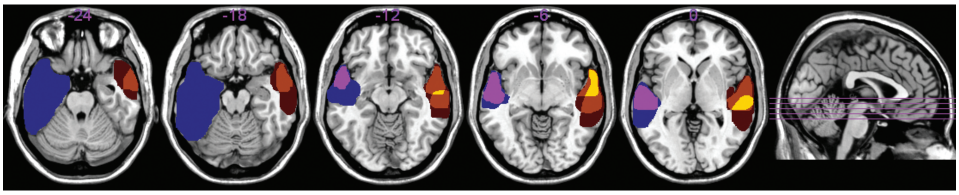

Figure 12.6: Overlay of the temporal lobe lesions of patients suffering from acquired amusia from the current study. Lesions are color coded; left sided TL lesions in blue-purple, right sided TL lesions in red-yellow. All patients show lesions in aSTG (BA 22) and all despite one in the temporal pole (BA 38).

study	Peretz 1994	e. g., Peretz 1997; Patel 1998	Peretz 1994, 1996; Patel, 1998	Griffiths 1997	Ayotte 2000		Satoh 2005	Merrill 2012				
patient	G. L.	I. R.	C. N.	H. V.	N. R.	R. C.	N. N.	P. B.	G. Z.	R. M.	S. E.	R. K.
STG	R/L	R/L	R/L	R	R	R	R/L	R	R	R	L	L
aSTG	R/L	R	R/L		R	R	R/L	R	R	R	L	L
pSTG/PT		L		R	R	R						
HG	L	L			R	R			R	R		
MTG	L	L			R		R/L	R	R	R	L	L
insula	R/L	R/L			R	R	L					
IPC	L	L		R	R							
frontal	R/L	R/L			R							

Table 12.5: Lesion overview of single cases with acquired amusia after brain damage. The cases from the present study are presented in the five columns on the far right, titled 'Merrill 2012'. R = right hemisphere; L = left hemisphere; STG = superior temporal gyrus; aSTG = anterior STG; pSTG = posterior STG; PT = planum temporale; HG = Heschl's Gyrus; MTG = middle temporal gyrus; IPC = inferior parietal cortex.

12.4 Conclusion

Two groups were tested on their song and speech discrimination ability with a forced-choice paradigm: one patient group with lesions in the temporal lobe (TL) from which half exhibited music processing deficits as evaluated by the MBEA, and a second group with congenital amusia; both in comparison with a matched control group.

Despite the lesions and the music processing deficits, none of the participants showed difficulties in classifying distinct and unambiguous sung and spoken sentences. Either a ceiling effect or different compensating processes could account for the almost perfect performance.

Classifying ambiguous stimuli into song and speech categories though, exhibited a surprising response bias: both, the TL and the amusics rated unclear and unusual vocal stimuli as *song*, while the controls' responses covered the whole rating spectrum – they rated AMB stimuli as song or speech in an almost balanced manner.

Three assumptions can be made following these results: Firstly, rating AMB vocal stimuli is an intraindividual process as has been shown by the controls' rating behavior. Secondly, TL lesions and music processing deficits lead participants to overcompensate their ratings toward singing. Therefore, their response behavior does *not* reflect their actual processing deficit, but rather a cognitive strategy to solve the task. Thirdly, TL and amusics seem to need clear sung and spoken stimuli to identify them as song or speech. Their tolerance of what is speech might be smaller than of individuals with normal music perception abilities.

Interestingly, the patients exhibiting music deficits revealed the aSTG and the temporal pole as a lesion overlap. Together with single case studies in the literature and findings from neuroimaging studies (also the current one), the aSTG bilaterally can be defined as a core area in music and most probably – *song* processing.

Chapter 13

General Discussion

The relationship between language and music is a well investigated field and partially overlapping networks for processing language and music have been suggested. The current work investigated both domains where they are closest to each other - song and speech - and aimed to map the neural underpinnings of our capacity to discriminate song from speech. The current work took two main approaches to unravel and compare the neural networks underlying the perception of sung and spoken stimuli: (1) the identification of relevant brain areas by means of a passive listening paradigm applied to healthy participants in a neuroimaging fMRI study, and (2) the specific assessment of temporal lobe involvement by means of a behavioral classification of song and speech in temporal lobe lesion patients and amusic individuals.

Experiment 1 aimed to gain further insights into the neural basis of song and speech perception by means of a novel hierarchical stimulus paradigm. The experiment comprised a behavioral pretest (1A) and an fMRI study (1B). The latter was analyzed with two different goals: (Part 1) the direct comparison of sung and spoken stimuli without paying attention to underlying parameters (i. e. words, melody and rhythm) and (Part 2) the perception of words and pitch patterns in song and speech while parceling out the other parameters.

The stimuli used in Experiment 1 constitute six conditions and were organized in a subtractive hierarchy with three levels: the first level comprised sung and spoken utterances with (the same) text, the second level excluded the text and only contained the underlying pitch-rhythm patterns (melodies) of the sung and spoken utterances, the third level further excluded pitch information and merely contained the underlying rhythm patterns. The use of stimuli which contained only the spoken prosody and the musical melody in a hummed manner (second level) allowed for a comparison of melodic and prosodic pitch patterns between both domains. The results of the pretest (Experiment 1A) confirmed the use of the stimuli in the fMRI experiment: the stimuli were rated as neutral and slightly arousing; the rhythm stimuli were only used as a control condition; the task was suitable for the scanner as evidenced by testing with background scanner noise. The rhythm condition was implemented as a control condition to control for the differential temporal information underlying musical melodies (regular beat) and prosody (less regular). This allowed for the investigation of the pure pitch contour and interval processing during song and speech perception.

Experiment 2 aimed at gaining further insights into the specific role of the temporal lobe as well as the influence of music processing deficits on song and speech classification and discrimination. In a behavioral experiment, patients and amusics were tested on their subjective classification of sung and spoken stimuli, which exhibited unambiguous and ambiguous characteristics. This experiment was conducted with patients displaying focal lesions in the temporal lobe and a group of congenital amusics (music disability) as well as a control group. The hypothesis was that ambiguous stimuli should lead to an indecisive classification. Lesion patients and amusics might show a bias in classifying ambiguous stimuli toward song or toward speech.

The stimuli used in Experiment 2 comprised (1) distinct and unambiguous spoken and sung utterances (with text) and (2) stimuli that were manipulated to bridge the continuum of acoustic parameters between song and speech (with text/lyrics). This special set of stimuli was perceived as being ‘halfway between’ song and speech (validated by a rating study). This circumstance allowed for the following approach: combining these ambiguous stimuli with a forced choice paradigm required participants to cognitively choose to perceive the stimuli as sung or spoken.

Perception of Song and Speech in the Healthy Brain

The findings of Experiment 1B Part 1 indicate that sung and spoken utterances involve the superior temporal gyrus and sulcus (STG/STS) bilaterally, the right premotor cortex (PMC) and the supplementary motor area (SMA). Song (with lyrics) engages these areas somewhat stronger than speech – with the anterior STG (aSTG) bilaterally as a distinct area. Prosodic and melodic pitch patterns recruit in addition to a bilateral STG/STS involvement the anterior intraparietal sulcus (IPS), which distinguishes between both melodic stimuli. These results lack the information about which parameter in song and speech the IPS and the aSTG specifically code for (e. g., words, melody, rhythm). A further approach using the hierarchy of the stimuli shed light on this issue.

The results of Experiment 1B Part 2 demonstrate that perceiving words and pitch patterns in song and speech recruit shared and distinct areas in the brain: the left inferior frontal gyrus (IFG) codes for spoken words and shows predominance over the right IFG for pitch in speech, whereas an opposite lateralization indicates word and pitch processing in song. Regarding pitch in song and speech, the IPS responds more strongly to the pitch patterns in song compared to speech. The STG and the PMC code for words and pitch independent of song and speech. A distinct pattern for words and pitch patterns is observed in the temporal lobe: activation for words extends more posteriorly and pitch more anteriorly to Heschl's gyrus (HG), most probably reflecting the complex pitch patterns in aSTG (Patterson et al., 2002) and the spectro-temporal complexity (Griffiths & Warren, 2002).

Taken together, the results from Part 1 and 2 demonstrate that the perception of song and speech involves temporo-frontal networks (see Figure 13.1), with the auditory cortices and motor-related regions bilaterally demonstrating overlap and the aSTG, the IFG and the IPS as areas that are differentially recruited during song and speech perception.

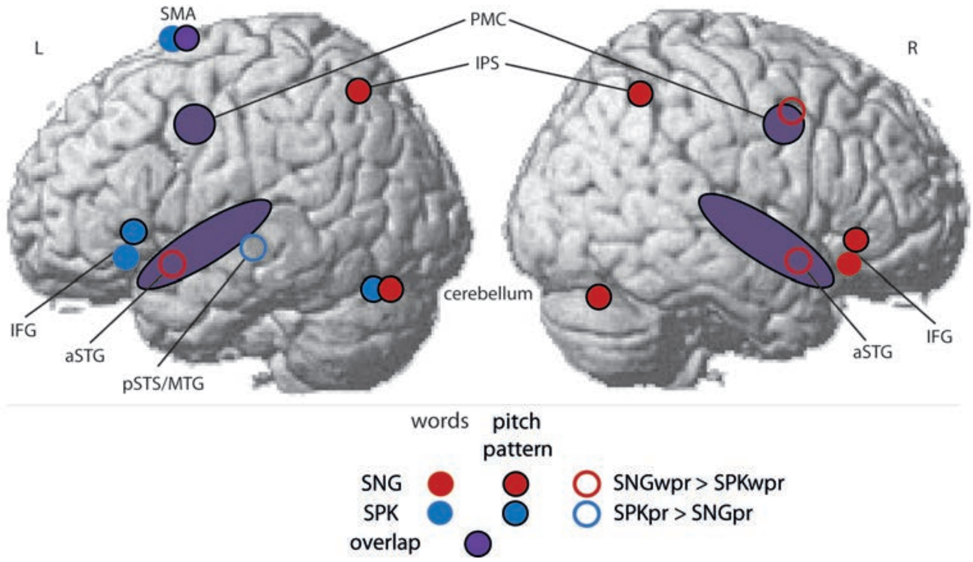

Figure 13.1: Model of the song (red) and speech (blue) networks. Legend below. Speech engages a left-hemispheric temporo-frontal network with the IFG coding for words and pitch patterns and the pSTS/MTG for prosody as core areas. Song engages a broader distributed bihemispheric network with the right IFG coding for words and pitch patterns, the IPS for the musical pitch patterns, and the aSTG coding for song in general (most likely due to its pitch pattern) as core areas. In the STG/STS, the PMC bilaterally and the SMA, overlap was suggested by the univariate analyses and a differential involvement by the multivariate analyses. SNG = sung; SPK = spoken; wpr = words, pitch, rhythm; pr = pitch, rhythm; IFG = inferior frontal gyrus; aSTG = anterior superior temporal gyrus; pSTS/MTG = posterior superior temporal sulcus/middle temporal gyrus; SMA = supplementary motor area; PMC = premotor cortex; IPS = intraparietal sulcus.

These networks are differentially lateralized: for speech to the left and for song stronger to the right. This is in line with the literature, where a lateralization of song to the right and speech to the left was found (Riecker et al., 2000; Jeffries et al., 2003; Callan et al., 2006) and suggested (Perry et al., 1999; Brown et al., 2006; Schön et al., 2010). It is interesting that authors speculate about a different lateralization of song and speech when (1) only singing was tested without a speaking condition (Perry et al., 1999; Brown et al., 2004) or (2) no lateralization effect between song and speech was tested (Schön et al., 2010; Brown et al., 2006) and (3) the speech over song contrast did not yield significant results (Wildgruber et al., 1996; Callan et al., 2006; Schön et al., 2010, in perception; Formby et al., 1989; Özdemir et al., 2006; Gunji et al., 2007, in production). Looking closely at the stimulus composition of song and speech it becomes obvious why a clear lateralization, especially for song, might not be evident: song, as well as speech, contains linguistic information in the form of lyrics, which is mostly expected to be processed in the left hemisphere. Furthermore, speech, as well as song, has a pitch pattern (prosody), which was found to be processed in the right hemisphere. However, prosodic pitch can be shifted to the left hemisphere due to task demands or semantic information (e. g., in tone languages). Some studies suggested that the 'musical aspect' (e. g., the spectral information) causes dominant processing in the right hemisphere for song, and the linguistic aspect causes dominant processing in the left hemisphere for speech. Only a few studies contrasted song with vocalized stimulus material – with contradicting results: slightly stronger in the left in perception (Schön et al., 2010) and bilaterally in production (Özdemir et al., 2006).

In the current dissertation, words in song are processed significantly stronger in the right IFG than the left, while words in speech engage the left IFG. This result might reflect a different way of vocalizing sung words (i. e. lengthening of the vowels) as the right hemisphere is known to process information at broader time scales than the left (Giraud et al., 2004; Poeppel et al., 2004; Boemio et al., 2005). Interestingly, as a results of *missing* linguistic segmental information, an area in the left posterior STS/middle temporal gyrus (MTG) was found for prosody compared to melody, possibly reflecting the participants' attempt to add words in order to make out the meaning in the prosodic stimulus. This is supported by studies using sine-wave speech, demonstrating a similar effect (e. g., Davis & Johnsrude, 2007). This process is not necessary in hummed melody as this is an independent stimulus.

Pitch patterns in song and speech also recruit areas in opposite hemispheres: pitch processing in speech engages the left IFG, pitch in song the right IFG. The lateralization for prosodic pitch is most probably determined by the *function* of pitch (and not its *form*, i. e. being pitch modulations) as pitch in tone languages is processed in left-lateralized areas when associated with semantics (Gandour et al., 2004).

Taken together, traditionally, song and speech are suggested to be processed in opposite hemispheres, but so are linguistic and melodic information. In the context of song and speech, these features are shown to involve the hemispheres according to the presumption about music and language distribution in the brain: speech in the left and song more in the right hemisphere.

In contrast to a clear lateralization in the IFG, the STG/STS and the PMC (as well as the medial SMA) are involved bilaterally in song and speech. While speech does not engage further areas in the right hemisphere, song engages additional areas bilaterally: the aSTG (in direct contrast with speech) and the IPS as well as the cerebellum for sung pitch patterns. Therefore, the findings in the current dissertation confirm a bihemispheric network for song (Brown et al., 2004; Özdemir et al., 2006; Gunji et al., 2007).

It is of note that the observation of song engaging a larger, more bihemispheric network than speech poses an obstacle for defining a complete speech network based on the difference to song. An observation of the speech network in contrast to song is overtaken by a usually stronger and bihemsipheric activation for song. The *direct* comparison of spoken over sung phrases has not yet revealed meaningful results in perception (Callan et al., 2006; Schön et al., 2010, and the current study). One can assume that song is the richer stimulus (acoustically and cognitively) and engages the brain more than speech. Therefore, the reduction of stimulus complexity achieved with the current paradigm is a first step in observing areas for speech over song, such as the left IFG for word and pitch pattern processing in speech.

An unexpected finding is that the IPS specifically responds to the pitch patterns in song and might therefore represent one of the core areas in song classification. The current dissertation discussed possible influences on working memory demands and selective attention processes which might have driven the activation. Speaking against WM demands is firstly that task trials were excluded from the analysis and secondly that the sung melodies were the easiest condition to match with other stimuli, revealed by a pilot testing. Therefore, a higher working memory load due to task difficulty was not obvious. On the other hand, participants might have paid selective attention to certain stimulus features in order to perform the upcoming task; in this case, it might be easier to pay attention to the discrete pitch in the sung stimulus than the gliding pitch in speech where no IPS involvement is observed. In conclusion, the IPS is only engaged in melodic pitch patterns – irrespective of whether the activation is driven by working memory demands or selective attention.

A novel, but intriguing interpretation for the IPS involvement was suggested by studies on musical pitch processing (e. g., Zatorre et al., 1994, 2009, 2010; Foster & Zatorre, 2010; Klein & Zatorre, 2011). The idea from visual studies of the role of the IPS in the systematic transformation of "any stimulus representation depending upon precise relationships among its elements" (Foster & Zatorre, 2010, p. 1357) was suggested, as the IPS is involved in perceiving transposed melodies (Foster & Zatorre, 2010), the categorical perception of major and minor chords (Klein & Zatorre, 2011) and currently discrete pitch relations – all requiring relative pitch encoding which is a fundamental element in music perception. More specifically, the IPS might be sensitive to the relationship between discrete pitches in music and not the gliding pitch in speech.

The specific involvement of parts of the temporal lobe remains open from the current analyses, which reveal overlap as well as a distinct pattern for song and speech. Overlap is suggested on a macroscopic level, but it remains unclear to what extent on a microscopic level the STG/STS (and the PMC) are involved in song and speech perception.

To conclude, a song and a speech network are proposed with overlap in the bilateral STG/STS, PMC, SMA and cerebellum, suggesting a shared sensorimotor network (see also Özdemir et al., 2006). Song involves a bihemispheric network with a right-hemispheric weighting, whereas speech involves a left-hemispheric network. This reflects the music-relatedness of song and the language-relatedness of speech on all levels: word and pitch processing. The left IFG (which detects words and prosodic pitch patterns in speech), the right IFG (which detects the lyrics and the melodies in song) and the IPS (which detects the musical pitch patterns) are responsible for the discrimination.

As it remains unclear from the current fMRI study how exactly the temporal lobe is involved in song and speech perception, the following study tested if lesions in the temporal lobe disrupt song and speech processing. Furthermore, the influence on musical ability was of interest as the left and right IFG involvement seemed to be caused by the function (being song or speech related) of the specific stimulus.

Perception of Song and Speech in the Musical Deficient Brain

In Experiment 2, the subjective classification of song and speech stimuli was tested behaviorally in patients with different etiologies: Two groups with either lesions in the temporal lobe (TL) or congenital amusia (amusics; without lesions) were tested. Patients with focal lesions in the temporal lobe were selected to further explore the function of the temporal lobe in song and speech processing. Individuals with music disabilities were selected based on their performance in the Montreal Battery of Evaluation of Amusia (MBEA; Peretz et al., 2003) in order to further investigate the 'musical' influence on song and speech discrimination. Both groups exhibited music perception deficits: all congenital amusics and half of the TL patients with left and right-hemispheric lesions in mainly the STG (acquired amusics). This proved the importance of the temporal lobe in music processing – but how it influences song and speech perception needed to be shown.

Both groups, TL and amusics, have no problems in classifying distinct and unambiguous stimuli as song or speech but show a response bias to classifying ambiguous stimuli as song. In contrast, healthy controls exhibit ratings in a balanced manner for speech and song. This probably reflects the use of a cognitive strategy that is indirectly related to their actual music processing deficit. Moreover, the acquired amusics (in line with cases reported in the literature, e. g., Peretz et al., 1994; Ayotte et al., 2000) revealed that the aSTG as well as the temporal pole (BA 38) was impaired in most of the cases and can therefore be taken as a crucial area in song processing (as also evident by the fMRI study, Experiment 1B Part 1).

The TL patients, with lesions in either the left or the right hemisphere, show a song-response bias. Despite their intact music perception (although at the lowest limit of the normal range), their perception of song and speech is different to healthy controls. This means that the lesion in the temporal lobe (regardless of the exact location) changes the perception of song and speech – in which direction though needs to be clarified. Therefore it is interesting that a music perception impairment (without lesions in the temporal lobe) leads to the same response pattern – toward song. This was the case in the acquired and the congenital amusics.

A lesion in the left or right temporal lobe and amusic symptoms altered the participants' perception of song and speech. Since previous studies with amusics have suggested amusia to be an impairment of the music domain only and speech to be unaffected (e. g., Peretz et al., 2012), a two-step mechanism might be assumed: the impaired music perception affects the amusics' song perception, possibly lowering their tolerance for what is considered speech in comparison to song perception in individuals with unimpaired music perception abilities. As a consequence, as soon as stimuli show a slight discrepancy from their mental model of speech in an utterance, the forced-choice task would lead amusics to classify these stimuli as 'non-speech', i. e. song. From the current experiment, this can only be speculated, as no complete speech-song continuum was used, where degrees of song and speech likeliness would be observed. A follow-up study could use a parametrical design. Such a design would probably also prevent the ceiling effect shown in the unambiguous stimulus rating. These results, combined with the results from Experiment 1, provide further insights in the neuroanatomical underpinnings of song and speech perception.

In the fMRI study it was assumed that the STG/STS was engaged differently in song and speech processing with the aSTG as a distinct area, coding for song (confirmed by Callan et al., 2006; Schön et al., 2010, and when speech was perceived as song, Tierney et al., 2012). Furthermore, song engaged the brain more strongly than speech (involving a bihemispheric network). Therefore, the differential patterns in the STG/STS bilaterally were accounted for by song processing. The acquired amusics exhibited a lesion overlap in the aSTG, which can be associated with the music processing deficit (confirmed by previous single case studies, also showing a consistent lesion in the aSTG). Additionally, the activation was found bilaterally and the aSTG lesion was found in the right and the left hemisphere, therefore it is confirmed by lesion studies that song is represented in the aSTG bilaterally – compared to speech.

The congenital amusics in this study were not tested for lesions, but in previous studies, congenital amusic exhibited gray and white matter anomalies in the right auditory and inferior frontal cortices and a reduced white matter connection between these regions (Hyde et al., 2006, 2007, 2011). This is in line with the finding in the current fMRI study where the right IFG has been found coding for song.

Taken together, TL and amusics are known to exhibit anomalies in brain areas which coded in the fMRI study for song: the aSTG bilaterally and the right IFG. As the TL and amusics' classification of ambiguous song and speech stimuli exhibited a different response bias toward song, this might reflect their disturbed song perception – in a way that they made their decision based on an over-compensatory strategy.

13.1 Summary

This dissertation exceeds the state-of-the-art research in song and speech by investigating song and speech perception on the levels of words, pitch patterns and rhythm, which lead to the following picture of song and speech networks: song engages a right-hemispheric temporo-frontal and bihemispheric temporo-parietal network, speech engages a left-hemispheric temporo-frontal network. Three regions code specifically for song: the bilateral aSTG, the IPS and the right IFG, while the left IFG specifically codes for speech. Right and left temporal lobe lesions as well as impaired music perception lead to a disturbed perception of song and speech, underlining the crucial role of an intact STG, specifically the aSTG, in song and speech perception.

13.2 Future Perspectives

Following the discussion on the data from the experiments presented in this dissertation, some further ideas resulting from the studies will now be reported.

Words in song. The major difference between song and speech on the linguistic level is the vowel length. In song, vowels carry the sound (discrete pitch) and remain longer in a stable position. In speech, sounds move more continuously. The fMRI study revealed a distinct neural pattern for words in speech compared to song in the left IFG – but not for words in song compared to speech. The lyrics in song showed predominance in the right IFG over the left, but did not become significant in contrast to speech – challenging the hypothesis that lyrics in song would also engage the left over the right IFG, reflecting segmental information processing. This issue remains unresolved – one explanation is that it was due to the specific way words in song are vocalized, i. e. characterized by a lengthening of the vowels. As the right hemisphere processes information at broader time scales (Giraud et al., 2004; Poeppel et al., 2004; Boemio et al., 2005), an fMRI-experiment with vocalized words where the vowels are stretched to different degrees, might yield more fine-grained insights. This would also be informative with respect to the Melodic Intonation Therapy, which treats non-fluent aphasics in their speaking ability – in one training step, words are articulated more slowly and therefore vowels remain in a stable position longer. An fMRI study would give further insights on this matter, especially regarding the lateralization of vowels in different articulation tempos. Once detected at which time point (which vowel length) the right hemisphere (STG and IFG) comes into play, the tempo of articulation in the MIT could be used more exactly and would treat most efficiently.

Furthermore, as a bihemisheric network for song was shown in production *and* perception, before a treatment with the MIT, a patient needs to be tested on his discrimination ability between song and speech first to test for unimpaired perception of song and speech – if impaired, the treatment might not be successful, which has been shown in some cases searching for explanations.

A patients study could also be informative regarding the relations between words and pitch patterns in song an speech. Patients with right IFG lesions could be tested on their song and speech perception, with special regard to the underlying words and pitch patterns.

Gliding and discrete pitch. These two pitch features constitute a main difference between song and speech and have been shown to be differentiated by the IPS. Amusics have often been tested with the comparison of pitch glides and discrete pitch – only two tones that varied in their transition to the next (e. g., Foxton et al., 2004; Liu, Xub, et al., 2012). An fMRI study with healthy participants on these simple tone relations would be informative, not only to investigate the fine-grained involvement of the IPS. A task could be included asking for the direction (was the second tone higher/lower). Furthermore, statements and questions could be included as they only differ in the pitch direction of the last tone/syllable (see also amusic studies, e. g., Patel et al., 2008; Liu, Jiang, et al., 2012). Also, glides and discrete pitches could be presented in comparison with the above mentioned syllables (‘Words in song’) – pitch contour in comparison with vowel length would comprise main acoustic differences between song and speech. These parameters varied step by step and presented in a parametric design, or with four conditions, presented in a 2×2 design with factors vowel (short/long) and pitch (discrete/glide) should give further insights into these fundamental constituents of music and language.

Anterior Superior Temporal Gyrus and Sulcus. The STG/STS is neither language nor music specific but may rather be involved in the processing of complex auditory features common to linguistic and musical sounds. Nonetheless, studies found the temporal pole (BA 38) to be active for song/music, but not for speech (Schön et al., 2010; Griffiths et al., 1998; Brown et al., 2004, and the current study). Presumably, different parts of the STG/STS respond differently to song and speech (so far only the aSTG has been defined as song-specific). Reducing the stimuli's complexity down to shorter and more precise sequences, for example using the above suggested stimuli on pitch type (glide and discrete) and vowel duration (to test for articulatory differences), might reveal a more fine-grained insight into the processes within the STG/STS.

A song and speech continuum. Different experimental opportunities derive from a song-speech continuum: For example, (1) the intelligibility decreases and the prosody 'increases' (becomes more important), (2) the vocal sound turns into an instrumental sound (see also Specht et al., 2009, who morphed stimuli from white noise to either music or speech sounds) and (3) the pitch tier (F0 frequency of a pitch pattern) of a spoken sentence becomes discrete when only pitch maxima, minima and averaged pitches are used. This would lead to a parametric design.

For example, Patel et al. (1998) used a technique for converting intonation patterns to tone analogs through extracting the median F0 (pitch tier) of each syllable (for detailed instructions see Patel et al., 1998; Liu et al., 2010). This approach could be widened by directing this more toward the music or speech domain: having longer (discrete) pitches for music and shorter and gliding pitches for speech. To bring speech closer to the music domain, these melodies could be sung on a syllable, hummed or replaced with an instrumental sound (strings or woods). This would lead to the idea of a song-speech continuum or music-speech continuum - this depends if vocal or instrumental sounds are used. One problem might be that the extracted pitch tier becomes such an abstract quality that it is no longer associated with speech. Therefore, the syllable/tone onset has to be quite clear in combination with dynamic variability to allow segmentation and grouping of the sounds.

Amusics and song and speech perception. As a follow-up study, amusics could be tested on their song and speech perception on a lower stimulus level: using syllables and testing syllable length, pitch height and interval distance with the categorical question of song or speech. Preliminary data from our group (Felber, 2009, diploma thesis) compared the perceptual behavior of non-musicians and singers. While non-musicians rated syllables as speech when they had a length between 200-400 ms, singers rated them toward speech only in between 100-200 ms – everything longer than that was rated as song. Both groups exhibited that with increasing interval distance syllables were rated less and less as speaking; up to a major second they were more likely considered as spoken and an interval-distance of a fourth turned out to be a good demarcation point for the differentiation of song and speech. This was also rated differently between the groups, up to a minor third, non-musicians revealed a rating response behavior more toward speaking than the singers. Taking these differences between non-musicians and singers as a baseline of musical expertise, it would be interesting to see how amusics would rate syllable length and pitch height regarding song and speech.

Appendix A

Appendix

A.1 Stimulus Material

Figure A.1 shows the sheet music of the 6 melodies with lyrics (Figure A.1A) used in the fMRI study (Experiment 1B, Part 1 and 2). Two additional melodies with lyrics (Figure A.1B) were used for training purposes in the behavioral (Experiment 1A) and the fMRI study (Experiment 1B).

A.2 CD Index

Track 1	Experiment 1. SNGwpr (sung sentence)
Track 2	Experiment 1. SNGpr (song melody)
Track 3	Experiment 1. SNGr (song rhythm)
Track 4	Experiment 1. SPKwpr (spoken sentence)
Track 5	Experiment 1. SPKpr (speech prosody)
Track 6	Experiment 1. SPKr (speech rhythm)

Track 7	Experiment 2. SPK (speech)
Track 8	Experiment 2. SNG, melody 1 (song)
Track 9	Experiment 2. SNG, melody 2 (song)
Track 10	Experiment 2. AMB, MT (ambiguous, monotonous)
Track 11	Experiment 2. AMB, MP (ambiguous, mimicked prosody
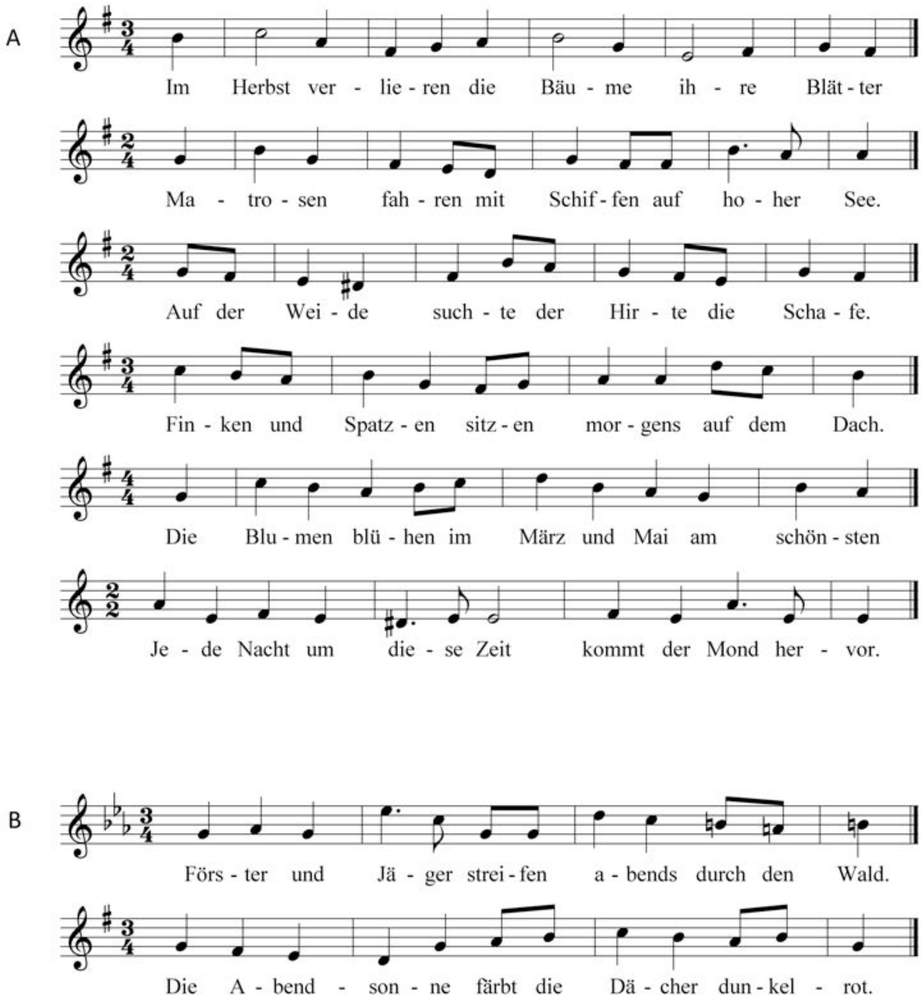
Figure A.1: Stimuli used in Experiment 1.

Bibliography

Abrams, D. A., Bhatara, A., Ryali, S., Balaban, E., Levitin, D. J., & Menon, V. (2011). Decoding temporal structure in music and speech relies on shared brain resources but elicits different fine-scale spatial patterns. *Cerebral Cortex*, *21*, 1507-1518.

Albert, M. L., Sparks, R. W., & Helm, N. A. (1973). Melodic intonation therapy for aphasia. *Archives of Neurology*, *29*, 130-131.

Aleman, A., Formisano, E., Koppenhagen, H., Hagoort, P., Haan, E. H. F., & Kahn, R. S. (2005). The functional neuroanatomy of metrical stress evaluation of perceived and imagined spoken words. *Cerebral Cortex*, *15*, 221-228.

Alexander, I., Cowey, A., & Walsh, V. (2005). The right parietal cortex and time perception: Back to critchley and the zeitraffer phenomenon. *Cognitive Neuropsychology*, *22(3)*, 306-315.

Allen, G. (1878). Note-deafness. *Mind*, *10*, 157-167.

Altenmüller, E., & Grossbach, M. (2002). Singen - die Ursprache? Zur Hirnphysiologie des Gesanges. In *Schriften zur Kulturwissenschaft, Band 46* (p. 58-66). Hamburg: Verlag Dr. Kovac.

Amaducci, L., Grassi, E., & Boller, F. (2002). Maurice Ravel and right-hemisphere musical creativity: Influence of disease on his last musical works? *European Journal of Neuroscience*, *9*, 75-82.

Ayotte, J., Peretz, I., & Hyde, K. (2002). Congenital amusia. A group study of adults afflicted with a music-specific disorder. *Brain*, *125*, 238-251.

Ayotte, J., Peretz, I., Rousseau, I., Bard, C., & Bojanowski, M. (2000). Patterns of music agnosia associated with middle cerebral artery infarcts. *Brain*, *123*, 1926-1938.

Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience*, *4*, 829-839.

Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychchology*, *63*, 1-29.

Bates, E., D'Amico, S., Jacobsen, T., Székely, A., & Andonova, E. (2003). Timed picture naming in seven languages. *Psychological Bulletin Review*, *10(2)*, 344-380.

Bautista, R. E. D., & Ciampetti, M. Z. (2003). Expressive aprosody and amusia as a manifestation of right hemisphere seizures. *Epilepsia*, *44(3)*, 466-467.

Belin, P., Eeckhout, P. van, Zilbovicius, M., Remy, P., François, C., Guillaume, S., et al. (1996). Recovery from nonfluent aphasia after melodic intonation therapy: A PET study. *Neurology*, *47*, 1504-1511.

Bendor, D., & Wang, X. (2005). The neuronal representation of pitch in primate auditory cortex. *Nature*, *436*, 1161-1165.

Bertinetto, P. M. (1989). Reflections on the dichotomy stress vs. syllable-timing. *Revue de Phonétique Appliquée*, *91-93*, 99-130.

Besson, M., Faïta, F., Peretz, I., Bonnel, A.-M., & Requin, J. (1998). Singing in the brain: Independence of lyrics and tunes. *Psychological Science*, *9*, 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.

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)*, 11-20.

Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S., Springer, J. A., Kaufman, J. N., et al. (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cerebral Cortex*, *10(5)*, 512-528.

Binder, J. R., Frost, J. A., Hammeke, T. A., Cox, R. W., Rao, S. M., & Prieto, T. (1997). Human brain language areas identified by functional magnetic resonance imaging. *Journal of Neuroscience*, *17(1)*, 353-362.

Blesser, R. (1972). Speech perception under conditions of spectral transformation: I. Phonetic characteristics. *Journal of Speech and Hearing Research*, *15*, 5-41.

Blonder, L. X., Bowers, D., & Heilman, K. M. (1991). The role of the right hemisphere in emotional communication. *Brain*, *114(3)*, 1115-1127.

Bode, S., Bogler, C., Soon, C. S., & Haynes, J. D. (2012). The neural encoding of guesses in the human brain. *NeuroImage*, *59(2)*, 1924-1931.

Boemio, A., Fromm, S., Braun, A., & Poeppel, D. (2005). Hierarchical and asymmetric temporal sensitivity in human auditory cortex. *Nature Neuroscience*, *3*, 389-395.

Bogler, C., Bode, S., & Haynes, J. D. (2011). Decoding successive computational stages of saliency processing. *Current Biology*, *21(19)*, 1667-1671.

Bolinger, D. (1989). *Intonation and its uses: Melody in grammar and discourse*. Stanford, CA: Stanford University Press.

Bookheimer, S. (2002). Functional MRI of language: New approaches to understanding the cortical organization of semantic processing. *Annual Review of Neuroscience*, *25*, 151-188.

Bregman, A. S., Liao, C., & Levitan, R. (1990). Auditory grouping based on fundamental frequency and formant peak frequency. *Canadian Journal of Psychology*, *44(3)*, 400-413.

Broca, P. P. (1861). Remarques sur le siége de la faculté du langage articulé, suivis d'une observation d'aphémie (perte de la parole). *Bulletin de la Société Anatomique*, *6*, 330-357.

Brown, S., Martinez, M., Hodges, D. A., Fox, P. T., & Parsons, L. M. (2004). The song system of the human brain. *Cognitive Brain Research*, *20*, 363-375.

Brown, S., & Martinez, M. J. (2007). Activation of premotor vocal areas during musical discrimination. *Brain and Cognition*, *63(1)*, 59-69.

Brown, S., Martinez, M. J., & Parsons, L. M. (2006). Music and language side by side in the brain: A PET study of the generation of melodies and sentences. *European Journal of Neuroscience*, *23(10)*, 2791-803.

Brown, S., Ngan, E., & Liotti, M. (2008). A larynx area in the human motor cortex. *Cerebral Cortex*, *18*, 837-845.

Brown, S., & Weishaar, K. (2010). Speech is heterometric: The changing rhythms of speech. *Speech Prosody*, *100074*, 1-4.

Buxton, R. (2002). *Introduction to functional magnetic resonance imaging: Prinziples and techniques*. Cambridge, MA: Cambridge University Press.

Cabeza, R., & Nyberg, L. (2000). Imaging cognition II: An empirical overview of 275 PET and fMRI studies. *Journal of Cognitive Neuroscience*, *12(1)*, 1-47.

Cadalbert, A., Landis, T., Regard, M., & Graves, R. E. (1994). Singing with and without words: Asymmetries in motor control. *Journal of Clinical and Experimental Neuropsychology*, *16(5)*, 664-70.

Callan, D. E., Kawato, M., Parsons, L., & Turner, R. (2007). Speech and song: The role of the cerebellum. *The Cerebellum*, *6*, 321-327.

Callan, D. E., Tsytsarev, V., Hanakawa, T., Callen, A. M., Katsuhara, M., Fukuyama, H., et al. (2006). Song and speech: Brain regions involved with perception and covert production. *NeuroImage*, *31*, 1327-1342.

Cancelliere, A. E., & Kertesz, A. (1990). Lesion localization in acquired deficits of emotional expression and comprehension. *Brain and Cognition*, *13(2)*, 133-147.

Carp, L. (1977). Gerorge Gershwin - illustrious American composer: His fatal glioblastoma. *American Journal of Surgery and Patholoy*, *3*, 473-478.

Caspers, S., Eickhoff, S., Geyer, S., Scheperjans, F., Mohlberg, H., Zilles, K., et al. (2008). The human inferior parietal lobule in stereotactic space. *Brain Structure and Function*, *212*, 481-495.

Caspers, S., Geyer, S., Schleicher, A., Mohlberg, H., Amunts, K., & Zilles, K. (2006). The human inferior parietal cortex: Cytoarchitectonic parcellation and interindividual variability. *NeuroImage*, *33(2)*, 430-448.

Caspers, S., Schleicher, A., Bacha-Trams, M., Palomero-Gallagher, N., Amunts, K., & Zilles, K. (2012). Organization of the human inferior parietal lobule based on receptor architectonics. *Cerebral Cortex*, 1-14.

Champod, A. S., & Petrides, M. (2007). Dissociable roles of the posterior parietal and the prefrontal cortex in manipulation and monitoring processes. *Proceedings of the National Academy of Sciences of the USA*, *104(37)*, 14837-14842.

Chang, C.-C., & Lin, C.-J. (2011). LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, *2(3)*, 27:1-27:27.

Chang, E. F., Rieger, J. W., Johnson, K., Berger, M. S., Barbaro, N. M., & Knight, R. T. (2010). Categorical speech representation in human superior temporal gyrus. *Nature Neuroscience*, *13(11)*, 1428-1432.

Chen, J. L., Penhune, V., & Zatorre, R. J. (2008). Listening to musical rhythms recruits motor regions of the brain. *Cerebral Cortex*, *18*, 2844-2854.

Chobert, J., Marie, C., François, C., Schön, D., & Besson, M. (2011). Enhanced passive and active processing of syllables in musician children. *Journal of Cognitive Neuroscience*, *23(12)*, 3874-3887.

Choi, H. J., Zilles, K., Mohlberg, H., Schleicher, A., Fink, G. R., & Armstrong, E. (2006). Cytoarchitectonic identification and probabilistic mapping of two distinct areas within the anterior ventral bank of the human intraparietal sulcus. *Journal of Computational Neurology*, *495(1)*, 53-69.

Confavreux, C., Croisile, B., Garassus, P., Aimard, G., & Trillet, M. (1992). Progressive amusia and aprosodia. *Archives of Neurology*, *49(9)*, 971-976.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, *3(3)*, 201-215.

Cowan, N., Li, D., Moffitt, A., Becker, T. M., Martin, E. A., Saults, J. S., et al. (2011). A neural region of abstract working memory. *Journal of Cognitive Neuroscience*, *23(10)*, 2852-2863.

Culham, J. C., & Kanwisher, N. G. (2001). Neuroimaging of cognitive functions in human parietal cortex. *Current Opinion in Neurobiology*, *11(2)*, 157-163.

Cusack, R. (2005). The intraparietal sulcus and perceptual organization. *Journal of Cognitive Neuroscience*, *17(4)*, 641-651.

Dalla Bella, S., Berkowska, M., & Sowiński, J. (2011). Disorders of pitch production in tone deafness. *Frontiers in Psychology*, *2(164)*, 1-11.

Dalla Bella, S., Giguere, J., & Peretz, I. (2009). Singing in congenital amusia. *Journal of the Acoustic Society of America*, *126*, 414-424.

Dalla Bella, S., Tremblay-Champoux, A., Berkowska, M., & Peretz, I. (2012). Memory disorders and vocal performance. *Annals of the New York Academy of Sciences*, *1252*, 338-344.

Damasio, H., & Damasio, A. R. (1989). *Lesion analysis in neuropsychology*. New York: Oxford University Press.

Davis, M. H., & Johnsrude, I. S. (2003). Hierarchical processing in spoken language comprehension. *Journal of Neuroscience*, *23*(8), 3423-31.

Davis, M. H., & Johnsrude, I. S. (2007). Hearing speech sounds: Top-down influences on the interface between audition and speech perception. *Hearing Research*, *229(1-2)*, 132-147.

Di Pietro, M., Laganaro, M., Leemann, B., & Schnider, A. (2004). Receptive amusia: Temporal auditory processing deficit in a professional musician following a left temporo-parietal lesion. *Neuropsychologia*, *42*(7), 868-877.

Donner, T. H., Kettermann, A., Diesch, E., Ostendorf, F., Villringer, A., & Brandt, S. A. (2002). Visual feature and conjunction searches of equal difficulty engage only partially overlapping frontoparietal networks. *NeuroImage*, *15(1)*, 16-25.

Dowling, W. (1968). Rhythmic fission and the perceptual organization of tone sequences. Unpublished Ph.D. Dissertation, Harvard University. Data puplished. In W. J. Dowling & D. Harwood (Eds.), *Music Cognition.* New York: Academic Press.

Dowling, W. (1999). The development of music perception and cognition. In D. Deutsch (Ed.), *The psychology of music* (p. 603-627). San Diego, CA: Academic Press.

Dronkers, N. F., Wilkins, D. P., van Valin, R. D., Redfern, B. B., & Jaeger, J. J. (2004). Lesion analysis of the brain areas involved in language comprehension. *Cognition*, *92(1-2)*, 145-177.

Eggert, G. H. (1977). *Wernicke's works on aphasia: A sourcebook and review*. The Hague: Mouton Publishers.

Eickhoff, S., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K., et al. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *NeuroImage*, *25(4)*, 1325-1335.

Farah, M. J. (1994). Neuropsychological inference with an interactive brain - a critique of the locality assumption. *Behavioral and Brain Sciences*, *17(1)*, 43-61.

Felber, M. (2009). *Perceptual factors contributing to song and speech distinction*. Unpublished master's thesis, Freie Universität Berlin.

Fitch, W. T. (2006). The biology and evolution of music: A comparative perspective. *Cognition*, *100*, 173-215.

Fodor, J. A. (1983). *The modularity of mind.* Cambridge, MA: The MIT Press.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). Mini-mental state (a practical method for grading the state of patients for the clinician). *Journal of Psychiatric Research*, *12*, 189-198.

Formby, C., Thomas, R. G., & Halsey, J. H. (1989). Regional cerebral blood flow for singers and nonsingers while speaking, singing, and humming a rote passage. *Brain and Language*, *36(4)*, 690-698.

Formisano, E., De Martino, F., Bonte, M., & Goebel, R. (2008). "Who" is saying "what"? Brain-based decoding of human voice and speech. *Science*, *322*, 970-973.

Formisano, E., Kim, D. S., Di Salle, F., Moortele, P. F. van de, Ugurbil, K., & Goebel, R. (2003). Mirror-symmetric tonotopic maps in human primary auditory cortex. *Neuron*, *40(4)*, 859-869.

Foster, N. E. V., & Zatorre, R. J. (2010). A role for the intraparietal sulcus in transforming musical pitch information. *Cerebral Cortex*, *20*, 1350-1359.

Foxton, J. M., Dean, J. L., Gee, R., Peretz, I., & Griffiths, T. D. (2004). Characterization of deficits in pitch perception underlying "tone deafness". *Brain*, *127*, 801-810.

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, *6*, 78-84.

Friederici, A. D. (2011). The brain basis of language: from structure to function. *Physiological Reviews*, *91*, 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*, 267-276.

Friederici, A. D., Kotz, S. A., Scott, S. K., & Obleser, J. (2010). Disentangling syntax and intelligibility in auditory language comprehension. *Human Brain Mapping*, *31*, 488-457.

Friederici, A. D., Meyer, M., & von Cramon, Y. D. (2000). Auditory language comprehension: An event-related fMRI study on the processing of syntactic and lexical information. *Brain and Language*, *75(3)*, 289-300.

Friederici, A. D., Pfeifer, E., & Hahne, A. (1993). Event-related brain potentials during natural speech processing: Effects of semantic, morphological and syntactic violations. *Cognitive Brain Research*, *1(3)*, 183-192.

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.

Gandour, J., Dzemidzic, M., Wong, D., Lowe, M., Tong, Y., Hsieh, L., et al. (2003). Temporal integration of speech prosody is shaped by language experience: An fMRI study. *Brain and Language*, *84*, 318-336.

Gandour, J., Tong, Y., Wong, D., Talavage, T., Dzemidzic, M., Xu, Y., et al. (2004). Hemispheric roles in the perception of speech prosody. *NeuroImage*, *23*(1), 344-357.

Gandour, J., Wong, D., Hsieh, L., Weinzapfel, B., Van Lancker, D., & Hutchins, G. D. (2000). A crosslinguistic PET study of tone perception. *Journal of Cognitive Neuroscience*, *12*, 207-222.

Gaser, C., & Schlaug, G. (2003). Brain structures differ between musicians and nonmusicians. *Journal of Cognitive Neuroscience*, *23(27)*, 9240-9245.

Geiser, E., Zaehle, T., Jäncke, L., & Meyer, M. (2008). The neural correlate of speech rhythm as evidenced by metrical speech processing. *Journal of Cognitive Neuroscience*, *20*, 541-552.

Geschwind, N. (1970). The organization of language and the brain. *Science*, *170(961)*, 940-944.

Giraud, A. L., Kell, C., Thierfelder, C., Sterzer, P., Russ, M. O., Preibisch, C., et al. (2004). Contributions of sensory input, auditory search and verbal comprehension to cortical activity during speech processing. *Cerebral Cortex*, *14(3)*, 247-255.

Gordon, H. W., & Bogen, J. E. (1974). Hemispheric lateralization of singing after intercarotid sodium amylobarbitone. *Journal of Neurology, Neurosurgery, and Psychiatry*, *37*, 727-738.

Gordon, R. L., Magne, C. L., & Large, E. W. (2011). EEG correlates of song prosody: A new look at the relationship between linguistic and musical rhythm. *Frontiers in Psychology*, *2(352)*, 1-13.

Gordon, R. L., Schön, D., Magne, C., Astésano, C., & Besson, M. (2010). Words and melody are intertwined in perception of sung words: EEG and behavioural evidence. *Plos One*, *5(3)*, e9889.

Goulet, G. M., Moreau, P., Robitaille, N., & Peretz, I. (2012). Congenital amusia persists in the developing brain after daily music listening. *Plos One*, *7(5)*, 1-9.

Grahn, J. A., & Brett, M. (2007). Rhythm and beat perception in motor areas of the brain. *Journal of Cognitive Neuroscience*, *19(5)*, 893-906.

Grefkes, C., & Fink, G. R. (2005). The functional organization of the intraparietal sulcus in humans and monkeys. *Journal of Anatomy*, *207*, 3-17.

Griffiths, T. D. (2003). Functional imaging of pitch analysis. *Annals of the New York Academy of Sciences*, *999*, 40-49.

Griffiths, T. D., Büchel, C., Frackowiak, R. S. J., & Patterson, R. D. (1998). Analysis of temporal structure in sound by the human brain. *Nature Neuroscience*, *1(5)*, 422-427.

Griffiths, T. D., Johnsrude, I. S., Dean, J. L., & Green, G. G. (1999). A common neural substrate for the analysis of pitch and duration pattern in segmented sound? *Neuroreport*, *10(18)*, 3825-3830.

Griffiths, T. D., Rees, A., Witton, C., Cross, P. M., Shakir, R. A., & Green, G. G. (1997). Spatial and temporal auditory processing deficits following right hemisphere infarction. A psychophysical study. *Brain*, *120(5)*, 785-794.

Griffiths, T. D., Uppenkamp, S., Johnsrude, I. S., Josephs, O., & Patterson, R. D. (2001). Encoding of the temporal regularity of sound in the human brainstem. *Nature Neuroscience*, *4(6)*, 633-637.

Griffiths, T. D., & Warren, J. D. (2002). The planum temporale as a computational hub. *Trends in Neurosciences*, *25(7)*, 348-353.

Gunji, A., Ishii, R., Chau, W., Kakigi, R., & Pantev, C. (2007). Rhythmic brain activities related to singing in humans. *NeuroImage*, *34*(1), 426-434.

Gutschalk, A., Patterson, R. D., Rupp, A., Uppenkamp, S., & Scherg, M. (2002). Sustained magnetic fields reveal separate sites for sound level and temporal regularity in human auditory cortex. *NeuroImage*, *15(1)*, 207-216.

Gutschalk, A., Patterson, R. D., Scherg, M., Uppenkamp, S., & Rupp, A. (2004). Temporal dynamics of pitch in human auditory cortex. *NeuroImage*, *22*(2), 755-766.

Hagoort, P. (2005). On Broca, brain, and binding: A new framework. *Trends in Cognitive Sciences*, *9(9)*, 416-423.

Hahne, A., & Friederici, A. D. (1999). Electrophysiological evidence for two steps in syntactic analysis. Early automatic and late controlled processes. *Journal of Cognitive Neuroscience*, *11(2)*, 194-205.

Halle, J., & Lerdahl, F. (1993). A generative textsetting model. *Current Musicology*, *55*, 3-23.

Halpern, A. R., & Zatorre, R. J. (1999). When that tune runs through your head: A PET investigation of auditory imagery for familiar melodies. *Cerebral Cortex*, *9*, 697-704.

Halpern, A. R., Zatorre, R. J., Bouffard, M., & Johnson, J. A. (2004). Behavioral and neural correlates of perceived and imagined musical timbre. *Neuropsychologia*, *42*, 1281-1292.

Halsband, U., Ito, N., Tanji, J., & Freund, H. (1993). The role of premotor cortex and the supplementary motor area in the temporal control of movement in man. *Brain*, *116*, 243-266.

Hanke, M., Halchenko, Y. O., Sederberg, P. B., Hanson, S. J., Haxby, J. V., & Pollmann, S. (2009). PyMVPA: A Python toolbox for multivariate pattern analysis of fMRI data. *Neuroinformatics*, *7*, 37-53.

Haxby, J., Gobbini, M., Furey, M., Ishai, A., Schouten, J., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, *293*, 2425-2430.

Haynes, J. D., & Rees, G. (2005). Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nature Neuroscience*, *8*, 686-691.

Haynes, J. D., & Rees, G. (2006). Decoding mental states from brain sensitivity in humans. *Nature Reviews Neuroscience*, *7*, 523-534.

Hebert, S., & Peretz, I. (2001). Are text and tunes of familiar songs seperable by brain damage? *Tennet*, *11*, 169-175.

Hebert, S., Racette, A., Gagnon, L., & Peretz, I. (2003). Revisiting the dissociation between singing and speaking in expressive aphasia. *Brain*, *126*, 1838-1850.

Henson, R. A. (1988). Maurice Ravel's illness: A tragedy of lost creativity. *British Medical Journal (Clinical research edition)*, *296(6636)*, 1585-1588.

Hickok, G., Buchsbaum, B., Humphries, C., & Muftuler, T. (2003). Auditory-motor interaction revealed by fMRI: Speech, music, and working memory in area Spt. *Journal of Cognitive Neuroscience*, *15*(5), 673-682.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*(5), 393-402.

Hill, K. T., & Miller, L. M. (2010). Auditory attentional control and selection during cocktail party listening. *Cerebral Cortex*, *20(3)*, 583-590.

Hsieh, L., Gandour, J., Wong, D., & Hutchins, G. D. (2001). Functional heterogeneity of inferior frontal gyrus is shaped by linguistic experience. *Brain and Language*, *76(3)*, 227-252.

Huber, W., Poeck, K., Weniger, D., & Willmes, K. (1993). *Aachener Aphasie Test (AAT)*. Göttingen: Hogrefe.

Hugdahl, K., Bronnick, K., Kyllingsbaek, S., Law, I., Gade, A., & Paulson, O. B. (1999). Brain activation during dichotic presentations of consonant-vowel and musical instrument stimuli: A 15O-PET study. *Neuropsychologia*, *37(4)*, 431-440.

Hugdahl, K., Law, I., Kyllingsbaek, S., Bronnick, K., Gade, A., & Paulson, O. B. (2000). Effects of attention on dichotic listening: An 15O-PET study. *Human Brain Mapping*, *10(2)*, 87-97.

Humphries, C., Binder, J. R., Medler, D. A., & Liebenthal, E. (2006). Syntactic and semantic modulation of neural activity during auditory sentence comprehension. *Journal of Cognitive Neuroscience*, *18(4)*, 665-679.

Husain, M., & Nachev, P. (2007). Space and the parietal cortex. *Trends in Cognitive Sciences*, *11(1)*, 30-36.

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.

Hutchins, S., & Peretz, I. (2011). Perception and action in singing. *Progress in Brain Research*, *191*, 103-118.

Hyde, K., Lerch, J., Zatorre, R. J., Griffiths, T. D., Evans, A. C., & Peretz, I. (2007). Cortical thickness in congenital amusia: When less is better than more. *Journal of Neuroscience*, *27*, 13028-13032.

Hyde, K., & Peretz, I. (2004). Brains that are out of tune but in time. *Psychological Science*, *15(5)*, 356-360.

Hyde, K., Peretz, I., & Zatorre, R. J. (2008). Evidence for the role of the right auditory cortex in fine pitch resolution. *Neuropsychologia*, *46*, 632-639.

Hyde, K., Zatorre, R. J., Griffiths, T. D., Lerch, J., & Peretz, I. (2006). Morphometry of the amusic brain: a two-site study. *Brain*, *129*, 2562-2570.

Hyde, K., Zatorre, R. J., & Peretz, I. (2011). Functional MRI evidence of an abnormal neural network for pitch processing in congenital amusia. *Cerebral Cortex*, *21*, 292-299.

Indefrey, P., & Levelt, W. J. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92(1-2)*, 101-144.

Ivry, R. B., & Hazeltine, E. (1995). Perception and production of temporal intervals across a range of durations: Evidence for a common timing mechanism. *Journal of Experimental Psychology: Human Perception & Performance*, *21*, 3-18.

Ivry, R. B., & Keele, S. W. (1989). Timing functions of the cerebellum. *Journal of Cognitive Neuroscience*, *1*, 136-152.

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.

Janata, P., & Grafton, S. T. (2003). Swinging in the brain: Shared neural substrates for behaviors related to sequencing and music. *Nature Neuroscience*, *6*, 682-687.

Janata, P., Tillmann, B., & Bharucha, J., Jamshed. (2002). Listening to polyphonic music recruits domain-general attention and working memory circuits. *Cognitive, Affective & Behavioral Neuroscience*, *2*(2), 121-140.

Jeffries, K. J., Fritz, J. B., & Braun, A. R. (2003). Words in melody: An H2 15O PET study of brain activation during singing and speaking. *NeuroReport*, *14*(5), 749-754.

Jernigan, T. L., Ostergaard, A. L., Law, I., Svarer, C., Gerlach, C., & Paulson, O. B. (1998). Brain activation during word identification and word recognition. *NeuroImage*, *8(1)*, 93-105.

Jezzard, P., Matthews, P. M., & Smith, S. M. (2001). *Functional MRI: An introduction to methods*. New York: Oxford University Press.

Jiang, C., Hamm, J. P., Lim, V. K., Kirk, I. J., & Yang, Y. (2010). Processing melodic contour and speech intonation in congenital amusics with mandarin chinese. *Neuropsychologia*, *48*, 2630-2639.

Jiang, C., Hamm, J. P., Lim, V. K., Kirk, I. J., & Yang, Y. (2011). Fine-grained pitch discrimination in congenital amusics with mandarin chinese. *Music Perception*, *28*, 519-526.

Joanisse, M. F., Zevin, J. D., & McCandliss, B. D. (2007). Brain mechanisms implicated in the preattentive categorization of speech sounds revealed using fMRI and a short-interval habituation trial paradigm. *Cerebral Cortex*, *17(9)*, 2084-2093.

Johnsrude, I. S., Penhune, V., & Zatorre, R. J. (2000). Functional specificity in the right human auditory cortex for perceiving pitch direction. *Brain*, *123(1)*, 155-163.

Jones, J., Zalewski, C., Brewer, C., Lucker, J., & Drayna, D. (2009). Widespread auditory deficits in tune deafness. *Ear and Hearing*, *30*, 63-72.

Kahnt, T., Heinzle, J., Park, S. Q., & Haynes, J. D. (2010). The neural code of reward anticipation in human orbitofrontal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *107(13)*, 6010-6015.

Kalmus, H., & Fry, D. B. (1980). On tune deafness (dysmelodia): Frequency, development, genetics and musical background. *Annals of Human Genetics*, *43(4)*, 369-382.

Kleber, B., Birbaumer, N., Veit, R., Trevorrow, T., & Lotze, M. (2007). Overt and imagined singing of an italian aria. *NeuroImage*, *36*, 889-900.

Kleber, B., Veit, R., Birbaumer, N., Gruzelier, J., & Lotze, M. (2010). The brain of opera singers: Experience-dependent changes in functional activation. *Cerebral Cortex*, *20(5)*, 1144-1152.

Klein, J. C., & Zatorre, R. J. (2011). A role for the right superior temporal sulcus in categorical perception of musical chords. *Neuropsychologia*, *49*, 878-887.

Knösche, T. R., Neuhaus, C., Haueisen, J., Alter, K., Maess, B., Witte, O. W., et al. (2005). The perception of phrase structure in music. *Human Brain Mapping*, *24*, 259-273.

Koelsch, S. (2011). Toward a neural basis of music perception - a review and updated model. *Frontiers in Psychology*, *2(110)*, 1-20.

Koelsch, S., Fritz, T., Schulze, K., Aslop, D., & Schlaug, G. (2005). Adults and children processing music: An fMRI study. *NeuroImage*, *25*, 1068-1076.

Koelsch, S., Gunter, T. C., Friederici, A. D., & Schroeger, E. (2000). Brain indices of music processing: “nonmusicians” are musical. *Journal of Cognitive Neuroscience*, *12(3)*, 520-541.

Koelsch, S., Gunter, T. C., von Cramon, D. Y., Zysset, S., Lohmann, G., & Friederici, A. D. (2002). Bach speaks: A cortical “language-network” serves the processing of music. *NeuroImage*, *17*(2), 956-966.

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*, 1-13.

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*(3), 302-307.

Koelsch, S., Schulze, K., Sammler, D., Fritz, T., Mueller, K., & Friederici, A. D. (2009). Functional architecture of verbal and tonal working memory: An fMRI study. *Human Brain Mapping*, *30*, 859-873.

Koelsch, S., & Siebel, W. A. (2005). Towards a neural basis of music perception. *Trends in Cognitive Sciences*, *9*(12), 578-584.

Kotz, S. A., Meyer, M., Alter, K., Besson, M., & von Cramon, Y. D. (2003). On the lateralization of emotional prosody: An event-related functional MR investigation. *Brain and Language*, *86*, 366-376.

Kotz, S. A., & Paulmann, S. (2007). When emotional prosody and semantics dance cheek to cheek: ERP evidence. *Brain Research*, *1151*, 107-118.

Kotz, S. A., & Schwartze, M. (2010). Cortical speech processing unplugged: a timely subcortico-cortical framework. *Trends in Cognitive Sciences*, *14(9)*, 392-399.

Kotz, S. A., Schwartze, M., & Schmidt-Kassow, M. (2009). Non-motor basal ganglia functions: A review and proposal for a model of sensory predictability in auditory language perception. *Cortex*, *45*, 982-990.

Kraemer, D. J. M., Macrae, C. N., Green, A. E., & Kelley, W. M. (2005). Sound of silence activates auditory cortex. *Nature*, *434*, 158.

Kriegeskorte, N., & Bandettini, P. (2007). Analyzing for information, not activation, to exploit high-resolution fMRI. *NeuroImage*, *38(4)*, 649-662.

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America*, *103(10)*, 3863-3868.

Krumhansl, C. L. (2004). The cognition of tonalitiy - as we know it today. *Journal of New Music Research*, *33(3)*, 253-268.

Langner, G. (1992). Periodicity coding in the auditory system. *Hearing Research*, *60(2)*, 115-142.

Langner, G., & Schreiner, C. E. (1988). Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanisms. *Journal of Neurophysiology*, *60(6)*, 1799-1822.

Large, E. W., & Kolen, J. F. (1994). Resonance and the perception of musical meter. *Connection Science*, *6(1)*, 177-208.

Lee, Y.-S., Janata, P., Frost, C., Hanke, M., & Granger, R. (2011). Investigation of melodic contour processing in the brain using multivariate pattern-based fMRI. *NeuroImage*, *57(1)*, 293-300.

Leff, A. P., Iverson, P., Schofield, T. M., Kilner, J. M., Crinion, J. T., Friston, K. J., et al. (2009). Vowel-specific mismatch responses in the anterior superior temporal gyrus: An fMRI study. *Cortex*, *45*, 517-526.

Lehiste, I. (1977). Isochrony reconsidered. *Journal of Phonetics*, *5*, 253-263.

Lehiste, I., & Meltzer, D. (1973). Vowel and speaker identification in natural and synthetic speech. *Language and Speech*, *16(4)*, 356-364.

Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge, MA: MIT Press.

Levelt, W. J. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Levitin, D. J., & Menon, V. (2003). Musical structure is processed in language areas of the brain: A possible role for Brodmann Area 47 in temporal coherence. *NeuroImage*, *20*, 2142-2152.

Levitin, D. J., & Tirovolas, A. K. (2009). Current advances in the cognitive neuroscience of music. *Annals of the New York Academy of Sciences*, *1156*, 211-231.

Liberman, M. (1975). *The intonational system of English.* Unpublished doctoral dissertation, MIT.

Lichtheim, L. (1885). On aphasia. *Brain*, *7*, 433-484.

Lidji, P., Jolicoeur, P., Kolinsky, R., Moreau, P., Connolly, J. F., & Peretz, I. (2010). Early integration of vowel and pitch processing: A mismatch negativity study. *Clinical Neurophysiology*, *121*, 533-541.

Liebenthal, E., Binder, J. R., Spitzer, S. M., Possing, E. T., & Medler, D. A. (2005). Neural substrates of phonemic perception. *Cerebral Cortex*, *15(10)*, 1621-1631.

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(10)*, 1853-1867.

Lindholm, B., & Sundberg, J. (2007). The human voice in speech and singing. In T. D. Rossing (Ed.), *Springer handbook of acoustics* (p. 669-712). New York: Springer.

Liu, F., Jiang, C., Thompson, W. F., Xu, Y., Yang, Y., & Stewart, L. (2012). The mechanism of speech processing in congenital amusia: Evidence from mandarin speakers. *Plos One*, *7(2)*, 1-11.

Liu, F., Patel, A. D., Fourcin, A., & Steward, L. (2010). Intonation processing in congenital amusia: Discrimination, identification and imitation. *Brain*, *133*, 1682-1693.

Liu, F., Xub, Y., Patel, A. D., Francart, T., & Jiang, C. (2012). Differential recognition of pitch patterns in discrete and gliding stimuli in congenital amusia: Evidence from mandarin speakers. *Brain and Cognition*, *79*, 209-215.

Logothetis, N. K. (2008). What we can do and what we cannot do with fMRI. *Nature*, *453(7197)*, 869-878.

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., & Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature*, *412*, 150-157.

Loui, P., Aslop, D., & Schlaug, G. (2009). Tone deafness: A new disconnection syndrome? *Journal of Neuroscience*, *29*(33), 10215-10220.

Loui, P., Guenther, F. H., Mathys, C., & Schlaug, G. (2008). Action-perception mismatch in tone-deafness. *Current Biology*, *18*, 331-332.

Loui, P., & Schlaug, G. (2009). Investigating musical disorders with diffusion tensor imaging. *Neurosciences and Music III - Disorders and Plasticity*, *1169*, 121-125.

Luria, A. R., Tsevetkova, L. S., & Futer, D. S. (1965). Aphasia in a composer. *Journal of the Neurological Sciences*, *2*, 288-292.

Macmillan, N. A., & Creelman, C. D. (2009). *Detection theory - a user's guide*. New York: Psychology Press.

Magne, C., Schön, D., & Besson, M. (2003). Prosodic and melodic processing in adults and children. Behavioral and electrophysiologic approaches. *Annals of the New York Academy of Sciences*, *999*, 461-476.

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.

Marler, P. (1970). Birdsong and speech development: Could there be parallels? *American Scientist*, *58*, 669-673.

Mattys, S. L. (1997). The use of time during lexical processing and segmentation: A review. *Psychonomic Bullettin & Review*, *4*, 310-329.

Mazoyer, B., Tzourio, N., Frak, V., & Syrota, A. (1993). The cortical representation of speech. *Journal of Cognitive Neuroscience*, *5*, 467-479.

McAuley, J. D., & Miller, N. S. (2007). Picking up the pace: Effects of global temporal context on sensitivity to the tempo of auditory sequences. *Perception & Psychophysics*, *69(5)*, 709-718.

McChesney-Atkins, S., Davies, K. G., Montouris, G. D., Silver, J. T., & Menkes, D. L. (2003). Amusia after right frontal resection for epilepsy with singing seizures: Case report and review of the literature. *Epilepsy & Behavior*, *4(3)*, 343-734.

McDermott, J., & Hauser, M. D. (2005). Probing the evolutionary origins of music perception. *Annals of the New York Academy of Sciences*, *1060*, 6-16.

McFarland, H. R., & Fortin, D. (1982). Amusia due to right temporoparietal infarct. *Archives of Neurology*, *39*, 725-727.

McMullen, E., & Saffran, J. R. (2004). Music and language: A developmental comparison. *Music Perception*, *21*, 289-311.

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. *Frontiers in Psychology*, *3(76)*, 1-13.

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, 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.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63(2)*, 81-97.

Milner, B. (1962). Laterality effects in audition. In V. B. Mountcastle (Ed.), *Interhemispheric relations and cerebral dominance* (p. 177-195). Baltimore: Johns Hopkins Press.

Moreau, P., Jolicoeur, P., & Peretz, I. (2009). Automatic brain responses to pitch changes in congenital amusia. *Annals of the New York Academy of Sciences*, *1169*, 191-194.

Moreno, S., & Besson, M. (2006). Musical training and language-related brain electrical activity in children. *Psychophysiology*, *43(3)*, 287-291.

Münte, T. F., Altenmüller, E., & Jäncke, L. (2002). The musician's brain as a model of neuroplasticity. *Nature Reviews Neuroscience*, *3(6)*, 473-478.

Murayama, J., Kashiwagi, T., Kashiwagi, A., & Mimura, M. (2004). Impaired pitch production and preserved rhythm production in a right brain-damaged patient with amusia. *Brain and Cognition*, *56*, 36-42.

Musacchia, G., Sams, M., Skoe, E., & Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proceedings of the National Academy of Sciences of the United States of America*, *104(40)*, 15894-15898.

Nan, Y., Sun, Y., & Peretz, I. (2010). Congenital amusia in speakers of a tone language: Association with lexical tone agnosia. *Brain*, *133(9)*, 2635-2642.

Narain, C., Scott, S. K., Wise, R. J., Rosen, S., Leff, A. A., Iversen, S. D., et al. (2003). Defining a left-lateralized response specific to intelligible speech using fMRI. *Cerebral Cortex*, *13(12)*, 1362-1368.

Narmour, E. (1990). *The analysis and cognition of basic melodic structures*. Chicago: University of Chicago Press.

Nee, D. E., Brown, J. W., Askren, M. K., Berman, M. G., Demiralp, E., Krawitz, A., et al. (2012). A meta-analysis of executive components of working memory. *Cerebral Cortex*.

Nichols, T., Brett, M., Andersson, J., Wager, T., & Poline, J.-B. (2005). Valid conjunction inference with the minimum statistic. *NeuroImage*, *25*, 653-660.

Nicholson, K. G., Baum, S. R., Cuddy, L. L., & Munhall, K. G. (2002). A case of impaired auditory and visual speech prosody perception after right hemisphere damage. *Neurocase*, *8(4)*, 314-322.

Nicholson, K. G., Baum, S. R., Kilgour, A., Koh, C. K., Munhall, K. G., & Cuddy, L. L. (2003). Impaired processing of prosodic and musical patterns after right hemisphere damage. *Brain and Cognition*, *52*(3), 382-389.

Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: Multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences*, *10(9)*, 424-430.

Norris, D. G. (2006). Principles of magnetic resonance assessment of brain function. *Journal of Magnetic Resonance Imaging*, *23(6)*, 794-807.

Norton, A., Zipse, L., Marchina, S., & Schlaug, G. (2009). Melodic intonation therapy. Shared insights on how it is done and why it might help. In S. Dalla Bella et al. (Eds.), *The neurosciences and music 3: Disorders and plasticity* (p. 431-436). New York: Academy of Sciences.

Obleser, J., Boecker, H., Drzezga, A., Haslinger, B., Hennenlotter, A., Roettinger, M., et al. (2006). Vowel sound extraction in anterior superior temporal cortex. *Human Brain Mapping*, *27*, 562-571.

Obleser, J., Eisner, F., & Kotz, S. A. (2008). Bilateral speech comprehension reflects differential sensitivity to spectral and temporal features. *Journal of Neuroscience*, *28(32)*, 8116-8124.

Obleser, J., & Kotz, S. A. (2010). Expectancy constraints in degraded speech modulate the language comprehension network. *Cerebral Cortex*, *20*, 633-640.

Obleser, J., Wise, R. J., Dresner, M. A., & Scott, S. K. (2007). Functional integration across brain regions improves speech perception under adverse listening conditions. *Journal of Neuroscience*, *27*, 2283-2289.

Okada, K., Rong, F., Venezia, J., Matchin, W., Hsieh, I. H., Saberi, K., et al. (2010). Hierarchical organization of human auditory cortex: Evidence from acoustic invariance in the response to intelligible speech. *Cerebral Cortex*, *20(10)*, 1486-1495.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9(1)*, 97-113.

Omigie, D., Pearce, M. T., & Stewart, L. (2012). Tracking of pitch probabilities in congenital amusia. *Neuropsychologia*, *50(7)*, 1483-1493.

Özdemir, E., Norton, A., & Schlaug, G. (2006). Shared and distinct neural correlates of singing and speaking. *NeuroImage*, *33*(2), 628-35.

Pannekamp, A., Toepel, U., Alter, K., Hahne, A., & Friederici, A. D. (2005). Prosody driven sentence processing. *Journal of Cognitive Neuroscience*, *17*, 407-421.

Pantev, C., Hoke, M., Lütkenhöner, B., & Lehnertz, K. (1989). Tonotopic organization of the auditory cortex: Pitch versus frequency representation. *Science*, *246(4929)*, 486-488.

Papanicolaou, A. (1998). *Fundamentals of functional brain imaging: A guide to the methods and their applications to psychology and behavioral neuroscience*. Lisse: Swets & Zeitlinger B. V.

Parsons, L. M. (2001). Exploring the functional neuroanatomy of music performance, perception, and comprehension. *Annals of the New York Academy of Sciences*, *930*, 211-231.

Patel, A. D. (2003a). Language, music, syntax and the brain. *Nature Neuroscience*, *6:7*.

Patel, A. D. (2003b). Rhythm in language and music, parallels and differences. *Annals of the New York Academy of Sciences*, *999*, 140-143.

Patel, A. D. (2008). *Language, music, syntax, 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*, 310-313.

Patel, A. D., Iversen, J. R., & Rosenberg, J. C. (2006). Comparing the rhythm and melody of speech and music: The case of British English and French. *Journal of the Acoustic Society of America*, *119*(5 Pt 1), 3034-3047.

Patel, A. D., Peretz, I., Tramo, M. J., & Labreque, R. (1998). Processing prosodic and musical patterns: A neuropsychological investigation. *Brain and Language*, *61*(1), 123-144.

Patel, A. D., Wong, M., Foxton, J. M., Lochy, A., & Peretz, I. (2008). Speech intonation perception deficits in musical tone deafness (congenital amusia). *Music Perception*, *25*(4), 357-368.

Patterson, R. D., Uppenkamp, S., Johnsrude, I. S., & Griffiths, T. D. (2002). The processing of temporal and melody information in auditory cortex. *Neuron*, *36*, 767-776.

Pell, M. D., & Baum, S. R. (1997a). The ability to perceive and comprehend intonation in linguistic and affective contexts by brain-damaged adults. *Brain and Language*, *57(1)*, 80-99.

Pell, M. D., & Baum, S. R. (1997b). Unilateral brain damage, prosodic comprehension deficits, and the acoustic cues to prosody. *Brain and Language*, *57(2)*, 195-214.

Penagos, H., Melcher, J. R., & Oxenham, A. J. (2004). A neural representation of pitch salience in nonprimary human auditory cortex revealed with functional magnetic resonance imaging. *Journal of Neuroscience*, *24(30)*, 6810-6815.

Penhune, V., Zatorre, R. J., & Evans, A. C. (1998). Cerebellar contributions to motor timing: A PET study of auditory and visual rhythm reproduction. *Journal of Cognitive Neuroscience*, *10(6)*, 752-765.

Perani, D., Saccuman, M., Scifo, P., Spada, D., Andreolli, G., Rovelli, R., et al. (2010). Functional specializations for music processing in the human newborn brain. *Proceedings of the National Academy of Sciences of the United States of America*, *107*, 4758-4763.

Peretz, I. (1990). Processing of local and global musical information by unilateral brain-damaged patients. *Brain*, *113*, 1185-1205.

Peretz, I. (1996). We can lose memory for music? A case of music agnosia in a nonmusician. *Journal of Cognitive Neuroscience*, *8(6)*, 481-496.

Peretz, I., Ayotte, J., Zatorre, R. J., Mehler, J., Ahad, P., Penhune, V. B., et al. (2002). Congenital amusia: A disorder of fine-grained pitch discrimination. *Neuron*, *33*, 185-191.

Peretz, I., & Belleville, S. (1997). Dissociations entre musique et langage aprés atteinte cérébrale: un nouveau cas d'amusie sans aphasie. *Revue canadienne de psychologie expérimentale*, *51(4)*, 354-367.

Peretz, I., Blood, A. J., Penhune, V., & Zatorre, R. J. (2001). Cortical deafness to dissonance. *Brain*, *124*, 928-940.

Peretz, I., Brattico, E., Järvenpää, M., & Tervaniemi, M. (2009). The amusic brain: In tune, out of key, and unaware. *Brain*, *132(5)*, 1-10.

Peretz, I., Champod, A. S., & Hyde, K. (2003). Varieties of musical disorders. The Montreal Battery of Evaluation of Amusia. *Annals of the New York Academy of Sciences*, *999*, 58-75.

Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nature Neuroscience*, *6(7)*, 688-691.

Peretz, I., Gagnon, L., Hebert, S., & Macoir, J. (2004). Singing in the brain: Insights from cognitive neuropsychology. *Music Perception*, *21(3)*, 1-18.

Peretz, I., & Hyde, K. (2003). What is specific to music processing? Insights from congenital amusia. *Trends in Cognitive Sciences*, *7*(8), 326-367.

Peretz, I., & Kolinsky, R. (1993). Boundaries of separability between melody and rhythm in music discrimination: A neuropsychological perspective. *Quarterly Journal of Experimental Psychology*, *46(2)*, 301-325.

Peretz, I., Kolinsky, R., Tramo, M. J., Labreque, R., Hubler, C., Demeurisse, G., et al. (1994). Functional dissociations following bilateral lesions of auditory cortex. *Brain*, *117*, 1283-1301.

Peretz, I., Saffran, J., Schön, D., & Gosselin, N. (2012). Statistical learning of speech, not music, in congenital amusia. *Annals of the New York Academy of Sciences*, *1252*, 361-367.

Peretz, I., & Zatorre, R. J. (2005). Brain organization for music processing. *Annual Review of Psychchology*, *56*, 89-114.

Perry, D. W., Zatorre, R. J., Petrides, M., Alivisatos, B., Meyer, E., & Evans, A. C. (1999). Localization of cerebral activity during simple singing. *NeuroReport*, *10*, 3979-3984.

Pfordresher, P. Q., Brown, S., Meier, K. M., Belyk, M., & Liotti, M. (2010). Imprecise singing is widespread. *Journal of the Acoustical Society of America*, *128(4)*, 2182-2190.

Phillips-Silver, J., Toiviainen, P., Gosselin, N., Piché, O., Nozaradan, S., Palmer, C., et al. (2011). Born to dance but beat deaf: A new form of congenital amusia. *Neuropsychologia*, *49(5)*, 961-969.

Piccirilli, M., Sciarma, T., & Luzzi, S. (2000). Modularity of music: Evidence from a case of pure amusia. *Journal of Neurology, Neurosurgery, and Psychiatry*, *69(4)*, 541-545.

Pike, K. N. (1945). *The intonation of American english*. Ann Arbor: University of Michigan Press.

Plack, C. J., Oxenham, A. J., Fay, R. R., & Popper, A. N. (2005). *Pitch: Neural coding and perception*. Berlin: Springer.

Plante, E., Creusere, M., & Sabin, C. (2002). Dissociating sentential prosody from sentence processing: Activation interacts with task demands. *NeuroImage*, *17*, 401-410.

Platel, H., Price, C., Baron, J. C., Wise, R., Lambert, J., Frackowiak, R. S., et al. (1997). The structural components of music perception. A functional anatomical study. *Brain*, *120 (2)*, 229-43.

Poeppel, D. (1997). A hierarchial model of temporal perception. *Trends in Cognitive Sciences*, *1*, 56-61.

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as “asymmetric sampling in time”. *Speech Communication*, *41*, 245-255.

Poeppel, D., Guillemin, A., Thompson, J., Fritz, J., Bavelier, D., & Braun, A. R. (2004). Auditory lexical decision, categorical perception, and FM direction discrimination differentially engage left and right auditory cortex. *Neuropsychologia*, *42*, 183-200.

Price, C. J. (2010). The anatomy of language: A review of 100 fMRI studies published in 2009. *Annals of the New York Academy of Sciences*, *1191*, 62-88.

Price, C. J., Thierry, G., & Griffiths, T. D. (2005). Speech-specific auditory processing: where is it? *Trends in Cognitive Sciences*, *9*(6), 271-276.

Price, P. J., Ostendorf, M., Shattuck-Hufnagel, S., & Fong, C. (1991). The use of prosody in syntactic disambiguation. *Journal of the Acoustic Society of America*, *90(6)*, 2956-2970.

Racette, A., Bard, C., & Peretz, I. (2006). Making non-fluent aphasics speak: Sing along! *Brain*, *129*, 2571-2584.

Randel, D. M. (1978). *The harvard concise dictionary of music*. Cambridge, MA: Harvard University Press.

Riecker, A., Ackermann, H., Wildgruber, D., Dogil, G., & Grodd, W. (2000). Opposite hemispheric lateralization effects during speaking and singing at motor cortex, insula and cerebellum. *NeuroReport*, *11*(9), 1997-2000.

Riecker, A., Wildgruber, D., Dogil, G., Grodd, W., & Ackermann, H. (2002). Hemispheric lateralization effects of rhythm implementation during syllable repetitions: An fMRI study. *NeuroImage*, *16*, 169-176.

Rinne, T., Kirjavainen, S., Salonen, O., Degerman, A., Kang, X., & Woods, D. L. (2007). Distributed cortical networks for focused auditory attention and distraction. *Neuroscience Letters*, *416(3)*, 247-251.

Rissman, J., Eliassen, J. C., & Blumstein, S. E. (2003). An event-related fMRI investigation of implicit semantic priming. *Journal of Cognitive Neuroscience*, *15(8)*, 1160-1175.

Robson, H., Sage, K., & Lambon, M. A. (2012). Wernicke's aphasia reflects a combination of acoustic-phonological and semantic control deficits: A case-series comparison of Wernicke's aphasia, semantic dementia and semantic aphasia. *Neuropsychologia*, *50*, 266-275.

Rodd, J. M., Davis, M. H., & Johnsrude, I. S. (2005). The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity. *Cerebral Cortex*, *15*, 1261-1269.

Ross, E. D. (1981). The aprosodias. Functional-anatomic organization of the affective components of language in the right hemisphere. *Archives of Neurology*, *38(9)*, 561-569.

Rothermich, K., Schmidt-Kassow, M., & Kotz, S. A. (2012). Rhythm's gonna get you: Regular meter facilitates semantic sentence processing. *Neurospychologia*, *50(2)*, 232-244.

Rottschy, C., Langner, R., Dogan, I., Reetz, K., Laird, A. R., Schulz, J. B., et al. (2012). Modelling neural correlates of working memory: A coordinate-based meta-analysis. *NeuroImage*, *60*, 830-846.

Russell, S. M., & Golfinos, J. G. (2003). Amusia following resection of a Heschl gyrus glioma. Case report. *Journal of Neurosurgery*, *98(5)*, 1109-1112.

Saito, Y., Ishii, K., Yagi, K., Tatsumi, I. F., & Mizusawa, H. (2006). Cerebral networks for spontaneous and synchronized singing and speaking. *NeuroReport*, *17*(18), 1893-1897.

Sammler, D., Baird, A., Valabreque, 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. *Journal of Neuroscience*, *30*, 3572-3578.

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.

Samson, F., Zeffiero, T. A., Toussaint, A., & Belin, P. (2011). Stimulus complexity and categorical effects in human auditory cortex: An activation likelihood estimation meta-analysis. *Frontiers in Psychology*, *1(241)*, 1-23.

Samson, S., & Zatorre, R. J. (1988). Melodic and harmonic discrimination following unilateral cerebral excision. *Brain and Cognition*, *7(3)*, 348-360.

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*, *17(4)*, 793-804.

Samson, S., & Zatorre, R. J. (1992). Learning and retention of melodic and verbal information after unilateral temporal lobectomy. *Neuropsychologia*, *30(9)*, 815-826.

Satoh, M., Takeda, K., Murakami, Y., Onouchi, K., Inoue, K., & Kuzuhara, S. (2005). A case of amusia caused by the infarction of anterior portion of bilateral temporal lobes. *Cortex*, *41*, 77-83.

Scheperjans, F., Eickhoff, S., Hömke, L., Mohlberg, H., Hermann, K., Amunts, K., et al. (2008). Probabilistic maps, morphometry, and variability of cytoarchitectonic areas in the human superior parietal cortex. *Cerebral Cortex*, *18(9)*, 2141-2157.

Schirmer, A., Fox, P. M., & Grandjean, D. (2012). On the spatial organization of sound processing in the human temporal lobe: A meta-analysis. *NeuroImage*, *63*, 137-147.

Schirmer, A., & Kotz, S. A. (2006). Beyond the right hemisphere: Brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences*, *10*(1), 24-30.

Schmahmann, J. D., Doyon, J., Toga, A. W., Petrides, M., & Evans, A. C. (2000). *MRI atlas of the human cerebellum*. San Diego, CA: Academic Press.

Schmidt-Kassow, M., & Kotz, S. A. (2009). Event-related brain potentials suggest a late interaction of meter and syntax in the P600. *Journal of Cognitive Neuroscience*, *21(9)*, 1693-1708.

Schmithorst, V. J. (2005). Separate cortical networks involved in music perception: Preliminary functional MRI evidence for modularity of music processing. *NeuroImage*, *25*, 444-451.

Schön, D., Gordon, R., Campagne, A., Magne, C., Astésano, C., Anton, J.-L., et al. (2010). Similar cerebral networks in language, music and song perception. *NeuroImage*, *51*, 450-461.

Schön, D., Gordon, R. L., & Besson, M. (2005). Musical and linguistic processing in song perception. *Annals of the New York Academy of Sciences*, *1060*, 71-81.

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.

Schönwiesner, M., & Zatorre, R. J. (2008). Depth electrode recording show double dissociation between pitch processing in lateral Heschl's gyrus and sound onset processing in medial Heschl's gyrus. *Experimental Brain Research*, *187*, 97-105.

Schulze, K., Zysset, S., Mueller, K., Friederici, A. D., & Koelsch, S. (2011). Neuroarchitecture of verbal and tonal working memory in nonmusicians and musicians. *Human Brain Mapping*, *32*, 771-783.

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.

Schwartze, M., Keller, P. E., Patel, A. D., & Kotz, S. A. (2011). The impact of basal ganglia lesions on sensorimotor synchronization, spontaneous motor tempo, and the detection of tempo changes. *Behavioural Brain Research*, *216(2)*, 685-691.

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(12)*, 2400-2406.

Scott, S. K., & Johnsrude, I. S. (2003). The neuroanatomical and functional organization of speech perception. *Trends in Neurosciences*, *26(2)*, 100-107.

Seidner, W., & Wendler, J. (1978). *Die Sängerstimme*. Berlin: Henschel Verlag.

Selkirk, E. O. (1984). *Phonology and syntax. The relation between sound and structure*. Cambridge, MA: MIT Press.

Sergent, J. (1993). Music, the brain and Ravel. *Trends in Neurosciences*, *16(5)*, 168-172.

Shafritz, K. M., Gore, J. C., & Marois, R. (2002). The role of the parietal cortex in visual feature binding. *Proceedings of the National Academy of Sciences USA*, *99(16)*, 10917-10922.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, *270*, 303-304.

Signoret, J. L., Van Eeckhout, P., Poncet, M., & Castaigne, P. (1987). [Aphasia without amusia in a blind organist. Verbal alexia-agraphia without musical alexia-agraphia in braille] Aphasie sans amusie chez un organiste aveugle. Alexie-agraphie verbale sans alexie-agraphie musicale en braille. *Review Neurology*, *143*, 172-181.

Smith, A. (1966). Speech and other functions after left (dominant) hemispherectomy. *Journal of Neurology, Neurosurgery, and Psychiatry*, *29*, 467.

Sparks, R. W., & Deck, J. W. (1994). Melodic intonation therapy. In R. Chapey (Ed.), *Language intervention strategies in adult aphasia* (p. 368-379). Baltimore: Williams & Wilkins.

Sparks, R. W., Helm, N. A., & Albert, M. L. (1974). Aphasia rehabilitation resulting from melodic intonation therapy. *Cortex*, *10*, 303-316.

Specht, K., Osnes, B., & Hugdahl, K. (2009). Detection of differential speech-specific processes in the temporal lobe using fMRI and a dynamic "sound morphing" technique. *Human Brain Mapping*, *30*, 3436-3444.

Stahl, B., Kotz, S. A., Henseler, I., Turner, R., & Geyer, S. (2011). Rhythm in disguise: Why singing may not hold the key to recovery from aphasia. *Brain*, *134*, 3083-3093.

Steele, J. (1775). *An essay towards establishing the melody and measure of speech to be expressed and perpetuated by peculiar symbols*. Menston: The Scholar Press, 1969. [Reprint].

Steinbeis, N., & Koelsch, S. (2008a). Comparing the processing of music and language meaning using EEG and fMRI provides evidence for similar and distinct neural representations. *Plos One*, *3*, 1-7.

Steinbeis, N., & Koelsch, S. (2008b). Shared neural resources between music and language indicate semantic processing of musical tension-resolution patterns. *Cerebral Cortex*, *18*, 1169-1178.

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.

Steinhauer, K., & Friederici, A. D. (2001). Prosodic boundaries, comma rules, and brain responses: The closure positive shift in ERPs as a universal marker for prosodic phrasing in listeners and readers. *Journal of Psycholinguistic Research*, *30(3)*, 267-295.

Steinke, W. R., Cuddy, L. L., & Jakobson, L. S. (2001). Dissociations among functional subsystems governing melody recognition after right-hemisphere damage. *Cognitive Neuropsychology*, *18*, 411-437.

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.

Stewart, L., von Kriegstein, K., Warren, D., & Griffiths, T. D. (2006). Music and the brain: Disorders of musical listening. *Brain*, *129*, 2533-2553.

Stewart, L., Walsh, V., Frith, U., & Rothwell, J. (2001). Transcranial magnetic stimulation produces speech arrest but not song arrest. *Annals of the New York Academy of Sciences*, 433-435.

Stoodley, C. J., & Schmahmann, J. D. (2009). Functional topography in the human cerebellum: A meta-analysis of neurimaging studies. *NeuroImage*, *44*, 489-501.

Straube, T., Schulz, A., Geipel, K., Mentzel, H.-J., & Miltner, W. H. (2008). Dissociations between singing and speaking in expressive aphasia: The role of song familiarity. *Neuropsychologia*, *46*, 1505-1512.

Sundberg, J. (1970). Formant structure and articulation of spoken and sung vowels. *Folia Phoniatrica et Logopaedica*, *22*, 28-48.

Sundberg, J. (1996). The human voice. In R. Greger & U. Windhorst (Eds.), *Comprehensive human physiology* (p. 1096-1104). Berlin: Spinger.

Sundberg, J. (2001). How vocal sounds are made. In J. Potter (Ed.), *Cambridge companion to singing* (p. 204-247). Cambridge, MA: Cambridge University Press.

Swain, J. P. (1997). *Musical languages*. UK: Norton.

Terao, Y., Mizuno, T., Shindoh, M., Sakurai, Y., Ugawa, Y., & Kobayashi, S. (2006). Vocal amusia in a professional tango singer due to a right superior temporal cortex infarction. *Neuropsychologia*, *44*, 479-488.

Tervaniemi, M., & Huotilainen, M. (2003). The promises of change-related brain potentials in cognitive neuroscience of music. *Annals of the New York Academy of Sciences*, *999*, 29-39.

Tervaniemi, M., Schröger, E., Saher, M., & Näätänen, R. (2000). Effects of spectral complexity and sound duration on automatic complex-sound pitch processing in humans - a mismatch negativity study. *Neuroscience Letters*, *290(1)*, 66-70.

Thompson, W. F., Schellenberg, E. G., & Husain, G. (2003). Perceiving prosody in speech. Effects of music lessons. *Annals of the New York Academy of Sciences*, *999*, 530-532.

Thompson-Schill, S., D'Esposito, M., Aguirre, G. K., & Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A reevaluation. *Proceedings of the National Academy of Sciences of the United States of America*, *94*, 14792-14797.

Thurman, L., & Welch, G. F. (2000). *Bodymind and voice: Foundations of voice education*. Iowa: National Centre for Voice and Speech.

Tierney, A., Dick, F., Deutsch, D., & Sereno, M. (2012). Speech versus song: Multiple pitch-sensitive areas revealed by a naturally occurring musical illusion. *Cerebral Cortex*.

Tillmann, B., & Bharucha, J. J. (2002). Effect of harmonic relatedness on the detection of temporal asynchronies. *Perception & Psychophysics*, *64*(4), 640-649.

Tillmann, B., Gosselin, N., Bigand, E., & Peretz, I. (2012). Priming paradigm reveals harmonic structure processing in congenital amusia. *Cortex*, *48(8)*, 1073-1078.

Tillmann, B., Koelsch, S., Escoffier, N., Bigand, E., Lalitte, P., Friederici, A. D., et al. (2006). Cognitive priming in sung and instrumental music: Activation of inferior frontal cortex. *NeuroImage*, *31*(4), 1771-82.

Tillmann, B., Peretz, I., Bigand, E., & Gosselin, N. (2007). Harmonic priming in an amusic patient: The power of implicit tasks. *Cognitive Neuropsychology*, *24(6)*, 603-622.

Tillmann, B., Schulze, K., & Foxton, J. M. (2009). Congenital amusia: A short-term memory deficit for non-verbal, but not verbal sounds. *Brain and Cognition*, *71*, 259-264.

Todd, N. P. M. (1985). A model of expressive timing in tonal music. *Music Perception*, *3*, 33-58.

Tramo, M. J., Shah, G. D., & Braida, L. D. (2002). Functional role of auditory cortex in frequency processing and pitch perception. *Journal of Neurophysiology*, *87(1)*, 122-139.

Trapp, S., & Lepsien, J. (2012). Attentional orienting to mnemonic representations: Reduction of load-sensitive maintenance-related activity in the intraparietal sulcus. *Neuropsychologia*, *50*, 2805-2811.

Trehub, S. E. (2001). Musical predispositions in infancy. *Annals of the New York Academy of Sciences*, *930*, 1-16.

Tremblay-Champoux, A., Dalla Bella, S., Phillips-Silver, J., Lebrun, M. A., & Peretz, I. (2010). Singing proficiency in congenital amusia: Imitation helps. *Cognitive Neuropsychology*, *27(6)*, 463-476.

Tusche, A., Bode, S., & Haynes, J. D. (2010). Neural responses to unattended products predict later consumer choices. *Journal of Neuroscience*, *30(23)*, 8024-8031.

Tyler, L. K., & Marslen-Wilson, W. (2008). Fronto-temporal brain systems supporting spoken language comprehension. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*, 1037-1054.

Tzourio, N., Massioui, F. E., Crivello, F., Joliot, M., Renault, B., & Mazoyer, B. (1997). Functional anatomy of human auditory attention studied with PET. *Neuroimage*, *5(1)*, 63-77.

Vaden, K. I., Muftuler, L. T., & G.Hickok. (2009). Phonological repetition-suppression in bilateral superior temporal sulci. *Neuroimage*, *49*, 1018-1023.

Vigneau, M., Beaucousin, V., Hervë, P. Y., Duffau, H., Crivello, F., Houdë, O., et al. (2006). Meta-analyzing left hemisphere language areas: Phonology, semantics, and sentence processing. *NeuroImage*, *30(4)*, 1414-1432.

Vignolo, L. A. (2003). Music agnosia and auditory agnosia. Dissociations in stroke patients. *Annals of the New York Academy of Sciences*, *999*, 50-57.

Vos, P. G., & Troost, J. M. (1989). Ascending and descending melodic intervals: Statistical findings and their perceptual relevance. *Music Perception*, *6*, 383-396.

Wallin, N., Merker, B., & Brown, S. (2000). *The origins of music*. Cambridge, MA: MIT Press.

Warren, J. D., & Griffiths, T. D. (2003). Distinct mechanisms for processing spatial sequences and pitch sequences in the human auditory brain. *Journal of Neuroscience*, *23*(13), 5799-5804.

Warren, J. D., Jennings, A. R., & Griffiths, T. D. (2005). Analysis of the spectral envelope of sounds by the human brain. *NeuroImage*, *24*, 1052-1057.

Warrier, C. M., & Zatorre, R. J. (2004). Right temporal cortex is critical for utilization of melodic contextual cues in a pitch constancy task. *Brain*, *127(7)*, 1616-1625. (patient study)

Wechsler, D. (1987). *Wechsler memory scale - revised (WMS-R)*. San Antonio, CA: Psychological Corporation/Harcourt Brace Jovanovich.

Wei, P., Müller, H. J., Pollmann, S., & Zhou, X. (2011). Neural correlates of binding features within- or cross-dimensions in visual conjunction search: An fMRIbr study. *NeuroImage*, *57(1)*, 235-241.

Welch, G. F. (2005). Singing as communication. In D. Miell, R. MacDonald, & D. J. Hargreaves (Eds.), *Musical communication* (p. 239-258). Oxford: Oxford University Press.

Wernicke, C. (1874). *Der aphasische Symtomencomplex. Eine psychologische Studie auf anatomischer Basis*. Berlin: Spinger.

Westbury, C. F., Zatorre, R. J., & Evans, A. C. (1999). Quantitative variability in the planum temporale: A probability map. *Cerebral Cortex*, *9*, 392-405.

Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., & Price, C. J. (1992). Segmental durations in the vicinity of prosodic boundaries. *Journal of the Acoustic Society of America*, *91*, 1707-1717.

Wildgruber, D., Ackermann, H., Klose, U., Kardatzki, B., & Grodd, W. (1996). Functional lateralization of speech production at primary motor cortex: A fMRI study. *NeuroReport*, *7*, 2791-2795.

Wildgruber, D., Pihan, H., Ackermann, M., Erb, M., & Grodd, W. (2002). Dynamic brain activation during processing of emotional intonation: Influence of acoustic parameters, emotional valence, and sex. *NeuroImage*, *15*, 856-869.

Williamson, V. J., & Stewart, L. (2010). Memory for pitch in congenital amusia: Beyond a fine-grained pitch perception problem. *Memory*, *18*, 657-669.

Wilson, S. M., Saygin, A. P., Sereno, M., & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature*, *7*, 701-702.

Winkler, I., Denham, S. L., & Nelken, I. (2009). Modeling the auditory scene: Predictive selection and integration in auditory sentence processing. *Trends in Cognitive Sciences*, *13*, 532-540.

Wong, P. C., Parsons, L. M., Martinez, M., & Diehl, R. L. (2004). The role of the insular cortex in pitch pattern perception: The effect of linguistic contexts. *Journal of Neuroscience*, *24*(41), 9153-60.

Yamadori, A., Osumi, y., Masuhara, S., & Okubo, M. (1977). Preservation of singing in Broca‘s aphasia. *Journal of Neurology, Neurosurgery, and Psychiatry*, *40*, 221-224.

Yoo, S.-S., Lee, C. U., & Choi, B. G. (2001). Human brain mapping of auditory imagery: Event-related functional MRI study. *Neuroreport*, *12*, 3045-3049.

Zacks, J. M. (2008). Neuroimaging studies of mental rotation: A meta-analysis and review. *Journal of Cognitive Neuroscience*, *20(1)*, 1-19.

Zaehle, T., Wüstenberg, T., Meyer, M., & Jäncke, L. (2004). Evidence for rapid auditory perception as the foundation of speech processing: A sparse temporal sampling fMRI study. *European Journal of Neuroscience*, *20*, 2447-2456.

Zarate, J. M., Wood, S., & Zatorre, R. J. (2010). Neural networks involved in voluntary and involuntary vocal pitch regulation in experienced singers. *Neuropsychologia*, *48(2)*, 607-618.

Zarate, J. M., & Zatorre, R. J. (2008). Experience-dependent neural substrates involved in vocal pitch regulation during singing. *NeuroImage*, *40(4)*, 1871-1887.

Zatorre, R. J. (1985). Discrimination and recognition of tonal melodies after unilateral cerebral excisions. *Neuropsychologia*, *23(1)*, 31-41.

Zatorre, R. J. (1988). Pitch perception of complex tones and human temporal-lobe functions. *Journal of the Acoustic Society of America*, *84(2)*, 566-572. (patient study)

Zatorre, R. J. (1998). Functional specialization of human auditory cortex for musical processing. *Brain*, *121(10)*, 1817-1818.

Zatorre, R. J., & Baum, S. R. (2012). Musical melody and speech intonation: Singing a different tune? *Plos Biology*, *10(7)*, e1001372.

Zatorre, R. J., & Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cerebral Cortex*, *11*, 946-953.

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.

Zatorre, R. J., Evans, A. C., & Meyer, E. (1994). Neural mechanisms underlying melodic perception and memory for pitch. *Journal of Neuroscience*, *14(4)*, 1908-19.

Zatorre, R. J., Evans, A. C., Meyer, E., & Gjedde, A. (1992). Lateralization of phonetic and pitch discrimination in speech processing. *Science*, *256*, 846-849.

Zatorre, R. J., Halpern, A. R., & Bouffard, M. (2009). Mental reversal of imagined melodies: A role for the posterior parietal cortex. *Journal of Cognitive Neuroscience*, *22(4)*, 775-789.

Zatorre, R. J., Halpern, A. R., & Bouffard, M. (2010). Mental reversal of imagined melodies: A role for the posterior parietal cortex. *Journal of Cognitive Neuroscience*, *22(4)*, 775-789.

Zatorre, R. J., Halpern, A. R., Perry, D. W., Meyer, E., & Evans, A. C. (1996). Hearing the mind's ear: A PET investigation of musical imagery and perception. *Journal of Cognitive Neuroscience*, *8*, 29-46.

Zatorre, R. J., Ptito, A., & Villemure, J. G. (1995). Preserved auditory spatial localization following cerebral hemispherectomy. *Brain*, *118(4)*, 879-889.

Zatorre, R. J., & Samson, S. (1991). Role of the right temporal neocortex in retention of pitch in auditory short-term memory. *Brain*, *114*, 2403-2417.

Zekveld, A. A., Heslenfeld, D. J., Festen, J. M., & Schoonhoven, R. (2006). Top-down and bottom-up processes in speech comprehension. *Neuroimage*, *32*, 1826-1836.

List of Abbreviations

AAT	Aachener Aphasie Test
AC	Auditory Cortex
ACC	Anterior Cingulate Cortex
ANOVA	Analysis of Variance
aSTG	Anterior Superior Temporal Gyrus
BA	Brodmann Area
BG	Basal Ganglia
BOLD	Blood Oxigenation Level Dependency
EEG	Electroencephalography
ERP	Event-related Potential
HG	Heschl's Gyrus
hIP	Human Intraparietal Area
IFG	Inferior Frontal Gyrus
IPC/IPL	Inferior Parietal Cortex/Lobule
IPL	Inferior Parietal Lobe
IPS	Intraparietal Sulcus
ITG	Inferior Temporal Gyrus
MBEA	Montreal Battery of Evaluation of Amusia
MEG	Middle Temporal Gyrus
MIT	Melodic Intonation Therapy
MRI	Magnetic Resonance Imaging
MTG	Middle Temporal Gyrus
MVPA	Multivariate Pattern Analysis

PAC	Primary Auditory Cortx
PET	Positron Emission Tomography
PMC	Premotor Cortex
PP	Planum Polare
pSTG	Posterior Superior Temporal Gyrus
PT	Planum Temporale
rCBF	Regional Cerebral Blood Flow
ROI	Region of Interest
SMA	Supplementary Motor Area
SMG	Supramarginal Gyrus
SNG	Song
SPC/SPL	Superior Parietal Cortex/Lobule
SPK	Speech
STG	Superior Temporal Gyrurs
STS	Superior Temporal Sulcus
TL	Temporal Lobe
TMS	Transcranial Magnetic Stimulation
UVA	Univariate Analysis
VLFC	Ventrolateral Frontal Cortex
WM	Working Memory
WMS	Wechsler Memory Scale

List of Figures

List of Tables

Summary

The interrelationship of music and language has been of scholarly interest for years. Singing is where both domains meet – song and speech are both vocal, bear linguistic and lexical content and carry pitch contour to convey meaning. Despite sharing similarities, apparently there are equally salient differences between song and speech; otherwise it would not be so easy to instantaneously differentiate whether someone is singing or speaking. In this context, the general research question of the present dissertation is: How does our brain code for differences in song and speech? This question will be tackled in more detail by systematically utilizing approaches covering (i) functional imaging in healthy adults, (ii) systemic perception alterations in lesion patients, and (iii) considerations along a theoretical framework of musical disorder.

Given the aforementioned degree of feature-similarity in song and speech, it is no surprise that the brain activity associated with the perception of song and speech demonstrated some overlap. Previous research, however, provides evidence for hemispheric specialization, reflecting the traditional view of music processing being predominantly lateralized to the right and language lateralized to the left hemisphere (Riecker et al., 2000; Jeffries et al., 2003; Callan et al., 2006). The specifics of a putative left/right separation remain unresolved as both sung and spoken sentences (i) express meaning through words thus bearing linguistic information, (ii) contain melodic information and (iii) have underlying rhythm patterns. Phenomenologically, song relies on melodies with discrete pitch relations and typically shows discrete rhythmic onsets at integer multiples of the underlying metric beat or its subdivisions. Speech, by contrast, does not show these discrete relations, neither in pitch pattern (which shows continuous transition or 'gliding' pitch) nor in the periodicity of rhythmic/metric timing. However, spectral and temporal features are known to be processed differently, with spectral information being processed in the right and temporal information in the left hemisphere (e.g., Zatorre et al., 2002). The possibility that different spectral and temporal aspects of song and speech lead to different lateralization patterns calls for an experiment that carefully dissects these aspects in order to draw a conclusive picture on the neural distinction of song and speech perception.

Under normal conditions it is easy to differentiate between song and speech. Yet, when the auditory perception is distorted, discrimination performance may show a notable drop. Distortion of the sensory percept can be a result of one of two changes: (i) modification of the acoustical input, which can be manipulated systematically in order to artificially induce a performance drop in discrimination tasks; (ii) impairments of the perception system itself, as a result of brain lesions or congenital amusia (i.e. a music perception disability). Both options will be exploited in the current dissertation. Introducing stimuli that are manipulated in ways to bridge the continuum of acoustic parameters between song and speech (being perceived as 'halfway between') in a forced choice paradigm allows investigating to what extent people show a bias in classifying those stimuli as song or speech. By testing patients with focal lesions, the specific role of the temporal lobe in speech and song perception (for which imaging studies suggested overlap) may be investigated. Moreover, individuals with congenital amusia can help to understand the influence of music processing on song and speech discrimination.

Experiments and Discussion

This dissertation sets out to investigate song and speech perception on the level of their underlying constituents, words and pitch patterns (Experiment 1) and the special role of the temporal lobe and music processing abilities on the discrimination of song and speech (Experiment 2).

Experiment 1 aimed to gain further insights into the neural basis of song and speech perception by means of a novel hierarchical stimulus paradigm in a functional magnetic resonance imaging (fMRI) study. The stimuli constitute six conditions and were organized in a subtractive hierarchy with three levels: The first level comprised sung and spoken utterances with identical text, the second level excluded the text and only contained the underlying pitch-rhythm patterns (hummed melodies) of the sung and spoken utterances, the third level further excluded pitch information and merely contained the underlying rhythm patterns (hummed without pitch variation). The stimulus constellation allowed for (i) direct comparisons of sung and spoken sentences as well as the pitch patterns of both (melodic vs. prosodic) and (ii) investigation of 'word' processing by parceling out the influence of pitch and rhythms, and the investigation of pitch processing by parceling out the influence of rhythm.

The combined results suggest that the perception of song and speech engages areas in the auditory cortices and motor-related regions, such as premotor cortices, supplementary motor area and cerebellum, suggesting a shared sensorimotor network engaged in subvocalic rehearsal and auditory-to-motor mapping.

Speech processing engages a left-lateralized fronto-temporal network, including areas typically associated with linguistic processing such as the superior temporal gyrus (STG) and sulcus (STS; bilaterally engaged), the posterior STS/middle temporal gyrus (MTG) and the inferior frontal gyrus (IFG). The left IFG was found for processing words in speech, most likely reflecting the focused processing of segmental linguistic information, such as lexical semantics and syntax (for a review see Bookheimer, 2002; Friederici, 2002, 2011), in order to decode the message of the perceived sentence. Furthermore, the left IFG coded for pitch patterns in speech. The lateralization for pitch patterns in speech to the left hemisphere is most probably determined by the 'speech-relatedness' of the stimulus, i.e. the function of pitch in contrasts to its form (i.e. pitch modulations in speech; Gandour et al., 2000). The posterior STS/MTG was activated for hummed prosody contrasted with hummed musical melody. The MTG, a region supporting the sound-to-meaning mapping (e.g., Hickok & Poeppel, 2007), might reflect the participants' attempt to add words in order to make out the meaning in the prosodic stimulus. This process was obviously not necessary in hummed musical melody as it is independent from words (i.e. instrumental music).

Song on the other hand engages a bihemispheric network with a right hemispheric weighting. Areas involved in the song network were the STG/STS bilaterally, the anterior STG (aSTG) bilaterally, the anterior intraparietal sulcus (IPS) bilaterally (stronger on the right) and the right IFG. A direct comparison of sung and spoken sentences yielded a bilateral activation in the aSTG coding for song. In the STG a pitch processing hierarchy toward the anterior portion has been suggested (e.g., Patterson et al., 2002). Hence, the activation can be interpreted as representing the musical aspects in song compared to speech. In anticipation of Experiment 2, this interpretation fits with song and speech processing deficits correlating with lesions in the aSTG.

The IPS responded specifically to the discrete pitch relations in song. It was found in the direct comparison of musical melody and prosody, i.e. discrete pitch vs. gliding pitch, and for musical melody corrected for the underlying rhythm pattern, i.e. discrete pitch vs. monotonous pitch. This finding is in line with a number of studies on musical pitch processing (e.g., Zatorre et al., 1994, 2009, 2010), thus supporting the idea originating from visual studies concerning the role of the IPS in the systematic transformation of stimulus representations, dependent upon precise relationships among its elements (Foster & Zatorre, 2010). Correspondingly, the IPS was found to be involved in perceiving transposed melodies (Foster & Zatorre, 2010), the categorical perception of major and minor chords (Klein & Zatorre, 2011) and currently discrete pitch relations – all requiring relative pitch encoding which is a fundamental element in music perception.

The right IFG has been shown to play a crucial role in music processing (e.g., Zatorre et al., 1994; Koelsch & Siebel, 2005; Schmithorst, 2005; Tillmann et al., 2006), supporting the current finding of involvement in pitch pattern processing in song. Interestingly, the right IFG was also found for word processing in song. This might be due to the specific way sung words are vocalized – as for example characterized by a lengthening of vowels. The right hemisphere processes information at broader time scales than the left (e.g., Poeppel et al., 2004), which might explain why the right IFG showed specific sensitivity to sung words.

Taken together, the current dissertation provides evidence for a left-lateralized network for speech and a more bihemispheric network for song processing. Language-related areas coded for speech, such as the STG/STS, the left MTG and the left IFG. Song on the other hand was found to recruit pitch-related areas, such as the aSTG due to pitch/melody-complexity, the IPS due to pitch-relations, and the right IFG due to melodic structure processing. Despite the aSTG coding for song, the STG/STS was involved in both song and speech processing, not showing a more fine-grained discriminating picture. Therefore, Experiment 2 tested to what extent lesions in the temporal lobe disrupt song and speech processing. Furthermore, the influence on musical ability was of interest as the left and right IFG involvement seemed to be caused by the function (being song- or speech-related) of the specific stimulus.

Experiment 2 aimed at gaining further insight into the specific role of the temporal lobe as well as the influence of music processing deficits on song and speech classification and discrimination. In a behavioral pilot experiment, participants were tested on their subjective classification of sung and spoken stimuli, which exhibited unambiguous, i.e. clear sung and spoken sentences, and ambiguous characteristics, i.e. stimuli 'halfway between' song and speech. This experiment was conducted with patients exhibiting focal lesions in the temporal lobe (TL), a group of congenital amusics (music disability, 'amusics' hereafter) and a control group. Crucially, half of the TL group also exhibited impaired music processing ('acquired amusics').

Both the TL patients and amusics had no problems in classifying distinct and unambiguous stimuli as song or speech but showed a response bias to classifying ambiguous stimuli as song. In contrast, healthy controls exhibited ratings in a balanced manner for song and speech. This probably reflects the use of a cognitive strategy that is indirectly related to their actual music processing deficit leading to their disturbed song perception – in this respect they made their decision based on an over-compensatory strategy. More specifically, a lesion in the left or right temporal lobe and amusic symptoms altered the participants' perception of song and speech. Since previous studies with amusics have suggested amusia to be an impairment of the music domain only and speech to be unaffected (e.g., Peretz et al., 2012), a two-step mechanism may be assumed: the impaired music perception affects the amusics' song perception, possibly lowering their tolerance of what is considered speech in comparison to song perception in individuals with unimpaired music perception abilities.

As a consequence, as soon as stimuli show a slight discrepancy from their mental model of speech in an utterance, the forced-choice task would lead amusics to classify these stimuli as ‘non-speech’, i.e. song. Moreover, the acquired amusics (from the TL group), together with cases reported in the literature (e.g., Peretz et al., 1994; Ayotte et al., 2000), exhibited a lesion overlap in the aSTG and the temporal pole, which can therefore be associated with the music processing deficit. Interestingly, the aSTG was the only distinct area in the temporal lobe found in the fMRI study for song and speech discrimination reflecting song processing. Additionally, the activation was found to be bilateral and the aSTG lesion was found in the right and the left hemisphere. Therefore it is confirmed by lesion studies that song is represented in the aSTG bilaterally – compared to speech.

Conclusion

This dissertation exceeded the state-of-the-art research in song and speech by investigating song and speech perception on the levels of words, pitch patterns and rhythm, which lead to the following picture of song and speech networks: song engages a right hemispheric temporo-frontal and bihemispheric temporo-parietal network. Speech engages a left hemispheric temporo-frontal network. Three regions code specifically for song: the bilateral aSTG, the IPS and the right IFG, while the left IFG specifically codes for speech. Right and left temporal lobe lesions as well as impaired music perception lead to a disturbed perception of song and speech, underlining the crucial role of an intact STG, specifically the aSTG, in song and speech perception.

Curriculum Vitae

Name: Julia Merrill (Groh)
Date of Birth: 18.09.1981
Place of Birth: Kassel, Germany

2008–2012	*Ph.D. student* at the Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany
2005–2008	*Diploma in Speech Science* at the Martin Luther University Halle-Wittenberg, Halle, Germany
2004–2005	Courses in German language and literature studies and musicology at the Heinrich Heine University, Düsseldorf, Germany
2000–2004	*Diploma in Church Music* at the Hochschule für Kirchenmusik, Heidelberg, Germany
2000	Abitur at the Jakob-Grimm-Schule Kassel, Germany

Selbständigkeitserklärung

Hiermit erkläre ich, dass die vorliegende Arbeit ohne unzulässige Hilfe und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt wurde und dass die aus fremden Quellen direkt oder indirekt übernommenen Gedanken in der Arbeit als solche kenntlich gemacht worden sind.

Julia Merrill
Leipzig, 18. September 2012

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 experimentalpsychologische 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 katecholaminerge 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 Benennungsalternative 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

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 Rüschemeyer
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
Aufgabeninhibition als Mechanismus der Konfliktreduktion zwischen Aufgabenrepräsentationen

76 Juliane Hofmann
Phonological, Morphological, and Semantic Aspects of Grammatical Gender Processing in German

77 Petra Augurzky
Attaching Relative Clauses in German – The Role of Implicit and Explicit Prosody in Sentence Processing

78 Uta Wolfensteller
Habituelle und arbiträre sensomotorische 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 ist? 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 Şükrü Barış 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 sensomotorischen System – Lerninduzierte Veränderungen in der Struktur und Funktion des menschlichen Gehirns

137 Patricia Garrido Vásquez
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