Check for updates

sempre:
Society for Education, Music and Psychology Research
MUSIC & SCIENCE

Conference Report

Music & Science
Volume 1: 1–17
© The Author(s) 2018
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/2059204318798831
journals.sagepub.com/home/mns
SAGE

Spring School on Language, Music, and Cognition: Organizing Events in Time

Rie Asano[1], Pia Bornus[1], Justin T. Craft[2], Sarah Dolscheid[3], Sarah E. M. Faber[4], Viviana Haase[5], Marvin Heimerich[1], Radha Kopparti[6], Marit Lobben[7], Ayumi M. Osawa[1], Kendra Oudyk[8], Patrick C. Trettenbrein[9,10], Timo Varelmann[1], Simon Wehrle[11], Runa Ya[12], Martine Grice[11] and Kai Vogeley[13,14]

Abstract

The interdisciplinary spring school "Language, music, and cognition: Organizing events in time" was held from February 26 to March 2, 2018 at the Institute of Musicology of the University of Cologne. Language, speech, and music as events in time were explored from different perspectives including evolutionary biology, social cognition, developmental psychology, cognitive neuroscience of speech, language, and communication, as well as computational and biological approaches to language and music. There were 10 lectures, 4 workshops, and 1 student poster session.

Overall, the spring school investigated language and music as neurocognitive systems and focused on a mechanistic approach exploring the neural substrates underlying musical, linguistic, social, and emotional processes and behaviors. In particular, researchers approached questions concerning cognitive processes, computational procedures, and neural mechanisms underlying the temporal organization of language and music, mainly from two perspectives: one was concerned with syntax or structural representations of language and music as neurocognitive systems (i.e., an intrapersonal perspective), while the other emphasized social interaction and emotions in their communicative function (i.e., an interpersonal perspective). The spring school not only acted as a platform for knowledge transfer and exchange but also generated a number of important research questions as challenges for future investigations.

Keywords

Computational approach, development, emotion, evolution, language and music cognition, mechanisms, mentalizing, prosody, social cognition, structure building

Submission date: 30 June 2018; Acceptance date: 2 August 2018

[1] Institute of Musicology, University of Cologne, Germany
[2] Department of Linguistics, University of Michigan, USA
[3] Department of Rehabilitation and Special Education, University of Cologne, Germany
[4] Rotman Research Institute, University of Toronto, Canada
[5] Institute for Philosophy II, Ruhr University Bochum, Germany
[6] Research Centre For Machine Learning, Department of Computer Science, City, University of London, UK
[7] Department of Psychology, University of Oslo, Norway
[8] Department of Music, Art, and Culture Studies, University of Jyväskylä, Finland
[9] Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, Germany
[10] International Max Planck Research School on Neuroscience of Communication: Function, Structure, and Plasticity (IMPRS NeuroCom), Germany
[11] IfL Phonetik, University of Cologne, Germany
[12] Institute for Musicology and Media Science, Humboldt-Universität zu Berlin, Germany
[13] Department of Psychiatry, University Hospital Cologne, Germany
[14] Institute for Neuroscience and Medicine—Cognitive Neuroscience (INM3), Research Center Juelich, Germany

Corresponding author:
Rie Asano, University of Cologne, Albertus-Magnus-Platz, 50923 Cologne, Germany.
Email: rie.asano@uni-koeln.de

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (http://www.creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).

Introduction

The spring school "Language, music, and cognition: Organizing events in time" was held from February 26 to March 2, 2018 at the Institute of Musicology, University of Cologne, Germany, as a part of a two-year education and research project titled *Language and music in cognition*.[1] Language and music cognition research involves a wide range of disciplines including musicology, linguistics, psychology, neuroscience, computer science, and biology, and thus requires close collaboration among different research fields (e.g., Arbib, 2013; Bannan, 2012; Honing, 2018; Patel, 2008; Peretz & Zatorre, 2003; Rebuschat, Rohrmeier, Hawkins, & Cross, 2012; Wallin, Merker, & Brown, 2000). Although language and music cognition research has been gaining ground, there is still little opportunity for students and young researchers to acquire knowledge of language and music cognition research in an interdisciplinary setting. Therefore, the spring school aimed at acting as a platform for knowledge transfer and exchange in this relatively new interdisciplinary research area.

This interdisciplinary spring school focused on language, speech, and music as "ways of ordering events in time" (Arbib, Verschure, & Seifert, 2013, p. 382). Words are integrated sequentially to understand a sentence, notes are integrated to make sense of a musical phrase, movements are integrated to generate a goal-directed action or behavior, and individual sentences or phrases are incorporated into the dynamics of conversational or joint musical co-construction. While this ordering of events in time may seem trivial, a number of questions arise that need addressing: What are the computational, cognitive, and neural mechanisms underlying temporal organization? How does the ability of temporal organization develop in ontogeny? How did the mechanisms underlying temporal organization evolve? What is the function (e.g., adaptive significance) of temporal organization?

These questions were discussed in an interdisciplinary fashion from the perspectives of the various contributing scientific disciplines. In particular, the current spring school investigated temporal organization in language, speech, and music by focusing on syntax, prosody (rhythm and pitch), action, parsing, and organization of verbal and nonverbal communication such as turn-taking. To explore the biological foundations of temporal organization in a full range, the scope of discussion was extended to other species such as non-human primates and birds. Importantly, all topics were discussed in light of comparative music and language research.

The main scientific program was organized around the following five topics: 1) comparative evolutionary biology; 2) social cognition; 3) developmental psychology; 4) cognitive neuroscience of speech, language, and communication; and 5) computational and biological approaches to language and music. Those topics were chosen on the basis of Tinbergen's four questions—causation (or mechanism), ontogeny, phylogeny, and function (or adaptive value) (Tinbergen, 1963)—as well as the "fifth question" concerning socio-affective and socio-cultural aspects of language and music (Fitch, 2010, 2015, 2018). Each topic was assigned to a group work session, two lectures, and a plenary discussion session. The lectures were given by Cedric Boeckx, Ian Cross, Maria Teresa Guasti, Barbara Höhle, Mathis Jording, Sonja Kotz, Chris Petkov, Daniela Sammler, Constance Scharff, Uwe Seifert, and David Vogel (for information on the lecturers and the topics presented, see Table 1 and the following section of this article).

In addition, there were four workshops and one student poster session. The workshops provided practical, hands-on activities such as programming computer simulation, tinkering hardware devices, building scientific hypotheses together, and playing traditional Japanese music instruments. The workshops were given by Rie Asano, Cedric Boeckx, Andreas Gernemann-Paulsen, Marvin Heimerich, Genta Toya, and the Cologne Gagaku Ensemble based at the University of Cologne (for more information see Table 1 and the "Workshops and posters" section). In the poster session, there were 23 presentations (see "Workshops and posters" section).

The spring school was attended by 73 participants, from undergraduate, graduate, and doctoral students to postdoc researchers and university faculty from 16 different countries. Their research backgrounds were wide-ranging, covering linguistics, musicology, neuroscience, psychology, cognitive science, and computer science. The organizing committee comprised Aria Adli, Rie Asano, Martine Grice, Marvin Heimerich, Sebastian Lammers, Doris Mücke, Ayumi Osawa, Lena Pagel, Martina Penke, Uwe Seifert, Volker Struckmeier, Sarah Verlage, and Kai Vogeley.

Lectures

Evolutionary biology

What makes us human? Structured sequence learning, language evolution, and the primate brain (Chris Petkov). Aiming at answering the fundamental anthropological question, Petkov reminded us that while many animals communicate, only humans have language. Given that our cognitive capacities are the product of evolution, he first focused on identifying select aspects of human capacities in the animal world. For example, while all attempts to teach non-human primates to speak have failed miserably, a project that sought to teach American Sign Language to a chimpanzee, aptly named Nim Chimpsky, demonstrated that Nim could learn more than 100 different symbolic associations (Terrace, Petitto, Sanders, & Bever, 1979). However, Nim's presumably "sentence-like" multi-sign utterances turned out to only superficially resemble an early stage of language acquisition in human infants, and re-analysis of this data with more rigorous statistical methods has since confirmed the lack of the expected productivity of a

Table 1. Overview of Lectures and Workshops.

Lecturers	Lecture titles
Professor Chris Petkov (Comparative Neuropsychology, Newcastle University, UK)	What makes us human? Structured sequence learning, language evolution and the primate brain
Professor Constance Scharff (Animal Behavior, Free University of Berlin, Germany)	Language, music, and birdsong—behavioral, neural and genetic similarities and differences
Professor Ian Cross (Centre for Music and Science, University of Cambridge, UK)	Music, speech, and the relational dimension of social interaction
Mathis Jording and David Vogel (Psychiatry, University Hospital Cologne, Germany)	Neural mechanisms of intersubjectivity
Professor Maria Teresa Guasti (Linguistics and Language Acquisition, University of Milano-Bicocca, Italy)	Predicting from rhythmic and syntactic representations
Professor Barbara Höhle (BabyLAB, University of Potsdam, Germany)	Prosodic cues in early first language acquisition
Professor Sonja Kotz (Neuropsychology, Maastricht University, The Netherlands)	Multimodal emotional speech perception: Why time and attention matters
Dr. Daniela Sammler (Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, Germany)	The melodic mind: neural bases of intonation in speech and music
Professor Cedric Boeckx (Catalan Institute for Advanced Studies (ICREA) and Universitat de Barcelona, Spain)	Language, music, and the brain: wrestling with granularity mismatch issues
Professor Uwe Seifert (Systematic Musicology, University of Cologne, Germany)	On musicology and computation
Workshop instructors	**Workshop titles**
Rie Asano (Systematic Musicology, University of Cologne) and Professor Cedric Boeckx	Evolution of vocal learning and rhythm
Genta Toya (JAIST, Japan) and Marvin Heimerich (Systematic Musicology, University of Cologne, Germany)	Evolutionary simulation using NetLogo
Cologne Gagaku Ensemble	Gagaku workshop
Andreas Gernemann-Paulsen (Systematic Musicology, University of Cologne, Germany)	Introduction to physical computing with Arduino for New Media Art in the context of empirical cognitive musicology

rule-based grammar (Yang, 2013). This tentatively confirms the observed dissociation between linguistic and communicative capacities in humans, as language must not be equated with speech (Friederici, Chomsky, Berwick, Moro, & Bolhuis, 2017).

Interestingly, while symbolic learning appears to be widespread in the animal world (consider bees and birds, in addition to Nim and other apes), the ability for vocal imitation is relatively rare and seems to have evolved independently in different species. Monkeys, birds, and whales all have vocal production abilities, and, upon closer inspection, some of their vocalizations turn out to be amazingly complex. For example, the songs of some songbirds exhibit complex structural organization in their phonology but lack the semantic component that enters into human linguistic vocalizations as well as the hierarchical phrase structure characteristic of human language.

This recurrent finding that animals lack syntactic capabilities has led to an enormous interest in artificial grammar learning studies in the field of comparative cognition, in order to explore similarities and differences in sequence and rule-learning abilities across species. According to Petkov, it remains unclear whether any animal can generalize and learn open-ended recursive structures such as A^nB^n in addition to the simpler $(AB)^n$, as current results may reflect problems with testing such abilities with animals. Petkov then turned to the neural substrates that enable sequence and syntax processing, pointing to the importance of inferior frontal regions, especially Brodmann areas 44 and 45 (constituting Broca's region in humans and its monkey homolog). Strikingly, only ventral frontal and opercular regions in both hemispheres are comparably engaged in processing adjacent relationships in macaques and humans (Wilson, Marslen-Wilson, & Petkov, 2017).

In conclusion, Petkov argued that almost all of the smaller building-blocks that make up human language and communicative capacities are not unique to humans (Hauser, Chomsky, & Fitch, 2002). Yet, due to the qualitative difference between human linguistic and animal communication, he left open the possibility that frontal regions in the human cortex may have differentiated to manage with "more complex sequencing demands [of spoken language]," thereby converging with conclusions drawn from independent work on language that has identified these regions and their connectivity profile, especially to the temporal region, as crucial for human syntactic capabilities (Friederici et al., 2017; Goucha, Zaccarella, & Friederici, 2017).

Language, music and birdsong—Behavioral, neural, and genetic similarities and differences (Constance Scharff). Scharff opened

the second lecture asking whether uniquely-human traits require unique components. She went on to discuss human language and music as capacities arising from behavioral, neural, and genetic components that can also be observed in songbirds, parrots and few other animals.

Regarding behavior, birdsong shares some design features with language and music, including the presence of both vocal and auditory channels. Like humans, songbirds are imitative vocal learners, depending on auditory feedback and an adult tutor (Williams, 2004) during critical developmental periods (Todt & Geberzahn, 2003). There is also evidence of innate learning constraints on birdsong, as demonstrated by the non-random syllable order of young songbirds tutored with syllable-randomized song (James & Sakata, 2017). Regarding precursors of human syntax, many bird vocalizations exhibit non-random, hierarchical structure in the ordering of song elements (Weiss, Hultsch, Adam, Scharff, & Kipper, 2014). Parrots can reach impressive levels of vocal production learning when imitating referential/propositional signals (Pepperberg, 1987). Interestingly, birds often also integrate song and dance (Dalziell et al., 2013; Soma & Iwama, 2017; Ullrich, Norton, & Scharff, 2016).

Social factors play a key role in music, dance, language, and birdsong alike. Most vocalizations have specific social purposes, for instance fighting and flirting. Like humans, many songbirds engage in vocal turn taking (Hultsch & Todt, 1982), and the young learn better from a live tutor and in the presence of a hearing mother, even if she herself does not sing, consistent with some form of active teaching (Williams, 2004). The presence of conspecifics influences the type and perception of human and songbird vocalizations (Woolley & Doupe, 2008). In the case of birds, the gene expression in the underlying brain areas also changes when males address their song to females (Jarvis, Scharff, Grossman, Ramos, & Nottebohm, 1998).

Diving deeper into the neural similarities between language, music, and birdsong, Scharff pointed out that, within the dedicated neural system for song, specific pathways subserve the auditory, motor, and learning components of song behavior. Petkov and Jarvis (2012) proposed that a similar logic underlies the neural circuits for birdsong and human speech. Accordingly, damage to specific brain areas or abnormal expression of certain genes cause similar behavioral deficits in humans and songbirds. For example, damage to the basal ganglia circuit leads to abnormal repetitions of song elements by songbirds (Kubikova et al., 2015) and to vocal-repetition-related disorders in humans (Ward, Connally, Pliatsikas, Bretherton-Furness, & Watkins, 2015). Comparative genetic research on the FoxP2 gene supports this connection, as mutations in humans result in developmental language and speech disorders (Lai, Fisher, Hurst, Vargha-Khadem, & Monaco, 2001) and experimental downregulation of FoxP2 in a striatal song nucleus disrupts normal song learning (Haesler et al., 2007; Murugan, Harward, Scharff, & Mooney, 2013).

Thus, parallels can be drawn between language, music, and birdsong at behavioral, neural, and genetic levels. Although it has been claimed that non-human animals lack the defining features of language (e.g., Bolhuis, Tattersall, Chomsky, & Berwick, 2014), comparative research may support a continuous view of human–animal communicative differences, particularly since non-human species are still under-studied, especially comparing birdsong with music. Further, such research highlights the gap in our understanding of the link between animal conceptualizations and their sensory-motor interfaces; befittingly, Scharff ended her talk with the Wittgenstein quote "If a lion could talk, we could not understand him" (Wittgenstein, 1953/1958, p. 225).

Social cognition

Music, speech, and the relational dimension of social interaction (Ian Cross). In his lecture, Cross conceptualized music as an interactive, communicative medium. He highlighted the relational dimension of communication and concluded that music and speech constitute overlapping but functionally and culturally differentiable components of the human communicative toolkit.

Music involves participatory activity and its nature is best understood as a communicative medium that both derives from, and is the result of, *reciprocity* and *affiliativeness.* Music's temporal predictability and the perception of music as an *honest signal* facilitate reciprocity and promote a sense of shared experience between participants. The manifestation of simultaneous *floating intentionality* (i.e., a plurality of "aboutnesses" (Cross, 1999)) allows interactants to share experience of musical events with others in idiosyncratically personal terms. This floating intentionality is a central feature of music which enables individuals to assign their own meanings to music without breaching social integrity (Cross, 2011). Thus, music can be considered as a fundamental affiliative mode of interaction that is optimal for managing social uncertainty (Cross, 2014).

Those features make music distinct from speech. Speech usually has shared consensual referentiality between interlocutors (Clark & Brennan, 1991), leading to mutual understanding and coordination of goal-directed joint action by establishing explicit common ground. Music, on the other hand, does not have consensual referentiality, but is instead experienced as simultaneously exhibiting floating intentionality and unmediated meaning. Explicit common ground is therefore unnecessary for successful interaction in music (Cross, 2014). That is, speech privileges the goal-directed, transactional dimension of communicative interaction, which allows the exercise of social power, whereas music privileges the relational dimension, which promotes mutual affiliation.

By focusing on the relational dimension of communication, Cross suggested that human interactions can be

interpreted within either affiliative or power/dominance frames with distinct neurobiological underpinnings. *Power/dominance* and *affiliation* are two aspects of sociality that define human relationships (Dillard, Solomon, & Palmer, 1999). Dominance indicates relational control (i.e., managing the assessments that others may make in respect of one's own behaviors (Dillard et al., 1999; Goffman, 1955)), and is associated with activation in the left prefrontal cortex (PFC) (Quirin et al., 2013). Affiliation is a much more complex construct that facilitates interpersonal attachment and social affect regulation (Coan, 2010; Dillard et al., 1999), correlating with an affiliative network that is largely subcortical and reward-centred (Feldman, 2017; Quirin et al., 2013).

Cross further elaborated the discussion of underlying common temporal processes on naturalistic interaction in speech and music by discussing ongoing research at Cambridge on interaction in spontaneous speech and in music. Preliminary results suggest that music and speech share common features of temporal processes such as periodicity (Hawkins, Cross, & Ogden, 2013), entrainment (Ogden & Hawkins, 2015), and characteristic pitch intervals (Robledo Del Canto, Hawkins, Cross, & Ogden, 2016; Robledo, Hurtado, Prado, Román, & Cornejo, 2016) in their communicative use.

In sum, the relationship between music, speech, and language is best investigated by treating them as interactive media. They are two overlapping and culturally reconfigurable manifestations of an underlying human communicative repertoire.

Neural mechanisms of intersubjectivity (Mathis Jording & David Vogel). Empathy is a fundamental part of intersubjectivity. Jording and Vogel introduced the distinction between affective empathy—re-experiencing of the inner condition of others—and cognitive empathy— recognizing and identifying the inner condition of others. Further, they discussed the distinction between "persons" and "things" in the seminal work by Fritz Heider (1958), suggesting that the perception of persons (or social perception) is probabilistic and concerned mainly with intentionality, whereas the perception of things (or non-social perception) is deterministic and concerned mainly with causality. Experimental research with moving geometric shapes has shown that different brain areas are recruited for these different kinds of perception: the more "personal" a stimulus, the more the mentalizing system is activated, the more "physical" a stimulus, the more the mirror neuron system is activated (Kuzmanovic et al., 2014; Santos, David, Bente, & Vogeley, 2008; Santos et al., 2010; Vogeley, 2017).

The second part of the talk highlighted the importance of nonverbal communication and of social gaze behavior in particular (see also Jording, Hartz, Bente, Schulte-Rüther, & Vogeley, 2018). Jording and Vogel pointed out that nonverbal communication conveys the majority of social meaning in human communication and that it is crucial for discourse and socio-emotional functions. Nonverbal communication is also implicit and unconscious, in contrast to verbal communication. Research was presented showing that for typically developing control persons, perceived likeability increased with the duration of direct gaze (eye contact) by a computer-generated human face, but that this pattern was not observed for participants diagnosed with autism spectrum disorder (ASD). fMRI data indicate a concurrent loss of activation in the "social" regions of the brain in the ASD subjects, corresponding to the mentalizing system (Georgescu et al., 2013; Kuzmanovic et al., 2009).

Subsequently, the so-called *Default Mode Network* (DMN) was introduced, a system of the brain that is active during resting states (i.e., when no particular task demand or sensory input is present and no motor output is required). It has been suggested that the function of the DMN is one of "inner rehearsal" or simulation, continuously reflecting on and optimizing behavior (Vogeley, 2017). As topographical analyses have shown a considerable overlap between the DMN and the mentalizing system, it can be speculated that the tendency to ascribe intentionality and to think about and imagine the perspective of others in itself represents a kind of default mode of the brain.

Finally, the main themes of the talk were tied together in the context of psychopathology, with the focus on ASD. Individuals with ASD both use and perceive social gaze differently, in directing gaze much less to the faces (and eyes) of others and ascribing comparatively less likeability to longer gazes. This is representative of a more general lack of social motivation, and of difficulties in understanding implicit meaning. Taken together with neurophysiological evidence for reduced activation in the mentalizing system, these findings reveal a potential neurophysiological basis for differences of behavior in ASD.

Developmental psychology

Predicting from rhythmic and syntactic representations (Maria Teresa Guasti). Guasti focused on developmental aspects of prediction, a fundamental competence related to comprehension and coordination in language and music that allows anticipation of abstract representations. In order to facilitate prediction, both systems make use of extrapolation of temporal regularities as well as semantic and (morpho)syntactic processing (Patel & Morgan, 2017). Addressing several comparative aspects of prediction development related to rhythm and morphosyntax, Guasti presented outcomes of ongoing behavioral studies by her research team.

The first part of her talk was on principles of rhythmic organization in handwriting and how developmental disorders may compromise them. Two rhythmic principles are utilized in handwriting in order to keep durations of motor acts constant across variations in letter size and writing speed: isochrony (related to the absolute movement duration in writing a word) and homothety (related to the relative durations in writing individual letters). In contrast to

typically developing (TD) children at nine years old, Pagliarini and colleagues (2015) showed that age-matched children with developmental dyslexia (DD) who do not meet the criteria for a diagnosis of dysgraphia systematically violate isochrony and homothety. Also, handwriting measures (average speed, dysfluency, and duration) correlated with reading measures (speed and errors in reading words and non-words, receptive vocabulary), suggesting that impairments of both types of competence may be related to deficits in abstract rhythmic representations affecting prediction. Importantly, TD children do not profit from additional handwriting practice, as they make use of these principles as soon as they start learning handwriting (Pagliarini et al., 2017).

DD also affects temporal prediction in the auditory modality. In an anticipation task using a warning-imperative paradigm, Guasti, Pagliarini, and Stucchi demonstrated that even if a beat is highly predictable from context, the synchronization error of children and adults with DD was larger compared with controls. Since rhythm is used to anticipate, these outcomes suggest that DD is associated with problems in exploiting temporal regularities and anticipating future events in time.

The effect of musical training and modality on anticipatory skills was investigated by Attardo, Guasti, and Stucchi (in prep), showing group differences in auditory and visual beat anticipation between professional classic musicians, professional jazz improvisers, and musically untrained controls. They demonstrated that intense training in music improvisation has a positive effect on temporal prediction in the auditory modality, but not in the visual modality.

The talk closed with findings regarding morphosyntax. Using a picture selection and warning-imperative task measuring reaction time (RT), Persici, Stucchi, and Arosio (2017) investigated effects of combinations of linguistic cues on the prediction of a target noun as a function of age. Altogether, results suggest that young preschoolers (< 5–6 years) make use of phonological cues for prediction of a noun, while older preschoolers and school children (9–11 years) make use of all information available in the linguistic context: grammatical, phonological, and semantic cues. Importantly, RT measures of young preschoolers also correlated with synchronization errors in temporal anticipation tasks. Guasti opened the ensuing discussion by concluding that predicting "what" in language tasks and predicting "when" in rhythm are related.

Prosodic cues in early first language acquisition (Barbara Höhle). The overarching topic of Höhle's lecture was the following question: To what extent is language acquisition (and speech perception) shaped by domain-general auditory principles, and to what extent do language-specific properties of the prosodic system play a role?

In her talk, Höhle addressed this question by focusing on a cross-linguistic comparison between German, a stress-based language, and French, a language without lexical stress. In the first part of her talk, she presented findings showing that the language-specific differences in the rhythmic grouping of German and French also seem to shape infants' word segmentation from very early on (even at the age of 6 months). German-learning infants were sensitive to a trochaic bias as was demonstrated in experiments using the head-turn preference procedure. At the same time, age-matched French-learning infants did not show this bias (Höhle, Bijeljac-Babic, Herold, Weissenborn, & Nazzi, 2009). These findings suggest that the trochaic listening bias is modulated by the specific rhythmic properties of the language acquired by a child.

In the second part of her talk, Höhle addressed the overarching question of domain-general versus language-specific mechanisms by focusing on adult speakers as well as on second language acquisition. She presented a study in which she tested how monolingual speakers of French, German, as well as French learners of German performed in a rhythmic grouping task. In line with her previous findings, she demonstrated cross-linguistic differences in the rhythmic grouping preferences for the two monolingual groups. French second language (L2)-learners of German displayed a similar grouping pattern as the monolingual French-speakers. However, L2-speakers' sensitivity to rhythm was enhanced by musical experience. Crucially, since only the L2-learners but not the monolingual speakers of French benefited from musical experience, this finding suggests that musical experience may influence the acquisition of word stress by French L2-learners of German (see also Boll-Avetisyan, Bhatara, Unger, Nazzi, & Höhle, 2016). The reported results can also be interpreted as an instance of an interrelation between the music and the language system.

Finally, Höhle emphasized that in addition to general auditory principles, language-specific properties influence infants' and adults' speech perception at various levels. She further stressed that a language-specific attunement happens early in development and concluded that this attunement has important consequences for language acquisition.

Cognitive neuroscience of speech, language, and communication

Multimodal emotional speech perception: why time and attention matters (Sonja Kotz). Kotz addressed questions on how we perceive, integrate, and adapt to multiple sensory events, how emotions affect the integration of dynamic sensory events and whether this integration requires attentional resources. Communication is multimodal, involving the body, face, and voice, leading to a sophisticated neural network involved in the prediction and integration of the information provided across the different dimensions. Kotz introduced various existing hypotheses about crossmodal prediction and integration together with a number of studies that addressed them empirically.

Prediction is a substantial part of action and perception in our dynamic environment. Kotz introduced the idea of forward models of prediction according to which there is a brain network that allows us to generate predictions and to assess whether they are fulfilled. This network includes the cerebellum (for the prediction of motor information) and the thalamus (for the prediction of somatosensory information). She raised the question of whether there is a similar system for the integration of auditory and visual information.

The so-called "crossmodal prediction hypothesis" states that in many audiovisual events, including the perception of others' emotions, the visual signal (e.g., lip movement) precedes the auditory signal and, furthermore, predicts where, when, and what type of sound will occur (Jessen & Kotz, 2013; Stekelenburg & Vroomen, 2012). This linearity is also referred to as "jitter of onset" and has impact on the processing and efficiency of the integration of these two information sources. The crossmodal prediction hypothesis is supported by evidence from event-related potential (ERP) studies revealing an alpha suppression (8–13 Hz) occurring around 500 ms before an auditory stimulus, while no such effect was found preceding visual or audiovisual stimuli.

The "early integration hypothesis" assumes that multimodal integration occurs independently of attention (Driver, 2001; Gelder, 2000). For attention, a chain reaction of auditory ERP responses has been observed: Among the early ERPs it is the N1 that occurs in response to repetitive versus changing stimulation—i.e., it is suppressed if a stimulus is expected, indicating that prediction and attention play a role in the perception of complex social signal processing such as multimodal dynamic emotion expressions (Besle, Fort, Delpuech, & Giard, 2004; van Wassenhove, Grant, & Poeppel, 2005; Vroomen & Stekelenburg, 2010). Among the later ERPs are the N2/P3, reflecting the updating of a model of the environment in response to an event that deviates from regularity. Based on the evidence, attention plays a role in multisensory integration. However, it is currently not possible to conclude whether the early integration hypothesis is correct.

In conclusion, crossmodal prediction facilitates the integration of multimodal information enabling faster adaptations in dynamic environments. Attention may furthermore alter this process, but further studies are necessary to determine its role in multimodal integration.

The melodic mind: Neural bases of intonation in speech and music (Daniela Sammler). Sammler's talk began by outlining the different ways meaning can be gleaned from melody and pitch in both language and music, and how this process can be quantified; including neural networks, segmentation of words and melody, and interpersonal dynamics.

In music, pitches are arranged according to music-syntactic rules that are learned simply through exposure to music (see Koelsch, 2011; Rohrmeier, 2011). Knowledge of these rules guides listeners' music perception, shown by early right anterior negativity (ERAN) responses evoked by syntactic violations in music (similar to (early) left anterior negativity ((E)LAN) responses in linguistic violations), localized to regions in the inferior frontal gyrus (IFG; for reviews, see Friederici, 2002; Koelsch, 2011; Koelsch & Siebel, 2005).

Syntax is equally important in music production, and a test of trained pianists reproducing chord sequences from video recordings of pianists' hands (without audio) showed longer reaction times for incongruent chord sequences compared with congruent sequences (Sammler, Novembre, Koelsch, & Keller, 2013). The authors concluded that pianists develop a syntactic representation of the music during play, even without auditory feedback, which anticipates motoric action. When the music deviates from this representation, an error response is triggered, and pianists must re-evaluate their planned action, resulting in increased reaction times. In a follow-up study with and without audio, activation of the IFG was observed in the processing of music-syntactic violations in both visual and auditory tasks, indicating its modality-independent role in music processing (Bianco et al., 2016).

In language, melodic aspects of speech (i.e., prosody) may be used to alter the meaning of an utterance. Sammler continued by introducing a model that proposes that prosody travels along dual dorsal and ventral pathways in the right hemisphere: The dorsal pathway supports subvocal articulation and evaluation of the sound, possibly as part of an action–perception network including areas in the premotor cortex and IFG (Sammler, Grosbras, Anwander, Bestelmeyer, & Belin, 2015). The ventral pathway links sound to meaning, possibly by forming prosodic units or gestalts from the auditory input. Sammler then noted that these right-hemispheric pathways have to communicate with syntactic processing streams in the left hemisphere during natural language comprehension. The corpus callosum—a fiber bundle that connects the temporal lobes of both hemispheres—was identified as a conduit for this binding occurring soon after a speech signal is detected (Sammler, Kotz, Eckstein, Ott, & Friederici, 2010).

Sammler concluded her lecture by detailing how prosodic features can influence perception of social-pragmatic meaning in speech. In a recent study, Hellbernd and Sammler (2016) found that participants were able to reliably classify a speaker's intention in prosodic utterances, irrespective of lexical word meaning. The medial prefrontal cortex (mPFC), posterior cingulate cortex (PCC), amygdala, and the brain region of the so-called temporoparietal junction (TPJ)—areas known from social cognition research—were found to be involved in addition to the ventral prosody pathway (Hellbernd & Sammler, 2018).

In sum, Sammler's talk illustrated the neurocognitive foundation that allows us to extract meaning from melodic and harmonic sequences in speech and music in terms of syntax and pragmatics. Our implicit knowledge about how

pitches are structurally arranged in music guides perception and action by recruiting similar brain areas and provides common ground for audiences and performers. This in turn helps foster mutual understanding and interaction. Prosody employs a dual-pathway network in the right hemisphere that interacts with linguistic processes in the left hemisphere. This network helps structure linguistic information while parsing interpersonal social and pragmatic meaning, recruiting additional brain areas, including the mPFC and amygdala. By recruiting multiple mutually interlinked neural pathways, the "melodic mind" supports nuanced, emotionally complex social interaction and communication.

Computational and biological approaches

Language, music, and the brain: Wrestling with granularity mismatch issues (Cedric Boeckx). Boeckx opened the final day of lectures at the spring school with a presentation on the nature of language and music and their neurobiological correlates. The lecture began with a summary of theoretically motivated neurolinguistic research that has attempted to identify the neural correlates of computational primitives and phrase-structural representations. The emphasis of this discussion was on conceptualizing localization; not on specific individual loci of computation but rather on the dynamic and complex relationships within the Broca-Wernicke network. Boeckx presented research from Angela Friederici's recent publication, *Language in our brain* (Friederici, 2017), where Broca's region (in particular BA44) and its white-matter connection to superior temporal regions were argued to be crucially involved in syntactic computation, in opposition to research which rather emphasizes a significant and potentially crucial role of the temporal cortex in syntactic computation (e.g., Brennan, Stabler, Van Wagenen, Luh, & Hale, 2016).

Before identifying an anatomical locus, Boeckx discussed the necessity of accurately defining the syntactic computation, jokingly referred to it as the "m-word" or *Merge*. Merge, Boeckx argued, has become a term that is misleading as does not refer to a singular entity. Rather, Merge consists of a *grouping step*, which puts two lexical items together and a *labelling step* that designates one of these items as the head of a phrase. Boeckx discussed how the term Merge was used misleadingly in Chomsky (1995) when the term was defined as these grouping and labelling steps but, additionally— and confusingly—the grouping step was also given the name "Merge". This is a potential pitfall as having any ambiguity about the definition of Merge and how it might be mapped onto the brain has the potential to lead toward "paradoxical results and unproductive controversies." Instead, Boeckx argued that by considering a clearly defined two-step (grouping and labelling) Merge, reminiscent of original proposals of phrase structure building utilizing P-markers and T-markers (Chomsky, 1955, 1957), linguists could look to research in memory architecture as having a possible solution for locating where syntactic computation takes place in the brain.

Boeckx proposed that P-markers (the grouping step of Merge) and T-markers (the labelling step of Merge) may utilize two different storage and retrieval mechanisms: The P-markers (groupings of lexical items) were suggested to utilize a stack and T-markers (a cache of items and manipulated symbols used in a derivation) were claimed to utilize a register. Previously, Broca's region has been associated with the stack (Fitch & Martins, 2014) and Wernicke's region with the register (Frankland, 2015; Frankland & Greene, 2015). Thus, an account of Merge (i.e., grouping and labelling) built on those mechanisms must take both regions into consideration. The controversy about where syntactic computation takes places in the brain may simply fall out from interactions between those two storage and retrieval mechanisms in Broca's region and Wernicke's region, which are connected via the arcuate fasciculus (AF). Such a theory could unify current computational and neurobiological proposals, predict developmental and evolutionary trajectories (e.g., the capacity for language and music, along with the ability to make and use tools), and raise new questions about the language faculty being specific to humans.

On musicology and computation (Uwe Seifert). In his talk on musicology and computation, Seifert raised the question of how linguistics, anthropology, cognitive neuroscience, computer science, psychology, philosophy, and social sciences can inform musicology. After dealing with general concerns of musicology and music theory, Seifert highlighted several challenges to the development of a formal theory of music.

The first is the *conceptualization challenge*, with music both as an abstract mathematical form and as a cognitive system, drawing parallelisms with language. The Chomsky hierarchy and concepts from different phases of generative syntax were discussed (Berwick & Chomsky, 2016; Chomsky, 1956, 1957, 1959, 1963, 1965, 1986, 1995) introducing concepts such as *I-language*, *Merge*, *discrete infinity*, *recursion*, and *hierarchical structure*. He argued a need for describing human music capacity in terms of effective procedures; that is, by descriptions falling within the theory of Turing machine computability (Levelt, 2008). In addition, he pointed out that, according to our current knowledge and taking the Church-Turing thesis into account, to date this computational framework provides an epistemological limit for our (explicit and communicable) scientific knowledge (Beeh, 1981).

The second challenge is the *mind—matter challenge*, where the gap between what is processed in our brain and the actual perception (qualia) was explained (Jackendoff, 1987). Seifert emphasized the importance of developing a functional architecture of a computational mind–brain system for bridging this explanatory gap (Levine, 1999)

between mind and brain, which led to the next challenge related to computation.

In answering what a computational framework should look like, Seifert identified three other challenges. One of them is the *affective motivational* challenge, relating to the concepts of musical significance, the functional role of affect and motivation, musical meaning, and the need to describe all of these computationally. Another is the *comparative challenge*, raising questions as to how brains and their neurocognitive functions evolved and how the perception of humans is different from other species (Petkov & Jarvis, 2012). Finally, the *collective mind challenge* emphasizes the importance of social interaction of human brains for cultural evolution (Arbib, 2012) and as an explication (Carnap, 1950) of the grounding of culture and history in a "collective consciousness/mind" or "objective mind" (Dahlhaus, 1971; Hartmann, 1962; Rothacker, 1947), following the tradition of 19th-century German Idealism and the methodological foundation of the "Geisteswissenschaften."

Overall, in explaining the computational challenge, Seifert highlighted the need to have a Turing computable framework (Gallistel & King, 2009) to represent and understand musical mind. He discussed various computational formalisms, starting from simple logical gate functions and McCulloch-Pitts neuron nets through automata theory and formal grammars (Nelson, 1989). He showed the formal equivalence of logical and McCulloch-Pitts nerve nets and how they are formally related to a finite-state transducer (a finite automata with an output). In addition, he discussed the formal relation between regular grammars and finite automata and showed how concepts from automata theory and formal grammar were used by Chomsky in his early work on transformational grammar to formalize and explicate the linguistic concept of grammar.

Some relevant approaches from the history of science of music research were mentioned, as examples the application of these ideas and concepts to formalize music and music theory in the 1970s and 1980s: transformational grammar, phrase structure grammars, and the generative theory of tonal music (Bernstein, 1976; Hughes, 1991; Lerdahl & Jackendoff, 1983; Sundberg & Lindblom, 1991). Seifert pointed out that today there are three important strands using the idea of grammar in music theory: categorial grammar (Steedman, 1996), generative syntax (Rohrmeier, 2011), and construction grammar (Zbikowski, 2017). He explained the rule types which constitute the basis of the four classes of grammars (and languages) of the Chomsky hierarchy and discussed their applications to artificial grammar learning and implicit learning (Fitch & Friederici, 2012; Petersson, Folia, & Hagoort, 2012; Rohrmeier, Zuidema, Wiggins, & Scharff, 2015). The question as to whether music is context-sensitive or context-free was raised at the end as an open question for discussion and methodological doubts concerning the artificial grammar learning paradigm as a useful empirical research method to study the musical mind/brain were raised.

In sum, Seifert's talk highlighted various challenges of computational cognitive musicology and provided a methodological framework for computational modeling of neurocognitive systems such as music and language. He highlighted that one must carefully bear in mind the different explanatory goals in both computational modeling approaches to music, as one must also in relation to music theories, and computational modeling approaches to the human music capacity, as one must in cognitive musicology: The explanatory goal of cognitive musicology is the musical mind/brain—rather than "musics."

Workshops and posters

Evolution of vocal learning and rhythm (Rie Asano & Cedric Boeckx)

Asano started the workshop by summarizing the current state of comparative animal studies. Previous research focused primarily on different species' capacity for vocal production learning and on contrasting other species' features with uniquely human aspects of musical rhythm. Rather than taking an a priori view of human uniqueness, the workshop concentrated on the neural substrates underlying vocal production learning and/or rhythmic capabilities in different species, and whether those two traits are linked in the brain.

According to the Kuyper/Jürgens' hypothesis, vocal production learning has a motor control origin; direct cortico-ambigual connections were exapted (Fitch, 2011) or duplicated (Jarvis, 2004) from the cortico-spinal tract in humans and other vocal production learners, while these connections were only indirect in non-human primates and felines. Further, vocal production learning depends on circuitry involving posterior motor and anterior vocal learning pathways that non-vocal production learning species lack (Jarvis, 2004). Recently, there was a preliminary consensus that the voluntary vocal control via direct cortico-ambigual connections constitutes one indispensable component of speech and song (although note also suboscine vocal learning bellbirds who are technically classified as "non-vocal"; see Kroodsma et al., 2013; Pepperberg, 2017).

Against this background, the workshop discussed the relationship between vocal production learning and rhythmic ability. For example, what were the selection pressures that led to the emergence of neural substrates for vocal production learning and rhythm abilities? Secondly, the classical chicken-and-egg question arises when asking, did the ability to dance along to music evolve as a by-product of vocal production learning or did rhythmic ability scaffold vocal learning?

It appears that, of the species investigated so far, most species that can perceive a beat and synchronize their movements to it are vocal production learners (Hasegawa,

Okanoya, Hasegawa, & Seki, 2011; Patel, Iversen, Bregman, & Schulz, 2009; Schachner, Brady, Pepperberg, & Hauser, 2009), with the exception of sea lions (Cook, Rouse, Wilson, & Reichmuth, 2013). Notably, human synchronization abilities differ from those of non-human primates because humans can predict a beat whereas non-human primates only react to the beat. Additionally, humans are better in the auditory-motor than the visuomotor domain (Merchant & Honing, 2014; Ravignani et al., 2013; Zarco, Merchant, Prado, & Mendez, 2009).

The evolution of beat induction was discussed by focusing on the auditory-motor circuits in the brain from between-species comparative perspectives. The basal ganglia and dorsal auditory pathways were probably modified for vocal learning and synchronization, resulting in auditory-motor coupling and action simulation of human auditory perception (Patel, 2006; Patel & Iversen, 2014). An increased audiomotor connectivity then evolved in the motor cortico-basal ganglia-thalamocortical circuits within the hominine lineage, giving the human auditory system privileged access to temporal and sequential mechanisms (Merchant, Grahn, Trainor, Rohrmeier, & Fitch, 2015; Merchant & Honing, 2014).

Finally, Asano summarized areas for future research: comparative analysis of neural circuitries subserving vocal/rhythm abilities and identification of evolutionary pressures and scenarios that could have led to the emergence of voluntary vocal control in different species. It was concluded that we need to extend current research to include more species, considering the evolutionary distance between vocal learners: mammals (humans, bats, pinnipeds, and cetaceans) and birds (parrots, hummingbirds, and songbirds).

Evolutionary simulation using NetLogo (Genta Toya & Marvin Heimerich)

The workshop introduced the concept of evolutionary simulation using NetLogo (Wilensky, 1999), an environment for multi-agent-based modeling that is suitable for simulating complex agent-based phenomena evolving over time. The participants were not required to have any prior experience in programming. The workshop consisted of three parts: 1) Introduction to evolutionary simulation; 2) A basic and interactive tutorial to NetLogo; and 3) Example of evolutionary simulation. The goal of this workshop was to provide participants with the knowledge and skills to understand the basics of multi-agent simulation. Further, this workshop instructed participants to construct simple simulation individually.The first part of the workshop introduced multi-agent simulation as a useful framework to investigate evolving social behavior between individual agents. Within this framework, language and music can be examined in terms of communicative and interactive phenomena. Evolutionary simulation uses algorithms inspired by biological dynamics that evolve adaptively to an environment to deal with optimization problems. That is, this approach deals with finding (quasi-)optimal solutions, which are given particular values based on a fitness function.

The second part involved hands-on experience in working with NetLogo. Following a simple tutorial, the participants became familiar with the programming environment and learned its possible applications for research. This part of the workshop ended with the participants creating their own simple computational model in NetLogo.

The workshop concluded with the introduction of an example of an evolutionary simulation using NetLogo. In this example, the agents were sheep that had the goal to survive and reproduce. The fitness value of the sheep indicated how successful a population of sheep was at surviving and reproducing, based on its genes. The genes of the sheep with high fitness value were passed on to the next generation, which resulted in the sheep becoming more adaptive to their environment.

Gagaku workshop (Cologne Gagaku Ensemble)

The Gagaku workshop was held by Violaine Mochizuki, Pia Bornus, Guido Schäfer, and Timothy Busch, all members of the Cologne Gagaku Ensemble led by Yoshiro Shimizu. The participants had an exceptional opportunity to engage in a hands-on workshop on Gagaku, the Cologne Gagaku Ensemble being Europe's only ensemble dedicated to performing the traditional Japanese court music of Gagaku, which has been placed on the United Nations Educational, Scientific and Cultural Organization (UNESCO) representative list of intangible cultural heritage of humanity (UNESCO, 2009).

The workshop began with a short overview of the history of Gagaku, which was imported from China to Japan around 1,300 years ago (McQueen Tokita & Hughes, 2008). Gagaku includes song and instrumental music, as well as music for dance. The lecturers highlighted the importance of social codes and performance procedures, which include those related to the musicians' seating arrangements, their clothing (e.g., hunting dress), and playing techniques.

In a practical session, the participants received hands-on experience of the playing techniques of the *ryūteki* (transverse flute), *hichiriki* (small, oboe-like wind instrument), *biwa* (plucked string instrument), and the *kakko* (small drum). The participants learnt how to tune the string instruments and became familiar with the scales used in Gagaku by playing the melodic instruments. In traditional Gagaku, the transmission of musical works does not rely exclusively on written notation; rather, beginners start learning pieces by orally imitating the nature of the instrument's playing, through the singing of syllables called *shōga* (McQueen Tokita & Hughes, 2008). With examples from the first notational phrases from the Gagaku piece *Etenraku*, the participants learned, with supervision, to sing the shōga

of the wind instruments hichiriki, ryūteki, and *shō* (a woodwind instrument).

Another challenge for the participants was learning and mastering the concept of counting in Gagaku, which is based on the timing of breathing (*inja*) and the mutual interaction of the musicians rather than on regular, isochronic beats. In principle, measures exist in Gagaku. By the changing the airflow direction, the shō player marks the transition to a new measure, after which the rhythmic playing of string instruments indicates the count. Within a measure, the wind instrument marks the beat, whereas the large drum emphasizes important structural events.

The workshop closed with an ensemble performance of the first notational phrases of *Etenraku*. Accompanied by the lecturers' instrumental performance, participants sung the shōga and learned to count and coordinate with each other using the inja.

Introduction to physical computing with Arduino for New Media Art in the context of empirical cognitive musicology (Andreas Gernemann-Paulsen)

Participants in this hands-on session were introduced to the Arduino UNO, a microcontroller board programmed using an open-source coding platform (Arduino IDE). Gernemann-Paulsen described an interactive artistic and scientific methodological approach using Arduino, and participants discussed how this technology could be applied to their own work. Participants assembled a lighting circuit using UNO boards, LEDs and additional parts, and programmed the board to execute simple flashing light sequences.

Poster session

Twenty-three posters were presented by 19 students (undergraduate, graduate, and doctoral level) and four additional researchers. In addition to the regular poster presentation settings, three-minute audio recordings of the presenters were made available on participants' mobile devices as a "poster audio guide" throughout the spring school to ensure that presenters had the opportunity to experience the other poster presentations. The posters addressed issues relating to pitch and rhythm—as constituent parts of both speech and music, structure and regularity processing in language and music, animal songs, and social cognition and interaction.

A variety of approaches was taken in relation to these issues. While linguistics and musicology provided structural analysis of speech and music, comparative animal studies dealt with structural analysis of animal songs that mirrored their cognitive capacity. Computational approaches were used to identify mechanisms underlying cognitive processes and evolutionary scenarios. Psychological and neuroscientific studies investigated the cognitive and neural basis for processing structures and different constituent parts in language and music. Studies with clinical populations provided knowledge on the potential influence of disorders on cognitive domains and implications for therapeutic applications.

Topics regarding language, speech, and music perception comprised melody recognition, perception of groove, tonal alignment in speech, speech sound classification, and the effect of modality (i.e., visual or auditory modality) on perception and learning. Some posters also examined a possible transfer effect or learning enhancement from the musical domain to the linguistic domain.

The commonalities between language and music in terms of their structures were investigated in works concerning recursive structure building operation, learning of an artificial language with both semantics and syntax, processing of affirmative and negative sentences, modality independence of structure building in the brain, and perception of statistical distribution of tones in musical sequences. In comparative animal research, recursive structure in whale songs was analyzed. Further posters examined the mechanism of sequence learning with a neural network model and the evolution of the cognitive capacity for the recursive operation with an evolutionary simulation.

Social aspects of music, such as joint actions in a group music performance and dyadic dance, were also investigated. Furthermore, a linguistic social aspect, the timing of turn taking in conversations, was analyzed in research on individuals with high-functioning autism. Other research on clinical populations included rhythmic ability and phonological awareness in children with developmental language impairment, and dyadic improvisation in music therapy.

The research presented in the poster session thus illustrated different approaches to tackling developmental, mechanistic, and evolutionary questions in comparative research on language and music cognition. The posters showed that musical experience enhances language learning as well as processing, and vice versa, and that they can be investigated in terms of shared learning and processing mechanisms. In addition, both language and music are influenced by factors such as social engagement and familiarity, and the changes in those factors can enhance or attenuate perception and production in both domains.

Conclusions and future perspectives

The spring school addressed questions regarding language, music, and (social) cognition from a multitude of different viewpoints—mechanistic, developmental, and evolutionary —concerning the organization of events in time. Overall, mechanistic questions exploring the neural structures and processes underlying mental processes and behavior were the focus of discussions; in particular, questions concerning cognitive, computational, and neural mechanisms underlying temporal organization of language and music. These were approached mainly from two perspectives.

The first approach was concerned with detailed studies of mechanisms underlying syntax in terms of sequencing complexity and grammar, computational principles of structure building, and learning and acquisition of structured sequences in the respective cognitive system. It was suggested that decomposition of syntax is required, especially when investigating the neural mechanisms of syntax. Two strategies were proposed. In the first strategy, syntax is divided into simple and complex structures on the basis of abstract structural properties. Processing of simple structures is presumably associated with the frontal operculum, while complex hierarchical structures are processed within Broca's region. The second strategy focused on two different mechanisms that are necessary for carrying out syntactic computations: a stack-like storage mechanism implemented in Broca's region and a register-like one in Wernicke's region. The AF, connecting Broca's and Wernicke's regions anatomically, was suggested as linking these storage and retrieval mechanisms, to yield the generative capacity of a mildly context-sensitive grammar. Shared neural resources of language, music, and action appear to be best investigated in terms of a stack-like mechanism implemented in Broca's region.

Another approach emphasized contextual and social aspects that often serve emotional functions. These include prosody, entrainment, joint action, social learning, empathy, and joint attention. Again, two main strategies for studying mental processes and neural mechanisms can be identified. The first was concerned with the intersubjective coordination of actions directed towards internal or external goals. For example, affective and cognitive empathy (the ability to respond with appropriate emotion and to recognize another's mental state, respectively) seem to be one key requirement for intersubjective coordination. Affect regulation in emotional and social domains plays an important role in communication and was suggested to rely on reward circuits neurally implemented in the ventral striatum as the neurobiological reward center. The second strategy was concerned with how emotional or social signals can convey "meaning." For example, both social gaze and prosodic modulation of the speech signal can encode social "meaning" such as speaker intentions, interest, and engagement, and both recruit the same mentalizing system, a network comprising the medial prefrontal cortex (mPFC), posterior cingulate cortex (PCC), amygdala, and the temporo-parietal junction (TPJ).

Candidate neural structural changes which presumably played central roles in the evolution of language, speech, and music comprise expansion of the PFC, strengthening of fronto-temporal connectivity via dorsal streams (especially via the AF connecting Broca's and Wernicke's regions), and emergence of the direct cortico-ambigual connections. Moreover, song bird studies showed that FoxP2 expression affects brain regions and networks, especially the basal ganglia and related circuits, which are crucial for song learning. As human basal ganglia and their circuitries work similarly to those of song birds, they are also good candidates to better understand the evolution of language, speech, and music; given the repeatedly emphasized social functions of speech and music, their evolvement could indicate substantial changes in brain regions recruited during social information processing. In this regard, the development of two social neural systems, namely the mentalizing system and the mirror-neuron system, needs to be taken into account (Vogeley, 2017). The role of prosody, especially rhythm, was repeatedly identified as crucial for the development of the ability to organize events in time; in particular, younger children make use of prosodic cues to structure incoming events.

This summary makes it clear that the current spring school has generated a number of important research questions as challenges for future investigations. For example, the intrapersonal perspective on syntactic processes focusing on "structure" and the interpersonal perspective focusing on "function" are in contrast with each other, and there is still no adequate framework to integrate both. But, indeed, this is a general challenge in cognitive science, where theories are either based on proposals that emphasize "the role of internal representations—paradigmatically internal models—of the agent's body and environment in explaining an agent's behavior," or are based on accounts that focus on "the role of high-bandwidth agent–environment interactions in producing adaptive behavior without much or any representation on the part of the agent" (Grush, 2005, p. 209).

At the neurobiological level, one promising line of research, which was not discussed in the spring school, investigates the different functional roles of the frontal cortex in terms of a so-called sensorimotor-to-cognitive gradient (Badre & D'Esposito, 2009; Fuster, 2008; Koechlin & Jubault, 2006; Koechlin & Summerfield, 2007). Is it possible to extend this well-known gradient to the social domain by including mPFC as one of the key regions in social cognition research? Is it possible to approach the apparent opposition of intra- and interpersonal perspectives by investigating cognitive systems within such a "sensorimotor-cognitive-socio-emotional" gradient framework? Although this concept of gradients could be fruitful for future debates (as well as for discussions on alternative non-hierarchical models), there is no general consensus among researchers—this issue should be elaborated in the next spring school. Notably, prosody and dance are two domains that integrate both structural as well as functional and social aspects, and thus could serve as good starting points for such an endeavor. Research on the action–perception cycle may also provide a good starting point to investigate language, speech, and music within an integrative framework that subsumes intra- and interpersonal perspectives (e.g., Arbib, 2013; Keller, 2012; Overy & Molnar-Szakacs, 2009; Preston & de Waal, 2002; Sebanz, Bekkering, & Knoblich, 2006).

Questions concerning ontogeny and phylogeny should be approached in tandem because they are tightly intertwined in terms of developmental plasticity influencing selection of traits, social learning, niche construction, that is, the altering of the environment by organisms, and gene–culture coevolution (e.g., Laland, Odling-Smee, Hoppitt, & Uller, 2013; Laland, Sterelny, Odling-Smee, Hoppitt, & Uller, 2011). Finally, the integration of computational and biological approaches is not yet straightforward in current comparative music and language research. Unified biological information processing frameworks were suggested by several authors (e.g., Brase, 2014; Krakauer, Ghazanfar, Gomez-Marin, MacIver, & Poeppel, 2017; Mobbs, Trimmer, Blumstein, & Dayan, 2018; Poggio, 2012), but the link between computation and neurobiology is not yet clear (Gallistel & King, 2009) and still requires further elaboration for comparative music and language research. As represented in the spring school, a mechanistic approach opens up many new avenues to relate the different fields of research presented here. Computational neurocognitive modeling and computational neuroethology or biology are possibilities to yield such a unified approach (e.g., Arbib, 2016; Asano & Seifert, 2018; Fitch, 2014).

Acknowledgments

Special thanks to all the lecturers and workshop instructors for their contributions to the spring school. We also would like to thank Cedric Boeckx, Andreas Gernemann-Paulsen, Barbara Höhle, Sonja Kotz, Chris Petkov, Daniela Sammler, Constance Scharff, Uwe Seifert, and the reviewer for their helpful comments on an earlier version of this article. Special thanks also to the Evolinguistics project members for providing RA with opportunities to discuss many of the ideas, reflected in the program contents of the spring school.

Contributorship

RA wrote the "Introduction" and "Conclusions and future perspectives" sections, edited, and was responsible for development of the entire manuscript. JTC wrote the summary of Cedric Boeckx's lecture, SD wrote the summary of Barbara Höhle's lecture, SEF wrote the summaries of Daniela Sammler's lecture and the workshop by Andreas Gernemann-Paulsen, VH wrote the summary of Sonja Kotz's lecture, MH wrote the summary of the workshop by Genta Toya and Marvin Heimerich, RK wrote the summary of Uwe Seifert's lecture, ML wrote the summary of the workshop by Rie Asano and Cedric Boeckx, AMO wrote the section on the poster session, KO wrote the summary of Constance Scharff's lecture, PCT wrote the summary of Chris Petkov's lecture, TV wrote the summary of Maria Teresa Guasti's lecture, SW wrote the summary of the lecture by Mathis Jording and David Vogel, and RY wrote the summary of Ian Cross's lecture. TV and PB wrote the summary of the Gagaku workshop. MG and KV contributed to the introductory and concluding sections, and also did editorial work. All authors critically revised and approved the manuscript for publication.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The spring school was made possible through the financial support of "QVM Innovation in der Lehre" and "Cologne Summer School."

ORCID iD

Patrick C. Trettenbrein http://orcid.org/0000-0003-2233-6720
Martine Grice http://orcid.org/0000-0003-4973-4059

Peer review

Maria Teresa Guasti, Università degli Studi di Milano-Bicocca, Department of Psychology.

Note

1. The homepage of the whole project can be found here: http://musikwissenschaft.phil-fak.uni-koeln.de/34666.html?&L=1

References

Arbib, M. A. (2012). *How the brain got language. The mirror system hypothesis*. Oxford, UK: Oxford University Press.

Arbib, M. A. (Ed.). (2013). *Language, music, and the brain. A mysterious relationship*. Cambridge, MA: MIT Press.

Arbib, M. A. (2016). Towards a computational comparative neuroprimatology: Framing the language-ready brain. *Physics of Life Reviews*, *16*, 1–54.

Arbib, M. A., Verschure, P. F. M. J., & Seifert, U. (2013). Action, language, and music. Events in time and models of the brain. In M. A. Arbib (Ed.), *Language, music, and the brain. A mysterious relationship* (pp. 357–391). Cambridge, MA/London, UK: MIT Press.

Attardo, L., Guasti, M. T., & Stucchi, N. (in prep) Anticipatory skills in Jazz and classical musicians.

Asano, R., & Seifert, U. (2018). Commentary: The Evolution of Musicality: What Can Be Learned from Language Evolution Research? *Frontiers in Neuroscience*, *12*. doi: 10.3389/fnins.2018.00640

Badre, D., & D'Esposito, M. (2009). Is the rostro-caudal axis of the frontal lobe hierarchical? *Nature Reviews Neuroscience*, *10*, 659–669.

Bannan, N. (Ed.). (2012). *Music, language, and human evolution*. Oxford, UK: Oxford University Press.

Beeh, V. (1981). *Sprache und Spracherlernung unter mathematisch-biologischer Perspektive*. Berlin: de Gruyter.

Bernstein, L. (1976). *The unanswered questions: Six talks at Harvard* (3rd ed.). Cambridge, MA: Harvard University Press.

Berwick, R. C., & Chomsky, N. (2016). *Why only us: Language and evolution*. Cambridge, MA: The MIT Press.

Besle, J., Fort, A., Delpuech, C., & Giard, M.-H. (2004). Bimodal speech: Early suppressive visual effects in human auditory cortex. *European Journal of Neuroscience*, *20*, 2225–2234.

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

Bolhuis, J. J., Tattersall, I., Chomsky, N., & Berwick, R. C. (2014). How could language have evolved? *PLoS Biology*, *12*, e1001934.

Boll-Avetisyan, N., Bhatara, A., Unger, A., Nazzi, T., & Höhle, B. (2016). Effects of experience with L2 and music on rhythmic grouping by French listeners. *Bilingualism: Language and Cognition*, *19*, 971–986.

Brase, G. L. (2014). Behavioral science integration: A practical framework of multi-level converging evidence for behavioral science theories. *New Ideas in Psychology*, *33*, 8–20.

Brennan, J. R., Stabler, E. P., Van Wagenen, S. E., Luh, W.-M., & Hale, J. T. (2016). Abstract linguistic structure correlates with temporal activity during naturalistic comprehension. *Brain and Language*, *157–158*, 81–94.

Carnap, R. (1950). *Logical foundations of probability*. Chicago: The University of Chicago Press.

Chomsky, N. (1955). *The logical structure of linguistic theory*. Cambridge, MA: Mimeo MIT.

Chomsky, N. (1956). Three models for the description of language. *IEEE Transactions on Information Theory*, *2*, 113–124.

Chomsky, N. (1957). *Syntactic structures*. Mouton: The Hague.

Chomsky, N. (1959). On certain formal properties of grammars. *Information and Control*, *2*, 137–167.

Chomsky, N. (1963). Formal properties of grammars. In R. D. Luce, R. R. Bush, & E. Galalnter (Eds.), *Handbook of mathematical psychology* (Vol. *II*, Chap. 9-14, pp. 323–418). New York: Wiley.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: The MIT Press.

Chomsky, N. (1986). *Knowledge of language: Its nature, origins, and use*. New York: Praeger.

Chomsky, N. (1995). *The minimalist program*. Cambridge, MA: The MIT Press.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127–149). Washington, DC: American Psychological Association.

Coan, J. A. (2010). Adult attachment and the brain. *Journal of Social and Personal Relationships*, *27*, 210–217.

Cook, P., Rouse, A., Wilson, M., & Reichmuth, C. (2013). A California sea lion (Zalophus californianus) can keep the beat: Motor entrainment to rhythmic auditory stimuli in a non vocal mimic. *Journal of Comparative Psychology*, *127*, 412–427.

Cross, I. (1999). Is music the most important thing we ever did? Music, development and evolution. In S. W. Yi (Ed.), *Music, mind and science* (pp. 10–39). Seoul: Seoul National University Press.

Cross, I. (2011). The meanings of musical meanings. Comment on "Towards a neural basis of processing musical semantics" by Stefan Koelsch. *Physics of Life Reviews*, *8*, 116–119.

Cross, I. (2014). Music and communication in music psychology. *Psychology of Music*, *42*, 809–819.

Dahlhaus, C. (1971). Musiktheorie. In C. Dahlhaus (Ed.), *Einführung in die systematische Musikwissenschaft* (pp. 93–132). Köln: Gerig.

Dalziell, A. H., Peters, R. A., Cockburn, A., Dorland, A. D., Maisey, A. C., & Magrath, R. D. (2013). Dance choreography is coordinated with song repertoire in a complex avian display. *Current Biology*, *23*, 1132–1135.

Dillard, J. P., Solomon, D. H., & Palmer, M. T. (1999). Structuring the concept of relational communication. *Communication Monographs*, *66*, 49–65.

Driver, J. (2001). A selective review of selective attention research from the past century. *British Journal of Psychology*, *92*, 53–78.

Feldman, R. (2017). The neurobiology of human attachments. *Trends in Cognitive Sciences*, *21*, 80–99.

Fitch, W. T. (2010). *The evolution of language*. Cambridge: Cambridge University Press.

Fitch, W. T. (2011). The evolution of syntax: An exaptationist perspective. *Frontiers in Evolutionary Neuroscience*, *3*, 9.

Fitch, W. T. (2014). Toward a computational framework for cognitive biology: Unifying approaches from cognitive neuroscience and comparative cognition. *Physics of Life Reviews*, *11*, 329–364.

Fitch, W. T. (2015). Four principles of bio-musicology. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *370*, 1664.

Fitch, W. T. (2018). Four principles of biomusicology. In H. Honing (Ed.), *The origins of musicality* (pp. 23–48). Cambridge, MA: The MIT Press.

Fitch, W. T., & Friederici, A. D. (2012). Artificial grammar learning meets formal language theory: An overview. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *367*, 1933–1955.

Fitch, W. T., & Martins, M. D. (2014). Hierarchical processing in music, language, and action: Lashley revisited. *Annals of the New York Academy of Sciences*, *1316*, 87–104.

Frankland, S. M. (2015). *Man bites dog: The representation of structured meaning in left-mid superior temporal cortex*. Doctoral dissertation, Harvard University, Cambridge.

Frankland, S. M., & Greene, J. D. (2015). An architecture for encoding sentence meaning in left mid-superior temporal cortex. *Proceedings of the National Academy of Sciences*, *112*, 11732–11737.

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

Friederici, A. D. (2017). *Language in our brain: The origins of a uniquely human capacity*. Cambridge, MA: The MIT Press.

Friederici, A. D., Chomsky, N., Berwick, R. C., Moro, A., & Bolhuis, J. J. (2017). Language, mind and brain. *Nature Human Behaviour*, *1*, 713–722.

Fuster, J. M. (2008). *The prefrontal cortex*. Amsterdam: Elsevier Academic Press.

Gallistel, C. R., & King, A. P. (2009). *Memory and the computational brain. Why cognitive science will transform neuroscience*. Chichester, UK: Wiley-Blackwell.

Gelder, B. d. (2000). Neuroscience: More to seeing than meets the eye. *Science*, *289*, 1148–1149.

Georgescu, A. L., Kuzmanovic, B., Schilbach, L., Tepest, R., Kulbida, R., Bente, G., & Vogeley, K. (2013). Neural

correlates of "social gaze" processing in high-functioning autism under systematic variation of gaze duration. *NeuroImage: Clinical*, *3*, 340–351.

Goffman, E. (1955). On face-work. *Psychiatry*, *18*, 213–231.

Goucha, T., Zaccarella, E., & Friederici, A. D. (2017). A revival of Homo loquens as a builder of labeled structures: Neurocognitive considerations. *Neuroscience & Biobehavioral Reviews*, *81*, 213–224.

Grush, R. (2005). Internal models and the construction of time: Generalizing from state estimation to trajectory estimation to address temporal features of perception, including temporal illusions. *Journal of Neural Engineering*, *2*, S209–S218.

Haesler, S., Rochefort, C., Georgi, B., Licznerski, P., Osten, P., & Scharff, C. (2007). Incomplete and inaccurate vocal imitation after knockdown of FoxP2 in songbird basal ganglia nucleus Area X. *PLoS Biology*, *5*, e321.

Hartmann, N. (1962). *Das Problem des geistigen Seins. Untersuchungen zur Grundlegung der Geschichtsphilosophie und der Geisteswissenschaften*. 3. (unveränder). Berlin: de Gruyter.

Hasegawa, A., Okanoya, K., Hasegawa, T., & Seki, Y. (2011). Rhythmic synchronization tapping to an audio-visual metronome in budgerigars. *Scientific Reports*, *1*, 120.

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science*, *298*, 1569–1579.

Hawkins, S., Cross, I., & Ogden, R. (2013). Communicative interaction in spontaneous music and speech. In M. Orwin, C. Howes, & R. Kempson (Eds.), *Language, music and interaction* (pp. 285–329). London, UK: College Publications.

Heider, F. (1958). *The psychology of interpersonal relations*. New York: Wiley.

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

Hellbernd, N., & Sammler, D. (2018). Neural bases of social communicative intentions in speech. *Social Cognitive and Affective Neuroscience, (June)*, 1–12.

Höhle, B., Bijeljac-Babic, R., Herold, B., Weissenborn, J., & Nazzi, T. (2009). Language specific prosodic preferences during the first half year of life: Evidence from German and French infants. *Infant Behavior and Development*, *32*, 262–274.

Honing, H. (Ed.). (2018). *The origins of musicality*. Cambridge, MA: The MIT Press.

Hughes, D. W. (1991). Grammars of non-western musics: A selective survey. In P. Howell, R. West, & I. Cross (Eds.), *Representing musical structure* (pp. 327–362). London: Academic Press.

Hultsch, H., & Todt, D. (1982). Temporal performance roles during vocal interactions in nightingales (Luscinia megarhynchos B.). *Behavioral Ecology and Sociobiology*, *11*, 253–260.

Jackendoff, R. (1987). *Consciousness and the computational mind*. Cambridge, MA: MIT Press.

James, L. S., & Sakata, J. T. (2017). Learning biases underlie "Universals" in Avian vocal sequencing. *Current Biology*, *27*, 3676–3682.

Jarvis, E. D. (2004). Learned birdsong and the neurobiology of human language. *Annals of the New York Academy of Sciences*, *1016*, 749–777.

Jarvis, E. D., Scharff, C., Grossman, M. R., Ramos, J. A., & Nottebohm, F. (1998). For whom the bird sings. *Neuron*, *21*, 775–788.

Jessen, S., & Kotz, S. A. (2013). On the role of crossmodal prediction in audiovisual emotion perception. *Frontiers in Human Neuroscience*, *7*, 369.

Jording, M., Hartz, A., Bente, G., Schulte-Rüther, M., & Vogeley, K. (2018). The "Social gaze space": A taxonomy for gaze-based communication in triadic interactions. *Frontiers in Psychology*, *9*, 226.

Keller, P. E. (2012). Rhythm and time in music epitomize the temporal dynamics of human communicative behavior: The broad implications of London's trinity. *Empirical Musicology Review*, *7*, 17–27.

Koechlin, E., & Jubault, T. (2006). Broca's area and the hierarchical organization of human behavior. *Neuron*, *50*, 963–974.

Koechlin, E., & Summerfield, C. (2007). An information theoretical approach to prefrontal executive function. *Trends in Cognitive Sciences*, *11*, 229–235.

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

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

Krakauer, J. W., Ghazanfar, A. A., Gomez-Marin, A., MacIver, M. A., & Poeppel, D. (2017). Neuroscience needs behavior: Correcting a reductionist bias. *Neuron*, *93*, 480–490.

Kroodsma, D., Hamilton, D., Sánchez, J. E., Byers, B. E., Fandiño-Mariño, H., Stemple, D. W., . . . Powell, G. V. N. (2013). Behavioral evidence for song learning in the suboscine bellbirds (Procnias spp.; Cotingidae). *The Wilson Journal of Ornithology*, *125*, 1–14.

Kubikova, L., Bosikova, E., Cvikova, M., Lukacova, K., Scharff, C., & Jarvis, E. D. (2015). Basal ganglia function, stuttering, sequencing, and repair in adult songbirds. *Scientific Reports*, *4*, 6590.

Kuzmanovic, B., Georgescu, A. L., Eickhoff, S. B., Shah, N. J., Bente, G., Fink, G. R., & Vogeley, K. (2009). Duration matters: Dissociating neural correlates of detection and evaluation of social gaze. *NeuroImage*, *46*, 1154–1163.

Kuzmanovic, B., Schilbach, L., Georgescu, A. L., Kockler, H., Santos, N. S., Shah, N. J., . . . Vogeley, K. (2014). Dissociating animacy processing in high-functioning autism: Neural correlates of stimulus properties and subjective ratings. *Social Neuroscience*, *9*, 309–325.

Lai, C. S. L., Fisher, S. E., Hurst, J. A., Vargha-Khadem, F., & Monaco, A. P. (2001). A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature*, *413*, 519–523.

Laland, K. N., Odling-Smee, J., Hoppitt, W., & Uller, T. (2013). More on how and why: Cause and effect in biology revisited. *Biology & Philosophy*, *28*, 719–745.

Laland, K. N., Sterelny, K., Odling-Smee, J., Hoppitt, W., & Uller, T. (2011). Cause and effect in biology revisited: Is

mayr's proximate-ultimate dichotomy still useful? *Science*, *334*, 1512–1516.

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

Levelt, W. J. M. (2008). *An introduction to the theory of formal languages and automata*. Amsterdam: Benjamins.

Levine, J. (1999). Explanatory gap. In R. A. Wilson & F. C. Keil (Eds.), *The MIT encyclopedia of the cognitive sciences* (pp. 304–305). Cambridge, MA: The MIT Press.

McQueen Tokita, A., & Hughes, D. W. (Eds.). (2008). *The Ashgate research companion to Japanese music*. Farnham, Surrey: Ashgate.

Merchant, H., Grahn, J. A., Trainor, L., Rohrmeier, M., & Fitch, W. T. (2015). Finding the beat: A neural perspective across humans and non-human primates. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 370*, 20140093.

Merchant, H., & Honing, H. (2014). Are non-human primates capable of rhythmic entrainment? Evidence for the gradual audiomotor evolution hypothesis. *Frontiers in Neuroscience*, *7*, 274.

Mobbs, D., Trimmer, P. C., Blumstein, D. T., & Dayan, P. (2018). Foraging for foundations in decision neuroscience: Insights from ethology. *Nature Reviews Neuroscience*, *19*, 419–427.

Murugan, M., Harward, S., Scharff, C., & Mooney, R. (2013). Diminished FoxP2 levels affect dopaminergic modulation of corticostriatal signaling important to song variability. *Neuron*, *80*, 1464–1476.

Nelson, R. J. (1989). *The logic of mind* (2nd ed.). Dordrecht: Reidel.

Ogden, R., & Hawkins, S. (2015). Entrainment as a basis for coordinated actions in speech. *Proceedings of the 18th International Congress of Phonetic Sciences, UK*. Retrieved from https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0599.pdf

Overy, K., & Molnar-Szakacs, I. (2009). Being together in time: Musical experience and the mirror neuron system. *Music Perception*, *26*, 489–504.

Pagliarini, E., Guasti, M. T., Toneatto, C., Granocchio, E., Riva, F., Sarti, D., . . . Stucchi, N. (2015). Dyslexic children fail to comply with the rhythmic constraints of handwriting. *Human Movement Science*, *42*, 161–182.

Pagliarini, E., Scocchia, L., Vernice, M., Zoppello, M., Balottin, U., Bouamama, S., . . . Stucchi, N. (2017). Children's first handwriting productions show a rhythmic structure. *Scientific Reports*, *7*, 5516.

Patel, A. D. (2006). Musical rhythm, linguistic rhythm, and human evolution. *Music Perception*, *24*(1), 99–104.

Patel, A. D. (2008). *Music, language, and the brain*. Oxford, New York: Oxford University Press.

Patel, A. D., & Iversen, J. R. (2014). The evolutionary neuroscience of musical beat perception: The Action Simulation for Auditory Prediction (ASAP) hypothesis. *Frontiers in Systems Neuroscience*, *8*, 57.

Patel, A. D., Iversen, J. R., Bregman, M. R., & Schulz, I. (2009). Experimental evidence for synchronization to a muscial beat in a nonhuman animal. *Current Biology*, *1169* 459–469.

Patel, A. D., & Morgan, E. (2017). Exploring cognitive relations between prediction in language and music. *Cognitive Science*, *41*, 303–320.

Pepperberg, I. M. (1987). Acquisition of the same/different concept by an African Grey parrot (Psittacus erithacus): Learning with respect to categories of color, shape, and material. *Animal Learning & Behavior*, *15*, 423–432.

Pepperberg, I. M. (2017). Animal language studies: What happened? *Psychonomic Bulletin & Review*, *24*, 181–185.

Peretz, I., & Zatorre, R. J. (Eds.). (2003). *The cognitive neuroscience of music*. Oxford: Oxford University Press.

Persici, V., Stucchi, N., & Arosio, F. (2017). Predicting the future in rhythm and language: The anticipation abilities of a group of Italian-speaking preschoolers. In *13th Generative Approaches to Language Acquisition (GALA13)* 7–9 September.

Petersson, K.-M., Folia, V., & Hagoort, P. (2012). What artificial grammar learning reveals about the neurobiology of syntax. *Brain and Language*, *120*, 83–95.

Petkov, C. I., & Jarvis, E. D. (2012). Birds, primates, and spoken language origins: Behavioral phenotypes and neurobiological substrates. *Frontiers in Evolutionary Neuroscience*, *4*, 12.

Poggio, T. (2012). The levels of understanding framework, revised. *Perception*, *41*, 1017–1023.

Preston, S. D., & de Waal, F. B. M. (2002). Empathy: Its ultimate and proximate bases. *Behavioral and Brain Sciences*, *25*, 1–20; discussion 20–71.

Quirin, M., Meyer, F., Heise, N., Kuhl, J., Küstermann, E., Strüber, D., & Cacioppo, J. T. (2013). Neural correlates of social motivation: An fMRI study on power versus affiliation. *International Journal of Psychophysiology*, *88*, 289–295.

Ravignani, A., Gingras, B., Asano, R., Sonnweber, R., Metellán, V., & Fitch, W. T. (2013). The evolution of rhythmic cognition: New perspectives and technologies in comparative research. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Cooperative minds: Social interaction and group dynamics. Proceedings of the 35th Annual Meeting of the Cognitive Science Society* (pp. 1199–1204). Austin, TX: Cognitive Science Society.

Rebuschat, P., Rohrmeier, M., Hawkins, J. A., & Cross, I. (Eds.). (2012). *Language and music as cognitive systems*. Oxford, New York: Oxford University Press.

Robledo Del Canto, J. P., Hawkins, S., Cross, I., & Ogden, R. (2016). Pitch-interval analysis of 'periodic' and 'aperiodic' Question+Answer pairs. *Proceedings of Speech Prosody 2016*, 1071–1075. Retrieved from https://www.isca-speech.org/archive/SpeechProsody_2016/pdfs/380.pdf

Robledo, J. P., Hurtado, E., Prado, F., Román, D., & Cornejo, C. (2016). Music intervals in speech: Psychological disposition modulates ratio precision among interlocutors' nonlocal f0 production in real-time dyadic conversation. *Psychology of Music*, *44*, 1404–1418.

Rohrmeier, M. (2011). Towards a generative syntax of tonal harmony. *Journal of Mathematics and Music*, *5*, 35–53.

Rohrmeier, M., Zuidema, W., Wiggins, G. A., & Scharff, C. (2015). Principles of structure building in music, language and animal song. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *370*, 20140097.

Rothacker, E. (1947). *Logik und Systematik der Geisteswissenschaften*. Bonn: Bouvier.

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

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

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

Santos, N. S., David, N., Bente, G., & Vogeley, K. (2008). Parametric induction of animacy experience. *Consciousness and Cognition*, *17*, 425–437.

Santos, N. S., Kuzmanovic, B., David, N., Rotarska-Jagiela, A., Eickhoff, S. B., Shah, J. N., . . . Vogeley, K. (2010). Animated brain: A functional neuroimaging study on animacy experience. *NeuroImage*, *53*, 291–302.

Schachner, A., Brady, T. F., Pepperberg, I. M., & Hauser, M. D. (2009). Spontaneous motor entrainment to music in multiple vocal mimicking species. *Current Biology*, *19*, 831–836.

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: Bodies and minds moving together. *Trends in Cognitive Sciences*, *10*, 70–76.

Soma, M., & Iwama, M. (2017). Mating success follows duet dancing in the Java sparrow. *PLOS One*, *12*, e0172655.

Steedman, M. (1996). The blues and the abstract truth: Music and mental models. In J. Oakhill & A. Garnham (Eds.), *Mental models in cognitive science: Essays in honour of Phil Johnson-Laird* (pp. 305–318). Hove: Psychology Press.

Stekelenburg, J. J., & Vroomen, J. (2012). Electrophysiological correlates of predictive coding of auditory location in the perception of natural audiovisual events. *Frontiers in Integrative Neuroscience*, *6*, 26.

Sundberg, J., & Lindblom, B. (1991). Generative theories for describing musical structure. In P. Howell, R. West, & I. Cross (Eds.), *Representing musical structure* (pp. 245–272). London: Academic Press.

Terrace, H., Petitto, L., Sanders, R., & Bever, T. (1979). Can an ape create a sentence? *Science*, *206*, 891–902.

Tinbergen, N. (1963). On aims and methods of Ethology. *Zeitschrift Für Tierpsychologie*, *20*, 410–433.

Todt, D., & Geberzahn, N. (2003). Age-dependent effects of song exposure: Song crystallization sets a boundary between fast and delayed vocal imitation. *Animal Behaviour*, *65*, 971–979.

Ullrich, R., Norton, P., & Scharff, C. (2016). Waltzing Taeniopygia: Integration of courtship song and dance in the domesticated Australian zebra finch. *Animal Behaviour*, *112*, 285–300.

UNESCO. (2009). *Gagaku*. Retrieved from https://ich.unesco.org/en/RL/gagaku-00265

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences*, *102*, 1181–1186.

Vogeley, K. (2017). Two social brains: Neural mechanisms of intersubjectivity. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*, 20160245.

Vroomen, J., & Stekelenburg, J. J. (2010). Visual anticipatory information modulates multisensory interactions of artificial audiovisual stimuli. *Journal of Cognitive Neuroscience*, *22*, 1583–1596.

Wallin, N. L., Merker, B., & Brown, S. (Eds.). (2000). *The origins of music*. Cambridge, MA: MIT Press.

Ward, D., Connally, E. L., Pliatsikas, C., Bretherton-Furness, J., & Watkins, K. E. (2015). The neurological underpinnings of cluttering: Some initial findings. *Journal of Fluency Disorders*, *43*, 1–16.

Weiss, M., Hultsch, H., Adam, I., Scharff, C., & Kipper, S. (2014). The use of network analysis to study complex animal communication systems: A study on nightingale song. *Proceedings of the Royal Society B: Biological Sciences*, *281*, 20140460.

Wilensky, U. (1999). *NetLogo*. Retrieved from https://ccl.northwestern.edu/netlogo/

Williams, H. (2004). Birdsong and singing behavior. *Annals of the New York Academy of Sciences*, *1016*, 1–30.

Wilson, B., Marslen-Wilson, W. D., & Petkov, C. I. (2017). Conserved sequence processing in primate Frontal Cortex. *Trends in Neurosciences*, *40*, 72–82.

Wittgenstein, L. (1953/1958). *Philosophical investigations* (G. E. M. Anscombe, trans.). Oxford: Basil Blackwell Ltd.

Woolley, S. C., & Doupe, A. J. (2008). Social context–induced song variation affects female behavior and gene expression. *PLoS Biology*, *6*, e62.

Yang, C. (2013). Ontogeny and phylogeny of language. *Proceedings of the National Academy of Sciences*, *110*, 6324–6327.

Zarco, W., Merchant, H., Prado, L., & Mendez, J. C. (2009). Subsecond timing in primates: Comparison of interval production between human subjects and rhesus monkeys. *Journal of Neurophysiology*, *102*, 3191–3202.

Zbikowski, L. M. (2017). *Foundations of musical grammar*. Oxford: Oxford University Press.