Check for updates

Emotional and Interactional Prosody across Animal Communication Systems: A Comparative Approach to the Emergence of Language

Piera Filippi*

Department of Artificial Intelligence, Vrije Universiteit Brussel, Brussels, Belgium

Across a wide range of animal taxa, prosodic modulation of the voice can express emotional information and is used to coordinate vocal interactions between multiple individuals. Within a comparative approach to animal communication systems, I hypothesize that the ability for emotional and interactional prosody (EIP) paved the way for the evolution of linguistic prosody – and perhaps also of music, continuing to play a vital role in the acquisition of language. In support of this hypothesis, I review three research fields: (i) empirical studies on the adaptive value of EIP in non-human primates, mammals, songbirds, anurans, and insects; (ii) the beneficial effects of EIP in scaffolding language learning and social development in human infants; (iii) the cognitive relationship between linguistic prosody and the ability for music, which has often been identified as the evolutionary precursor of language.

Keywords: language evolution, musical protolanguage, prosody, interaction, turn-taking, arousal, infant-directed speech, entrainment

PROSODY IN HUMAN COMMUNICATION

Whenever listeners comprehend spoken speech, they are processing sound patterns. Traditionally, studies on language processing assume a two-level hierarchy of sound patterns, a property called "duality of pattern" or "double articulation" (Hockett, 1960; Martinet, 1980). The first dimension is the concatenation of meaningless phonemes into larger discrete units, namely morphemes, in accordance to the phonological rules of the given language. At the next level, these phonological structures are formed into words and morphemes with semantic content and arranged within hierarchical structures (Hauser et al., 2002), according to morpho-syntactical rules. Surprisingly, this line of research has often overlooked prosody, the "musical" aspect of the speech signal, i.e., the so-called "suprasegmental" dimension of the speech stream, which includes timing, frequency spectrum, and amplitude (Lehiste, 1970). Taken together, these values outline the overall prosodic contour of words and/or sentences. According to the source–filter theory of voice production (Fant, 1960; Titze, 1994), vocalizations in humans - and in mammals more generally- are generated by airflow interruption through vibration of the vocal folds in the larynx ('source'). The signal produced at the source is subsequently filtered in the vocal tract ('filter'). The source determines the fundamental frequency of the call ($F0$), and the filter shapes the source signal, producing a concentration of acoustic energy around particular frequencies in the speech wave, i.e., the formants. Thus, it is

important to highlight that in producing vocal utterances, speakers across cultures and languages modulate *both* segmental, and prosodic information in the signal. In humans, prosodic modulation of the voice affects language processing at multiple levels: linguistic (lexical and morpho-syntactic), emotional, and interactional.

Linguistic Prosody

Prosody has a key role in word recognition, syntactic structure processing, and discourse structure comprehension (Cutler et al., 1997; Endress and Hauser, 2010; Wagner and Watson, 2010; Shukla et al., 2011). Prosodic cues such as lexical stress patterns specific to each natural language are exploited to segment words within speech streams (Mehler et al., 1988; Cutler, 1994; Jusczyk and Aslin, 1995; Jusczyk, 1999; Johnson and Jusczyk, 2001; Curtin et al., 2005). For instance, many studies of English have indicated that segmental duration tends to be longest in word-initial position and shorter in word-final position (Oller, 1973). Newborns use stress patterns to classify utterances into broad language classes defined according to global rhythmic properties (Nazzi et al., 1998). The effect of prosody in word processing is distinctive in tonal languages, where F0 variations on the same segment results in totally different meanings (Cutler and Chen, 1997; Lee, 2000). For instance, the Cantonese consonant-vowel sequence [si] can mean "poem," "history," or "time," based on the specific tone in which it is uttered.

Prosodic variations such as phrase-initial strengthening through pitch rise, phrase-final lengthening, or pitch discontinuity at the boundaries between different phrases mark morpho-syntactic connections within sentences (Soderstrom et al., 2003; Johnson, 2008; Männel et al., 2013). These prosodic variations mark phrases within sentences, favoring syntax acquisition in infants (Steedman, 1996; Christophe et al., 2008) and guiding hierarchical or embedded structure comprehension in continuous speech in adults (Müller et al., 2010; Langus et al., 2012; Ravignani et al., 2014b; Honbolygo et al., 2016). Moreover, these prosodic cues enable the resolution of global ambiguity in sentences like "flying airplanes can be dangerous" – which can mean that *the act* of flying airplanes can be dangerous or that the objects *flying airplanes* can be dangerous – or "I read about the repayment with interest," where "with interest" can be directly referred to the act of reading or to the repayment. Furthermore, sentences might be characterized by local ambiguity, i.e., ambiguity of specific words, which can be resolved by semantic integration with the following information within the same sentence, as in "John believes Mary implicitly" or "John believes Mary to be a professor." Here, the relationship between "believes" and "Mary" depends on what follows. In the case of both global and local ambiguity, prosodic cues to the syntactical structure of the sentence aid the understanding of the utterance meaning as intended by the speaker (Cutler et al., 1997; Snedeker and Trueswell, 2003; Nakamura et al., 2012).

Prosodic features of the signal are used to mark questions (Hedberg and Sosa, 2002; Kitagawa, 2005; Rialland, 2007), and in some languages, prosody serves as a marker of salient (Bolinger, 1972) or new (Fisher and Tokura, 1995) information. Consider for instance, "MARY gave the book to John" vs. "Mary gave the book to JOHN," in which the accented word is the one the speaker wants to drive the listener's attention to in the conversational context.

Emotional Prosody In Humans

The prosodic modulation of the utterance can signal the emotional state of the speaker, independently from her/his intention to express an emotion. Research suggests that specific patterns of voice modulation can be considered a "biological code" for both linguistic and paralinguistic communication (Gussenhoven, 2002). Indeed, physiological changes might cause tension and action of muscles used for phonation, respiration, and speech articulation (Lieberman, 1967; Scherer, 2003). For instance, physiological variations in an emotionally aroused speaker might cause an increase of the subglottal pressure (i.e., the pressure generated by the lungs beneath the larynx), which might affect voice amplitude and frequency, thus expressing his/her emotional state. Crucially, in cases of emotional communication, prosody can prime or guide the perception of the semantic meaning (Ishii et al., 2003; Schirmer and Kotz, 2003; Pell, 2005; Pell et al., 2011; Newen et al., 2015; Filippi et al., 2016). Moreover, the expression of emotions through prosodic modulation of the voice, in combination with other communication channels, is crucial for affective and attentional regulation in social interactions both in adults (Sander et al., 2005; Schore and Schore, 2008) and infants (see section "EIP in Language Acquisition" below).

Prosody for Interactional Coordination In Humans

A crucial aspect of spoken language is its interactional nature. In conversations, speakers typically use prosodic cues for interactional coordination, i.e., implicit turn-taking rules that aid the perception of who is to speak next and when, predicting the content and timing of the incoming turn (Roberts et al., 2015). The typical use of a turn-taking system might explain why language is organized into short phrases with an overall prosodic envelope (Levinson, 2016). Within spoken interactions, prosodic features such as low pitch or final word lengthening are used for turn-taking coordination, determining the rhythm of the conversations among speakers (Ward and Tsukahara, 2000; Ten Bosch et al., 2005; Levinson, 2016). These prosodic features in the signal are used to recognize opportunities for turn transition and appropriate timing to avoid gaps and overlaps between speakers (Sacks et al., 1974; Stephens and Beattie, 1986). Wilson and Wilson (2005) suggested that both the listener and the speaker engage in an oscillator-based cycle of readiness to initiate a syllable, which is at a minimum in the middle of syllable production, at the point of greatest syllable sonority, and at a maximum when the prosodic values of the syllable lessen, typically in the final part of the syllable. The listener is entrained by the speaker's rate of syllable production, but her/his cycle is counterphased to the speaker's cycle. Therefore, the listener will be able to take turn in speaking if s/he detects that the speaker is not initiating a new cycle of syllable production. In accordance to this model, Stivers et al. (2009) provided evidence for biologically

rooted timings in replying to speakers on the base of prosodic features in the signal, a finding that is indicative of a strong universal basis for turn-taking behavior. Specifically, this study provides evidence for a similar distribution of response offsets (unimodal peak of response within 200 ms of the end of the utterance) across conversations in ten languages drawn from traditional indigenous communities to major world languages. The authors observed a general avoidance of overlapping talk and minimal silence between conversational turns across all tested languages.

A Comparative Approach to Emotional and Interactional Prosody

Given the centrality of prosody in spoken communication, it is worth addressing the adaptive role of prosody on both an evolutionary and a developmental level. Here, I hypothesize that prosodic modulation of the voice marking emotional communication and interactional coordination (hereafter EIP, emotional and interactional prosody), as we observe it nowadays across multiple animal taxa, evolved into the ability to modulate prosody for language processing – and might have played an important role in the emergence of music (**Figure 1**) (Phillips-Silver et al., 2010; Bryant, 2013; Zimmermann et al., 2013). In support of this hypothesis, within a comparative approach, I will review studies on the adaptive use of prosodic modulation of the voice for emotional communication and interactional coordination in animals.
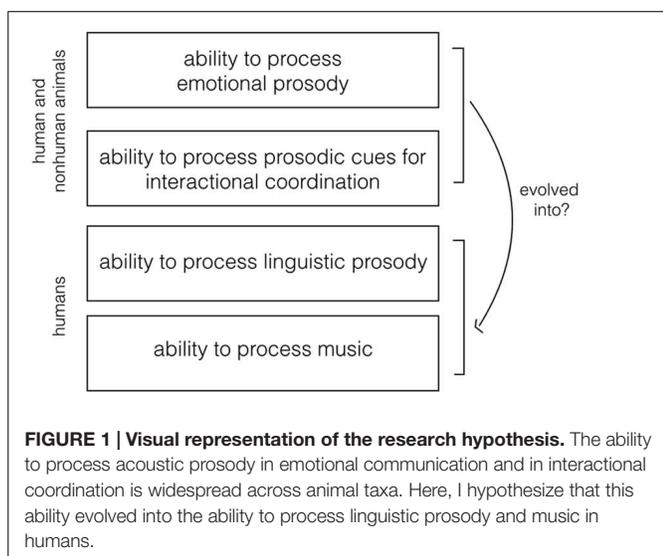
Importantly, following Morton (1977) and Owren and Rendall (1997), I aim to address the behavioral and functional effects of emotional vocalizations in animals, as conveyed by their prosodic characteristics and by the interactional dynamics of communication act. Therefore, I will adopt the very basic, but fundamental assumption that the prosodic structure of calls (which reflects the physiological/emotional state of the signaler) and call-answer dynamics induce nervous-system and physiological responses in the receiver. For instance, a call

might induce an increased level of emotional arousal or of attention. These physiological responses might trigger specific types of behaviors in the listeners, for instance escape or physical approaching (Nesse, 1990; Frijda, 2016). Ultimately, these behaviors are the immediate functional effect of the communication act (Owren and Rendall, 2001; Rendall et al., 2009).

A crucial dimension, constitutive of a multiple communicative behaviors across animal species, is interactional coordination. Examples of interactional coordination are widespread across animal classes, including unrelated taxa. This suggests that this ability has evolved independently in a number of species under similar selective pressures (Ravignani, 2014). There are three main types of interactional coordination in animal acoustic communication: choruses, antiphonal calling, and duets (Yoshida and Okanoya, 2005). In choruses, males simultaneously emit a signal for sexual advertisement or as an anti-predator defensive behavior. Antiphonal calling occurs when more than two members of a group exchange calls within an interactive context. Duets occur when members of a pair (e.g., sexual mates, caregiver-juvenile) exchange calls within a precise time window. Importantly, the modulation of the prosodic features of the vocal signals is key to coordinating these communicative behaviors.

Based on Tinbergen (1963), in order to grasp an integrative understanding of animal vocal communication, I will go through four levels of description: mechanisms, functional effects (**Table 1**), phylogenetic history, and ontogenetic development. Two strands of analysis are relevant in the context of comparative investigation on the adaptive advantages of prosody in relation to the origins of language: (a) research on the evolutionary 'homologies,' which provides information on the phylogenetic traits that humans and other primates share with their common ancestor; (b) investigations on "analogous" traits, aimed at finding the evolutionary pressures that guided the emergence of the same biological traits that evolved independently in phylogenetically distant species (Gould and Eldredge, 1977; Hauser et al., 2002). As to the ontogenetic level of explanation, I will review empirical data on the beneficial effects of EIP for the development of social and vocal learning skills in multiple animal species.

Within this line of research, it is important to highlight that extensive research has identified the evolutionary precursor of language in a general ability to produce music (Brown, 2001; Mithen, 2005; Patel, 2006; Fitch, 2010, 2012). There are at least two orders of argumentation supporting the hypothesis that aspects of musical processing were involved in human language evolution: (a) research on the cognitive link between music and verbal language processing; (b) comparative data on animal communication systems, suggesting that this ability, already in place in different primate as well as in many non-primate species, might have evolved into an adaptive ability in the first hominins. Based on the reviews of (a) and (b), I propose to identify the emotional and interactional functions of prosody as dimensions that are sufficient to an account for the "musical" origins of language. This conceptual operation will provide a parsimonious account for the investigation of the origins of language as well as of

FIGURE 1 | Visual representation of the research hypothesis. The ability to process acoustic prosody in emotional communication and in interactional coordination is widespread across animal taxa. Here, I hypothesize that this ability evolved into the ability to process linguistic prosody and music in humans.

TABLE 1 | Overview of the functional effects of emotional and interactional prosody across diverse animal taxa.

		Insects	Anurans	Birds	Non-human mammals	Non-human primates	Humans
Emotional prosody		Expression of the signaler's physiological state					Expression of the signaler's physiological state, affective regulation of interpersonal interactions
Prosody for interactional coordination in auditory communication	Chorus	Sexual advertisement, anti-predator behavior	Sexual advertisement, anti-predator behavior	Social bonding, synchronization of activities and group or territory defense	[Not reported]	[Not reported]	Social entrainment, group cohesion, cooperation
	Antiphonal calling	[Not reported]	[Not reported]	Aggressive/submissive signaling in territorial contests	Spatial location, social bonding, identity signaling	Group cohesion	[Not reported]
	Duet	Sexual advertisement	Sexual advertisement, male–male competition	Adults: pair bonding, spacing of males, reunification of separated mates; Tutor-juvenile: song learning	Sexual advertisement [reported only in Cape-mole rats]	Adults: pair bonding, territory and resource defense; Caregiver-juvenile: interpersonal bonding, social development, vocal development	Adults: inter-individual affective regulation; Caregiver-Infant: socio-cognitive development, sense of agency, language development

language acquisition at a developmental level, keeping this research close to both ethological and cognitive principles of explanation.

MUSICAL ORIGINS OF LANGUAGE: REVISITING DARWIN'S HYPOTHESIS

A close look to the empirical studies on animal communication reveals how EIP is widespread across a broad range of animal taxa. A comparative investigation will provide us with relevant information on the adaptive valence, and therefore on the evolutionary role, of such crucial dimensions in the domain of animal communication. Darwin provides an important insight on this topic:

> Primeval man, or rather some early progenitor of man, probably first used his voice in *producing true musical cadences, that is in singing*, as do some of the gibbon-apes at the present day; and we may conclude, from a wide-spread analogy, that this power would have been especially exerted during the courtship between sexes, – would have expressed various emotions, such as love, jealousy, triumph, – and would have served as a challenge to rivals.

> (Darwin, 1871, pp. 56–57; my emphasis).

Darwin's hypothesis that early humans were *singing*, as gibbons do today, has called for a comparative investigation into the ability to make "music" as a precursor of language (Rohrmeier et al., 2015). In order to gain a clearer understanding of the adaptive value of *musical* vocalizations in animals, and of its adaptive role for the emergence of human language, we need to examine: (i) to what extent it is correct to attribute musical abilities to non-human animals, and (ii) whether the ability to process EIP, rather then a general ability for music in non-human animals, can be considered an adaptive prerequisite necessary for the emergence of human language. I believe that making the distinction between a general aptitude for music and the use of EIP, might improve the investigation of the origins of language. This line of investigation will shed light on the adaptive role of EIP for the emergence of language, and perhaps of the ability for music itself in both human and non-human animals.

The question, then, is: Are gibbons, and non-human animals in general, able to make music in a way that is comparable to humans? Recent research has shown that birds, monkeys, and humans share the predisposition to distinguish consonant vs. dissonant music (Hulse et al., 1995; Izumi, 2000; Sugimoto et al., 2009). Moreover, studies suggest that rhesus macaques, *Macaca mulatta* (Wright et al., 2000), rats, *Rattus norvegicus* (Blackwell and Schlosberg, 1943), and dolphins, *Tursiops truncatus* (Ralston and Herman, 1995) are able to recognize two melodies as the "same" melody even when transposed one octave up or down. Songbirds, which in contrast miss this ability, have been shown to rely on absolute frequency over relative pitch within a scale (Cynx, 1995; Hoeschele et al., 2013). Furthermore, as Patel (2010) suggests, birdsong has a rhythm that, despite violating human metric conventions, is nonetheless stable and internally consistent. Recent research has also established

that some parrot species (*Cacatua galerita and Melopsittacus undulatus*) and a California sea lion (*Zalophus californianus*) are able to extract the pulse from musical rhythm, moving along with it (Fitch, 2013 for a review). Hence, we can accept that a biological inclination toward the ability for music is also present, to a certain extent, in non-human animals (Doolittle and Gingras, 2015; Fitch, 2015; Hoeschele et al., 2015).

However, non-human animals' ability to modulate sounds in courtship or rivalry contexts, which Darwin identified as a precursor of language, might be described, more parsimoniously, as an instance of EIP. Here, I suggest that the ability to modulate prosody in emotional communication and within turn-taking contexts (rather than the ability for music), as enough to describe the emergence of vocal utterances in the early *Homo*. Darwin's hypothesis may thus be updated in light of contemporary research and read in the following terms: the first hominins communicated exploiting prosody for emotional expression and communicative coordination. As I will clarify in the following sections, extensive research indicates that in different animal species the ability to vary prosodic features in the voice, in conjunction with the ability to coordinate sound production with others – expressing emotions, and possibly triggering emotional reactions – has an adaptive value. This use of prosody has positive effects in relation to sexual partner attraction, territory defense, group cohesion, parental care (Searcy and Andersson, 1986). Thus, the investigation of prosodic modulation of the voice provides an excellent, and surprisingly overlooked paradigm for a comparative approach addressing the adaptive features grounding the emergence of language. In the next sections, I will review studies reporting on EIP in non-human primates, non-primate mammals, birds, insects, and anurans.

EIP in Non-human Primates

The ability to modulate the prosodic features of a signal can be considered a homologous trait, i.e., a trait that humans and other primates share with their common ancestor. Experiments conducted both in the field and in captivity suggest that several species of prosimians and anthropoids are able to modulate spectro-temporal features of a call (frequency, tempo, and amplitude) as noise-induced vocal modifications (Hotchkin and Parks, 2013 for an extensive review). Research on chimpanzees' (*Pan troglodytes*) panthoots, a type of long-distance calls emitted while traveling or in the presence of abundant food sources, reveals individual and contextual modulation of the prosodic structure of this call (Notman and Rendall, 2005). De la Torre and Snowdon (2002) found that also pygmy marmosets, *Cebuella pygmaea*, adjust the frequency and temporal structure of their contact calls in a way appropriate to the frequency distortion effects of the habitats where they are located in order to maintain the acoustic structure of the long distance vocalization.

Studies provide evidence on arousal-related modulation of the call structure in non-human primates (Morton, 1977; Briefer, 2012). Specifically, it has been shown that high call rate (tempo), number of calls, and elevated fundamental frequency range correlate positively with high levels of arousal in chimpanzees, *Pan troglodytes* (Riede et al., 2007), squirrel monkeys, *Saimiri*

sciureus (Fichtel et al., 2001), bonnet macaques, *Macaca radiata* (Coss et al., 2007), vervet monkeys, *Macaca mulatta* (Seyfarth et al., 1980), rhesus monkeys, *Chlorocebus pygerythrus* (Hauser and Marler, 1993; Jovanovic and Gouzoules, 2001; Hall, 2009), baboons, *Papio papio* (Rendall et al., 1999; Seyfarth and Cheney, 2003), mouse lemurs, *Microcebus* spp. (Zimmermann, 2010), tree shrews, *Tupaia belangeri* (Schehka et al., 2007). It is important to stress that the modulation of these acoustic features of the signal derives from arousal-based physiological changes, thus these modulations are not under the voluntary control of the signaler. For instance, emotionally induced changes in muscular tone and coordination can affect the tension in the vocal folds, and consequently the fundamental frequency range of the vocalization and the voice quality of the caller (Rendall, 2003). Crucially, although the transmission of the emotional content of the signal is not intentional, the receivers are nonetheless sensitive to it, and are able to perceive, for instance, the level of urgency of the situation in which the call is produced, behaving in the most adaptive way (Zuberbühler et al., 1999; Seyfarth and Cheney, 2003). Further research is required to investigate whether different levels of arousal are encoded in (or decoded from) the structure of the interactive calls between conspecifics (Filippi et al., submitted), and whether the dynamics of alternate calling affects the emotional or attentive state of the signalers themselves.

Evidence suggests that non-human primates can coordinate the production of a signal with the vocal behavior of a mate or of other individuals of a group, modulating the acoustic features of vocalizations for communicative purposes. For instance, the ability for antiphonal calling, i.e., to flexibly respond to conspecifics in order to maintain contact between group members, has been reported in recent work conducted across prosimians, monkeys, and lesser apes: chimpanzees, *Pan troglodytes* (Fedurek et al., 2013), barbary macaques, *Macaca sylvanus* (Hammerschmidt et al., 1994), Campbell's monkeys, *Cercopithecus campbelli* (Lemasson et al., 2010), Diana monkeys, *Cercopithecus diana* (Candiotti et al., 2012), pygmy marmosets, *Cebuella pygmaea* (Snowdon and Cleveland, 1984), common marmosets, *Callithrix jacchus* (Miller et al., 2009), cotton-top tamarins, *Saguinus oedipus* (Ghazanfar et al., 2002), squirrel monkeys, *Saimiri sciureus* (Masataka and Biben, 1987), vervet monkeys, *Macaca mulatta* and *Chlorocebus pygerythrus* (Hauser, 1992), geladas, *Theropithecus gelada* (Richman, 2000) and Japanese macaques, *Macaca fuscata* (Sugiura, 1993; Lemasson et al., 2013). These so-called antiphonal vocalizations are guided by a sort of "turn taking" conversational rule system employed within an interactive and reciprocal dynamic between the calling individuals. Versace et al. (2008) found that cotton top tamarins, *Saguinus oedipus* can detect and wait for silent windows to vocalize. Call alternation in monkeys promotes social bonding and keeps the members of a group in vocal contact when visual access is precluded.

Turn-taking duet-like activities have been reported in caregiver-juvenile pairs in gibbons (Koda et al., 2013) and marmosets (Chow et al., 2015). In both species, caregivers interact with their juveniles, engaging in time-coordinated vocal feedback. This behavior scaffolds the development of

turn-taking and social competences in the juvenile marmosets, *Callithrix jacchus* (Chow et al., 2015), and seems to enhance vocal development in juvenile gibbons, *Hylobates agilis agilis* (Koda et al., 2013). Vocal duets in male-female pairs have been reported in: gibbons, *Hylobates* spp (Geissmann, 2000a), lemurs, *Lepilemur edwardsi*; (Méndez-Cárdenas and Zimmermann, 2009), common marmosets, *Callithrix jacchus* (Takahashi et al., 2013), the coppery titi, *Callicebus cupreus* (Müller and Anzenberger, 2002), squirrel monkeys, *Saimiri* spp. (Symmes and Biben, 1988), Campbell's monkeys, *Cercopithecus campbelli* (Lemasson et al., 2011), siamangs *Hylobates syndactylus* (Haimoff, 1981; Geissmann and Orgeldinger, 2000). Duets constitute a remarkable instance of interactional prosody, where members of a pair coordinate their sex-specific calls, effectively *composing* a single 'song' with two voices. Duets are interactive processes that involve time- and pattern-specific coordination among vocalizations flexibly exchanged between two individuals. Such a level of vocal coordination requires extensive practice over a long period of time. It seems that this investment strengthens the bond between the partners, since the quantity of duets performed is positively correlated with the pair bonding quality (measured by with grooming practice and physical proximity). In turn, the strength of the pair bonding also has positive adaptive effects on the management of parental care, territory defense, or foraging activities (Geissmann, 2000b; Geissmann and Orgeldinger, 2000; Müller and Anzenberger, 2002; Méndez-Cárdenas and Zimmermann, 2009).

From this set of studies we can infer that non-human primates possess the ability to process EIP, which is linked to group cohesion, territory defense, pair bonding, parental care, and social development. In conclusion, comparative review of studies on EIP in primates supports the hypothesis that these abilities have a functional role, and can thus be considered adaptive "homologous" traits in non-human primates.

EIP in Non-primate Mammals

Comparative research on non-primate mammals addressed the ability to modulate prosodic features of the voice, which express different levels of emotional arousal, and are used in interactive communications. These studies, focused on traits that are analogous in humans and non-primate mammals, are crucial within a comparative frame of research, as they may shed light on the selective pressures favoring the emergence of the human ability to process prosody as cue to language comprehension and maybe also of the human inclination for music.

Evidence has been reported on the ability to modulate the prosodic features of the vocal signals in several non-primate mammals: bottlenose dolphins, *Tursiops truncatus* (Buckstaff, 2004), humpback whales, *Megaptera novaeangliae* (Doyle et al., 2008), killer whales, *Orcinus orca* (Holt et al., 2009, 2011), right whales, *Eubalaena glacialis* (Parks et al., 2007, 2011), free-tailed bat, *Tadarida brasiliensis* (Tressler and Smotherman, 2009), mouse-tailed bat, *Rhinopoma microphillum* (Schmidt and Joermann, 1986), Californian ground squirrel, *Spermophilus beecheyi* (Rabin et al., 2003), and domestic cats, *Felis catus* (Nonaka et al., 1997). Only little attention has been devoted to the emotional content of calls in the species mentioned above.

However, recent research conducted on giant pandas, *Ailuropoda melanoleuca* (Stoeger et al., 2012) and on African elephants, *Loxodonta africana* (Soltis et al., 2005b; Stoeger et al., 2011) provides evidence that in mammals high levels of arousal can be expressed through specific acoustic features in the signal, namely: noisy and aperiodic segments, increased call duration and elevated fundamental frequency. The effective expression and perception of emotional arousal may allow individuals to respond appropriately, based on the degree of urgency or distress encoded in the call. Thus, the ability to process these calls correctly may be crucial for survival under natural conditions.

In addition, studies indicate that, in the case of conflicts or separation from the group and when visual cues are not available, the following species of mammals produce antiphonal calls to signal their identity or spatial location: African elephants, *Loxodonta africana* (Soltis et al., 2005a), Atlantic spotted dolphins, *Stenella frontalis* (Dudzinski, 1998), bottlenose dolphins, *Tursiops truncatus* (Janik and Slater, 1997; Kremers et al., 2014), white-winged vampire bats, *Diaemus youngi* (Carter et al., 2008, 2009; Vernes, 2016), horseshoe bats, *Rhinolophus ferrumequinum nippon* (Matsumura, 1981), killer whales, *Orcinus orca* (Miller et al., 2004), sperm whales, *Physeter macrocephalus* (Schulz et al., 2008), and naked mole-rats, *Heterocephalus glaber* (Yosida et al., 2007). Individuals in all these species alternate calls, following specific patterns of response timing to maintain group cohesion and bonding relationships. Furthermore, vocal duets have been reported in Cape-mole rats, *Georychus capensis* (Narins et al., 1992). Members of this species alternate seismic signals (generated by drumming their hind legs on the burrow floor) to attract sexual mates.

In sum, the studies reviewed in this section indicate that the ability to process EIP is present also in non-primate mammals, where it might have evolved as adaptive "analogous traits," i.e., under the same selective pressures (group cohesion, territory defense, pair bonding, parental care) that triggered its emergence in primates.

EIP in Birds

The study of mechanisms and processes underlying EIP in birds has revealed multiple analogous traits, i.e., strong evolutionary convergences, with vocal communication in humans. By shedding light on the selective pressures grounding the emergence of EIP in species that are phylogenetically distant, as it is the case for humans and birds, this line of research may enhance our understanding of the evolutionary path of the ability to process linguistic prosody (and perhaps also music) in humans.

Differently to mammalians, in birds, sounds are produced by airflow interruption through vibration of the labia in the syrinx (Gaunt and Nowicki, 1998). Modulation in bird vocalization is thought to originate predominantly from the sound source (Greenewalt, 1968), while the resonance filter shapes the complex time-frequency patterns of the source (Nowicki, 1987; Hoese et al., 2000; Beckers et al., 2003). For instance, songbirds are able to change the shape of their vocal tract, tuning it to the fundamental frequency of their song (Riede et al., 2006; Amador et al., 2008).

Importantly, variations in the prosodic features of the calls may be indicative of the emotional state of the signaler. The expression of arousal and/or emotional information through the modulation of prosody in birds has been shown in chickens, *Gallus gallus* (Marler and Evans, 1996), ring doves, *Streptopelia risoria* (Cheng and Durand, 2004), Northern Bald Ibis, *Geronticus eremita* (Szipl et al., 2014), black-capped chickadees, *Poecile atricapillus* (Templeton et al., 2005; Avey et al., 2011). The ability to process different levels of emotional arousal in bird vocalizations serve numerous functions including signaling type and degree of potential threats, dominance in agonistic contexts, or the presence of high quality food (Ficken and Witkin, 1977; Evans et al., 1993; Griffin, 2004; Templeton et al., 2005).

As to the interactional dimension of prosody, evidence for choruses has been reported in: Common mynas, *Acridotheres tristis* (Counsilman, 1974), Australian magpies, *Gymnorhina tibicen* (Brown and Farabaugh, 1991), and in black-capped chickadees, *Poecile atricapillus* (Foote et al., 2008). This activity has been shown to favor social bonding, synchronization of activities, and group or territory defense.

Research has described the capacity to modulate and coordinate vocal productions in antiphonal calling between individuals of different groups in European starlings, *Sturnus vulgaris* (Hausberger et al., 2008) and in nightingales, *Luscinia megarhynchos* (Naguib and Mennill, 2010). Crucially, Henry et al. (2015) found that prosodic features of vocal interactions in starlings are influenced by the immediate social context, the individual history, and the emotional state of the signaler. Camacho-Schlenker et al. (2011) suggest that in winter wrens, *Troglodytes troglodytes*, call exchanges among neighbors might have different aggressive/submissive values. Thus, these antiphonal calls can escalate in territorial contests, influencing females' mate choice.

Multiple studies report duets in songbirds. Indeed, duets among sexual partners, which coordinated their phrases by alternation or overlap, are widespread among songbirds. As in non-human primates, they help to maintain pair bonds and are used to defend territories or resources. Duets have been reported in: fred-backed fairy-wrens, *Malurus melanocephalus* (Baldassarre et al., 2016; reviews: Langmore, 1998; Hall, 2009; Dahlin and Benedict, 2014). Notably, the capacity to coordinate the production of sounds with the vocalizations of a partner requires control over the modulation of phonation in frequency, tempo, and amplitude. Dilger (1953) suggests that in crimson-breasted barbets, *Psilopogon haemacephalus*, the coordination of two sexual mates in duetting could affect the production of reproductive hormones, thereby ensuring synchrony in the reproductive status of the breeding partners. Thus, the ability to coordinate or synchronize vocal sounds has an adaptive value that may have guided the evolution of song complexity and plasticity in songbirds (Kroodsma and Byers, 1991). Indeed, the ability to produce complex sequences of sounds is indicative of an individual's capacity to memorize complex sequences and how fine a caller's motor and neural control is over the sounds of the song (Searcy and Andersson, 1986; Langmore, 1998). This strong index of mental and physical skills is shown to be important in a mate choice context in zebra finches, *Taeniopygia*

guttata (Neubauer, 1999) and Bengalese finches, *Lonchura striata* (Okanoya, 2004). Similarly, recent research conducted on humans suggest that women have sexual preferences during peak conception times for men that are able to create more complex sequences of sounds (Charlton et al., 2012; Charlton, 2014).

Importantly, both in humans and songbirds vocal learning has an interactive dimension. Interestingly, in both groups, the ability to alternate and coordinate vocalizations with conspecifics is acquired by interactive tutoring with adult conspecifics (Poirier et al., 2004; Feher et al., 2009; see section "EIP in Language Acquisition" below). Goldstein et al. (2003) argue that such convergence reveals that the social dimension is an important adaptive pressure that favored the acquisition of complex vocalizations in humans and songbirds (Syal and Finlay, 2011).

Taken together, studies reporting on EIP in songbirds support the hypothesis that the ability to modulate prosodic features of the calls, marking emotional expression and interactional coordination, can be identified as an analogous and adaptive trait that humans and songbirds share. Thus, based on these data, we can infer that the abilities involved in EIP might have set the ground for the emergence of language in humans.

EIP in Anurans

The adaptive and functional value of EIP emerges quite clearly also considering research on a variety of anurans' species, which are notably phylogenetically very distant to the *Homo* line. As in humans, and generally, similarly to mammals, the source of vocal sounds in anurans is airflow interruption through vibration of the vocal folds in the larynx (Dudley and Rand, 1991; Prestwich, 1994; Fitch and Hauser, 1995). Calls emitted in different contexts, such as sexual advertisement and male-male aggression, show clear spectral, and acoustic differences (Pettitt et al., 2012; Reichert, 2013). Although it has never been tested empirically, it is plausible that these different call features reflect differences in the level of emotional arousal in the signaler.

In most species of anurans investigated so far, males acoustically compete for females under conditions of high background noise produced by conspecifics. As a consequence, males have developed calling strategies for improving their conspicuousness, i.e., the ability to fine-tune the timing of their calls according to the prosodic and spectral characteristics of the acoustic context (Grafe, 1999).

Anurans aggregate in choruses. The ability for simultaneous acoustic signaling in choruses might have evolved as an anti-predator behavior – specifically, to confuse the predators' auditory localization abilities (Tuttle and Ryan, 1982) and under sexual selection pressures, as females prefer collective calls to individual male calls. In fact, besides being heard as a group, males have to produce a signal that could stand out from the collective sound in order to attract the female. In order to be heard as a "leader," advertising individual qualities (Fitch and Hauser, 2003), each signaler has to emit a signal faster than his neighbor. This "time pressure" eventually results in a very tight overlap or synchronization of signals between calling individuals. Females in most species of anurans prefer

the calls of "leaders," individuals that emit more prominent calls (Klump and Gerhardt, 1992), flexibly adjusting their onsets accordingly. Evidence suggests that females in the Afrotropical species *Kassina fusca* prefer leading male calls when the degree of call overlap with the other signallers is high (75 and 90%). However, intriguingly, in this species, females prefer follower male calls when the degree of call overlap is low (10 and 25%). Thus, follower males in *K. fusca* actively adjust their overlap timing in accordance to their vocalizing neighbors, in order to attract females (Grafe, 1999). Ryan et al. (1981) found that in the neotropical frog *Physalaemus pustulosus*, singing in a chorus is adaptive as it decreases the risk of being attacked by a predator and at the same time, increases mating opportunities.

Antiphonal calling in anurans has never been reported. In contrast, duets are described in: the Neotropical *Caphiopus bombifrons* and *Pternohyla fodiens* (Bogert, 1960), the common Mexican treefrogs, *Smilisca baudinii* and in the genuses *Eleutherodactylus* and *Phyllobates* (Duellman, 1967). Tobias et al. (1998) reported remarkable duetting behaviors in the South African clawed frog, *Xenopus* laevis. Females in this species have a very short sexual receptivity time window, in which they have to accurately locate a potential sexual mate. This is not an easy task, considering the high population density and the low visibility in their natural habitat. These constraints may have led to fertility advertisement call by females (rapping) when oviposition is imminent. Tobias et al. (1998) found that females swim to an advertising male and produce the rapping call, which stimulates male approach and elicits an answer call. Thus, the two sexes respond to each other's calls (which partially overlap), a behavior that results in a rapping–answer interaction. Interestingly, Bosch and Márquez (2001) found that in midwife toads, *Alytes obstetricans*, males engage in duets in competitive contexts. This research suggests that, when duetting, males adjust the temporal structure of their calls, increasing calling rate. This variations correlates with the caller's body size and seems to affect females' mate choice.

EIP In Insects

Crucial implications for the understanding of EIP in humans may derive from research on insects. Notably, this animal taxon is phylogenetically quite distant to humans. Therefore, comparative work on EIP in humans and insects is a perfect candidate to highlight selective pressures underlying the ability to process the prosodic modulation of sounds marking emotional expression and interactional coordination.

It is worth remarking that the mechanisms underlying sound production in insects are extremely different than the ones possessed by the animal taxa reviewed so far. In fact, insects produce advertising or aggressive sounds through stridulation, i.e., vibration of a specific sound source generating by rubbing two body structures against each other, for instance, the forewings in crickets and katydids, or the legs across a sclerotized plectrum, in grasshoppers (Prestwich, 1994; Bennet-Clark, 1999; Hartbauer and Römer, 2016). In the *Expression of the emotions in man and animals*, Darwin (1872) observed that although stridulation is generally used to emit a sexual advertisement signal, bees may

vary the degree of stridulation to express different emotional intensities. However, to my knowledge, the auditory expression of emotional arousal in insects has received only little empirically investigation to date (Brüggemeier et al., submitted; Rezával et al., 2016). In contrast, much research on this class of animals has addressed the ability for interactional coordination in sound production.

As to the study of inter-individual coordination as an adaptive analogous trait in humans and insects, it is important to refer to a striking phenomenon in the visual domain: fireflies, winged beetles in the family of *Lampyridae*, use their ability for bioluminescence in courtship or mating contexts (Greenfield, 2005; Ravignani et al., 2014a). Several species of this family are able to entrain in highly precise synchronized flashing, probably to create a more prominent signal to potential mates from a remote location (Buck and Buck, 1966).

Similarly to the case of bioluminescent signals in fireflies, several species of insects have the ability to coordinate timing patterns of their acoustic signals. Specifically, male individuals tend to synchronize their signal within choruses. In ratter ants (genus: *Camponotus*), the ability to entrain in synchronized signal production has evolved as an anti-predator behavior (Merker et al., 2009). However, in most species of studied insects, this ability seems to have evolved under sexual selection pressures (Alexander, 1975; Greenfield, 1994a,b; Yoshida and Okanoya, 2005; Ravignani et al., 2014a). Typically, only males generate acoustic signals, and the mute females approach the singing males. To produce a louder signal that has a better chance to be heard by (and attract) females from a greater distance, advertising males of the tropical katydid species *Mecopoda elongata* tend to synchronize the production of acoustic sounds (Hartbauer and Römer, 2016). Synchrony maximizes the peak signal amplitude of group display, an emergent property known as the "beacon effect" (Buck and Buck, 1966).

In the Neotropical katydid *Neoconocephalus spiza*, females display a strong preference for males that produce a signal after a slight lag, or alternatively, to coincide with, but slightly lead, the other males (Greenfield and Roizen, 1993). As for anurans, male insects have to produce prominent signals to stand out from the group and attract a sexual mate. In the *M. elongata*, in order to lead the chorus, thus being heard by the female, each signaler has to emit a signal before another individual, and at a higher amplitude (Hartbauer et al., 2014). Thus, each male's emission rate becomes increasingly faster, resulting in synchronization of signals. This suggests that time-coordinated (in this case, synchronized) collective signal is an epiphenomenon created by competitive interactions between males within sexual advertisement contexts (Greenfield and Roizen, 1993). Sismondo (1990) has shown that in *M. elongata*, the dynamic of sound production between leaders and followers has oscillator properties, a finding that echoes data from research on turn-taking dynamics in human conversations.

Antiphonal calling in insects has never been reported. Nonetheless, in multiple orders of insects, individuals of opposite sex engage in time-coordinated duets initiated by the male, with the female replying within a time window that is often species-specific (Zimmermann et al., 1989; Bailey, 2003). Males initiating

a duet often insert a trigger pulse at the conclusion of their call, and the females might use this as a cue to which they may reply (Bailey and Field, 2000). Bailey (2003) hypothesized that, in duetting species, females evolved the ability to reply to males to counterbalance predation risk and energy consumption linked to the production of complex and long sounds in males. The author suggested that signal prominence decrease as a result of a counter-selection pressure from male costs.

EIP IN LANGUAGE ACQUISITION

As detailed in the previous sections, much research reports on the ability for EIP across a diverse range of animal taxa providing data on both homologous and analogous traits involved in EIP, thus on their adaptive and functional value. These data, combined with evidence on the pervasive use of prosodic modulation of the voice in linguistic communications in modern humans, support the hypothesis that the ability to process EIP might have evolved into the human ability to process linguistic prosody. This holds true not only on a phylogenetic scale, but also for human language development, i.e., on an ontogenetic scale.

When talking to infants, parents of different languages and cultures typically use vocal patterns that are distinct from speech directed at adults: this special kind of speech, commonly referred to as infant-directed speech (hereafter IDS), is often characterized by shorter utterances, longer pauses, higher pitch, exaggerated intonational contours (Fernald and Simon, 1984; Fernald et al., 1989) and expanded vowel space (Kuhl, 1997; de Boer, 2005). IDS is a good example for the ontogenetic role of EIP in humans, with striking effects both on children's acquisition of language and their development of social cognition. Recent research suggests that caregivers across multiple cultures instinctively adjust their speech prosodic features to their infants (Kitamura et al., 2001; Burnham et al., 2002).

As Fernald (1992) observes, by intuitively moving to a pitch range that an infant is more sensitive to (i.e., where the perceived loudness of the signal is increased), mothers compensate for the infants' auditory limitations. Indeed, it has been shown that infants' threshold of auditory brainstem responses (ABR) are higher by 3–25 dB than adult ABR thresholds (Sininger et al., 1997). Given that neonates have greater auditory limitations than adults (Schneider et al., 1979), the speech addressed to neonates needs to be more intense in order to be effectively perceived. A sound of 500 Hz has a higher frequency and will be perceived by the human hearing system as louder than a sound at 150 Hz with the same intensity. It follows that speech with a higher frequency will be more salient to the infant. Therefore, frequency changes seem to be particularly salient: infants tested in an operant auditory preference procedure showed a strong listening preference for the frequency contours of IDS, but not for other associated patterns such as amplitude or duration (Fernald and Kuhl, 1987; Cooper et al., 1997).

The prosodic features typical of IDS modulate the infants' attention and emotional engagement (Fernald and Simon, 1984; Locke, 1995), scaffolding language development. The specific acoustic parameters used in IDS are very effective in communicating prohibition, approval, comfort, and attention bid (Papoušek et al., 1990; Fernald, 1992; Bryant and Barrett, 2007) and also in conveying emotional content such as love, fear, and surprise (Trainor et al., 2000). Therefore, sound modulation typical of IDS elicits attention and emotional responses in the infants, and conveys crucial information about the speaker's communicative intent (Fernald, 1989). In addition, the exaggerated pitch parameter cross-culturally employed in IDS provides markers that have the following uses: (a) to highlight target words (Grieser and Kuhl, 1988; Fernald and Mazzie, 1991), (b) to convey language-specific phonological information (Burnham et al., 2002; Kuhl, 2004), (c) as cues to word learning (Thiessen et al., 2005; Filippi et al., 2014), or (d) as cues to the syntactic structure of sentences (Sherrod et al., 1977; Fernald and McRoberts, 1996).

Crucially, caregivers combine sounds and modulate the intonation (frequency, tempo, and amplitude) of speech, engaging in time-coordinated vocal interactions with the children. Contingent responsiveness from caregivers, thus interactive coordination, facilitates language learning (Goldstein et al., 2003; Kuhl et al., 2003; Gros-Louis et al., 2006; Goldstein and Schwade, 2008; Rasilo et al., 2013), and improves the child's accuracy in speech production. Moreover, caregiver-child interactional coordination scaffolds the child's social development (Todd and Palmer, 1968; Fernald et al., 1989; Goldstein et al., 2003; Goldstein and Schwade, 2008; Brandt et al., 2012), and her/his acquisition of social conventions, such as turn-taking in conversations (Weisberg, 1963; Kuhl, 1997; Jaffe et al., 2001). Keitel et al. (2013) found that 3-year-old children strongly rely on prosodic information to process conversational turn-taking. Thus, prosodic intonation, in combination with lexico-syntactic information is used by adults and infants as cues to anticipate upcoming turn transitions (Lammertink et al., 2015). In summary, a number of studies indicate that IDS promotes the social and emotional development of infants and favors the acquisition of language. Based on these findings, we can conclude that IDS constitutes a relevant biological signal (Fernald, 1992).

Bringing together comparative data on caregiver-infant communication in humans and chimpanzees with paleo-anthropological evidence, Falk (2004) suggested that it is very likely that the first forms of IDS in the early hominins evolved as the trend for enlarging brain size, which made parturition increasingly difficult. This caused a selective shift toward females that gave birth to neonates with relatively small and underdeveloped brains who were, consequently, strongly dependent on caretakers for survival. According to this hypothesis, humans started to make use of prosodic modulations in order to engage infants' attention, and to convey affective messages to them while engaging in other activities. Interestingly, this would explain why humans are the only species where tutors exaggerate the prosodic features of the signal when speaking to immature offspring. Based on this research, I propose that the use of prosody for emotional communication and interactional coordination was critical for the evolutionary emergence of the first vocalizations in humans. EIP can thus be considered a critical biological ability adopted by humans on both a phylogenetic and an ontogenetic scale.

COGNITIVE LINK BETWEEN LINGUISTIC PROSODY AND MUSIC: IS EIP THEIR EVOLUTIONARY COMMON GROUND?

Music is a universal ability performed in all human cultures (Honing et al., 2015; Trehub et al., 2015) and has often been identified as an evolutionary precursor of language (Brown, 2001; Mithen, 2005; Fitch, 2006, 2010, 2012; Patel, 2006). A number of studies hypothesize that the musical abilities attested in different species of animals constitute homologous or analogous traits, which paved the evolution of language in humans (Geissmann, 2000a; Marler, 2000; Fitch, 2005; Berwick et al., 2011). This line of research follows up on Darwin's hypothesis on the musical origins of language (see section: "Darwin's Hypothesis: In the Beginning Was the Song").

The studies on EIP across animal taxa reviewed in the previous sections, taken together, have crucial implications for this line of research on the origins of language: it is plausible that the ability for EIP evolved into the ability to process linguistic prosody (namely prosodic cues to lexical units, syntactic structure, and discourse structure comprehension), and perhaps also into the ability for music itself. If this is true, than shared traits between the human abilities for linguistic prosody and music should be empirically observable. Indeed, multiple studies show a large overlap between these two domains. Koelsch (2012) suggested that music and language can be positioned along a continuum in which the boundary distinguishing one from the other is quite blurry (Jackendoff, 2009; Patel, 2010). Two interesting cases in which the ability to process linguistic prosody overlaps with music are the so-called talking drums and the whistled languages (Meyer, 2004): the talking drums are instruments whose frequencies can be modulated to mimic tone and prosody of human spoken languages. Whistled language speakers use whistles to emulate the tones or vowel formants of their natural language, keeping its prosody contours (Remez et al., 1981), as well as its full lexical and syntactic information (Carreiras et al., 2005; Güntürkün et al., 2015). Intriguingly, although left-hemisphere superiority has been reported for atonal and tonal languages, click consonants, writing, and sign languages (Best and Avery, 1999; Levänen et al., 2001; Marsolek and Deason, 2007; Gu et al., 2013), recent brain studies (Carreiras et al., 2005; Güntürkün et al., 2015) suggest that whistled language comprehension relies on symmetric hemispheric activation. In addition, empirical evidence from brain imaging research indicates that the ability to process prosodic variations in language plays a vital role in the comprehension of both verbal and musical expressions. For instance, amusic subjects show deficits in fine-grained perception of pitch (Peretz and Hyde, 2003), failing to distinguish a question from a statement solely on the basis of changes in pitch direction (Patel et al., 2008; Liu et al., 2010). This observed difficulty in a sample of amusic patients supports the hypothesis that music and prosody share specific neural resources for processing pitch patterns (Ayotte et al., 2002). Further brain imaging studies report a considerable overlap in the brain areas involved in the perception of pitch and rhythm patterns in words and songs (Zatorre et al., 2002; Patel,

2003; Merill et al., 2012), and in sound patterns processing in melodies and linguistic phrases (Brown et al., 2006). Therefore, based on the outcome of this line of research, we can conclude that the abilities underpinning linguistic prosody and music share cognitive and neural resources. However, is it plausible to identify in EIP an evolutionary common ground for both abilities?

To date, the cognitive and evolutionary link between the ability to process prosody as cue to the emotional state of the signaler and the ability to use prosody as guide to word recognition, or to syntactic and discourse structure, remains open to empirical investigation. In contrast much research has examined the cognitive link between the ability to process emotional prosody and music in humans, showing that in both music and language, specific emotions (e.g., happiness, sadness, fear, or anger) are expressed through similar patterns of pitch, tempo, and intensity (Scherer, 1995; Juslin and Laukka, 2003; Fritz et al., 2009; Bowling et al., 2012; Cheng et al., 2012). For instance, in both channels, happiness is expressed by fast speech rate/tempo, medium-high voice intensity/sound level, medium-high frequency energy, high F0/pitch level, much F0/pitch variability, rising F0/pitch contour, fast voice onsets/tone attacks (Juslin and Laukka, 2003). Research on this topic suggests that musical melodies and emotional prosody are two channels that use the same acoustic code for expressing emotional and affective content.

As to the evolutionary link between the ability to use prosodic cues to coordinated interactions in auditory communication and social entrainment in music, studies conducted on humans suggests that a strong motivation to engage in frames of coordinated activities such as social entrainment or synchronization, favor adaptive behaviors, and specifically, the inclination to cooperate (Hagen and Bryant, 2003; Wiltermuth and Heath, 2009; Kirschner and Tomasello, 2010; Koelsch, 2013; Manson et al., 2013; Morley, 2013; Launay et al., 2014; Tarr et al., 2014). Consistent with these findings, Phillips-Silver et al. (2010) suggest that the ability for coordinated rhythmic movement, and thus entrainment, applies to music and dance as well as to other socially coordinated activities. From their perspective, the ability for music and dance might be rooted in a broader ability for social entrainment to rhythmic signals, which spans across communicative domains and animal species.

Social engagement in time-coordinated activities, as interactive communications or music – promotes prosocial behaviors (Cirelli et al., 2014; Ravignani, 2015). These adaptive behaviors might have favored the evolution of language, including the ability to process and exchange prosodically modulated linguistic utterances within coordinated interactions – on a phylogenetic scale (Noble, 1999; Smith, 2010).

Crucially, in line with this hypothesis, recent findings suggest that social coordination favors word learning also in modern human adults (Verga et al., 2015). Within this frame of research, empirical evidence indicates that children with communicative disorders benefit from music therapy for social skills such as initiative, response, vocalization within an interactive frame of communication (Müller and Warwick, 1993; Bunt and Marston-Wyld, 1995; Elefant, 2002; Oldfield et al., 2003). These findings

are consistent with comparative work on the brain neuroanatomy in humans and birds suggesting that social motivation and affect played a key role in the emergence of language at both a developmental and a phylogenetic scale (Syal and Finlay, 2011).

Taken together, these studies point to the existence of a biologically rooted link between (i) the ability to use prosody for the expression of emotions, interactional coordination between multiple individuals, and language processing, and (ii) the ability to process music. However, the hypothesis that the ability for EIP played a crucial role in the emergence of a fully-blown linguistic and musical abilities in humans is currently open to empirical investigation (Bryant, 2013).

CONCLUSIONS

Theories on the origins of language often identify the musical aspect of speech as a critical component that might have favored, or perhaps triggered, its emergence (Rousseau, 1781; Darwin, 1871; Jespersen, 1922; Livingstone, 1973; Richman, 1993; Brown, 2000; Merker, 2000). Indeed, evidence of shared cognitive processes in music and human language has led to the hypothesis that these two faculties were intertwined during their evolution (Brown, 2001; Mithen, 2005; Fitch, 2006, 2010, 2012; Patel, 2006). Crucially, multiple studies have identified musical behaviors shown in different species of animals (Geissmann, 2000a; Marler, 2000; Fitch, 2005; Berwick et al., 2011), as precursors for the evolution of language.

However, in this article I proposed to address the focus of research on language evolution and development toward the ability to process prosody for emotional communication and interactional coordination. This ability, which is widespread across animal taxa, might have evolved into the ability to process prosodic modulation of the voice as cue to language processing, and perhaps also into the biological inclination to music. In support of this hypothesis, I reviewed a number of studies reporting adaptive uses of EIP in non-human animals, where it evolved as anti-predator defense, social development, sexual advertisement, territory defense, and group cohesion. Based on these studies, we can infer that EIP provided the same adaptive advantages to early hominins (Pisanski et al., 2016). In addition, I reviewed research pointing to the processes involved in EIP as common evolutionary traits grounding the abilities to process linguistic prosody and music.

In the course of speech evolution, an increased control of pitch contour might have enabled a greater vocal versatility and expressiveness, building on the limited pitch-control used for emotive, social vocalizations already in use amongst higher primates (Morley, 2013).

This hypothesis is consistent with the "prosodic protolanguage" version of Darwin's musical protolanguage suggested by Fitch (2005). According to this model, the first linguistic utterances produced by humans, similar to birdsong, were internally complex, lacked propositional meaning, but could be learned and culturally transmitted. The prosodic protolanguage hypothesis harmonizes with the "holistic protolanguage" model (Jespersen, 1922; Wray, 1998), according

to which early humans modulated the prosodic values of their vocalizations, conveying messages as whole utterances that were strongly dependent on the context of use. By this model, this first stage was then followed by a process of gradual fractionation of these holistic, prosodically modulated units into smaller items. It is plausible that this process paved the emergence of propositions ruled by combinatorial principles that would increase their learnability, thus the possibility of their cultural transmission (Kirby et al., 2008; Verhoef, 2012). The identification of the cognitive mechanisms underlying EIP has implications for our understanding of the processes involved in the production and perception of such songbird-like protolanguage, thus of the evolutionary process that led to language.

The beneficial value of EIP is evident in modern humans, particularly in the case of speech addressed to preverbal infants, where it favors the developmental process of language learning and emotional bonding. The comparative studies reviewed in this paper indicate that the prosodic modulation of sounds within an interactive and emotion-related dynamic is a critical ability that might have favored the evolution of spoken language (aiding emotion processing, group coordination, and social bonding), and continues to play a striking role in the acquisition of language in humans (Syal and Finlay, 2011). Further empirical research is required to analyze how the ability to modulate prosody for emotional communication and interactional coordination favors the production and perception of the constitutive building blocks of language (phonemes and morphemes) and of the syntactic connections between words or phrases. This line of research might be conducted on infants, by investigating the developmental benefits of EIP on language processing.

Comparative studies have addressed the ability to process linguistic prosody, e.g., trochaic vs. iambic stress patterns in non-human animals (Ramus et al., 2000; Toro et al., 2003; Yip, 2006; Naoi et al., 2012; de la Mora et al., 2013; Spierings and ten Cate, 2014; Hoeschele and Fitch, 2016; Toro and Hoeschele, submitted). Moreover, research has examined non-human animals' ability to perceive or produce phonemes (Bowling and Fitch, 2015; Kriengwatana et al., 2015). Nonetheless, to my knowledge, the effect of EIP on the perception of the building blocks of heterospecific or conspecific communication systems in non-human animals is still open to empirical examination.

The integration of these studies within a research framework focused on the functional valence of prosodic modulation of the voice in animals, i.e., to its emotional, motivational, and socially coordinated dimensions – will favor a deeper understanding of the evolutionary roots of human emotional and linguistic interactions (Anderson and Adolphs, 2014). Additionally, comparative research on non-human animals and pre-verbal infants, combined with new methods to explore emotional and interactive sound modulation in music and language from a neural and behavioral perspective, promise empirical, and theoretical progress. This investigative framework may ultimately result into new empirical questions targeted at a deeper understanding of the inter-individual, multimodal dimension of communication.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and approved it for publication.

FUNDING

ACKNOWLEDGMENTS

REFERENCES

Alexander, R. D. (1975). "Natural selection and specialized chorusing behavior in acoustical insects," in *Insects, Science and Society*, ed. D. Pimentel (New York, NY: Academic Press), 35–77.

Amador, A., Goller, F., and Mindlin, G. B. (2008). Frequency modulation during song in a suboscine does not require vocal muscles. *J. Neurophysiol.* 99, 2383–2389. doi: 10.1152/jn.01002.2007

Anderson, D. J., and Adolphs, R. (2014). A framework for studying emotions across species. *Cell* 157, 187–200. doi: 10.1016/j.cell.2014.03.003

Avey, M. T., Hoeschele, M., Moscicki, M. K., Bloomfield, L. L., and Sturdy, C. B. (2011). Neural correlates of threat perception: neural equivalence of conspecific and heterospecific mobbing calls is learned. *PLoS ONE* 6:e23844. doi: 10.1371/journal.pone.0023844

Ayotte, J., Peretz, I., and Hyde, K. (2002). Congenital amusia: a group study of adults afflicted with a music-specific disorder. *Brain* 125, 238–251. doi: 10.1093/brain/awf028

Bailey, W. J. (2003). Insect duets: underlying mechanisms and their evolution. *Physiol. Entomol.* 28, 157–174. doi: 10.1046/j.1365-3032.2003.00337.x

Bailey, W. J., and Field, G. (2000). Acoustic satellite behaviour in the Australian bushcricket *Elephantodeta nobilis* (Phaneropterinae, Tettigoniidae, Orthoptera). *Anim. Behav.* 59, 361–369. doi: 10.1006/anbe.1999.1325

Baldassarre, D. T., Greig, E. I., and Webster, M. S. (2016). The couple that sings together stays together: duetting, aggression and extra-pair paternity in a promiscuous bird species. *Biol. Lett.* 12:20151025. doi: 10.1098/rsbl.2015.1025

Beckers, G. J., Suthers, R. A., and Ten Cate, C. (2003). Pure-tone birdsong by resonance filtering of harmonic overtones. *Proc. Natl. Acad. Sci. U.S.A.* 100, 7372–7376. doi: 10.1073/pnas.1232227100

Bennet-Clark, H. C. (1999). Resonators in insect sound production: how insects produce loud pure-tone songs. *J. Exp. Biol.* 202, 3347–3357.

Berwick, R. C., Okanoya, K., Beckers, G. J. L., and Bolhuis, J. J. (2011). Songs to syntax: the linguistics of birdsong. *Trends Cogn. Sci.* 15, 113–121.

Best, C. T., and Avery, R. A. (1999). Left-hemisphere advantage for click consonants is determined by linguistic significance and experience. *Psychol. Sci.* 10, 65–70. doi: 10.1111/1467-9280.00108

Blackwell, H. R., and Schlosberg, H. (1943). Octave generalization, pitch discrimination, and loudness thresholds in the white rat. *J. Exp. Psychol.* 33, 407–419. doi: 10.1037/h0057863

Bogert, C. M. (1960). "The influence of sound on the behavior of amphibians and rep- tiles," in *Animal Sounds and Communication*, eds W. E. Lanyon and W. N. Tavolga (Washington, DC: American Institute of Biological Sciences), 137–320.

Bolinger, D. (1972). Accent is predictable (if you're a mind-reader). *Language* 48, 633–644. doi: 10.2307/412039

Bosch, J., and Márquez, R. (2001). Call timing in male-male acoustical interactions and female choice in the midwife toad *Alytes obstetricans*. *Copeia* 2001, 169–177. doi: 10.1643/0045-8511(2001)001%5B0169:CTIMMA%5D2.0.CO;2

Bowling, D. L., and Fitch, W. T. (2015). Do animal communication systems have phonemes? *Trends Cogn. Sci.* 19, 555–557. doi: 10.1016/j.tics.2015.08.011

Bowling, D. L., Sundararajan, J., Han, S., and Purves, D. (2012). Expression of emotion in eastern and western music mirrors vocalization. *PLoS ONE* 7:e31942. doi: 10.1371/journal.pone.0031942

Brandt, A., Gebrian, M., and Slevc, L. R. (2012). Music and early language acquisition. *Front. Psychol.* 3:327. doi: 10.3389/fpsyg.2012.00327

Briefer, E. F. (2012). Vocal expression of emotions in mammals: mechanisms of production and evidence. *J. Zool.* 288, 1–20. doi: 10.1111/j.1469-7998.2012.00920.x

Brown, E. D., and Farabaugh, S. M. (1991). Song sharing in a group-living songbird, the Australian magpie, *Gymnorhina tibicen*. Part III. Sex specificity and individual specificity of vocal parts in communal chorus and duet songs. *Behaviour* 118, 244–274.

Brown, S. (2000). "The 'Musilanguage' model of music evolution", In *The Origins of Music*, eds N. L. Wallin, B. Merker, and S. Brown (Cambridge, MA: The MIT Press), 271–300.

Brown, S. (2001). Are music and language homologues? *Biol. Found. Music* 930, 372–374.

Brown, S., Martinez, M. J., and Parsons, L. M. (2006). Music and language side by side in the brain: a PET study of the generation of melodies and sentences. *Eur. J. Neurosci.* 23, 2791–2803. doi: 10.1111/j.1460-9568.2006.04785.x

Bryant, G. A. (2013). Animal signals and emotion in music: coordinating affect across groups. *Front. Psychol.* 4:990. doi: 10.3389/fpsyg.2013.00990

Bryant, G. A., and Barrett, H. C. (2007). Recognizing intentions in infant-directed speech evidence for universals. *Psychol. Sci.* 18, 746–751. doi: 10.1111/j.1467-9280.2007.01970.x

Buck, J., and Buck, E. (1966). Mechanisms of rhythmic synchronous flashing of fireflies. *Polymer* 7:232.

Buckstaff, K. C. (2004). Effects of watercraft noise on the acoustic behavior of bottlenose dolphins, *Tursiops truncatus*, in Sarasota Bay, Florida. *Mar. Mamm. Sci.* 20, 709–725. doi: 10.1111/j.1748-7692.2004.tb01189.x

Bunt, L., and Marston-Wyld, J. (1995). Where words fail music takes over: a collaborative study by a music therapist and a counselor in the context of cancer care. *Music Ther. Perspect.* 13, 46–50. doi: 10.1093/mtp/13.1.46

Burnham, D., Kitamura, C., and Vollmer-Conna, U. (2002). What's new, pussycat? On talking to babies and animals. *Science* 296, 1435–1435.

Camacho-Schlenker, S., Courvoisier, H., and Aubin, T. (2011). Song sharing and singing strategies in the winter wren *Troglodytes troglodytes*. *Behav. Process.* 87, 260–267. doi: 10.1016/j.beproc.2011.05.003

Candiotti, A., Zuberbühler, K., and Lemasson, A. (2012). Convergence and divergence in Diana monkey vocalizations. *Biol. Lett.* 8, 382–385. doi: 10.1098/rsbl.2011.1182

Carreiras, M., Lopez, J., Rivero, F., and Corina, D. (2005). Neural processing of a whistled language. *Nature* 433, 31–32. doi: 10.1038/433031a

Carter, G. G., Fenton, M. B., and Faure, P. A. (2009). White-winged vampire bats (*Diaemus youngi*) exchange contact calls. *Can. J. Zool.* 87, 604–608. doi: 10.1371/journal.pone.0038791

Carter, G. G., Skowronski, M. D., Faure, P. A., and Fenton, B. (2008). Antiphonal calling allows individual discrimination in white-winged vampire bats. *Anim. Behav.* 76, 1343–1355. doi: 10.1016/j.anbehav.2008.04.023

Charlton, B. D. (2014). Menstrual cycle phase alters women's sexual preferences for composers of more complex music. *Proc. Biol. Sci.* 281, 20140403. doi: 10.1098/rspb.2014.0403

Charlton, B. D., Filippi, P., and Fitch, W. T. (2012). Do women prefer more complex music around ovulation? *PLoS ONE* 7:e35626. doi: 10.1371/journal.pone.0035626

Cheng, M. F., and Durand, S. E. (2004). Song and the limbic brain: a new function for the bird's own song. *Ann. N. Y. Acad. Sci.* 1016, 611–627. doi: 10.1196/annals.1298.019

Cheng, Y., Lee, S.-Y., Chen, H.-Y., Wang, P.-Y., and Decety, J. (2012). Voice and emotion processing in the human neonatal brain. *J. Cogn. Neurosci.* 24, 1411–1419. doi: 10.1162/jocn_a_00214

Chow, C. P., Mitchell, J. F., and Miller, C. T. (2015). Vocal turn-taking in a non-human primate is learned during ontogeny. *Proc. R. Soc. Lond. B* 282, 20150069.

Christophe, A., Millotte, S., Bernal, S., and Lidz, J. (2008). Bootstrapping lexical and syntactic acquisition. *Lang. Speech* 51, 61–75. doi: 10.1177/0023830908085105010501

Cirelli, L. K., Einarson, K. M., and Trainor, L. J. (2014). Interpersonal synchrony increases prosocial behavior in infants. *Dev. Sci.* 17, 1003–1011. doi: 10.1111/desc.12193

Cooper, R. P., Abraham, J., Berman, S., and Staska, M. (1997). The development of infants' preference for motherese. *Infant Behav. Dev.* 20, 477–488. doi: 10.1016/S0163-6383(97)90037-0

Coss, R. G., McCowan, B., and Ramakrishnan, U. (2007). Threat-related acoustical differences in alarm calls by wild Bonnet Macaques (*Macaca radiata*) elicited by Python and Leopard models. *Ethology* 113, 352–367. doi: 10.1111/j.1439-0310.2007.01336.x

Counsilman, J. J. (1974). Waking and roosting behaviour of the Indian Myna. *Emu* 74, 135–148. doi: 10.1071/MU974135

Curtin, S., Mintz, T. H., and Christiansen, M. H. (2005). Stress changes the representational landscape: evidence from word segmentation. *Cognition* 96, 233–262. doi: 10.1016/j.cognition.2004.08.005

Cutler, A. (1994). Segmentation problems, rhythmic solutions. *Lingua* 92, 81–104. doi: 10.1016/0024-3841(94)90338-7

Cutler, A., and Chen, H. C. (1997). Lexical tone in Cantonese spoken-word processing. *Percept. Psychophys.* 59, 165–179. doi: 10.3758/BF03211886

Cutler, A., Oahan, D., and Van Donselaar, W. (1997). Prosody in the comprehension of spoken language: a literature review. *Lang. Speech* 40, 141–201.

Cynx, J. (1995). Similarities in absolute and relative pitch perception in songbirds (starling and zebra finch) and a nonsongbird (pigeon). *J. Comp. Psychol.* 109, 261–267. doi: 10.1037/0735-7036.109.3.261

Dahlin, C. R., and Benedict, L. (2014). Angry birds need not apply: a perspective on the flexible form and multifunctionality of avian vocal duets. *Ethology* 120, 1–10. doi: 10.1111/eth.12182

Darwin, C. (1871). *The Descent of Man, and Selection in Relation to Sex.* London: John Murray.

Darwin, C. (1872). *The Expression of the Emotions in Man and Animals.* London: John Murray.

de Boer, B. (2005). Evolution of speech and its acquisition. *Adapt. Behav.* 13, 281–292. doi: 10.1177/105971230501300405

de la Mora, D. M., Nespor, M., and Toro, J. M. (2013). Do humans and nonhuman animals share the grouping principles of the iambic–trochaic law? *Atten. Percept. Psychophys.* 75, 92–100. doi: 10.3758/s13414-012-0371-3

De la Torre, S., and Snowdon, C. T. (2002). Environmental correlates of vocal communication of wild pygmy marmosets, *Cebuella pygmaea. Anim. Behav.* 63, 847–856. doi: 10.1006/anbe.2001.1978

Dilger, W. C. (1953). Duetting in the crimson-breasted barbet. *Condor* 55, 220–221.

Doolittle, E., and Gingras, B. (2015). Zoomusicology. *Curr. Biol.* 25, R819–R820. doi: 10.1016/j.cub.2015.06.039

Doyle, L. R., McCowan, B., Hanser, S. F., Chyba, C., Bucci, T., and Blue, J. E. (2008). Applicability of information theory to the quantification of responses to anthropogenic noise by Southeast Alaskan Humpback Whales. *Entropy* 10, 33–46. doi: 10.3390/entropy-e10020033

Dudley, R., and Rand, A. S. (1991). Sound production and vocal sac inflation in the túngara frog, *Physalaemus pustulosus* (Leptodactylidae). *Copeia* 1991, 460–470. doi: 10.2307/1446594

Dudzinski, K. M. (1998). Contact behavior and signal exchange in Atlantic spotted dolphins. *Aquat. Mamm.* 24, 129–142.

Duellman, W. E. (1967). Social organization in the mating calls of some neotropical anurans. *Am. Midl. Nat.* 77, 156–163.

Elefant, C. (2002). *Enhancing Communication in Girls with Rett Syndrome through Songs in Music Therapy.* Ph.D. thesis, Aalborg University, Aalborg.

Endress, A. D., and Hauser, M. D. (2010). Word segmentation with universal prosodic cues. *Cogn. Psychol.* 61, 177–199. doi: 10.1016/j.cogpsych.2010.05.001

Evans, C. S., Evans, L., and Marler, P. (1993). On the meaning of alarm calls: functional reference in an avian vocal system. *Anim. Behav.* 46, 23–38. doi: 10.1006/anbe.1993.1158

Falk, D. (2004). Prelinguistic evolution in early hominins: whence motherese? *Behav. Brain Sci.* 27, 491–502. doi: 10.1017/S0140525X04000111

Fant, G. (1960). *Acoustic Theory of Speech Production.* The Hague: Mouton & Co.

Fedurek, P., Schel, A. M., and Slocombe, K. E. (2013). The acoustic structure of chimpanzee pant-hooting facilitates chorusing. *Behav. Ecol. Sociobiol.* 67, 1781–1789. doi: 10.1007/s00265-013-1585-7

Feher, O., Wang, H., Saar, S., Mitra, P. P., and Tchernichovski, O. (2009). De novo establishment of wild-type song culture in the zebra finch. *Nature* 459, 564–568. doi: 10.1038/nature07994

Fernald, A. (1989). Intonation and communicative intent in mothers' speech to infants: is the melody the message? *Child Dev.* 60, 1497–1510.

Fernald, A. (1992). "Meaningful melodies in mothers' speech to infants," in *Comparative and Developmental Approaches*, eds H. Papoušek, U. Jurgens, and M. Papoušek (Cambridge: Cambridge University Press), 262–282.

Fernald, A., and Kuhl, P. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behav. Dev.* 10, 279–293. doi: 10.1016/0163-6383(87)90017-8

Fernald, A., and Mazzie, C. (1991). Prosody and focus in speech to infants and adults. *Dev. Psychol.* 27, 209–221. doi: 10.1037/0012-1649.27.2.209

Fernald, A., and McRoberts, G. (1996). "Prosodic bootstrapping: a critical analysis of the argument and the evidence," in *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition*, eds J. L. Morgan and K. Demuth (Hillsdale, NJ: Erlbaum Associates), 365–388.

Fernald, A., and Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Dev. Psychol.* 20, 104–113. doi: 10.1037/0012-1649.20.1.104

Fernald, A., Taeschner, T., Dunn, J., Papoušek, M., de Boysson-Bardies, B., and Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *J. Child Lang.* 16, 477–501. doi: 10.1017/S0305000900010679

Fichtel, C., Hammerschmidt, K., and Jurgens, U. (2001). On the vocal expression of emotion. A multi-parametric analysis of different states of aversion in the squirrel monkey. *Behaviour* 138, 97–116. doi: 10.1163/15685390151067094

Ficken, M. S., and Witkin, S. R. (1977). Responses of black-capped chickadee flocks to predators. *Auk* 94, 156–157.

Filippi, P., Gingras, B., and Fitch, W. T. (2014). Pitch enhancement facilitates word learning across visual contexts. *Front. psychol.* 5:1468. doi: 10.3389/fpsyg.2014.01468

Filippi, P., Ocklenburg, S., Bowling, D. L., Heege, L., Güntürkün, O., Newen, A., et al. (2016). More than words (and faces): evidence for a Stroop effect of prosody in emotion word processing. *Cogn. Emot.* 1–13. doi: 10.1080/02699931.2016.1177489

Fisher, C., and Tokura, H. (1995). The given-new contract in speech to infants. *J. Mem. Lang.* 34, 287–310. doi: 10.1006/jmla.1995.1013

Fitch, W., and Hauser, M. D. (1995). Vocal production in nonhuman primates: acoustics, physiology, and functional constraints on "honest" advertisement. *Am. J. Primatol.* 37, 191–219. doi: 10.1002/ajp.1350370303

Fitch, W. T. (2005). The evolution of language: a comparative review. *Biol. Philos.* 20, 193–230. doi: 10.1007/s10539-005-5597-1

Fitch, W. T. (2006). The biology and evolution of music: a comparative perspective. *Cognition* 100, 173–215. doi: 10.1016/j.cognition.2005.11.009

Fitch, W. T. (2010). *The Evolution of Language.* Cambridge: Cambridge University Press.

Fitch, W. T. (2012). "The biology and evolution of rhythm: unravelling a paradox," in *Language and Music as Cognitive Systems*, eds P. Rebushat, M. Rohrmeier, J. A. Hawkins, and I. Cross (Oxford: Oxford University Press), 73–95.

Fitch, W. T. (2013). Rhythmic cognition in humans and animals: distinguishing meter and pulse perception. *Front. Syst. Neurosci.* 7:68. doi: 10.3389/fnsys.2013.00068

Fitch, W. T. (2015). Four principles of bio-musicology. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 370:20140091. doi: 10.1098/rstb.2014.0091

Fitch, W. T., and Hauser, M. D. (2003). "Unpacking "honesty": vertebrate vocal production and the evolution of acoustic signals," in *Acoustic Communication*, eds A. Simmons, R. R. Fay, and A. N. Popper (New York, NY: Springer), 65-137.

Foote, J. R., Fitzsimmons, L. P., Mennill, D. J., and Ratcliffe, L. M. (2008). Male chickadees match neighbors interactively at dawn: support for the social dynamics hypothesis. *Behav. Ecol.* 19, 1192–1199. doi: 10.1093/beheco/arn087

Frijda, N. H. (2016). The evolutionary emergence of what we call "emotions". *Cogn. Emot.* 30, 609–620. doi: 10.1080/02699931.2016.1145106

Fritz, T., Jentschke, S., Gosselin, N., Sammler, D., Peretz, I., Turner, R., et al. (2009). Universal recognition of three basic emotions in music. *Curr. Biol.* 19, 573–576. doi: 10.1016/j.cub.2009.02.058

Gaunt, A. S., and Nowicki, S. (1998). "Sound production in birds: acoustics and physiology revisited," in *Animal Acoustic Communication*, eds S. L. Hopp, M. J. Owren, and C. S. Evans (Berlin: Springer), 291–321.

Geissmann, T. (2000a). "Gibbon songs and human music from an evolutionary perspective," in *The Origins of Music*, eds N. L. Wallin, B. Merker, and S. Brown (Cambridge, MA: The MIT Press), 103–123.

Geissmann, T. (2000b). The relationship between duet songs and pair bonds in siamangs, *Hylobates* syndactylus. *Anim. Behav.* 60, 805–809. doi: 10.1006/anbe.2000.1540

Geissmann, T., and Orgeldinger, M. (2000). The relationship between duet songs and pair bonds in siamangs, *Hylobates* syndactylus. *Anim. Behav.* 60, 805–809. doi: 10.1006/anbe.2000.1540

Ghazanfar, A. A., Smith-Rohrberg, D., Pollen, A. A., and Hauser, M. D. (2002). Temporal cues in the antiphonal long-calling behaviour of cottontop tamarins. *Anim. Behav.* 64, 427–438. doi: 10.1006/anbe.2002.3074

Goldstein, M. H., King, A. P., and West, M. J. (2003). Social interaction shapes babbling: testing parallels between birdsong and speech. *Proc. Natl. Acad. Sci. U.S.A.* 100, 8030–8035. doi: 10.1073/pnas.1332441100

Goldstein, M. H., and Schwade, J. A. (2008). Social feedback to infants' babbling facilitates rapid phonological learning. *Psychol. Sci.* 19, 515–523.

Gould, S. J., and Eldredge, N. (1977). Punctuated equilibria: the tempo and mode of evolution reconsidered. *Paleobiology* 3, 115–151. doi: 10.1017/S0094837300005224

Grafe, T. U. (1999). A function of synchronous chorusing and a novel female preference shift in an anuran. *Proc. R. Soc. Lond. B Biol. Sci.* 266, 2331–2336. doi: 10.1098/rspb.1999.0927

Greenewalt, C. H. (1968), *Bird Song: Acoustics and Physiology*. Washington, DC: Smithsonian Institution Press.

Greenfield, M. D. (1994a). Cooperation and conflict in the evolution of signal interactions. *Annu. Rev. Ecol. Syst.* 25, 97–126. doi: 10.1146/annurev.es.25.110194.000525

Greenfield, M. D. (1994b). Synchronous and alternating choruses in insects and anurans: common mechanisms and diverse functions. *Am. Zool.* 34, 605–615. doi: 10.1093/icb/34.6.605

Greenfield, M. D. (2005). Mechanisms and evolution of communal sexual displays in arthropods and anurans. *Adv. Study Behav.* 35, 1–62. doi: 10.1016/S0065-3454(05)35001-7

Greenfield, M. D., and Roizen, I. (1993). Katydid synchronous chorusing is an evolutionarily stable outcome of female choice. *Nature* 364, 618–620. doi: 10.1038/364618a0

Grieser, D. L., and Kuhl, P. K. (1988). Maternal speech to infants in a tonal language: support for universal prosodic features in motherese. *Dev. Psychol.* 24:14. doi: 10.1037/0012-1649.24.1.14

Griffin, A. S. (2004). Social learning about predators: a review and prospectus. *Anim. Learn. Behav.* 32, 131–140. doi: 10.3758/BF03196014

Gros-Louis, J., West, M. J., Goldstein, M. H., and King, A. P. (2006). Mothers provide differential feedback to infants' prelinguistic sounds. *Int. J. Behav. Dev.* 30, 509–516. doi: 10.1177/0165025406071914

Gu, F., Zhang, C., Hu, A., and Zhao, G. (2013). Left hemisphere lateralization for lexical and acoustic pitch processing in Cantonese speakers as revealed by mismatch negativity. *Neuroimage* 83, 637–645. doi: 10.1016/j.neuroimage.2013.02.080

Güntürkün, O., Güntürkün, M., and Hahn, C. (2015). Whistled Turkish alters language asymmetries. *Curr. Biol.* 25, R706–R708. doi: 10.1016/j.cub.2015.06.067

Gussenhoven, C. (2002). "Intonation and biology," in *Liber Amicorum Bernard Bichakjian (Festschrift for Bernard Bichakjian)*, eds H. Jakobs and L. Wetzels (Maastricht: Shaker), 59–82.

Hagen, E. H., and Bryant, G. A. (2003). Music and dance as a coalition signaling system. *Hum. Nat.* 14, 21–51. doi: 10.1007/s12110-003-1015-z

Haimoff, E. H. (1981). Video analysis of siamang (*Hylobates* syndactylus) songs. *Behaviour* 76, 128–151. doi: 10.1163/156853981X00040

Hall, M. L. (2009). A review of vocal duetting in birds. *Adv. Study Behav.* 40, 67–121. doi: 10.1016/S0065-3454(09)40003-2

Hammerschmidt, K., Ansorge, V., Fischer, J., and Todt, D. (1994). Dusk calling in barbary macaques (*Macaca sylvanus*): demand for social shelter. *Am. J. Primatol.* 32, 277–289. doi: 10.1002/ajp.1350320405

Hartbauer, M., Haitzinger, L., Kainz, M., and Römer, H. (2014). Competition and cooperation in a synchronous bushcricket chorus. *R. Soc. Open Sci.* 1:140167. doi: 10.1098/rsos.140167

Hartbauer, M., and Römer, H. (2016). Rhythm generation and rhythm perception in insects: the evolution of synchronous choruses. *Front. Neurosci.* 10:223. doi: 10.3389/fnins.2016.00223

Hausberger, M., Henry, L., Testé, B., and Barbu, S. (2008). "Contextual sensitivity and bird songs: a basis for social life" in *Evolution of Communicative Flexibility*, eds O. Kimbrough, and U. Griebel (Cambridge, MA: The MIT Press), 121–138.

Hauser, M. D. (1992) A mechanism guiding conversational turn-taking in vervet monkeys and rhesus macaques. *Top. Primatol.* 1, 235–248.

Hauser, M. D., Chomsky, N., and Fitch, W. T. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science* 298, 1569–1579. doi: 10.1126/science.298.5598.1569

Hauser, M. D., and Marler, P. (1993). Food-associated calls in rhesus macaques (*Macaca mulatta*): II. Costs and benefits of call production and suppression. *Behav. Ecol.* 4, 206–212.

Hedberg, N., and Sosa, J. M. (2002). "The prosody of questions in natural discourse," in *Proceeding of Speech Prosody 2002*. Aix-en-Provence: Université de Provence, 275–278.

Henry, L., Craig, A. J., Lemasson, A., and Hausberger, M. (2015). Social coordination in animal vocal interactions. Is there any evidence of turn-taking? The starling as an animal model. *Front. Psychol.* 6:1416. doi: 10.3389/fpsyg.2015.01416

Hockett, C. (1960). The origin of speech. *Sci. Am.* 203, 88–111. doi: 10.1038/scientificamerican0960-88

Hoeschele, M., and Fitch, W. T. (2016). Phonological perception by birds: budgerigars can perceive lexical stress. *Anim. Cogn.* 19, 643–654. doi: 10.1007/s10071-016-0968-3

Hoeschele, M., Merchant, H., Kikuchi, Y., Hattori, Y., and ten Cate, C. (2015). Searching for the origins of musicality across species. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 370:20140094. doi: 10.1098/rstb.2014.0094

Hoeschele, M., Weisman, R. G., Guillette, L. M., Hahn, A. H., and Sturdy, C. B. (2013). Chickadees fail standardized operant tests for octave equivalence. *Anim. Cogn.* 16, 599–609. doi: 10.1007/s10071-013-0597-z

Hoese, W. J., Podos, J., Boetticher, N. C., and Nowicki, S. (2000). Vocal tract function in birdsong production: experimental manipulation of beak movements. *J. Exp. Biol.* 203, 1845–1855.

Holt, M. M., Noren, D. P., and Emmons, C. K. (2011). Effects of noise levels and call types on the source levels of killer whale calls. *J. Acoust. Soc. Am.* 130, 3100–3106. doi: 10.1121/1.3641446

Holt, M. M., Noren, D. P., Veirs, V., Emmons, C. K., and Veirs, S. (2009). Speaking up: killer whales (*Orcinus orca*) increase their call amplitude in response to vessel noise. *J. Acoust. Soc. Am.* 125, EL27–EL32. doi: 10.1121/1.3040028

Honbolygo, F., Török, Á., Bánréti, Z., Hunyadi, L., and Csépe, V. (2016). ERP correlates of prosody and syntax interaction in case of embedded sentences. *J. Neurol.* 37, 22–33.

Honing, H., ten Cate, C., Peretz, I., and Trehub, S. E. (2015). Without it no music: cognition, biology and evolution of musicality. *Philos. Trans. R. Soc. Lond B Biol. Sci.* 370: 20140088. doi: 10.1098/rstb.2014.0088

Hotchkin, C., and Parks, S. (2013). The Lombard effect and other noise-induced vocal modifications: insight from mammalian communication systems. *Biol. Rev.* 88, 809–824. doi: 10.1111/brv.12026

Hulse, S. H., Bernard, D. J., and Braaten, R. F. (1995). Auditory discrimination of chord-based spectral structures by European starlings (*Sturnus vulgaris*). *J. Exp. Psychol. Gen.* 124, 409–423.

Ishii, K., Reyes, J. A., and Kitayama, S. (2003). Spontaneous attention to word content versus emotional tone differences among three cultures. *Psychol. Sci.* 14, 39–46. doi: 10.1111/1467-9280.01416

Izumi, A. (2000). Japanese monkeys perceive sensory consonance of chords. *J. Acoust. Soc. Am.* 108, 3073–3078. doi: 10.1121/1.1323461

Jackendoff, R. (2009). Parallels and Nonparallels between Language and Music. *Music Percept.* 26, 195–204. doi: 10.1525/mp.2009.26.3.195

Jaffe, J., Beebe, B., Feldstein, S., Crown, C. L., Jasnow, M. D., Rochat, P., et al. (2001). Rhythms of dialogue in infancy: coordinated timing in development. *Monogr. Soc. Res. child Dev.* 66, i–viii, 1–132.

Janik, V. M., and Slater, P. J. B. (1997). Vocal learning in mammals. *Adv. Study Behav.* 26, 59–99. doi: 10.1016/S0065-3454(08)60377-0

Jespersen, O. (1922). *Language: Its Nature, Developement and Origin*. London: Allen and Unwin.

Johnson, E. K. (2008). Infants use prosodically conditioned acoustic-phonetic cues to extract words from speech. *J. Acoust. Soc. Am.* 123, EL144–EL148. doi: 10.1121/1.2908407

Johnson, E. K., and Jusczyk, P. W. (2001). Word segmentation by 8-month-olds: when speech cues count more than statistics. *J. Mem. Lang.* 44, 548–567. doi: 10.1006/jmla.2000.2755

Jovanovic, T., and Gouzoules, H. (2001). Effects of nonmaternal restraint on the vocalizations of infant rhesus monkeys (*Macaca mulatta*). *Am. J. Primatol.* 53, 33–45.

Jusczyk, P. W. (1999). How infants begin to extract words from speech. *Trends Cogn. Sci.* 3, 323–328.

Jusczyk, P. W., and Aslin, R. N. (1995). Infants detection of the sound patterns of words in fluent speech. *Cogn. Psychol.* 29, 1–23. doi: 10.1006/cogp.1995.1010

Juslin, P. N., and Laukka, P. (2003). Communication of emotions in vocal expression and music performance: different channels, same code? *Psychol. Bull.* 129, 770–814. doi: 10.1037/0033-2909.129.5.770

Keitel, A., Prinz, W., Friederici, A. D., von Hofsten, C., and Daum, M. M. (2013). Perception of conversations: the importance of semantics and intonation in children's development. *J. Exp. Child Psychol.* 116, 264–277. doi: 10.1016/j.jecp.2013.06.005

Kirby, S., Cornish, H., and Smith, K. (2008). Cumulative cultural evolution in the laboratory: an experimental approach to the origins of structure in human language. *Proc. Natl. Acad. Sci. U.S.A.* 105, 10681–10686. doi: 10.1073/pnas.0707835105

Kirschner, S., and Tomasello, M. (2010). Joint music making promotes prosocial behavior in 4-year-old children. *Evol. Hum. Behav.* 31, 354–364. doi: 10.1016/j.evolhumbehav.2010.04.004

Kitagawa, Y. (2005). Prosody, syntax and pragmatics of wh-questions in Japanese. *Engl. Linguist.* 22, 302–346. doi: 10.9793/elsj1984.22.302

Kitamura, C., Thanavishuth, C., Burnham, D., and Luksaneeyanawin, S. (2001). Universality and specificity in infant-directed speech: pitch modifications as a function of infant age and sex in a tonal and non-tonal language. *Infant Behav. Dev.* 24, 372–392. doi: 10.1016/S0163-6383(02)00086-3

Klump, G. M., and Gerhardt, H. C. (1992). "Mechanisms and function of call-timing in male–male interactions in frogs," in *Playback and Studies of Animal Communication*, ed. P. K. McGregor (New York, NY: Plenum Press), 153–174.

Koda, H., Lemasson, A., Oyakawa, C., Pamungkas, J., and Masataka, N. (2013). Possible role of mother-daughter vocal interactions on the development of species-specific song in gibbons. *PLoS ONE* 8:e71432. doi: 10.1371/journal.pone.0071432

Koelsch, S. (2012). *Brain and Music*. Hoboken, NY: John Wiley & Sons.

Koelsch, S. (2013). From social contact to social cohesion—the 7 Cs. *Music Med.* 5, 204–209. doi: 10.1177/1943862113508588

Kremers, D., Briseño-Jaramillo, M., Böye, M., Lemasson, A., and Hausberger, M. (2014). Nocturnal vocal activity in captive bottlenose dolphins (*Tursiops truncatus*): could dolphins have presleep choruses. *Anim. Behav. Cogn.* 1, 464–469.

Kriengwatana, B., Escudero, P., and ten Cate, C. (2015). Revisiting vocal perception in non-human animals: a review of vowel discrimination, speaker

voice recognition, and speaker normalization. *Front. Psychol.* 5:1543. doi: 10.3389/fpsyg.2014.01543

Kroodsma, D. E., and Byers, B. E. (1991). The function(s) of bird song. *Am. Zool.* 31, 318–328. doi: 10.1093/icb/31.2.318

Kuhl, P. K. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science* 277, 684–686. doi: 10.1126/science.277.5326.684

Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nat. Rev. Neurosci.* 5, 831–843.

Kuhl, P. K., Tsao, F. M., and Liu, H. M. (2003). Foreign-language experience in infancy: effects of short-term exposure and social interaction on phonetic learning. *Proc. Natl. Acad. Sci. U.S.A.* 100, 9096–9101. doi: 10.1073/pnas.1532872100

Lammertink, I., Casillas, M., Benders, T., Post, B., and Fikkert, P. (2015). Dutch and english toddlers' use of linguistic cues in predicting upcoming turn transitions. *Front. Psychol.* 6:495. doi: 10.3389/fpsyg.2015.00495

Langmore, N. E. (1998). Functions of duet and solo songs of female birds. *Trends Ecol. Evol.* 13, 136–140. doi: 10.1016/S0169-5347(97)01241-X

Langus, A., Marchetto, E., Bion, R. A. H., and Nespor, M. (2012). Can prosody be used to discover hierarchical structure in continuous speech? *J. Mem. Lang.* 66, 285–306. doi: 10.1016/j.jml.2011.09.004

Launay, J., Dean, R. T., and Bailes, F. (2014). Synchronising movements with the sounds of a virtual partner enhances partner likeability. *Cogn. Process.* 15, 491–501. doi: 10.1007/s10339-014-0618-0

Lee, C. Y. (2000). Lexical tone in spoken word recognition: a view from Mandarin Chinese. *J. Acoust. Soc. Am.* 108, 2480–2480. doi: 10.1121/1.4715150

Lehiste, I. (1970). *Suprasegmentals*. Cambridge, MA: MIT Press.

Lemasson, A., Gandon, E., and Hausberger, M. (2010). Attention to elders' voice in non-human primates. *Biol. Lett.* 6:328. doi: 10.1098/rsbl.2009.0875

Lemasson, A., Glas, L., Barbu, S., Lacroix, A., Guilloux, M., Remeuf, K., et al. (2011). Youngsters do not pay attention to conversational rules: is this so for nonhuman primates? *Sci. Rep.* 1:22. doi: 10.1038/srep00022

Lemasson, A., Guilloux, M., Barbu, S., Lacroix, A., and Koda, H. (2013). Age-and sex-dependent contact call usage in Japanese macaques. *Primates* 54, 283–291. doi: 10.1007/s10329-013-0347-5

Levänen, S., Uutela, K., Salenius, S., and Hari, R. (2001). Cortical representation of sign language: comparison of deaf signers and hearing non-signers. *Cereb. Cortex* 11, 506–512. doi: 10.1093/cercor/11.6.506

Levinson, S. C. (2016). Turn-taking in human communication–origins and implications for language processing. *Trends Cogn. Sci.* 20, 6–14. doi: 10.1016/j.tics.2015.10.010

Lieberman, P. (1967). *Intonation, perception, and language*. Cambridge, MA: MIT Press.

Liu, F., Patel, A. D., Fourcin, A., and Stewart, L. (2010). Intonation processing in congenital amusia: discrimination, identification and imitation. *Brain* 133, 1682–1693. doi: 10.1093/brain/awq089

Livingstone, F. B. (1973). Did the Australopithecines sing? *Curr. Anthropol.* 14, 25–29. doi: 10.1086/201402

Locke, J. L. (1995). *The Child's Path to Spoken Language*. Cambridge, MA: Harvard University Press.

Männel, C., Schipke, C. S., and Friederici, A. D. (2013). The role of pause as a prosodic boundary marker: language ERP studies in German 3- and 6-year-olds. *Dev. Cogn. Neurosci.* 5, 86–94. doi: 10.1016/j.dcn.2013.01.003

Manson, J. H., Bryant, G. A., Gervais, M. M., and Kline, M. A. (2013). Convergence of speech rate in conversation predicts cooperation. *Evol. Hum. Behav.* 34, 419–426. doi: 10.1016/j.evolhumbehav.2013.08.001

Marler, P. (2000). "Origins of music and speech: insights from animals," in *The Origins of Music*, eds N. L. Wallin, B. Merker, and S. Brown (Cambridge, MA: The MIT Press), 31–48.

Marler, P., and Evans, C. (1996). Bird calls: just emotional displays or something more? *Ibis* 138, 26–33. doi: 10.1111/j.1474-919X.1996.tb04765.x

Marsolek, C. J., and Deason, R. G. (2007). Hemispheric asymmetries in visual word-form processing: progress, conflict, and evaluating theories. *Brain Lang.* 103, 304–307. doi: 10.1016/j.bandl.2007.02.009

Martinet, A. (1980). *Eléments de Linguistique Générale*. Paris: Armand Collin.

Masataka, N., and Biben, M. (1987). Temporal rules regulating affiliative vocal exchanges of squirrel monkeys. *Behaviour* 101, 311–319. doi: 10.1163/156853987X00035

Matsumura, S. (1981). Mother-infant communication in a horseshoe bat (Rhinolophus ferrumequinum nippon): vocal communication in three-week-old Infants. *J. Mammal.* 62, 20–28. doi: 10.2307/1380474

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., and Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition* 29, 143–178. doi: 10.1016/0010-0277(88)90035-2

Méndez-Cárdenas, M. G., and Zimmermann, E. (2009). Duetting—A mechanism to strengthen pair bonds in a dispersed pair-living primate (*Lepilemur edwardsi*)? *Am. J. Phys. Anthropol.* 139, 523–532. doi: 10.1002/ajpa.21017

Merill, J., Sammler, D., Bangert, M., Goldhahn, D., Turner, R., and Friederici, A. D. (2012). Perception of words and pitch patterns in song and speech. *Front. Psychol.* 3:76. doi: 10.3389/fpsyg.2012.00076

Merker, B. (2000). Synchronous chorusing and the origins of music. *Musi. Sci.* 3, 59–73. doi: 10.1177/10298649000030S105

Merker, B. H., Madison, G. S., and Eckerdal, P. (2009). On the role and origin of isochrony in human rhythmic entrainment. *Cortex* 45, 4–17. doi: 10.1016/j.cortex.2008.06.011

Meyer, J. (2004). Bioacoustics of human whistled languages: an alternative approach to the cognitive processes of language. *An. Acad. Bras. Ciênc.* 76, 406–412. doi: 10.1590/S0001-37652004000200033

Miller, C. T., Beck, K., Meade, B., and Wang, X. (2009). Antiphonal call timing in marmosets is behaviorally significant: interactive playback experiments. *J. Comp. Physiol. A* 195, 783–789. doi: 10.1007/s00359-009-0456-1

Miller, P. J. O., Shapiro, A. D., Tyack, P. L., and Solow, A. R. (2004). Call-type matching in vocal exchanges of free-ranging resident killer whales, *Orcinus orca. Anim. Behav.* 67, 1099–1107. doi: 10.1016/j.anbehav.2003.06.017

Mithen, S. (2005). *The Singing Neanderthals: The Origins of Music, Language, Mind, and Body*. Cambridge, MA: Harvard University Press.

Morley, I. (2013). A multi-disciplinary approach to the origins of music: perspectives from anthropology, archaeology, cognition and behaviour. *J. Anthropol. Sci.* 92, 147–177. doi: 10.4436/JASS.92008

Morton, E. S. (1977). On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *Am. Nat.* 111, 855–869. doi: 10.1016/j.beproc.2009.04.008

Müller, A. E., and Anzenberger, G. (2002). Duetting in the titi monkey *Callicebus cupreus*: structure, pair specificity and development of duets. *Folia Primatol.* 73, 104–115. doi: 10.1159/000064788

Müller, J. L., Bahlmann, J., and Friederici, A. D. (2010). Learnability of embedded syntactic structures depends on prosodic cues. *Cogn. Sci.* 34, 338–349. doi: 10.1111/j.1551-6709.2009.01093.x

Müller, P., and Warwick, A. (1993). "Autistic children and music therapy. The effects of maternal involvement in therapy," in *Music Therapy in Health and Education*, eds M. Heal and T. Wigram (London: Jessica Kingsley), 214–243.

Naguib, M., and Mennill, D. J. (2010). The signal value of birdsong: empirical evidence suggests song overlapping is a signal. *Anim. Behav.* 80, e11–e15. doi: 10.1016/j.anbehav.2010.06.001

Nakamura, C., Arai, M., and Mazuka, R. (2012). Immediate use of prosody and context in predicting a syntactic structure. *Cognition* 125, 317–323. doi: 10.1016/j.cognition.2012.07.016

Naoi, N., Watanabe, S., Maekawa, K., and Hibiya, J. (2012). Prosody discrimination by songbirds (*Padda oryzivora*). *PLoS ONE* 7:e47446. doi: 10.1371/journal.pone.0047446

Narins, P. M., Reichman, O. J., Jarvis, J. U., and Lewis, E. R. (1992). Seismic signal transmission between burrows of the Cape mole-rat, *Georychus capensis. J. Comp. Physiol. A* 170, 13–21.

Nazzi, T., Bertoncini, J., and Mehler, J. (1998). Language discrimination by newborns: toward an understanding of the role of rhythm. *J. Exp. Psychol. Hum. Percept. Perform.* 24:756–766.

Nesse, R. M. (1990). Evolutionary explanations of emotions. *Hum. Nat.* 1, 261–289. doi: 10.1007/BF02733986

Neubauer, R. (1999). Super-normal length song preferences of female zebra finches (*Taeniopygia guttata*) and a theory of the evolution of bird song. *Evol. Ecol.* 13, 365–380. doi: 10.1023/A:1006708826432

Newen, A., Welpinghus, A., and Juckel, G. (2015). Emotion recognition as pattern recognition: the relevance of perception. *Mind Lang.* 30, 187–208. doi: 10.1111/mila.12077

Noble, J. (1999). Cooperation, conflict and the evolution of communication. *Adapt. Behav.* 7, 349–369. doi: 10.1177/105971239900700308

Nonaka, S., Takahashi, R., Enomoto, K., Katada, A., and Unno, T. (1997). Lombard reflex during PAG-induced vocalization in decerebrate cats. *Neurosci. Res.* 29, 283–289. doi: 10.1016/S0168-0102(97)00097-7

Notman, H., and Rendall, D. (2005). Contextual variation in chimpanzee pant hoots and its implications for referential communication. *Anim. Behav.* 70, 177–190.

Nowicki, S. (1987). Vocal tract resonances in oscine bird sound production: evidence from birdsongs in a helium atmosphere. *Nature* 325, 53–55. doi: 10.1038/325053a0

Okanoya, K. (2004). Song syntax in Bengalese finches: proximate and ultimate analyses. *Adv. Study Behav.* 34, 297–346. doi: 10.1016/S0065-3454(04)34008-8

Oldfield, A., Adams, M., and Bunce, L. (2003). An investigation into short-term music therapy with mothers and young children. *Br. J. Music Ther.* 17, 26–45. doi: 10.1177/135945750301700105

Oller, D. K. (1973). The effect of position in utterance on speech segment duration in English. *J. Acoust. Soc. Am.* 54, 1235–1246. doi: 10.1121/1.1914393

Owren, M. J., and Rendall, D. (1997). "An affect-conditioning model of nonhuman primate vocal signaling," in *Perspectives in Ethology, Communication*, Vol. 12, eds D. W. Owings, M. D. Beecher, and N. S. Thompson (New York, NY: Plenum Press), 299–346.

Owren, M. J., and Rendall, D. (2001). Sound on the rebound: bringing form and function back to the forefront in understanding nonhuman primate vocal signaling. *Evol. Anthropol.* 10, 58–71. doi: 10.1002/evan.1014.abs

Papoušek, M., Bornstein, M. H., Nuzzo, C., Papoušek, H., and Symmes, D. (1990). Infant responses to prototypical melodic contours in parental speech. *Infant Behav. Dev.* 13, 539–545. doi: 10.1016/0163-6383(90)90022-Z

Parks, S. E., Clark, C. W., and Tyack, P. L. (2007). Short- and long-term changes in right whale calling behavior: The potential effects of noise on acoustic communication. *J. Acoust. Soc. Am.* 122, 3725–3731. doi: 10.1121/1.2799904

Parks, S. E., Johnson, M., Nowacek, D., and Tyack, P. L. (2011). Individual right whales call louder in increased environmental noise. *Biol. Lett.* 7, 33–35. doi: 10.1098/rsbl.2010.0451

Patel, A. D. (2003). Language, music, syntax and the brain. *Nat. Neurosci.* 6, 674–681. doi: 10.1038/nn1082

Patel, A. D. (2006). Musical rhythm, linguistic rhythm, and human evolution. *Music Percept.* 24, 99–104. doi: 10.1525/mp.2006.24.1.99

Patel, A. D. (2010). *Music, Language, and the Brain*. Oxford: Oxford University Press.

Patel, A. D., Wong, M., Foxton, J., Lochy, A., and Peretz, I. (2008). Speech intonation perception deficits in musical tone deafness (congenital amusia). *Music Percept.* 25, 357–368. doi: 10.1525/mp.2008.25.4.357

Pell, M. D. (2005). Nonverbal emotion priming: evidence from the facial affect decision task. *J. Nonverbal Behav.* 29, 45–73. doi: 10.1007/s10919-004-0889-8

Pell, M. D., Jaywant, A., Monetta, L., and Kotz, S. A. (2011). Emotional speech processing: disentangling the effects of prosody and semantic cues. *Cogn. Emot.* 25, 834–853. doi: 10.1080/02699931.2010.516915

Peretz, I., and Hyde, K. L. (2003). What is specific to music processing? Insights from congenital amusia. *Trends Cogn. Sci.* 7, 362–367. doi: 10.1016/S1364-6613(03)00150-5

Pettitt, B. A., Bourne, G. R., and Bee, M. A. (2012). Quantitative acoustic analysis of the vocal repertoire of the golden rocket frog (*Anomaloglossus beebei*). *J. Acoust. Soc. Am.* 131, 4811–4820. doi: 10.1121/1.4714769

Phillips-Silver, J., Aktipis, C. A., and Bryant, G. A. (2010). The ecology of entrainment: Foundations of coordinated rhythmic movement. *Music Percept.* 28, 3–14. doi: 10.1525/mp.2010.28.1.3

Pisanski, K., Cartei, V., McGettigan, C., Raine, J., and Reby, D. (2016). Voice modulation: a window into the origins of human vocal control? *Trends Cogn. Sci.* 20, 304–318. doi: 10.1016/j.tics.2016.01.002

Poirier, C., Henry, L., Mathelier, M., Lumineau, S., Cousillas, H., and Hausberger, M. (2004). Direct social contacts override auditory information in the song-learning process in starlings (*Sturnus vulgaris*). *J. Comp. Psychol.* 118, 179–193. doi: 10.1037/0735-7036.118.2.179

Prestwich, K. N. (1994). The energetics of acoustic signaling in anurans and insects. *Am. Zool.* 34, 625–643. doi: 10.1093/icb/34.6.625

Rabin, L. A., McCowan, B., Hooper, S. L., and Owings, D. H. (2003). Anthropogenic noise and its effect on animal communication: an interface between comparative psychology and conservation biology. *Int. J. Comp. Psychol.* 16, 172–192.

Ralston, J. V., and Herman, L. M. (1995). Perception and generalization of frequency contours by a bottlenose dolphin (*Tursiops truncatus*). *J. Comp. Psychol.* 109, 268–277.

Ramus, F., Hauser, M. D., Miller, C., Morris, D., and Mehler, J. (2000). Language discrimination by human newborns and by cotton-top tamarin monkeys. *Science* 288, 349–351. doi: 10.1126/science.288.5464.349

Rasilo, H., Räsänen, O., and Laine, U. K. (2013). Feedback and imitation by a caregiver guides a virtual infant to learn native phonemes and the skill of speech inversion. *Speech Commun.* 55, 909–931. doi: 10.1016/j.specom.2013.05.002

Ravignani, A. (2014). Chronometry for the chorusing herd: hamilton's legacy on context-dependent acoustic signalling – a comment on Herbers (2013). *Biol. Lett.* 10:20131018.

Ravignani, A. (2015). Evolving perceptual biases for antisynchrony: a form of temporal coordination beyond synchrony. *Front. Neurosci.* 9:339. doi: 10.3389/fnins.2015.00339

Ravignani, A., Bowling, D. L., and Fitch, W. (2014a). Chorusing, synchrony, and the evolutionary functions of rhythm. *Front. Psychol.* 5:1118. doi: 10.3389/fpsyg.2014.01118

Ravignani, A., Martins, M., and Fitch, W. T. (2014b). Vocal learning, prosody, and basal ganglia: don't underestimate their complexity. *Behav. Brain Sci.* 37, 570–571. doi: 10.1017/S0140525X13004184

Reichert, M. S. (2013). Patterns of variability are consistent across signal types in the treefrog *Dendropsophus ebraccatus*. *Biol. J. Linn. Soc.* 109, 131–145. doi: 10.1111/bij.12028

Remez, R. E., Rubin, P. E., Pisoni, D. B., and Carrell, T. D. (1981). Speech perception without traditional speech cues. *Science* 212, 947–949. doi: 10.1126/science.7233191

Rendall, D. (2003). Acoustic correlates of caller identity and affect intensity in the vowel-like grunt vocalizations of baboons. *J. Acoust. Soc. Am.* 113, 3390–3402. doi: 10.1121/1.1568942

Rendall, D., Owren, M. J., and Ryan, M. J. (2009). What do animal signals mean? *Anim. Behav.* 78, 233–240. doi: 10.1016/j.anbehav.2009.06.007

Rendall, D., Seyfarth, R. M., Cheney, D. L., and Owren, M. J. (1999). The meaning and function of grunt variants in baboons. *Anim. Behav.* 57, 583–592. doi: 10.1006/anbe.1998.1031

Rezával, C., Pattnaik, S., Pavlou, H. J., Nojima, T., Brüggemeier, B., D'Souza, L. A., et al. (2016). Activation of latent courtship circuitry in the brain of *Drosophila* females induces male-like behaviors. *Curr. Biol.* doi: 10.1016/j.cub.2016.07.021 [Epub ahead of print].

Rialland, A. (2007). Question prosody: an African perspective. *Tones Tunes* 1, 35–62.

Richman, B. (1993). On the evolution of speech: singing as the middle term. *Curr. Anthropol.* 34:721. doi: 10.1086/204217

Richman, B. (2000). "How music fixed "nonsense" into significant formulas: on rhythm, repetition, and meaning," in *The Origins of Music*, eds N. L. Wallin, B. Merker, and S. Brown (Cambridge, MA: The MIT Press), 301–314.

Riede, T., Arcadi, A. C., and Owren, M. J. (2007). Nonlinear acoustics in the pant hoots of common chimpanzees (Pan troglodytes): vocalizing at the edge. *J. Acoust. Soc. Am.* 121, 1758–1767. doi: 10.1121/1.2427115

Riede, T., Suthers, R. A., Fletcher, N. H., and Blevins, W. E. (2006). Songbirds tune their vocal tract to the fundamental frequency of their song. *Proc. Natl. Acad. Sci. U.S.A.* 103, 5543–5548. doi: 10.1073/pnas.0601262103

Roberts, S. G., Torreira, F., and Levinson, S. C. (2015). The effects of processing and sequence organization on the timing of turn taking: a corpus study. *Front. Psychol.* 6:509. doi: 10.3389/fpsyg.2015.00509

Rohrmeier, M., Zuidema, W., Wiggins, G. A., and Scharff, C. (2015). Principles of structure building in music, language and animal song. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 370, 20140097. doi: 10.1098/rstb.2014.0097

Rousseau, J. J. (1781). *Essay on the Origin of Languages and Writings Related to Music*. Hanover: University Press of New England.

Ryan, M. J., Tuttle, M. D., and Taft, L. K. (1981). The costs and benefits of frog chorusing behavior. *Behav. Ecol. Sociobiol.* 8, 273–278. doi: 10.1007/BF00299526

Sacks, H., Schegloff, E. A., and Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language* 50, 696–735. doi: 10.2307/412243

Sander, D., Grandjean, D., Pourtois, G., Schwartz, S., Seghier, M. L., Scherer, K. R., et al. (2005). Emotion and attention interactions in social cognition: brain regions involved in processing anger prosody. *Neuroimage* 28, 848–858. doi: 10.1016/j.neuroimage.2005.06.023

Schehka, S., Esser, K.-H., and Zimmermann, E. (2007). Acoustical expression of arousal in conflict situations in tree shrews (*Tupaia belangeri*). *J. Comp. Physiol. A Sens. Neural Behav. Physiol.* 193, 845–852. doi: 10.1007/s00359-007-0236-8

Scherer, K. R. (1995). Expression of emotion in voice and music. *J. Voice* 9, 235–248. doi: 10.1016/S0892-1997(05)80231-0

Scherer, K. R. (2003). Vocal communication of emotion: a review of research paradigms. *Speech commun.* 40, 227–256. doi: 10.1016/S0167-6393(02)00084-5

Schirmer, A., and Kotz, S. A. (2003). ERP evidence for a sex-specific Stroop effect in emotional speech. *J. Cogn. Neurosci.* 15, 1135–1148. doi: 10.1162/089892903322598102

Schmidt, U., and Joermann, G. (1986). The influence of acoustical interferences on echolocation in bats. *Mammalia* 50, 379–390. doi: 10.1515/mamm.1986.50.3.379

Schneider, B. A., Trehub, S. E., and Bull, D. (1979). The development of basic auditory processes in infants. *Can. J. Psychol.* 33, 306–319. doi: 10.1037/h0081728

Schore, J. R., and Schore, A. N. (2008). Modern attachment theory: the central role of affect regulation in development and treatment. *Clin. Soc. Work J.* 36, 9–20. doi: 10.1007/s10615-007-0111-7

Schulz, T. M., Whitehead, H., Gero, S., and Rendell, L. (2008). Overlapping and matching of codas in vocal interactions between sperm whales: insights into communication function. *Anim. Behav.* 76, 1–12. doi: 10.1016/j.anbehav.2008.07.032

Searcy, W. A., and Andersson, M. (1986). Sexual selection and the evolution of song. *Annu. Rev. Ecol. Syst.* 17, 507–533. doi: 10.1146/annurev.es.17.110186.002451

Seyfarth, R. M., and Cheney, D. L. (2003). Meaning and emotion in animal vocalizations. *Ann. N. Y. Acad. Sci.* 1000, 32–55. doi: 10.1196/annals.1280.004

Seyfarth, R. M., Cheney, D. L., and Marler, P. (1980). Monkey responses to three different alarm calls: evidence of predator classification and semantic communication. *Science* 210, 801–803. doi: 10.1126/science.7433999

Sherrod, K. B., Friedman, S., Crawley, S., Drake, D., and Devieux, J. (1977). Maternal language to prelinguistic infants: syntactic aspects. *Child Dev.* 48, 1662–1665. doi: 10.2307/1128531

Shukla, M., White, K. S., and Aslin, R. N. (2011). Prosody guides the rapid mapping of auditory word forms onto visual objects in 6-mo-old infants. *Proc. Natl. Acad. Sci. U.S.A.* 108, 6038–6043. doi: 10.1073/pnas.1017617108

Sininger, Y. S., Abdala, C., and Cone-Wesson, B. (1997). Auditory threshold sensitivity of the human neonate as measured by the auditory brainstem response. *Hear. Res.* 104, 27–38. doi: 10.1016/S0378-5955(96)00178-5

Sismondo, E. (1990). Synchronous, alternating, and phase-locked stridulation by a tropical katydid. *Science* 249, 55–58. doi: 10.1126/science.249.4964.55

Smith, E. A. (2010). Communication and collective action: language and the evolution of human cooperation. *Evol. Hum. Behav.* 31, 231–245. doi: 10.1016/j.evolhumbehav.2010.03.001

Snedeker, J., and Trueswell, J. (2003). Using prosody to avoid ambiguity: effects of speaker awareness and referential context. *J. Mem. Lang.* 48, 103–130. doi: 10.1016/S0749-596X(02)00519-3

Snowdon, C. T., and Cleveland, J. (1984). 'Conversations' among pygmy marmosets. *Am. J. Primatol.* 7, 15–20. doi: 10.1002/ajp.1350070104

Soderstrom, M., Seidl, A., Kemler Nelson, D. G., and Jusczyk, P. W. (2003). The prosodic bootstrapping of phrases: evidence from prelinguistic infants. *J. Mem. Lang.* 49, 249–267. doi: 10.1016/S0749-596X(03)00024-X

Soltis, J., Leong, K., and Savage, A. (2005a). African elephant vocal communication I: antiphonal calling behaviour among affiliated females. *Anim. Behav.* 70, 579–587. doi: 10.1016/j.anbehav.2004.11.015

Soltis, J., Leong, K., and Stoeger, A. (2005b). African elephant vocal communication II: rumble variation reflects the individual identity and emotional state of caller. *Anim. Behav.* 70, 589–599. doi: 10.1016/j.anbehav.2004.11.016

Spierings, M. J., and ten Cate, C. (2014). Zebra finches are sensitive to prosodic features of human speech. *Proc. R. Soc. Lond. B* 281, 20140480. doi: 10.1098/rspb.2014.0480

Steedman, M. (1996). "Phrasal intonation and the acquisition of syntax," in *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition*, eds J. L. Morgan and K. Demuth (Hillsdale, NJ: Erlbaum), 331–342.

Stephens, J., and Beattie, G. (1986). On judging the ends of speaker turns in conversation. *J. Lang. Soc. Psychol.* 5, 119–134. doi: 10.1177/0261927X8652003

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., et al. (2009). Universals and cultural variation in turn-taking in conversation. *Proc. Natl. Acad. Sci. U.S.A.* 106, 10587–10592. doi: 10.1073/pnas.0903616106

Stoeger, A. S., Baotic, A., Li, D., and Charlton, B. D. (2012). Acoustic features indicate arousal in infant Giant Panda vocalisations. *Ethology* 118, 896–905. doi: 10.1111/j.1439-0310.2012.02080.x

Stoeger, A. S., Charlton, B. D., Kratochvil, H., and Fitch, W. T. (2011). Vocal cues indicate level of arousal in infant African elephant roars. *J. Acoust. Soc. Am.* 130, 1700–1710. doi: 10.1121/1.3605538

Sugimoto, T., Kobayashi, H., Nobuyoshi, N., Kiriyama, Y., Takeshita, H., Nakamura, T., et al. (2009). Preference for consonant music over dissonant music by an infant chimpanzee. *Primates* 51, 7–12. doi: 10.1007/s10329-009-0160-3

Sugiura, H. (1993). Temporal and acoustic correlates in vocal exchange of coo calls in Japanese macaques. *Behaviour* 124, 207–225. doi: 10.1163/156853993X00588

Syal, S., and Finlay, B. L. (2011). Thinking outside the cortex: social motivation in the evolution and development of language. *Dev. Sci.* 14, 417–430. doi: 10.1111/j.1467-7687.2010.00997.x

Symmes, D., and Biben, M. (1988). "Conversational vocal exchanges in squirrel monkeys," in *Primate Vocal Communication*, eds D. Todt, P. Goedeking, and D. Symmes (Berlin: Springer Verlag), 123–132.

Szipl, G., Boeckle, M., Werner, S. A. B., and Kotrschal, K. (2014). Mate recognition and expression of affective state in croop calls of Northern Bald Ibis (*Geronticus eremita*). *PLoS ONE* 9:e88265. doi: 10.1371/journal.pone.0088265

Takahashi, D. Y., Narayanan, D. Z., and Ghazanfar, A. A. (2013). Coupled oscillator dynamics of vocal turn-taking in monkeys. *Curr. Biol.* 23, 2162–2168. doi: 10.1016/j.cub.2013.09.005

Tarr, B., Launay, J., and Dunbar, R. I. (2014). Music and social bonding: "self-other" merging and neurohormonal mechanisms. *Front. Psychol.* 5:1096. doi: 10.3389/fpsyg.2014.01096

Templeton, C. N., Greene, E., and Davis, K. (2005). Allometry of alarm calls: black-capped chickadees encode information about predator size. *Science* 308, 1934–1937. doi: 10.1126/science.1108841

Ten Bosch, L., Oostdijk, N., and Boves, L. (2005). On temporal aspects of turn taking in conversational dialogues. *Speech Commun.* 47, 80–86. doi: 10.1016/j.specom.2005.05.009

Thiessen, E. D., Hill, E. A., and Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy* 7, 53–71. doi: 10.1207/s15327078in0701_5

Tinbergen, N. (1963). On aims and methods of ethology. *Z. Tierpsychol.* 20, 410–433. doi: 10.1111/j.1439-0310.1963.tb01161.x

Titze, I. R. (1994). *Principles of Voice Production*. Upper Saddle River, NJ: Prentice Hall.

Tobias, M. L., Viswanathan, S. S., and Kelley, D. B. (1998). Rapping, a female receptive call, initiates male–female duets in the South African clawed frog. *Proc. Natl. Acad. Sci. U.S.A.* 95, 1870–1875. doi: 10.1073/pnas.95.4.1870

Todd, G. A., and Palmer, B. (1968). Social reinforcement of infant babbling. *Child Dev.* 39, 591–596. doi: 10.2307/1126969

Toro, J. M., Trobalon, J. B., and Sebastián-Gallés, N. (2003). The use of prosodic cues in language discrimination tasks by rats. *Anim. Cogn.* 6, 131–136. doi: 10.1007/s10071-003-0172-0

Trainor, L. J., Austin, C. M., and Desjardins, R. N. (2000). Is infant-directed speech prosody a result of the vocal expression of emotion? *Psychol. Sci.* 11, 188–195. doi: 10.1111/1467-9280.00240

Trehub, S. E., Becker, J., and Morley, I. (2015). Cross-cultural perspectives on music and musicality. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 370, 20140096. doi: 10.1098/rstb.2014.0096

Tressler, J., and Smotherman, M. S. (2009). Context-dependent effects of noise on echolocation pulse characteristics in free-tailed bats. *J. Comp. Physiol.* 195, 923–934. doi: 10.1007/s00359-009-0468-x

Tuttle, M. D., and Ryan, M. J. (1982). The role of synchronized calling, ambient light, and ambient noise, in anti-bat-predator behavior of a treefrog. *Behav. Ecol. Sociobiol.* 11, 125–131. doi: 10.1007/BF00300101

Verga, L., Bigand, E., and Kotz, S. A. (2015). Play along: effects of music and social interaction on word learning. *Front. Psychol.* 6:1316. doi: 10.3389/fpsyg.2015.01316

Verhoef, T. (2012). The origins of duality of patterning in artificial whistled languages. *Lang. Cogn.* 4, 357–380. doi: 10.1515/langcog-2012-0019

Vernes, S. C. (2016). What bats have to say about speech and language. *Psychon. Bull. Rev.* 1–7. doi: 10.3758/s13423-016-1060-3 [Epub ahead of print].

Versace, E., Endress, A. D., and Hauser, M. D. (2008). Pattern recognition mediates flexible timing of vocalizations in nonhuman primates: experiments with cottontop tamarins. *Anim. Behav.* 76, 1885–1892. doi: 10.1016/j.anbehav.2008.08.015

Wagner, M., and Watson, D. G. (2010). Experimental and theoretical advances in prosody: a review. *Lang. Cogn. Process.* 25, 905–945. doi: 10.1080/01690961003589492

Ward, N., and Tsukahara, W. (2000). Prosodic features which cue back-channel responses in English and Japanese. *J. Pragmat.* 32, 1177–1207. doi: 10.1016/S0378-2166(99)00109-5

Weisberg, P. (1963). Social and nonsocial conditioning of infant vocalizations. *Child Dev.* 34, 377–388. doi: 10.1111/j.1467-8624.1963.tb05145.x

Wilson, M., and Wilson, T. P. (2005). An oscillator model of the timing of turn-taking. *Psychon. Bull. Rev.* 12, 957–968. doi: 10.3758/BF03206432

Wiltermuth, S. S., and Heath, C. (2009). Synchrony and Cooperation. *Psychol. Sci.* 20, 1–5. doi: 10.1111/j.1467-9280.2008.02253.x

Wray, A. (1998). Protolanguage as a holistic system for social interaction. *Lang. commun.* 18, 47–67. doi: 10.1016/S0271-5309(97)00033-5

Wright, A. A., Rivera, J. J., Hulse, S. H., Shyan, M., and Neiworth, J. J. (2000). Music perception and octave generalization in rhesus monkeys. *J. Exp. Psychol.* 129, 291–307. doi: 10.1037/0096-3445.129.3.291

Yip, M. J. (2006). The search for phonology in other species. *Trends Cogn. Sci.* 10, 442–446. doi: 10.1016/j.tics.2006.08.001

Yoshida, S., and Okanoya, K. (2005). Animal cognition evolution of turn-taking: a bio-cognitive perspective. *Cogn. Stud.* 12, 153–165.

Yosida, S., Kobayasi, K. I., Ikebuchi, M., Ozaki, R., and Okanoya, K. (2007). Antiphonal vocalization of a subterranean rodent, the Naked Mole-Rat (*Heterocephalus glaber*). *Ethology* 113, 703–710. doi: 10.1111/j.1439-0310.2007.01371.x

Zatorre, R. J., Belin, P., and Penhune, V. B. (2002). Structure and function of auditory cortex: music and speech. *Trends Cogn. Sci.* 6, 37–46. doi: 10.1016/S1364-6613(00)01816-7

Zimmermann, E. (2010). Vocal expression of emotion in a nocturnal prosimian primate group, mouse lemurs. *Handb. Behav. Neurosci.* 19, 215–225. doi: 10.1016/B978-0-12-374593-4.00022-X

Zimmermann, E., Leliveld, L. M. C., and Schehka, S. (2013). "Toward the evolutionary roots of affective prosody in human acoustic communication: a comparative approach to mammalian voices," in *Evolution of Emotional Communication: From Sounds in Nonhuman Mammals to Speech and Music in Man*, eds E. Altenmüller, S. Schmidt, and E. Zimmermann (Oxford: Oxford University Press), 116–132.

Zimmermann, U., Rheinlaender, J., and Robinson, D. (1989). Cues for male phonotaxis in the duetting bushcricketLeptophyes punctatissima. *J. Comp. Physiol. A* 164, 621–628. doi: 10.1007/BF00614504

Zuberbühler, K., Jenny, D., and Bshary, R. (1999). The predator deterrence function of primate alarm calls. *Ethology* 105, 477–490. doi: 10.1046/j.1439-0310.1999.00396.x