THEORIA ET HISTORIA SCIENTIARUM, VOL. XVI
Ed. Nicolaus Copernicus University 2019

DOI: http://dx.doi.org/10.12775/ths.2019.005

Piera Filippi*

Institute of Language, Communication and the Brain, Aix-en-Provence, France;
Laboratoire Parole et Langage LPL UMR 7309, Centre National de la Recherche Scientifique, Aix-Marseille Université, Aix-en-Provence, France;
Laboratoire de Psychologie Cognitive LPC UMR7290, Centre National de la Recherche Scientifique, Aix-Marseille Université, Marseille, France
pie.filippi@gmail.com

Emotion Communication Through Voice Modulation: Insights on Biological and Evolutionary Underpinnings of Language

Abstract. The aim of this review is to enhance our understanding of the role of emotional communication in the emergence of language. I provide data on the following research topics: 1) Cross-species comparative approach to the anatomical principles governing emotional vocal production. 2) Analysis of acoustic parameters conveying emotional arousal and valence through voice modulation across human cultures and a wide variety of vocalizing nonhuman animals. On this regard, I will describe the evolutionary advantage of being able to identify emotional content in both heterospecific and conspecific vocalizations. 3) The relative salience of emotional voice modulation and verbal content in emotional meaning processing, as an indicator of the biological role of voice modulation in the emergence of language. Finally, I propose that co-evolutionary dynamics between genetic transmission of the cognitive mechanisms underpinning language and socio-cultural transmission of vocal behaviors are responsible for the emergence of the abilities involved in language.

Keywords: language evolution; co-evolutional; emotion; prosody; word meaning; interactions.

* Piera Filippi is supported by grants ANR-16-CONV-0002 (ILCB), ANR-11-LABX-0036 (BLRI) and the Excellence Initiative of Aix-Marseille University (A*MIDEX). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

> I cannot doubt that language owes its origin to the imitation and modification, aided by signs and gestures, of various natural sounds, the voices of other animals, and man's own distinctive cries [. . .] we may conclude from a widely spread analogy that this power would have been especially exerted during the courtship of the sexes, serving to express various emotions, as love, jealousy, triumph, and serving as a challenge to their rivals. The imitation by articulate sounds of musical cries might have given rise to words expressive of various complex emotions.
>
> Darwin (1871)

Introduction

In *The descent of man, and Selection in Relation to Sex* (1871), Darwin hypothesized that the ability to modulate the voice to express emotions, which is shared across animal species, might have evolved into the ability to express emotional content in words, thus starting the path of language evolution. Yet, our understanding of the role of vocal expression of emotions in language evolution remains the same today as when Darwin first tackled it in 1871. The overarching aim of this review is to describe recent findings on vocal emotion communication in animal communication systems and in language. If taken together, these findings may provide insights into future research aimed at addressing the expression of emotion through voice modulation as a biologically universal factor underpinning the emergence of language.

A growing body of research has focused on the expression of emotions across animal species, using multiple models for emotion classification. Here, it is important to describe two key theoretical paradigms that have been used to investigate emotions, and explain which one of these models is more suitable for a cross-species comparison approach. The first model belongs to a long-standing research tradition centered on the study of emotions as discrete categories such as happiness or anger (Ekman, 1992). A different approach, which includes *dimensional emotion models,* argues that emotional states can be classified based on their *valence* (positive or negative) and their *arousal* level (i.e., activation or responsiveness levels, typically classified as low/high or calm/excited) (Mendl, Burman, & Paul, 2010; Mendl, Paul, & Chittka, 2011; Russell, 1980). Crucially, unlike discrete emotions and emotional valence, different levels of arousal can be directly linked to the physiological state of the signaler, enabling quantitative mapping of physiological,

behavioral and acoustic data (Briefer, Tettamanti, & Mcelligott, 2015a; Briefer et al., 2015b; Maigrot et al., 2017). This is particularly suitable for precise quantitative comparisons within and across species, aimed at exploring biologically universal aspects of emotional expression. For this reason, in this work, I adopt a comparative approach on emotional voice modulation across nonhuman animal and human communication systems, focusing specifically on emotional arousal. Within this approach, I will provide empirical data on the following three research topics: 1) the anatomical principles governing vocal production; 2) analysis of acoustic parameters conveying emotional arousal through voice modulation. This analysis will provide insights into acoustic parameters constituting a biologically universal vocal code that enables the perception of emotional states across human cultures and a wide variety of vocalizing nonhuman animals; 3) the relative salience of emotional voice modulation and verbal content in emotional meaning processing, as an indicator of the biological role of voice modulation in the emergence of language. I will highlight the centrality of interactional dynamics as the natural place where, most likely, emotional voice modulation favored language evolution. This review will ultimately provide insights on empirical findings supporting Darwin's hypotheses on the effect of emotional expression through voice modulation in driving the evolution of the ability for arbitrary word-meaning associations in humans (Darwin, 1871).

Anatomical Mechanisms Underpinning Emotional Voice Modulation Across Animal Species: A Comparative Approach

Two strands of analysis are relevant in the context of comparative investigation on animal vocal communication: (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 & Eldredge, 1977; Hauser, Chomsky, & Fitch, 2002). In this review, I will report empirical data on mechanisms governing voice modulation, which are shared across a wide variety of species across animal classes, thus constituting evolutionary ancient homolog traits. In order to understand how distinct vocal sounds are produced across animal species, and the connection between emotional states and the voice, it is crucial to consider recent research, which has provided an

explanatory theory on voice production, the so-called source – filter theory of voice production (Taylor & Reby, 2010).

According to the source-filter theory, vocalizations are generated by tissue vibrations stimulated by the passage of air in the sound "source": the larynx in mammals, amphibians and reptiles, and the syrinx in birds. The signal produced by the source is subsequently filtered by the resonances of the supralaryngeal vocal tract (the "filter") with certain frequencies being enhanced or attenuated. Source vibration determines the fundamental frequency of the vocalization (F0), in other words how low or high a voice sounds. The filter resonances shape its spectral content, producing concentrations of acoustic energy in particular frequency bands (called 'formants'), which are perceived as vowel-like sounds. For instance, when humans vocalize, air passes from the lungs through an opening between the vocal folds, causing them to vibrate. These vibrations are transmitted through the air in the vocal tract to the openings of the mouth and nose, where they are broadcast into the environment. This theory has crucial implications for exploring the link between emotional physiology and the physical mechanisms of voice production. Indeed, physiological changes associated with change in emotional states affect voice production by acting on the muscles required for vocal production. For instance, the diaphragm, intercostals and vocalis muscles, which are critical in sound production, can be affected by muscular tension, and this alters the way air flows through the system and thus the quality of the sounds produced (Titze, 1994). This may induce vocal folds to vibrate at their natural limit, generating sound waves at heightened amplitude. These sound waves may be perceived as harsh sounds (Taylor & Reby, 2010).

Biologically Universal Meanings at the Origins of Language

The presence of disturbance or danger in the environment, for instance the imminent attack of a predator, activates the sympathetic nervous system of an individual. The consequent tension in the body of this individual may affect acoustic parameters in her voice, which, thus, reflect heightened levels of emotional intensity. In parallel, correct identification of heightened levels of emotional arousal from voice modulation activates the sympathetic nervous system in listeners. This enables them to react to the given vocal signal appropriately, but also in an automatic and fast way (fight-or-flight response) to imminent life threats (Scherer, 1986; Scherer, 2003). These types of responses are adaptive and may be universally shared across animals. For

instance, an animal may attempt to avoid an attacking predator by fleeing, by using defensive postures, or aggressive counterattack (Edmunds, 1974). Crucially, these reactions are determined by a change in the physiological state, which could alter the probability that an animal will be able to outrun a potential predator, and ultimately, survive. Indeed, states of heightened emotional intensity may induce heightened muscular tension, which prepares the signaler for immediate action (Arnal, Flinker, Kleinschmidt, Giraud, & Poeppel, 2015).

These physiological changes linked to fight-or-flight responses, which typically correspond to heightened levels of emotional arousal, can be reflected in the acoustic features of the vocalizations, as detailed in the previous section. In turn, the ability to recognize heightened levels of arousal in vocalizations may help avoiding threats or disturbances in the surrounding environment, as, for instance, the imminent approach of a predator. A vocalization produced by a signaler that is in a heightened emotional intensity state may induce fear or alertness in the listeners, thus prompting them to avoid dangers or disturbances in the surroundings. Hence, a high arousal vocalization may have been shaped by selection to affect others' behavior in an urgent manner (Fitch, Neubauer, & Herzel, 2002).

Importantly, survival may be facilitated by the ability to identify emotions not only in vocalizations emitted by conspecifics, but also by members of other species (Nesse, 1990). This ability may provide information that is crucial to responding appropriately. It has been shown that nonhuman animals' "eavesdropping" on another species alarm calls increases opportunities for survival (de Boer, Wich, Hardus, & Lameira, 2015; Fallow, Gardner, & Magrath, 2011; Kitchen, Bergman, Cheney, Nicholson, & Seyfarth, 2010; Magrath, Pitcher, & Gardner, 2009; Owings & Morton, 1998). Generally, the ability to respond appropriately to heterospecific calls, which may presuppose the ability to recognize their level of emotional arousal (Mendl et al., 2010), is the result of a signaling system that affords inter-specific beneficial outcomes in dangerous contexts (Adolphs, 2013).

Recent research has explored the acoustic parameters of voice modulation that enable the perception of emotional arousal states across human cultures and nine animal species spanning across all classes of terrestrial vertebrates (i.e. amphibia, reptilia, aves, and mammalia). In this study, Filippi et al. (2017a) provided empirical data showing that humans from three language groups (English, German and Mandarin) use information related to the frequency domain of vocalizations to identify emotional content in vocalizations across all species included in the study. These results suggest that fundamental mechanisms of vocal emotional expression are biologically

rooted in humans and widely shared among vocalizing vertebrates. Hence, modulation of frequency values in the voice might represent a cross-cultural universal signaling system. Question then is as to whether emotional voice modulation represents a biologically universal code used across species. Filippi et al. (2017a) point to a positive answer to this question. In fact, this study suggests that the ability to process changes in voice modulation as indicators of emotional arousal in animal calls may have emerged in the early stages of the evolution of vocalizing animals and have been preserved across a broad range of animal species (cf. Filppi, Gogoleva, Volodina, Volodin, & de Boer, 2017b). In line with the outcome of this study, multiple studies on arousal perception show that humans rate human, piglet, cat and dog vocalizations with higher F0 as expressing higher emotional arousal (Faragó et al., 2014; Laukka, Juslin, & Bresin, 2005; Maruščáková et al., 2015; McComb, Taylor, Wilson, & Charlton, 2009).

Taken together, these studies suggest that the ability to express and identify heightened levels of emotional arousal in both conspecific and heterospecific vocalizations, is evolutionary adaptive and ancient (Charlton & Reby, 2016; Darwin, 1871). Hence, this line of research provides evidence for a phylogenetic continuity of emotional communication across species, in terms of acoustic parameters involved in vocal production (Bowling, Gingras, Han, Sundararajan, & Opitz, 2013; Briefer, 2012; Linhart, Ratcliffe, Reby, & Špinka, 2015; Morton, 1977; Reichert, 2013; Stoeger, Baotic, Li, & Charlton, 2012; Stoeger, Charlton, Kratochvil, & Fitch, 2011; Templeton, Greene, & Davis, 2005) and in the perception of emotional content in these vocalizations (Belin et al., 2008; Faragó et al., 2014; McComb et al., 2009; Pongrácz, Molnár, & Miklósi, 2006; Sauter, Eisner, Ekman, & Scott, 2010).

Crucially, this line of research contributes to the study of the evolutionary precursors of human language in animal communication systems, which typically focuses on animals' ability for sound-meaning associations (Engesser, Ridley, & Townsend, 2016). Indeed, comparative research on animal communication has explored animal calls as "functionally referential" (see Hauser, 1992). Functionally referential calls are traditionally described as calls that provide listeners with sufficient information to determine the individual object *denoted* by the signal. For instance, in a very influential study Seyfarth, Cheney, & Marler (1980) suggested that the vervet monkey alarm calls denote "snake", "eagle", or "leopard", and trigger appropriate responses in the listeners, such as looking up upon hearing the call denoting "eagle". However, this research has overlooked meaning effects of emotional voice modulation in nonhuman communication systems. In fact, the expression of emotional content, which enables to imply the presence of potential threat

in the environment has a strong communicative effect (Fischer & Price, 2017; Manser, Seyfarth, & Cheney, 2002; Price et al., 2015). Hence, it is plausible that the ability to associate emotional content to specific changes in the voice, which goes beyond the association between discrete units and their denotation, constitutes a fundamental evolutionary precursor of language. Future research shall further address this line of investigation, which will advance the study of the evolutionary vocal precursors of language, placing it into a fruitful perspective (see Liao, Zhang, Cai, & Ghazanfar, 2018).

Emotion Communication Through Voice Modulation: A Biological Universal Underpinning Language Evolution?

Humans typically combine two sources of information when speaking: linguistic (e.g., lexical information, morphology or syntax) and paralinguistic information (e.g., body posture, facial expression, prosodic modulation of the voice and pragmatic context) (Hockett, 1960). For the purposes of this review, we will focus on the cognitive link between two auditory channels in speech: *prosodic modulation*, which includes timing, frequency spectrum and amplitude (Lehiste, 1970) – and *lexical information*. Specifically, the question I will attempt to address in this section is whether there is evidence for an evolutionary continuity link between emotional modulation of the voice and lexical information. Studies on the relative salience of these two channels in word meaning processing are relevant for this research question, as prominence of one channel over the other on both cognitive and neural levels may provide insights into which channel is more ancient than the other. Thus, this line of research may provide empirical evidence on the evolutionary role of emotional voice modulation on the ability to articulate and understand spoken lexical units.

In fact, emotional communication can take place by integrating prosody and lexical information, which can interact with each other, for instance, through priming or simultaneous interaction. Research has shown that lexical information and prosodic modulation of spoken units prime the interpretation of a following target word in an emotion-congruent manner. When the two channels are congruent, emotional prosody strengthens memory of affective words (Schirmer, Kotz, & Friederici, 2002). Furthermore, phonetic information and prosodic modulation of the voice can simultaneously express contrasting contents. For instance, this is the case when someone says "I'm sad!" with happy prosody. Filippi et al. (2017c) found that when the two expressive channels are incongruent, prosody dominates over phonetic

information in recruiting cognitive resources for emotion identification. Accordingly, studies have found that emotional modulation of the voice speeds up lexical decision tasks (Nygaard & Lunders, 2002) and orients word identification in cases of lexical ambiguity (Filippi, Gingras, & Fitch, 2014; Kim & Sumner, 2017). In addition, much research has addressed the effect of emotional prosody in verbal language from a neuroscientific approach. This research shows that the vocal expression of emotional states deeply impacts language processing, involving the most ancient brain circuitries (Dalgleish, 2004; Kotz & Paulmann, 2011). In sum, cross-disciplinary studies on the role of emotional prosody in language contribute to a relatively consistent picture: the ability for emotional communication through prosodic modulation of the voice is evolutionary older than the ability to process lexical information (Brown, 2017; Filippi, 2016; Fitch, 2010; Mithen, 2005). Furthermore, findings from studies on the role of prosodic modulation of the voice in language acquisition are consistent with the hypothesis that it may have facilitated the emergence of the ability for language. Indeed, research shows that prosody drives words' segmentation (Johnson & Jusczyk, 2001), favors accurate word-meaning mapping (Filippi, Gingras, & Fitch, 2014; Filippi, Laaha, & Fitch, 2017) and is used for syntactic disambiguation (Soderstrom, Seidl, Kemler Nelson, & Jusczyk, 2003). Accordingly, research on language development parallels these works, showing that prosodic cues favor lexical access and syntactic analysis at an ontogenetic level, orienting language acquisition in preverbal children (de Carvalho, Dautriche, Lin, & Christophe, 2017; Gout, Christophe, & Morgan, 2004).

It is worth emphasizing that these studies constitute an initial step towards an increasing understanding of the dynamics driving the evolution of language. Future studies should address the effect of emotional prosody on at least the following core abilities involved in language: a) phonetic articulation, b) process syntactical connections and c) understand the interlocutors' state of mind. Crucially, extensive research on each of these abilities suggests that they are present in nonhuman animals and may constitute a scaffold for the emergence of language. In fact, computational models of the vocal tract of rhesus macaques (*Macaca mulatta*) (Fitch, de Boer, Mathur, & Ghazanfar, 2016), as well as anatomical analyses of baboons' (*Papio papio*) vocal tract, combined with the acoustic analyses of their vocalizations (Boë et al., 2017) suggest that monkey vocal tracts are predisposed to produce vowel-like sounds and a range of consonants (cf. Lameira, Maddieson, & Zuberbühler, 2014). What monkeys are missing to be able to speak is a human-like neural control over vocal tract muscles, which would enable vocal learning and combinatorial operations over speech

sounds, and thus, clearly intelligible speech. Secondly, multiple studies report that the ability for meaningful combination of calls is found across species of monkeys (Ouattara, Lemasson, & Zuberbühler, 2009; Collier et al., 2014; Zuberbühler, 2002), as well as in songbirds (Engesser, 2016). Researchers refer to this ability as the precursor of the human ability for syntax, stressing its rudimentary nature: In fact, nonhuman animals' ability for syntax typically involves two or three units, and it never generates a potentially infinite set of novel utterances, as it is the case for humans. Finally, debate on nonhuman sensitivity to the listeners' state of *knowledge* is, to date, still ongoing and far from a cross-disciplinary shared perspective on the topic. Nevertheless, extensive agreement has been reached on the assumption that nonhuman animals are able to act with specific behavioral goals, or to affect the *attentional* state of the listener (Fitch, Huber, & Bugnyar, 2010; Townsend et al., 2016). Taken together, these studies suggest that the abovementioned three core abilities constitute a tight evolutionary link between nonhuman animals' communication systems and language. However, the question of how emotional modulation of the voice affected the evolution of these core abilities into the human ability for language remains, to date, open to question.

Similarly, the question of how the ability for emotional voice modulation evolved into the ability to use prosodic features in the voice to modulate linguistic information remains open to future studies. In fact, in language, the so-called "linguistic prosody" modulates the information conveyed in the signal, by orienting the perception of phonetic information (Bosker, 2017), lexical items (van Donselaar, Koster, & Cutler, 2005), and morpho-syntactical connections (Soderstrom et al., 2003). Prosodic modulation of the voice affects perception of phrase boundaries, of a word (*lexical stress*), or of specific words within a sentence (*sentence focus*). Consider for instance, "MARY gave the book to John" vs. "Mary gave the book to JOHN". Here, the two sentences are identical from a phonetic point of view. However, by accenting one word or the other, the speaker guides the listener's perception of the sentence. In addition, linguistic prosody may be used to distinguish different meanings in phonetically identical words in tone languages or to infer statement types, for instance, to infer an assertion from a question or a command (Cutler, Dahan, & van Donselaar, 1997).

It is plausible that the ability to modulate the vocal signal to express emotional content evolved into the ability to modulate language-specific prosodic parameters in the voice. The emergence of all the abilities involved in language may have been affected by co-evolutionary dynamics between cognitive mechanisms – as, for instance, increasingly fine-tuned neural

control over the muscles involved in vocal production – and socio-cultural processes of language transmission (Figure 1). Crucially, as I will argue in the next section, emotional modulation of the voice within social interactions was the cradle and the main force orienting the evolution of a biologically universal code into the whole set of cognitive mechanisms and processes enabling the human ability for language.

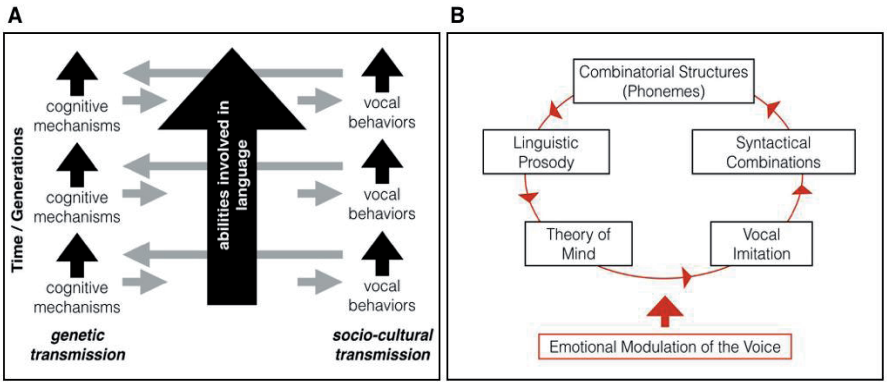

Figure 1. Co-evolutionary dynamics underpinning language.

A. The arrows indicate the inter-connection between the evolution of cognitive mechanisms, the abilities involved in language and vocal behaviors. Figure inspired by Deacon (1998). Over time and generations, modifications in cognitive mechanisms **(left)** and vocal behaviors **(right)** are transmitted through genetic and socio-cultural transmission, respectively. For a given generation or time period, the existing cognitive mechanisms enable specific sets of abilities involved in language. These abilities result into specific vocal behaviors, which in turn affect the evolution of cognitive mechanisms throughout socio-cultural transmission. Behavior-driven changes in the social environment may be adaptive for humans, influencing which genes will be passed on to the next generation. **B.** Emotional modulation of the voice may have triggered the emergence of language, which includes a broad set of inter-connected abilities (central black arrow in A), as for instance the ones included in black rectangles. In addition, these abilities retro-act on each other, pushing the evolution of language forward. The hypothesis proposed in this paper is that emotional modulation of the voice has a strong cognitive effect in the dynamics underpinning the evolution of these inter-connected abilities (red circle with arrows).

Concluding Remarks: Social Interactions at the Origins of Language

Prosodic modulation of the voice dramatically affects dynamics of social interactions. For instance, in humans, acoustic modulation of the voice conveys social affect such as politeness (Brown & Levinson, 2006; Ohala, 1983), and, as in nonhuman animals, it expresses the dominance state of the speaker (Owren & Rendall, 2001; Tusing, 2000). Furthermore, multiple comparative studies on animal vocal communication systems report on the widely shared ability to use voice modulation for vocal coordination in species spanning all classes of animals. Specifically, vocal coordination in animals results in the following types of behaviors: *choruses*, *duets* and *antiphonal calling* (Yoshida & Okanoya, 2005). Crucially, these vocal behaviors occur in contexts that involve various degrees of emotional arousal, as, for instance, territorial defense, social bonding and sexual advertisement. Choruses, which are typically produced by males, are commonly performed in anurans and insects for sexual advertisement or as an anti-predator defensive behavior. Duets are performed by members of a pair (e.g., sexual mates, caregiver-juvenile), who coordinate vocal interactions within a precise time window to strengthen and display pair bonding. Duets are observed in insects, anurans, birds, and mammals. Finally, antiphonal calls are exchanged between multiple individuals, independently from their sex, favoring group cohesion and diverting outsiders. These calls are reported in several species of birds and mammals. This evidence suggests that the ability for emotional vocal coordination, which is widely shared across animal species, might have scaffolded the evolution of language. Importantly, this ability is central in the context of linguistic interactions, where humans, use prosodic modulation of the voice to coordinate the exchange of vocal utterances (see Filippi, 2016, for a review). In addition, contingent turn-taking in speech addressed to infants is fundamental in the development of linguistic and social competences of the child (Romeo et al., 2018).

Recent research on marmoset monkeys (*Callithrix jacchus*) investigated whether monkey calls are automatic reflexes intrinsically linked to internal states such as emotional arousal, or whether they result from a degree of volitional vocal control, and can thus be strategically used to manipulate the listeners' reactions (Liao et al., 2018). Interestingly, the authors found that physical distance from a conspecific and visual access to her/him affected the level of arousal, resulting into different vocal behaviors. However, these changes in internal states were not reflected into systematic variation on voice

modulations of the calls or in the production of different call types. In fact, a close examination of the data revealed that vocal production in this species of monkeys results from the combination of variations of the emotional state of the caller with extrinsic factors for social coordination, such as timing of a conspecific vocalization.

In conclusion, future work on further species of nonhuman animals should address the relative role of social factors and emotional arousal state of callers (in both positively- and negatively-valenced contexts) in vocal production. Ideally, this research should include precise measurements of emotional states of the callers, which can be collected through physiological data (e.g. heart rate, respiration rate, and skin conductance, see Briefer et al., 2015). Notably, the measurements, which reflect into acoustic features of the vocal signal, can be analyzed and compared across human cultures and nonhuman animal species. Based on this methodological advantage, this research will significantly advance our understanding of the line of continuity between acoustic modulation of the voice in animal vocalizations and in language. This will help pinpoint a biologically universal code for emotional vocal communication, which is used across all vocal species. Ultimately, this will provide insights into the evolutionary role of emotional voice modulation in the transition from primate-like calls to human speech and on its cognitive role in modern humans' language.

References

Adolphs, R. (2013). The biology of fear. *Current Biology*, 23(2), R79–R93.

Arnal, L. H., Flinker, A., Kleinschmidt, A., Giraud, A. L., & Poeppel, D. 2015. Human screams occupy a privileged niche in the communication soundscape. *Current Biology*, 25(15), 2051–2056. https://doi.org/10.1016/j.cub.2015.06.043

Belin, P., Fecteau, S., Charest, I., Nicastro, N., Hauser, M. D., & Armony, J. L. (2008). Human cerebral response to animal affective vocalizations. *Proceedings of the the Royal Society B*, *275*(1634), 473–481. https://doi.org/10.1098/rspb.2007.1460

Boë, L.-J., Berthommier, F., Legou, T., Captier, G., Kemp, C., Sawallis, T. R., … Payan, Y. (2017). Evidence of a vocalic proto-system in the baboon (*Papio papio*) suggests pre-hominin speech precursors. *PLOS One*, 12(1), e0169321. https://doi.org/10.1371/journal.pone.0169321

Bosker, H. R. (2017). Accounting for rate-dependent category boundary shifts in speech perception. *Attention, Perception, & Psychophysics*, 79(1), 333–343. https://doi.org/10.3758/s13414-016-1206-4

Bowling, D. L., Gingras, B., Han, S., Sundararajan, J., & Opitz, E. C. L. (2013). Tone of voice in emotional expression: Relevance for the affective character of musical mode. *Journal of Interdisciplinary Music Studies*, 7, 29–44. https://doi.org/10.4407/jims.2014.06.002

Briefer, E. (2012). Vocal expression of emotions in mammals: Mechanisms of production and evidence. *Journal of Zoology*, 288(1), 1–20.

Briefer, E. F., Tettamanti, F., & Mcelligott, A. G. (2015a). Animal studies repository emotions in goats: Mapping physiological, behavioural and vocal profiles. *Animal Behaviour*, 99, 131–143.

Briefer, E. F., Maigrot, A. L., Mandel, R., Freymond, S. B., Bachmann, I., & Hillmann, E. (2015b). Segregation of information about emotional arousal and valence in horse whinnies. *Scientific Reports*, 4, 9989.

Brown, P., & Levinson, S. C. (2006). Chapter 22: Politeness: Some universals in language usage. In A. Jaworski & N. Coupland (Eds.), *The discourse reader* (pp. 311–323). Abingdon: Routledge.

Brown, S. (2017). A joint prosodic origin of language and music. *Frontiers in Psychology*, *8*, 1894. https://doi.org/10.3389/fpsyg.2017.01894

Charlton, B. D., & Reby, D. (2016). The evolution of acoustic size exaggeration in terrestrial mammals. *Nature Communications*, *7*, 12739. https://doi.org/10.1038/ncomms12739

Collier, K., Bickel, B., van Schaik, C. P., Manser, M. B., & Townsend, S. W. (2014). Language evolution: Syntax before phonology? *Proceedings of the Royal Society B*, 281(1788), 20140263. https://doi.org/10.1098/rspb.2014.0263

Cutler, A., Dahan, D., & van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech*, 40, 141–201. https://doi.org/10.1177/002383099704000203

Dalgleish, T. (2004). The emotional brain. *Nature Reviews Neuroscience*, 5(7), 583.

Darwin, C. (1871). *The descent of man, and selection in relation to sex*. London: John Murray.

Deacon, T. W. (1998). *The symbolic species: The co-evolution of language and the brain*. New York City, NY: W. W. Norton & Company.

de Boer, B., Wich, S. A., Hardus, M. E., & Lameira, A. R. (2015). Acoustic models of orangutan hand-assisted alarm calls. *The Journal of Experimental Biology*, 218(6), 907–914. https://doi.org/10.1242/jeb.110577

de Carvalho, A., Dautriche, I., Lin, I., & Christophe, A. (2017). Phrasal prosody constrains syntactic analysis in toddlers. *Cognition*, 163, 67–79. https://doi.org/10.1016/j.cognition.2017.02.018

Edmunds, M. (1974). *Defence in animals: A survey of anti-predator defences*. Harlow: Longman Publishing Group.

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, *6*(3–4), 169–200. https://doi.org/10.1080/02699939208411068

Engesser, S., Ridley, A. R., & Townsend, S. W. (2016). Meaningful call combinations and compositional processing in the southern pied babbler. *Proceedings of the*

National Academy of Sciences, 113(21), 5976–5981. https://doi.org/10.1073/pnas.1600970113

Fallow, P. M., Gardner, J. L., & Magrath, R. D. (2011). Sound familiar? Acoustic similarity provokes responses to unfamiliar heterospecific alarm calls. *Behavioral Ecology*, 22(2), 401–410. https://doi.org/10.1093/beheco/arq221

Faragó, T., Andics, A., Devecseri, V., Kis, A., Gácsi, M., & Miklósi, D. (2014). Humans rely on the same rules to assess emotional valence and intensity in conspecific and dog vocalizations. *Biology Letters*, 10(1), 20130926. https://doi.org/10.1098/rsbl.2013.0926

Filippi, P. (2016). Emotional and interactional prosody across animal communication systems: A comparative approach to the emergence of language. *Frontiers in Psychology*, 7, 1393. https://doi.org/10.3389/fpsyg.2016.01393

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

Filippi, P., Congdon, J. V., Hoang, J., Bowling, D. L., Reber, S. A., Pašukonis, A., [...] Güntürkün, O. (2017a). Humans recognize emotional arousal in vocalizations across all classes of terrestrial vertebrates: Evidence for acoustic universals. *Proceedings of the Royal Society B*, 284(1859), 20170990.

Filippi, P., Gogoleva, S. S., Volodina, E. V., Volodin, I. A., & de Boer, B. (2017b). Humans identify negative (but not positive) arousal in silver fox vocalizations: Implications for the adaptive value of interspecific eavesdropping. *Current Zoology*, 63(4), 445–456. https://doi.org/10.1093/cz/zox035

Filippi, P., Ocklenburg, S., Bowling, D. L., Heege, L., Güntürkün, O., Newen, A., & de Boer, B. (2017c). More than words (and faces): Evidence for a Stroop effect of prosody in emotion word processing. *Cognition and Emotion*, 31(5), 879–891.

Filippi, P., Laaha, S., & Fitch, W. T. (2017d). Utterance-final position and pitch marking aid word learning in school-age children. *Royal Society Open Science*, 4(8), 161035.

Fischer, J., & Price, T. (2017). Meaning, intention, and inference in primate vocal communication. *Neuroscience and Biobehavioral Reviews*, 82, 22–31. https://doi.org/10.1016/j.neubiorev.2016.10.014

Fitch, W. T., de Boer, B., Mathur, N., & Ghazanfar, A. A. (2016). Monkey vocal tracts are speech-ready. *Science Advances*, *2*(12), e1600723. https://doi.org/10.1126/sciadv.1600723

Fitch, W. T., Huber, L., & Bugnyar, T. (2010). Social cognition and the evolution of language: Constructing cognitive phylogenies. *Neuron*, 65(6), 795–814. https://doi.org/10.1016/j.neuron.2010.03.011

Fitch, W. T., Neubauer, J., & Herzel, H. (2002). Calls out of chaos: The adaptive significance of nonlinear phenomena in mammalian vocal production. *Animal Behaviour*, 63(3), 407–418. https://doi.org/10.1006/anbe.2001.1912

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

Gould, S. J., & Eldredge, N. (1977). Puncuated equilibria: The tempo and mode of evolution reconsidered. *Paleobiology*, 3(2), 115–151. https://doi.org/10.1017/S0094837300005224

Gout, A., Christophe, A., & Morgan, J. L. (2004). Phonological phrase boundaries constrain lexical access II. Infant data. *Journal of Memory and Language*, 51(4), 548–567. https://doi.org/10.1016/j.jml.2004.07.002

Hauser, M. D. (1996). *The evolution of communication*. Cambridge, MA: The MIT Press.

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science*, 298(5598), 1569–1579.

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

Johnson, E. K., & Jusczyk, P. W. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language*, 44(4), 548–567. https://doi.org/10.1006/jmla.2000.2755

Kim, S. K., & Sumner, M. (2017). Beyond lexical meaning: The effect of emotional prosody on spoken word recognition. *The Journal of the Acoustical Society of America*, 142(1), EL49–EL55.

Kitchen, D. M., Bergman, T. J., Cheney, D. L., Nicholson, J. R., & Seyfarth, R. M. (2010). Comparing responses of four ungulate species to playbacks of baboon alarm calls. *Animal Cognition*, *13*(6), 861–870. https://doi.org/10.1007/s10071-010-0334-9

Kotz, S. A., & Paulmann, S. (2011). Emotion, language, and the brain. *Language and Linguistics Compass*, 5(3), 108–125.

Lameira, A. R., Maddieson, I., & Zuberbühler, K. (2014). Primate feedstock for the evolution of consonants. *Trends in Cognitive Sciences*, 18(2), 60–62. https://doi.org/10.1016/j.tics.2013.10.013

Laukka, P., Juslin, P., & Bresin, R. (2005). A dimensional approach to vocal expression of emotion. *Cognition & Emotion*, 19(5), 633–653. https://doi.org/10.1080/02699930441000445

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

Liao, D. A., Zhang, Y. S., Cai, L. X., & Ghazanfar, A. A. (2018). Internal states and extrinsic factors both determine monkey vocal production. *Proceedings of the National Academy of Sciences*, 201722426. https://doi.org/10.1073/pnas.1722426115

Linhart, P., Ratcliffe, V. F., Reby, D., & Špinka, M. (2015). Expression of emotional arousal in two different piglet call types. *PLoS One*, 10(8), e0135414. https://doi.org/10.1371/journal.pone.0135414

Magrath, R. D., Pitcher, B. J., & Gardner, J. L. (2009). Recognition of other species' aerial alarm calls: Speaking the same language or learning another? *Proceedings of the Royal Society B: Biological Sciences*, 276(1657), 769–774. https://doi.org/10.1098/rspb.2008.1368

Maigrot, A. L., Hillmann, E., Anne, C., & Briefer, E. F. (2017). Vocal expression of emotional valence in Przewalski's horses (Equus przewalskii). *Scientific Reports*, 7(1), 8779.

Manser, M. B., Seyfarth, R. M., & Cheney, D. L. (2002). Suricate alarm calls signal predator class and urgency. *Trends in Cognitive Sciences*, 6(2), 55–57. https://doi.org/10.1016/S1364-6613(00)01840-4

Maruščáková, I. L., Linhart, P., Ratcliffe, V. F., Tallet, C., Reby, D., & Špinka, M. (2015). Humans (Homo sapiens) judge the emotional content of piglet (Sus scrofa domestica) calls based on simple acoustic parameters, not personality, empathy, nor attitude toward animals. *Journal of Comparative Psychology*, 129(2), 121–131.

McComb, K., Taylor, A. M., Wilson, C., & Charlton, B. D. (2009). The cry embedded within the purr. *Current Biology*, 19(13), R507–R508.

Mendl, M., Burman, O. H. P., & Paul, E. S. (2010). An integrative and functional framework for the study of animal emotion and mood. *Proceedings. Biological Sciences / The Royal Society*, 277(1696), 2895–2904. https://doi.org/10.1098/rspb.2010.0303

Mendl, M., Paul, E. S., & Chittka, L. (2011). Animal behaviour: Emotion in invertebrates? *Current Biology*, 21(12), R463–R465. https://doi.org/10.1016/j.cub.2011.05.028

Mithen, S. J. (2005). *The singing Neanderthals: The origins of music, language, mind, and body*. Cambridge, MA: Harvard University Press.

Morton, E. S. (1977). On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *The American Naturalist*, 111(981), 855–869. https://doi.org/10.1086/283219

Nesse, R. M. (1990). Evolutionary explanations of emotions. *Human Nature*, 1(3), 261–289.

Nygaard, L. C., & Lunders, E. R. (2002). Resolution of lexical ambiguity by emotional tone of voice. *Memory and Cognition*, 30(4), 583–593. https://doi.org/10.3758/BF03194959

Ouattara, K., Lemasson, A., & Zuberbühler, K. (2009). Campbell's monkeys concatenate vocalizations into context-specific call sequences. *Proceedings of the National Academy of Sciences*, pnas-0908118106.

Ohala, J. J. (1983). Cross-language use of pitch: An ethological view. *Phonetica*, 40(1), 1–18. https://doi.org/10.1159/000261678

Owings, D. H., & Morton, E. S. (1998). *Animal vocal communication: A new approach*. Cambridge: Cambridge University Press.

Owren, M. J., & Rendall, D. (2001). Sound on the rebound: Bringing form and function back to the forefront in understanding nonhuman primate vocal signaling. *Evolutionary Anthropology: Issues, News, and Reviews*, 10(2), 58–71. https://doi.org/10.1002/evan.1014

Pongrácz, P., Molnár, C., & Miklósi, Á. (2006). Acoustic parameters of dog barks carry emotional information for humans. *Applied Animal Behaviour Science*, 100(3–4), 228–240.

Price, T., Wadewitz, P., Cheney, D., Seyfarth, R., Hammerschmidt, K., & Fischer, J. (2015). Vervets revisited: A quantitative analysis of alarm call structure and context specificity. *Scientific Reports*, 5(13220), 1–11. https://doi.org/10.1038/srep13220

Reichert, M. S. (2013). Patterns of variability are consistent across signal types in the treefrog Dendropsophus ebraccatus. *Biological Journal of the Linnean Society*, 109(1), 131–145. https://doi.org/10.1111/bij.12028

Romeo, R. R., Leonard, J. A., Robinson, S. T., West, M. R., Mackey, A. P., Rowe, M. L., & Gabrieli, J. D. E. (2018). Beyond the 30-million-word gap: Children's conversational exposure is associated with language-related brain function. *Psychological Science*, 29(5), 700–710. https://doi.org/10.1177/0956797617742725

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6), 1161.

Sauter, D. A., Eisner, F., Ekman, P., & Scott, S. K. (2010). Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proceedings of the National Academy of Sciences of the United States of America*, 107(6), 2408–2412. https://doi.org/10.1073/pnas.0908239106

Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin March*, 99(2), 143–165. https://doi.org/10.1037/0033-2909.99.2.143

Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication*, 40(1), 227–256. https://doi.org/10.1016/S0167-6393(02)00084-5

Schirmer, A., Kotz, S. A., & Friederici, A. D. (2002). Sex differentiates the role of emotional prosody during word processing. *Cognitive Brain Research*, 14(2), 228–233.

Soderstrom, M., Seidl, A., Kemler Nelson, D. G., & Jusczyk, P. W. (2003). The prosodic bootstrapping of phrases: Evidence from prelinguistic infants. *Journal of Memory and Language*, 49(2), 249–267.

Stoeger, A. S., Baotic, A., Li, D., & Charlton, B. D. (2012). Acoustic features indicate arousal in infant giant panda vocalisations. *Ethology*, 118(9), 896–905. https://doi.org/10.1111/j.1439-0310.2012.02080.x

Stoeger, A. S., Charlton, B. D., Kratochvil, H., & Fitch, W. T. (2011). Vocal cues indicate level of arousal in infant African elephant roars. *The Journal of the Acoustical Society of America*, 130(3), 1700–1710. https://doi.org/10.1121/1.3605538

Taylor, A. M., & Reby, D. (2010). The contribution of source-filter theory to mammal vocal communication research. *Journal of Zoology*, 280(3), 221–236.

Templeton, C. N., Greene, E., & Davis, K. (2005). Allometry of alarm calls: Black-capped chickadees encode information about predator size. *Science*, 308(5730), 1934–1937.

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

Townsend, S. W., Koski, S. E., Byrne, R. W., Slocombe, K. E., Bickel, B., Boeckle, M., [...] Glock, H. J. (2017). Exorcising G rice's ghost: An empirical approach to studying intentional communication in animals. *Biological Reviews*, 92(3), 1427–1433. https://doi.org/10.1111/brv.12289

Tusing, K. (2000). The sounds of dominance. Vocal precursors of perceived dominance during interpersonal influence. *Human Communication Research*, 26(1), 148–171. https://doi.org/10.1093/hcr/26.1.148

Van Donselaar, W., Koster, M., & Cutler, A. (2005). Exploring the role of lexical stress in lexical recognition. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, 58(2), 251–273. https://doi.org/10.1080/02724980343000927

Yoshida, S., & Okanoya, K. (2005). Evolution of turn-taking: A bio-cognitive perspective. *Cognitive Studies*, 12(3), 153–165.

Zuberbühler, K. (2002). A syntactic rule in forest monkey communication. *Animal Behaviour*, 63(2), 293–299.