Brain networks of emotional prosody processing

Grandjean, Didier Maurice

How to cite

GRANDJEAN, Didier Maurice. Brain networks of emotional prosody processing. In: Emotion Review, 2020, p. 175407391989852. doi: 10.1177/1754073919898522

This publication URL: https://archive-ouverte.unige.ch/unige:134036
Publication DOI: 10.1177/1754073919898522

emotion review

Emotion Review
1–10
© The Author(s) 2020
ISSN 1754-0739
DOI: 10.1177/1754073919898522
https://journals.sagepub.com/home/emr

Brain Networks of Emotional Prosody Processing

Didier Grandjean
Department of Psychology and Educational Sciences and Swiss Center for Affective Sciences, University of Geneva, Switzerland

Abstract

The processing of emotional nonlinguistic information in speech is defined as emotional prosody. This auditory nonlinguistic information is essential in the decoding of social interactions and in our capacity to adapt and react adequately by taking into account contextual information. An integrated model is proposed at the functional and brain levels, encompassing 5 main systems that involve cortical and subcortical neural networks relevant for the processing of emotional prosody in its major dimensions, including perception and sound organization; related action tendencies; and associated values that integrate complex social contexts and ambiguous situations.

Keywords
affective science, emotion, emotional prosody, vocalization

Introduction

During everyday social interactions, nonhuman and human primates are sensitive to affective signals conveyed through vocalizations. Primate emotional vocalizations can be divided in two broad domains: nonverbal emotional vocalizations and verbal emotional vocalizations. Among the nonverbal emotional vocalizations, affective bursts are important during social interactions, especially for nonhuman primates. In humans, however, the emotional aspects in the auditory domain are often embedded in linguistic content, which has been labeled "emotional prosody." This concept can be defined as suprasegmental and segmental modifications of the speech produced during an emotional episode (Bänziger, Patel, & Scherer, 2014; Bänziger & Scherer, 2005; Grandjean, Bänziger, & Scherer, 2006; Scherer, 1995). The modifications of suprasegmental aspects are largely related to the perceived melody of the voice and are correlated with acoustic modifications such as the fundamental F0 (F0 corresponds to the physical fundamental frequency of a sound and is related to pitch, which is a perceptual quality of the frequency), which is perceived as pitch; and the energy envelope (Banse & Scherer, 1996), which is perceived as loudness variability. The segmental modifications are more related to the distribution of energy in the spectral domain and are correlated with the resonance modifications that impact on, for example, the acoustical structure of the formants. These acoustical features are also important for the distinction of nonverbal emotional vocalizations (Sauter, Eisner, Calder, & Scott, 2010). Different mechanisms are involved in the genesis of these acoustic and perceived vocal modifications during an emotional episode, including respiration, phonation, and articulation (Scherer, Johnstone, & Klasmeyer, 2003). The ability of humans to produce speech is crucial for information exchange at the linguistic, semantic, and pragmatic levels (Wagner & Watson, 2010; Wilson & Wharton, 2006), but the ability to build up a representation of the emotional mental states of others during social interactions is also an important function of emotional prosody. Our cognitive equipment is crucial for this ability and is based on a series of distributed neural networks that are central to the perception and interpretation of perceived vocal changes during social interactions (Grandjean et al., 2006).

Brain Mechanisms

Two brain mechanisms are essential for perceiving and interpreting these acoustic and perceptual changes during an emotional episode in others: feedforward and feedback information flows (Fries, 2005). The first mechanism is based on the capacity of the central nervous system to organize information provided by peripheral sensory systems and is related to feedforward information processing along the ascending auditory pathway including different nuclei in the brainstem (the cochlear nucleus and the superior olivary complex), the midbrain with the inferior colliculus, and subcortical regions including the medial geniculate nucleus in the thalamus and cortical areas including primary and secondary auditory regions. The second mechanism is related to the ability to infer contextual meanings and is organized as feedback information processing that is able to influence the first mechanism, that is, the feedforward information flow. This second mechanism is essentially implemented in

Corresponding author: Didier Grandjean, Department of Psychology and Educational Sciences, University of Geneva, 40 bd du Pont d'Arve, Geneva, 1205, Switzerland.
Email: didier.grandjean@unige.ch

cortical areas including the frontal and temporal regions. These two mechanisms influence each other dynamically to provide a global sense of perceived speech during social interactions and allows humans to interpret the complex intentions of others in a sophisticated manner, especially during challenging situations.

One important characteristic of the auditory system in animals is that it can represent information that is not directly accessible through the other senses. If you are exposed to new information, for instance, when somebody approaches you from behind, your visual system cannot provide relevant information, and your tactile system can provide information only if you come into physical contact with the person or are near enough, for example, to sense the heat of the other person's body. Smell can be useful, but this system is not well adapted to precise spatial localization. In contrast, the auditory system can detect sounds all around you, even if this information is behind you or hidden by obstacles or light levels, for example, during the night. This capacity of the auditory system to represent events in a 360° sphere is thus crucial for survival.

Voice-Sensitive Temporal Brain Areas

Human voice production is a complex system that has been instrumentalized to exchange information through organized sound shaping: speech. A large corpus of literature has documented how temporal and frontal brain areas are organized in ventral and dorsal streams involved in speech processing in humans (e.g., Hickok & Poeppel, 2007, for a review). However, it is only in 2000 that Belin and colleagues (Belin, Zatorre, Lafaille, Ahad, & Pike, 2000) showed that a specific dedicated temporal brain region is more sensitive to the human voice in general, not restricted to speech per se, compared to environmental sounds. Initially, these temporal voice areas (TVAs) were thought to be *specific* to the human voice. One can argue, however, that this specialization might not be specific to the voice object itself, that is, the brain mechanisms centered on *objects*, but instead be related to the kinds of brain mechanisms necessary to encode and represent auditory dynamic information in which specific aspects are especially relevant to achieve a specific task, for example, to recognize the human voice compared with other kinds of sounds—*a functional perspective*. Another important aspect of human voice is the relevance of such information; you have to be efficient to detect and represent vocal information from your conspecifics with whom you have complex social interactions, more complex than interactions with other species present around you in the environment. Of course, through your own vocal production you have also trained your brain systems to be efficient to represent such sounds (vocalizations) in order to, for example, be able to compare what you want to produce with what you actually produced. Schirmer and colleagues (Schirmer, Fox, & Grandjean, 2012) showed that a large part of this temporal system, the voice areas, are not specific to the human voice but can also be modulated by music. In their meta-analysis, they reported that only a tiny part of TVAs show significant greater activity in response to human voices than they do in response to music. Rather than considering such brain areas to be specific to a particular object, that is, the human voice, I propose that a large part of these temporal brain areas are necessary to extract and represent specific dynamic auditory aspects that are shared between the human voice and music or other kinds of complex organized sound streams. I suggest that, instead of focusing on an auditory object, one should consider these brain areas to be selective in representing dynamic complex sound patterns with special relevance. Note that these voice-sensitive areas have also been documented in other species for pairs' vocalizations (e.g., Belin, 2006).

A corpus of empirical studies has also shown that these temporal brain regions can be modulated by the emotion perceived in the voice (Alba-Ferrara, Ellison, & Mitchell, 2012; Alba-Ferrara, Hausmann, Mitchell, & Weis, 2011; Ethofer, Anders, Wiethoff, et al., 2006; Ethofer, De Ville, Scherer, & Vuilleumier, 2009; Mitchell, Elliott, Barry, Cruttenden, & Woodruff, 2003), even if such stimuli are outside the voluntary focus of attention (Grandjean et al., 2005; Sander et al., 2005). This phenomenon is called attentional capture, which can bias information processing to relevant stimuli even if the subject is occupied with another task or information processing, as has been shown, for example, in vision (Vuilleumier, 2005). The attentional capture can be ear-dependent: different information is provided at the same time to both ears, with a task (e.g., sex or emotion discrimination) on the auditory material presented on the left or the right ear; the so-called dichotic listening task (Grandjean et al., 2005; Sander et al., 2005). But attentional capture is usually, in an ecological context, related to spatial information, for example, an angry voice appearing on your left spatial side while you are performing a task on your right spatial side. We manipulated the origin of the emotional voice (left and right sides) and its closeness by using a head-related transfer function[1] (Ceravolo, Fruhholz, & Grandjean, 2016a, 2016b): in these studies, we showed that emotional voices can attract attention even if they are located outside the voluntary spatial focus of attention, with the modulation of the TVAs as a function of distance. Researchers have also reported that the emotional intensity of happy and angry vocalizations modulates TVAs even when the blood-oxygen-level-dependent (BOLD) signal is controlled for low-level acoustic features relevant for emotional discrimination, such as the mean and variability of volume and F0 (Ethofer, Anders, Wiethoff, et al., 2006). These authors also showed that these modulations are independent of the explicit (discrimination of emotional prosody) or implicit (semantic task on the same auditory material) processing of emotion. It has moreover been reported that these regions are not the only regions to be modulated by the human emotional voice.

Subcortical Modulations Related to Emotion

Beyond the modulation by emotional voices of auditory temporal cortical areas, especially the primary auditory regions and the TVAs, a series of subcortical regions are modulated by emotion in the voice (Fruhholz et al., 2016; Sander et al., 2005). This is the case for the amygdalae (Bestelmeyer, Maurage, Rouger, Latinus, & Belin, 2014; Ethofer, Kreifelts, et al., 2009; Schirmer et al., 2008; Simon, Becker, Mothes-Lasch, Miltner, & Straube, 2017; Wiethoff, Wildgruber, Grodd, & Ethofer, 2009) and other grey nuclei,[2] especially the basal ganglia (Breitenstein, Daum, & Ackermann, 1998; Dara, Monetta, & Pell, 2008; Kluger & Heilman, 2007; Paulmann, Pell, & Kotz, 2009; Pell & Leonard, 2003; Peron et al., 2015; Peron et al., 2017; Schroder, Nikolova, & Dengler, 2010; Sidtis, Pachana, Cummings, & Sidtis, 2006). The amygdalae are considered to be crucial for biasing information at the cortical and subcortical levels, allowing organisms to react quickly to relevant information (Sander, Grafman, & Zalla, 2003) and to dedicate more resources at the cortical level to represent this relevant information in a fined-grained manner (Fruhholz & Grandjean, 2013a; Fruhholz, Hofstetter, et al., 2015; Vuilleumier, 2005). The amygdalae are also involved in the modulation of other brain areas related to the autonomic and hormonal systems, especially the hypothalamus. These peripheral activations are essential for action tendencies and behaviors. The connectivity pattern of the amygdalae is efficiently organized to modulate primary and secondary sensory regions, as well as different attentional brain networks, to prompt the decision–action system to avoid, for example, a threat behind you. We have also shown, using multivariate voxel pattern analysis, that the brain signal recorded at different levels—including the auditory thalamus, the core of the auditory region, and the belt and parabelt regions—is able to distinguish different kinds of emotional voices (Fruhholz et al., 2016). The low acoustical feature differences between different emotional or neutral voices are likely to explain these modulations at this stage of processing; it would be interesting to investigate these early responses at the subcortical and cortical levels with a paradigm which does not confound emotional aspects and low-level acoustical feature differences between neutral and emotional auditory stimuli. As stated earlier, the basal ganglia, especially the subthalamic regions, are also modulated by emotional voices compared with neutral voices (Peron, Fruhholz, Ceravolo, & Grandjean, 2016; Peron et al., 2009); through neuronal synchronization, these regions are believed to be involved in the orchestration of other subcortical and cortical regions that are involved in the perception of a specific object and related tasks or actions. The recurrence of such coupling between perception and action is related to the concept of habits and the cognitive flexibility needed, for example, in task switching or during inhibition of prepotent responses or habits (Peron, Fruhholz, Verin, & Grandjean, 2013).

Connectivity Between Subcortical Regions and Temporal Cortical Areas

The brain connectivity can be twofold: anatomical connectivity, in which research demonstrates an anatomical link between two or several brain regions, for example through white fibers, or functional connectivity. This latter refers to statistical patterns of brain activities between different and distant brain regions; this statistical patterning can be correlational between two different brain regions, that is, functional connectivity, or can be causal, that is, effective connectivity, in this latter case, the activity in one region predicts the modulations of another brain region. The anatomical and functional brain connectivity between TVAs and other cortical brain regions and subcortical regions has been documented in several studies that have shown a stronger functional or effective connectivity for emotional prosodic stimuli than for neutral stimuli between TVAs and thalamic subregions, the inferior and medial frontal regions, the supramarginal gyrus, and the basal ganglia and cerebellum (Ethofer, Anders, Erb, et al., 2006; Ethofer et al., 2012). The functional connectivity between TVAs and the amygdalae is thought to be mediated through the medial geniculate body (Pannese, Grandjean, & Fruhholz, 2015). Moreover, amygdala lesions impact on the cortical responses to emotional prosody in TVAs (Fruhholz, Hofstetter, et al., 2015). Direct anatomical connectivity between auditory areas and the amygdalae has been reported in animals, suggesting a possible functional modulation of cortical auditory processing through amygdala inputs (Reser, Burman, Richardson, Spitzer, & Rosa, 2009). The functional connectivity between the amygdalae, primary auditory regions, and TVAs is considered important for the detection of relevant auditory information. In the first stage, auditory inputs are processed by the amygdalae at a coarse level, and then the amygdalae induce cortical activity in the auditory regions to achieve a more fine-grained and in-depth analysis of the relevant auditory stimulus. As mentioned earlier, other subcortical regions of importance are the basal ganglia. They are believed to be involved in the organization of sound patterns that are especially important for rhythm (e.g., Schwartze, Keller, Patel, & Kotz, 2011; Trost et al., 2014). The dorsal regions of the basal ganglia are important in this case, especially the dorsal caudate nucleus. It is likely that the involvement of these regions, well-known in the context of motor organization (i.e., procedural learning), in rhythmic organization is related to the propensity to move in rhythm when exposed to music. Beyond this motor-rhythmic role of the dorsal parts of the basal ganglia, the two other territories, that is, the medial and ventral regions, are also involved in emotional prosody decoding. The medial parts are anatomically and functionally connected to associated cortical regions and are probably crucial in the integration of information across modalities. The ventral parts are more connected to the subcortical limbic regions, especially the amygdalae, and are involved in the emotional patterning of prosody. The anatomical and functional connectivity between the basal ganglia is organized in three different networks: the direct, indirect, and hyperdirect pathways. These pathways connect the different parts of the basal ganglia with specific cortical or subcortical regions: the dorsal parts with motor regions, the medial parts with associative regions (e.g., temporal areas), and the ventral parts with brain regions known for their involvement in emotional processing (e.g., amygdala). In animal studies, the subthalamic nucleus (STN) has been shown to be anatomically connected to frontal regions through the hyperdirect pathway in order to activate or inhibit neural networks responsible for habits related to emotions (e.g., Weintraub &

Zaghloul, 2013). We have shown a functional connection between the STN and the orbitofrontal regions in humans by using functional magnetic resonance imaging (fMRI); participants had to inhibit the prepotent response to anger prosody and to perform a sex discrimination task of the stimuli rather than an emotion task (Peron et al., 2016). The psycho-physiological interaction analysis used in this study revealed a specific functional connectivity pattern involving the STN, the orbitofrontal regions, the TVAs, and the amygdalae. We interpreted the results in light of the previously formulated hypothesis for the involvement of the basal ganglia in the genesis or inhibition of overlearned patterning between subcortical and cortical areas to achieve a specific task, such as detecting emotion in the environment in order to produce a habitual response, for example, in the fight/flight/freezing systems.

Dynamics Between Frontal Areas and TVAs

The perception of emotional prosody at the brain level is organized in a complex network that includes cross-talking between different frontal brain areas and TVAs. The first reported study of the connectivity pattern between TVAs and the frontal areas used dynamic causal modeling to document the bilateral connectivity between TVAs and the inferior frontal cortices for the processing of emotional prosody (Ethofer, Anders, Erb, et al., 2006). More recently, Frühholz and colleagues also demonstrated modulation of the connectivity between the inferior frontal cortices and different parts of TVAs for the processing of anger prosody (Fruhholz & Grandjean, 2012). From fMRI and diffusion tensor imaging studies, we proposed that the inferior frontal gyrus, together with TVAs, are involved in the ability of humans to extract statistical regularities in sound patterns for emotion categorization and discrimination at different time scales (Dricu, Ceravolo, Grandjean, & Fruhholz, 2017; Fruhholz & Grandjean, 2013b; Fruhholz, Gschwind, & Grandjean, 2015). The medial frontal areas, especially the medial and lateral orbitofrontal areas, have also been reported to be modulated by emotional prosody (Paulmann, Seifert, & Kotz, 2010; Peron et al., 2016; Peron et al., 2009; Quadflieg, Mohr, Mentzel, Miltner, & Straube, 2008; Ross & Monnot, 2011; Sander et al., 2005; Schirmer et al., 2008; Wildgruber, Ackermann, Kreifelts, & Ethofer, 2006). These frontal regions seem to be crucial in the context of ambiguous information when the impact of a specific context necessitates in-depth analysis of the relationships between the contextual information and the emotional prosodic aspects; in such situations, humans need to infer contextual values, as in the case of sarcasm or irony, for example (Kipps, Nestor, Acosta-Cabronero, Arnold, & Hodges, 2009; Matsui et al., 2016; Uchiyama et al., 2012; Uchiyama et al., 2006). The medial orbitofrontal cortex (OFC) regions have also been reported to be modulated by emotional prosody, especially anger prosody (Sander et al., 2005); it is likely that a complex pattern of connectivity between these medial parts of the OFC and the amygdalae, together with TVAs, are involved in the decoding of contextual meaning of auditory social stimuli, especially in ambiguous situations.

Timing Matters

One important feature of the auditory system is its ability to process temporal information at the millisecond level. Auditory information is always dynamic and the temporal dimension cannot be avoided in empirical studies. This is true for voice and speech, but it is also the case for all sound processing especially when slow temporal fluctuations are essential for the decoding and the recognition of the sound. This is not always the case for visual information, which can be presented statically. The complexity of the processing of unfolding information in the auditory domain is probably why scientific investigations of emotional prosody have been fewer than empirical studies in the domain of facial emotional expression. One method that has recently been used is the gating paradigm. In such a paradigm, auditory information is systematically increased through different gate durations in order to assess the ability of humans to decode unfolding auditory information as they perform, for example, an emotion categorization task. Using such a gating paradigm with five basic emotions, Pell and Kotz (2011) demonstrated that some emotional prosodies were recognized at shorter gate intervals than others, with different identification points: neutral, fear, and sadness at ~500 to 600 ms; anger at ~700 ms; happiness at ~1,000 ms; and disgust at ~1,500 ms; the latter being characterized as having the worst recognition level, as demonstrated in previous studies (Banse & Scherer, 1996). These authors also identified the critical importance of F0 mean, variation, and range in the emotion identification point. Recently, we also showed that such a paradigm can be useful in understanding the processing of unfolding information in nonverbal emotional vocalizations (Schaerlaeken & Grandjean, 2018). In that study, we showed that such vocalizations can be recognized in less than 500 ms for anger, disgust, and fear, whereas sadness, joy, and neutral recognition necessitated more information. The investigation of the impact of acoustic features on these recognition functions included parameters related to the frequency domain (F0 indicators), energy, and amplitude, as well as spectral parameters, such as harmonic differences and the Hammarberg index (a measure related to the spectral distribution of energy), which in previous studies were shown to be crucial for the decoding of such nonverbal emotional vocalizations (Sauter et al., 2010). Although the gating paradigm has been useful in the context of behavioral studies, electroencephalography (EEG) has been used to investigate the timing of brain processing in the context of emotional prosody (Grandjean et al., 2006; Grandjean, Ducommun, Bernard, & Scherer, 2002; Kotz & Paulmann, 2007; Paulmann & Kotz, 2008; Paulmann et al., 2009; Paulmann et al., 2010). Paulmann and Kotz (Kotz & Paulmann, 2007; Paulmann & Kotz, 2008) showed, using EEG at the scalp level, that emotional prosody modulated the N200 and P300, demonstrating how such emotional information can be processed at an early stage. Through the use of intracranial recordings in the STN in humans, we have also demonstrated that emotional prosody can modulate these subcortical regions at an early stage, around 200–300 ms, as well as at later stages (Peron et al., 2017).

The ability of the central nervous system to represent different time scales might be strongly related to brain fluctuations at different frequencies, as is the case for brain oscillations. For example, gamma activity (high frequencies: > 35 Hz) has been shown to be especially crucial for short-term auditory perception, while theta (low frequencies: ~4.5 to 8 Hz) is more involved in larger scales such as syllabic perception (Baltus & Herrmann, 2015; Morillon, Liegeois-Chauvel, Arnal, Benar, & Giraud, 2012; Palva et al., 2002; Shamir, Ghitza, Epstein, & Kopell, 2009). We can argue that such oscillatory systems are also crucial for emotional prosody perception, as Symons and colleagues have suggested (Symons, El-Deredy, Schwartze, & Kotz, 2016). Auditory slow fluctuations would be encoded by theta and delta (low frequencies: < 4 Hz) brain waves, while a higher time-scale resolution would be encoded at higher frequencies of brain waves, such as alpha (~8.5 to 12 Hz) and gamma bands. More studies are needed in this field to better understand how the brain integrates different time-scale information processing for representing emotional prosody stimuli. Cross-frequency coupling[3] between gamma and theta or alpha bands, for example, is a strong candidate for such a mechanism of integration for building up complex sound objects.

Lateralization

A large corpus of neurological studies involving patients with brain damage has suggested an important role of the right hemisphere in the production and perception of prosody, especially emotional prosody (Behrens, 1989; Blonder, Bowers, & Heilman, 1991; Blonder, Pickering, Heath, Smith, & Butler, 1995; Bradvik et al., 1991; Geigenberger & Ziegler, 2001; Gibson, Atchley, Voyer, Diener, & Gregersen, 2016; Grandjean, Sander, Lucas, Scherer, & Vuilleumier, 2008; Heilman, Bowers, Speedie, & Coslett, 1984; Kucharska-Pietura, Phillips, Gernand, & David, 2003; Pell, 1998, 1999; Pell & Baum, 1997a, 1997b; Ryalls, Joanette, & Feldman, 1987; Schmidt, 2004; Shapiro & Danly, 1985; Sherratt, 2007; Tompkins & Flowers, 1985; Trauner, Ballantyne, Friedland, & Chase, 1996). However, Schirmer and Kotz (2006) have proposed that even though the right hemisphere is crucial for emotional prosody decoding, the left hemisphere is especially crucial for the processing of short segmental information, which is essential in speech perception. Indeed, lateralization would be driven massively by the time scale that is relevant to the decoding of unfolding auditory information. Low fluctuations that drive the ability to integrate large-scale information, for example, the pitch dynamics of voice, would be more related to the right hemisphere, whereas drivers of the ability to discriminate short-scale information, that is, the phonemes, would preferentially recruit the left hemisphere. I propose that the left hemisphere is also important in the context of emotional prosody because short-scale information processing is important in representing roughness or other kinds of short-scale spectral fluctuations. The lateralization of emotional prosody decoding is also related to the complexity of the utterances presented to the subject: Mitchell and Ross demonstrated that the involvement of the left hemisphere was dependent on the syllabic complexity of the verbal material (Mitchell & Ross, 2008).

Summary of Functional Aspects and Neural Networks

The processing of emotional prosody can be described in five major neuronal networks subserving different functional aspects (see Figure 1). The first subcortical network, after the initial brainstem and midbrain relays, represents the ascendant flow of information involving the auditory thalamus and the core of the primary auditory regions, including the belt, the parabelt, and the amygdalae (Fruhholz & Grandjean, 2013a; Fruhholz, Hofstetter, et al., 2015; Fruhholz et al., 2016; Pannese, Grandjean, & Fruhholz, 2016; see Figure 1A) . This system is primarily related to the coarse encoding of the emotional auditory stimulus and is involved, especially for the amygdalae, in the modulation of the second system. This second system, massively related to feedforward information flow from the first subcortical network, is composed of the complex temporal cortical system that includes the TVAs (Belin, Bestelmeyer, Latinus, & Watson, 2011; Belin et al., 2000; Bestelmeyer, Rouger, DeBruine, & Belin, 2010; Grandjean et al., 2005; Ponsot, Burred, Belin, & Aucouturier, 2018). They are implemented in the middle superior temporal gyri and sulci with a gradient of sound complexity from the medial part toward the anterior parts of these regions, and a second gradient with multimodal integration toward the back of the system that includes the ascendant branch of the superior temporal gyrus and the superior temporal sulcus (Kreifelts, Ethofer, Grodd, Erb, & Wildgruber, 2007). This system is involved in the building up of the auditory object as an integrated percept and its relationships, for example, with visual information (Watson et al., 2014; see Figure 1B). Even if this second system is massively dependent on feedforward information flow, some studies have shown that it can be modulated by feedback information flow from the frontal regions, for example related to the task that the subject has to perform (Ethofer, Anders, Erb, et al., 2006; Fruhholz, Ceravolo, & Grandjean, 2012; Fruhholz & Grandjean, 2012). The third brain network, a frontal system, is closely connected with the TVA system. It is composed of the inferior frontal gyri and sulci, which are involved in the implicit or explicit categorization and discrimination processing of auditory emotional information (see Figure 1C). This system would be important in its ability to modulate the use of information extracted from the acoustical signal to perform a specific task and then would be involved in feedback information flow on the previous networks. The fourth system, which would be also important in the feedback information flow, is composed of the OFC, especially the medial part, which is closely connected to the subcortical system that is composed of the amygdalae and the cortical TVA system. This system is essential in the integration of contextual aspects for the valuation of a specific sound object, such as in an ambiguous situation (e.g., sarcasm; see Figure 1D). The fifth system is composed of subcortical nuclei, the basal ganglia, and the cerebellum, which are especially important at the functional level for the organization of sound processing (e.g.,

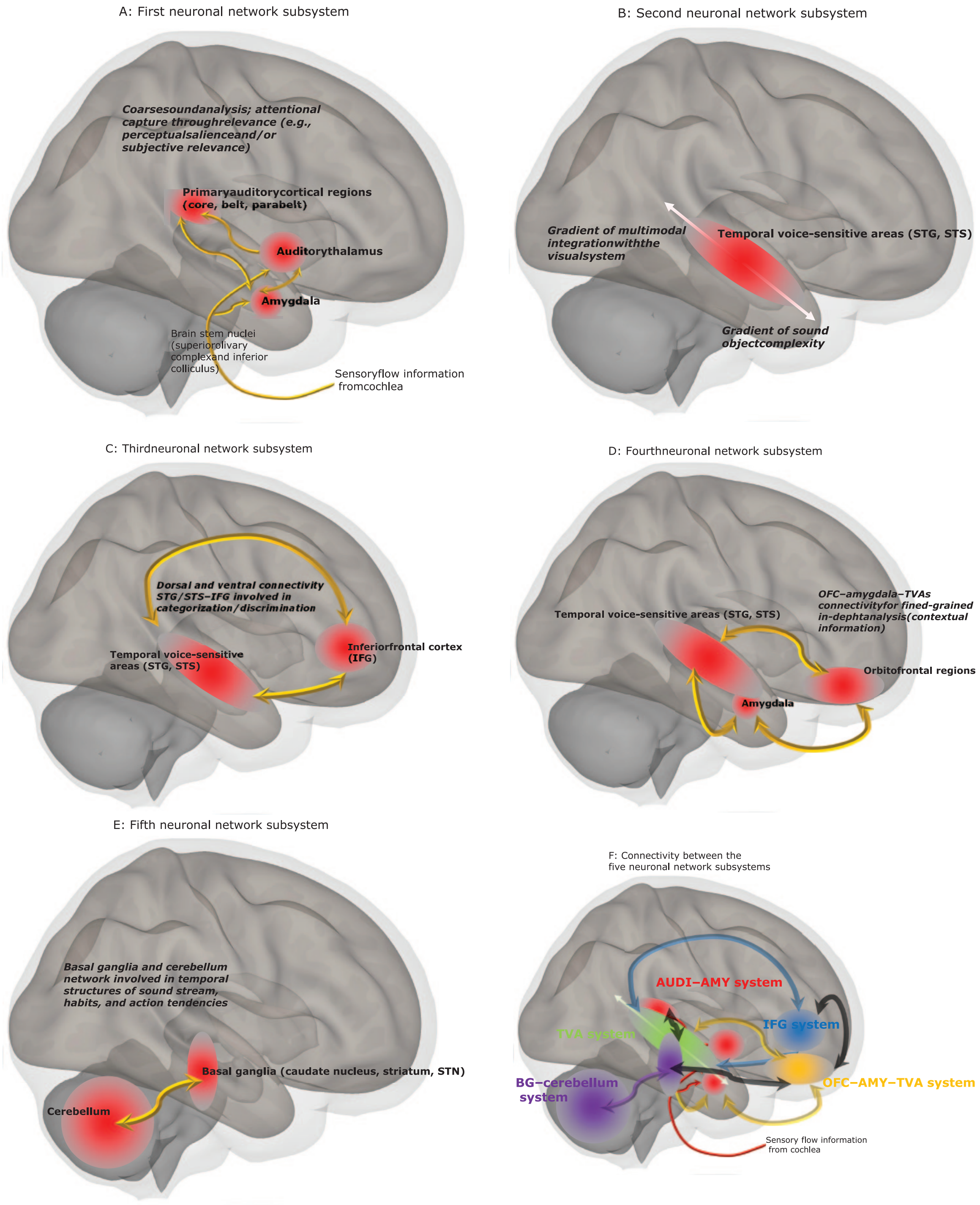

Figure 1. The five anatomic and functional brain networks involved in emotional prosody decoding: 1A: Neural network related to the first feedforward information flow from the peripheral sensory auditory system to the subcortical nuclei (e.g., auditory thalamus), the amygdala, and the cortical primary auditory regions involved in coarse sound analysis and relevance. 1B: Neural network including the different subparts of the temporal voice areas (TVAs) with two gradients: sound complexity and multimodal visual integration. 1C: Inferior frontal cortical neural network with both dorsal and temporal functional connectivity with the TVAs. 1D: Orbitofrontal neural network functionally connected with amygdala and TVAs regions for fine-grained depth sound analysis. 1E: Basal ganglia and cerebellum network involved in unfolding auditory stream organization, habits, and action tendencies related to emotional prosody processing. 1F: Connectivity between the five anatomically and functionally distinct neural networks (black arrows) between the OFC–AMY–TVAs and IFG–TVAs networks, OFC–AMY–TVAs and BG–cerebellum networks, and the AUDI–AMY and the TVA networks.

Note. AMY = amygdala; AUDI: auditory; BG: basal ganglia; IFG: inferior frontal gyrus; OFC: orbitofrontal area; STG: superior temporal gyrus; STN: subthalamic nucleus; STS: superior temporal sulcus; TVA: temporal voice area.

rhythm) and temporal predictions; it would be important in both feedforward and feedback information flows. This system is also involved in the formation of habits in the context of auditory emotion perception and its relationships with action tendencies and behaviors. Moreover, it is important in the inhibition of prepotent responses in the context of auditory information processing (Peron et al., 2016; see Figure 1F). These systems are involved bilaterally in the processing of emotional prosody, but lateralization depends on the time scale necessary to represent the relevant information and/or to achieve a specific goal or task (implicitly or explicitly).

Conclusions and Perspectives

We have defined a model that includes five different neural and functional systems involved in the decoding of emotional prosody. These neural systems are involved in different functional subprocesses (see previous lines) and are not totally independent from one another, as they exchange information at the organism level, but they constitute highly connected subsystems that are anatomically and functionally segregated. The brain lateralization of these different systems is largely explained by the timescale resolution necessary to achieve a specific perception, representation, or task or goal. In future studies, the differential contribution of the brain waves and oscillatory systems at different frequencies should be addressed in order to build a computational model of the perception of emotional prosody. The formal comparison between the neural networks involved in the decoding of emotional prosody and other kinds of emotional vocalizations, like affective bursts, might be also the purpose of a specific series of experiments; for example, the question of lateralization in this context would be very interesting: one can argue that temporal lobe modulations would be more lateralized on the left side for the processing of emotional prosody compared to other kinds of emotional vocalizations. The involvement of the inferior frontal regions might be also more distributed, with bilateral activations, during emotional prosody decoding compared to other kinds of emotional vocalizations. It would also be interesting to approach the phenomenon of emotional decoding of verbal and nonverbal auditory information from a phylogenetic perspective with an investigation of vocalization processing in apes and monkeys, especially for distress, agonistic, or affiliative components (Gruber & Grandjean, 2017). For example, to what extent humans are able to infer emotional content from nonhuman primate vocalizations or even in mammalian or nonmammalian animals? Actually, it has been shown that humans are able to infer the level of arousal in a large variety of animal vocalizations (Filippi et al., 2017), suggesting a phylogenetic continuity in the decoding of arousal in auditory signals produced by vertebrates. It would be also interesting to quantify the importance of phylogenetic proximity versus acoustical proximity factors in the recognition of emotional vocalizations in nonhuman primates. Fritz and collaborators (Fritz et al., 2018) showed that vocalizations produced by chimpanzees induced an increase of the superior temporal gyri in humans compared to macaques' vocalizations. Furthermore, an ontogenetic approach to human and nonhuman primates would also shed light on the mechanism of the origins of emotional voices at the behavioral and brain levels, at both perception and production levels. Indeed, the production of the phenomenon of emotional vocalizations should also be addressed systematically in behavioral, peripheral, and neuroimaging studies in order to better understand its components. As stated by embodied cognition theory (e.g., Gibbs, 2006; Harrison, Gray, Gianaros, & Critchley, 2010; Niedenthal, 2007; Niedenthal, Mondillon, Winkielman, & Vermeulen, 2009; Thompson & Varela, 2001) and by earlier works done by Piaget (e.g., Piaget, 1950, 1953), with a developmental perspective, it is likely that production and perception are closely intertwined; then, research on the production side of emotional prosody might provide new ways to better understand the mechanisms involved in the perception and recognition of emotional prosody.

Declaration of Conflicting Interests

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

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Didier Grandjean https://orcid.org/0000-0001-6125-4520

Notes

1 Head-related transfer function is a method able to temporally shift the auditory information of few milliseconds between the two ears in order to create an illusion of a lateralized sound. For example, when a sound comes from your right side, the information reaches the right ear before the left ear, note that the sound pressure is also less strong on the left side compared the right side in such situation.

2 The term grey nuclei refers to all group of neurons anatomically organized in a specific nucleus such as amygdalae and basal ganglia.

3 Cross-frequency coupling refers to the fact that, for example, gamma bursts can be temporally organized at the phase of slower frequency bands such as alpha or theta bands (see Jensen & Colgin, 2007, for a review).

References

Alba-Ferrara, L., Ellison, A., & Mitchell, R. L. C. (2012). Decoding emotional prosody: Resolving differences in functional neuroanatomy from fMRI and lesion studies using TMS. *Brain Stimulation*, *5*(3), 347–353.

Alba-Ferrara, L., Hausmann, M., Mitchell, R. L., & Weis, S. (2011). The neural correlates of emotional prosody comprehension: Disentangling simple from complex emotion. *PLoS ONE*, *6*(12). https://doi.org/10.1371/journal.pone.0028701

Baltus, A., & Herrmann, C. S. (2015). Auditory temporal resolution is linked to resonance frequency of the auditory cortex. *International Journal of Psychophysiology*, *98*(1), 1–7.

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, *70*(3), 614–636.

Bänziger, T., Patel, S., & Scherer, K. R. (2014). The role of perceived voice and speech characteristics in vocal emotion communication. *Journal of Nonverbal Behavior*, *38*(1), 31–52.

Bänziger, T., & Scherer, K. R. (2005). The role of intonation in emotional expressions. *Speech Communication*, *46*(3–4), 252–267.

Behrens, S. J. (1989). Characterizing sentence intonation in a right hemisphere-damaged population. *Brain and Language*, *37*(2), 181–200.

Belin, P. (2006). Voice processing in human and non-human primates. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *361*(1476), 2091–2107.

Belin, P., Bestelmeyer, P. E. G., Latinus, M., & Watson, R. (2011). Understanding voice perception. *British Journal of Psychology*, *102*, 711–725.

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, *403*(6767), 309–312.

Bestelmeyer, P. E., Maurage, P., Rouger, J., Latinus, M., & Belin, P. (2014). Adaptation to vocal expressions reveals multistep perception of auditory emotion. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *34*(24), 8098–8105.

Bestelmeyer, P. E., Rouger, J., DeBruine, L. M., & Belin, P. (2010). Auditory adaptation in vocal affect perception. *Cognition*, *117*(2), 217–223.

Blonder, L. X., Bowers, D., & Heilman, K. M. (1991). The role of the right hemisphere in emotional communication. *Brain*, *114*(2, Pt. 3), 1115–1127.

Blonder, L. X., Pickering, J. E., Heath, R. L., Smith, C. D., & Butler, S. M. (1995). Prosodic characteristics of speech pre- and post-right hemisphere stroke. *Brain and Language*, *51*(2), 318–335.

Bradvik, B., Dravins, C., Holtas, S., Rosen, I., Ryding, E., & Ingvar, D. H. (1991). Disturbances of speech prosody following right hemisphere infarcts. *Acta Neurologica Scandinavica*, *84*(2), 114–126.

Breitenstein, C., Daum, I., & Ackermann, H. (1998). Emotional processing following cortical and subcortical brain damage: Contribution of the fronto-striatal circuitry. *Behavioural Neurology*, *11*(1), 29–42.

Ceravolo, L., Fruhholz, S., & Grandjean, D. (2016a). Modulation of auditory spatial attention by angry prosody: An fMRI auditory dot-probe study. *Frontiers in Neuroscience*, *10*. http://doi.org/10.3389/fnins.2016.00216

Ceravolo, L., Fruhholz, S., & Grandjean, D. (2016b). Proximal vocal threat recruits the right voice-sensitive auditory cortex. *Social Cognitive and Affective Neuroscience*, *11*(5), 793–802.

Dara, C., Monetta, L., & Pell, M. D. (2008). Vocal emotion processing in Parkinson's disease: Reduced sensitivity to negative emotions. *Brain Research*, *1188*, 100–111.

Dricu, M., Ceravolo, L., Grandjean, D., & Fruhholz, S. (2017). Biased and unbiased perceptual decision-making on vocal emotions. *Scientific Reports*, *7*(1). http://doi.org/10.1038/s41598-017-16594-w

Ethofer, T., Anders, S., Erb, M., Herbert, C., Wiethoff, S., Kissler, J., . . . Wildgruber, D. (2006). Cerebral pathways in processing of affective prosody: A dynamic causal modeling study. *NeuroImage*, *30*(2), 580–587.

Ethofer, T., Anders, S., Wiethoff, S., Erb, M., Herbert, C., Saur, R., . . . Wildgruber, D. (2006). Effects of prosodic emotional intensity on activation of associative auditory cortex. *Neuroreport*, *17*(3), 249–253.

Ethofer, T., Bretscher, J., Gschwind, M., Kreifelts, B., Wildgruber, D., & Vuilleumier, P. (2012). Emotional voice areas: Anatomic location, functional properties, and structural connections revealed by combined fMRI/DTI. *Cerebral Cortex*, *22*(1), 191–200.

Ethofer, T., De Ville, D. V., Scherer, K., & Vuilleumier, P. (2009). Decoding of emotional information in voice-sensitive cortices. *Current Biology*, *19*(12), 1028–1033.

Ethofer, T., Kreifelts, B., Wiethoff, S., Wolf, J., Grodd, W., Vuilleumier, P., & Wildgruber, D. (2009). Differential influences of emotion, task, and novelty on brain regions underlying the processing of speech melody. *Journal of Cognitive Neuroscience*, *21*(7), 1255–1268.

Filippi, P., Congdon, J. V., Hoang, J., Bowling, D. L., Reber, S. A., Pasukonis, A., . . .Gunturkun, O. (2017). Humans recognize emotional arousal in vocalizations across all classes of terrestrial vertebrates: Evidence for acoustic universals. *Proceedings. Biological Sciences*, *284*(1859). http://doi.org/10.1098/rspb.2017.0990

Fries, P. (2005). A mechanism for cognitive dynamics: Neuronal communication through neuronal coherence. *Trends in Cognitive Sciences*, *9*(10), 474–480.

Fritz, T., Mueller, K., Guha, A., Gouws, A., Levita, L., Andrews, T. J., & Slocombe, K. E. (2018). Human behavioural discrimination of human, chimpanzee and macaque affective vocalisations is reflected by the neural response in the superior temporal sulcus. *Neuropsychologia*, *111*, 145–150.

Fruhholz, S., Ceravolo, L., & Grandjean, D. (2012). Specific brain networks during explicit and implicit decoding of emotional prosody. *Cerebral Cortex*, *22*(5), 1107–1117.

Fruhholz, S., & Grandjean, D. (2012). Towards a fronto-temporal neural network for the decoding of angry vocal expressions. *NeuroImage*, *62*(3), 1658–1666.

Fruhholz, S., & Grandjean, D. (2013a). Amygdala subregions differentially respond and rapidly adapt to threatening voices. *Cortex*, *49*(5), 1394–1403.

Fruhholz, S., & Grandjean, D. (2013b). Processing of emotional vocalizations in bilateral inferior frontal cortex. *Neuroscience and Biobehavioral Reviews*, *37*(10, Pt. 2), 2847–2855.

Fruhholz, S., Gschwind, M., & Grandjean, D. (2015). Bilateral dorsal and ventral fiber pathways for the processing of affective prosody identified by probabilistic fiber tracking. *NeuroImage*, *109*, 27–34.

Fruhholz, S., Hofstetter, C., Cristinzio, C., Saj, A., Seeck, M., Vuilleumier, P., & Grandjean, D. (2015). Asymmetrical effects of unilateral right or left amygdala damage on auditory cortical processing of vocal emotions. *Proceedings of the National Academy of Sciences of the USA*, *112*(5), 1583–1588.

Fruhholz, S., van der Zwaag, W., Saenz, M., Belin, P., Schobert, A. K., Vuilleumier, P., & Grandjean, D. (2016). Neural decoding of discriminative auditory object features depends on their socio-affective valence. *Social Cognitive and Affective Neuroscience*, *11*(10), 1638–1649.

Geigenberger, A., & Ziegler, W. (2001). The processing of prosodic patterns in patients with unilateral brain lesions. *Clinical Linguistics & Phonetics*, *15*(1–2), 85–89.

Gibbs, R. W. (2006). *Embodiment and cognitive science*. New York, NY: Cambridge University Press.

Gibson, L., Atchley, R. A., Voyer, D., Diener, U. S., & Gregersen, S. (2016). Detection of sarcastic speech: The role of the right hemisphere in ambiguity resolution. *Laterality*, *21*(4–6), 549–567.

Grandjean, D., Bänziger, T., & Scherer, K. R. (2006). Intonation as an interface between language and affect. *Progress in Brain Research*, *156*, 235–247.

Grandjean, D., Ducommun, C., Bernard, P. J., & Scherer, K. (2002). Comparison of cerebral activation patterns in identifying affective prosody, semantic prosody, and phoneme differences. *International Journal of Psychophysiology*, *45*(1–2), 147–148.

Grandjean, D., Sander, D., Lucas, N., Scherer, K. R., & Vuilleumier, P. (2008). Effects of emotional prosody on auditory extinction for voices in patients with spatial neglect. *Neuropsychologia*, *46*(2), 487–496.

Grandjean, D., Sander, D., Pourtois, G., Schwartz, S., Seghier, M. L., Scherer, K. R., & Vuilleumier, P. (2005). The voices of wrath: Brain responses to angry prosody in meaningless speech. *Nature Neuroscience*, *8*(2), 145–146.

Gruber, T., & Grandjean, D. (2017). A comparative neurological approach to emotional expressions in primate vocalizations. *Neuroscience and Biobehavioral Reviews*, *73*, 182–190.

Harrison, N. A., Gray, M. A., Gianaros, P. J., & Critchley, H. D. (2010). The embodiment of emotional feelings in the brain. *Journal of Neuroscience*, *30*(38), 12878–12884.

Heilman, K. M., Bowers, D., Speedie, L., & Coslett, H. B. (1984). Comprehension of affective and nonaffective prosody. *Neurology*, *34*(7), 917–921.

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

Jensen, O., & Colgin, L. L. (2007). Cross-frequency coupling between neuronal oscillations. *Trends in Cognitive Sciences*, *11*(7), 267–269.

Kipps, C. M., Nestor, P. J., Acosta-Cabronero, J., Arnold, R., & Hodges, J. R. (2009). Understanding social dysfunction in the behavioural variant of frontotemporal dementia: The role of emotion and sarcasm processing. *Brain*, *132*(Pt. 3), 592–603.

Kluger, B. M., & Heilman, K. M. (2007). Dysfunctional facial emotional expression and comprehension in a patient with corticobasal degeneration. *Neurocase*, *13*(3), 165–168.

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

Kreifelts, B., Ethofer, T., Grodd, W., Erb, M., & Wildgruber, D. (2007). Audiovisual integration of emotional signals in voice and face: An event-related fMRI study. *NeuroImage*, *37*(4), 1445–1456.

Kucharska-Pietura, K., Phillips, M. L., Gernand, W., & David, A. S. (2003). Perception of emotions from faces and voices following unilateral brain damage. *Neuropsychologia*, *41*(8), 1082–1090.

Matsui, T., Nakamura, T., Utsumi, A., Sasaki, A. T., Koike, T., Yoshida, Y., . . .Sadato, N. (2016). The role of prosody and context in sarcasm comprehension: Behavioral and fMRI evidence. *Neuropsychologia*, *87*, 74–84.

Mitchell, R. L., Elliott, R., Barry, M., Cruttenden, A., & Woodruff, P. W. (2003). The neural response to emotional prosody, as revealed by functional magnetic resonance imaging. *Neuropsychologia*, *41*(10), 1410–1421.

Mitchell, R. L., & Ross, E. D. (2008). fMRI evidence for the effect of verbal complexity on lateralisation of the neural response associated with decoding prosodic emotion. *Neuropsychologia*, *46*(12), 2880–2887.

Morillon, B., Liegeois-Chauvel, C., Arnal, L. H., Benar, C. G., & Giraud, A. L. (2012). Asymmetric function of theta and gamma activity in syllable processing: An intra-cortical study. *Frontiers in Psychology*, *3*. http://doi.org/10.3389/fpsyg.2012.00248

Niedenthal, P. M. (2007). Embodying emotion. *Science*, *316*(5827), 1002–1005.

Niedenthal, P. M., Mondillon, L., Winkielman, P., & Vermeulen, N. (2009). Embodiment of emotion concepts. *Journal of Personality and Social Psychology*, *96*(6), 1120–1136.

Palva, S., Palva, J. M., Shtyrov, Y., Kujala, T., Ilmoniemi, R. J., Kaila, K., & Naatanen, R. (2002). Distinct gamma-band evoked responses to speech and non-speech sounds in humans. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *22*(4). http://doi.org/10.1523/JNEUROSCI.22-04-j0003.2002

Pannese, A., Grandjean, D., & Fruhholz, S. (2015). Subcortical processing in auditory communication. *Hearing Research*, *328*, 67–77.

Pannese, A., Grandjean, D., & Fruhholz, S. (2016). Amygdala and auditory cortex exhibit distinct sensitivity to relevant acoustic features of auditory emotions. *Cortex*, *85*, 116–125.

Paulmann, S., & Kotz, S. A. (2008). An ERP investigation on the temporal dynamics of emotional prosody and emotional semantics in pseudo- and lexical-sentence context. *Brain and Language*, *105*(1), 59–69.

Paulmann, S., Pell, M. D., & Kotz, S. A. (2009). Comparative processing of emotional prosody and semantics following basal ganglia infarcts: ERP evidence of selective impairments for disgust and fear. *Brain Research*, *1295*, 159–169.

Paulmann, S., Seifert, S., & Kotz, S. A. (2010). Orbito-frontal lesions cause impairment during late but not early emotional prosodic processing. *Social Neuroscience*, *5*(1), 59–75.

Pell, M. D. (1998). Recognition of prosody following unilateral brain lesion: Influence of functional and structural attributes of prosodic contours. *Neuropsychologia*, *36*(8), 701–715.

Pell, M. D. (1999). Fundamental frequency encoding of linguistic and emotional prosody by right hemisphere-damaged speakers. *Brain and Language*, *69*(2), 161–192.

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

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

Pell, M. D., & Kotz, S. A. (2011). On the time course of vocal emotion recognition. *PLoS ONE*, *6*(11). https://doi.org/10.1371/journal.pone.0027256

Pell, M. D., & Leonard, C. L. (2003). Processing emotional tone from speech in Parkinson's disease: A role for the basal ganglia. *Cognitive, Affective and Behavioral Neuroscience*, *3*(4), 275–288.

Peron, J., Cekic, S., Haegelen, C., Sauleau, P., Patel, S., Drapier, D., . . . Grandjean, D. (2015). Sensory contribution to vocal emotion deficit in Parkinson's disease after subthalamic stimulation. *Cortex*, *63*, 172–183.

Peron, J., Fruhholz, S., Ceravolo, L., & Grandjean, D. (2016). Structural and functional connectivity of the subthalamic nucleus during vocal emotion decoding. *Social Cognitive and Affective Neuroscience*, *11*(2), 349–356.

Peron, J., Fruhholz, S., Verin, M., & Grandjean, D. (2013). Subthalamic nucleus: A key structure for emotional component synchronization in humans. *Neuroscience and Biobehavioral Reviews*, *37*(3), 358–373.

Peron, J., Grandjean, D., Le Jeune, F., Sauleau, P., Haegelen, C., Drapier, D., . . .Verin, M. (2009). Recognition of emotional prosody is altered after subthalamic nucleus deep brain stimulation in Parkinson's disease. *Neuropsychologia*. Advance online publication. http://doi.org/10.1016/j.neuropsychologia.2009.12.003

Peron, J., Renaud, O., Haegelen, C., Tamarit, L., Milesi, V., Houvenaghel, J. F., . . .Grandjean, D. (2017). Vocal emotion decoding in the subthalamic nucleus: An intracranial ERP study in Parkinson's disease. *Brain and Language*, *168*, 1–11.

Piaget, J. (1950). *La construction du réel chez l'enfant* [The construction of reality in children] (2nd ed.). Neuchâtel, Switzerland: Delachaux et Niestlé.

Piaget, J. (1953). *The origin of intelligence in the child*. London, UK: Routledge & Paul.

Ponsot, E., Burred, J. J., Belin, P., & Aucouturier, J. J. (2018). Cracking the social code of speech prosody using reverse correlation. *Proceedings of the National Academy of Sciences of the USA*, *115*(15), 3972–3977.

Quadflieg, S., Mohr, A., Mentzel, H. J., Miltner, W. H., & Straube, T. (2008). Modulation of the neural network involved in the processing of anger prosody: The role of task-relevance and social phobia. *Biological Psychology*, *78*(2), 129–137.

Reser, D. H., Burman, K. J., Richardson, K. E., Spitzer, M. W., & Rosa, M. G. (2009). Connections of the marmoset rostrotemporal auditory area: Express pathways for analysis of affective content in hearing. *The European Journal of Neuroscience*, *30*(4), 578–592.

Ross, E. D., & Monnot, M. (2011). Affective prosody: What do comprehension errors tell us about hemispheric lateralization of emotions, sex and aging effects, and the role of cognitive appraisal. *Neuropsychologia*, *49*(5), 866–877.

Ryalls, J., Joanette, Y., & Feldman, L. (1987). An acoustic comparison of normal and right-hemisphere-damaged speech prosody. *Cortex*, *23*(4), 685–694.

Sander, D., Grafman, J., & Zalla, T. (2003). The human amygdala: An evolved system for relevance detection. *Reviews in the Neurosciences*, *14*(4), 303–316.

Sander, D., Grandjean, D., Pourtois, G., Schwartz, S., Seghier, M. L., Scherer, K. R., & Vuilleumier, P. (2005). Emotion and attention interactions in social cognition: Brain regions involved in processing anger prosody. *NeuroImage*, *28*(4), 848–858.

Sauter, D. A., Eisner, F., Calder, A. J., & Scott, S. K. (2010). Perceptual cues in nonverbal vocal expressions of emotion. *Quarterly Journal of Experimental Psychology (Hove)*, *63*(11), 2251–2272.

Schaerlaeken, S., & Grandjean, D. (2018). Unfolding and dynamics of affect bursts decoding in humans. *PLoS ONE*, *13*(10). https://doi.org/10.1371/journal.pone.0206216

Scherer, K. R. (1995). Expression of emotion in voice and music. *Journal of Voice*, *9*(3), 235–248.

Scherer, K. R., Johnstone, T., & Klasmeyer, G. (2003). Vocal expression of emotion. In R. J. Davidson, H. H. Goldsmith & K. Scherer (Eds.), *Handbook of the affective sciences* (pp. 433–456). Oxford, UK: Oxford University Press.

Schirmer, A., Escoffier, N., Zysset, S., Koester, D., Striano, T., & Friederici, A. D. (2008). When vocal processing gets emotional: On the role of social orientation in relevance detection by the human amygdala. *NeuroImage*, *40*(3), 1402–1410.

Schirmer, A., Fox, P. M., & Grandjean, D. (2012). On the spatial organization of sound processing in the human temporal lobe: A meta-analysis. *NeuroImage*, *63*(1), 137–147.

Schirmer, A., & Kotz, S. A. (2006). Beyond the right hemisphere: Brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences*, *10*(1), 24–30.

Schmidt, J. M. (2004). Emotional prosody production in brain-damaged populations: Human judgments and acoustical analysis. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, *64*(8, Pt. B), 4102.

Schroder, C., Nikolova, Z. T., & Dengler, R. (2010). Changes of emotional prosody in Parkinson's disease. *Journal of the Neurological Sciences*, *289*(1–2), 32–35.

Schwartze, M., Keller, P. E., Patel, A. D., & Kotz, S. A. (2011). The impact of basal ganglia lesions on sensorimotor synchronization, spontaneous motor tempo, and the detection of tempo changes. *Behavioral Brain Research*, *216*(2), 685–691.

Shamir, M., Ghitza, O., Epstein, S., & Kopell, N. (2009). Representation of time-varying stimuli by a network exhibiting oscillations on a faster time scale. *PLoS Computational Biology*, *5*(5). https://doi.org/10.1371/journal.pcbi.1000370

Shapiro, B. E., & Danly, M. (1985). The role of the right hemisphere in the control of speech prosody in propositional and affective contexts. *Brain and Language*, *25*(1), 19–36.

Sherratt, S. (2007). Right brain damage and the verbal expression of emotion: A preliminary investigation. *Aphasiology*, *21*(3–4), 320–339.

Sidtis, D. V., Pachana, N., Cummings, J. L., & Sidtis, J. J. (2006). Dysprosodic speech following basal ganglia insult: Toward a conceptual framework for the study of the cerebral representation of prosody. *Brain and Language*, *97*(2), 135–153.

Simon, D., Becker, M., Mothes-Lasch, M., Miltner, W. H., & Straube, T. (2017). Loud and angry: Sound intensity modulates amygdala activation to angry voices in social anxiety disorder. *Social Cognitive and Affective Neuroscience*, *12*(3), 409–416.

Symons, A. E., El-Deredy, W., Schwartze, M., & Kotz, S. A. (2016). The functional role of neural oscillations in non-verbal emotional communication. *Frontiers in Human Neuroscience*, *10*. https://doi.org/10.3389/fnhum.2016.00239

Thompson, E., & Varela, F. J. (2001). Radical embodiment: Neural dynamics and consciousness. *Trends in Cognitive Sciences*, *5*(10), 418–425.

Tompkins, C. A., & Flowers, C. R. (1985). Perception of emotional intonation by brain-damaged adults: The influence of task processing levels. *Journal of Speech and Hearing Research*, *28*(4), 527–538.

Trauner, D. A., Ballantyne, A., Friedland, S., & Chase, C. (1996). Disorders of affective and linguistic prosody in children after early unilateral brain damage. *Annals of Neurology*, *39*(3), 361–367.

Trost, W., Fruhholz, S., Schon, D., Labbe, C., Pichon, S., Grandjean, D., & Vuilleumier, P. (2014). Getting the beat: Entrainment of brain activity by musical rhythm and pleasantness. *NeuroImage*, *103*, 55–64.

Uchiyama, H. T., Saito, D. N., Tanabe, H. C., Harada, T., Seki, A., Ohno, K., . . .Sadato, N. (2012). Distinction between the literal and intended meanings of sentences: A functional magnetic resonance imaging study of metaphor and sarcasm. *Cortex*, *48*(5), 563–583.

Uchiyama, H. T., Seki, A., Kageyama, H., Saito, D. N., Koeda, T., Ohno, K., & Sadato, N. (2006). Neural substrates of sarcasm: A functional magnetic-resonance imaging study. *Brain Research*, *1124*(1), 100–110.

Vuilleumier, P. (2005). How brains beware: Neural mechanisms of emotional attention. *Trends in Cognitive Sciences*, *9*(12), 585–594.

Wagner, M., & Watson, D. G. (2010). Experimental and theoretical advances in prosody: A review. *Language and Cognitive Processes*, *25*(7–9), 905–945.

Watson, R., Latinus, M., Noguchi, T., Garrod, O., Crabbe, F., & Belin, P. (2014). Crossmodal adaptation in right posterior superior temporal sulcus during face–voice emotional integration. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *34*(20), 6813–6821.

Weintraub, D. B., & Zaghloul, K. A. (2013). The role of the subthalamic nucleus in cognition. *Reviews in the Neurosciences*, *24*(2), 125–138.

Wiethoff, S., Wildgruber, D., Grodd, W., & Ethofer, T. (2009). Response and habituation of the amygdala during processing of emotional prosody. *Neuroreport*, *20*(15), 1356–1360.

Wildgruber, D., Ackermann, H., Kreifelts, B., & Ethofer, T. (2006). Cerebral processing of linguistic and emotional prosody: fMRI studies. *Progress in Brain Research*, *156*, 249–268.

Wilson, D., & Wharton, T. (2006). Relevance and prosody. *Journal of Pragmatics*, *38*(10), 1559–1579.