Chapter 5
Brain Mechanisms in Emotional Voice Production and Perception and Early Life Interactions

Didier Grandjean

Abstract *Introduction*. The understanding of social human interactions through vocalizations, especially at an early stage of development, necessitates characterization of the brain mechanisms that produce the mother's emotional vocalizations and of perception, that is, how the brain is able to perceive emotion in the mother's vocalizations.

Main aims. In this chapter, we review how emotions can impact on voice production during social interactions, as described in peripheral and central neurophysiological studies. We discuss how emotions are characterized by peripheral reactions modulating the vocal tractus, thus influencing the way people vocalize with others. The perception of emotion in the voice has been extensively studied in adult populations, and the neuronal networks involved in the different stages of emotion perception, categorization, and valuation are discussed in this chapter.

Conclusions. On the basis of empirical evidence in adults and infants, as well as in newborns, the mechanisms of emotional perception in the voice and its development are discussed and extended to an early stage of development, including in premature newborns.

How Emotion Affects Vocal Production

The main goal of this chapter is to present an integrated framework from the production of emotion in the voice to its perception, including how emotion can impact on vocalizations; how these vocalizations, emotionally connoted, are processed at the brain level by the auditory system; and how they affect interpersonal communication. Of course, this dynamic view of human interaction through the auditory

D. Grandjean (✉)
Department of Psychology and Educational Sciences and Swiss Center for Affective Sciences, University of Geneva, Geneva, Switzerland
e-mail: didier.grandjean@unige.ch

M. Filippa et al. (eds.), *Early Vocal Contact and Preterm Infant Brain Development*, DOI 10.1007/978-3-319-65077-7_5

sensory system is not restricted to adults but evolves from the early stages beginning in pregnancy to later stages of development. For example, mothers or fathers talking to their newborns produce what is called infant-directed speech or baby talk, exaggerating the prosodic aspects in order to help segmentation or recruit the attention of the newborn. Like the other interactive sensory systems, such as vision or a sense of touch, that allow humans to develop an ensemble of representations, the auditory domain is crucial for emotional and social development.

Emotions can be defined as massive temporally organized changes (e.g., through synchronization) of the five different organismic components (Grandjean, Sander, & Scherer, 2008; Sander, Grandjean, & Scherer, 2005; Scherer, 1984a, 1984b): (i) the cognitive or appraisal component through modulations in the central nervous system that allow the organism to process information and integrate it in a series of representations or as a unified representation – whether consciously accessible or not (Grandjean et al., 2008); (ii) the autonomic component, which is functionally related to homeostasis; (iii) the expressive component (facial, vocal, gestural, and postural aspects); (iv) the motivational component, especially related to action tendencies and performed actions; and finally (v) the so-called feeling component. The latter component has been conceptualized as an integrated representation or a series of representations that are potentially consciously accessible and can be verbally expressed (Grandjean et al., 2008). Feelings can also be a way to modulate or regulate emotions, for example, by controlling one or several components (e.g., controlling facial expression in a specific social context) or by reevaluating events, that is, what researchers have defined as reappraisal (e.g., reevaluating a specific context). Reappraisal is thought to be characterized as a different way of appraising events, for example, changing the interpretation or the agency of a negative event. If your newborn has a crisis of rage, you can interpret it as "his or her will to manipulate me to obtain something" or "he or she is suffering and needs something"; these different ways of understanding and conceptualizing the same event induce very different emotions (K. R. Scherer, Dan, & Flykt, 2006; Smith & Ellsworth, 1985) and related actions.

When you are exposed to a difficult situation, the way you appraise it affects the different components related to emotions mentioned earlier (van Reekum et al., 2004). The domain of vocalization is directly related to the expressive component. When you express vocalizations, different organismic systems are involved. The respiratory, phonatory, and articulatory systems are the main systems (Fig. 5.1) that can be modulated by emotions in the vocal domain, which in turn modify the vocalizations produced (e.g., P. J. Davis, Zhang, Winkworth, & Bandler, 1996). Typically, the intensity, that is, the loudness, of your vocalization is strongly influenced by the quantity of air that you are able to expel through your vocal tract. The vibration of the vocal chords is directly related to phonation. Finally, all the muscles in the lower part of your face and throat allow you to articulate and then organize the bursts of air that you expel from your body. All of these systems can be affected by emotions, especially through the autonomic system, which can affect, for example, the phonatory system and your ability to control the air that you expel. As an example, in stressful situations, humans have the tendency to breathe superficially and more rapidly compared with how they breathe in a relaxed state (e.g., Boiten, 1998; Butler, Wilhelm, & Gross, 2006).

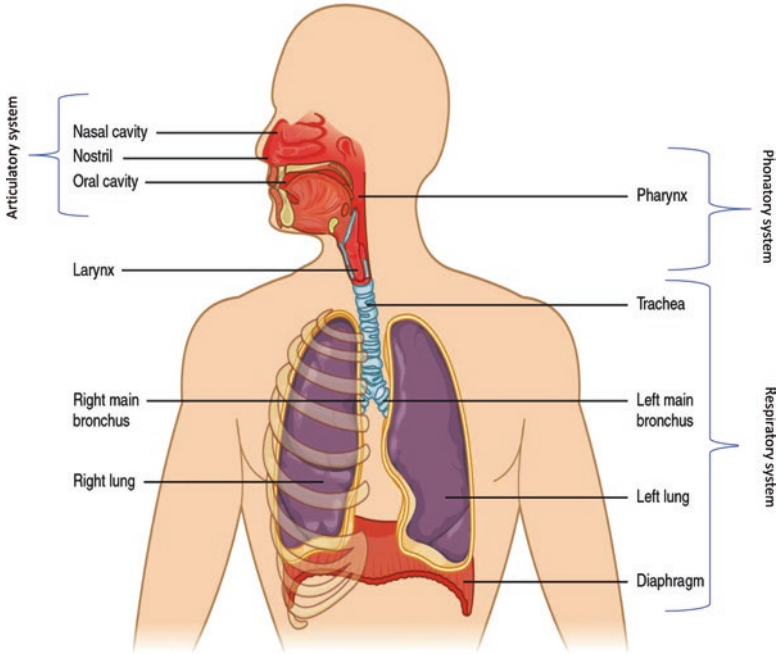

Fig. 5.1 Three main body systems are involved in vocalization: (i) the respiratory system, (ii) the phonatory system, and (iii) the articulatory system

Mechanisms Involved in the Production of Affective Vocalizations

The production of emotional vocalizations is a complex phenomenon involving a series of mechanisms at different levels, including the peripheral and central systems. These components of vocalization production can be described in at least three different interacting peripheral systems: (i) the respiratory system, mainly related to the energy of production through the level of subglottal air pressure; (ii) the phonation system, defined as how and at which frequency the vocal chords vibrate and structure the airflow, mainly related to vocal frequency, type of phonation, and register of the voice; and (iii) the articulatory system, which contributes to the shape and variance of the vocal tract (e.g., length, volume), impacting on voice quality and resonance (e.g., organization of the formants) (Banse & Scherer, 1996); see Fig. 5.1. All of these peripheral systems can be modulated by direct or indirect influences of the sympathetic and parasympathetic systems, which are strongly modulated by emotional processes. More specifically, respiratory characteristics have been shown to be affected by emotion (e.g., Etzel, Johnsen, Dickerson, Tranel, & Adolphs, 2006). The phonatory and articulatory systems can also be modulated by the emotional state of the speaker. In cats, for example, as shown from periaqueductal gray stimulations, every specific call or vocalization induces complex

specific muscle patterning, allowing the animal to produce specific prototypical vocalizations (Subramanian, Arun, Silburn, & Holstege, 2016). During an emotional episode, all of these subsystems are influenced by the impact of brain regions known to be crucial in emotion, such as the amygdala and the orbitofrontal regions, which modulate the areas involved in the control or regulation of respiration, phonation, and articulation.

Two main brain motor systems are thought to be crucial for the generation of vocalization; the first system is composed of the mesencephalic periaqueductal gray matter, and the second involves the motor and premotor cortical areas and the subcortical nuclei, especially the basal ganglia (Fruhholz, Klaas, Patel, & Grandjean, 2015; Holstege & Subramanian, 2016); see Fig. 5.2. While the periaqueductal gray

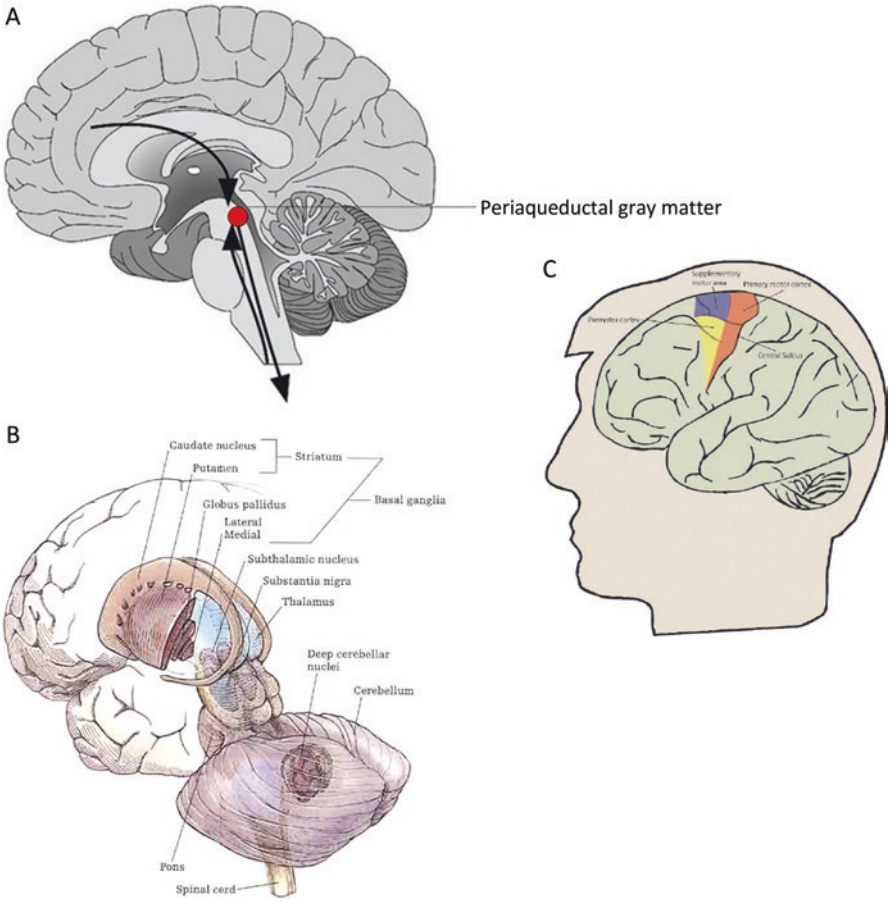

Fig. 5.2 Two main brain systems are involved in vocalization: the periaqueductal gray matter (**a**) and the basal ganglia (**b**), as well as cortical regions, including the primary motor cortex, the supplementary motor area, and the premotor cortex (**c**)

matter is essential for prototypical and automatized vocal production, the cortical areas are more related to planned and controlled motor actions. The periaqueductal gray matter is influenced by limbic and cortical regions such as the amygdala, anterior cingulate cortex, insula, and orbitofrontal cortical areas and constitutes a neuronal network involved in emotional vocalizations. This region impacts strongly on the caudal medullary nucleus retroambiguus. It has direct access to the motor neurons involved in vocalization innervating the upper part of the vocal tract, including the soft palate, pharynx, and larynx, as well as the muscle groups in the lower part of the body, including the diaphragm and the intercostal, abdominal, and pelvic floor muscles. All of these systems determine the muscle configuration related to intra-abdominal, intrathoracic, and subglottic pressure, which is necessary for generating vocalization (Holstege & Subramanian, 2016).

To date, only a few brain imaging studies have investigated the brain regions involved in emotional vocalization production in humans. The first evidence about the role of the different brain regions involved in emotional prosody decoding comes from brain lesion studies. In these studies, the authors have shown that a right hemispheric lesion impacts strongly on the production of emotional prosody, inducing a syndrome called dysprosodic or aprosodic production (Blonder et al., 2005; Cancelliere & Kertesz, 1990; Grela & Gandour, 1998; Guranski & Podemski, 2015; Nakhutina, Borod, & Zgaljardic, 2006; Schirmer, Alter, Kotz, & Friederici, 2001). Other brain lesions can also greatly impact on the production of vocalization and prosody, such as those in the periaqueductal gray matter (P. J. Davis et al., 1996), medial frontal cortex (Frysztak & Neafsey, 1991), and basal ganglia (Blonder, Pickering, Heath, Smith, & Butler, 1995; Cohen, Riccio, & Flannery, 1994). Using functional magnetic resonance imaging, Pichon and Kell (Pichon & Kell, 2013) showed that during the preparation of emotional vocalization, the bilateral ventral striatum is more activated than it is during the preparation of nonemotional vocalization. They also highlighted that these specific basal ganglia ventral regions are functionally connected to the temporal poles and the insular cortices; see Fig. 5.2. In addition, they demonstrated that the dorsal parts of the striatum are more involved in the preparation and production of cognitive and motor components of emotional vocalization production and that these dorsal regions are more connected to the motor network of speech production. These authors also discussed how emotional prosody processing involved the right posterior part of the temporal sulcus and gyrus, proposing that the classic view of right lateralization of emotional prosody production might be related to the ability of these right lateralized neuronal populations to process slow acoustical changes through the ventral cortical pathway. In one study comparing repeated and evoked angry and neutral production (Fruhholz et al., 2015), we revealed that angry voice production, compared to neutral production, induced an increased *bold* signal in the temporal voice-sensitive areas, including the left middle superior temporal gyrus (STG) and posterior STG, right STG, anterior cingulate cortex, bilateral basal ganglia (left putamen and right caudate nucleus), and bilateral inferior frontal gyri (IFG). We also showed a significant increase in amygdala response for the angry-evoked condition compared to angry repetition or neutral productions. In a follow-up study (Klaas, Fruhholz, & Grandjean, 2015)

investigating functional connectivity patterns during emotional vocalization production, we revealed an increase in functional connectivity between the bilateral auditory cortices during affective vocalizations, which points to a bilateral exchange of relevant acoustic information of produced vocalizations. We also showed that bilateral motor cortices involved in the control of vocal motor behavior revealed functional connectivity to the right IFG and right STG. Moreover, we confirmed that different parts of the basal ganglia presented both positive and negative modulatory connectivity with the anterior cingulate cortex and the IFG, as well as with different parts of the STG.

The studies investigating the production of affective vocalizations and affective prosody in newborns are scarce even though the ability of newborns to produce affective bursts starts at birth. For example, Stewart and collaborators (Stewart et al., 2013) have shown significant correlations between autonomic markers (reductions in heart period and respiratory sinus arrhythmia) and the acoustical characteristics of affective vocalizations in newborns, highlighting the possible early coupling between the reactivity of the autonomic system and the maxima and minima of prosodic cues. These results favor a functional role for the vocalizations produced by newborns during emotional states (as marked by autonomic reactions) to induce phasic reactions in parental care (attentional focus toward the baby's state) in order to promote actions to reduce the newborn's negative emotional states (i.e., distress), for example, related to a hungry condition. These kinds of early distress vocalizations are thought to be mainly associated with the periaqueductal gray system. An interesting study done by Mampe and collaborators has revealed that early vocalizations such as cries are influenced by the surrounding spoken language, probably by passive exposure during the last 3 months of pregnancy (Mampe, Friederici, Christophe, & Wermke, 2009). They analyzed 60 crying patterns of newborns (2–5 days of life), 30 French and 30 German, revealing that French newborns produced a pattern of rising melodies, whereas German newborns preferentially produced a falling melodic contour, like those produced by the respective adults (French versus German) in everyday life in terms of usual prosodic contours. During development, newborns, and later on infants, learn progressive modulations of this periaqueductal gray system in order to increase control of their vocalizations as a result of the complex relationships between motor cortical areas and the basal ganglia systems promoting, in the context of vocalization, what scientists call "habits". Two main components have been proposed as being crucial to the control of vocalization, also called the "principle of efficient modulation". The first is the vertical component described in the frame/content theory in which MacNeilage and collaborators proposed that the first training for vocalization control is related to mandibular oscillation (i.e., openness/closing mouth movements; B. L. Davis, MacNeilage, & Matyear, 2002; MacNeilage, 1998a, 1998b), allowing newborns to progressively use a more fine-tuned way to shape their vocal production. The second component, described as the horizontal component, has been called "constriction control" in which fine control of the upper part of the vocal tract allows newborns/infants to shape the resonances useful for speech production (Boe et al., 2013). Such progressive con-

trol and instrumentalization of vocalizations in newborns is, of course, contingent on early life interactions in the context of child caring by adults.

To summarize, the studies about emotional prosody production have revealed that depth structures, especially the periaqueductal gray matter, are essential in the context of automatic vocalization production, often occurring in the context of emotions. During their development, children progressively take control of their vocalization in more structured communication contexts as a result of different cortical areas, including not only the pre-supplementary motor area, supplementary motor area, and motor cortices but also the ventro-dorsal prefrontal cortex (Aboitiz & Garcia, 2009), the subcortical areas such as the basal ganglia nuclei for habits (Graybiel, 2008; Peron, Fruhholz, Verin, & Grandjean, 2013), and the basic periaqueductal gray system. As mentioned earlier, these systems are also functionally connected to the auditory system (including the primary auditory cortex, STG, and superior temporal sulcus (STS), especially for auditory feedback) and to different structures involved in emotion such as the amygdala, insula, and anterior cingulate cortex, among others. Finally, the role of the inferior frontal areas is also crucial in speech production and emotional categorization in the context of prosodic modulations (Fruhholz et al., 2015; Klaas et al., 2015).

Relationships Between Produced Emotional Vocalizations and Their Perception

The perception of emotional prosody can be described in a systematic fashion, as Brunswik (Brunswik, 1956) has suggested for other mechanisms of human characteristic attribution (e.g., intelligence), by using a dedicated model between the speaker and the listener and taking into consideration different processing steps. In an adapted Brunswik model for the perception of the emotional voice (Grandjean, Banziger, & Scherer, 2006), we have proposed that the perception of emotion in vocalization is related to a kind of schemata recognition based on the probability that specific acoustic features are specifically correlated to specific states, as has been shown by the systematic analysis of acoustical profiles of different emotions (Banse & Scherer, 1996); see Fig. 5.3.

The initial step is about the production and how the vocal message is encoded – unintentionally or intentionally – from the speaker's perspective. This means that in a specific emotional state, the different components of emotion affect the way that the speaker produces a specific vocalization. At the end of the speaker's vocal tract, one can systematically characterize the acoustic features for a specific vocalization. This available information can be used by the listener's cognitive system in order to build up a representation of this vocal percept. As in the other sensory domains, our auditory perceptions do not correspond one on one with how researchers can characterize the produced vocalizations at the physical level (i.e., objective). For example, we are more sensitive to intensity level (i.e., energy) in high frequencies compared with low frequencies, meaning that our sensory and cognitive systems are

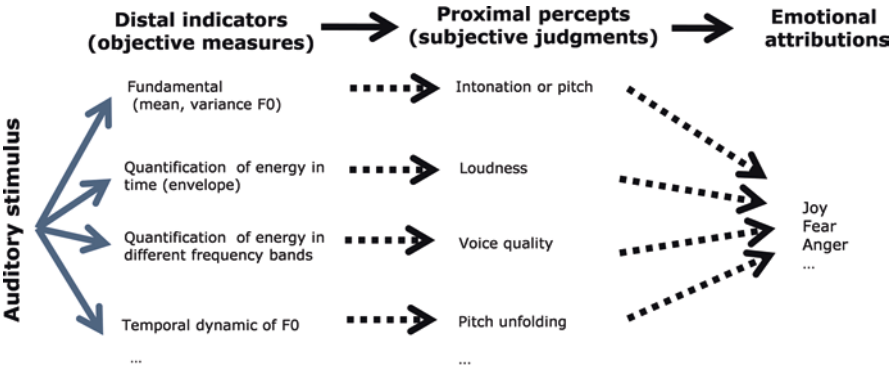

Fig. 5.3 Adapted Brunswik's model for the perception of emotional prosody (Adapted from Grandjean et al. (2006))

not linearly sensitive to intensity and frequency. Physical analysis of the fundamental frequency of a sound (f0) does not correspond perfectly to the perceived f0, called pitch. This perception step has been conceptualized in a Brunswikian approach as the first stage of decoding from the listener's perspective. Of course, the environment in which a specific vocalization is produced can also have a large impact on perception, for example, in the context of a noisy situation such as that experienced by premature babies in neonatology environments. This first stage of perception is thought to be followed by a process of percept integration in order to build up a dynamic integrated percept, such as that categorized as a human voice, or corresponding to the sound produced by a specific object, such as a glass breaking on the floor. The so-called perceived auditory object is also the subject of a third processing step that we can describe as the processing of categorization. In emotion recognition, it is related to a series of processes, based on a probabilistic approach, to characterize the percept as a sign of joy or irritation, for example. It is also related to the ability of the cognitive system to extract the invariants between sensory information and inferences based on past experiences; these inferences are also used by humans and other animals such as apes to attribute emotional mental states of others (Gruber & Grandjean, 2017).

A large corpus of literature has been developed to systematically study how the human brain is able to extract emotional information from voice signals. Our auditory peripheral system is characterized by a series of mechanisms, from air pressure to the vibration of the tympanic membrane to a mechanical structure comprising three ossicles (malleus, incus, and stapes). The last part of this mechanical system is able to modulate the vibration of the little oval membrane in our inner ear, which in turn produces movement of the liquid contained in the cochlear system. The waves of this liquid induce movement of the specific appendages of nerve cells, the so-called hair cells. A metabolic cascade, the so-called transduction phenomenon (not described here in detail) produces the first neural signal in our nervous system. The firing of these neurons entrains a series of brain processes along what is called

the ascending auditory pathway. This neural pathway is organized into several cerebral structures, starting from the ventral and dorsal cochlear nuclei located in the brain stem and projecting to the superior olivary complex and the inferior colliculus, respectively. The latter projects to the medial geniculate nuclei of the thalamus, which then projects to the primary auditory cortical regions. Brain imaging studies that investigated the perception of emotional prosody have revealed a complex neuronal network. This network is composed of different parts of temporal areas, including the STG and the STS. These voice-sensitive areas, described primarily by Belin and collaborators (Belin, Zatorre, Lafaille, Ahad, & Pike, 2000), are modulated by emotional prosody (Ceravolo, Fruhholz, & Grandjean, 2016a, 2016b; Ethofer et al., 2006, 2012; Ethofer, De Ville, Scherer, & Vuilleumier, 2009; Fruhholz, Ceravolo, & Grandjean, 2012; Fruhholz & Grandjean, 2013a; Fruhholz et al., 2016; Grandjean et al., 2005; Sander et al., 2005; Wiethoff et al., 2008; Wildgruber et al., 2005) even in the context of unvoiced emotional prosody, that is, whispered voicing (Fruhholz, Trost, & Grandjean, 2016), which is often used in the context of intimate communication. These temporal regions are thought to be crucial in building up complex auditory objects and are significant in the context of emotional communication. The amygdala complex is also thought to be essential in emotional prosody decoding, especially for rapid reaction to such emotional signals and for organization of the emotional response to such relevant positive or negative auditory stimuli, as evidenced by brain imaging (Bach et al., 2008; Fruhholz et al., 2012; Fruhholz & Grandjean, 2012a; Johnstone, van Reekum, Oakes, & Davidson, 2006; Sander et al., 2005; Schirmer et al., 2008; Wiethoff, Wildgruber, Grodd, & Ethofer, 2009; Wildgruber, Ackermann, Kreifelts, & Ethofer, 2006) and amygdala lesion studies (Dellacherie, Hasboun, Baulac, Belin, & Samson, 2011; Fruhholz et al., 2015; Scott et al., 1997). These temporal and subcortical regions are parts of a distributed neural network also involving the frontal regions, especially the inferior frontal regions (IFG) and the orbitofrontal regions (OFC). Although the right IFG is thought to be especially important for discrimination and categorization of emotional prosody that is based on dynamic acoustic invariants (Beaucousin et al., 2007; Buchanan et al., 2000; Ethofer et al., 2006, 2012; Eviatar & Just, 2006; Fruhholz et al., 2012; Fruhholz & Grandjean, 2012b, 2013b; Fruhholz, Gschwind, & Grandjean, 2015; Johnstone et al., 2006; Leitman et al., 2010; Mizuno & Sugishita, 2007; Tzourio-Mazoyer et al., 2007; Wildgruber et al., 2004; 2006), the OFC is especially important for contextual valuation of emotional auditory stimuli such as in the case of sarcasm (Paulmann, Seifert, & Kotz, 2010; Sander et al., 2005; Schirmer & Kotz, 2006; Schirmer et al., 2008; Wildgruber et al., 2006); see Fig. 5.4.

Empirical evidence for auditory emotional perception at an early stage of life is scarce. It has been demonstrated, using behavioral measures, that newborns have a preference for the human voice compared to nonvocal auditory stimuli (EcklundFlores & Turkewitz, 1996; Hutt, Vonbernu, Lenard, Hutt, & Prechtl, 1968). Using electroencephalography and mismatch negativity, an event-related component elicited in the context of specific rule violations, Cheng and collaborators showed in 1- to 5-day-old newborns that, compared to neutral or to matched control stimuli, happy, angry, and fearful intonations (using the word "dada") elicited a significant stronger

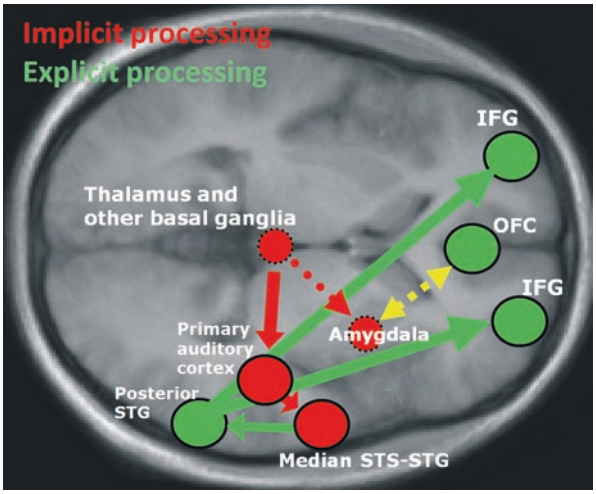

Fig. 5.4 Cerebral network involved in the processing of emotional vocalization: implicit processing (*red*), including the thalamus and other basal ganglia (e.g., subthalamic nucleus, striatum), primary auditory cortex, median superior temporal sulcus (*STS*) and superior temporal gyrus (*STG*), and amygdala complex, and explicit processing (*green*), including mainly the posterior STG, inferior frontal gyri (*IFG*), and orbitofrontal cortex (*OFC*). The median STS and STG are parts of the sensitive voice regions described by Belin et al. in 2000 (Belin et al., 2000) (Adapted from Wildgruber, Ethofer, Grandjean, and Kreiflets, 2009)

mismatch response in the right hemisphere (Cheng, Lee, Chen, Wang, & Decety, 2012). Furthermore, Mastropieri and Turkewitz showed that happy intonations more often elicited open-eye behaviors when newborns were exposed to a typical mother speech pattern, which was not the case for a foreign language (Mastropieri & Turkewitz, 1999). This observation is compatible with the concept of sensitivity of language context during pregnancy (see also Chap. 2 by Moon). The early ability to react to the mother's voice has also been documented in premature newborns. Filippa and collaborators (Filippa, Devouche, Arioni, Imberty, & Gratier, 2013) showed that the physiological state of a premature newborn is significantly modulated during exposure to the mother's live voice compared to situations without such exposure. Furthermore, they showed that exposure to the mother's voice significantly decreases the number of critical events during this challenging period of development. Using functional magnetic resonance imaging, Abrams and colleagues (Abrams et al., 2016) showed that at a later stage of development, exposure of infants to the biological mother's voice (around 10 years), compared to other unknown mothers' voices, induced an increase in activity (*bold* signal) in a large neuronal network, including the primary auditory network in the midbrain and cortical regions, the voice-sensitive temporal areas (STG/STS), and a series of regions known for their implications in emotional processes, including the amygdala, nucleus accumbens (in the ventral part of the striatum), OFC, and anterior insula and cingulate. They also demonstrated the implications of the fusiform gyrus, especially the part known to be modulated by face perception; the authors interpreted this finding in the context of facial

mental imagery related to the perception of infants to their own mother's voice. The brain connectivity patterns have also revealed significant correlations between the infant's social communication skills and the strength of brain connectivity between the voice-sensitive areas (especially the STS) and the amygdala, nucleus accumbens, OFC, dorsal part of the cingulate, and anterior insula.

The Concept of Scheme of Production and Perception in the Context of Social Communication

One important issue in the developmental perspective is how the cognitive system and, by extension, the brain mechanisms are able to evolve in ontogenetic time to at least be a representative part of reality and then render the organism able to interact with his/her complex environment, not only at the physical level but also at the social and emotional levels. Based on genetic and early epigenetic influences and related phylogenetic determinants, the brain and other body systems (e.g., muscular, skeletal, sensory systems) have to be able to learn how to guide behavior in order to achieve numerous different goals, the first being to survive. In order to stabilize the relationships between environmental features, interoceptive and proprioceptive states, and the interactions with motor actions and related explicit or implicit goals, the system has to be able to extract the invariants at different levels. Several researchers have proposed that a Bayesian approach might be a way to explain this ability of biological systems to build up stable coupling between internal states, representations, and actions. One brain system seems to be an excellent candidate to achieve this: the complex system formed by the dynamic interactions between the basal ganglia, other subcortical structures (such as hippocampal formation), and cortical territories. The basal ganglia are a complex system that include the brain origins of the dopaminergic system through the ventral tegmental area and the substantia nigra. These regions are characterized by a huge number of dopaminergic neurons that impact on the rest of the brain. Different nuclei comprise the basal ganglia system: the subthalamic nucleus, globus pallidus, striatum composed of the nucleus accumbens and the caudate nucleus, and thalamus. These nuclei have numerous connections with other subcortical structures such as the amygdala, hippocampal formation, claustrum, and colliculus, among others. Graybiel (2008) has proposed that this system is crucial to build up what she called a chunking process. This concept, chunking, from our point of view, seems crucial to explain how, starting from a hyperconnected and relatively unstructured system, which is the state of the early brain-body system in newborns, the system is able to extract invariants between the sensory and motor systems and environmental elements. Chunking is defined as a series of simple steps, with processing action becoming a kind of integrated series of evaluation-representation-actions able to achieve a specific function, for example, grasping an object. In this case, the coupling of sensory/representation/action is simple, but the concept has been extended to more complex behavior such as that involved in complex learning, for example, being able to drive a car automatically

(i.e., after learning to drive, it is no longer necessary to think effortfully about each action for driving). Graybiel has proposed the concept of habits, which is opposed to what researchers have defined as goal-oriented actions. We think that the early ability to perceive the world is not very different in terms of the mechanisms involved to what we described for learning to drive. This means that the system has to learn and then extract the invariants between sensory information and representation/action related to them. From such a perspective, the ability of the newborn to learn about this environment and how to interact with it necessitates automatization of some procedures involving specific brain region interactions and then the progressive shaping of the neuronal system through the synaptic weights, which is a kind of crystallization of neuronal networks to achieve specific functions. The learning of social communication, especially through the vocal system, can be understood in this context of learning habits, both at the production and the perception levels. The central nervous system of the newborn is then able to learn progressively, by systematic invariant extraction in the context of action/perception, how to produce specific vocalizations and adapt them to the interpersonal context. With the systematic repetition or special relevance of some situations, the neural connectivity pattern, thanks to the basal ganglia system and its interaction with cortical areas, is progressively shaped to respond to specific functions, for example, social functions, both at the production and the perception levels.

In light of the studies performed in adults and newborns in the vocal domain, we can reasonably predict that during early parental emotional vocal interactions, similar brain regions are involved in both newborns and adults. Of course, the ability of newborns to extract invariants for both production and perception levels is progressively shaped by their experience. Early life interactions are crucial to promote the best functioning of the newborn and to shape the complex brain-body systems for the best adaptation of the physical environment but also, and perhaps even more crucially, for the social context.

The investigation of the brain mechanisms involved in the production and perception of emotional voice processing at an early stage of development, including in premature populations, should be extended and developed further. Such studies may better characterize how and when future interventions, such as exposure to the mother's live voice, have the best impact on promoting the development of the social brain during this crucial period, which is strongly characterized by brain maturation processes, and on long-term effects on the development of social skills.

Key Messages

- The production of reflex-like vocalizations at an early stage of development is mainly related to the periaqueductal gray matter. Progressively, newborns learn to control the production of emotional vocalization through interactions between this brain region as well as the subcortical nuclei, especially the basal ganglia, and motor cortical areas. This progressive control is essential in the social life of newborns and later on during childhood.

- The perception of emotion in the voice involves a series of specific brain mechanisms, mainly including the voice-sensitive areas (superior temporal regions), basal ganglia, amygdala, orbitofrontal regions, and inferior frontal regions.
- The development of habits for emotional voice processing involves complex functional neuronal loops between the basal ganglia and other subcortical and cortical areas, both at the production and the perception level, and is crucial for social communication skills.

References

Aboitiz, F., & Garcia, R. (2009). Merging of phonological and gestural circuits in early language evolution. *Reviews in the Neurosciences*, *20*(1), 71–84.

Abrams, D. A., Chen, T. W., Odriozola, P., Cheng, K. M., Baker, A. E., Padmanabhan, A., … Menon, V. (2016). Neural circuits underlying mother's voice perception predict social communication abilities in children. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(22), 6295–6300.

Bach, D. R., Grandjean, D., Sander, D., Herdener, M., Strik, W. K., & Seifritz, E. (2008). The effect of appraisal level on processing of emotional prosody in meaningless speech. *NeuroImage*, *42*(2), 919–927.

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

Beaucousin, V., Lacheret, A., Turbelin, M. R., Morel, M., Mazoyer, B., & Tzourio-Mazoyer, N. (2007). FMRI study of emotional speech comprehension. *Cerebral Cortex*, *17*(2), 339–352.

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

Blonder, L. X., Heilman, K. M., Ketterson, T., Rosenbek, J., Raymer, A., Crosson, B., … Roth, L. G. (2005). Affective facial and lexical expression in aprosodic versus aphasic stroke patients. *Journal of the International Neuropsychological Society*, *11*(6), 677–685.

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.

Boe, L. J., Badin, P., Menard, L., Captier, G., Davis, B., MacNeilage, P., … Schwartz, J. L. (2013). Anatomy and control of the developing human vocal tract: A response to Lieberman. *Journal of Phonetics*, *41*(5), 379–392.

Boiten, F. A. (1998). The effects of emotional behaviour on components of the respiratory cycle. *Biological Psychology*, *49*(1–2), 29–51.

Brunswik, E. (1956). *Perception and the representative design of psychological experiments* (2nd ed.). Berkeley, CA: University of California Press.

Buchanan, T. W., Lutz, K., Mirzazade, S., Specht, K., Shah, N. J., Zilles, K., & Jancke, L. (2000). Recognition of emotional prosody and verbal components of spoken language: An fMRI study. *Brain Research. Cognitive Brain Research*, *9*(3), 227–238.

Butler, E. A., Wilhelm, F. H., & Gross, J. J. (2006). Respiratory sinus arrhythmia, emotion, and emotion regulation during social interaction. *Psychophysiology*, *43*(6), 612–622.

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

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*, 216.

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.

Cheng, Y. W., Lee, S. Y., Chen, H. Y., Wang, P. Y., & Decety, J. (2012). Voice and emotion processing in the human neonatal brain. *Journal of Cognitive Neuroscience*, *24*(6), 1411–1419.

Cohen, M. J., Riccio, C. A., & Flannery, A. M. (1994). Expressive aprosodia following stroke to the right basal ganglia: A case report. *Neuropsychology*, *8*(2), 242–245.

Davis, B. L., MacNeilage, P. F., & Matyear, C. L. (2002). Acquisition of serial complexity in speech production: A comparison of phonetic and phonological approaches to first word production. *Phonetica*, *59*(2–3), 75–107.

Davis, P. J., Zhang, S. P., Winkworth, A., & Bandler, R. (1996). Neural control of vocalization: Respiratory and emotional influences. *Journal of Voice*, *10*(1), 23–38.

Dellacherie, D., Hasboun, D., Baulac, M., Belin, P., & Samson, S. (2011). Impaired recognition of fear in voices and reduced anxiety after unilateral temporal lobe resection. *Neuropsychologia*, *49*(4), 618–629.

EcklundFlores, L., & Turkewitz, G. (1996). Asymmetric headturning to speech and nonspeech in human newborns. *Developmental Psychobiology*, *29*(3), 205–217.

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.

Etzel, J. A., Johnsen, E. L., Dickerson, J., Tranel, D., & Adolphs, R. (2006). Cardiovascular and respiratory responses during musical mood induction. *International Journal of Psychophysiology*, *61*(1), 57–69.

Eviatar, Z., & Just, M. A. (2006). Brain correlates of discourse processing: An fMRI investigation of irony and conventional metaphor comprehension. *Neuropsychologia*, *44*(12), 2348–2359.

Filippa, M., Devouche, E., Arioni, C., Imberty, M., & Gratier, M. (2013). Live maternal speech and singing have beneficial effects on hospitalized preterm infants. *Acta Paediatrica*, *102*(10), 1017–1020.

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. (2012a). Amygdala subregions differentially respond and rapidly adapt to threatening voices. *Cortex, 49*, 1394–1403.

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

Fruhholz, S., & Grandjean, D. (2013a). Multiple subregions in superior temporal cortex are differentially sensitive to vocal expressions: A quantitative meta-analysis. *Neuroscience and Biobehavioral Reviews*, *37*(1), 24–35.

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 United States of America*, *112*(5), 1583–1588.

Fruhholz, S., Klaas, H. S., Patel, S., & Grandjean, D. (2015). Talking in fury: The cortico-subcortical network underlying angry vocalizations. *Cerebral Cortex*, *25*(9), 2752–2762.

Fruhholz, S., Trost, W., & Grandjean, D. (2016). Whispering – The hidden side of auditory communication. *NeuroImage*, *142*, 602–612.

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.

Frysztak, R. J., & Neafsey, E. J. (1991). The effect of medial frontal cortex lesions on respiration, "freezing," and ultrasonic vocalizations during conditioned emotional responses in rats. *Cerebral Cortex*, *1*(5), 418–425.

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

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.

Grandjean, D., Sander, D., & Scherer, K. R. (2008). Conscious emotional experience emerges as a function of multilevel, appraisal–driven response synchronization. *Consciousness and Cognition*, *17*(2), 484–495.

Graybiel, A. M. (2008). Habits, rituals, and the evaluative brain. *Annual Review of Neuroscience*, *31*, 359–387.

Grela, B., & Gandour, J. (1998). Locus of functional impairment in the production of speech rhythm after brain damage: A preliminary study. *Brain and Language*, *64*(3), 361–376.

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

Guranski, K., & Podemski, R. (2015). Emotional prosody expression in acoustic analysis in patients with right hemisphere ischemic stroke. *Neurologia i Neurochirurgia Polska*, *49*(2), 113–120.

Holstege, G., & Subramanian, H. H. (2016). Two different motor systems are needed to generate human speech. *The Journal of Comparative Neurology*, *524*(8), 1558–1577.

Hutt, C., Vonbernu, H., Lenard, H. G., Hutt, S. J., & Prechtl, H. F. R. (1968). Habituation in relation to state in human neonate. *Nature*, *220*(5167), 618–620.

Johnstone, T., van Reekum, C. M., Oakes, T. R., & Davidson, R. J. (2006). The voice of emotion: An FMRI study of neural responses to angry and happy vocal expressions. *Social Cognitive and Affective Neuroscience*, *1*(3), 242–249.

Klaas, H. S., Fruhholz, S., & Grandjean, D. (2015). Aggressive vocal expressions-an investigation of their underlying neural network. *Frontiers in Behavioral Neuroscience*, *9*, 121.

Leitman, D. I., Wolf, D. H., Ragland, J. D., Laukka, P., Loughead, J., Valdez, J. N., … Gur, R. C. (2010). "It's not what you say, but how you say it": A reciprocal temporo-frontal network for affective prosody. *Frontiers in Human Neuroscience*, *4*, 19.

MacNeilage, P. F. (1998a). The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences*, *21*(4), 499–511.

MacNeilage, P. F. (1998b). The frame/content view of speech: What survives, what emerges. *Behavioral and Brain Sciences*, *21*(4), 532–546.

Mampe, B., Friederici, A. D., Christophe, A., & Wermke, K. (2009). Newborns' cry melody is shaped by their native language. *Current Biology*, *19*(23), 1994–1997.

Mastropieri, D., & Turkewitz, G. (1999). Prenatal experience and neonatal responsiveness to vocal expressions of emotion. *Developmental Psychobiology*, *35*(3), 204–214.

Mizuno, T., & Sugishita, M. (2007). Neural correlates underlying perception of tonality-related emotional contents. *Neuroreport*, *18*(16), 1651–1655.

Nakhutina, L., Borod, J. C., & Zgaljardic, D. J. (2006). Posed prosodic emotional expression in unilateral stroke patients: Recovery, lesion location, and emotional perception. *Archives of Clinical Neuropsychology*, *21*(1), 1–13.

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.

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.

Pichon, S., & Kell, C. A. (2013). Affective and sensorimotor components of emotional prosody generation. *Journal of Neuroscience*, *33*(4), 1640–1650.

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.

Sander, D., Grandjean, D., & Scherer, K. R. (2005). A systems approach to appraisal mechanisms in emotion. *Neural Networks*, *18*(4), 317–352.

Scherer, K. R. (1984a). Emotions – Functions and components. *Cahiers De Psychologie Cognitive-Current Psychology of Cognition*, *4*(1), 9–39.

Scherer, K. R. (1984b). On the nature and function of emotion. A component process approach. In K. R. Scherer & P. Ekman (Eds.), *Approaches to emotion* (pp. 293–317). Hillsdale, MI: Erlbaum.

Scherer, K. R., Dan, E. S., & Flykt, A. (2006). What determines a feeling's position in affective space? A case for appraisal. *Cognition & Emotion*, *20*(1), 92–113.

Schirmer, A., Alter, K., Kotz, S. A., & Friederici, A. D. (2001). Lateralization of prosody during language production: A lesion study. *Brain and Language*, *76*(1), 1–17.

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

Scott, S. K., Young, A. W., Calder, A. J., Hellawell, D. J., Aggleton, J. P., & Johnson, M. (1997). Impaired auditory recognition of fear and anger following bilateral amygdala lesions. *Nature*, *385*(6613), 254–257.

Smith, C. A., & Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology*, *48*(4), 813–838.

Stewart, A. M., Lewis, G. F., Heilman, K. J., Davila, M. I., Coleman, D. D., Aylward, S. A., & Porges, S. W. (2013). The covariation of acoustic features of infant cries and autonomic state. *Physiology & Behavior*, *120*, 203–210.

Subramanian, H. H., Arun, M., Silburn, P. A., & Holstege, G. (2016). Motor organization of positive and negative emotional vocalization in the cat midbrain periaqueductal gray. *The Journal of Comparative Neurology*, *524*(8), 1540–1557.

Tzourio-Mazoyer, N., Beaucousin, V., Lacheret, A., Turbelin, M. R., More, M., & Mazoyer, B. (2007). FMRI study of emotional speech comprehension. *Cerebral Cortex*, *17*(2), 339–352.

van Reekum, C. M., Johnstone, T., Banse, R., Etter, A., Wehrle, T., & Scherer, K. R. (2004). Psychophysiological responses to appraisal dimensions in a computer game. *Cognition & Emotion*, *18*(5), 663–688.

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.

Wiethoff, S., Wildgruber, D., Kreifelts, B., Becker, H., Herbert, C., Grodd, W., & Ethofer, T. (2008). Cerebral processing of emotional prosody – Influence of acoustic parameters and arousal. *NeuroImage*, *39*(2), 885–893.

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.

Wildgruber, D., Ethofer, T., Grandjean, D., & Kreiflets, B. (2009). A cerebral network model of speech prosody comprehension. *International Journal of Speech-Language Pathology*, *11*(4), 277–281.

Wildgruber, D., Hertrich, I., Riecker, A., Erb, M., Anders, S., Grodd, W., & Ackermann, H. (2004). Distinct frontal regions subserve evaluation of linguistic and emotional aspects of speech intonation. *Cerebral Cortex*, *14*(12), 1384–1389.

Wildgruber, D., Riecker, A., Hertrich, I., Erb, M., Grodd, W., Ethofer, T., & Ackermann, H. (2005). Identification of emotional intonation evaluated by fMRI. *NeuroImage*, *24*(4), 1233–1241.