



## Review article

# The sound of emotions—Towards a unifying neural network perspective of affective sound processing



Sascha Frühholz<sup>a,b,c,\*</sup>, Wiebke Trost<sup>d</sup>, Sonja A. Kotz<sup>e,f</sup>

<sup>a</sup> Department of Psychology, University of Zurich, Zurich, Switzerland

<sup>b</sup> Neuroscience Center Zurich, University of Zurich and ETH Zurich, Zurich, Switzerland

<sup>c</sup> Center for Integrative Human Physiology (ZIHP), University of Zurich, Switzerland

<sup>d</sup> Swiss Center for Affective Sciences, University of Geneva, Geneva, Switzerland

<sup>e</sup> Faculty of Psychology and Neuroscience, Department of Neuropsychology and Psychopharmacology, Maastricht University, Maastricht, Netherlands

<sup>f</sup> Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

## ARTICLE INFO

## Article history:

Received 7 January 2016

Received in revised form 1 May 2016

Accepted 4 May 2016

Available online 14 May 2016

## Keywords:

Affect

Sound

Voice

Music

Limbic system

Auditory cortex

Basal ganglia

Cerebellum

## ABSTRACT

Affective sounds are an integral part of the natural and social environment that shape and influence behavior across a multitude of species. In human primates, these affective sounds span a repertoire of environmental and human sounds when we vocalize or produce music. In terms of neural processing, cortical and subcortical brain areas constitute a distributed network that supports our listening experience to these affective sounds. Taking an exhaustive cross-domain view, we accordingly suggest a common neural network that facilitates the decoding of the emotional meaning from a wide source of sounds rather than a traditional view that postulates distinct neural systems for specific affective sound types. This new integrative neural network view unifies the decoding of affective valence in sounds, and ascribes differential as well as complementary functional roles to specific nodes within a common neural network. It also highlights the importance of an extended brain network beyond the central limbic and auditory brain systems engaged in the processing of affective sounds.

© 2016 Elsevier Ltd. All rights reserved.

## Contents

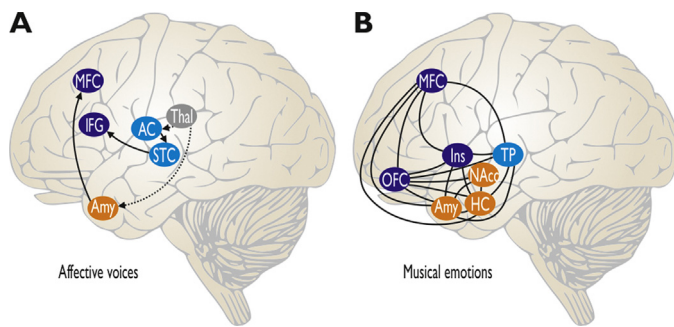
1. Introduction.....	96
2. A typology of affective sounds.....	97
3. Central brain systems mediating the processing of affective sounds across domains.....	98
3.1. The amygdala and the auditory cortex—complementary regions for the decoding of the emotional meaning from sound.....	102
3.2. Mediation and evaluation in the fronto-insular brain system.....	103
3.3. Temporal prediction and response adaptation in the basal ganglia and the cerebellum.....	104
4. An extended brain network for complex affective sounds.....	104
5. A neurocognitive model for the distributed processing of affective sounds.....	105
6. Hemispheric lateralization of affective sound processing.....	106
7. Conclusions.....	106
Acknowledgments.....	107
References.....	107

\* Corresponding author at: Department of Psychology, University of Zurich, Binzmühlestrasse 14, Box 18, Zurich 8050, Switzerland.

E-mail address: [sascha.fruehholz@uzh.ch](mailto:sascha.fruehholz@uzh.ch) (S. Frühholz).

## 1. Introduction

Sounds we hear often carry emotional meaning that affects us consciously or unconsciously. Imagine you are at a shopping mall – subtle background music plays and whether you like it or not you buy more than you want and feel good about it. The melody, rhythm, tempo, and timbre of this background music, the lyrics,



**Fig. 1.** Previous domain-specific models for affective sound processing. (A) Neural models for processing affective voices (i.e. nonverbal expressions, affective prosody) proposed a hierarchical processing from an acoustic analysis in low-level auditory cortex (AC) and the representation of auditory percepts in higher-level auditory cortex (STC). From the STC, auditory information is fed forward to the IFC for a cognitive evaluation and classification of the emotional meaning of voices (Schirmer and Kotz, 2006; Wildgruber et al., 2009). For processing affective voices outside the focus of attention, the amygdala decodes emotional information, which is fed forward to the MFC (Wildgruber et al., 2009). The thalamus (thal) is shown here as it is assumed to provide basic acoustic information both for the AC and the amygdala (dashed lines). (B) A recent neural model for processing musical emotions especially focused on regions of the limbic system, the limbic striatum, and the temporal pole (TP) (Koelsch, 2010). The model proposes anatomical connections between these regions, but seems underspecified in terms of exact the functional role of these connections.

and the acoustic features of the singing voice seem to capture our attention and induce an emotional reaction. Both the singing and the speaking voice carry emotional meaning, and the human voice expresses emotions in laughter, cries, or screams as well as affective speech melody also called affective prosody (Banse and Scherer, 1996; Patel et al., 2011; Sundberg et al., 2011). Further, natural environmental or animal sounds such as a car screeching or a monkey screaming in the zoo also induce emotional reactions in us.

These examples show that different simple and complex sound types can convey emotional meaning. Interestingly, these diverse sound types seem to recruit a vastly similar brain network when we listen to them (Escoffier et al., 2012; Frühholz et al., 2014). Core regions of this network (Fig. 1) involve not only cortical but also subcortical brain areas during the decoding of the affective valence from sounds (Frühholz et al., 2014; Koelsch, 2014; Schirmer and Kotz, 2006), and these regions also regulate emotional reactions in the listener (LeDoux, 2012). In contrast, distinct activation patterns for affective sound types were observed, leading to divergent descriptions of this neural network. These differences may originate in theoretical perspectives and specifically derived research questions in different domains of affective sound research including nonhuman sounds, human vocalizations, and music. Many of these differences can be comprehensively integrated by introducing a cross-domain view that motivates and initiates a meaningful cross-talk between research domains. For example, a predominant research issue in studies on musical emotions is the felt pleasure while listening to enjoyable music. However, certain types of human vocalization, such as vocal joy or elation (Bachorowski and Owren, 2001), are similarly pleasurable, but studies of voice processing have yet to consider this perspective.

We propose an integrative and unifying cognitive and neural network perspective of how the brain decodes emotional meaning from various sound types and suggest that this will critically advance the field of auditory emotional neuroscience on three accounts: First, the convergence of a core neural network for affective sound processing allows to explain how affective sounds that differ in complexity and temporal structure, integrate information in time, support the evaluation of their emotional content, and lead to behavioral adaptation. Second, a cross-domain network view integrates different functional processes underlying affective

sound processing. For example, auditory cortex activity in response to affective sounds is considered to reflect the acoustic analysis of these sounds. However, from a cross-domain perspective, such activity may be generic to the decoding of affective sound beyond a simple acoustic analysis. Third, a unifying network view highlights the importance of currently functionally underspecified brain regions involved in affective sound processing.

## 2. A typology of affective sounds

Affective sounds considerably differ in *how* they signal meaning. They also differ in the *origin* of their emotional meaning, which can be in the sound source (i.e. a signal or the sender) or in the receiver (i.e. the listener). Sebeok (1985) introduced a general typology of “natural” and “cultural” signs in the field of semiotics. He distinguished *signals* (arbitrary signs, which trigger reactions in the receiver by mechanistic or conventional mechanisms), *symptoms* (automatic and non-arbitrary signs, which naturally link with the signified), *icons* (a topological resemblance between the sign and the signified), *indexes* (contiguity between the sign and signified, or the sign is a sample of the signified), *symbols* (intentional or cultural conventions), and *names*.

An important issue concerns the distinction between “signals” and “symptoms”. Both types of signs work on a very basic level in terms of triggering (automatic) responses in listeners. A major difference, however, concerns the link between a signal and a symptom, and what they signify. While it has been proposed that there is a natural link between a symptom and its signified, no such link exists for signals. For example, nonhuman sounds can trigger automatic responses in listeners, such as a startle response to a loud sound and sounds with sudden onsets; however, there is nothing in the sound or its features that predefines that individuals startle in response to sudden loud sound onset (i.e. organisms could potentially show any type of response to aversive sounds). Symptoms, however, are proposed to have a natural link in different forms of social communication, such that an affective sound (i.e. a certain vocal affective burst) both potentially induces a certain response in listeners, and is produced with the intention to induce such a response. From an evolutionary perspective, organisms may have learned how the production of certain vocal sounds (i.e. barking of dog) can induce certain reactions in listeners (i.e. defensive reaction in listeners). Consequently, we mainly refer to basic non-human sounds as “signals”, while basic affective bursts in animal, infant, and adult vocalizations are largely “symptoms” (given their “natural” link to the intentional production of the sender, and the expected reaction in the listener), though some of their sound features can also be a basic “signal”.

This typology of signs may be applied to classify affective sounds as different types of “affective signs” (Table 1). For example, the aversive quality of many environmental sounds, such as the sound of chalk on a blackboard, are basic “signals” of emotional meaning, where the meaning originates from the affective reaction of the listener. Some of these environmental sounds are used as an unconditioned stimulus (UCS) in studies of classical conditioning to produce a conditioned stimulus (CS). The CS then is an “index” of the emotional meaning of the UCS. Some other emotional sounds could equate to a “symptom” such as animal sounds or nonverbal human vocalizations originating from and signifying inner emotional states of a sender, but they can also stand for a basic “signal” as we may be evolutionary prepared to show habitual affective responses to animal sounds or infant vocalizations, and especially to some of their sound features. They could also be an “index” for our previous pleasant (e.g. nice dogs) or unpleasant experience (e.g. being bitten by a dog) with such sounds or with the objects producing such sounds.

**Table 1**

A taxonomy of affective sounds originating from different sound sources and which might figure as “affective signs” for potential perceivers of these sounds.

Source	Categories	Examples	Type of sign	Origin of affective meaning
Environmental sounds	Technical sound, natural sound	Chalk on a board, noise bursts, thunder, sudden onset tones	Signal	Perceiver
	Conditioned sound (CS)	Classical conditioned tones	Index	Perceiver
	Animal sound	Barking (dog), growling (monkey), twittering (songbird)	Symptom (sender), signal/index (listener)	Source, Perceiver
Nonverbal expressions	Infant vocalizations	Baby cries, baby laughs, distress	Symptom (sender), signal/index (listener)	Source, (Perceiver)
	Adult vocalizations	Laughing, screaming, moaning	Symptom (sender), signal/index (listener)	Source, (Perceiver)
	Interjections	“hurrah”, “pah”, “wow”	(Symptom), index, symbol	Source
Speech intonations	Prosody	Angrily inflected speech, happy speech intonations	(Symptom), index, symbol	Source
Artificial human speech	Analog	Vocoded speech/voice	Icon	Source
	Digital	Computer generated synthetic speech	Icon	Source
Singing human voice	Nonverbal	Crooning, phonated intonations	Symbol	Source, Perceiver
	Verbally based	Lyrical singing	Symbol	Source, Perceiver
Music	(–)	Classical music, folk music, popular music, ...	Symbol (affective), symbol (aesthetic), signal/symptom (sudden sound changes)	Source, Perceiver

Unlike human nonverbal vocalizations, individuals produce many emotional signs while speaking, such as interjections or affective speech prosody, mainly representing an “index” (contingent to the emotional period) or “symbol” (conventional rules of vocal expression) of the emotional state of a sender. The affective signals during speaking can also be simulated in artificial human speech (Murray and Arnott, 1993) representing an “icon” of the original human speech. Finally, musical expressions of emotions mainly adhere to conventional rules of *emotional* and *aesthetic* expressions and therefore may be considered as a “symbol” even though some musical techniques make use to the fundamental affective properties of basic sounds as “signals” (e.g. sudden onsets of musical events and sudden sound changes).

The typology of affective sounds aims at providing a comprehensive overview of the different types of affective sounds. The typology especially provides a schema to estimate domains of auditory research in which we find ample recent and ongoing research (i.e. affective voices and musical emotions), and areas of research, which so far have been neglected, such as the neural processing of artificial affective sounds (Beaucousin et al., 2007).

### 3. Central brain systems mediating the processing of affective sounds across domains

The neural network supporting the decoding of a number of affective sound types described above engages both cortical and subcortical brain regions. In addition, views put forward by different within-domain neural network models (Koelsch, 2010; Schirmer and Kotz, 2006; Wildgruber et al., 2009) seem to describe similar functional processing of affective sound types, but with differential neural correlates (Fig. 2). Consequently, in a first step, we take a cross-domain neural network view and aim at providing a functional description of the core regions involved in affective sound processing.

To specify, which brain regions establish this network, we reviewed recent literature and assessed how consistently activation occurred for four different affective sound types (Fig. 1A, Table 2). Consistency was assessed as the ratio between studies

reporting activation in a certain brain region and those studies that did not report such activation. Using a minimum consistency level within and across different affective sound types, we identified a core network of brain regions that underlies the decoding of affective meaning from sound (Fig. 1B–C). The consistency level was set to a minimum of 25%, which corresponds to the four different affective types.

We searched for all neuroimaging studies (fMRI and PET), which involved the neural processing of emotional sounds across the main four types of emotional sounds, that is nonhuman environmental sounds, human nonverbal expressions, emotional inflected speech (affective prosody), and musical emotions (Table 2). We also searched for studies including (affective) animal vocalizations and synthetic vocalizations based on our comprehensive typology of affective sounds (Table 1), but this latter search revealed only a minor number of studies and thus were not considered here for the meta-analysis. The search was done with PubMed (<http://www.ncbi.nlm.nih.gov/pubmed/>) using BrainMap database using Sleuth (<http://brainmap.org/sleuth/>) (Laird et al., 2005), PubMed ([www.pubmed.org](http://www.pubmed.org)), and the ISI Web of knowledge (June 2014) using the following search strings: “(vocal OR voice OR auditory OR sound OR tone OR music) AND (emotion OR expressions OR affect OR aversive OR pleasure OR reward) AND (fMRI OR PET)”.

We included only those studies (Table 2) that were published in English before June 2014, included only healthy young and middle-aged human adults, used auditory stimulation, and were original research papers (i.e. no meta-analytic papers and previous reviews). We also included studies with a multimodal stimulation, but only when they asked participants to focus on auditory processing. Studies or reported activations were excluded if they resulted from a simple comparison against a baseline or rest condition, except for PET studies reporting this comparison, because brain activity can be directly quantified here. For studies including nonhuman environmental sounds we also looked for imaging studies, which used a classical conditions paradigm including emotional sounds as the unconditioned stimulus (UCS) or as the conditioned stimulus (CS+), and which reported brain activity to either of these.

**Table 2**

List of studies included in the meta-analysis shown in Fig. 1.

Study	Method	Number of participants	Task characteristics	Contrasts
<b>(A) Nonhuman environmental sounds</b>				
Bach et al. (2008a)	fMRI	35	Passive listening	Rising > falling intensity
Buchel et al. (1999)	fMRI	11	Passive listening	CS+ > CS–
Knight et al. (2005)	fMRI	9	Passive listening	Activity associated with SCR production to CS+
Kumar et al. (2012)	fMRI	16	Pleasantness rating	Interaction between acoustic features; rating of unpleasantness
Mirz et al. (2000)	PET	12	Passive listening	Sound > baseline
Morris et al. (2001)	fMRI	6	Target face detection	UCS > (CS+ + CS–)
Viinikainen et al. (2012)	fMRI	17	Count inanimate/animate entities per block	Valence dependency
Zald and Pardo (2002)	PET	13	Passive listening	Aversive sound > white noise
<b>(B) Nonverbal expressions</b>				
Dietrich et al. (2007)	fMRI	16	Passive listening	affect bursts > vocal gesture; affect bursts > vegetative sound
Bestelmeyer et al. (2014)	fMRI	19	Emotional discrimination	Physical difference and perceptual difference between consecutive affective voices
Dietrich et al. (2008)	fMRI	16	Passive listening	Emotion > neutral
Fecteau et al. (2005)	fMRI	15	Gender discrimination	Emotion > neutral
Fecteau et al. (2007)	fMRI	14	Gender discrimination	Emotion > neutral; positive > neutral; negative > neutral; negative > positive; positive > negative
Meyer et al. (2005)	fMRI	12	Target sound detection	Laughter > speech; laughter > sounds
Morris et al. (1999)	PET	6	Gender discrimination	Emotion > neutral; fear > (happy, sad, neutral)
Phillips et al. (1998)	fMRI	6	Gender discrimination	Fear > neutral; disgust > neutral
Sander et al. (2003)	fMRI	13	Target pitch detection	ROI analysis: Laughing > silence; crying > silence; Laughing > crying; crying > laughing
Sander and Scheich (2005)	fMRI	20	Target pitch detection	ROI analysis: Laughing vs. crying, forward vs. backward
Sander et al. (2007)	fMRI	17	Target sine wave detection	ROI analysis: Laughing vs. crying, female vs. male participants; natural vs. unnatural
Szameitat et al. (2010)	fMRI	18	Emotional discrimination; laugh bouts count	Joy > tickle, taunt > tickle; tickle > joy; conjunction; explicit > implicit
Warren et al. (2006)	fMRI	20	Passive listening; voluntary smiling	Main effect of emotion; combined auditory and motor effects (i.e. smiling)
<b>(C) Affective prosody</b>				
George et al. (1996)	PET	13	Emotion discrimination; content discrimination; repeat word vocally (control task)	Content > control task; prosody > control task; prosody > content task
Alba-Ferrara et al. (2011)	fMRI	19	Emotion discrimination	Emotional > neutral; simple > complex emotion (pitch controlled);
Bach et al. (2008a)	fMRI	16	Emotion discrimination; gender discrimination	Implicit > explicit task; explicit > implicit task; emotion > neutral; anger > fear
Beaucousin et al. (2007)	fMRI	23	Emotion discrimination; grammar discrimination	Emotion > grammar task; emotion > neutral
Beaucousin et al. (2011)	fMRI	23	Emotion discrimination; grammar discrimination	Emotion > grammar task; task × listener gender
Buchanan et al. (2000)	fMRI	10	Emotional target detection; Target word detection	Sad > happy; happy > sad; emotion > phonemic task; phonemic > emotion task
Dietrich et al. (2008)	fMRI	16	Passive listening	Emotion > neutral; high > low lexicality
Ethofer et al. (2006a)	fMRI	24	Emotion discrimination; word discrimination	Emotion > content task; content > emotion task
Ethofer et al. (2006b)	fMRI	24	Emotion discrimination; word discrimination	Emotional intensity correlation
Ethofer et al. (2007)	fMRI	24	Passive listening	Emotion > neutral
Ethofer et al. (2009)	fMRI	24	Emotion discrimination; word class discrimination	Angry > neutral; emotion > word task
Frühholz et al. (2012)	fMRI	17	Emotion discrimination; gender discrimination	Angry > neutral; explicit: angry > neutral; Implicit: angry > neutral; task × emotion
Grandjean et al. (2005)	fMRI	15	Gender discrimination	Anger > neutral
Johnstone et al. (2006)	fMRI	40	Emotional face discrimination; Emotional voice discrimination	Happy > angry; angry > happy; happy × task; angry × task
Kotz et al. (2003)	fMRI	12	Emotion discrimination	Normal > prosodic speech; prosodic > normal speech; positive > neutral; negative > neutral
Kotz et al. (2013)	fMRI	20	Emotion discrimination	Angry > neutral; happy > neutral; sad > neutral; surprised > neutral; MVPA <sup>b</sup> : emotion
Kreifelts et al. (2010)	fMRI	24	Gender discrimination	Audiovisual > auditory/visual; correlation with trait emotional intelligence

Table 2 (Continued)

Study	Method	Number of participants	Task characteristics	Contrasts
Leitman et al. (2010)	fMRI	20	Emotion discrimination	Emotion > neutral; emotion $\times$ cue salience; anger $\times$ cue salience; happy/fear $\times$ cue salience
Mitchell et al. (2003)	fMRI	13	Passive listening; content discrimination; emotion discrimination	Congruent > neutral prosody; neutral > congruent prosody; congruent > incongruent prosody; incongruent > congruent prosody; emotion > semantic task; semantic > emotion task
Mitchell (2006)	fMRI	28	Emotion discrimination	Incongruent > prosody only; congruent > prosody only
Mitchell (2007)	fMRI	26	N-back task; Emotion discrimination; lexical discrimination	Increasing working memory load; working memory $\times$ listener gender
Mothes-Lasch et al. (2011)	fMRI	24	Gender discrimination; symbol discrimination	Explicit: angry > neutral; explicit > implicit task
Quadflieg et al. (2008)	fMRI	12 <sup>a</sup>	Emotion discrimination; gender discrimination	Angry > neutral
Rota et al. (2008)	fMRI	10	Emotion discrimination	Emotion > neutral; congruent > incongruent prosody
Sander et al. (2005)	fMRI	15	Gender discrimination	Anger > neutral; anger attended > neutral; anger unattended > neutral; anger attended > anger unattended; attention right > attention left; attention left > attention right
Schirmer et al. (2004)	fMRI	24	Word emotion discrimination; Emotional voice discrimination	Incongruent > congruent prosody
Schirmer et al. (2008)	fMRI	14	Passive listening	Incongruent $\times$ listener gender
Wiethoff et al. (2008)	fMRI	24	Emotion discrimination	Correlation with acoustic and arousal features
Wildgruber et al. (2002)	fMRI	12	Emotion discrimination; expressivity discrimination	Discrimination of expressiveness
Wildgruber et al. (2004)	fMRI	10	Emotion discrimination; linguistic discrimination	Emotion > linguistic task; linguistic > emotion task
Wildgruber et al. (2005)	fMRI	10	Emotion discrimination; vowel discrimination	Emotion > vowel task; vowel > emotion task
Wittfoth et al. (2010)	fMRI	20	Emotional voice discrimination	Incongruent > congruent prosody; emotion > neutral; positive > negative; negative > positive
(D) Musical emotions				
Baumgartner et al. (2006)	fMRI	24	Pleasantness rating, personal involvement rating	Audiovisual > visual; visual > audiovisual
Brattico et al. (2011)	fMRI	15	Familiarity rating; sad-happy rating; feeling sad-happy rating; like-dislike rating; pleasantness rating; ugly-beautiful rating	Sad > happy; happy > sad; music + lyrics > music; music > music + lyrics; sad music + lyrics > sad music; sad music > sad music + lyrics; happy music + lyrics > happy music; sad $\times$ lyrics; happy $\times$ lyrics
Blood et al. (1999)	PET	10	Valence rating; intensity rating	Regression with dissonance level; high dissonance > low dissonance; high dissonance > noise; low dissonance > noise
Blood and Zatorre (2001)	PET	10	Chill intensity rating	Regression with chill intensity; subject-selected > control music
Caria et al. (2011)	fMRI	14	Assess own emotional state prior to experiment; passive listening	Standard/favorite > control; favorite > standard
Chapin et al. (2010)	fMRI	21	Arousal rating; Valence rating	Expression > mechanical; experienced > inexperienced; correlation with tempo; correlation with arousal
Eldar et al. (2007)	fMRI	12	Arousal rating; valence rating	Negative: audiovisual > visual/audio; positive: audiovisual > visual/audio; neutral: audiovisual > visual/audio; Minor > major
Green et al. (2008)	fMRI	21	Like-dislike rating	Correlation with positive affect
Janata (2009)	fMRI	13	Familiarity rating; autobiographical salience rating; positive affect rating	
Khalifa et al. (2005)	fMRI	13	Emotion discrimination	Minor > major; mode $\times$ tempo
Koelsch et al. (2006)	fMRI	11	Tap beat; pleasantness rating	Unpleasant > pleasant; pleasant > unpleasant
Koelsch et al. (2008)	fMRI	24	Valence discrimination	Irregular > regular chord
Koelsch et al. (2013)	fMRI	18	Passive listening	Joy > fear; fear > joy; emotion $\times$ time
Lee et al. (2011)	fMRI	12	Catch trial detection	MVPA <sup>b</sup> : ascending/descending; MVPA <sup>b</sup> : major/minor
Lehne et al. (2014)	fMRI	25	Dynamic tension ratings	Music > rating; correlation with loudness; correlation with tension; dynamic tension > static tension; tension increase > tension decrease
Menon and Levitin (2005)	fMRI	13	Passive listening	Music > scrambled; scrambled > music



Table 2 (Continued)

Study	Method	Number of participants	Task characteristics	Contrasts
Mitterschiffthaler et al. (2007)	fMRI	16	Mood rating	Happy > neutral; sad > neutral; neutral > sad/happy
Mizuno and Sugishita (2007)	fMRI	18	Emotion discrimination	Major > neutral; minor > neutral; major > minor; minor > major
Muller et al. (2011)	fMRI	20	Passive listening	nonmanipulated > manipulated
Salimpoor et al. (2011)	fMRI/PET	10	Indicate chill	Pleasurable > neutral; anticipation > neutral; chill > neutral
Salimpoor et al. (2013)	fMRI	19	Music auction	Desirable > undesirable music; increasing reward value
Suzuki et al. (2008)	fMRI	13	Rate beauty of chord sequence	Consonant (major) > consonant (minor); consonant > dissonant
Trost et al. (2012)	fMRI	16	GEMS ratings <sup>a</sup> ; arousal rating; valence rating; familiarity rating	Music > control; correlation with 4 major emotional dimensions; correlation with arousal/valence/familiarity

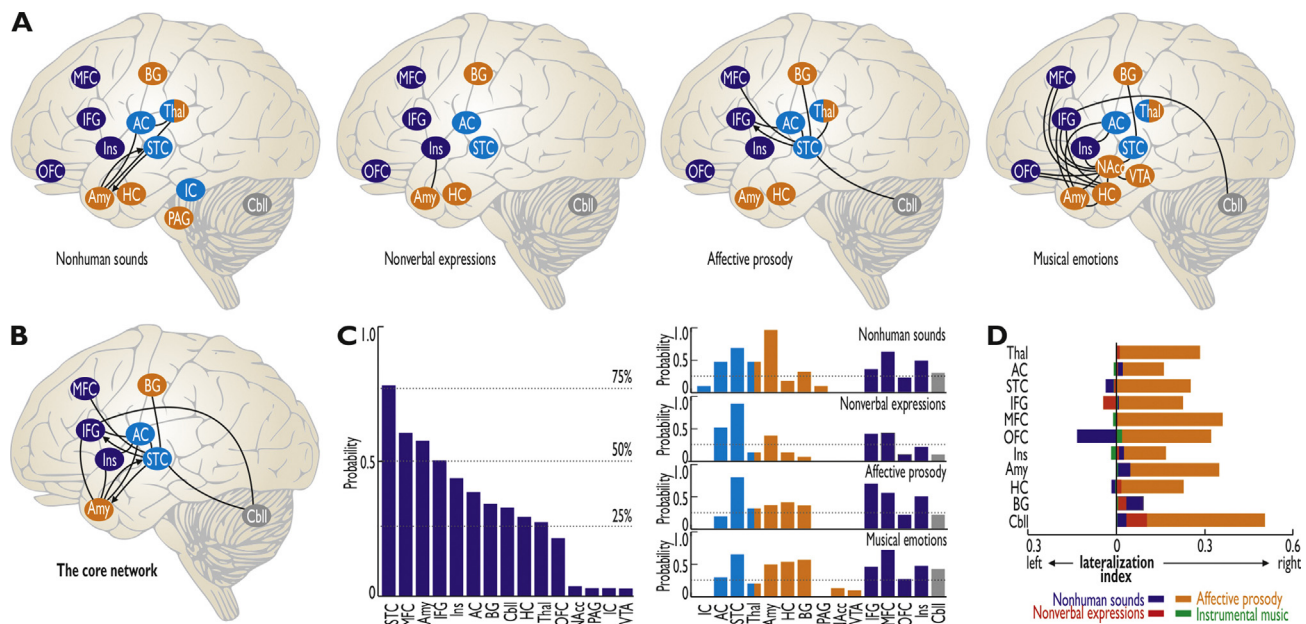
<sup>a</sup> Results were only included for the group of healthy participants.

<sup>b</sup> MVPA = multi-voxel pattern analysis; performed using a searchlight analysis (Kriegeskorte et al., 2006).

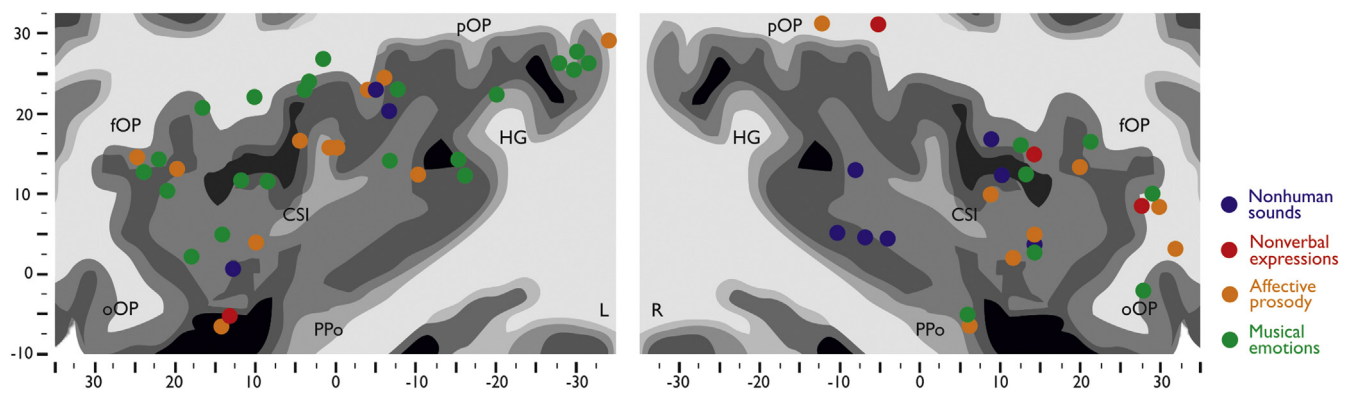
<sup>c</sup> Based on Zentner et al. (2008).

From the studies we included, peak activations from a diversity of contrasts between experimental conditions arose. Peak activations were included if they resulted from the comparison of affective to non-affective sounds or from dimensional affective ratings and analysis. We also included peak activations resulting from different experimental tasks and conditions. Next to explicit emotion discrimination tasks, we included peak activations from tasks and conditions that were largely orthogonal to the emotional dimension of sound, such as gender discrimination tasks (e.g. Frühholz et al., 2012; Grandjean et al., 2005), linguistic decision tasks (e.g. Ethofer et al., 2009; Wildgruber et al., 2005), irrelevant target detection tasks (e.g. Lee et al., 2011; Meyer et al., 2005; Morris et al., 2001), or (in-)congruence decisions (e.g. Schirmer

et al., 2004; Wittfoth et al., 2010). The rationale for including peak activations from these latter conditions and tasks was two-fold: first, several studies used different tasks to modulate the attentional orienting towards or away from the emotional dimension of sound. Even if attention is oriented away from the emotional meaning of sound, it can be assumed that this meaning is still processed at an “implicit” level (Vuilleumier, 2005). Second, we hypothesized that even if different tasks introduce variable neural effects, the effects of task variation may be more random compared to the common denominator of all studies presenting affective sounds, which are explicitly or implicitly processed by listeners. Our proposed consistency index should be largely sensitive to the neural effects of this common denominator and less to task variations.



**Fig. 2.** (A) Brain regions that have been most consistently involved in the processing of four different types of affective sounds, that is, nonhuman environmental sounds, human nonverbal expressions, emotional inflected speech (affective prosody), and music. The black lines indicate functional connections between brain regions if they were reported to underlie the decoding of emotional meaning from sounds. Arrows indicate unidirectional connections. Light blue regions belong to the ascending auditory system, dark blue regions correspond to fronto-insular regions, and orange regions belong to other subcortical structures. (B) The core network of brain regions, which is generally sensitive to several affective sounds. This core network shows a high interconnectivity, and connections are included if they were reported to underlie emotional decoding from any type of affective sound. (C) The core network was determined as those regions that showed consistent activity across all types of affective sounds (left panel), that showed activity in at least 25% (corresponding to the four different types of affective sounds) of all studies in the respective stimulus type (dashed line in the right panel), and which exceeded this threshold for at least 3 different types of affective sounds. (D) Laterality index for several regions of the core network and of the extended network, for which a reliable laterality distribution could be established. Abbreviations: AC auditory cortex, Amy amygdala, BG basal ganglia, Cbll cerebellum, IC inferior colliculi, IFG inferior frontal gyrus, HC hippocampus, NAcc nucleus accumbens, Ins insula, MFC medial frontal cortex, STC superior temporal cortex, Thal thalamus, VTA ventral tegmental area. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



**Fig. 3.** (A) Functional activity in the left (left panel) and right insula (right panel) for the different types of affective sounds in Montreal Neurological Institute (MNI) space. Every type of affective sound shows activity that spreads across the insula without a clear predominance of activity in the anterior or posterior insula for a certain type of sound. Abbreviations: CSI central sulcus of the insula, fOP frontal operculum, HG Heschl's gyrus, oOP orbital operculum, pOP parietal operculum, PPO planum polare.

By reviewing the functional activations reported in these studies we were able to identify core neural network for affective sound processing. Unsurprisingly, the core neural network involved in affective sound processing comprises the limbic and the auditory system, but extends to frontal and subcortical brain regions that are not typically and consistently included in previous domain-specific models (Fig. 2). Previous domain-specific models concerned either the processing of affective voices (Schirmer and Kotz, 2006; Wildgruber et al., 2009) or the processing of musical emotions (Koelsch, 2010, 2014), and they only marginally overlap in their description of underlying neural brain regions. Neurocognitive models of processing affective voices were so far designed as hierarchical feedforward models concerning the auditory-inferior frontal (Schirmer and Kotz, 2006) and auditory-amygdala-inferior frontal information flow (Wildgruber et al., 2009), while leaving several other important brain regions underspecified, such as the cerebellum, the basal ganglia, or more superior frontal regions. Existing neurocognitive models of processing musical emotions were designed as multidirectional neural model that were largely based on anatomical connectivity in the limbic system, especially of limbic parts of the temporal and frontal lobe, and more specifically of the limbic striatum (Koelsch, 2010, 2014). Models of processing musical emotions are sometimes underspecified in terms of the functional meaning of the neural limbic connections, and in terms of the important role of the auditory system (Koelsch, 2010). Similarly to affective voice processing models, these models also leave several other important brain regions underspecified. Finally, a common property of existing models of affective voice and music emotion processing is that they propose that different functional processes for the decoding of the emotional meaning of sounds are sharply separated and supported by single brain regions.

Based on the identified shortcomings of these domain specific models, and in an attempt to overcome these shortcomings, we propose a unifying model that considers that all affective sounds consistently induce activity in cortical and subcortical regions in a common core network. This common neural network is largely non-hierarchical with important functional connections between brain regions to share information and functions. Although these core network regions generally respond to the affective meaning of sounds, and this response is not confined to a simple acoustic analysis, their functional role in the decoding of the affective valence may be both *differential* and *complementary*. By differential we mean that a region has a functional role that is unique to this region, but nevertheless contributes to overall processing via its structural and functional connectivity to other regions. By complementary we imply that more than one region is supporting a function such that subroutines of this function are located in a single region.

In the following, we briefly outline the core regions of the neural network of affective sound processing and highlight and expand their functional role from a cross-domain perspective.

### 3.1. The amygdala and the auditory cortex—complementary regions for the decoding of the emotional meaning from sound

The limbic system has long been considered as central and uniquely important for the processing of affective stimulus valence (LeDoux, 2012), especially for affective sounds (Fröhholz et al., 2015b; Koelsch, 2014; Milesi et al., 2014) of different valence (Fig. 2A). The amygdala's involvement in the processing of affective sounds is quite consistent across all affective sound types, but with higher consistency for less complex affective sounds, such as non-human environmental sounds and nonverbal vocalizations, which may be based on their basic nature of affective signs representing an emotional “signal” or “symptom” to (automatically) trigger emotional responses (see above). However, less complex sounds (i.e. short high intensity sounds or aversive sounds) can also be a feature of more complex sounds (i.e. sudden acoustic changes in music) and thus contribute to emotional processing in response to more complex sounds (Trost et al., 2015).

The notion that less complex affective sounds consistently elicit amygdala activation seems to point to a fast afferent auditory signal transmission to the amygdala (Fröhholz et al., 2014; Pannese et al., 2015). This transmission may be accomplished via the medial geniculate nucleus (MGN) as part of the ascending auditory pathway (Viinikainen et al., 2012) anatomically projecting to the basolateral amygdala (Ball et al., 2007; Fröhholz and Grandjean, 2013a; LeDoux et al., 1990). The amygdala anatomically projects back to both subcortical and cortical auditory regions (Fröhholz et al., 2014). The MGN shows activity in response to complex sounds (Wenstrup, 1999), to conspecific vocalizations (Syka et al., 1997), and to affective communication signals (Cappe et al., 2009). Thus, the MGN can provide a reliable acoustic profile of affective sounds, and this acoustic profile may be transferred directly to the amygdala given the anatomical MGB-amygdala connections while bypassing the auditory cortex (Pannese et al., 2015).

It is generally assumed that the amygdala plays a central role in the decoding of affective sounds, but cortical regions within the auditory system seem to play a similarly important role. Cortical activation is reported in primary and secondary auditory cortex (AC) and in the superior temporal cortex (STC) (Fröhholz and Grandjean, 2013b; Koelsch, 2014; Kumar et al., 2012; Szameitat et al., 2010). Surprisingly, compared to the amygdala, the auditory cortex more consistently responds to emotional sounds and the total number of studies reporting STC and AC activity in response

to affective sounds is more than two times higher than reports on amygdala activity.

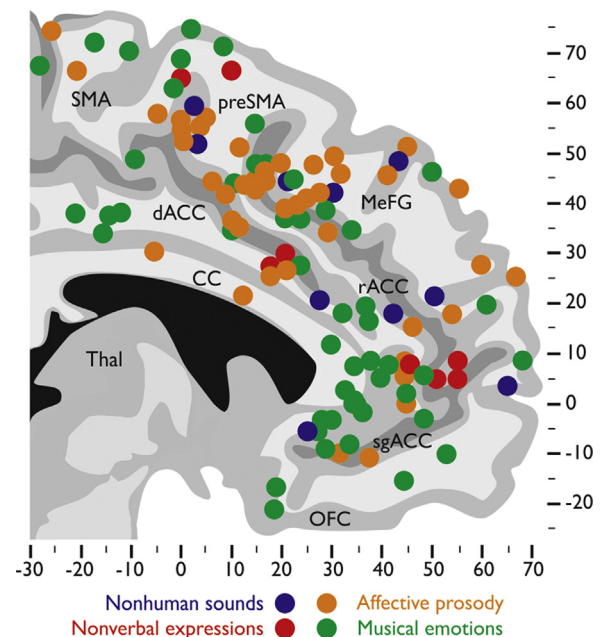
This difference supports the notion of multiple routes engaged in emotional processing and that sensory cortical areas can be similarly important for emotional processing as regions of the limbic system (Pessoa and Adolphs, 2010). While increased activation in sensory cortical regions has been equated with increased sensory processing, and may be remotely driven by increased emotional decoding in the amygdala (Vuilleumier et al., 2004), one could also think of genuine and complementary emotion processing in the auditory cortex (Pessoa and Adolphs, 2010, 2011) that is different from affective sound processing in the amygdala. This difference is apparent in studies that reported enhanced activity in AC without a corresponding signal increase in the amygdala, especially for vocal emotions (Alba-Ferrara et al., 2011; Beaucousin et al., 2011; Dietrich et al., 2007, 2008; Ethofer et al., 2006a,b, 2007; Kotz et al., 2013, 2003; Meyer et al., 2005; Mitchell, 2006, 2007; Mitchell et al., 2003; Rota et al., 2008; Warren et al., 2006; Wiethoff et al., 2008; Wildgruber et al., 2004, 2005; Wittfoth et al., 2010) and musical emotions (Blood et al., 1999; Brown et al., 2004; Caria et al., 2011; Chapin et al., 2010; Janata, 2009; Lee et al., 2011; Mueller et al., 2011; Suzuki et al., 2008; Trost et al., 2012). It is also worth noting that this enhanced activity in AC is usually reported for affective relative to neutral sounds or relative to other affective sounds (Frühholz and Grandjean, 2013b). It may thus reflect generic affect processing beyond a simple acoustic analysis of sound.

The central role of the STC in processing affective cues from sound may be related to the specific nature of sound. Sound evolves over time, and so does relevant acoustic information of these sounds. Relevant sound information thus needs to be integrated in time (Koelsch, 2014; Pell and Kotz, 2011), which is especially relevant when listening to music (Salimpoor et al., 2011). A central role of the STC may therefore be to register the unfolding of relevant acoustic affective cues over time, and to integrate this information in form of an auditory percept (Boemio et al., 2005; Schonwiesner et al., 2005). This information may be shared with the amygdala (Kumar et al., 2012), but also with frontal brain regions (Morris et al., 1999) via functional connections.

### 3.2. Mediation and evaluation in the fronto-insular brain system

Several frontal brain regions also strongly respond to affective sounds. For example, the inferior frontal cortex (IFC) is an integral part of a fronto-temporal network involved in the processing of affective sounds in social communication (Friederici, 2012; Frühholz and Grandjean, 2013c; Rauschecker and Scott, 2009). Complementary to the STC, the IFC is thought to integrate emotionally relevant sound features provided by the STC via dorsal and ventral connections (Frühholz and Grandjean, 2012; Frühholz et al., 2015a; Glasser and Rilling, 2008; Saur et al., 2008; Schirmer and Kotz, 2006). These connections support the categorization of sounds according to their social meaning and affective value (Frühholz and Grandjean, 2012; Hoekert et al., 2008, 2010; Rauschecker and Scott, 2009). The IFG is also involved in adaptive response preparation in a perception-action cycle (Frühholz et al., 2016; Rauschecker and Scott, 2009) due to its vicinity to the primary motor cortex that controls vocal motor plans (Schubotz et al., 2003).

Next to the IFC, the insula (Bamiou et al., 2003) also responds to most types of affective sounds (Frühholz et al., 2012; Kumar et al., 2012; Sander and Scheich, 2005; Trost et al., 2012). The insula has been proposed to have a differential role in its anterior and posterior part (Taylor et al., 2009), such that the anterior insula represents an emotional salience detector and the posterior part underlies environmental monitoring and supports response selection (Fig. 3). However, no clear anterior-to-posterior distinction

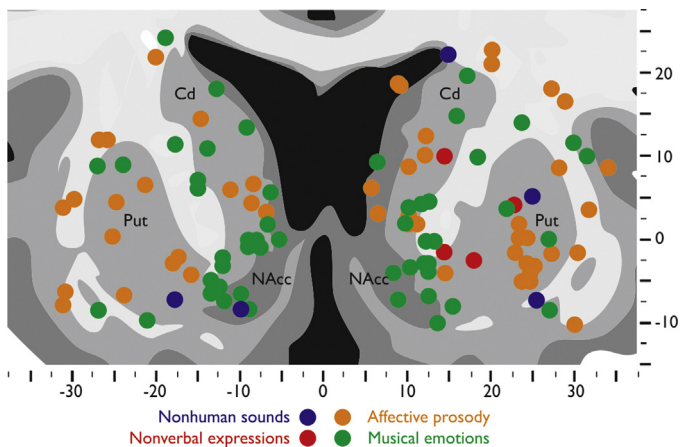


**Fig. 4.** Functional activity in the medial frontal cortex (MFC) for the different types of affective sounds in Montreal Neurological Institute (MNI) space. Every type of affective sound shows activity that spreads across the whole medial frontal cortex. However, activity for affective prosody seems to be denser in the dACC and the MeFG, whereas activity for musical emotions seems to be denser in the rACC and the sgACC. Abbreviations: CC corpus callosum, dACC dorsal anterior cingulate cortex, MeFG medial frontal gyrus, OFC orbitofrontal cortex, preSMA presupplementary motor area, SMA supplementary motor area, rACC rostral anterior cingulate cortex, sgACC subgenual anterior cingulate cortex, Thal thalamus.

was found for the processing of different types of affective sounds, indicating that the insula seems both involved in a general environmental monitoring and the salience detection of affective sounds or sound features. Besides its monitoring and salience detection functions, the insula has also been proposed as a mediator between sensory and affective brain systems in the perception of affective sounds (Mirz et al., 2000). This mediation may translate affective cue perception of sounds into a self-experience of emotions (Kotz et al., 2013; Wildgruber et al., 2004) when sounds are evaluated (Mothes-Lasch et al., 2011). The self-experience of emotions in terms of a subjective feeling is a general and important component of an emotional episode. This transposition of the auditory perception of sounds into a subjective feeling in the insula seems to apply to a variety of sounds. While the insula seems to be most responsive to aversive sounds (Buchel et al., 1999; Knight et al., 2005; Mirz et al., 2000; Zald and Pardo, 2002) and vocalizations of disgust (Calder, 2004; Calder et al., 2000; Phillips et al., 1998), it also responds to pleasant and self-selected music (Blood and Zatorre, 2001; Brown et al., 2004; Koelsch et al., 2006; Menon and Levitin, 2005; Salimpoor et al., 2013) including music with lyrics (Brattico et al., 2011).

Finally, studies reported activity in several subregions of the medial frontal cortex (MFC) (Fig. 4). The MFC supports several social and emotional functions related to interpersonal communication and understanding (Amodio and Frith, 2006). The dorso-caudal part may serve emotional appraisal and evaluation processes (Bach et al., 2008a; Ethofer et al., 2009; Szameitat et al., 2010; Wildgruber et al., 2009), whereas the ventro-rostral part may support the retrieval of emotional memories and associations (Euston et al., 2012; Trost et al., 2012) as well as the regulation of adaptive emotional behavior (Etkin et al., 2011; Euston et al., 2012; Frühholz et al., 2010; Garcia et al., 1999). All types of affective sounds seem to elicit activity in all subregions of the MFC. However, activity in





**Fig. 5.** (B) Functional activity in the BG and the NAcc. Peak activations show a wide distribution across the dorsal and ventral basal ganglia for all types of affective sounds, while only activations for musical emotions additionally show some unique and dense peak activity in the NAcc. Abbreviations: CC corpus callosum, Cd caudate nucleus, dACC dorsal anterior cingulate cortex, MeFG medial frontal gyrus, NAcc nucleus accumbens, OFC orbitofrontal cortex, preSMA presupplementary motor area, Put putamen, SMA supplementary motor area, rACC rostral anterior cingulate cortex, sgACC subgenual anterior cingulate cortex, Thal thalamus.

response to affective prosody seems to be denser in the dorso-caudal part of the MFC, while musical emotions seem to have stronger peak activity in the ventro-rostral part of the MFC. This structural distinction may indicate a differential response in the MFC to different types of affective sounds. Emotional cues from speech are predominantly subjected to increased evaluation in appraisal, while music may predominantly relate to regulatory processes and enrichment by memory-associated emotional processes (Blood and Zatorre, 2001; Panksepp, 1995; Panksepp and Bernatzky, 2002). However, both functional processes may also be relevant for any other sound type.

### 3.3. Temporal prediction and response adaptation in the basal ganglia and the cerebellum

Musical emotions not only elicit activity in the MFC, but also in the ventral striatum of the basal ganglia (BG) (Fig. 5). Activity in the ventral striatum and especially in the nucleus accumbens (NAcc) suggests that listening to music is a pleasurable and rewarding experience (Koelsch, 2014) based on dopamine release controlled by the ventral tegmental area (VTA) (Chanda and Levitin, 2013). Within the ventral striatum, the NAcc is almost exclusively active when listening to music, and thus seems to play a specific role in inducing feelings of pleasure while listening to enjoyable music. Nevertheless, it is conceivable that, for example, laughter also leads to pleasurable feelings in a listener in a social context, but evidence on this topic is currently sparse. Besides music, ventral BG involvement in emotional processing from affective voices comes from studies in psychiatric and neurologic patients, leading to mood disorders (Paradiso et al., 2013), or to deficits in processing emotional stimuli (Paulmann et al., 2005, 2008; Pell and Leonard, 2003; Peron et al., 2011). However this ventral BG activity was located outside the NAcc, and thus might only be indirectly linked to pleasurable feelings when listening to voices.

Unlike the ventral BG, the dorsal BG contribute to the temporal decoding and anticipation of acoustic events (Kotz and Schwartz, 2010; Kotz et al., 2009; Salimpoor et al., 2011). The dorsal BG may decode temporal variations and patterns of acoustic features especially in affective prosody that benefit the decoding of emotional meaning (Grahn and Brett, 2007; Hass and Herrmann, 2012), but may also represent temporal anticipations of a musical chill

(Salimpoor et al., 2011). The dorsal BG may functionally orchestrate the anticipation and decoding of temporal patterns in concert with the auditory cortex (Ethofer et al., 2012; Frühholz and Grandjean, 2012; Salimpoor et al., 2013), but also with the cerebellum (Cbll) especially concerning vocal affective patterns (Ethofer et al., 2012; Kotz et al., 2013) and music (Tolgyesi and Evers, 2014; Trost et al., 2012).

So far, the Cbll has been largely neglected in all domains of affective sound processing. However, it seems particularly sensitive to discrete temporal events and sudden sound changes (Schwartz and Kotz, 2013) leading to uncontrolled, tensed, and defensive motor responses when hearing aversive sounds (Zald and Pardo, 2002), but also to an urge to move when listening to music (Janata et al., 2012; Trost et al., 2014). The functional role of the Cbll thus may be to transpose the perception of sound into an emotional output behavior, which is another major component of an emotional episode next to subjective feelings. This transposition might be especially relevant for some “automatic” and uncontrolled emotional responses to affective sounds that represent simple signals in terms of our proposed typology. We especially classified technical and environmental sounds as signals given that they can automatically induce emotional responses mediated by the Cbll, and some of these can be classified as “emotional reflexes”, such as the startle response to loud and aversive sounds (Zald and Pardo, 2002). The Cbll accordingly also showed the highest activations consistency for nonhuman sounds. Given this evidence for the functional roles of the BG and Cbll, both seem to be involved in controlled and uncontrolled emotional output behavior that is supported by the close link between the sensorimotor and affective components of the BG (Pichon and Kell, 2013). Thus, the functional role of the BG and the Cbll in affective sound processing can be considered two-fold: first, temporal prediction and integration of affective sound and, second, the integration of perceptual decoding and behavioral response preparation.

So far we have described a core network of brain regions underlying affective sound processing. This core network involves regions, whose contributions to affective sound processing are well known, but needs to be extended to brain regions (e.g. Cbll) that are traditionally reported outside the context of affective sound processing. These regions ascribe to different functions in the emotional decoding of sound. Some functions are differential (e.g. pleasure in the ventral BG) or complementary (e.g. emotional decoding in the amygdala and auditory cortex, sound integration in the STC, dorsal BG, and IFC), and some functions are split within an area (e.g. appraisal/evaluation in the dorsal MFC, regulation/memory retrieval in the ventral MFC). Altogether this points to an extended core neural network for affective sound processing, which we will elaborate on below.

## 4. An extended brain network for complex affective sounds

A short excursion is warranted before we describe an integrative network of affective sound processing. Next to the described “core” network that shows activity in response to nearly all affective sound types, there are distinct brain regions that only respond to specific types of affective sounds, such as the hippocampus (HC), the orbitofrontal cortex (OFC), and most likely the NAcc. We propose that these brain regions represent orbits of an extended neural network for affective sound processing. Further empirical and cross-domain evidence will have to show whether these orbits indeed reflect domain specificity or domain-general. As outlined below, these regions predominantly respond to complex affective sounds, such as affective prosody and especially to musical emotions, which may be related to their more complex nature in terms of an affective sign at the level of “symbol”.

Although the hippocampus (HC) and the parahippocampal gyrus (PHG) generally respond to all types of affective sounds (Fig. 2), they primarily show consistent activity in response to affective prosody and to musical emotions (Frühholz et al., 2014). The contribution of the HC has only rarely been found for the processing of environmental sounds (Mirz et al., 2000) and emotional voices (Frühholz et al., 2014), but is supposed to play a generic role in the processing of musical emotions (Koelsch, 2010, 2014; Trost et al., 2012) associated with its role in the regulation of the hypothalamic-pituitary-adrenal (HPA) axis in emotional episodes (Jacobson and Sapolsky, 1991) and in the retrieval of emotional musical associations (Frühholz et al., 2014). These functions may also apply to the regulation of adaptive behavior while listening and responding to affectively inflected speech. The latter is a socially complex stimulus that may require the retrieval of episodic or semantic memory associations, because their accurate interpretation often depends on social or cultural contexts, in which individuals were socialized and which qualifies them as “symbol” in terms of an affective sign.

These HC functions may be supported by connections of the HC to the OFC (Koelsch et al., 2013; Salimpoor et al., 2013). A number of studies reported consistent activation of the OFC for musical emotions (Lehne et al., 2014; Salimpoor et al., 2013; Trost et al., 2012), but less to other types of sound. The OFC especially encodes secondary and learned values (Kringelbach, 2004, 2005), and the latter may be primarily encoded (Sescousse et al., 2013, 2010) and valued in music (Blood and Zatorre, 2001; Blood et al., 1999; Lehne et al., 2014; Mas-Herrero et al., 2014; Salimpoor et al., 2013; Suzuki et al., 2008) as a complex social and cultural symbol. These regions preferentially respond to complex sounds during evaluation, as observed in music studies. Future studies thus may determine whether these regions are also responsive to other types of affective sounds by applying a research paradigm (i.e. the pleasure of listening to sounds) that is dominant in one domain (music) to another domain (sound and voice processing).

## 5. A neurocognitive model for the distributed processing of affective sounds

Above we summarized evidence in favor of a generalized neural network underlying affective sound processing and we proposed that additional network nodes need to be integrated into this network. We put forward a unifying neural network, consisting of a core and extension regions that aims at integrating similarities and differences across research domains on affective sound processing under one hooding (Fig. 6). The relative weight of certain regions and functions in the core network, however, dynamically changes with variability in quality and complexities of affective sounds, but the core network provides a common ground most probably for the immediacy of emotional reactions to affective sound types. Furthermore, this cross-domain perspective also expands and alters the functional description of these brain regions by highlighting the functional role of so far neglected brain regions.

The core network comprises the amygdala as one of the prime areas in the limbic system (Dalglish, 2004; Pessoa and Adolphs, 2011; Popescu et al., 2009; Sergerie et al., 2008; Shabel and Janak, 2009) that specifically responds to brief sounds such as affective burst and sound features that can immediately signal relevance and elicit emotional responses (Bach et al., 2008b; Zald and Pardo, 2002). We consider that the auditory cortex plays an equivalent and complementary role to the amygdala, but decodes more complex sounds and sound features that evolve in time (Frühholz et al., 2014). The auditory cortex seems to integrate emotional information over time (Frühholz et al., 2014), with the left and right auditory cortex integrating acoustic information of finer and larger time-scale, respectively (Zatorre and Belin, 2001). The emotional

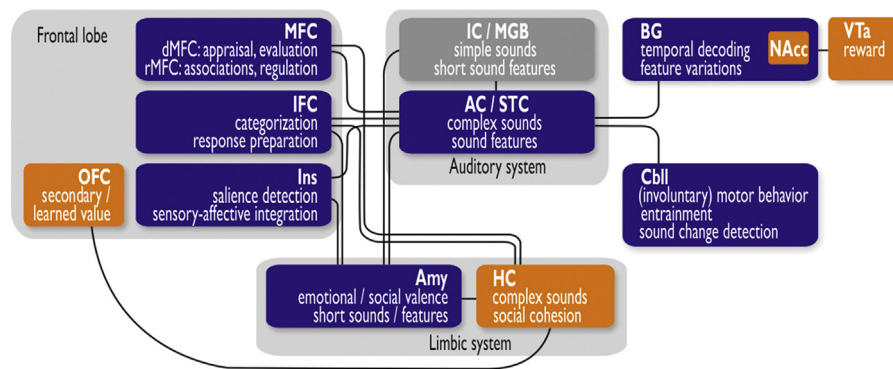
decoding in the amygdala and in the auditory cortex does not have to be exclusive as there are functional connections between both regions (Kumar et al., 2012; Viinikainen et al., 2012). More specifically, we propose that both regions should be generically involved in sound-based emotional processing, but their relative weight depends on the specific sound type. For example, musical emotions acoustically evolve over a large and variable time scale, and thus may primarily activate the auditory cortex during the integration of affectively sound information. However, sometimes, short musical features such as sudden changes, evoke strong emotions. These features seem more likely to be decoded in the amygdala as they do for nonhuman sounds, which also activate the amygdala.

In addition to this emotional decoding in the amygdala and the auditory cortex, we propose different functional roles for regions in the frontal cortex that may be accomplished in concert with the previously described regions. Regions in the frontal lobe (MFC, IFC, Ins) show functional connectivity both with the amygdala (Morris et al., 1999) and the auditory cortex (Ethofer et al., 2006a, 2012; Frühholz and Grandjean, 2012; Koelsch et al., 2013) during processing of emotional meaning from different sound types, and show highly consistent activity within the core network. The insula may play a specific role in salience detection and relates sensory acoustic information to an internal emotional representation of affective sound (Menon and Uddin, 2010), while other regions of the lateral and medial frontal cortex evaluate and appraise affective sound and enrich these representations with (episodic) memory associations. A general role of specific subregions of the frontal cortex is higher-order representation and evaluation of affective sounds that may likely support emotional categorization and adaptive behavioral preparation in the IFC in response to these sounds.

Thus, regions in the frontal cortex have rather differential functions in affective sound processing, but some of these functions may also be complementary to functions in other regions. For example, the amygdala may also serve as a salience detector of meaningful sound information and sound features (Armony and LeDoux, 1997), but may do so especially for novel salient events (Blackford et al., 2010), while the insula may respond to and integrate salient acoustic sound information in general. Furthermore, we propose that the auditory cortex may temporally accumulate and integrate relevant affective sound features as they are perceived over time (Boemio et al., 2005; Schonwiesner et al., 2005). In contrast, the IFC may integrate selected sound information that is relevant for an accurate emotional classification of sound independent of their time course (Frühholz and Grandjean, 2013c). The integration of sound information in the IFC might happen in terms of weighting available dynamic and static sound information in preparation for adaptive behavior. As such, the IFC is sensitive to the acoustic profile of vocalizations (Romanski et al., 2005) and can differentiate vocalizations according to behavioral relevance (Gifford 3rd et al., 2005).

Complementary to the auditory cortex and the IFC, the dorsal BG, functionally connected to auditory cortex (Ethofer et al., 2012), are sensitive to temporal patterns in sound and thus provides temporal predictions, serving as acoustic cue to the emotional meaning of a sound (Banse and Scherer, 1996; Juslin and Laukka, 2003; Patel et al., 2011; Salimpoor et al., 2011; Weninger et al., 2013). The dorsal BG appear to anticipate pleasurable events, especially in music, causing an increase in activity in the VTA-NAcc system. So far, the latter has been shown exclusively in music listening, but may inspire future studies investigating other affective sound types, for example, the potential of voices or to nonhuman sounds to induce pleasurable responses.

Our cross-domain perspective also identified brain regions, which have been largely neglected due to strong within-domain positions. The most important example is the Cbll, which has been only discussed in nonhuman sound processing (Zald and Pardo,



**Fig. 6.** Based on this core network we propose a unifying neural model for the processing of affective sounds in the core (blue) and the extended brain network (orange). The regions of the core network are essential for the decoding of the affective value from many different sounds, and have either differential or complementary functional roles. The extended network consists of the OFC, the HC, and probably also the reward system (VTA-NAcc system), and seems to be important for the decoding of the affective value only from complex sounds (i.e. affective prosody and predominantly musical emotions). Additionally to the core and extended network, some subcortical regions of the auditory system (IC, MGB) might receive more attention and evidence in future studies along with advancements in neuroimaging methods (dark grey). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2002). The Cbll may engage in behavioral responses to affective sounds, especially regarding involuntary motor responses. Motor responses to aversive sounds typically involve defensive motor reactions and increased muscle tension (Zald and Pardo, 2002). This points to a role of the Cbll in executing adaptive responses to affective sounds and sound features (Laricchiuta et al., 2014).

Above we described a neurocognitive model for the distributed processing of affective sounds in a core and an extended network of brain regions. We proposed complementary and differential functional roles for each of these brain regions that was inspired by recent neuroimaging studies on affective sound processing. However, we note that some of the proposed functions may still be speculative as comprehensive evidence from patient and TMS studies is scarce. We hope that our proposed model and framework nevertheless is fruitful and inspires future research exploring the functional role of specific brain regions in this concerted network.

## 6. Hemispheric lateralization of affective sound processing

A long lasting debate in auditory emotional neuroscience concerns the hemispheric lateralization of emotion processing. Hence, a final notion concerns the lateralization of functions in the core network identified for affective sound processing. As a rough estimate of laterality we quantified the relative number of peak activations reported in the left and the right hemisphere for each type of affective sound (Fig. 1D). This laterality index does not determine the strength of left and right hemispheric activity, but only takes the number of reported peak activations as a measure of lateralization strength. For most types of affective sounds, one only finds marginal support of lateralization such as a stronger involvement of the right BG for environmental sounds, and the right Cbll for non-verbal expressions. This indicates that both the left and the right hemisphere are equally involved in decoding the affective meaning from sounds at all functional levels across an extended brain network. However certain subfunctions of this processing may be lateralized to one hemisphere. For example, it has been recently suggested that bilateral IFC receives acoustic information especially about the dynamic variation of certain sound features (e.g. pitch) from the STC, whereas left IFC mainly receives information about fast features variations and right IFC receives information about slowly varying features with a high range of variability (Frühholz and Grandjean, 2013c).

Besides this general relevance of both hemispheres in the processing of affective sounds, affective prosody processing seems to be different. Affective prosody seems to be strongly lateralized to

the right hemisphere across almost all discussed brain regions, except for BG, which shows similar activity across hemispheres. The observed lateralization suggests two important neural characteristics for the processing of affective prosody. First, as indicated by recent patient studies, right brain lesions lead to more severe impairments in the processing of affective prosody compared to left brain lesions (Pell, 2006; Ross and Monnot, 2008), pointing to a predominance of the right hemisphere. Second, it also indicates that the right hemisphere may be more diversified even at a local level for specific brain regions, which would lead to higher local peak activity in the right hemisphere (Frühholz et al., 2012).

## 7. Conclusions

Previous neural models have provided domain-specific views on neural network mechanisms underlying the decoding of the emotional meaning from affective sounds (Koelsch, 2010; Schirmer and Kotz, 2006). Accordingly, while the importance of the auditory cortex and partly of the amygdala is acknowledged by all previous models, they overlap only minimally in terms of other neural brain system and functions that are grouped around these two major network components. However, the comprehensive review of evidence from different domains provided here allows the unification of diverse network descriptions to a common core neural network that is able to provide all functional mechanisms necessary for the decoding of the emotional meaning from sounds. This common neural network provides a neural “toolkit” for functional mechanisms that are used to process any type of affective sound.

Thus, unlike suggested by recent models proposing that the same functional mechanism (i.e. decoding of the emotional meaning) is similarly accomplished by distinct neural networks for each sound type separately, we put forward a common neural network view. The latter suggests multiple functional mechanisms that are distributed across this common neural network. This view points to a sparse neural decoding, because functional mechanisms in a common neural network are shared for the affective processing of sounds. This sparse coding is reasonable, first, given the similarities in functional processing requirements. Second, from a sparse coding point of view some functional mechanisms of certain brain areas seem more diverse than suggested in domain-specific models. Third, a sparse neural network usually also displays larger learning diversity (Gripone and Berrou, 2011) and the ability to transfer and generalize network performance (Awano et al., 2011), which may be reflected in the common neural network that underlies the processing of all types of affective sounds. This sparse neural



network view thus also has the potential to generalize to certain domains of affective sound processing, which so far have been neglected in this research, such as the neural processing of artificial affective sounds (Beaucousin et al., 2007).

Based on this common neural network, we propose a unifying neural network view on affective sound processing both in terms of sounds as auditory signs for emotional meaning and in terms of their underlying neural network. This unifying view highlights three key aspects. First, rather than a single brain region or brain system we proposed a core neural network underlying any type of affective sound processing, but the relative weight of specific regions changes across the different sound types. The functional roles of brain regions in this network can be differential or complementary. For example, we identified relevant complementary roles of the amygdala and the auditory cortex. The amygdala plays a central role in the decoding of the affective meaning from sound, but the auditory cortex seems to be an equivalent and complementary player in affective sound decoding beyond its role of an acoustic sound analysis (Pessoa and Adolphs, 2011). This corroborates recent suggestions about multiple roads to the neural decoding of emotional meaning from sensory stimuli (Pessoa and Adolphs, 2010, 2011).

Second, the unifying neural network view highlights the importance of partly underspecified brain regions involved in affective sound processing. For example, even though the Cbll has at times been associated with normal and dysfunctional emotional processing (Schmahmann and Sherman, 1998; Strata et al., 2011; Tomlinson et al., 2013; Villanueva, 2012), it has not been considered to be an integral part of emotion processing. However, the consistent evidence for Cbll activity in response to affective sounds reported and reviewed here highlight its essential role in coordinating motor output behavior in response to affective sounds, which seems also relevant for the processing of nonauditory emotional stimuli (Adamaszek et al., 2014; Ferrucci et al., 2012; Laricchiuta et al., 2014; Schutter et al., 2009).

Third, a cross-modal view on the neural processing of affective sounds may point to commonalities across acoustic domains, also given that acoustic cues to emotions in sounds are often similar or are often mimicked across domains (Weninger et al., 2013). This cross-modal view should inspire future research by encouraging researchers to pose cross-domain questions and to transfer research paradigms across domains. For example, the perspective of reward and pleasure while listening to enjoyable music may also be experienced when listening to positive nonhuman sounds or vocal expressions. Especially the latter are a major medium of auditory social communication, which conveys explicit and implicit social reward.

Finally, an important question concerns the specificity of the proposed neural network model for the processing of auditory emotions as individuals also perceive emotions in the visual modality (Pessoa and Adolphs, 2011), especially from facial expression in social interactions. Neural models of (emotional) face processing similarly highlight the important role of the amygdala, the insula, and the IFC for decoding emotions from faces (Adolphs, 2002; Duchaine and Yovel, 2015; Haxby et al., 2000). This is also confirmed by a recent meta-analysis on emotional face processing, which also points to contributions of the Cbll and the BG (Fusar-Poli et al., 2009). Thus, parts of the common neural network proposed here that extend beyond sensory auditory areas, may also play an important role in processing emotions from other sensory modalities. An exact comparison of sensory-specific contributions in the proposed network beyond primary and secondary sensory areas should be determined in future research. A cross-modality perspective may specifically determine the relative functional weight or

importance of several brain regions involved in the processing of emotions from one modality compared to another.

## Acknowledgments

S.F. was supported by a grant from the Swiss National Science Foundation (SNSF PP00P1.157409/1). We thank Cristina Soriano for helpful comments on the manuscript.

## References

- Adamaszek, M., D'Agata, F., Kirkby, K.C., Trenner, M.U., Sehm, B., Steele, C.J., Berneiser, J., Strecker, K., 2014. Impairment of emotional facial expression and prosody discrimination due to ischemic cerebellar lesions. *Cerebellum* 13, 338–345.
- Adolphs, R., 2002. Neural systems for recognizing emotion. *Curr. Opin. Neurobiol.* 12, 169–177.
- 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, e28701.
- Amodio, D.M., Frith, C.D., 2006. Meeting of minds: the medial frontal cortex and social cognition. *Nat. Rev. Neurosci.* 7, 268–277.
- Armony, J.L., LeDoux, J.E., 1997. How the brain processes emotional information. *Ann. N. Y. Acad. Sci.* 821, 259–270.
- Awano, H., Nishide, S., Arie, H., Tani, J., Takahashi, T., Okuno, H., Ogata, T., 2011. Use of a sparse structure to improve learning performance of recurrent neural networks. In: Lu, B.-L., Zhang, L., Kwok, J. (Eds.), *Neural Information Processing*. Springer, Berlin Heidelberg, pp. 323–331.
- Bach, D.R., Grandjean, D., Sander, D., Herdener, M., Strik, W.K., Seifritz, E., 2008a. The effect of appraisal level on processing of emotional prosody in meaningless speech. *Neuroimage* 42, 919–927.
- Bach, D.R., Schachinger, H., Neuhoﬀ, J.G., Esposito, F., Di Salle, F., Lehmann, C., Herdener, M., Scheﬄer, K., Seifritz, E., 2008b. Rising sound intensity: an intrinsic warning cue activating the amygdala. *Cereb. Cortex* 18, 145–150.
- Bachorowski, J.A., Owren, M.J., 2001. Not all laughs are alike: voiced but not unvoiced laughter readily elicits positive affect. *Psychol. Sci.* 12, 252–257.
- Ball, T., Rahm, B., Eickhoﬀ, S.B., Schulze-Bonhage, A., Speck, O., Mutschler, I., 2007. Response properties of human amygdala subregions: evidence based on functional MRI combined with probabilistic anatomical maps. *PLoS One* 2, e307.
- Bamiou, D.E., Musiek, F.E., Luxon, L.M., 2003. The insula (Island of Reil) and its role in auditory processing: literature review. *Brain Res. Rev.* 42, 143–154.
- Banise, R., Scherer, K.R., 1996. Acoustic profiles in vocal emotion expression. *J. Pers. Soc. Psychol.* 70, 614–636.
- Baumgartner, T., Lutz, K., Schmidt, C.F., Jancke, L., 2006. The emotional power of music: how music enhances the feeling of affective pictures. *Brain Res.* 1075, 151–164.
- Beaucousin, V., Lacheret, A., Turbelin, M.-R., Morel, M., Mazoyer, B., Tzourio-Mazoyer, N., 2007. fMRI study of emotional speech comprehension. *Cereb. Cortex* 17, 339–352.
- Beaucousin, V., Zago, L., Herve, P.Y., Strelnikov, K., Crivello, F., Mazoyer, B., Tzourio-Mazoyer, N., 2011. Sex-dependent modulation of activity in the neural networks engaged during emotional speech comprehension. *Brain Res.* 1390, 108–117.
- Bestelmeyer, P.E.G., Maurage, P., Rouger, J., Latinus, M., Belin, P., 2014. Adaptation to vocal expressions reveals multistep perception of auditory emotion. *J. Neurosci.* 34, 8098–8105.
- Blackford, J.U., Buckholz, J.W., Avery, S.N., Zald, D.H., 2010. A unique role for the human amygdala in novelty detection. *Neuroimage* 50, 1188–1193.
- Blood, A.J., Zatorre, R.J., 2001. Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *Proc. Natl. Acad. Sci. U. S. A.* 98, 11818–11823.
- Blood, A.J., Zatorre, R.J., Bermudez, P., Evans, A.C., 1999. Emotional responses to pleasant and unpleasant music correlate with activity in paralimbic brain regions. *Nat. Neurosci.* 2, 382–387.
- Boemio, A., Fromm, S., Braun, A., Poeppel, D., 2005. Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nat. Neurosci.* 8, 389–395.
- Brattico, E., Alluri, V., Bogert, B., Jacobsen, T., Vartiainen, N., Nieminen, S., Tervaniemi, M., 2011. A functional MRI study of happy and sad emotions in music with and without lyrics. *Front. Psychol.* 2, 308.
- Brown, S., Martinez, M.J., Parsons, L.M., 2004. Passive music listening spontaneously engages limbic and paralimbic systems. *Neuroreport* 15, 2033–2037.
- 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. *Cogn. Brain Res.* 9, 227–238.
- Buchel, C., Dolan, R.J., Armony, J.L., Friston, K.J., 1999. Amygdala-hippocampal involvement in human aversive trace conditioning revealed through event-related functional magnetic resonance imaging. *J. Neurosci.* 19, 10869–10876.
- Calder, A.J., Keane, J., Manes, F., Antoun, N., Young, A.W., 2000. Impaired recognition and experience of disgust following brain injury. *Nat. Neurosci.* 3, 1077–1078.



- Calder, A.J., 2004. Impaired recognition of anger following damage to the ventral striatum. *Brain* 127, 1958–1969.
- Cappe, C., Morel, A., Barone, P., Rouiller, E.M., 2009. The thalamocortical projection systems in primate: an anatomical support for multisensory and sensorimotor interplay. *Cereb. Cortex* 19, 2025–2037.
- Caria, A., Venuti, P., de Falco, S., 2011. Functional and dysfunctional brain circuits underlying emotional processing of music in autism spectrum disorders. *Cereb. Cortex* 21, 2838–2849.
- Chanda, M.L., Levitin, D.J., 2013. The neurochemistry of music. *Trends Cogn. Sci.* 17, 179–193.
- Chapin, H., Jantzen, K., Kelso, J.A., Steinberg, F., Large, E., 2010. Dynamic emotional and neural responses to music depend on performance expression and listener experience. *PLoS One* 5, e13812.
- Dalgleish, T., 2004. The emotional brain. *Nat. Rev. Neurosci.* 5, 583–589.
- Dietrich, S., Hertrich, I., Alter, K., Ischebeck, A., Ackermann, H., 2007. Semiotic aspects of human nonverbal vocalizations: a functional imaging study. *Neuroreport* 18, 1891–1894.
- Dietrich, S., Hertrich, I., Alter, K., Ischebeck, A., Ackermann, H., 2008. Understanding the emotional expression of verbal interjections: a functional MRI study. *Neuroreport* 19, 1751–1755.
- Duchaine, B., Yovel, G., 2015. A revised neural framework for face processing. *Annu. Rev. Vis. Sci.* 1, 393–416.
- Eldar, E., Ganor, O., Admon, R., Bleich, A., Hendler, T., 2007. Feeling the real world: limbic response to music depends on related content. *Cereb. Cortex* 17, 2828–2840.
- Escoffier, N., Zhong, J., Schirmer, A., Qiu, A., 2012. Emotional expressions in voice and music: same code, same effect? *Hum. Brain Mapp.* 34, 1796–1810.
- Ethofer, T., Anders, S., Erb, M., Herbert, C., Wiethoff, S., Kissler, J., Grodd, W., Wildgruber, D., 2006a. Cerebral pathways in processing of affective prosody: a dynamic causal modeling study. *Neuroimage* 30, 580–587.
- Ethofer, T., Anders, S., Wiethoff, S., Erb, M., Herbert, C., Saur, R., Grodd, W., Wildgruber, D., 2006b. Effects of prosodic emotional intensity on activation of associative auditory cortex. *Neuroreport* 17, 249–253.
- Ethofer, T., Wiethoff, S., Anders, S., Kreifelts, B., Grodd, W., Wildgruber, D., 2007. The voices of seduction: cross-gender effects in processing of erotic prosody. *Soc. Cogn. Affect. Neurosci.* 2, 334–337.
- 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. *J. Cogn. Neurosci.* 21, 1255–1268.
- 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. *Cereb. Cortex* 22, 191–200.
- Etkin, A., Egner, T., Kalisch, R., 2011. Emotional processing in anterior cingulate and medial prefrontal cortex. *Trends Cogn. Sci.* 15, 85–93.
- Euston, D.R., Gruber, A.J., McNaughton, B.L., 2012. The role of medial prefrontal cortex in memory and decision making. *Neuron* 76, 1057–1070.
- Fecteau, S., Armony, J.L., Joannette, Y., Belin, P., 2005. Sensitivity to voice in human prefrontal cortex. *J. Neurophysiol.* 94, 2251–2254.
- Fecteau, S., Belin, P., Joannette, Y., Armony, J.L., 2007. Amygdala responses to nonlinguistic emotional vocalizations. *Neuroimage* 36, 480–487.
- Ferrucci, R., Giannicola, G., Rosa, M., Fumagalli, M., Boggio, P.S., Hallett, M., Zago, S., Priori, A., 2012. Cerebellum and processing of negative facial emotions: cerebellar transcranial DC stimulation specifically enhances the emotional recognition of facial anger and sadness. *Cogn. Emot.* 26, 786–799.
- Frühholz, S., Grandjean, D., 2012. Towards a fronto-temporal neural network for the decoding of angry vocal expressions. *Neuroimage* 62, 1658–1666.
- Frühholz, S., Grandjean, D., 2013a. Amygdala subregions differentially respond and rapidly adapt to threatening voices. *Cortex* 49, 1394–1403.
- Frühholz, S., Grandjean, D., 2013b. Multiple subregions in superior temporal cortex are differentially sensitive to vocal expressions: a quantitative meta-analysis. *Neurosci. Biobehav. Rev.* 37, 24–35.
- Frühholz, S., Grandjean, D., 2013c. Processing of emotional vocalizations in bilateral inferior frontal cortex. *Neurosci. Biobehav. Rev.* 37, 2847–2855.
- Frühholz, S., Prinz, M., Herrmann, M., 2010. Affect-related personality traits and contextual interference processing during perception of facial affect. *Neurosci. Lett.* 469, 260–264.
- Frühholz, S., Trost, W., Grandjean, D., 2014. The role of the medial temporal limbic system in processing emotions in voice and music. *Prog. Neurobiol.* 123, 1–17.
- Frühholz, S., Gschwind, M., Grandjean, D., 2015a. Bilateral dorsal and ventral fiber pathways for the processing of affective prosody identified by probabilistic fiber tracking. *Neuroimage* 109, 27–34.
- Frühholz, S., Hofstetter, C., Cristinzio, C., Saj, A., Seeck, M., Vuilleumier, P., Grandjean, D., 2015b. Asymmetrical effects of unilateral right or left amygdala damage on auditory cortical processing of vocal emotions. *Proc. Natl. Acad. Sci. U. S. A.* 112, 1583–1588.
- Frühholz, S., Sander, D., Grandjean, D., 2016. Functional neuroimaging of human vocalizations and affective speech. *Behav. Brain Sci.* 37, 554–555.
- Friederici, A.D., 2012. The cortical language circuit: from auditory perception to sentence comprehension. *Trends Cogn. Sci.* 16, 262–268.
- Fruhholz, S., Ceravolo, L., Grandjean, D., 2012. Specific brain networks during explicit and implicit decoding of emotional prosody. *Cereb. Cortex* 22, 1107–1117.
- Fusar-Poli, P., Placentino, A., Carletti, F., Landi, P., Allen, P., Surguladze, S., Benedetti, F., Abbamonte, M., Gasparotti, R., Barale, F., Perez, J., McGuire, P., Politi, P., 2009. Functional atlas of emotional faces processing: a voxel-based meta-analysis of 105 functional magnetic resonance imaging studies. *J. Psychiatry Neurosci.* 34, 418–432.
- Garcia, R., Vouimba, R.M., Baudry, M., Thompson, R.F., 1999. The amygdala modulates prefrontal cortex activity relative to conditioned fear. *Nature* 402, 294–296.
- George, M.S., Parekh, P.I., Rosinsky, N., Ketter, T.A., Kimbrell, T.A., Heilman, K.M., Herscovitch, P., Post, R.M., 1996. Understanding emotional prosody activates right hemisphere regions. *Arch. Neurol.* 53, 665–670.
- Gifford 3rd, G.W., MacLean, K.A., Hauser, M.D., Cohen, Y.E., 2005. The neurophysiology of functionally meaningful categories: macaque ventrolateral prefrontal cortex plays a critical role in spontaneous categorization of species-specific vocalizations. *J. Cogn. Neurosci.* 17, 1471–1482.
- Glasser, M.F., Rilling, J.K., 2008. DTI tractography of the human brain's language pathways. *Cereb. Cortex* 18, 2471–2482.
- Grahn, J.A., Brett, M., 2007. Rhythm and beat perception in motor areas of the brain. *J. Cogn. Neurosci.* 19, 893–906.
- 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 meaningful speech. *Nat. Neurosci.* 8, 145–146.
- Green, A.C., Baerentsen, K.B., Stodkilde-Jorgensen, H., Wallentin, M., Roepstorff, A., Vuust, P., 2008. Music in minor activates limbic structures: a relationship with dissonance? *Neuroreport* 19, 711–715.
- Gripon, V., Berrou, C., 2011. Sparse neural networks with large learning diversity. *IEEE Trans. Neural Netw.* 22, 1087–1096.
- Hass, J., Herrmann, J.M., 2012. The neural representation of time: an information-Theoretic perspective. *Neural Comput.* 24, 1519–1552.
- Haxby, J.V., Hoffman, E.A., Gobbini, M.I., 2000. The distributed human neural system for face perception. *Trends Cogn. Sci.* 4, 223–233.
- Hoekert, M., Bais, L., Kahn, R.S., Aleman, A., 2008. Time course of the involvement of the right anterior superior temporal gyrus and the right fronto-parietal operculum in emotional prosody perception. *PLoS One* 3, e2244.
- Hoekert, M., Vingerhoets, G., Aleman, A., 2010. Results of a pilot study on the involvement of bilateral inferior frontal gyri in emotional prosody perception: an fTMS study. *BMC Neurosci.* 11, 93.
- Jacobson, L., Sapolsky, R., 1991. The role of the hippocampus in feedback regulation of the hypothalamic-pituitary-adrenocortical axis. *Endocr. Rev.* 12, 118–134.
- Janata, P., Tomic, S.T., Haberman, J.M., 2012. Sensorimotor coupling in music and the psychology of the groove. *J. Exp. Psychol. Gen.* 141, 54–75.
- Janata, P., 2009. The neural architecture of music-evoked autobiographical memories. *Cereb. Cortex* 19, 2579–2594.
- 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. *Soc. Cogn. Affect. Neurosci.* 1, 242–249.
- Juslin, P.N., Laukka, P., 2003. Communication of emotions in vocal expressions and music performance: different channels, same code? *Psychol. Bull.* 129, 770–814.
- Khalifa, S., Schon, D., Anton, J.L., Liegeois-Chauvel, C., 2005. Brain regions involved in the recognition of happiness and sadness in music. *Neuroreport* 16, 1981–1984.
- Knight, D.C., Nguyen, H.T., Bandettini, P.A., 2005. The role of the human amygdala in the production of conditioned fear responses. *Neuroimage* 26, 1193–1200.
- Koelsch, S., Fritz, T., Cramon, V., Müller, K., Friederici, A.D., 2006. Investigating emotion with music: an fMRI study. *Hum. Brain Mapp.* 27, 239–250.
- Koelsch, S., Fritz, T., Schlaug, G., 2008. Amygdala activity can be modulated by unexpected chord functions during music listening. *Neuroreport* 19, 1815–1819.
- Koelsch, S., Skouras, S., Fritz, T., Herrera, P., Bonhage, C., Kussner, M.B., Jacobs, A.M., 2013. The roles of superficial amygdala and auditory cortex in music-evoked fear and joy. *Neuroimage* 81, 49–60.
- Koelsch, S., 2010. Towards a neural basis of music-evoked emotions. *Trends Cogn. Sci.* 14, 131–137.
- Koelsch, S., 2014. Brain correlates of music-evoked emotions. *Nat. Rev. Neurosci.* 15, 170–180.
- Kotz, S.A., Schwartz, M., 2010. Cortical speech processing unplugged: a timely subcortico-cortical framework. *Trends Cogn. Sci.* 14, 392–399.
- Kotz, S.A., Meyer, M., Alter, K., Besson, M., von Cramon, D.Y., Friederici, A.D., 2003. On the lateralization of emotional prosody: an event-related functional MR investigation. *Brain Lang.* 86, 366–376.
- Kotz, S.A., Schwartz, M., Schmidt-Kassow, M., 2009. Non-motor basal ganglia functions: a review and proposal for a model of sensory predictability in auditory language perception. *Cortex* 45, 982–990.
- Kotz, S.A., Kalberlah, C., Bahlmann, J., Friederici, A.D., Haynes, J.D., 2013. Predicting vocal emotion expressions from the human brain. *Hum. Brain Mapp.* 34, 1971–1981.
- Kreifelts, B., Ethofer, T., Huberle, E., Grodd, W., Wildgruber, D., 2010. Association of trait emotional intelligence and individual fMRI-activation patterns during the perception of social signals from voice and face. *Hum. Brain Mapp.* 31, 979–991.
- Kriegeskorte, N., Goebel, R., Bandettini, P., 2006. Information-based functional brain mapping. *Proc. Natl. Acad. Sci. U. S. A.* 103, 3863–3868.
- Kringelbach, M., 2004. The functional neuroanatomy of the human orbitofrontal cortex: evidence from neuroimaging and neuropsychology. *Prog. Neurobiol.* 72, 341–372.
- Kringelbach, M.L., 2005. The human orbitofrontal cortex: linking reward to hedonic experience. *Nat. Rev. Neurosci.* 6, 691–702.

- Kumar, S., von Kriegstein, K., Friston, K., Griffiths, T.D., 2012. Features versus feelings: dissociable representations of the acoustic features and valence of aversive sounds. *J. Neurosci.* 32, 14184–14192.
- Laird, A.R., Lancaster, J.L., Fox, P.T., 2005. BrainMap: the social evolution of a human brain mapping database. *Neuroinformatics* 3, 65–78.
- Laricchiuta, D., Petrosini, L., Picerni, E., Cutuli, D., Iorio, M., Chiapponi, C., Caltagirone, C., Piras, F., Spalletta, G., 2014. The embodied emotion in cerebellum: a neuroimaging study of alexithymia. *Brain Struct. Funct.* 220, 2275–2287.
- LeDoux, J.E., Cicchetti, P., Xagoraris, A., Romanski, L.M., 1990. The lateral amygdaloid nucleus: sensory interface of the amygdala in fear conditioning. *J. Neurosci.* 10, 1062–1069.
- LeDoux, J., 2012. Rethinking the emotional brain. *Neuron* 73, 653–676.
- Lee, Y.S., Janata, P., Frost, C., Hanke, M., Granger, R., 2011. Investigation of melodic contour processing in the brain using multivariate pattern-based fMRI. *Neuroimage* 57, 293–300.
- Lehne, M., Rohmeier, M., Koelsch, S., 2014. Tension-related activity in the orbitofrontal cortex and amygdala: an fMRI study with music. *Soc. Cogn. Affect. Neurosci.* 9, 1515–1523.
- Leitman, D.I., Wolf, D.H., Ragland, J.D., Laukka, P., Loughhead, J., Valdez, J.N., Javitt, D.C., Turetsky, B.I., Gur, R.C., 2010. It's not what you say, but how you say it: a reciprocal temporo-frontal network for affective prosody. *Front. Hum. Neurosci.* 4, 1–13.
- Mas-Herrero, E., Zatorre, R.J., Rodriguez-Fornells, A., Marco-Pallares, J., 2014. Dissociation between musical and monetary reward responses in specific musical anhedonia. *Curr. Biol.* 24, 699–704.
- Menon, V., Levitin, D.J., 2005. The rewards of music listening: response and physiological connectivity of the mesolimbic system. *Neuroimage* 28, 175–184.
- Menon, V., Uddin, L.Q., 2010. Saliency, switching, attention and control: a network model of insula function. *Brain Struct. Funct.* 214, 655–667.
- Meyer, M., Zysset, S., von Cramon, D.Y., Alter, K., 2005. Distinct fMRI responses to laughter, speech, and sounds along the human peri-sylvian cortex. *Brain Res. Cogn. Brain Res.* 24, 291–306.
- Milesi, V., Cekic, S., Peron, J., Frühholz, S., Cristinzio, C., Seeck, M., Grandjean, D., 2014. Multimodal emotion perception after anterior temporal lobectomy (ATL). *Front. Hum. Neurosci.* 8, 275.
- Mirz, F., Gjedde, A., Sodkilde-Jrgensen, H., Pedersen, C.B., 2000. Functional brain imaging of tinnitus-like perception induced by aversive auditory stimuli. *Neuroreport* 11, 633–637.
- 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, 1410–1421.
- Mitchell, R.L., 2006. How does the brain mediate interpretation of incongruent auditory emotions? The neural response to prosody in the presence of conflicting lexico-semantic cues. *Eur. J. Neurosci.* 24, 3611–3618.
- Mitchell, R.L., 2007. fMRI delineation of working memory for emotional prosody in the brain: commonalities with the lexico-semantic emotion network. *Neuroimage* 36, 1015–1025.
- Mitterschiffthaler, M.T., Fu, C.H., Dalton, J.A., Andrew, C.M., Williams, S.C., 2007. A functional MRI study of happy and sad affective states induced by classical music. *Hum. Brain Mapp.* 28, 1150–1162.
- Mizuno, T., Sugishita, M., 2007. Neural correlates underlying perception of tonality-related emotional contents. *Neuroreport* 18, 1651–1655.
- Morris, J.S., Scott, S.K., Dolan, R.J., 1999. Saying it with feeling: neural responses to emotional vocalizations. *Neuropsychologia* 37, 1155–1163.
- Morris, J.S., Buchel, C., Dolan, R.J., 2001. Parallel neural responses in amygdala subregions and sensory cortex during implicit fear conditioning. *Neuroimage* 13, 1044–1052.
- Mothes-Lasch, M., Mentzel, H.J., Miltner, W.H., Straube, T., 2011. Visual attention modulates brain activation to angry voices. *J. Neurosci.* 31, 9594–9598.
- Mueller, K., Mildner, T., Fritz, T., Lepsien, J., Schwarzbauer, C., Schroeter, M.L., Moller, H.E., 2011. Investigating brain response to music: a comparison of different fMRI acquisition schemes. *Neuroimage* 54, 337–343.
- Muller, V.I., Habel, U., Derntl, B., Schneider, F., Zilles, K., Turetsky, B.I., Eickhoff, S.B., 2011. Incongruence effects in crossmodal emotional integration. *Neuroimage* 54, 2257–2266.
- Murray, I.R., Arnott, J.L., 1993. Toward the simulation of emotion in synthetic speech: a review of the literature on human vocal emotion. *J. Acoust. Soc. Am.* 93, 1097–1108.
- Panksepp, J., Bernatzky, G., 2002. Emotional sounds and the brain: the neuro-affective foundations of musical appreciation. *Behav. Processes* 60, 133–155.
- Panksepp, J., 1995. The emotional sources of chills induced by music. *Music Percept.* 13, 171–207.
- Pannese, A., Grandjean, D., Frühholz, S., 2015. Subcortical processing in auditory communication. *Hear. Res.* 328, 67–77.
- Paradiso, S., Ostedgaard, K., Vaidya, J., Ponto, L.B., Robinson, R., 2013. Emotional blunting following left basal ganglia stroke: the role of depression and fronto-limbic functional alterations. *Psychiatry Res.* 211, 148–159.
- Patel, S., Scherer, K.R., Bjorkner, E., Sundberg, J., 2011. Mapping emotions into acoustic space: the role of voice production. *Biol. Psychol.* 87, 93–98.
- Paulmann, S., Pell, M.D., Kotz, S.A., 2005. Emotional prosody recognition in BG-patients: disgust recognition revisited. *Brain Lang.* 95, 143–144.
- Paulmann, S., Pell, M.D., Kotz, S.A., 2008. Functional contributions of the basal ganglia to emotional prosody: evidence from ERPs. *Brain Res.* 1217, 171–178.
- Pell, M.D., Kotz, S.A., 2011. On the time course of vocal emotion recognition. *PLoS One* 6, e27256.
- Pell, M.D., Leonard, C.L., 2003. Processing emotional tone from speech in Parkinson's disease: a role for the basal ganglia. *Cogn. Affect. Behav. Neurosci.* 3, 275–288.
- Pell, M.D., 2006. Cerebral mechanisms for understanding emotional prosody in speech. *Brain Lang.* 96, 221–234.
- Peron, J., El Tamer, S., Grandjean, D., Leray, E., Travers, D., Drapier, D., Verin, M., Millet, B., 2011. Major depressive disorder skews the recognition of emotional prosody. *Prog. Neuropsychopharmacol. Biol. Psychiatry* 35, 987–996.
- Pessoa, L., Adolphs, R., 2010. Emotion processing and the amygdala: from a 'low road' to 'many roads' of evaluating biological significance. *Nat. Rev. Neurosci.* 11, 773–783.
- Pessoa, L., Adolphs, R., 2011. Emotion and the brain: multiple roads are better than one. *Nat. Rev. Neurosci.* 12, 425.
- Phillips, M.L., Young, A.W., Scott, S.K., Calder, A.J., Andrew, C., Giampietro, V., Williams, S.C., Bullmore, E.T., Brammer, M., Gray, J.A., 1998. Neural responses to facial and vocal expressions of fear and disgust. *Proc. Biol. Sci.* 265, 1809–1817.
- Pichon, S., Kell, C.A., 2013. Affective and sensorimotor components of emotional prosody generation. *J. Neurosci.* 33, 1640–1650.
- Popescu, A.T., Popa, D., Pare, D., 2009. Coherent gamma oscillations couple the amygdala and striatum during learning. *Nat. Neurosci.* 12, 801–807.
- 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. *Biol. Psychol.* 78, 129–137.
- Rauschecker, J.P., Scott, S.K., 2009. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat. Neurosci.* 12, 718–724.
- Romanski, L.M., Averbach, B.B., Diltz, M., 2005. Neural representation of vocalizations in the primate ventrolateral prefrontal cortex. *J. Neurophysiol.* 93, 734–747.
- Ross, E.D., Monnot, M., 2008. Neurology of affective prosody and its functional-anatomic organization in right hemisphere. *Brain Lang.* 104, 51–74.
- Rota, G., Veit, R., Nardo, D., Weiskopf, N., Birbaumer, N., Dogil, G., 2008. Processing of inconsistent emotional information: an fMRI study. *Exp. Brain Res.* 186, 401–407.
- Salimpoor, V.N., Benovoy, M., Larcher, K., Dagher, A., Zatorre, R.J., 2011. Anatomically distinct dopamine release during anticipation and experience of peak emotion to music. *Nat. Neurosci.* 14, 257–262.
- Salimpoor, V.N., van den Bosch, I., Kovacevic, N., McIntosh, A.R., Dagher, A., Zatorre, R.J., 2013. Interactions between the nucleus accumbens and auditory cortices predict music reward value. *Science* 340, 216–219.
- Sander, K., Scheich, H., 2005. Left auditory cortex and amygdala, but right insula dominance for human laughing and crying. *J. Cogn. Neurosci.* 17, 1519–1531.
- Sander, K., Brechmann, A., Scheich, H., 2003. Audition of laughing and crying leads to right amygdala activation in a low-noise fMRI setting. *Brain Res. Brain Res. Protoc.* 11, 81–91.
- 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, 848–858.
- Sander, K., Frome, Y., Scheich, H., 2007. fMRI activations of amygdala, cingulate cortex, and auditory cortex by infant laughing and crying. *Hum. Brain Mapp.* 28, 1007–1022.
- Saur, D., Kreher, B.W., Schnell, S., Kummerer, D., Kellmeyer, P., Vry, M.S., Umarova, R., Musso, M., Glauche, V., Abel, S., Huber, W., Rijntjes, M., Hennig, J., Weiller, C., 2008. Ventral and dorsal pathways for language. *Proc. Natl. Acad. Sci. U. S. A.* 105, 18035–18040.
- Schirmer, A., Kotz, S.A., 2006. Beyond the right hemisphere: brain mechanisms mediating vocal emotional processing. *Trends Cogn. Sci.* 10, 24–30.
- Schirmer, A., Zysset, S., Kotz, S.A., Yves von Cramon, D., 2004. Gender differences in the activation of inferior frontal cortex during emotional speech perception. *Neuroimage* 21, 1114–1123.
- 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, 1402–1410.
- Schmahmann, J.D., Sherman, J.C., 1998. The cerebellar cognitive affective syndrome. *Brain* 121 (Pt 4), 561–579.
- Schonwiesner, M., Rubsamen, R., von Cramon, D.Y., 2005. Hemispheric asymmetry for spectral and temporal processing in the human antero-lateral auditory belt cortex. *Eur. J. Neurosci.* 22, 1521–1528.
- Schubotz, R.I., von Cramon, D.Y., Lohmann, G., 2003. Auditory what, where, and when: a sensory somatotopy in lateral premotor cortex. *Neuroimage* 20, 173–185.
- Schutter, D.J., Enter, D., Hoppenbrouwers, S.S., 2009. High-frequency repetitive transcranial magnetic stimulation to the cerebellum and implicit processing of happy facial expressions. *J. Psychiatry Neurosci.* 34, 60–65.
- Schwartz, M., Kotz, S.A., 2013. A dual-pathway neural architecture for specific temporal prediction. *Neurosci. Biobehav. Rev.* 37, 2587–2596.
- Sebeok, T., 1985. Contributions to The Doctrine of Signs. University Press of America, Lanham, Md.
- Sergerie, K., Chochol, C., Armony, J.L., 2008. The role of the amygdala in emotional processing: a quantitative meta-analysis of functional neuroimaging studies. *Neurosci. Biobehav. Rev.* 32, 811–830.
- Sescousse, G., Redoute, J., Dreher, J.C., 2010. The architecture of reward value coding in the human orbitofrontal cortex. *J. Neurosci.* 30, 13095–13104.

- Sescousse, G., Caldu, X., Segura, B., Dreher, J.C., 2013. Processing of primary and secondary rewards: a quantitative meta-analysis and review of human functional neuroimaging studies. *Neurosci. Biobehav. Rev.* 37, 681–696.
- Shabel, S.J., Janak, P.H., 2009. Substantial similarity in amygdala neuronal activity during conditioned appetitive and aversive emotional arousal. *Proc. Natl. Acad. Sci. U. S. A.* 106, 15031–15036.
- Strata, P., Scelfo, B., Sacchetti, B., 2011. Involvement of cerebellum in emotional behavior. *Physiol. Res.* 60 (Suppl. 1), S39–48.
- Sundberg, J., Patel, S., Bjorkner, E., Scherer, K.R., 2011. Interdependencies among voice source parameters in emotional speech. *IEEE Trans. Affect. Comput.* 2, 162–174.
- Suzuki, M., Okamura, N., Kawachi, Y., Tashiro, M., Arao, H., Hoshishiba, T., Gyoba, J., Yanai, K., 2008. Discrete cortical regions associated with the musical beauty of major and minor chords. *Cogn. Affect. Behav. Neurosci.* 8, 126–131.
- Syka, J., Popelar, J., Kvasnak, E., Suta, J., Jilek, M., 1997. Processing of species-specific vocalizations in the inferior colliculus and medial geniculate body of the guinea pig. *Acoust. Signal Process. Central Audit. Syst.*, 431–441.
- Szameitat, D.P., Kreifelts, B., Alter, K., Szameitat, A.J., Sterr, A., Grodd, W., Wildgruber, D., 2010. It is not always tickling: distinct cerebral responses during perception of different laughter types. *Neuroimage* 53, 1264–1271.
- Taylor, K.S., Seminowicz, D.A., Davis, K.D., 2009. Two systems of resting state connectivity between the insula and cingulate cortex. *Hum. Brain Mapp.* 30, 2731–2745.
- Tolgyesi, B., Evers, S., 2014. The impact of cerebellar disorders on musical ability. *J. Neurol. Sci.* 343, 76–81.
- Tomlinson, S.P., Davis, N.J., Bracewell, R.M., 2013. Brain stimulation studies of non-motor cerebellar function: a systematic review. *Neurosci. Biobehav. Rev.* 37, 766–789.
- Trost, W., Ethofer, T., Zentner, M., Vuilleumier, P., 2012. Mapping aesthetic musical emotions in the brain. *Cereb. Cortex* 22, 2769–2783.
- Trost, W., Frühholz, 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.
- Trost, W., Frühholz, S., Cochrane, T., Cojan, Y., Vuilleumier, P., 2015. Temporal dynamics of musical emotions examined through intersubject synchrony of brain activity. *Soc. Cogn. Affect. Neurosci.* 10, 1705–1721.
- Viinikainen, M., Katsyri, J., Sams, M., 2012. Representation of perceived sound valence in the human brain. *Hum. Brain Mapp.* 33, 2295–2305.
- Villanueva, R., 2012. The cerebellum and neuropsychiatric disorders. *Psychiatry Res.* 198, 527–532.
- Vuilleumier, P., Richardson, M.P., Armony, J.L., Driver, J., Dolan, R.J., 2004. Distant influences of amygdala lesion on visual cortical activation during emotional face processing. *Nat. Neurosci.* 7, 1271–1278.
- Vuilleumier, P., 2005. How brains beware: neural mechanisms of emotional attention. *Trends Cogn. Sci.* 9, 585–594.
- Warren, J.E., Sauter, D.A., Eisner, F., Wiland, J., Dresner, M.A., Wise, R.J., Rosen, S., Scott, S.K., 2006. Positive emotions preferentially engage an auditory-motor mirror system. *J. Neurosci.* 26, 13067–13075.
- Weninger, F., Eyben, F., Schuller, B.W., Mortillaro, M., Scherer, K., 2013. On the acoustics of emotion in audio: what speech, music, and sound have in common. *Front. Psychol.* 4, 1–12.
- Wenstrup, J.J., 1999. Frequency organization and responses to complex sounds in the medial geniculate body of the mustached bat. *J. Neurophysiol.* 82, 2528–2544.
- 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, 885–893.
- Wildgruber, D., Pihan, H., Ackermann, H., Erb, M., Grodd, W., 2002. Dynamic brain activation during processing of emotional intonation: influence of acoustic parameters, emotional valence, and sex. *Neuroimage* 15, 856–869.
- 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. *Cereb. Cortex* 14, 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, 1233–1241.
- Wildgruber, D., Ethofer, T., Grandjean, D., Kreifelts, B., 2009. A cerebral network model of speech prosody comprehension. *Int. J. Speech Lang. Pathol.* 11, 277–281.
- Wittfoth, M., Schroder, C., Schardt, D.M., Dengler, R., Heinze, H.J., Kotz, S.A., 2010. On emotional conflict: interference resolution of happy and angry prosody reveals valence-specific effects. *Cereb. Cortex* 20, 383–392.
- Zald, D.H., Pardo, J.V., 2002. The neural correlates of aversive auditory stimulation. *Neuroimage* 16, 746–753.
- Zatorre, R.J., Belin, P., 2001. Spectral and temporal processing in human auditory cortex. *Cereb. Cortex* 11, 946–953.
- Zentner, M., Grandjean, D., Scherer, K.R., 2008. Emotions evoked by the sound of music: characterization, classification, and measurement. *Emotion* 8, 494–521.