Cognitive, Affective, & Behavioral Neuroscience (2023) 23:17–29
https://doi.org/10.3758/s13415-022-01030-y

THEORETICAL REVIEW

Disentangling emotional signals in the brain: an ALE meta-analysis of vocal affect perception

Maël Mauchand[1] · Shuyi Zhang[1]

Accepted: 24 July 2022 / Published online: 9 August 2022

Abstract

Recent advances in neuroimaging research on vocal emotion perception have revealed voice-sensitive areas specialized in processing affect. Experimental data on this subject is varied, investigating a wide range of emotions through different vocal signals and task demands. The present meta-analysis was designed to disentangle this diversity of results by summarizing neuroimaging data in the vocal emotion perception literature. Data from 44 experiments contrasting emotional and neutral voices was analyzed to assess brain areas involved in vocal affect perception in general, as well as depending on the type of voice signal (speech prosody or vocalizations), the task demands (implicit or explicit attention to emotions), and the specific emotion perceived. Results reassessed a consistent bilateral network of Emotional Voices Areas consisting of the superior temporal cortex and primary auditory regions. Specific activations and lateralization of these regions, as well as additional areas (insula, middle temporal gyrus) were further modulated by signal type and task demands. Exploring the sparser data on single emotions also suggested the recruitment of other regions (insula, inferior frontal gyrus, frontal operculum) for specific aspects of each emotion. These novel meta-analytic results suggest that while the bulk of vocal affect processing is localized in the STC, the complexity and variety of such vocal signals entails functional specificities in complex and varied cortical (and potentially subcortical) response pathways.

Keywords Emotion · Brain · Prosody · Neuroimaging · fMRI

Introduction

In person, on the phone, or in a video call, the voice is one (if not the) most important medium to convey emotions; it also is one of the least understood, especially at the neural level. While the neurocognitive literature on emotions is still largely dominated by the perception of facial expressions, research on vocal affect processing has been steadily growing over the last two decades. This increased interest has shifted the treatment of the voice in the literature from being an "auditory face" to a more independent modality of emotion perception (Schirmer, 2018; Schirmer & Adolphs, 2017). In particular, the use of functional Magnetic Resonance Imaging (fMRI) has revealed a consistent network of "Emotional Voice Areas" or EVA (Ethofer et al., 2012), consisting mainly of the bilateral mid-superior temporal cortex (STC), primary auditory regions, and surrounding areas that are highly sensitive to human voice. More anterior and frontal regions such as the insula and the inferior and middle frontal gyri (IFG, MFG), as well as emotion-related subcortical structures like the amygdala and basal ganglia are also often found to be activated for emotional compared with neutral voices (Belyk & Brown, 2013; Brück et al., 2011b).

Despite these consistent findings, vocal emotion processing is likely not as uniform as a simple aggregation of voice-sensitive areas. Indeed, affective signals in the voice are not simple biological markers of emotions but function as interpersonal social information, mobilizing a wide variety of perceptual, affective, and cognitive processes (Pell & Kotz, 2021; van Berkum, 2019; Van Kleef, 2009). While it has been treated as a right-lateralized mechanism (Ross & Monnot, 2008), increasing evidence suggest that the lateralization—and more generally, the localization—of vocal emotion perception is more relative than absolute and may depend on more specific factors

✉ Maël Mauchand
mael.mauchand@mail.mcgill.ca

[1] School of Communication Sciences and Disorders, McGill University, 2001 McGill College - Rm. 850, Montréal, QC H3A 1G1, Canada

(Kotz & Paulmann, 2011). The motivational relevance of a signal, for example, plays a critical role: studies comparing explicit to implicit vocal emotion perception suggest that the IFG as well as prefrontal regions are involved in motivated, goal-driven emotion processing (Grandjean, 2021; Schirmer, 2018). This activity may play a role in a complex cortico-subcortical network by suppressing limbic responses, such as amygdala, in a top-down emotion regulation process during explicit evaluation (Brück et al., 2011a; Mitchell et al., 2007).

Several reviews and meta-analyses of emotional voice perception have already been published. Most ALE meta-analyses were published in the early 2010s on less than 20 studies (Belyk & Brown, 2013; Frühholz & Grandjean, 2013; Witteman et al., 2012); the latest one was performed in 2017 but focused on differences between facial and vocal perception of affect (Schirmer, 2018). As such, the literature now requires not only an update, but deeper detailing of specific neural mechanisms for emotional voice processing. The definition of EVA remains very broad as it relates to the perception of vocal affect in general; this broadness does not align well with the rich multidimensional characterization of emotion communication in the psychology and pragmatics literature (Frick, 1985; Grandjean et al., 2006; Pell & Kotz, 2021; Scherer & Bänziger, 2004). The expression of emotions in the voice is complex and full of specificities and presupposes a similar degree of complexity and specificity in its perceptual processing.

There are usually two ways to vocally express affect: nonlinguistic vocalizations and speech prosody. Vocalizations are short affective bursts, such as laughter, sobs, or screams, while prosody refers to the tone of voice accompanying the production of linguistic speech. Because they share the same channel, both vocalizations and prosody extensively use acoustic features of the voice, such as pitch (fundamental frequency), voice quality (noise, jitter, harmonicity), or loudness (intensity) to convey emotions (Banse & Scherer, 1996; Eyben et al., 2016). However, fundamental differences between the two modes arise from the different constraints they are limited by. As short bursts, vocalizations are temporally restricted, and thus express emotions through large deviations in the acoustic signal that make their affective quality quickly identifiable (Castiajo & Pinheiro, 2019; Pell et al., 2015). In contrast, prosody is constrained by the linguistic content it accompanies, which already has specific acoustic signatures that make large deviations more difficult. As such, prosodic expressions of emotions are more subtle and rely more on rhythm (e.g., speech rate) and variations in the acoustic signal (e.g., a rising pitch); while this makes affective prosody harder to identify than vocalizations, it allows a more fine-grained and complex differentiation of the emotions it conveys (Grandjean et al., 2006; Pell et al., 2011).

Although the neuroimaging literature provides data from both prosody (Kotz & Paulmann, 2011) and vocalizations (Scott et al., 2010) perception, studies actively considering the distinction between the two modes are scarce. Meta-analytic work by Frühholz et al. (2016) points toward a common, but functionally diverse network for the decoding of different affective sounds. Centered around the STC, this model describes the key role of the amygdala for encoding affective salience, especially for short sounds (vocalizations), whereas frontal areas (IFG, MFG, insula) are involved in more complex mechanisms of emotional appraisal, categorization, and integration. Affective prosody also appears as preferentially right-lateralized and involving temporal decoding from the basal ganglia (Frühholz et al., 2016). It is however important to note that these meta-analytic results come from a systematic, but qualitative observation of peak activations; whether and how this network appears in an ALE analysis remains to be determined.

Emotions can then be characterized either as discrete constructs (anger, happiness, sadness, fear, disgust, and surprise) to which can be added a variety of more “complex” ones (Ekman, 1992) or as a continuous, multidimensional space, usually across the two dimensions of valence and arousal (Bradley & Lang, 1994; Russel, 1980; Schubert, 1999). While the biphasic, continuous distinction is now often preferred to the discrete categorization of emotion (Bradley & Lang, 2007), it is relatively difficult to incorporate in an ALE meta-analysis aggregating independent studies; the present study will thus consider a discrete characterization of emotion which provide a simpler and more intuitive methodological framework. Regardless of the theory considered, it is widely accepted that emotion communication is not a uniform phenomenon and that two distinct emotions can involve very distinct processes at both production and perception stages (Mauss & Robinson, 2009). For example, anger often is expressed with high intensity, increased voice noise, and a rising pitch, whereas sadness is produced at a lower intensity, decreased speech rate, and with a more restricted, falling pitch (Juslin & Laukka, 2003; Laukka et al., 2016). Perceptually, emotions are well recognized from the voice (both in discrete categories and on valence-arousal dimensions), serve a wide range of purposes (threat signaling, affiliation, mentalizing), and elicit various levels of contagion, empathy, aversion, and social behaviors (Pell & Kotz, 2021; Scarantino, 2017; Van Kleef, 2009).

As the neuroimaging literature on vocal emotion perception was still in its infancy and restricted by technical (e.g., scanner noise) and methodological (simple contrast analyses, limited power) limitations, these complex distinctions were seldom considered. This created an imbalance in terms of how different disciplines were apprehending emotion processing in the voice, with most neuroimaging studies simply considering how a collection of emotions differed from neutral speech. More recently, research has started to shift from an affect-general to an emotion-specific focus: more experiments begin to consider single emotions or disentangle multiple emotions through novel multivariate methods (Kotz et al., 2013; Whitehead & Armony, 2019). Still, a potential one-to-one mapping between specific emotions and brain mechanisms remains unclear in the field. While some studies point toward segregation within the STC (Frühholz & Grandjean, 2013), other evidence suggests that different emotions recruit a collection of additional regions depending on their various characteristics (Scott et al., 2010). In a recent review, Grandjean (2021) proposes a five-network model of vocal emotion processing centered around the EVA in which connectivity to subcortical, inferior frontal, and orbitofrontal regions reflect the temporal decoding, emotional categorization, and contextual interpretation of emotions.

The present study was designed to both update and detail the neuroimaging literature on vocal emotion processing. As described below, the current sample of 44 studies—an increase of 10 since 2017 (Schirmer, 2018), can not only strengthen and reaffirm previous meta-analytic data but also begins to provide enough power to assess more specific aspects of our perception of affect in the voice. As such, this ALE meta-analysis is the first to thoroughly evaluate how whole-brain networks may respond to different affective signals and contexts using a novel approach in the meta-analytic literature on vocal emotion perception. Beyond reassessing a network of "Emotional Voice Areas" in the Superior Temporal Cortex, expected to be robustly activated for the processing of vocal affect in general, its goal was to investigate how this network might be differentially activated based on how emotions were produced (i.e., the type of stimulus), how they were apprehended (i.e., participant task), and the specific emotion conveyed. Several other participating regions were expected to show stimulus- or context-specific activations, including the insula, middle temporal cortex, and inferior frontal gyrus, known to index secondary processing of affect in various contexts, as well as subcortical structures related to emotion (basal ganglia, amygdala).

Methods

Literature search

The present meta-analysis was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA, see Moher et al., 2009; Page et al., 2021). A systematic literature search was conducted on April 6, 2022 to find whole-brain neuroimaging articles contrasting responses to emotional vocal stimuli versus neutral vocal stimuli (Fig. 1). The search was performed on both PubMed and Scopus. PubMed gathers an important collection of neuroimaging studies, the multidisciplinary nature the Scopus database allowed the detection of several of additional studies that did not fit the PubMed catalogue, as previously used in a similar meta-analysis (Schirmer, 2018). Both databases were searched with the following term:

> ("emotion*" OR "affect*" OR "anger" OR "angry" OR "sadness" OR "sad" OR "happiness" OR "happy" OR "fear*" OR "surprise*") AND ("prosod*" OR "voice" OR "vocal" OR "spoken" OR "speech") AND ("fMRI" OR "neuroimaging" OR "brain imaging" OR "functional magnetic resonance imaging" OR "PET" OR "Positron Emission Tomography") AND ("human*")

The first segment of the term restricted the search to emotion-related articles, the second to prosody and vocal communication, and the third to neuroimaging studies. The search yielded 1,113 results in PubMed and 2,001 results in Scopus for a total of 3,114 articles, reduced to 2,246 after removing duplicates. Titles and abstracts of those articles were screened to remove irrelevant studies and unfit article types (such as reviews or commentaries). For this step, authors each screened half of the items, and cross-verified the other half to ensure all relevant articles were identified. Then, 164 potential articles were assessed for eligibility through full-text screening. Authors first assessed all articles independently, noting whether each article was eligible (with a reason in the case it was not), then compared their screening results to reach a final selection decision. Assessments conflicted for 29 articles (82% agreement); after a common reassessment, 27 were found ineligible, and 2 were eligible. Within the screened articles, 27 did not report whole-brain analyses; 8 did not fit the population of interest (neurotypical adults aged 18-50 years); 33 were reporting results from a task inadequate to our research questions (e.g., comparing the same stimuli set in different task conditions); 21 used stimuli in which an emotional aspect was absent or unclear; 24 were multimodal experiments (e.g., face + voice, semantics + prosody), which did not report contrasts for prosody effects alone, and 7 did not report coordinates for the contrasts of interest. In total, 120 articles were excluded from further analysis, leaving 44 eligible articles from which data were extracted. Authors of 12 articles were contacted to ask for potentially unreported data (e.g., analyses for controls in patient experiments, single emotions contrasts), but none were able to deliver data that could be included. A complete list of the articles with references is available in the supplementary Table A.

Data extraction

Analyses were conducted in two phases: a main affect-general phase and an exploratory emotion-specific phase.

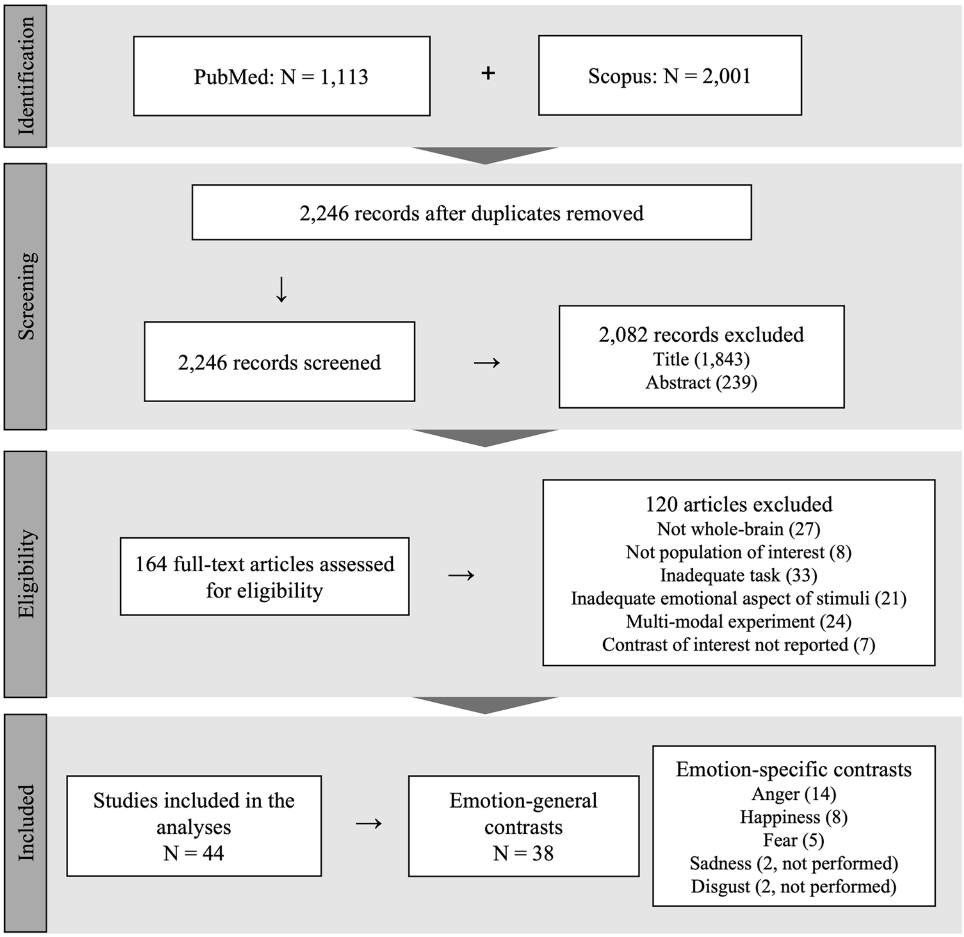
Identification
PubMed: N = 1,113
+
Scopus: N = 2,001
Screening
2,246 records after duplicates removed
2,246 records screened
2,082 records excluded
Title (1,843)
Abstract (239)
Eligibility
164 full-text articles assessed for eligibility
120 articles excluded
Not whole-brain (27)
Not population of interest (8)
Inadequate task (33)
Inadequate emotional aspect of stimuli (21)
Multi-modal experiment (24)
Contrast of interest not reported (7)
Included
Studies included in the analyses
N = 44
Emotion-general contrasts
N = 38
Emotion-specific contrasts
Anger (14)
Happiness (8)
Fear (5)
Sadness (2, not performed)
Disgust (2, not performed)

Fig. 1 Flow diagram of the systematic literature search

In each phase, data was extracted from all selected studies by one author and cross-verified by the other. In addition to the foci coordinates, sample size, and age information, the following information was extracted:

- The discrete emotions present in the contrasts of interest: while most studies focused on basic emotions (anger, happiness, fear, sadness, disgust), some also included more complex emotions or attitudes in their analysis. When present, these stimuli were only a minority among a large collection of emotions. Individual analyses on complex emotions and attitudes were not included.
- The type of vocal signal: vocalizations, including affective bursts and vowels, or prosody, including spoken words, sentences, and pseudo-sentences.
- The task demands: explicit (attending directly to the emotional nature of the stimuli), implicit (attending to an orthogonal aspect of the stimuli, e.g., speaker gender), passive (no task performed), or mixed (combining explicit and explicit tasks).

A summary of the data extracted for each analysis is displayed in Table 1. In the main phase, contrasts between emotional and neutral speech, regardless of emotion type, were included. In order to avoid sample overlap, only one contrast was selected for each included study. For studies that performed contrasts for several emotions, we selected the contrast that grouped the most emotions possible. If that was not possible, the most underrepresented emotion contrast was selected. For example, for a study reporting contrasts Anger>Neutral and Fear>Neutral, the Fear>Neutral was preferred as Anger contrasts were more frequent in the rest of the literature. A total of 38 contrasts (294 foci, 695 subjects) were extracted for this phase. Three analyses were conducted on these contrasts. The first analysis was a single-dataset analysis on all contrasts to assess brain areas activated by affective versus neutral speech in general. The second analysis aimed to compare how the way (e.g., through vocalizations or prosody) emotions were conveyed could affect these activations. The set of contrasts was divided into two stimulus types: vocalizations and speech prosody. There were 20 vocalization

Table 1 Summary of included studies in each analysis

Number of studies	Analysis					
	Affect-general	Anger	Happiness	Fear	Sadness	Disgust
	38	14	9	5	2	2
Participants M (SD)	18.62 (5.57)	19.07 (7.45)	24.89 (7.35)	17.33 (10.46)	23.50 (5.19)	12.00 (8.49)
Age M (SD)	26.64 (4.17)	26.00 (3.14)	26.69 (3.04)	28.47 (5.17)	28.25 (2.47)	31.50 (7.78)
Emotion						
Anger	25	14	-	-	-	-
Happiness	22	-	9	-	-	-
Fear	12	-	-	5	-	-
Sadness	16	-	-	-	2	-
Disgust	6	-	-	-	-	2
Surprise	3	-	-	-	-	-
Other	13	-	-	-	-	-
Stimulus type						
Vocalizations	20	5	5	5	1	2
Prosody	18	9	3	0	1	0
Task						
Explicit	19	3	2	3	0	1
Implicit	11	6	4	2	2	1
Passive	4	4	0	0	0	0
Mixed	4	1	2	0	0	0

contrasts (149 foci, 371 subjects) and 18 prosody contrasts (145 foci, 324 subjects). The third analysis assessed the effect of task on the activations. There were 19 contrasts with explicit tasks: 11 with implicit tasks, 4 with passive tasks, and 4 mixed tasks. The latter four were excluded from the analysis. Because the interest of this analysis was to assess how attending actively to emotions might affect brain activity, passive tasks were grouped together with implicit tasks as they both implied no direct attention to the affective quality of stimuli. This also allowed a more balanced contrast with the higher number of explicit tasks, yielding a comparison between 19 explicit contrasts (152 foci, 373 subjects) and 15 implicit (and passive) contrasts (114 foci, 257 subjects).

The exploratory phase aimed to investigate specific activations for each basic emotion (anger, happiness, fear, sadness, and disgust). For each emotion, a set of contrasts specific to this emotion was created; because there was less data than expected, contrasts between the emotion of interest and other emotions also were included (e.g., Anger > Happiness was included in the Anger set). In total, there were 14 contrasts (13 non-null) for Anger (101 foci, 248 subjects), 9 contrasts (8 non-null) for Happiness (53 foci, 191 subjects), 5 contrasts for Fear (28 foci, 71 subjects), 2 contrasts for Sadness, and 2 contrasts for Disgust. No analysis was performed for Sadness and Disgust.

Meta-analytic method

All coordinates were converted to MNI space before analysis: contrasts reported in Talaraich space were converted using the tal2icbm command of GingerALE software (Lancaster et al., 2007). Activation Likelihood Estimation (ALE) analyses were then performed using GingerALE v3.0.2 software (Eickhoff et al., 2009, 2012). In this method, the activation likelihood of each voxel is determined by using the peak coordinates of each included contrast. For each contrast, foci images are thus created with a three-dimensional Gaussian blur and a FWHM depending on the sample size of the corresponding study. The combination of the images from one contrast yields a modeled activation map, and the union of all maps yields an ALE image. This ALE image is then tested for above-chance clustering of activated foci between experiments against an empirically defined null distribution map. In the present analyses, the resulting maps were thresholded using a Family-Wise Error (FWE) method with cluster-level inference set at $p < 0.05$ with 1,000 permutations, and a cluster-forming threshold at an uncorrected $p < 0.001$. The FWE correction set a minimum cluster size of 888 mm^3 for the main analysis, 600 mm^3 for the anger-specific analysis, 832 mm^3 for the fear-specific analysis, and 520 mm^3 for the happiness-specific analysis.

Additionally, contrast analyses (stimuli type and task type) were performed by subtracting previously thresholded maps from one another (Eickhoff et al., 2011). To correct for dataset sizes, pooled images were created and randomly permuted into two sets of the same size before subtraction. The analyses were performed with 10,000 permutations and thresholded at a level $p < 0.05$ with a 200 mm^3 minimum cluster size.

Resulting clusters for each analysis were visualized and identified with the Harvard-Oxford atlas in FSLeyes (McCarthy, 2021) combined with the Nearest Grey Matter atlas of Mango software (Lancaster, Martinez; www.ric.uthscsa.edu/mango). A checklist summarizing the meta-analytic method (Müller et al., 2018) is available in the supplementary materials.

Results

Affect-general analyses

The whole-brain analysis of experiments contrasting emotional prosody with neutral prosody provided two main clusters that distributed bilaterally in the brain, both centered around the superior temporal cortices (Fig. 2). The first significant ALE-cluster was observed the left STG, planum temporale, Heschl's gyrus, and extended into the left insula. Thirty-nine foci from 24 experiments contributed to this 9,424 mm^3 cluster. Another main cluster was located in the right STG, MTG, and planum temporale. The cluster spanned 9,136 mm^3 and was defined by 37 contributing foci from 24 experiments. MNI coordinates and statistics of each cluster are provided in Table 2. To assess the robustness of these clusters against publication bias, a Fail-Safe N (FSN) method adapted to ALE analyses, consisting of adding randomly generated "null studies" to the dataset (Acar et al., 2018) was performed. Running the analysis with the

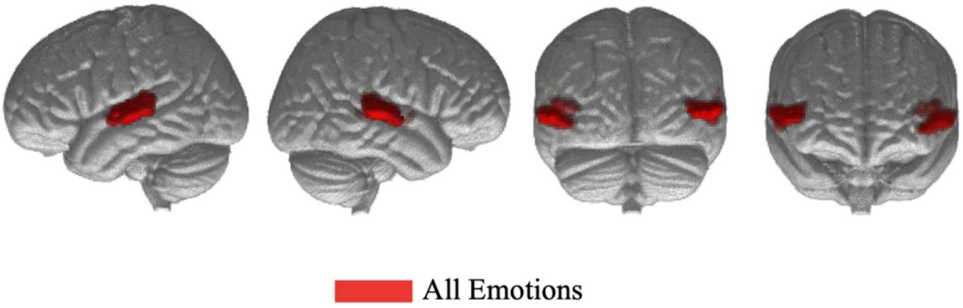

Fig. 2 ALE cluster activations for emotional compared to neutral voice

Table 2 Brain regions activated by emotions vs. neutral (FWE p <.05). Coordinates are reported in MNI space

Cluster	Peak voxel coordinates			ALE-value (x10^{-3})	P	Z	Brain region
	x	y	z				
1 *9,424 mm^3*	-56	-20	2	34.94	<.001	6.60	L. Planum Temporale
	-50	-16	-2	30.89	<.001	6.08	L. Heschl's Gyrus
	-64	-28	4	28.17	<.001	5.72	L. Superior Temporal Gyrus
	-44	-32	8	21.60	<.001	4.79	L. Planum Temporale
	-52	-28	12	21.52	<.001	4.78	L. Planum Temporale
	-40	-4	-6	18.43	<.001	4.30	L. Insular Cortex
	-36	-22	-6	15.23	<.001	3.78	L. Insular Cortex
2 *9,136 mm^3*	62	-26	4	37.52	<.001	6.99	R. Superior Temporal Gyrus
	56	-16	2	35.98	<.001	6.80	R. Planum Temporale
	50	-24	4	35.79	<.001	6.77	R. Planum Temporale
	50	-36	4	17.34	<.001	4.19	R. Middle Temporal Gyrus
	66	-32	18	15.26	<.001	3.84	R. Superior Temporal Gyrus

L = left hemisphere; R = right hemisphere.

standard guideline of 5k+10 = 200 additional null studies (Rosenthal, 1979) showed that both clusters remained significant, suggesting the robustness of these results to publication bias.

In addition to this affect-general analysis, contrast analyses revealed more novel emotion-related activations dependent on stimuli type (prosody vs. vocalization) and participant task (explicit vs. implicit). Contrasting emotion > neutral responses to prosody versus vocalization revealed that affective prosody yielded stronger activations in the right posterior STG, bilateral anterior STG, right MTG, and right insula. On the other hand, vocalizations only elicited greater activation in the left Heschl's gyrus (Fig. 3). Comparing emotion > neutral responses during explicit versus implicit tasks showed a hemispheric specialization of brain activations. While explicit tasks involved more activations in the left STC, implicit tasks responses were more right-lateralized and were restricted to the primary auditory areas: right planum temporale, central opercular cortex, and Heschl's gyrus (Fig. 4). Summaries of the contrasts for stimuli and task type are displayed in Table 3.

Exploratory emotion-specific analyses

Analyses focusing on single emotion contrasts showed activations in the EVA but also pointed toward a more diversified response based on emotion type (Table 4; Fig. 5). Analysis of anger-specific analysis revealed three clusters across both hemispheres of the brain. One cluster lay in the right hemisphere (superior temporal gyrus, Heschl's gyrus). The other two were located in the left hemisphere (superior temporal gyrus and Heschl's gyrus, and frontal operculum cortex and inferior frontal gyrus).

Specific analyses of fear and happiness did not reveal any significant cluster with FWE correction. With a more relaxed threshold of uncorrected $p = 0.001$ (minimum cluster size 200 mm^3), we found two clusters each for fear and happiness. For the vocal perception of fear, activation was found within the regions of the left insula, left frontal operculum cortex, and right paracingulate gyrus. The vocal activations for happiness were more right-lateralized with ALE foci located mostly in the EVA, including the

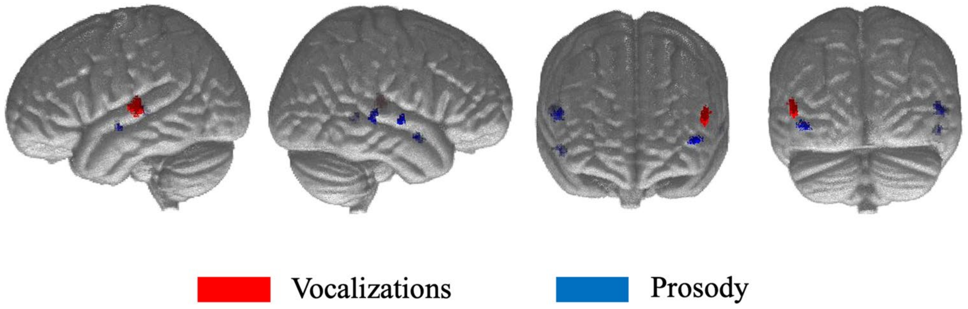

Fig. 3 Clusters showing a significant contrast between prosody and vocalization

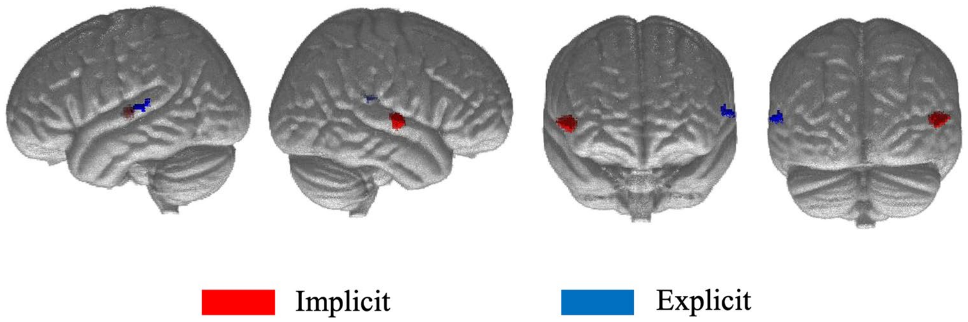

Fig. 4 Clusters showing a significant contrast between explicit and implicit tasks

Table 3 Brain regions that showed emotional activation contrasts between prosody and vocalizations, and between explicit and implicit tasks. Coordinates are reported in MNI space

Cluster	Peak voxel coordinates			P	Z	Brain region
	x	y	z			
Prosody vs Vocalizations						
1	-66	-20	2	.010	2.35	L. Superior Temporal Gyrus
776 mm^3	-64	-18	-4	.010	2.33	L. Superior Temporal Gyrus
2	62	-46	0	.008	2.40	R. Middle Temporal Gyrus
360 mm^3	62	-42	0	.008	2.40	R. Middle Temporal Gyrus
3	54	0	-16	.016	2.16	R. Superior Temporal Gyrus
288 mm^3	60	0	-14	.027	1.93	R. Middle Temporal Gyrus
4 *288 mm^3*	56	-32	0	.031	1.87	R. Middle Temporal Gyrus
5 *216 mm^3*	-46	-6	-8	.013	2.24	R. Insular Cortex
Vocalizations vs Prosody						
1 *544 mm^3*	-50	-22	8	.009	2.38	L. Heschl's Gyrus
Explicit vs Implicit						
1	-60	-36	10	0.01	2.19	L. Superior Temporal Gyrus
328 mm^3	-66	-26	8	0.04	1.78	L. Superior Temporal Gyrus
Implicit vs Explicit						
1	56	-10	4	.004	2.63	R. Planum Temporale
976 mm^3	50	-10	6	.007	2.47	LR. Central Opercular Cortex

right STG and planum temporale as well as the caudate nucleus. Only a couple foci contributed to each cluster in these analyses.

However, running FSN analyses on these clusters revealed a potential for publication biases. No cluster survived the minimum FSN of 2k null studies (Acar et al., 2018) in any of the emotion-specific analyses.

Discussion

Emotional voice areas

The present results provide an updated summary of how brain activity relates to the perception of emotions in the voice. This update was twofold: it reassessed a consistent network of Emotional Voice Areas, and further detailed how these and additional areas participate in various aspects of vocal emotion processing.

Overall, perceiving emotional relative to neutral speech appears to activate a large temporal network in both hemispheres, as assessed in previous work (Ethofer et al., 2012; Frühholz & Grandjean, 2013; Schirmer, 2018; Witteman et al., 2012). These EVA are first composed of primary auditory regions, Heschl's gyrus and planum polare, reflecting the initial appraisal of auditory cues related to emotions (Formisano et al., 2003; Griffiths & Warren, 2002). Then, the larger portion of the EVA is represented by the superior temporal cortices, centered around the mid-STG but extending posteriorly, anteriorly, and into the superior temporal sulcus and MTG. This secondary auditory region constitutes the bulk of emotion processing, potentially extracting affective information and participating in mentalizing processes (Kotz & Paulmann, 2011; Molenberghs et al., 2016). Finally, emotional voices also activated the insula, a nonlanguage area involved in more abstract emotional representations in the brain as well as empathy (Kanel et al., 2019; Molenberghs et al., 2016).

Functional specificities in this network are further revealed when comparing tasks that require or not explicit attention to emotional information. In implicit or passive tasks, where emotions are irrelevant to the task demands, primary auditory areas in the right hemisphere showed

Table 4 Brain regions activated by single emotions. Coordinates are reported in MNI space

Cluster	Peak voxel coordinates			ALE-value (x10^{-3})	P	Z	Brain region
	x	y	z				
Anger (FWE p < .05)							
1	66	-30	16	15.91	<.001	4.76	R. Superior Temporal Gyrus
2,080 mm^3	60	-34	10	13.10	<.001	4.16	R. Superior Temporal Gyrus
2	-52	-10	-4	22.25	<.001	5.85	L. Superior Temporal Gyrus
1,472 mm^3	-50	-20	6	9.73	<.001	3.51	L. Heschl's Gyrus
	-48	-16	4	9.25	<.001	3.42	L. Heschl's Gyrus
3	-44	20	0	14.90	<.001	4.56	L. Inferior Frontal Gyrus
696 mm^3	-44	18	8	14.90	<.001	4.56	L. Frontal Operculum Cortex
Fear (uncorrected p < .001)							
1	-34	20	4	8.76	<.001	4.21	L. Insular Cortex
512 mm^3	-44	18	2	8.15	<.001	3.89	L. Frontal Operculum Cortex
2	8	52	4	8.91	<.001	4.24	R. Paracingulate Gyrus
296 mm^3							
Happiness (uncorrected p < .001)							
1	46	-36	4	9.24	<.001	3.64	R. Planum Temporale
336 mm^3	46	-28	8	9.11	<.001	3.59	R. Superior Temporal Gyrus
	46	-28	2	8.69	<.001	3.50	R. Superior Temporal Gyrus
2	26	-36	12	9.71	<.001	3.78	Caudate nucleus
216 mm^3	30	-30	14	8.76	<.001	3.51	Caudate nucleus

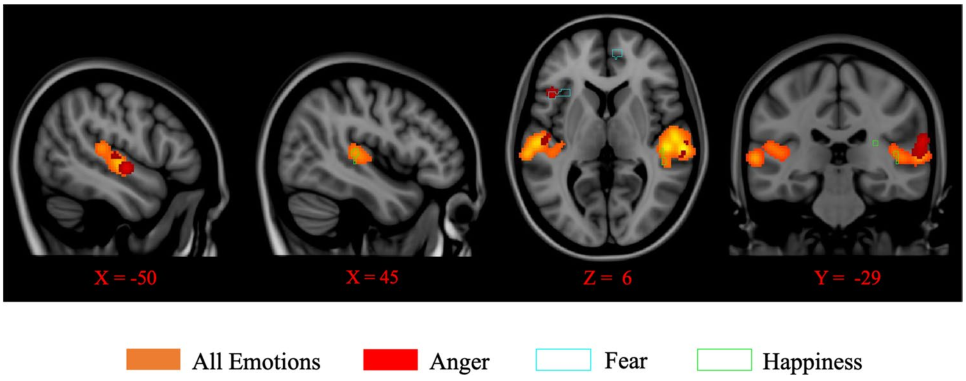

Fig. 5 Comparing affect-general and emotion-specific activation clusters. Note that fear and happiness clusters are only significant at an uncorrected p <.001

stronger responses, indicating an automatic extraction of basic acoustic properties. On the other hand, active listening with attention to emotions preferentially activated the left STC, reinforcing the role of this region in *processing* rather than just *perceiving* emotions in the voice. Note that a previous meta-analysis comparing explicit to implicit emotional tasks (involving both neutral and emotional speech) also reported activity in prefrontal areas (Schirmer, 2018). These areas were likely related to the attention required by explicit tasks, irrespective of the emotionality of the stimuli. The present results suggest that the active processing of affect in general (compared with neutral speech) involves mainly STC and not prefrontal cortices. Specific emotions requiring more attention and task control may still involve prefrontal regulation, as described later in the discussion. This result is somewhat limited by the fact that both implicit and passive tasks were grouped into one single group of contrasts; while both conditions can be seen as emotion-inattentive, they may imply different emotional perception processes that should be investigated in the future.

Different signal, different response

Comparing different types of vocal stimulus has begun to reveal distinct neural pathways dependent on the type of signal transmitted by the voice. These pathways appear to differ from the beginning of perception, in the form of a lateralized specificity of primary auditory regions. Affective vocalizations, such as laughter, sobs, or screams, elicited more activity in the left Heschl's gyrus, whereas emotional speech prosody showed stronger activation in the right posterior STG. A possible interpretation of this lateralization is related to the melodic aspect of prosody. The perception of music and singing, which are acoustically and perceptually very similar to the emotional tone of voice (Juslin & Laukka, 2003), often is reported to preferentially activate right auditory areas (Tervaniemi & Hugdahl, 2003). In contrast, the short and salient vocalizations, closer to speech phoneme perception could be perceived in a more categorical and heuristic manner through leftward activity (Grandjean, 2021; Tervaniemi & Hugdahl, 2003).

Further differences between the two modes arise in the secondary processing areas, where emotional prosody processing was enhanced bilaterally in more anterior parts of the STG, as well as in the right MTG and insula compared to vocalizations. As mentioned, the subtle and temporal aspect of prosody provides more emotional information in a less salient manner, thus requiring a deeper processing after primary acoustic feature extraction. Where vocalizations can be easily categorized in the primary perception phase as "raw" emotional cues, speech prosody carries social and interpersonal meanings that are not always immediate and unequivocal (Pell & Kotz, 2021). In particular, the parallel production of semantic information needs to be integrated in a meaningful way (Kotz & Paulmann, 2007; Pell et al., 2011): how much does prosody relate to speech content? How does the speaker feel about what they are saying? Processing vocalizations without context is a simple *what* question; processing prosody in spoken sentences is a more complex *why*, implicating intricate mentalizing processes in secondary auditory and emotion representation regions. These results align with the functional affective sound processing network from Frühholz et al. (2016), although several key regions, such as the amygdala, basal ganglia, and frontal regions, did not survive FWE correction in this analysis. This absence may be due to a lack of power combined with a very diversified collection of affective sounds, as further described below.

Specificity of vocal emotion perception

The literature search and data extraction has revealed that whole-brain neuroimaging data on the perception of specific emotions remain very scarce. In addition, no cluster FSN analysis, showing a lack of robustness of these data to the file-drawer problem and other publication biases. As such, the results of the exploratory single-emotion analysis need to be interpreted with caution; they constitute a lead toward further research rather than a conclusive summary of specific emotional brain areas.

Anger appeared to be the most extensively studied basic emotion, with almost a third of selected studies reporting anger-specific contrasts. While the analysis naturally highlighted localized activation in the EVA, it also suggests activity in STC areas beyond those highlighted in the affect-general analysis, as well as in the inferior frontal gyrus, a region often attributed to emotional prosody integration and categorization (Grandjean, 2021; Witteman et al., 2012) and emotional contagion as part of the mirror neuron system (Baird et al., 2011; Prochazkova & Kret, 2018). This activity suggests that emotion processing may extend beyond the EVA in a more partitioned manner dependent on different aspects of emotions. This partitioning may be common for multiple emotions, as fear perception showed overlapping activations with anger around the frontal operculum, which is associated with prefrontal cortex and have been found to implicate in cognitive and explicit processing (Higo et al., 2011). Because negative emotions often contain important or even life-threatening signals, those vocal cues may easily attract people's attention and reach to the frontal regions for quick decision-making and top-down control (Mitchell et al., 2007). Fear also showed marginal activity in the insula, previously associated with this specific emotion (Scott et al., 2010) but also with empathy (Jauniaux et al., 2019; Lang et al., 2011), and in the paracingulate cortex, again hinting at other possible specificities of emotion processing. In contrast to these negative emotions, happiness perception appeared to be more localized in the right STC, pointing to a valence-related segregation of vocal emotion pathways. Happiness also was the only analysis that showed suggestive subcortical activation in the caudate nucleus. Basal ganglia are known for their role in the temporal encoding of emotional signals (Frühholz et al., 2016; Paulmann et al., 2011); happiness-specific activation may be reflective of its fast and regular temporal signature, as well as its high-frequency variability (Juslin & Laukka, 2003), but this activation remains too weak to be conclusive.

These results and interpretations are not definitive but suggest that emotion processing may involve a complex network covering various aspects of affect, instead of restricting to a specialized voice-sensitive area (Ethofer et al., 2009; Grandjean, 2021). Unfortunately, the present emotion-specific analyses remain underpowered and subject to publication bias, as demonstrated by their subthreshold FSN. As such, it is not possible from this meta-analysis to exactly point out functional differences in the cortical processing of individual emotions, let alone identify emotion-specific regions within or beyond the EVA. Future research should focus on how this network is activated by single emotions rather than affect in general, and explore lesser investigated emotional constructs, such as disgust, sadness, surprise, and more complex emotions. As of now, most neuroimaging studies on vocal emotions use stimuli expressing more than four different emotions. Consequently, each individual emotion is represented by very few items, reducing any power to detect specific activations and creating a potentially false impression that all emotions activate emotional areas in the same way. Restricting investigations to fewer emotions would allow a richer, more detailed picture of how each emotion might be individually processed. As vocal emotion research spans many disciplines, it is crucial that the neuroimaging literature catches up to the more developed theoretical and empirical findings in other domains to create a unified framework for emotion processing in the voice.

Limited subcortical evidence

Interestingly, apart from the marginal caudate activation from happiness perception, the analyses performed did not show any strong subcortical activation despite the important role attributed to the amygdala and the basal ganglia in emotion processing (Paulmann et al., 2011; Pessoa, 2017). Amygdala, for example, is reported to show stronger activation for implicit compared with explicit emotion processing (Ethofer et al., 2009; Mitchell et al., 2007). One factor that could explain the lack of subcortical activity for these contrasts in the present results may be the complexity and diversity of vocal emotions: while these ancestral subcortical regions can process coarse emotional representations (felt or perceived), decoding more subtle and varied signals of the voice requires more advanced regions of the neocortex (Liebenthal et al., 2016). In line with this hypothesis, previous work reveals that subcortical structures are preferentially activated by facial rather than vocal emotion, again segregating the two modes (Schirmer, 2018). Still, it seems unlikely that vocal affect processing would not recruit the areas historically identified as emotional hubs; indeed, many experiments and reviews do report amygdala and basal ganglia activity when perceiving emotion in the voice, especially with simpler vocalizations (Scott et al., 2010; Witteman et al., 2012), and often through indirect cortical routes (Liebenthal et al., 2016).

A second factor explaining the absence of subcortical activations in the present results pertains to the lack of power and inconsistencies across the studies. The amygdala and basal ganglia are small structures that can easily not survive cluster thresholding in fMRI, especially with small sample sizes. Given the restricted number of whole-brain studies on the vocal emotions and their clear disparity in stimulus type, task requirements, and emotions conveyed, it is possible that a real amygdala (or basal ganglia) activity may not be detected. The individual variance (e.g., personality, sex) across studies also may impact the amygdala response to emotional stimuli (Brück et al., 2011a; Witteman et al., 2012). In contrast, ROI (Region of Interest) analyses of these structures (which were not included here) tend to show stronger patterns of activations (Pannese et al., 2016; Wiethoff et al., 2009). Future research focusing on more specific aspects of vocal emotion perception may provide novel insight of the roles of the subcortex in these processes.

Conclusions

The neuroimaging literature on vocal affect perception has revealed both consistent and divergent patterns in human brains' responses to emotional voices. The present meta-analysis underlines the now well-established Emotional Voices Areas, revolving around bilateral primary auditory areas and superior temporal cortices, which constitute the bulk of vocal affect processing. More importantly, it also begins to disentangle the many specific features of emotional signals in the voice and their corresponding neural signatures, both in the EVA and in key regions of the temporal and frontal lobe. The type of signal (prosody, vocalizations), the contextual demands (implicit or explicit processing), and the discrete nature of the emotion perceived show a number of specificities in the lateralization of neural activity and in the brain areas involved. In line with functional models of vocal emotion processing, it suggests that different affective signals will require diverse perceptual, affective, and cognitive functions and mobilize different brain regions accordingly. However, these analyses are limited by their low power, especially regarding individual emotion contrasts, leaving specific emotional processing pathways unclear. Unraveling these possible pathways will require detailed investigations of the acoustic, contextual, and social underpinnings of vocal emotion perception, as informed by the rich pragmatics and psychology standpoints on the subject.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.3758/s13415-022-01030-y.

Acknowledgments The authors thank Dr. Nathan Spreng for his precious help and advice regarding the meta-analytic method. This meta-analysis has not been pre-registered. All data and results files are available on the Open Science Framework (Foster & Deardorff, 2017) via this link: https://osf.io/4bcy5/?view_only=69fe95a01f3e45d18f05cd4525e1fb62.

References

Acar, F., Seurinck, R., Eickhoff, S. B., & Moerkerke, B. (2018). Assessing robustness against potential publication bias in activation likelihood estimation (ALE) meta-analyses for fMRI. *PLoS One, 13*(11), e0208177. https://doi.org/10.1371/JOURNAL.PONE.0208177

Baird, A. D., Scheffer, I. E., & Wilson, S. J. (2011). Mirror neuron system involvement in empathy: A critical look at the evidence Mirror neuron system in empathy. *Social Neuroscience, 6*(4), 327–335. https://doi.org/10.1080/17470919.2010.547085

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology Psychological Association, Inc, 70*(3), 614–636. https://pdfs.semanticscholar.org/6c2b/79b0afd32329d3e69f232aabd0521cecc484.pdf.

Belyk, M., & Brown, S. (2013). Perception of affective and linguistic prosody: An ALE meta-analysis of neuroimaging studies. *Social Cognitive and Affective Neuroscience, 9*(9), 1395–1403. https://doi.org/10.1093/scan/nst124

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry, 25*(1), 49–59. https://doi.org/10.1016/0005-7916(94)90063-9

Bradley, M. M., & Lang, P. J. (2007). Emotion and motivation. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (pp. 581–607). Cambridge University Press. https://doi.org/10.1017/CBO9780511546396.025

Brück, C., Kreifelts, B., Kaza, E., Lotze, M., & Wildgruber, D. (2011a). Impact of personality on the cerebral processing of emotional prosody. *NeuroImage, 58*(1), 259–268. https://doi.org/10.1016/j.neuroimage.2011.06.005

Brück, C., Kreifelts, B., & Wildgruber, D. (2011b). Emotional voices in context: A neurobiological model of multimodal affective information processing. *Physics of Life Reviews, 8*(4), 383–403. https://doi.org/10.1016/j.plrev.2011.10.002

Castiajo, P., & Pinheiro, A. P. (2019). Decoding emotions from nonverbal vocalizations: How much voice signal is enough? *Motivation and Emotion, 43*(5), 803–813. https://doi.org/10.1007/s11031-019-09783-9

Eickhoff, S. B., Bzdok, D., Laird, A. R., Kurth, F., & Fox, P. T. (2012). Activation likelihood estimation meta-analysis revisited. *NeuroImage, 59*(3), 2349–2361. https://doi.org/10.1016/j.neuroimage.2011.09.017

Eickhoff, S. B., Bzdok, D., Laird, A. R., Roski, C., Caspers, S., Zilles, K., & Fox, P. T. (2011). Co-activation patterns distinguish cortical modules, their connectivity and functional differentiation. *NeuroImage, 57*(3), 938–949. https://doi.org/10.1016/j.neuroimage.2011.05.021

Eickhoff, S. B., Laird, A. R., Grefkes, C., Wang, L. E., Zilles, K., & Fox, P. T. (2009). Coordinate-based activation likelihood estimation meta-analysis of neuroimaging data: A random-effects approach based on empirical estimates of spatial uncertainty. *Human Brain Mapping, 30*(9), 2907–2926. https://doi.org/10.1002/hbm.20718

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

Ethofer, T., Bretscher, J., Gschwind, M., Kreifelts, B., Wildgruber, D., & Vuilleumier, P. (2012). Emotional voice areas: Anatomic location, functional properties, and structural connections revealed by combined fMRI/DTI. *Cerebral Cortex, 22*(1), 191–200. https://doi.org/10.1093/cercor/bhr113

Ethofer, T., Kreifelts, B., Wiethoff, S., Wolf, J., Grodd, W., Vuilleumier, P., & Wildgruber, D. (2009). Differential influences of emotion, task, and novelty on brain regions underlying the processing of speech melody. *Journal of Cognitive Neuroscience, 21*(7), 1255–1268. https://doi.org/10.1162/jocn.2009.21099

Eyben, F., Scherer, K. R., Schuller, W., Sundberg, J., Andr, E., Busso, C., Devillers, L. Y., Epps, J., Laukka, P., Narayanan, S. S., & Truong, K. P. (2016). The Geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing. *IEEE Transactions on Affective Computing, 7*(2). https://doi.org/10.1109/TAFFC.2015.2457417

Formisano, E., Kim, D. S., Di Salle, F., Van De Moortele, P. F., Ugurbil, K., & Goebel, R. (2003). Mirror-symmetric tonotopic maps in human primary auditory cortex. *Neuron, 40*(4), 859–869. https://doi.org/10.1016/S0896-6273(03)00669-X

Foster, E. D., & Deardorff, A. (2017). Open Science framework (OSF). *Journal of the Medical Library Association, 105*(2), 203–206. https://doi.org/10.5195/JMLA.2017.88

Frick, R. W. (1985). Communicating emotion: The role of prosodic features. *Psychological Bulletin, 97*(3), 412–429. https://doi.org/10.1037/0033-2909.97.3.412

Frühholz, S., & Grandjean, D. (2013). Multiple subregions in superior temporal cortex are differentially sensitive to vocal expressions: A quantitative meta-analysis. *Neuroscience & Biobehavioral Reviews, 37*(1), 24–35. https://doi.org/10.1016/j.neubiorev.2012.11.002

Frühholz, S., Trost, W., & Kotz, S. A. (2016). The sound of emotions—Towards a unifying neural network perspective of affective sound processing. *Neuroscience & Biobehavioral Reviews, 68*, 96–110. https://doi.org/10.1016/J.NEUBIOREV.2016.05.002

Grandjean, D. (2021). Brain networks of emotional prosody processing. *Emotion Review, 13*(1), 34–43. https://doi.org/10.1177/1754073919898522

Grandjean, D., Bänziger, T., & Scherer, K. R. (2006). Intonation as an interface between language and affect. *Progress in Brain Research, 156*, 235–247. https://doi.org/10.1016/S0079-6123(06)56012-1

Griffiths, T. D., & Warren, J. D. (2002). The planum temporale as a computational hub. In *trends in neurosciences* (Vol. 25, issue 7, pp. 348–353). Elsevier ltd. https://doi.org/10.1016/S0166-2236(02)02191-4.

Higo, T., Mars, R. B., Boorman, E. D., Buch, E. R., & Rushworth, M. F. S. (2011). Distributed and causal influence of frontal operculum in task control. *Proceedings of the National Academy of Sciences of the United States of America, 108*(10), 4230–4235. https://doi.org/10.1073/pnas.1013361108

Jauniaux, J., Khatibi, A., Rainville, P., & Jackson, P. (2019). A meta-analysis of neuroimaging studies on pain empathy: Investigating the role of visual information and observers' perspective. *Social Cognitive and Affective Neuroscience*.

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin, 129*(5), 770–814. https://doi.org/10.1037/0033-2909.129.5.770

Kanel, D., Al-Wasity, S., Stefanov, K., & Pollick, F. E. (2019). Empathy to emotional voices and the use of real-time fMRI to enhance activation of the anterior insula. *NeuroImage, 198*, 53–62. https://doi.org/10.1016/J.NEUROIMAGE.2019.05.021

Kotz, S. A., Kalberlah, C., Bahlmann, J., Friederici, A. D., & Haynes, J. D. (2013). Predicting vocal emotion expressions from the human brain. *Human Brain Mapping, 34*(8), 1971–1981. https://doi.org/10.1002/hbm.22041

Kotz, S. A., & Paulmann, S. (2007). When emotional prosody and semantics dance cheek to cheek: ERP evidence. *Brain Research, 1151*, 107–118. https://doi.org/10.1016/j.brainres.2007.03.015

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

Lancaster, J. L., Tordesillas-Gutié Rrez, D., Martinez, M., Salinas, F., Evans, A., Zilles, K., Mazziotta, J. C., & Fox, P. T. (2007). Bias between MNI and Talairach coordinates analyzed using the ICBM-152 brain template. *Human Brain Mapping, 28*, 1194–1205. https://doi.org/10.1002/hbm.20345

Lang, S., Yu, T., Markl, A., Müller, F., & Kotchoubey, B. (2011). Hearing others' pain: Neural activity related to empathy. *Cognitive, Affective, & Behavioral Neuroscience, 11*(3), 386–395. https://doi.org/10.3758/s13415-011-0035-0

Laukka, P., Elfenbein, H. A., Thingujam, N. S., Rockstuhl, T., Iraki, F. K., Chui, W., & Althoff, J. (2016). The expression and recognition of emotions in the voice across five nations: A lens model analysis based on acoustic features. *Journal of Personality and Social Psychology, 111*(5), 686–705. https://doi.org/10.1037/pspi0000066

Liebenthal, E., Silbersweig, D. A., & Stern, E. (2016). The language, tone and prosody of emotions: Neural substrates and dynamics of spoken-word emotion perception. *Frontiers in Neuroscience, 10*(506). https://doi.org/10.3389/fnins.2016.00506

Mauss, I. B., & Robinson, M. D. (2009). Measures of emotion: A review. In *cognition and emotion* (Vol. 23, issue 2, pp. 209–237). Taylor & Francis Group. https://doi.org/10.1080/02699930802204677.

McCarthy, P. (2021). *FSLeyes*. https://doi.org/10.5281/ZENODO.4704476

Mitchell, D. G. V., Nakic, M., Fridberg, D., Kamel, N., Pine, D. S., & Blair, R. J. R. (2007). The impact of processing load on emotion. *NeuroImage, 34*(3), 1299–1309. https://doi.org/10.1016/j.neuroimage.2006.10.012

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Group, T. P. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine, 6*(7), e1000097. https://doi.org/10.1371/JOURNAL.PMED.1000097

Molenberghs, P., Johnson, H., Henry, J. D., & Mattingley, J. B. (2016). Understanding the minds of others: A neuroimaging meta-analysis. *Neuroscience & Biobehavioral Reviews, 65*, 276–291. https://doi.org/10.1016/j.neubiorev.2016.03.020

Müller, V. I., Cieslik, E. C., Laird, A. R., Fox, P. T., Radua, J., Mataix-Cols, D., Tench, C. R., Yarkoni, T., Nichols, T. E., Turkeltaub, P. E., Wager, T. D., & Eickhoff, S. B. (2018). Ten simple rules for neuroimaging meta-analysis. *Neuroscience and Biobehavioral Reviews, 84*(November 2017), 151–161. https://doi.org/10.1016/j.neubiorev.2017.11.012

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., & Moher, D. (2021). Updating guidance for reporting systematic reviews: Development of the PRISMA 2020 statement. *Journal of Clinical Epidemiology, 134*, 103–112. https://doi.org/10.1016/J.JCLINEPI.2021.02.003

Pannese, A., Grandjean, D., & Frühholz, S. (2016). Amygdala and auditory cortex exhibit distinct sensitivity to relevant acoustic features of auditory emotions. *Cortex, 85*, 116–125. https://doi.org/10.1016/J.CORTEX.2016.10.013

Paulmann, S., Ott, D. V. M., & Kotz, S. A. (2011). Emotional speech perception unfolding in time: The role of the basal ganglia. *PLoS One, 6*(3), 13–17. https://doi.org/10.1371/journal.pone.0017694

Pell, M. D., Jaywant, A., Monetta, L., Kotz, S., & a. (2011). Emotional speech processing: Disentangling the effects of prosody and semantic cues. *Cognition & Emotion, 25*(5), 834–853. https://doi.org/10.1080/02699931.2010.516915

Pell, M. D., & Kotz, S. A. (2021). The next frontier: Prosody research gets interpersonal. *Emotion Review, 13*(1), 51–56. https://doi.org/10.1177/1754073920954288

Pell, M. D., Rothermich, K., Liu, P., Paulmann, S., Sethi, S., & Rigoulot, S. (2015). Preferential decoding of emotion from human non-linguistic vocalizations versus speech prosody. *Biological Psychology, 111*, 14–25. https://doi.org/10.1016/j.biopsycho.2015.08.008

Pessoa, L. (2017). A network model of the emotional brain. In *trends in cognitive sciences* (Vol. 21, Issue 5, pp. 357–371). Elsevier Ltd. https://doi.org/10.1016/j.tics.2017.03.002

Prochazkova, E., & Kret, M. E. (2018). Connecting minds and sharing emotions through mimicry: A neurocognitive model of emotional contagio. *Neuroscience and Biobehavioral Reviews, 80*, 99–114. https://doi.org/10.1016/j.neubiorev.2017.05.013

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86*(3), 638–641. https://doi.org/10.1037/0033-2909.86.3.638

Ross, E. D., & Monnot, M. (2008). Neurology of affective prosody and its functional-anatomic organization in right hemisphere. *Brain and Language, 104*(1), 51–74. https://doi.org/10.1016/j.bandl.2007.04.007

Russel, J. A. (1980). A circumplex model of affect. *Journal of Personality and Psychology*, *39*(6), 1161–1178 https://psycnet.apa.org/fulltext/1981-25062-001.pdf.

Scarantino, A. (2017). How to do things with emotional expressions: The theory of affective pragmatics. *Psychological Inquiry, 28*(3), 165–185. https://doi.org/10.1080/1047840X.2017.1328951

Scherer, K. R., & Bänziger, T. (2004). Emotional expression in prosody: a review and an agenda for future research. *Proc. Speech Prosody*, 359–366. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.95.7677&rep=rep1&type=pdf

Schirmer, A. (2018). Is the voice an auditory face? An ALE meta-analysis comparing vocal and facial emotion processing. *Social Cognitive and Affective Neuroscience, 13*(1), 1–13. https://doi.org/10.1093/scan/nsx142

Schirmer, A., & Adolphs, R. (2017). Emotion perception from face, voice, and touch: Comparisons and convergence. *Trends in Cognitive Sciences, 21*(3), 216–228. https://doi.org/10.1016/j.tics.2017.01.001

Schubert, E. (1999). Measuring emotion continuously: Validity and reliability of the two-dimensional emotion-space. *Australian Journal of Psychology, 51*(3), 154–165. https://doi.org/10.1080/00049539908255353

Scott, S. K., Sauter, D., & McGettigan, C. (2010). Brain mechanisms for processing perceived emotional vocalizations in humans. In *handbook of behavioral neuroscience* (Vol. 19, issue C, pp. 187–197). https://doi.org/10.1016/B978-0-12-374593-4.00019-X.

Tervaniemi, M., & Hugdahl, K. (2003). Lateralization of auditory-cortex functions. In *brain research reviews* (Vol. 43, issue 3, pp. 231–246). Elsevier. https://doi.org/10.1016/j.brainresrev.2003.08.004.

van Berkum, J. J. A. (2019). Language comprehension and emotion: Where are the interfaces, and who cares? In G. I. de Zubicaray & N. O. Schiller (Eds.), *The Oxford handbook of neurolinguistics (pp. 735–766)*. Oxford University Press. https://doi.org/10.1093/OXFORDHB/9780190672027.013.29

Van Kleef, G. A. (2009). How emotions regulate social life. *Current Directions in Psychological Science, 18*(3), 184–188. https://doi.org/10.1111/j.1467-8721.2009.01633.x

Whitehead, J. C., & Armony, J. L. (2019). Multivariate fMRI pattern analysis of fear perception across modalities. *European Journal of Neuroscience, 49*(12), 1552–1563. https://doi.org/10.1111/ejn.14322

Wiethoff, S., Wildgruber, D., Grodd, W., & Ethofer, T. (2009). Response and habituation of the amygdala during processing of emotional prosody. *NeuroReport, 20*(15), 1356–1360. https://doi.org/10.1097/WNR.0b013e328330eb83

Witteman, J., Van Heuven, V. J. P., & Schiller, N. O. (2012). Hearing feelings: A quantitative meta-analysis on the neuroimaging literature of emotional prosody perception. *Neuropsychologia, 50*(12), 2752–2763. https://doi.org/10.1016/j.neuropsychologia.2012.07.026

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.