archive ouverte UNIGE

http://archive-ouverte.unige.ch

Article

Amygdala and auditory cortex exhibit distinct sensitivity to relevant acoustic features of auditory emotions

PANNESE, Alessia, GRANDJEAN, Didier Maurice, FRUEHHOLZ, Sascha

Abstract

Discriminating between auditory signals of different affective value is critical to successful social interaction. It is commonly held that acoustic decoding of such signals occurs in the auditory system, whereas affective decoding occurs in the amygdala. However, given that the amygdala receives direct subcortical projections that bypass the auditory cortex, it is possible that some acoustic decoding occurs in the amygdala as well, when the acoustic features are relevant for affective discrimination. We tested this hypothesis by combining functional neuroimaging with the neurophysiological phenomena of repetition suppression (RS) and repetition enhancement (RE) in human listeners. Our results show that both amygdala and auditory cortex responded differentially to physical voice features, suggesting that the amygdala and auditory cortex decode the affective quality of the voice not only by processing the emotional content from previously processed acoustic features, but also by processing the acoustic features themselves, when these are relevant to the identification of the voice's affective value. Specifically, we found [...]

Reference

UNIVERSITÉ
DE GENÈVE

Research report

Amygdala and auditory cortex exhibit distinct sensitivity to relevant acoustic features of auditory emotions

Alessia Pannese [a,b,c], *Didier Grandjean* [a,b] and *Sascha Frühholz* [a,d,e,f,*]

[a] *Swiss Center for Affective Sciences, University of Geneva, Switzerland*
[b] *Neuroscience of Emotion and Affective Dynamics Laboratory, University of Geneva, Geneva, Switzerland*
[c] *Department of Fundamental Neuroscience, Geneva University Hospital, Geneva, Switzerland*
[d] *Department of Psychology, University of Zurich, Zurich, Switzerland*
[e] *Neuroscience Center Zurich, University of Zurich and ETH Zurich, Zurich, Switzerland*
[f] *Center for Integrative Human Physiology (ZIHP), University of Zurich, Switzerland*

ARTICLE INFO

ABSTRACT

Discriminating between auditory signals of different affective value is critical to successful social interaction. It is commonly held that acoustic decoding of such signals occurs in the auditory system, whereas affective decoding occurs in the amygdala. However, given that the amygdala receives direct subcortical projections that bypass the auditory cortex, it is possible that some acoustic decoding occurs in the amygdala as well, when the acoustic features are relevant for affective discrimination. We tested this hypothesis by combining functional neuroimaging with the neurophysiological phenomena of repetition suppression (RS) and repetition enhancement (RE) in human listeners. Our results show that both amygdala and auditory cortex responded differentially to physical voice features, suggesting that the amygdala and auditory cortex decode the affective quality of the voice not only by processing the emotional content from previously processed acoustic features, but also by processing the acoustic features themselves, when these are relevant to the identification of the voice's affective value. Specifically, we found that the auditory cortex is sensitive to spectral high-frequency voice cues when discriminating vocal anger from vocal fear and joy, whereas the amygdala is sensitive to vocal pitch when discriminating between negative vocal emotions (i.e., anger and fear). Vocal pitch is an instantaneously recognized voice feature, which is potentially transferred to the amygdala by direct subcortical projections. These results together provide evidence that, besides the auditory cortex, the amygdala too processes acoustic information, when this is relevant to the discrimination of auditory emotions.

* *Corresponding author*. University of Zurich, Department of Psychology, Binzmühlestrasse 14, Box 18, 8050 Zurich, Switzerland.
E-mail address: sascha.fruehholz@uzh.ch (S. Frühholz).

1. Introduction

The human voice is a rich source of information about the affective state of the speaker. Independently of their semantic content, vocal signals convey important indications about a person's emotions and intentions through their intonation. Different emotional states are expressed by distinct profiles of vocal acoustic features (Banse & Scherer, 1996). The correct interpretation of these combined acoustic profiles and discriminative voice features is critical for the correct identification and discrimination of the affective value of the vocalization, which is crucial for the implementation of appropriate behavioral responses (Marsh, Ambady, & Kleck, 2005).

A central question in vocal communication is how the decoding of these acoustic features relates to the discrimination between different emotional states. Vocal emotions engage the auditory cortex (Ethofer et al., 2012; Frühholz & Grandjean, 2013b) and the amygdala (Fecteau, Belin, Joanette, & Armony, 2007; Frühholz & Grandjean, 2013a). The amygdala is thought to decode the overall affective value based on acoustic features that are pre-processed in other brain regions, especially in the auditory cortex (Kumar, von Kriegstein, Friston, & Griffiths, 2012). However, besides receiving information from the auditory cortex, the amygdala also receives direct projections from subcortical regions of the ascending auditory pathway (Frühholz, Trost, & Grandjean, 2014). The functional significance of this direct structural connection is largely unknown. It is possible that these direct projections supply information about the acoustic features of auditory signals. If this were the case, the amygdala might not only perform an emotional analysis on previously processed acoustic features, but might also process some of these acoustic features itself, when these are relevant for the affective discrimination of the vocalization.

We addressed the question of whether, and to what extent, the *amygdala* and the *temporal voice area* (TVA, a voice-sensitive area within the auditory cortex; see Belin & Zatorre, 2000) discern different types of affective vocalizations based on acoustic features. The study focused on these two regions-of-interest (ROI) because of their supposed role on emotional and auditory processing from auditory emotions (Frühholz, Trost et al., 2014; Pannese, Grandjean, & Frühholz, 2015), respectively, and because a recent study proposed a functional segregation between these two regions (Kumar et al., 2012). We furthermore focused on three acoustic features known to differ across, and to be used by listeners to discriminate between, vocal emotions (Banse & Scherer, 1996): the mean of the fundamental frequency (F0m) and its variability (F0sd) (both F0m and F0sd contribute to the perception of vocal pitch) as well as the spectral high-frequencies (HF) cues. The latter are represented by the energy or power of fast sound wave oscillations approximately above 1 kHz. Likewise, we chose three vocal emotions known to differ in these discriminative acoustic features (Banse & Scherer, 1996): anger, fear, and happiness. This combination allows testing neural discrimination mechanisms within and across valences requiring different adaptive responses (Marsh et al.,

2005). Here, we specifically tested how fearful and happy voices are neurally discriminated from angry voices, which is a highly relevant decision for survival.

We tested our hypothesis by combining functional magnetic resonance imaging (fMRI) in human listeners with the neurophysiological phenomena of repetition suppression (RS) and repetition enhancement (RE), consisting in a decreased or increased neural response, respectively, to a given stimulus feature following repeated exposure to it (Grill-Spector, Henson, & Martin, 2006). A given stimulus may induce RS or RE depending on various factors, including perceptual strength (e.g., visibility) and number of repetitions (Muller, Strumpf, Scholz, Baier, & Melloni, 2013). RS is thought to reflect neural coding specificity (Naccache & Dehaene, 2001), and the reduced response is thought to reflect prime-induced facilitation in the neural processing of the target due to synaptic changes in the stimulus-response pathway. RE is instead thought to reflect strengthening of the stimulus representation (Rainer, Lee, & Logothetis, 2004), and is often observed when the repeated stimuli are degraded (Turk-Browne, Yi, Leber, & Chun, 2007). RS and RE can co-occur within the same cortical region (de Gardelle, Waszczuk, Egner, & Summerfield, 2013). The presence of RS or RE in a given brain region for trials in which prime and target share a given feature (as opposed to for trials in which prime and target do not share that feature) indicates that the given brain region is involved in the processing of that feature.

2. Materials and methods

2.1. Technique

We identified the TVA through a functional voice localizer scan, and the amygdala through anatomical labeling. We then tested TVA and amygdala responses to pairs of stimuli (S1 and S2; Fig. 1A) during three experimental conditions: (a) S1—S2 expressed the same emotion ('repeated' condition), (b) S1—S2 expressed different emotions ('unrepeated' condition), or (c) S1—S2 expressed different emotions, but were made equal or similar in terms of the F0m, F0sd, or HF cues (Fig. 1B). Equating S1—S2 for F0m, F0sd, or HF cues eliminates a distinguishing feature between different emotions, and makes the emotions expressed in S1—S2 acoustically more similar. Thus, the 'repeated' and 'unrepeated' conditions tested for TVA and amygdala overall sensitivity to vocal emotions ('emotion' discrimination effect, Fig. 1B, lower panel), whereas the 'F0m', 'F0sd', and 'HF' conditions tested for TVA and amygdala sensitivity to specific discriminative acoustic features of vocal emotions ('acoustic' discrimination effect Fig. 1B, upper panel).

2.2. Participants

Twenty-two healthy adults (14 females; mean age 24.7 years, age range 19—36 years) participated in the study. All participants had normal or corrected-to-normal vision, normal hearing abilities, and no history of neurological or psychiatric disorders. All participants gave informed written consent in accordance
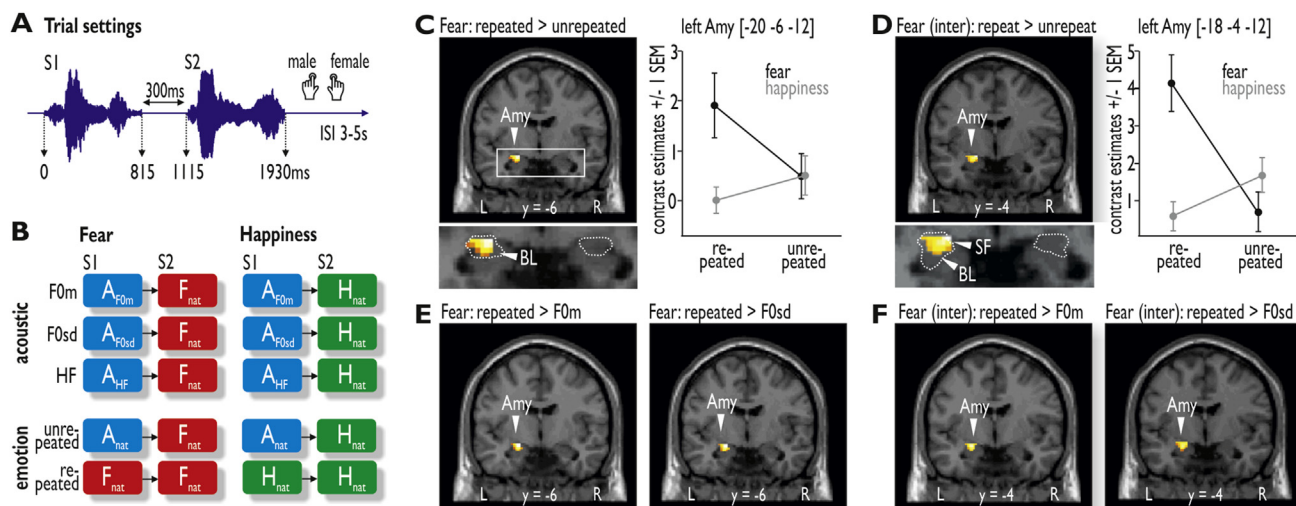
Fig. 1 — Emotion and acoustic discrimination in the amygdala. (A) Participants heard pairs (S1 and S2) of semantically meaningless two-syllable pseudo-words spoken in an emotional tone, and pressed a button to indicate the gender of the speaker (S1 and S2 always had the same gender). (B) In 'acoustic' conditions (upper part), S1 was manipulated to have the same or similar mean pitch ('F0m'), pitch variation ('F0sd'), or high-frequency cues ('HF') as S2. In 'emotion' conditions (lower part), S1 expressed the same ('repeated') or a different ('unrepeated') emotion from S2. S2 expressed either fear (F_{nat}) or happiness (H_{nat}). This resulted in ten different experimental conditions, such as for example F0m-manipulated angry voices (S1) followed by native fearful voices (S2) (referred to as [$A_{F0m}F_{nat}$] trials, upper left part). (C) An *emotion* discrimination effect was found in a region within the left amygdala (MNI xyz [−20 −6 −12], likely to be located in the superficial nucleus, the basolateral complex, and the centromedian nucleus (80%, 70%, and 60% probability, respectively, based on the SPM Anatomy toolbox)) illustrated by a repetition enhancement (RE) effect for repeated fear ([$F_{nat}F_{nat} > A_{nat}F_{nat}$]), but not repeated happiness ([$H_{nat}H_{nat} > A_{nat}H_{nat}$]). This effect is also shown in contrasts estimates in the right panel (error bars ± 1SEM). (D) A very similar region (MNI xyz [−18 −4 −12], likely to be located in the superficial nucleus (80% probability based on the SPM Anatomy toolbox)) was found to exhibit an interaction effect between repetition and emotion ([$F_{nat}F_{nat} > A_{nat}F_{nat}$] > [$H_{nat}H_{nat} > A_{nat}H_{nat}$]; p = .033). No main effect of repetition (i.e., across emotions) was observed ([$F_{nat}F_{nat} + H_{nat}H_{nat}$] > [$A_{nat}F_{nat} + A_{nat}H_{nat}$]). (E) An *acoustic* discrimination effect, illustrated by an RE effect, was also found in the left amygdala for discriminating vocal anger and fear along the mean of vocal pitch (i.e., F0m; [$F_{nat}F_{nat} > A_{F0m}F_{nat}$]; [−20 −6 −12]) and along the variation of vocal pitch (i.e., F0sd; [$F_{nat}F_{nat} > A_{F0sd}F_{nat}$]; [−20 −6 −14]). (F) These acoustic discrimination effects for fear-anger discrimination along vocal pitch features (F0m left panel, F0sd right panel) were further specified by interaction contrasts showing enhanced activity for fear relative to happy voices ([$F_{nat}F_{nat} > A_{F0m}F_{nat}$] > [$H_{nat}H_{nat} > A_{F0m}H_{nat}$], [−22 −6 −14] (left panel); ([$F_{nat}F_{nat} > A_{F0sd}F_{nat}$] > [$H_{nat}H_{nat} > A_{F0sd}H_{nat}$], [−20 −6 −14] (right panel). For all data: N = 22; peak activations are small volume corrected (SVC) at a FWE-corrected threshold of p < .05 applied to an anatomical bilateral amygdala mask.

with the Helsinki declaration. The study was approved by the local ethics committee of the University of Geneva.

2.3. Voice localizer scan

For the purpose of localizing auditory cortical regions sensitive to human voice, on each participant a voice localizer run was included in the fMRI experimental session (Belin & Zatorre, 2000). The localizer run consisted in a passive listening block design including 32 stimulation epochs and 16 silent epochs (8 sec duration each). The stimulation epochs included two conditions: (a) 'vocal', including 16 blocks with human voices (e.g., speaking, laughing, singing); and (b) 'non-vocal', including 16 blocks with other sounds (of which 8 blocks were cries of various animals, and 8 blocks were environmental sounds, e.g., train whistling, telephone ringing, bells tolling) (Fig. 2A). Participants were instructed to passively listen to the stimuli.

2.4. Main experiment

2.4.1. Pre-evaluation
The experimental stimuli consisted in speech-like but semantically meaningless two-syllable pseudo-words spoken in an emotional tone by two male and two female actors (Frühholz, Klaas, Patel, & Grandjean, 2014). Four pseudo-words ('belam', 'lagod', 'namil', 'nolan') were selected based on their high discrimination rate (on a continuous scale ranging from "0" (low) to "100" (high)) in a pre-evaluation. Each pseudo-word was spoken two times by two speakers (one male, one female), in three target emotional expressions: anger, fear, and happiness.

For each target emotion, a one-way repeated-measure ANOVA (5 emotion rating scales between 0 and 100: neutral, anger, fear, sadness, and happiness) showed that the pseudo-words were evaluated significantly more strongly on the target emotion scale than for the other scales: (a) *Anger*:
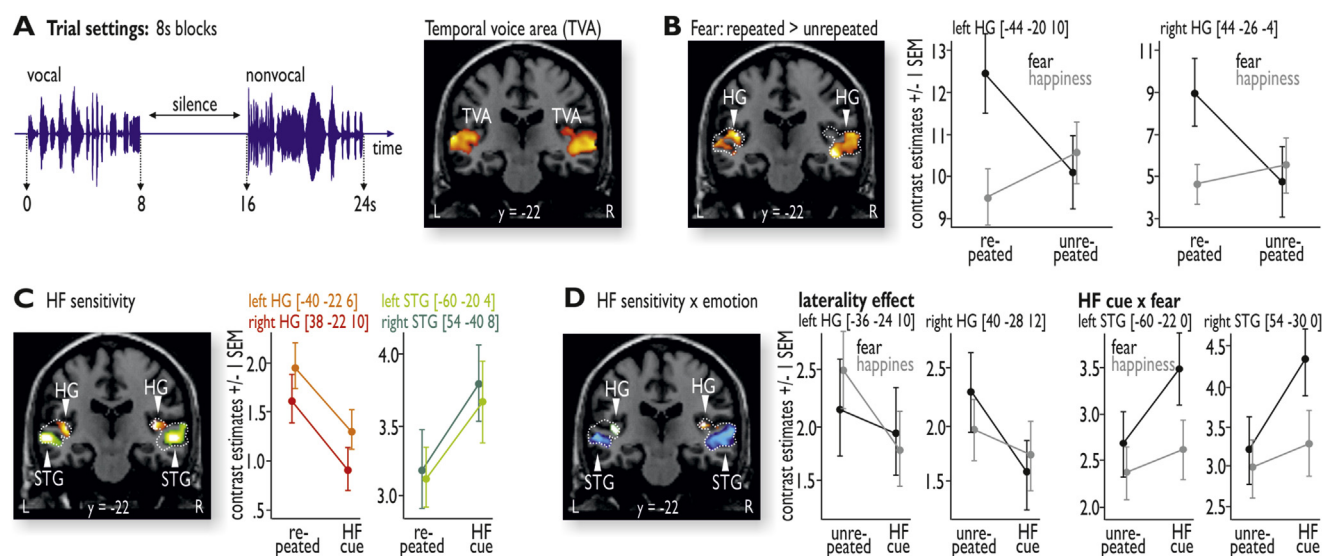
Fig. 2 − Emotion and acoustic discrimination in voice-sensitive auditory cortex. **(A)** The voice localizer consisted of alternating 'vocal', 'nonvocal', and 'silent' 8 sec epochs. Contrasting vocal relative to nonvocal sounds yielded the temporal voice areas (TVA; $p < .05$, cluster-level corrected), which was used as ROI for all subsequent analyses. The TVA is indicated by the white dotted outline in B-D. **(B)** An *emotion* discrimination effect was found in bilateral HG as indicated by a RE effect for repeated fear ([$F_{nat}F_{nat} > A_{nat}F_{nat}$]), but not for happiness. This selective effect for vocal fear was further specified by an interaction contrasts indicating significantly increased effects for fear relative to happy voices. These interaction effects are also shown in contrasts estimates in the middle and right panel (error bars ± 1SEM). **(C)** A general *acoustic* discrimination effect was also found in bilateral HG, which exhibited repetition suppression (RS) for the discrimination of vocal anger from both fear and happiness along HF cues ([$F_{nat}F_{nat} + H_{nat}H_{nat}] < [A_{HF}F_{nat} + A_{HF}H_{nat}$]), whereas for the same discrimination along HF cues the bilateral STG exhibited RE ([$F_{nat}F_{nat} + H_{nat}H_{nat}] > [A_{HF}F_{nat} + A_{HF}H_{nat}$]). **(D)** The *acoustic* discrimination effect for HF cues in the HG also showed an *emotion* specific lateralization effect. The left HG showed RS for discriminating vocal anger from happiness ([$A_{nat}H_{nat} > A_{HF}H_{nat}$]) along HF cues, while the right HG showed RS for discriminating vocal anger from fear ([$A_{nat}F_{nat} > A_{HF}F_{nat}$]) along HF cues. This laterality effect was not found in the STG, since the bilateral STG showed RE for discriminating vocal anger from fear along the HF cues ([$A_{nat}F_{nat} < A_{HF}F_{nat}$]). For all data: N = 22; peak activations are small-volume corrected (SVC) at a FWE-corrected threshold of $p < .05$ applied to the TVA functional mask.

significant main effect ($F_{1.31,18.31} = 501.675$, $p < .001$) with a Greenhouse-Geisser (GG) correction (because Mauchly's test detected a violation of sphericity). Post-hoc pairwise comparisons using the Bonferroni correction showed that angry stimuli were rated significantly higher on the angry scale [$M = 64.36$ (2.37), SEM always in brackets] than on any of the other 4 scales [happy: $M = .80$ (.28), $p < .001$; sad: $M = 1.85$ (.44), $p < .001$; neutral: $M = 2.74$ (.60), $p < .001$; fearful: $M = 2.67$ (.74), $p < .001$]; (b) *Fear*: significant main effect ($F_{1.69,23.71} = 97.86$, $p < .001$; GG corrected). Post-hoc pairwise comparisons using the Bonferroni correction showed that fearful stimuli were rated significantly higher on the fear scale [$M = 41.70$ (2.83)] than on any of the other 4 scales [anger: $M = 1.17$ (.45), $p < .001$; sadness: $M = 16.22$ (1.42), $p < .001$; neutrality: $M = 6.05$ (1.39), $p < .001$; happiness: $M = 2.34$ (.46), $p < .001$]; (c) *Happiness*: significant main effect ($F_{1.58, 22.10} = 117.491$, $p < .001$; GG corrected). Post-hoc pairwise comparisons using the Bonferroni correction showed that happy stimuli were rated significantly higher on the happy scale [$M = 53.83$ (3.56)] than on any of the other 4 scales [angry: $M = 4.64$ (1.00), $p < .001$; sad: $M = 6.02$ (1.40), $p < .001$; neutral: $M = 5.13$ (1.28), $p < .001$; fearful: $M = 2.40$ (.53), $p < .001$].

There was a main effect for arousal ratings between angry, happy, and fearful stimuli using a one-way repeated-measures ANOVA ($F_{2,28} = 18.748$, $p < .001$), and post-hoc pairwise comparisons using the Bonferroni correction showed that the

arousal rating for fearful stimuli [$M = 42.46$ (1.38)] was significantly lower compared to the arousal ratings for angry [$M = 56.49$ (1.49)], and happy stimuli [$M = 52.91$ (1.69)] ($p < .001$, and $p = .005$, respectively). No difference was detected between the arousal ratings of angry and happy stimuli ($p = .471$).

2.4.2. Rationale of the procedure

Our approach exploited the phenomena of RS and RE. RS is thought to reflect fatigue, or facilitation due to synaptic changes in the stimulus-response pathway (Grill-Spector et al., 2006). RE is instead thought to reflect strengthening of the stimulus representation (Rainer et al., 2004), often in response to poor perceptual conditions (Turk-Browne et al., 2007). The presence of RS or RE in a given brain region for trials in which two subsequent stimuli share a given feature indicates that the given brain region is involved in the processing of that feature. In this study, we exploited this neurophysiological phenomenon in order to probe TVA's and amygdala's sensitivity to discriminative acoustic voice features of affective vocalizations.

2.4.3. Stimulus manipulation

Stimuli were cropped to a standardized duration of 815 msec, and were presented in pairs, with first (S1) and second stimulus (S2) separated by a 300 msec gap. S1 was either native or

manipulated (as described below), and S2 was always native (i.e., in its original, natural version). The manipulation of S1 consisted in the alteration of one of the following three physical features of the sound: (a) the mean of the fundamental frequency (F0m), (b) the standard deviation of the fundamental frequency (F0sd), or (c) the high-frequency spectrum (HF). In the 'F0m' manipulation, the mean of the fundamental frequency of S2 was applied to S1, resulting in a condition in which the F0m was the repeated acoustic feature across the stimulus pair. In the 'F0sd' manipulation, the standard deviation of the fundamental frequency of S2 was applied to S1, resulting in a condition in which the F0sd was the repeated acoustic feature across the stimulus pair. The STRAIGHT toolbox for Matlab (www.wakayama-u.ac.jp/~kawahara/STRAIGHTadv/index_e.html) was used to manipulate the stimuli according to F0m and F0sd. Specifically, we extracted the F0 contour, changed either the F0m or F0sd of this extracted contour, and re-synthesized the stimuli using the F0 parameter. Changing angry voice to match the F0m of fearful (ΔF0m = 98.90 Hz, t_{15} = 7.754, p < .001) and happy (ΔF0m = 69.23 Hz, t_{15} = 5.168, p < .001) led to significant changes in the F0m; the same significant changes were found for changing the F0sd of angry voices to match the F0sd of fearful (ΔF0sd = 25.07 Hz, t_{15} = 5.200, p < .001) and happy voices (ΔF0sd = 20.90 Hz, t_{15} = 3.822, p = .002). In the 'HF' manipulation, a low-pass filter (Hann pass band filter of 0–1 kHz, including a roll-off of 100 Hz) was created in Matlab that eliminated almost all frequencies above 1 kHz was applied to S1, leading to a drop of its alpha ratio (Δalpha = 128.60 Pa^2/sec, t_{15} = 3.133, p = .007). The F0m and F0sd manipulations resulted in vocalizations that sounded marginally artificial as introduced by the resynthesis procedure. We therefore subjected all final stimuli to the resynthesis procedure as used for the F0m and F0sd manipulated voices to equate all stimuli in terms of artificiality. The acoustic manipulation of the voices did not affect the emotional quality of the voice in terms of their emotional intensity and in terms of their emotional valence, as indicated by evaluation data reported in one of our previous studies using identical acoustic manipulation procedures (Frühholz et al., 2016).

These manipulations were motivated by the fact that vocal emotions have been shown to differ in specific acoustic features (Banse & Scherer, 1996). Specifically, F0sd is one of the most distinguishing feature between angry and fearful vocalizations, and frequencies above 1 kHz are one the most distinguishing feature between angry and happy vocalizations (Banse & Scherer, 1996; Juslin & Laukka, 2003; Patel, Scherer, Bjorkner, & Sundberg, 2011). Equating the F0sd of S1 with that of S2, or filtering out of S1 the high-frequency cues, eliminate a distinguishing feature between S1 and S2, which should make the affective vocalizations expressed in S1 and S2 acoustically more similar. As a consequence of this feature-elimination approach, even though fearful or happy voices (S2) presented immediately after angry voices (S1) are 'unrepeated' stimuli at the emotional level (and should therefore not lead to RS or RE in regions sensitive to the emotional quality of the stimulus as a whole, independently of its acoustic features), the elimination of a distinguishing feature makes them 'repeated' at the feature level (and should therefore lead to RS or RE in regions sensitive to acoustic features).

2.4.4. Conditions

There were two sets of five conditions each: (a) in the five 'happiness' conditions, S2 was always a natural happy voice (H_{nat}); (b) in the five 'fear' conditions, S2 was always a natural fearful voice (F_{nat}). On every trial, S1 and S2 were vocalizations on the same pseudo-word from the same speaker, such that speaker, gender, and pseudo-word were held constant across S1–S2.

In each of the two sets of conditions, S2 (H_{nat} or F_{nat}) was preceded by an S1 in one of the following five configurations: (a) in the 'F0m' conditions ([$A_{F0m}H_{nat}$] and [$A_{F0m}F_{nat}$]), S1 was an angry voice whose F0m had been manipulated to match the F0m of S2 (A_{F0m}); (b) in the 'F0sd' conditions ([$A_{F0sd}H_{nat}$] and [$A_{F0sd}F_{nat}$]), S1 was an angry voice whose F0sd had been manipulated to match the F0sd of S2 (A_{F0sd}). In both (a) and (b), the manipulation was done on a trial-by-trial basis, and was therefore specific to the individual pair of S1–S2, which ensured that the manipulated feature (F0m or F0sd) was identical (hence 'repeated') across the pair; (c) in the 'HF' conditions ([$A_{HF}H_{nat}$] and [$A_{HF}F_{nat}$]), S1 was an angry voice whose high-frequency components (above 1 kHz) had been removed (A_{HF}); (d) in the 'unrepeated' conditions ([$A_{nat}H_{nat}$] and [$A_{nat}F_{nat}$]), S1 was a native (unmanipulated) angry voice (A_{nat}); (e) in the 'repeated' conditions ([$H_{nat}H_{nat}$] and [$F_{nat}F_{nat}$]), S1 was a native voice expressing the same emotion as S2 (H_{nat} or F_{nat}). Hence, 'repeated' trials consisted of two successive sound files in which the same speaker uttered the same pseudo-word with the same emotion. Since in our stimulus database we had two recordings for any given combination of speaker/pseudo-word/emotion, in half of the 'repeated' trials S1 and S2 were also the exact same sound file.

Conditions (a), (b), and (c) consisted therefore in 'acoustic' manipulations, and served to test for important acoustic voice features that influence the discrimination between angry and fearful or happy voices. Please note that the acoustic manipulation did not change the emotional content of the voices as determined by an evaluation of the stimuli (Frühholz et al., 2016), thus the voices were different in their emotional valence, but were identical or very similar in terms of one important acoustic feature. Conditions (d) and (e) consisted in 'emotion' manipulations, and served to test for overall emotion discrimination under conditions in which S1 and S2 expressed different ('unrepeated') or same ('repeated') affective value. This design resulted in a 2 (emotions: 'fear' and 'happiness') × 5 (conditions: 'F0m', 'F0sd', 'HF', 'unrepeated', and 'repeated') factorial design, with ten total conditions, schematically summarized in Fig. 1B.

2.4.5. Procedure

Thirty-two stimulus-pairs were presented in each of these ten conditions, resulting in a total of 320 trials, administered in four counterbalanced runs of 80 pseudo-randomized trials each. The average duration of each trial was approximately 6 sec (variable between 4930 and 7330 msec), including pre-stimulus fixation cross (700 ± 200 msec), S1 (815 msec), interstimulus gap (300 msec), S2 (815 msec), and interstimulus interval (3500 ± 1000 msec). Participants were instructed to pay attention to the voices and classify the gender of the speaker as quickly and accurately as possible through a right or left button press (order and side were counterbalanced

across participants) after the presentation of S2 (see Fig. 1A). The voice gender of S1 and S2 was always equal. This study is necessarily limited to the emotions and specific configuration of the design. Technical considerations related to the overall duration of the fMRI scanning session (approximately 1 h of net scanning time) prevented a more comprehensive testing across other emotions and stimulus combinations.

2.5. fMRI

2.5.1. Data acquisition
fMRI data were acquired on a 3T Siemens Trio System (Erlangen, Germany). For the experimental runs and the voice localizer scan, we used a T2*-weighted multiplexed echo-planar imaging sequence (Feinberg et al., 2010) with an acceleration factor of four. We used a partial volume acquisition of 28 slices (thickness/gap = 2/.4 mm, field of view [FoV] = 192 mm, in-plane resolution 2×2 mm, flip angle = $54°$) to cover the temporal lobe and the amygdalae. The sequence had a time to repetition (TR)/time to echo (TE) of 650/30 msec. Finally, high-resolution structural images were obtained through a T1-weighted sequence (192 contiguous 1 mm slices, TR/TE/TI = 1900/2.27/900 msec, FoV = 296 mm, in-plane resolution of 1×1 mm).

2.5.2. Data analysis
fMRI images were analyzed through the Statistical Parametric Mapping (SPM) software (version 8; Wellcome Trust Centre for Neuroimaging, London). Functional images were realigned and coregistered to the anatomical image. Anatomical images were segmented using the "new segment" option in SPM, and normalizations parameters were estimated using the unified approach in DARTEL. During the application of normalization parameters, functional images were resampled at a 2 mm isometric voxel size, and normalized to the Montreal Neurological Institute (MNI) space. A Gaussian smoothing of full-width at half maximum 6 mm isometric was applied.

For the individual subject (first-level) analysis, we used a general linear model (GLM) where a boxcar function (defined by the onset of S1 and the offset of S2) was convolved with the canonical hemodynamic response function. For the main experiment, separate regressors were created for each of the ten conditions (Fig. 1B).

For the voice localizer, two regressors were created for the two conditions ('vocal', and 'non-vocal'). In both the main experiment and the localizer, six additional regressors resulting from the motion correction parameters, as well as four regressors resulting from heart rate and respiration (measured during the fMRI through the BIOPAC system; www.biopac.com) were included in the model, using the RetroIcor function (Glover, Li, & Ress, 2000), as regressors of no interest, in order to minimize false positive activations due to head movements and physiological activity. The imaging data from the localizer were interrogated through the contrast [vocal > non-vocal] in order to identify voice-sensitive clusters in the superior temporal cortex (temporal voice area, TVA, Fig. 2A). These activation clusters were then used as regions-of-interest (ROI) for the analysis of the imaging data from the main experiment. We were specifically interested in examining the effect of repetition either of the emotion or of specific

acoustic features of the stimulus on the activity of the amygdala and of the TVA. For the ROI analysis on the amygdala, a mask corresponding to the bilateral amygdalae was created anatomically from the Anatomical Automatic Labeling (AAL) atlas (Tzourio-Mazoyer et al., 2002).

Functional activations were initially thresholded at a voxel-threshold of $p < .05$ uncorrected, but here we report only functional peak activations that survived a small volume correction (SVC) based on a family-wise error (FWE) correction threshold of $p < .05$, and a cluster extend threshold of $k = 10$. The procedure of SVC was chosen give our region-of-interest approach on the amygdala and the TVA. SVC provides a method for accounting for the multiple comparison problem, while increase the sensitivity for a local signal in selected ROIs.

For the ROI analysis on the TVA, a mask corresponding to the bilateral TVA was created functionally from the voice localizer ([vocal > non-vocal] contrast), applying a cluster-extent threshold of 100 voxels and a voxel-level correction of $p = .001$, which, together, correspond to a cluster-level correction of $p < .05$ based on the estimated smoothness of the data (Fig. 2A). Clusters were anatomically labeled through the SPM Anatomical Toolbox (cytoarchitectonic probabilities) based on the MNI space. For the ROI analysis on the amygdala, a bilateral anatomical mask was applied as described above.

The analysis of the fMRI data included two major types of the analysis. First, in order to test the ROIs' sensitivity to *emotion* discrimination (i.e., emotion discrimination effect), we compared functional activation within our two ROIs during trials in which S1 and S2 expressed the same emotion as opposed to during trials in which S1 and S2 expressed a different emotion. For example, we compared activity in trials with repeated native vocal fear ([$F_{nat}F_{nat}$]) to trials where native vocal fear was preceded by a native vocal anger ([$A_{nat}F_{nat}$]). The same comparisons were applied to trials involving vocal happiness.

Second, in order to test the ROIs' sensitivity to discriminate emotions along *acoustic* features (i.e., acoustic discrimination effect), we compared the activation during trials in which S1 and S2 expressed the same emotion (e.g., [$F_{nat}F_{nat}$]) to trials in which S1 and S2 expressed a different emotion but the same or similar acoustic feature (e.g., [$A_{F0m}F_{nat}$]). In order to further characterize the interaction between vocal emotion and acoustic manipulation we compared activation within our ROIs during trials in which S1 and S2 expressed different emotions (e.g., [$A_{nat}F_{nat}$]) to trials in which S1 and S2 expressed different emotions but the same acoustic feature (e.g., [$A_{HF}F_{nat}$]). These different contrasts revealed locations of peak activations in the TVA. Accordingly, contrast estimates in selected ROIs of the TVA were derived from the mean signal in a 3 mm sphere around these peak voxel (see Fig. 1C and Fig. 2B–D).

3. Results

3.1. Behavioral performance

A 2 (emotion of S2: 'fear' or 'happiness') $\times$ 5 (condition: 'F0m', 'F0sd', 'HF', 'repeated', 'unrepeated') repeated-measures

ANOVA on response times for the gender decision task revealed a main effect of emotion ($F_{1,21} = 153.32$, $p < .001$), with faster responses to stimulus pairs expressing fear [$M = 953$ msec (55)] than to stimulus pairs expressing happiness [$M = 1272$ msec (68)]. There was no main effect of condition ($F_{1.75,36.74} = 1.69$, $p = .200$, GG corrected). This main effect of emotion was accompanied by a marginal interaction effect emotion × condition ($F_{4,84} = 2.57$, $p = .044$), with faster responses to fearful voices preceded by F0m-manipulated angry voices ([$A_{F0m}F_{nat}$]; $M = 941$ msec [56]) compared to fearful voices preceded by HF-manipulated angry voices ([$A_{HF}F_{nat}$]; $M = 993$ msec [55]; $t_{21} = 2.157$, $p = .043$).

3.2. Functional definition of the TVA

Comparing functional activations for human voices ('vocal') versus non-human sounds ('non-vocal') produced extended bilateral activation within the superior temporal gyrus (STG) and superior temporal sulcus (STS), bilaterally (Fig. 2A). These clusters served to define the TVA on which we conducted the region-of-interest (ROI) analysis of the main experiment.

3.3. Functional data of the main experiment

We focused on the response of two ROIs. These ROIs were the TVA, defined functionally as described above, and the amygdala, defined anatomically through a mask created with the Anatomical Automatic Labeling (AAL) atlas (Tzourio-Mazoyer et al., 2002). Functional activations maps for determining 'acoustic' and 'emotion' effects were initially set at a voxel-threshold of $p < .05$. However, here we report only functional peak activations that survived a small volume correction (SVC) based on a family-wise error (FWE) corrected threshold of $p < .05$, and a cluster extend threshold of $k = 10$.

In order to test for 'emotion' discrimination effects, we compared functional activation within our two ROIs (i.e., TVA and amygdala) during trials in which S1 and S2 expressed the same emotion ('repeated' condition) as opposed to during trials in which S1 and S2 expressed a different emotion ('unrepeated' condition) (Fig. 1B, lower panel). We found that a region of the left amygdala showed RE upon repeated exposure to fearful voices ([$F_{nat}F_{nat} > A_{nat}F_{nat}$]; MNI xyz [−20 −6 −12], $p = .003$), but not to happy voices ([$H_{nat}H_{nat} > A_{nat}H_{nat}$]; Fig. 1C). A very similar region was found to exhibit an interaction effect between repetition and emotion ([$F_{nat}F_{nat} > A_{nat}F_{nat}$] > [$H_{nat}H_{nat} > A_{nat}H_{nat}$]; MNI xyz [−18 −4 −12], $p = .033$) (Fig. 1D). These combined data indicate that amygdala responses to repeated fearful and happy voices are significantly different from each other, as opposed to the former (and not the latter) being significantly different from zero. No main effect of repetition (i.e., across emotions) was observed ([$F_{nat}F_{nat} + H_{nat}H_{nat}$] > [$A_{nat}F_{nat} + A_{nat}H_{nat}$]).

Visual observation of the two largely overlapping regions of the left amygdala found to selectively respond to fearful voices ([−20 −6 −12] and [−18 −4 −12]) suggests that both clusters are likely to lie mainly within the basolateral complex, although parts of this cluster of activation also fell within the SF complex of the amygdala. However, the involvement of neighbouring subregions is also possible. Probabilistic labelling based on the SPM Anatomy toolbox (Eickhoff et al., 2005)

identifies the clusters (peak-intensity voxels [−20 −6 −12] and [−18 −4 −12]) as likely to be located within the superficial nucleus (SF, 80% probability for both clusters), the basolateral complex, and the centromedian nucleus (70%, and 60% probability, respectively, for [−20 −6 −12]).

The fear-related RE in the amygdala was detected in the 'repeated' condition not only in comparison to the 'unrepeated' condition, in which S1–S2 differed both in emotion and in acoustic features (i.e., the emotion discrimination effect), but also in comparison to conditions in which S1–S2 differed in emotion but shared acoustic features, such as the 'F0m' condition ([$F_{nat}F_{nat} > A_{F0m}F_{nat}$], [−20 −6 −12] (probabilistic labelling based on the SPM Anatomy toolbox: 80%, 70%, and 60% for superficial nucleus, basolateral complex, and centromedian nucleus respectively), $z = 4.76$, $p < .001$) or the 'F0sd' condition ([$F_{nat}F_{nat} > A_{F0sd}F_{nat}$], [−20 −6 −14] (probabilistic labelling: 80% for basolateral complex), $z = 4.20$, $p = .002$) (Fig. 1E). The latter two contrasts represent an acoustic discrimination effect, and indicate that the amygdala is sensitive to physical stimulus features and discriminates angry from fearful voices according to vocal pitch features. There was no main effect of the 'F0m' of 'F0sd' condition in general for fearful and happy voices ([$F_{nat}F_{nat} + H_{nat}H_{nat} > A_{F0m}F_{nat} + A_{F0m}H_{nat}$] and [$F_{nat}F_{nat} + H_{nat}H_{nat} > A_{F0sd}F_{nat} + A_{F0sd}H_{nat}$]). An interaction effect was instead found between condition and emotion, with the left amygdala discriminating fearful from angry voices (but not happy from angry voices) both in the F0m ([$F_{nat}F_{nat} > A_{F0m}F_{nat}$] > [$H_{nat}H_{nat} > A_{F0m}H_{nat}$], [−22 −6 −14] (probabilistic labelling: 80% for basolateral complex), $z = 3.62$, $p = .020$) and the F0sd conditions ([$F_{nat}F_{nat} > A_{F0sd}F_{nat}$] > [$H_{nat}H_{nat} > A_{F0sd}H_{nat}$], [−20 −6 −14] [probabilistic labelling: 80% for basolateral complex], $z = 3.82$, $p = .010$) (Fig. 1F).

Similarly to the emotion discrimination effect found in the amygdala, a region within the bilateral Heschl's gyrus (HG) located in the TVA showed RE upon repeated exposure to fearful voices ([$F_{nat}F_{nat} > A_{nat}F_{nat}$]; left [−44 −28 2], $z = 4.28$, $p = .012$; right [44 −26 −6], $z = 5.04$, $p < .001$), but not to happy voices ([$H_{nat}H_{nat} > A_{nat}H_{nat}$]) (Fig. 2B). Again no main effect of repeating affective voices was observed ([$F_{nat}F_{nat} + H_{nat}H_{nat}$] > [$A_{nat}F_{nat} + A_{nat}H_{nat}$]). An interaction effect was instead found between repetition and emotion, with repeated fearful voices eliciting larger activation of the bilateral TVA compared to repeated happy voices ([$F_{nat}F_{nat} > A_{nat}F_{nat}$] > [$H_{nat}H_{nat} > A_{nat}H_{nat}$]; [−44 −20 10], $z = 4.22$, $p = .016$; [44 −26 −1], $z = 4.03$, $p = .016$) (Fig. 2B, right panel).

However, unlike the amygdala's sensitivity to vocal pitch features while discriminating negative voices, the TVA showed sensitivity only to HF voice cues, but during the discrimination of both negative and positive voices. Specifically, we found that the elimination of the HF cues in S1 generally led to RE within the bilateral STG [$F_{nat}F_{nat} + H_{nat}H_{nat}$] < [$A_{HF}F_{nat} + A_{HF}H_{nat}$]; left [−60 −20 4], $z = 4.04$, $p = .030$; right [54 −40 8], $z = 4.50$, $p = .005$) (Fig. 2C) while discriminating vocal anger from both vocal fear and happiness. No such effects were found for the pitch-related cues. The acoustic discrimination effect related to the HF cues in the STG was especially significant for discriminating vocal anger from fear ([$A_{nat}F_{nat} < A_{HF}F_{nat}$]; left [−60 −22 0] $z = 4.46$, $p = .006$; right [54 −30 0], $z = 5.81$, $p < .001$) (Fig. 2D), but

not for discriminating vocal anger from happiness ([$A_{nat}H_{nat} < A_{HF}H_{nat}$] along HF cues. No further interaction effects were observed.

Whilst the STG overall showed RE, the HG exhibited mainly RS effects. We found RS in bilateral HG to HF cues ([$F_{nat}F_{nat} + H_{nat}H_{nat}$] > [$A_{HF}F_{nat} + A_{HF}H_{nat}$], left [−40 −22 6], $z = 5.00$, $p = .001$; right [38 −22 10], $z = 5.13$, $p < .001$) while discriminating vocal anger from other affective voices (Fig. 2C). No such effects were found for the pitch-related cues. The effect concerning the HF cues was differentially lateralized. The left HG showed RS while discriminating vocal anger from *happiness* along HF cues ([$A_{nat}H_{nat} > A_{HF}H_{nat}$]; [−36 −24 10], $z = 3.90$, $p = .049$), whereas the right HG showed RS while discriminating vocal anger from *fear* along HF cues ([$A_{nat}F_{nat} > A_{HF}F_{nat}$]; [40 −28 12], $z = 4.35$, $p = .009$) (Fig. 2D, 'laterality effect' graphs). Again, no further interaction effects were observed.

3.4. Control analysis in visual area V1

To finally test if the RS and RE results that we found in the TVA and the amygdala were specific to these regions, we computed the same contrasts (as reported in the results above) for area V1 (BA17). We chose V1 as a control because it is a visual area, hence not expected to respond to emotional and acoustic manipulations of vocal stimuli. We again used SVC in V1 based on an initial voxel-threshold of $p = .05$, and an FWE-corrected threshold of $p = .05$ in this ROI. None of the above contrasts revealed activity in V1, indicating that the effects reported in our results are specific to the TVA and the amygdala.

4. Discussion

The purpose of our study was to test whether and to what extent the amygdala and TVA respond to the voice's emotional quality and/or physical features. In order to address our region-specific question, we took a ROI approach, comparing and contrasting the responses of amygdala and TVA to affective vocalisation. Summarizing our results, we found: first, that the left amygdala and the voice-sensitive auditory cortex (i.e., the TVA) discriminated vocal anger from fear, but not from happiness (emotion discrimination effect); second, that the amygdala discriminated angry from fearful voices based on the mean and the variation of vocal pitch, whereas the bilateral TVA discriminated angry from other affective voices based on high-frequency (HF) cues (acoustic discrimination effect). The latter effect was specifically significant for discriminating angry from fearful voices, and this effect was predominant in the STG. Third, unlike the STG the HG showed a lateralization effect. The left HG used HF cues while discriminating vocal anger from happiness, and the right HG used HF cues while discriminating vocal anger from fear. Taken together, these results indicate that both the amygdala and the TVA are differentially involved in the decoding of the affective quality of the voice, and that this decoding includes the processing of physical voice features.

Concerning the emotion effect in the amygdala, our results show that only the left amygdala responded to the affective quality of the voice. This laterality effect is consistent with previous reports (Frühholz, Ceravolo, & Grandjean, 2012; Frühholz et al., 2015). Our results also showed that this left-specific responsiveness is emotion-specific, as the neural adaptation effects occurred only upon repeated exposure to stimuli expressing fear, whereas no adaptation was observed upon repeated exposure to stimuli expressing happiness. This emotion-specificity for fear suggests that the timing and stage of involvement of the amygdala in the processing of affective information may vary according to the significance of the stimulus for survival. Specifically, it is possible that the amygdala processes perceptual information involving threatening vocal information (i.e., anger and fear) at an earlier stage of processing compared to perceptual information involving happiness, which is likely to be less determinant for survival. The discrimination of threatening information especially of anger and fear is vital because both emotions signal potential danger but each requires a different behavioral response. Note that, although fearful and happy stimuli differed in arousal (see pre-test evaluation results in Methods section), this difference should not affect the significance of the findings in the amygdala, because arousal was *lower* for fearful voices. The lower arousal would normally be expected to associate with weaker (not stronger) response in the amygdala. Therefore, our observation of a stronger response in the amygdala suggests that arousal is unlikely to have driven the activation. On the contrary, the lower level of arousal induced by the fearful stimuli makes our findings more conservative, as it points to a preferential sensitivity of the amygdala to negative emotions.

This preferential sensitivity is consistent with our finding that the left amygdala discriminated vocal anger and fear according to central vocal pitch features, which might be to be relayed directly from the auditory thalamus and thus bypassing the auditory cortex (Pannese et al., 2015). This transfer of acoustic pitch features from the auditory thalamus to the amygdala seems likely, because the auditory cortex did not show sensitivity to vocal pitch features, and is therefore unlikely to have been the source that provided this information to the amygdala. The mean (Patel et al., 2011) and especially the variation and dynamics of vocal pitch (Banse & Scherer, 1996; Juslin & Laukka, 2003; Leitman et al., 2010) are two of the most important cues, which listeners use to discriminate vocal anger and fear. The perceptual discrimination between these two emotions might therefore be supported by a rapid acoustic feature analysis in the amygdala along discriminative vocal pitch cues. These results point to a sensitivity of the amygdala to physical features of relevant stimuli beyond its sensitivity to their emotional quality.

Both the amygdala and the TVA showed lateralized sensitivity to the repetition of affective and acoustic qualities of the voice. Lateralized repetition priming in the amygdala has already been reported in the visual modality (Bourne, Vladeanu, & Hole, 2009). Our results show that a similar effect occurs in the auditory domain as well, and that it extends to the TVA. We found that the left and right medial TVA responded differently to different affective qualities, the left being sensitive to stimuli expressing happiness, the right to stimuli expressing fear. This sensitivity was especially found when discriminating these emotions from vocal anger. Moreover, different subregions within the TVA were selectively

sensitive to physical HF voice cues, and differentially responded to their elimination either with RE or RS when discriminating vocal anger from vocal happiness and vocal fear. This finding is consistent with previously proposed patterns of left-right hemispheric dominance for positive and negative emotions (Davidson, Ekman, Saron, Senulis, & Friesen, 1990), and extends the relevance of this lateralization to include the discrimination of emotions along central voice features.

The adaptation effects related to RE/RS have been used as a methodological tool to determine the stage at which different levels of neural processing occur along afferent perceptual pathways (Rotshtein, Henson, Treves, Driver, & Dolan, 2005; Vuilleumier, Henson, Driver, & Dolan, 2002), from low-level perceptual processing (Weigelt, Limbach, Singer, & Kohler, 2012), to higher-level cortical mechanisms (Pannese & Hirsch, 2011, 2013). Regions exhibiting RS receive predominantly bottom-up information concerning low-level physical stimulus properties (Kriegeskorte, Formisano, Sorger, & Goebel, 2007), whereas regions exhibiting RE receive both bottom-up and top-down information to strengthen object representation and establish perceptual expectations (de Gardelle et al., 2013). The clusters we found in HG and STG were both contained within the TVA. Hence, their different types of response (RS in HG, RE in STG) suggest that, whilst both HG and STG are sensitive to the voice, their sensitivity rests upon different mechanisms. It specifically points to a complementary acoustic processing in these two cortical subregions (Leaver & Rauschecker, 2010), whereby the low-level HG presumably treats the voice as a collection of acoustic features, whereas the higher-level STG represents it as an integrated auditory object, whose representation is strengthened under unfamiliar or degraded perceptual conditions. This notion is consistent with the fact that the stimuli treated with the HF manipulation (to which the STG responded with RE, hence, with an enhanced activity) lacked a significant component of the energy spectrum, resulting in poor clarity of the acoustic signal (which became comparable to a muffled sound). The RE may therefore be a manifestation of this strengthening of the perceptual representation within the STG. This response is likely to involve computing the ratio between HF and low-frequency cues (i.e., the so-called 'alpha-ratio') as a discriminative second-order affective voice cue (Attwood et al., 2011), which is based on prior auditory cortical processing (Leaver & Rauschecker, 2010).

The patterns of RS and RE observed at the neural level were not matched by similar patterns at the behavioral level, where the only effect was faster responses to stimuli expressing fear than to stimuli expressing happiness (independently of the conditions). This mismatch indicates that the difference in brain activity found in the various regions expressing RS and RE are not due to differences in behavioral response (i.e., overall duration of engagement with trials). This finding is also consistent with previous evidence that adaptation effects in object-selective cortex are independent of task performance (Sayres & Grill-Spector, 2006).

5. Conclusions

Taken together, these results challenge a previously held view whereby acoustic and affective processing of auditory information occurs in distinct regions: the former in the auditory cortex, and the latter in the amygdala, once acoustic features have been previously processed in auditory cortical areas (Kumar et al., 2012; Wiethoff, Wildgruber, Grodd, & Ethofer, 2009). By showing that both the amygdala and the auditory cortex respond to the repetition of physical features of the vocal signal, and that this response is modulated by the vocal signal's emotional content, our data suggest instead that both amygdala and auditory cortex decode the affective quality of the voice, and that the amygdala does so not only by extracting the emotional quality from previously processed acoustic information, but also by processing the acoustic information itself, when this is relevant to the identification of the voice's affective quality.

REFERENCES

Attwood, B. K., Bourgognon, J. M., Patel, S., Mucha, M., Schiavon, E., Skrzypiec, A. E., et al. (2011). Neuropsin cleaves EphB2 in the amygdala to control anxiety. *Nature, 473,* 372–375.

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

Belin, P., & Zatorre, R. J. (2000). Voice-selective areas in human auditory cortex. *Nature, 403,* 309.

Bourne, V. J., Vladeanu, M., & Hole, G. J. (2009). Lateralised repetition priming for featurally and configurally manipulated familiar faces: Evidence for differentially lateralised processing mechanisms. *Laterality, 14,* 287–299.

Davidson, R. J., Ekman, P., Saron, C. D., Senulis, J. A., & Friesen, W. V. (1990). Approach-withdrawal and cerebral asymmetry: Emotional expression and brain physiology. I. *Journal of Personality and Social Psychology, 58,* 330–341.

Eickhoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K., et al. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *NeuroImage, 25,* 1325–1335.

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,* 191–200.

Fecteau, S., Belin, P., Joanette, Y., & Armony, J. L. (2007). Amygdala responses to nonlinguistic emotional vocalizations. *NeuroImage, 36,* 480–487.

Feinberg, D. A., Moeller, S., Smith, S. M., Auerbach, E., Ramanna, S., Glasser, M. F., et al. (2010). Multiplexed echo planar imaging for sub-second whole brain FMRI and fast diffusion imaging. *PLoS One, 5,* e15710.

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

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. *Neuroscience and Biobehavioral Reviews, 37*, 24–35.

Frühholz, S., Hofstetter, C., Cristinzio, C., Saj, A., Seeck, M., Vuilleumier, P., et al. (2015). Asymmetrical effects of unilateral right or left amygdala damage on auditory cortical processing of vocal emotions. *Proceedings of the National Academy of Sciences of the United States of America, 112*, 1583–1588.

Frühholz, S., Klaas, H. S., Patel, S., & Grandjean, D. (2014). Talking in Fury: The Cortico-Subcortical Network Underlying Angry Vocalizations. *Cerebral Cortex, 25*, 2752–2762.

Frühholz, S., Trost, W., & Grandjean, D. (2014). The role of the medial temporal limbic system in processing emotions in voice and music. *Progress in Neurobiology, 123*, 1–17.

Frühholz, S., van der Zwaag, W., Seanz, M., Belin, P., Schobert, A.-K., Vuilleumie, P., et al. (2016). Neural decoding of discriminative auditory object features depends on their socio-affective value. *Social, Cognitive, and Affective Neuroscience, 11*, 1638–1649.

de Gardelle, V., Waszczuk, M., Egner, T., & Summerfield, C. (2013). Concurrent repetition enhancement and suppression responses in extrastriate visual cortex. *Cerebral Cortex, 23*, 2235–2244.

Glover, G. H., Li, T. Q., & Ress, D. (2000). Image-based method for retrospective correction of physiological motion effects in fMRI: RETROICOR. *Magnetic Resonance in Medicine, 44*, 162–167.

Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: Neural modles of stimulus-specific effects. *Trends in Cognitive Science, 10*, 14–23.

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expressions and music performance: Different channels, same code? *Psychological Bulletin, 129*, 770–814.

Kriegeskorte, N., Formisano, E., Sorger, B., & Goebel, R. (2007). Individual faces elicit distinct response patterns in human anterior temporal cortex. *Proceedings of the National Academy of Sciences of the United States of America, 104*, 20600–20605.

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. *Journal of Neuroscience, 32*, 14184–14192.

Leaver, A. M., & Rauschecker, J. P. (2010). Cortical representation of natural complex sounds: Effects of acoustic features and auditory object category. *Journal of Neuroscience, 30*, 7604–7612.

Leitman, D. I., Laukka, P., Juslin, P. N., Saccente, E., Butler, P., & Javitt, D. C. (2010). Getting the cue: Sensory contributions to auditory emotion recognition impairments in schizophrenia. *Schizophrenia Bulletin, 36*, 545–556.

Marsh, A. A., Ambady, N., & Kleck, R. E. (2005). The effects of fear and anger facial expressions on approach- and avoidance-related behaviors. *Emotion, 5*, 119–124.

Muller, N. G., Strumpf, H., Scholz, M., Baier, B., & Melloni, L. (2013). Repetition suppression versus enhancement—it's quantity that matters. *Cerebral Cortex, 23*, 315–322.

Naccache, L., & Dehaene, S. (2001). The priming method: Imaging unconscious repetition priming reveals an abstract representation of number in the parietal lobes. *Cerebral Cortex, 11*, 966–974.

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

Pannese, A., & Hirsch, J. (2011). Self-face enhances processing of immediately preceding invisible faces. *Neuropsychologia, 49*, 564–573.

Pannese, A., & Hirsch, J. (2013). Unconscious neural specificity for 'Self' and the brainstem. *Journal of Consciousness Studies, 20*, 169–179.

Patel, S., Scherer, K. R., Bjorkner, E., & Sundberg, J. (2011). Mapping emotions into acoustic space: The role of voice production. *Biological Psychology, 87*, 93–98.

Rainer, G., Lee, H., & Logothetis, N. K. (2004). The effect of learning on the function of monkey extrastriate visual cortex. *PLoS Biology, 2*, E44.

Rotshtein, P., Henson, R. N., Treves, A., Driver, J., & Dolan, R. J. (2005). Morphing Marilyn into Maggie dissociates physical and identity face representations in the brain. *Nature Neuroscience, 8*, 107–113.

Sayres, R., & Grill-Spector, K. (2006). Object-selective cortex exhibits performance-independent repetition suppression. *Journal of Neurophysiology, 95*, 995–1007.

Turk-Browne, N. B., Yi, D. J., Leber, A. B., & Chun, M. M. (2007). Visual quality determines the direction of neural repetition effects. *Cerebral Cortex, 17*, 425–433.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., et al. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage, 15*, 273–289.

Vuilleumier, P., Henson, R. N., Driver, J., & Dolan, R. J. (2002). Multiple levels of visual object constancy revealed by event-related fMRI of repetition priming. *Nature Neuroscience, 5*, 491–499.

Weigelt, S., Limbach, K., Singer, W., & Kohler, A. (2012). Orientation-selective functional magnetic resonance imaging adaptation in primary visual cortex revisited. *Human Brain Mapping, 33*, 707–714.

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