OXFORD

Social Cognitive and Affective Neuroscience, 2016, 793–802

doi: 10.1093/scan/nsw004
Advance Access Publication Date: 8 January 2016
Original article

Proximal vocal threat recruits the right voice-sensitive auditory cortex

Leonardo Ceravolo,[1,2] Sascha Frühholz,[1,2,3] and Didier Grandjean[1,2]

[1]Neuroscience of Emotion and Affective Dynamics Lab, Department of Psychology, [2]Swiss Center for Affective Sciences, University of Geneva, CH-1202 Geneva, Switzerland and [3]Department of Psychology, University of Zurich, 8050 Zurich, Switzerland

Correspondence should be addressed to Leonardo Ceravolo, Swiss Center for Affective Sciences Biotech Campus 9, Chemin des Mines, CH-1202, Geneva, Switzerland. E-mail: leonardo.ceravolo@unige.ch.

Abstract

The accurate estimation of the proximity of threat is important for biological survival and to assess relevant events of everyday life. We addressed the question of whether proximal as compared with distal vocal threat would lead to a perceptual advantage for the perceiver. Accordingly, we sought to highlight the neural mechanisms underlying the perception of proximal *vs* distal threatening vocal signals by the use of functional magnetic resonance imaging. Although we found that the inferior parietal and superior temporal cortex of human listeners generally decoded the spatial proximity of auditory vocalizations, activity in the right voice-sensitive auditory cortex was specifically enhanced for proximal aggressive relative to distal aggressive voices as compared with neutral voices. Our results shed new light on the processing of imminent danger signaled by proximal vocal threat and show the crucial involvement of the right mid voice-sensitive auditory cortex in such processing.

Key words: emotion; fMRI; proximal; spatial hearing; voice

Introduction

Living beings undergo a constant evolutionary process according to which behavioral as well as neural mechanisms allowing the species' survival is favored. In the human literature, research on the visual sensory modality brought to light several mechanisms crucially involved in processing environmental threat, whereas research on the auditory modality was underrepresented in such context. In fact, while threat proximity can be assessed visually, the accurate detection and localization of threat often relies on our auditory system, a mechanism that is more specifically efficient in the case of vocal threat processing.

Human neuroimaging studies highlighted the role of the superior temporal gyrus (STG) and the superior temporal sulcus in processing vocal as opposed to non-vocal signals (Belin *et al.*, 2000). Subparts of these temporal voice-sensitive areas were further shown to respond specifically to voices signaling threat (Buchanan *et al.*, 2000; Grandjean *et al.*, 2005; Leitman *et al.*, 2010; Ethofer *et al.*, 2011; Frühholz *et al.*, 2012; Witteman *et al.*, 2012; Frühholz and Grandjean 2013a), highlighting the fundamental meaning and high importance of such events for biological survival in humans. This line of argument was further strengthened by studies showing that emotional expressions, despite not in the focus of attention could still be processed notably by the amygdala and could hence have a significant influence on the distribution of spatial attention both in the visual (Vuilleumier and Schwartz, 2001) and the vocal domain (Sander *et al.*, 2003; Wambacq *et al.*, 2004) or more generally in the auditory domain, involving the lateral parietal cortex (Griffiths and Warren, 2002). Such survival mechanisms involving an enhanced processing of biologically relevant events are known to be shared by most mammals and animals (Öhman, 1986) and a better understanding of their role in the human spatial hearing system is of high importance. In fact, the fast and accurate decoding and localization of vocal signals of threat is especially relevant for humans, even more when these signals are relatively close to the perceiver, as they require fast and adaptive behavioral responses. Proximal vocal threat may imply imminent danger and its processing is thus of high importance for

Received: 16 July 2015; **Revised:** 16 November 2015; **Accepted:** 4 January 2016

biological survival (McNally and Westbrook, 2006), even though this topic is rather poorly studied in the literature. In fact, while the general framework behind emotional processing was studied in detail, literature on the specific perceptual and neural decoding of the proximity of relevant vocalizations in terms of auditory threat has been largely unexplored. The common behavioral and neural mechanisms underlying threat distance were in fact investigated in other sensory modalities such as vision (Mobbs *et al.*, 2007), but the vocal couterpart of such studies was surprisingly not investigated.

No study to date investigated the perception of vocal threat in space using fMRI (functional magnetic resonance imaging) and the general aim of the present study was thus to reveal the perceptual and neural mechanisms underlying the processing of proximal (i.e. imminent danger) relative to distal auditory threat (i.e. potential danger). We asked 14 healthy human listeners to evaluate the distance of virtually spatialized aggressive and neutral voices while we recorded their brain activity by using fMRI. Using this paradigm, we were able to test our predictions according to which the distance of aggressive voices would be more accurately perceived than that of neutral voices. Data from an additional, independent behavioral control group strengthened the results from the fMRI group. Regarding neuroimaging data, we expected enhanced activity in brain regions involved both in voice- and space-related perception. More specifically, we hypothesized an involvement of brain regions known for their perceptual role in emotion and voice processing (the superior part of the temporal cortex, the amygdala) as well as brain regions involved in spatial hearing (the lateral/posterior parietal cortex). Regarding the interaction between emotion and proximity, we hypothesized a specific involvement of the lateral parietal and superior temporal cortices as the neural mechanisms underlying the perception of proximal compared with distal aggressive voices.

Materials and methods

Participants, stimuli, task and procedure during the fMRI experiment

Participants. Fourteen right-handed, healthy, native or highly proficient French-speaking participants (six male, eight female, mean age 23.07 years, s.d. 3.95) were included in an fMRI study. All participants were naïve to the experimental design and study, had normal or corrected-to-normal vision, normal hearing and no history of psychiatric or neurologic incidents. Participants gave written informed consent for their participation in accordance with ethical and data security guidelines of the University of Geneva. The study was approved by the Ethical Committee of the University of Geneva and was conducted according to the Declaration of Helsinki.

Stimuli. Ten professional actors (five male and five female) pronounced the vowel /a/ in a neutral or an aggressive vocal tone, leading to 20 stimuli in total (five aggressive and five neutral /a/'s from the male actors; the same from the female actors). These stimuli, presented binaurally, were taken from the large and validated Geneva Multimodal Expression Portrayals (GEMEP) database (Bänziger and Scherer, 2007) and normalized in terms of mean intensity for each distance across emotions. To ensure a lateralized and proximal-to-distant stimulus presentation, we performed a lateralization process by using a semi-individual head-related transfer function (HRTF) with the Panorama 5 toolbox implemented in Sony SoundForge software (Sony Creative Software Inc., Middleton, WI, USA). This convolution takes into account head size and ear shape and uses fine modulations of wave amplitude and interaural time difference in order to virtually spatialize sounds. We thus used slightly delayed interaural time differences to virtually lateralize the voice stimuli, meaning that even though the sound was perceived in the left auditory space, it was actually presented to both ears with a slight delay to the ear opposite to the space of presentation (the right ear in this example). This convolution represents the most accurate and ecologically valid means for matching the physics of real-life sound spatialization. The azimuthal angle was 20° (10° to the front and 10° to the back of the ear/head of the participant) and elevation was kept to ear-plane level, namely to 0°. The use of HRTFs to create a diotic as opposed to a dichotic stimulus presentation significantly improved ecological validity and took into account a double dissociation, suggesting that different neural networks serve the detection ability of auditory space *vs* that of the ears (Clarke and Thiran, 2004).

We used nine different HRTFs all implemented in our Eprime (Psychology Software Tools, Pittsburgh, PA, USA) script; each participant was first trained on a short demonstration of the task, including all nine HRTFs. The optimal HRTF among the nine was selected for each participant according to the accuracy in evaluating the distance of the presented voices (minimum 75% of correct responses). Stimuli used for the demonstration were excluded from the distance evaluation task. The 1100-ms \a\'s were lateralized at four virtual distances (1, 5, 10 and 20 m) in a total of eight locations in the left/right auditory spaces. The total number of trials for one participant was 384, with 192 aggressive and 192 neutral voices (Emotion factor). Each emotion was split into four distances (Distance factor), giving a total of 48 trials for each emotion and distance. For each distance, stimuli were lateralized either in the left or the right auditory space (Space laterality factor), leading to 24 trials for each emotion, distance and space laterality.

Even though a large cohort had already evaluated the auditory stimuli in terms of emotions, we asked participants to evaluate them at the end of the experiment in order to have a subjective value with their own judgments (Supplementary Figures S1 and S2).

Experimental design. The experiment consisted of one session of ~20 min during which the spatialized /a/'s were presented binaurally, each at a time, through pneumatic, MR-compatible headphones (MR confon GmbH, Germany). All stimuli were presented with the same intensity throughout the experiment, namely at a 70 dB sound-pressure level. Participants were asked to respond as quickly and accurately as possible regarding the evaluation of the distance of the presented voice by a button press on an MR-compatible response box, without paying attention to the emotional tone, the direction or the location in space of the presented stimuli. The four buttons represented the following distances: Button 1 ('Very close', corresponding to 1 m); Button 2 ('Close', corresponding to 5 m); Button 3 ('Far', corresponding to 10 m); Button 4 ('Very far', corresponding to 20 m).

In an effort to keep the number of trials per condition at a reasonable level to improve the statistical power and, more importantly, to distinguish between proximal and distal auditory space, we merged correct trials for 'Very close' and 'Close' and for 'Far' and 'Very far' together, creating the 'Proximal' and 'Distal' conditions used in the analyses for each emotion category. This led us to the following independent (χ^2=1.07, P = 0.785) conditions: 'Aggressive proximal' (mean number of trials 46.29, s.d. 7.91), 'Aggressive distal' (mean number of trials 44.71, s.d. 8.33), 'Neutral proximal' (mean number of trials 41.14,

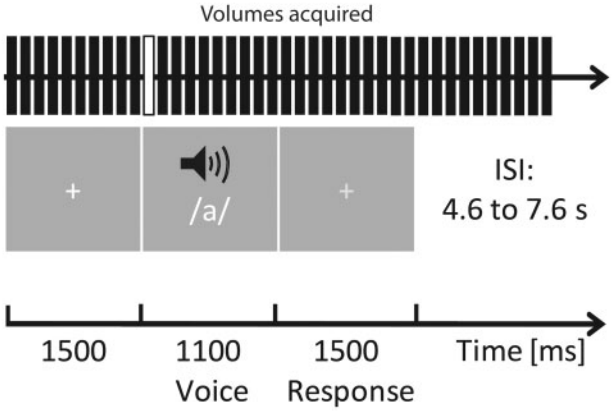

Fig. 1. Experimental design of the neuroimaging distance evaluation task. Participants were placed in a supine position in the scanner and instructed to focus on the distance of the auditorily presented prosody stimuli (/a/). They were instructed to focus on a white central fixation cross displayed via a rear-mounted projector and viewed through a head coil-mounted mirror. After the prosody offset, the white crosshair turned light gray, indicating that the participants had to respond. Auditory stimuli were presented through MR-compatible pneumatic headphones (MR confon GmbH, Germany). ISI represents inter-stimulus interval. The white bar of the 'volumes acquired' represents the onset used for the fMRI analyses.

s.d. 6.22) and 'Neutral distal' (mean number of trials 37.57, s.d. 12.80). The order of presentation was fully randomized and stimulus onset was jittered in steps of 1000 ms (2000 ± 1000 ms). The inter-stimulus interval ranged from 4.6 to 7.6 s (Figure 1).

Behavioral data analysis. Following normality estimation, data were analyzed by using a 2 (emotion) × 2 (distance) × 2 (space laterality) repeated measures analysis of variance (ANOVA) with Statistica 12 software (StatSoft Inc., Tulsa, OK, USA). Additional *t*-tests were used to follow up significant interaction effects between the factors. *Post hoc* correction for multiple comparisons was applied by using a Bonferroni correction. Pearson correlations were used to assess a relation between responses and reaction times and the threshold was set at $P < 0.05$.

Temporal voice area functional localizer task. Auditory stimuli consisted of sounds from a variety of sources. Vocal stimuli were obtained from 47 speakers: 7 babies, 12 adults, 23 children and 5 older adults. Stimuli included 20 blocks of vocal sounds and 20 blocks of non-vocal sounds. Vocal stimuli within a block could be either speech (33%: words, non-words, foreign language) or non-speech (67%: laughs, sighs, various onomatopoeia). Non-vocal stimuli consisted of natural sounds (14%: wind, streams), animals (29%: cries, gallops), human environment (37%: cars, telephones, airplanes) or musical instruments (20%: bells, harp, instrumental orchestra). The paradigm, design and stimuli were obtained through the Voice Neurocognition Laboratory website (http://vnl.psy.gla.ac.uk/resources.php). Stimuli were presented at an intensity that was kept constant throughout the experiment (70 dB sound-pressure level). Participants were instructed to actively listen to the sounds. The silent inter-block interval was 8 s long.

Image acquisition. Structural and functional brain imaging data were acquired by using a 3 T scanner (Siemens Trio, Erlangen, Germany) with a 32-channel coil. A magnetization prepared rapid acquisition gradient echo sequence was used to acquire high-resolution ($1 \times 1 \times 1\,\text{mm}^3$) T1-weighted structural images (TR = 1900 ms, TE = 2.27 ms, TI = 900 ms). Functional images were acquired by using a multislice echo planar imaging sequence (36 transversal slices in descending order, slice thickness 3.2 mm, TR = 2100 ms, TE = 30 ms, field of view = $205 \times 205\,\text{mm}^2$, 64×64 matrix, flip angle = 90°, bandwidth 1562 Hz/Px).

Image analysis. Functional images were analyzed with Statistical Parametric Mapping software (SPM12, Wellcome Trust Centre for Neuroimaging, London, UK). Preprocessing steps included realignment to the first volume of the time series, slice timing, normalization into the Montreal Neurological Institute (MNI) (Collins *et al.*, 1994) space using the DARTEL toolbox (Ashburner, 2007) and spatial smoothing with an isotropic Gaussian filter of 8 mm full width at half maximum. To remove low frequency components, we used a high-pass filter with a cutoff frequency of 128 s. Anatomical locations were defined with a standardized coordinate database (Talairach Client, http://www.talairach.org/client.html) by transforming MNI coordinates to match the Talairach space and transforming it back into MNI for display purposes.

For the fMRI distance evaluation task, we used a general linear model including hits only, where each correct trial (Aggressive proximal: 48.95%, s.d. 8.24; Aggressive distal: 46.87%, s.d. 8.67; Neutral proximal: 42.70%, s.d. 6.48; Neutral distal: 39.58%, s.d. 13.33; chance level was 25% according to the four possible response keys corresponding to the four to-be-evaluated distances) was modeled by using a stick function and was convolved with the hemodynamic response function. Events were time-locked to the onset of the voice stimuli. Separate regressors were created for each experimental condition and for the mean spectrum energy of each voice stimulus. Sound energy, a measure of sound pressure, was included as a parametric modulator on a trial-by-trial basis in our analyses. An additional regressor included errors and missed trials, as well as trials with reaction times outside the limits of a 98% confidence interval (these trials were also excluded from the behavioral data analyses). Finally, six motion parameters were included as regressors of no interest to account for movement in the data.

The condition regressors were used to compute linear contrasts for each participant and were then taken to a second-level, flexible factorial analysis. The second-level analysis was performed with a $2 \times 2 \times 2$ factorial design with the factors 'Emotion' (aggressive, neutral), 'Distance' (proximal, distal) and 'Space laterality' (left and right spaces). The emotion factor aimed at uncovering enhanced brain activity for aggressive relative to neutral trials. As we were mostly interested in brain activity between aggressive and neutral trials relative to their proximal *vs* distal space, the distance factor was also included in the design. Finally, to control for any bias related to spatial lateralization in the left/right auditory space, the space laterality factor was included in the design. The flexible factorial design assumed that participants were independent while conditions (Distance and Space laterality factor) were not. Finally, variance estimation was set to unequal for both participants and conditions.

For the temporal voice area localizer session, we used a general linear model in which each block was modeled by using a block function and was convolved with the hemodynamic response function, time-locked to the onset of each block. Separate regressors were created for each condition. Finally, six motion parameters were included as regressors of no interest to account for movement in the data. The condition regressors

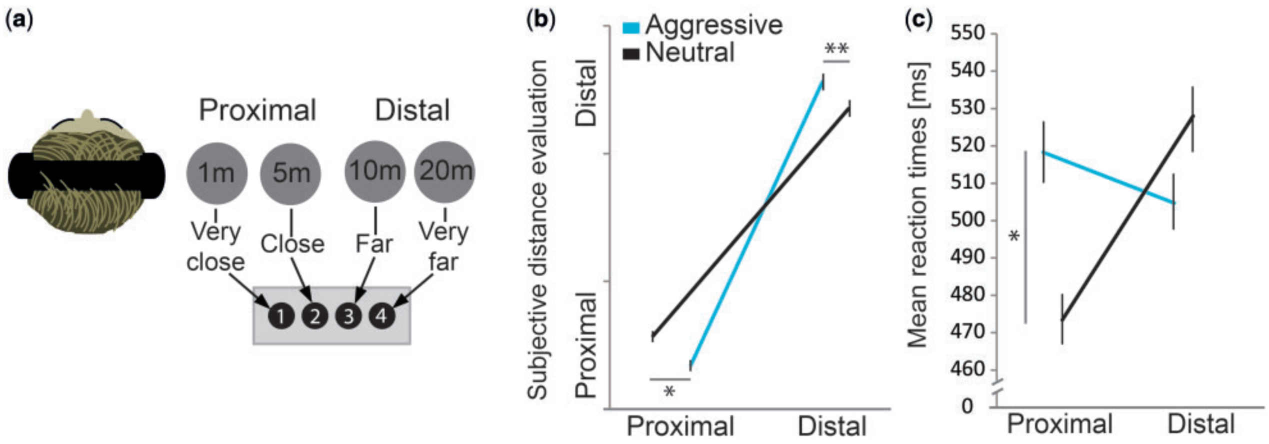

Fig. 2. Task design (see Materials and methods), mean perceived distance and reaction times across all participants (*N* = 14). (a) For each trial, the vowel /a/ spoken in an aggressive or neutral tone had to be evaluated as very close (1 m), close (5 m), far (10 m) or very far (20 m). Vocalizations were spatialized at these distances in the left or right space (right space shown only). (b) Interaction between emotion (aggressive, neutral) and perceived distance (proximal, distal) (*F*(1,13) = 13.09, *P* = 0.003), with more accurate evaluation for aggressive *vs* neutral voices both in the proximal (*t*(13) = −1.98, *P* = 0.048) and distal space (*t*(13)=3.80, *P*=0.002). (c) Reaction times for the voice distance evaluation, showing an interaction between emotion and perceived distance (*F*(1,13)=9.58, *P*=0.008), with slower reaction times for aggressive compared with neutral proximal voices (*t*(13)=2.80, *P*=0.015). No interaction with auditory space laterality was observed (all *P*s > 0.05, Supplementary Table S1). Error bars ± 1 SEM. **P* < 0.05; ***P* < 0.01. *Post hoc* correction for multiple comparisons was applied to *t*-statistics by using a Bonferroni correction.

were used to compute linear contrasts for each participant. These contrasts were then taken to a second-level analysis in which a one-sample *t*-test was used to contrast vocal against non-vocal stimuli.

All neuroimaging activations were thresholded in SPM12 using voxel-wise false discovery rate (FDR) correction at $P < 0.05$. To remove single voxels or very small clusters, an arbitrary cluster extent of $k > 30$ voxels was used. All percentage of signal change analyses (for each peak separately) were performed by using repeated-measure ANOVAs, and *post hoc* correction for multiple comparisons was applied by using a Bonferroni correction following normality estimation.

Behavioral distance evaluation experiment

Participants

Seventeen right-handed, healthy, native or highly proficient French-speaking participants (eight male, nine female, mean age 21.41 years, s.d. 3.76) were included in an independent behavioral experiment conducted at the University of Geneva in a dedicated behavioral testing room. All participants were naïve to the experimental design and study, and they had normal or corrected-to-normal vision, normal hearing and no history of psychiatric or neurologic incidents. Participants gave written informed consent for their participation in accordance with ethical and data security guidelines of the University of Geneva. The study was approved by the Ethical Committee of the University of Geneva and conducted according to the Declaration of Helsinki.

Experimental design

Stimuli were identical to those used in the fMRI distance evaluation task. The behavioral distance evaluation experiment consisted of one session of ∼20 min during which the spatialized /a/ stimuli were presented binaurally, each at a time, through headphones (Sennheiser electronic GmbH & Co., KG, Germany). All stimuli were presented with the same intensity throughout the experiment, namely at a 70 dB sound-pressure level. Participants had to respond as quickly and accurately as possible and evaluate the distance of the presented voice by a button press on a response box, without paying attention to the emotional tone, direction or location in space of the presented stimuli. The button mapping and conditions were identical to those in the fMRI distance evaluation experiment.

Data analysis

Following normality estimation, data were analyzed by using a 2 (emotion) × 2 (distance) × 2 (space laterality) repeated measures ANOVA with Statistica 12 software (StatSoft Inc., Tulsa, OK, USA). Additional *t*-tests were used to follow up significant interaction effects within/between the factors. *Post hoc* correction for multiple comparisons was applied by using a Bonferroni correction. Pearson correlations were used to assess a relation between responses and reaction times, and the threshold was set at $P < 0.05$.

Results

Behavioral data of the fMRI experiment

A 2 × 2 × 2 repeated measures ANOVA was performed on distance evaluation (accuracy) data with factors emotion (aggressive, neutral), distance (proximal, distal) and space laterality (left, right auditory space). While the main effect of emotion ($F(1,13) = 7.80$, $P = 0.015$), distance ($F(1,13)=2464.9$, $P = 0.000001$) and their interaction ($F(1,13)=13.09$, $P = 0.003$) were significant with more accurate evaluation for aggressive *vs* neutral voices both in the proximal ($t(13) = -1.98$, $P = 0.048$) and distal space ($t(13)=3.80$, $P = 0.002$), the triple interaction between emotion, distance and space laterality was not ($F(1,13)=0.332$, $P = 0.574$) (Figure 2B and Supplementary Table S1).

Due to the relatively small sample size ($N = 14$) of our experimental fMRI group, we decided to conduct the same study in an independent behavioral experiment including 17 participants, age-matched to the fMRI group. The analyses were performed using a similar 2 × 2 × 2 repeated measures ANOVA. The main result of the fMRI group study was the significant interaction between emotion and distance, and we found this significant interaction in our behavioral experiment group as well

Table 1. Mean cluster location and local maxima of BOLD signal change for aggressive compared with neutral voices in the fMRI distance evaluation task ($P < 0.05$, FDR corrected)

			MNI			
Region name (Brodmann area)	Left/Right	Z score	x	y	z	Size (voxels)
Insula (13)	L	7.18	−36	−34	18	1782
Inferior parietal lobule (40)	L	5.53	−50	−36	24	
Superior temporal gyrus (22)	L	4.09	−48	6	−8	
Superior temporal gyrus (41)	R	5.80	42	−30	14	1067
Superior temporal gyrus (42)	R	5.69	66	−30	16	
Superior temporal gyrus (22)	R	4.67	68	−34	10	
Inferior temporal gyrus (20)	R	4.20	44	−6	−26	41
Thalamus	L	3.89	−6	−14	2	37
Thalamus	R	3.80	10	−28	−4	48
Pulvinar	L	3.67	−6	−30	−2	75
Amygdala	R	3.54	24	−4	−20	30

(F(1,16)=11.48, $P = 0.003$). Again, distance evaluation was more accurate for aggressive compared with neutral proximal (t(16) = −1.99, $P = 0.045$) and for aggressive compared with neutral distal voices (t(16) = 4.66, $P = 0.0003$), without any interaction between emotion, distance and space laterality (F(1,16)=1.39, $P = 0.254$) (Supplementary Figure S3a and Table S1).

Reaction times for the fMRI, voice distance evaluation task were also analyzed using a 2 × 2 × 2 repeated measures ANOVA with factors emotion × distance × space laterality. Emotion and distance did not reveal any significant main effect (F(1,13)=1.10, $P = 0.312$ and F(1,13)=1.85, $P = 0.196$, respectively) but the interaction between these factors was significant (F(1,13)=9.58, $P = 0.008$). In fact, slower reaction times for aggressive compared with neutral proximal voices (t(13)=2.80, $P = 0.015$) were found to drive this interaction between emotion and distance (Figure 2C). No triple interaction with auditory space laterality was observed (F(1,13)=3.97, $P = 0.078$; Supplementary Table S1).

Regarding the independent behavioral experiment group, reaction times results were again similar to the fMRI group, with no main effect of emotion (F(1,16)=0.016, $P = 0.899$) or distance (F(1,16)=0.527, $P = 0.478$), but a significant interaction between these factors (F(1,16)=7.06, $P = 0.017$). This interaction showed faster reaction times for aggressive distal than for neutral distal voices (t(16) = −2.34, $P = 0.033$) (Supplementary Figure S3b). Again, no triple interaction with auditory space laterality was observed (F(1,16)=0.32, $P = 0.860$; Supplementary Table S1).

We finally computed Pearson correlations for both groups separately between accuracy and reaction times, but found no significant relation between the two measures (all $Ps > 0.05$).

Functional data

Regarding blood-oxygen-level dependent (BOLD) measures and related neural activity, we contrasted correctly evaluated aggressive as opposed to neutral voices, across all distances/space laterality. This contrast yielded to enhanced BOLD signal in large portions of the STG, bilaterally as well as in subcortical regions such as the thalamus and the amygdala (Figure 1A and C and Table 1 and Supplementary Figure S4). While these brain areas exhibited enhanced activity for the emotion factor, only some of them showed a significantly enhanced response to proximal as compared with distal voices (distance factor, Figure 3D and Table 2), such as the left insula (MNI *xyz* −36 −34 18; proximal > distal: t(13)=6.37, $P = 0.00002$) and inferior parietal lobule (IPL) (MNI *xyz* −50 −30 24; proximal > distal: t(13)=2.50, $P = 0.026$) and several subregions of the superior part of the right temporal gyrus (Proximal > distal: MNI *xyz* 42 −30 14, t(13)=6.08, $P = 0.00004$; MNI *xyz* 66 −30 16, t(13)=3.90, $P = 0.0018$; MNI *xyz* 68 −34 10, t(13)=2.88, $P = 0.013$). Using the present paradigm, we were specifically interested in brain regions showing an interaction between emotion and distance. By quantifying the BOLD signal change in the peaks of brain areas found in the aggressive > neutral contrast, we found an interaction effect for aggressive *vs* neutral and proximal *vs* distal voices only in the right mid-STG (MNI *xyz* 66 −30 16; F(1,13) = 6.134, $P = 0.036$) (Figure 3A and Table 3). More specifically, signal extraction in the right mid-STG showed stronger activation for aggressive proximal than aggressive distal voices (t(13) = 3.40, $P = 0.005$) and the same effect was observed for neutral voices (t(13) = 3.25, $P = 0.006$). As mentioned, the full interaction was significant in this region (F(1,13) = 6.134, $P = 0.036$), and it also showed stronger activation for aggressive proximal compared with neutral proximal (t(13) = 4.86, $P = 0.0003$) and for aggressive distal compared with neutral distal voices (t(13) = 3.28, $P = 0.006$). It should be noted here that these BOLD increases are independent of the energy differences of auditory signals across our experimental manipulations as we used these measures as regressors of non-interest in our design.

Interestingly, this right mid-STG region [MNI *xyz* 66 –30 16] was located in the voice-sensitive auditory cortex (Figure 3A, Supplementary Figure S5 and Table S2). This result further extends the role attributed so far to voice-sensitive areas and sheds new light on a subpart of the right mid temporal voice-sensitive areas as a crucial hub for processing proximal relative to distal aggressive voices.

Discussion

The study of threat distance is relatively underrepresented in the human auditory literature, even though the underlying biological and psychological mechanisms of this system directly impact on our species' survival. The aim of this study was hence to shed new light on the underpinnings of vocal threat perception in a spatial hearing paradigm involving a fine-tuned convolution to create virtually spatialized stimuli. We found a crucial involvement of a subpart of the right mid voice-sensitive cortex in accurately perceiving proximal as opposed to distal vocal threat. This neural mechanism underlied a distance evaluation advantage for aggressive over neutral vocal signals, further strengthening the abovementioned argument according to which vocal threat is a fundamental vector of biological relevance.

In the present paradigm, the evaluation of voice distance highlighted a higher accuracy for aggressive as compared with

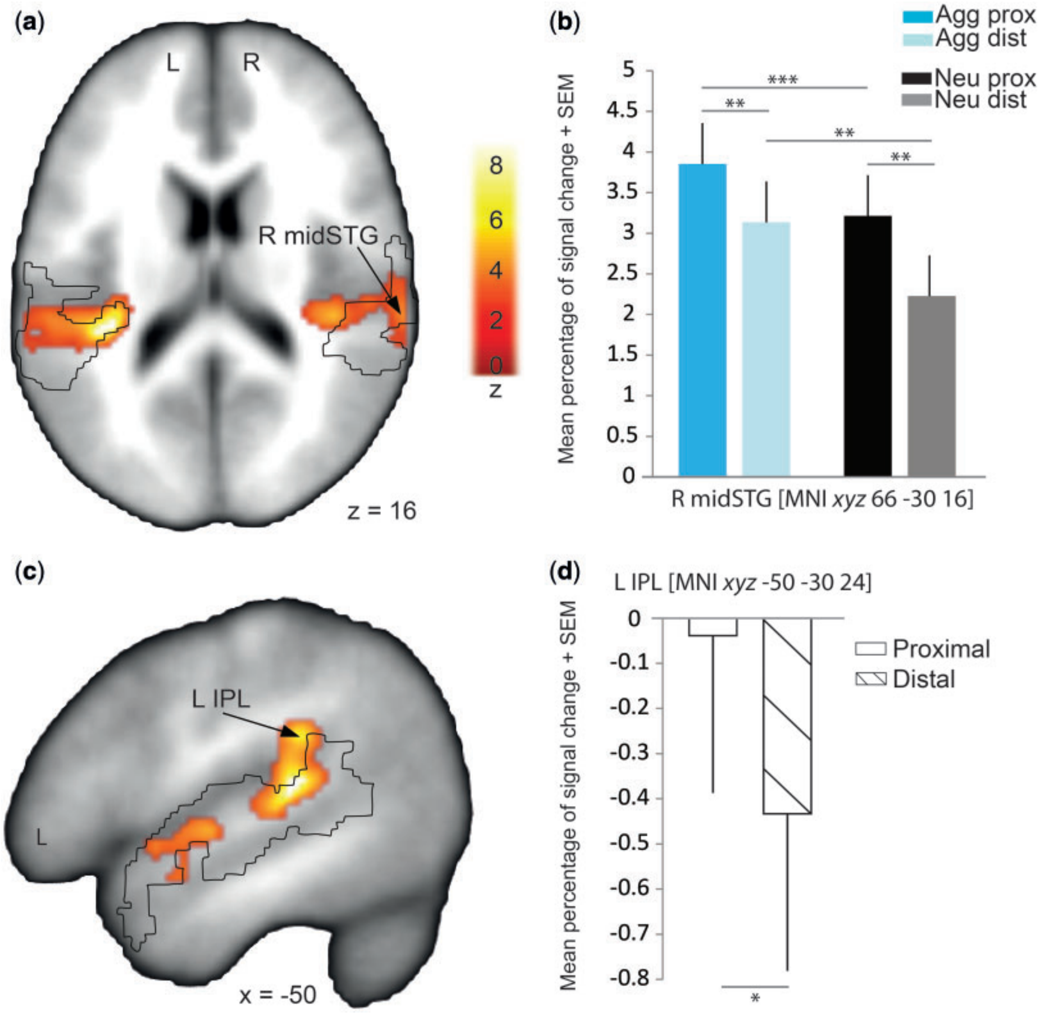

Fig. 3. Neural effects of aggressive relative to neutral voices for correctly evaluated distance. (a and c) Stronger activations for aggressive compared with neutral voices across all distances. The black outline represents voice-sensitive areas (Supplementary Figure S5 and Table S2). (b) Signal extraction in the right mid-STG (R mid-STG) [MNI *xyz*: 66 −30 16], showing stronger activations for proximal compared with distal (Aggressive voices: $t(13) = 3.40$, $P = 0.005$; Neutral voices: $t(13) = 3.25$, $P = 0.006$) and for aggressive proximal (Agg prox) and distal (Agg dist) compared with neutral proximal (Neu prox) and distal (Neu dist) voices (Agg prox >Neu prox: $t(13)$=4.86, P=0.0003; Agg dist > Neu dist: $t(13)$=3.28, P=0.006). (d) Signal in the left IPL (L IPL) [MNI *xyz*: −50 −30 24], showing a stronger activation decrease for distal than for proximal voices. No region showed an interaction with auditory space laterality (all Ps > 0.05, Supplementary Table S1). Activations are thresholded at $P < 0.05$ (FDR corrected, voxel-wise) with $k > 30$ voxels. Error bars ± 1 SEM. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$. *Post hoc* correction for multiple comparisons was applied to *t*-statistics by using a Bonferroni correction.

Table 2. Significant effect of distance (proximal, distal) in the regions obtained by contrasting aggressive with neutral voices in the fMRI distance evaluation task ($P < 0.05$, FDR corrected).

	MNI			
Region name (Brodmann area)	x	y	z	Statistical effect of distance
Insula (13)	−36	−34	18	$F(1,13) = 40.63$, $P = 0.00001$
Inferior parietal lobule (40)	−50	−36	24	$F(1,13) = 6.25$, $P = 0.026$
Superior temporal gyrus (42)	42	−30	14	$F(1,13) = 32.93$, $P = 0.00004$
Superior temporal gyrus (22)	66	−30	16	$F(1,13) = 15.19$, $P = 0.001$
Superior temporal gyrus (22)	68	−34	10	$F(1,13) = 8.278$, $P = 0.013$

neutral stimuli and this result was found in both the fMRI group and the behavioral control experiment group. Reaction time data emphasized an attentional capture effect only in the proximal space showing a difficulty to disengage from aggressive voices, while facilitation was observed in the distal space. This result is in line with observations according to which aggressive/angry vocal signals take more time to be processed with regard to the cognitive appraisal model of emotion (Scherer, 1999) but shows that it can be more specific to proximal than distal space, reflecting the imminence of danger when the event is near the perceiver. Again, it also emphasizes the impact of threat in the perception of vocal signals. These

Table 3. Interaction between emotion (aggressive, neutral) and distance (proximal, distal) factors in the regions obtained by contrasting aggressive with neutral voices in the fMRI distance evaluation task ($P < 0.05$, FDR corrected)

	MNI			
Region name (Brodmann area)	x	y	z	Statistical interaction (emotion × distance)
Insula (13)	−36	−34	18	$F(1,13) = 0.039$, $P = 0.847$
Inferior parietal lobule (40)	−50	−36	24	$F(1,13) = 0.396$, $P = 0.540$
Superior temporal gyrus (22)	−48	6	−8	$F(1,13) = 1.600$, $P = 0.228$
Superior temporal gyrus (41)	42	−30	14	$F(1,13) = 0.129$, $P = 0.725$
Superior temporal gyrus (42)	66	−30	16	$F(1,13) = 6.134$, $P = 0.036$
Superior temporal gyrus (22)	68	−34	10	$F(1,13) = 0.013$, $P = 0.908$
Inferior temporal gyrus (20)	44	−6	−26	$F(1,13) = 1.513$, $P = 0.241$
Thalamus	−6	−14	2	$F(1,13) = 1.424$, $P = 0.254$
Thalamus	10	−28	−4	$F(1,13) = 0.571$, $P = 0.463$
Pulvinar	−6	−30	−2	$F(1,13) = 0.103$, $P = 0.753$
Amygdala	24	−4	−20	$F(1,13) = 0.434$, $P = 0.522$

results are also coherent with, and at the same time interestingly differ from behavioral studies on distance perception. In fact, while proximal stimuli were shown to be evaluated more accurately than distal stimuli (Little *et al.*, 1992), our results emphasize the accurate evaluation of distance for aggressive voices in both proximal/distal spaces. One interpretation would be that voices have a more complex structure than the tones used by Little *et al.* (1992), and their intrinsic value also involves a higher biological relevance notably for aggressive voices. This biological relevance aspect could explain the accurate perception of vocal threat in the auditory space as compared with neutral voices or tones. However, pitch, frequency and loudness were shown to greatly impact on the perceived urgency of a sound and one cannot exclude that low level acoustic cues would partly account for the observed advantage of aggressive over neutral stimuli notably due to their acoustical differences (Haas and Edworthy, 1996). Such effects can thus be partly confounded in our results, even though loudness was constant across participants and conditions, and intensity was also normalized for our voice stimuli. One should also take into account the setup used in this study. In fact, participants had to explicitly evaluate the voice's perceptual distance; hence it is not impossible that our results would significantly change in a more natural context where no auditory evaluation is expected. Moreover, in a situation where someone is whispering while very near or shouting while very far away from the listener, other cues and other mechanisms would allow us to distinguish between distance and salience (Philbeck and Mershon, 2002), as intensity evaluation alone would not be sufficient in order to accurately localize these events. Our results and paradigm cannot exhaustively account for these crucial aspects and more studies need to be conducted in order to disentangle this matter, notably by the use of carefully selected auditory vocal/non-vocal and unvoiced stimuli with varying distance and intensity.

Another interesting body of literature investigated the effect of size-changing stimuli, such as looming in the visual domain and motion in the auditory domain. While looming refers to a fast increase in the size of a visual stimulus on the retina, auditory motion is created by the appearance of approaching *vs* receding sounds. Both looming and auditory motion (more specifically in the case of approaching stimuli) were shown to capture the perceiver's attention, a result interpreted as beneficial for the processing of urgent, approaching behavioral events (Haas and Edworthy, 1996; Gabbiani *et al.*, 2002; Franconeri and Simons, 2003; Tajadura-Jiménez *et al.*, 2010). These results further highlighted the sense of warning implicitly conveyed by looming sounds (Bach *et al.*, 2009) and such intensity-varying stimuli were shown to enhance activity in temporal brain regions such as the superior temporal sulcus and middle temporal gyri (Seifritz *et al.*, 2002). In the light of these previous results, the biological relevance of our aggressive voice stimuli would be additive to the inherent salience of closer or approaching auditory stimuli for correct distance evaluation. In-line with this interpretation, a mechanism of overestimation of intensity change, namely approaching sounds estimated as closer than their actual distance, was also observed and proposed as a mechanism underlying the expectancy of an event coming from a spatial source, hence leading to an advantage in detecting and reacting to such events (Neuhoff, 2010). The use of approaching *vs* receding voices in addition to static voice stimuli could elegantly address this point in order to distinguish between distance and motion in vocal threat perception. In addition, moving voices would more accurately represent our everyday vocal experience as generally people are not static when speaking. Finally, visual input was also shown to bias auditory motion and this manipulation could also be interesting to uncover an influence of multimodal stimulation on space perception in the context of voice processing (Kitagawa and Ichihara, 2002).

Regarding neuroimaging data, we should consider statistical thresholding as a major filter of our findings, because we used multiple-comparison correction with FDR at 0.05. In fact, using this threshold is less conservative than family wise error and we cannot exclude that 5% of activated voxels are actually false-positives, while using a cluster size of $k > 30$ should nevertheless slightly decrease this percentage value. That being said, our results emphasize the involvement of the bilateral superior temporal and lateral parietal cortex in processing virtually spatialized aggressive as opposed to neutral voices. Abovementioned regions are part of the planum temporale, a broad region repeatedly shown to play a fundamental role in all types of acoustical processing including tone and voice perception, in standard as well as in more specific spatial hearing paradigms (Griffiths and Warren, 2002). The superior temporal regions we observed in our study also overlap with the temporal voice areas (Belin *et al.*, 2000), the emotional voice areas (Ethofer *et al.*, 2011) and with temporal regions specifically responding to angry as opposed to neutral voices (Grandjean *et al.*, 2005). Amygdala activity is also in-line with literature on emotion processing using both visual (Vuilleumier *et al.*, 2004) and auditory, vocal stimuli (Sander *et al.*, 2003). The spatial nature of our convolved voice stimuli did also recruit the IPL, a region that can be seen as the core area underlying spatial hearing (Griffiths and Warren, 2002; Lewald *et al.,* 2002; At *et al.*, 2011) and auditory attention in humans (Shomstein and Yantis, 2006). The left insula and IPL as well as several right STG regions we obtained showed a specific effect of distance. Such result is coherent with the role of the temporal and critically, the lateral parietal cortex in spatial hearing. In fact, the posterior/lateral parietal cortex was shown to underlie spatial location processing in primates (Bremmer *et al.*, 2001) and humans (Andersen *et al.*, 1985), in addition to coding intention (Andersen, 1995; Snyder *et al.*, 1997; Andersen and Buneo, 2002) and attentional processes such as spatial orienting (Corbetta *et al.*, 2000; Buschman and Miller, 2007) and auditory spatial attention (Kong *et al.*, 2014). However, the result concerning the STG regions improves our understanding of the role of subregions of the voice-sensitive cortex in processing proximal as opposed to distal voices. More specifically, it highlights the role of these auditory regions in the perception and processing of proximal/distal voices and extends the ability of temporal voice areas to process spatial vocal events. This result also interestingly adds up to the neuropsychological literature on auditory space processing in which it was shown that different mechanisms and underlying brain areas were recruited in the 'where' and 'what' auditory pathways (Clarke *et al.*, 2002; Thiran and Clarke, 2003). The present results somehow lack an involvement of the inferior frontal cortex (IFC) that was shown to process emotional voices as well as spatial aspects of voices. In fact, while the IFG was shown to respond to emotional vocalizations such as emotional prosody (Fruhholz and Grandjean, 2013b), its involvement in processing spatial information was more specifically related to spatial (Courtney *et al.*, 1998) and non-spatial memory in the visual (Bushara *et al.*, 1999) and auditory domain (Alain *et al.*, 2008). Future work on the role of the IFG in coding auditory spatial events have yet to be performed together with an investigation of the link between spatial processing and spatial memory, because these processes were mainly studied separately in the literature. The lack of a memory-related or of a higher level of vocal processing in the present task may also account for this absence of an involvement of the IFC in our results.

Most importantly, we observed a clear dissociation between abovementioned brain regions and the right mid-STG. This finding is of high importance since to the best of our knowledge, we show for the first time that the right mid-STG is able to respond specifically to the concurrent emotional tone and proximity of vocal signals, literally exhibiting enhanced activity for proximal aggressive voices as opposed to distal or neutral voices. While the right STG was already shown to be involved in spatial hearing by the use of transcranial magnetic stimulation (Lewald *et al.*, 2004) and in processing angry voices even when not in the focus of attention (Grandjean *et al.*, 2005), its involvement in emotional, vocal spatial hearing was never reported before. As mentioned earlier, this result could be interpreted in terms of low level auditory cues such as pitch, frequency or intensity triggering urgency (Haas and Edworthy, 1996) and activity in the right temporal cortex, as this region extends to the primary auditory cortex that is highly involved in low level acoustic cues processing (Kaas and Hackett, 2000; Schirmer and Kotz, 2006). However, knowing that across stimuli intensity was normalized, loudness constant (70 dB) and intensity-related energy used as a regressor of non-interest in our design, we think it is fair to interpret this result as a dynamic neural processing resulting from higher level, emotion- and distance-related perceptual mechanisms. In fact, the spatial convolution of our voice stimuli uses a fine-tuned algorithm impacting mostly on interaural time difference, a procedure that mimics the statistics and physics of real-life auditory events (Wightman and Kistler, 1992) and we thus assume that low level cues such as intensity have less impact on our results than emotion or distance *per se*. Finally, we should take into account that auditory threat is not expressed exclusively by aggressive/angry voices and the impact of fearful vocal expressions should be studied in a detailed spatial paradigm as well. A further comparison of approach and escape behaviors in the proximal and distal space could also be investigated in this context, potentially leading to a more general understanding of the perceptual influence vocal threat can exhibit on behavior.

Conclusion

Taken together, the present results emphasize the involvement of a network of brain regions comprising temporal, parietal and subcortical brain regions as an underlying cerebral mechanism in processing, perceiving and correctly evaluating proximal and distal vocal signals. Our results further highlight the crucial role of the right mid-STG in processing and potentially reacting to imminent danger signaled by proximal vocal threat. Finally, this study emphasizes the biological relevance of aggressive voices for humans, especially when these vocal signals of threat are relatively close to the perceiver and can thus negatively impact on the species' survival.

Supplementary data

Supplementary data are available at *SCAN* online.

Acknowledgements

We would like to acknowledge Andy Christen and Nicolas Burra for their valuable insights into this work and Eva Pool for helping improve the manuscript. The research was funded by the Swiss National Science Foundation (SNSF project no 105314_124572/1), the National Center of Competence in Research in Affective Sciences of the University of Geneva (51NF40-104897—D.G.), Switzerland.

Conflict of interest: None declared.

References

Alain, C., He, Y. Grady, C. (2008). The contribution of the inferior parietal lobe to auditory spatial working memory. *Journal of Cognitive Neuroscience*, **20**(2), 285–95.

Andersen, R.A. (1995). Encoding of intention and spatial location in the posterior parietal cortex. *Cerebral Cortex*, **5**(5), 457–69.

Andersen, R.A., Buneo, C.A. (2002). Intentional maps in posterior parietal cortex. *Annual Review of Neuroscience*, **25**(1), 189–220.

Andersen, R.A., Essick, G.K., Siegel, R.M. (1985). Encoding of spatial location by posterior parietal neurons. *Science*, **230**(4724), 456–8.

Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. *Neuroimage*, **38** (1), 95–113.

At, A., Spierer, L., Clarke, S. (2011). The role of the right parietal cortex in sound localization: a chronometric single pulse transcranial magnetic stimulation study. *Neuropsychologia*, **49**(9), 2794–7.

Bach, D.R., Neuhoff, J.G., Perrig, W., Seifritz, E. (2009). Looming sounds as warning signals: the function of motion cues. *International Journal of Psychophysiology*, **74**(1), 28–33.

Bänziger, T., Scherer, K.R. (2007). Affective computing and intelligent interaction. In: Paiva, A.C.R., Prada R., Picard R. W., editors *Using Actor Portrayals to Systematically Study Multimodal Emotion Expression: The Gemep Corpus*. Springer, Springer Berlin Heidelberg, 476–87.

Belin, P., Zatorre, R.J., Lafaille, P., Ahad, P., Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, **403**(6767), 309–12.

Bremmer, F., Schlack, A., Duhamel, J.R., Graf, W., Fink, G.R. (2001). Space coding in primate posterior parietal cortex. *Neuroimage*, **14**(1), S46–51.

Buchanan, T.W., Lutz, K., Mirzazade, S., *et al.* (2000). Recognition of emotional prosody and verbal components of spoken language: an fMRI study. *Cognitive Brain Research*, **9**(3), 227–38.

Buschman, T.J., Miller, E.K. (2007). Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. *Science*, **315**(5820), 1860–2.

Bushara, K.O., Weeks, R.A., Ishii, K., *et al.* (1999). Modality-specific frontal and parietal areas for auditory and visual spatial localization in humans. *Nature Neuroscience*, **2**(8), 759–66.

Clarke, S., Thiran, A.B. (2004). Auditory neglect: what and where in auditory space. *Cortex*, **40**(2), 291–300.

Clarke, S., Thiran, A.B., Maeder, P., *et al.* (2002). What and where in human audition: selective deficits following focal hemispheric lesions. *Experimental Brain Research*, **147**(1), 8–15.

Collins, D.L., Neelin, P., Peters, T.M., Evans, A.C., (1994). Automatic 3D intersubject registration of MR volumetric data in standardized talairach space. *Journal of Computer Assisted Tomography,* **18**(2), 192–205.

Corbetta, M., Kincade, J.M., Ollinger, JM., McAvoy, M.P., Shulman, G.L. (2000). Voluntary orienting is dissociated from target detection in human posterior parietal cortex. *Nature Neuroscience*, **3**(3), 292–7.

Courtney, S.M., Petit, L., Maisog, J.M., Ungerleider, L.G., Haxby, J.V. (1998). An area specialized for spatial working memory in human frontal cortex. *Science*, **279**(5355), 1347–51.

Ethofer, T., Bretscher, J., Gschwind, M., Kreifelts, B., Wildgruber, D., Vuilleumier, P. (2011). Emotional voice areas: anatomic location, functional properties, and structural connections revealed by combined fMRI/DTI. *Cerebral Cortex*, **22**(1):191–200.

Franconeri, S.L., Simons, D.J. (2003). Moving and looming stimuli capture attention. *Perception Psychophysics*. **65**(7), 999–1010.

Frühholz, S., Ceravolo, L., Grandjean, D. (2012). Specific brain networks during explicit and implicit decoding of emotional prosody. *Cerebral Cortex*, **22**(5):1107–17.

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

Fruhholz, S., Grandjean, D. (2013b). Processing of emotional vocalizations in bilateral inferior frontal cortex. *Neuroscience and Biobehavioral Reviews*, **37**(10 Pt 2), 2847–55.

Gabbiani, F., Krapp, H.G., Koch, C., Laurent, G. (2002). Multiplicative computation in a visual neuron sensitive to looming. *Nature*, **420**(6913), 320–4.

Grandjean, D., Sander, D., Pourtois, G., *et al.* (2005). The voices of wrath: brain responses to angry prosody in meaningless speech. *Nature Neuroscience*, **8**(2), 145–6.

Griffiths, T.D., Warren, J.D. (2002). The planum temporale as a computational hub. *Trends in Neuroscience*, **25**(7), 348–53.

Haas, E.C., Edworthy, J. (1996). Designing urgency into auditory warnings using pitch, speed and loudness. *Computuing and Control Engineering Journal*, **7**(4), 193–8.

Kaas, J.H., Hackett, T.A. (2000). Subdivisions of auditory cortex and processing streams in primates. *Proceedings of the National Academy of Science of the United States of America*. **97**(22), 11793–9.

Kitagawa, N., Ichihara, S. (2002). Hearing visual motion in depth. *Nature*, **416**(6877), 172–4.

Kong, L., Michalka, S.W., Rosen, M.L., *et al.* (2014). Auditory spatial attention representations in the human cerebral cortex. *Cerebral Cortex*, **24**(3), 773–84.

Leitman, D.I., Wolf, D.H., Ragland, J.D., *et al.* (2010). “It’s not what you say, but how you say it": a reciprocal temporo-frontal network for affective prosody. *Frontiers in Human Neuroscience*, **4**, 19.

Lewald, J., Foltys, H., Töpper, R. (2002). Role of the posterior parietal cortex in spatial hearing. *Journal of Neuroscience*, **22**(3), RC207.

Lewald, J., Meister, I.G., Weidemann, J., Topper, R. (2004). Involvement of the superior temporal cortex and the occipital cortex in spatial hearing: evidence from repetitive transcranial magnetic stimulation. *Journal of Cognitive Neuroscience*, **16**(5), 828–38.

Little, AD., Mershon, D.H., Cox, P.H. (1992). Spectral content as a cue to perceived auditory distance. *Perception*, **21**(3), 405–16.

McNally, G.P., Westbrook, R.F. (2006). Predicting danger: the nature, consequences, and neural mechanisms of predictive fear learning. *Learning and Memory*, **13**(3), 245–53.

Mobbs, D., Petrovic, P., Marchant, J.L., *et al* (2007). When fear is near: threat imminence elicits prefrontal-periaqueductal gray shifts in humans. *Science*, **317**(5841), 1079–83.

Neuhoff, J.G. (2001). An adaptive bias in the perception of looming auditory motion. *Ecological Psychology*, **13**(2), 87–110.

Öhman, A. (1986). Face the beast and fear the face: animal and social fears as prototypes for evolutionary analyses of emotion. *Psychophysiology*, **23**(2), 123–45.

Philbeck, J.W., Mershon, D.H. (2002). Knowledge about typical source output influences perceived auditory distance. *Journal of the Acoustical Society of America*, **111**(5), 1980–3.

Sander, D., Grafman, J., Zalla, T. (2003). The human amygdala: an evolved system for relevance detection. *Reviews in the Neuroscience*, **14**(4), 303–16.

Scherer, K.R. (1999). Appraisal theory. In T. D. M. J. Power, editor. *Handbook of Cognition and Emotion*. New York: John Wiley & Sons Ltd, 637–63.

Schirmer, A., Kotz, S.A. (2006). Beyond the right hemisphere: brain mechanisms mediating vocal emotional processing. *Trends in cognitive science*, **10**(1), 24–30.

Seifritz, E., Neuhoff, J.G., Bilecen, D., *et al.* (2002). Neural processing of auditory looming in the human brain. *Current Biology*, **12**(24), 2147–51.

Shomstein, S., Yantis, S. (2006). Parietal cortex mediates voluntary control of spatial and nonspatial auditory attention. *Journal of Neuroscience*, **26**(2), 435–9.

Snyder. L., Batista, A., Andersen, R. (1997). Coding of intention in the posterior parietal cortex. *Nature*, **386**(6621), 167–70.

Tajadura-Jiménez, A., Väljamäe, A., Asutay, E., Västfjäll, D. (2010). Embodied auditory perception: the emotional impact of approaching and receding sound sources. *Emotion*, **10**(2), 216.

Thiran, A.B., Clarke, S. (2003). Preserved use of spatial cues for sound segregation in a case of spatial deafness. *Neuropsychologia*, **41**(9), 1254–61.

Vuilleumier, P., Richardson, M.P., Armony, J.L., Driver, J., Dolan, R.J. (2004). Distant influences of amygdala lesion on visual cortical activation during emotional face processing. *Nature Neuroscience*, **7**(11), 1271–8.

Vuilleumier, P., Schwartz, S. (2001). Emotional facial expressions capture attention. *Neurology*, **56**(2), 153–8.

Wambacq, I.J., Shea-Miller, K.J., Abubakr, A. (2004). Non-voluntary and voluntary processing of emotional prosody: an event-related potentials study. *Neuroreport*, **15**(3), 555–9.

Wightman, F.L., Kistler, D.J. (1992). The dominant role of low-frequency interaural time differences in sound localization. *Journal of the Acoustical Society of America*, **91**(3), 1648–61.

Witteman, J., Van Heuven, V.J., Schiller, N.O. (2012). Hearing feelings: a quantitative meta-analysis on the neuroimaging literature of emotional prosody perception. *Neuropsychologia*, **50**(12), 2752–3.