Cognition and Emotion

ISSN: 0269-9931 (Print) 1464-0600 (Online) Journal homepage: https://www.tandfonline.com/loi/pcem20

Spatial location and emotion modulate voice perception

Ana P. Pinheiro, Diogo Lima, Pedro B. Albuquerque, Andrey Anikin & César F. Lima

COGNITION AND EMOTION
2019, VOL. 33, NO. 8, 1577–1586
https://doi.org/10.1080/02699931.2019.1586647

Spatial location and emotion modulate voice perception

Ana P. Pinheiro [a], Diogo Lima [b], Pedro B. Albuquerque [b], Andrey Anikin [c] and César F. Lima [d]

[a]Faculdade de Psicologia, Universidade de Lisboa, Lisboa, Portugal; [b]School of Psychology, University of Minho, Braga, Portugal; [c]Division of Cognitive Science, Department of Philosophy, Lund University, Lund, Sweden; [d]Instituto Universitário de Lisboa (ISCTE-IUL), Lisboa, Portugal

ABSTRACT
How do we perceive voices coming from different spatial locations, and how is this affected by emotion? The current study probed the interplay between space and emotion during voice perception. Thirty participants listened to nonverbal vocalizations coming from different locations around the head (left *vs.* right; front *vs.* back), and differing in valence (neutral, positive [amusement] or negative [anger]). They were instructed to identify the location of the vocalizations (Experiment 1) and to evaluate their emotional qualities (Experiment 2). Emotion-space interactions were observed, but only in Experiment 1: emotional vocalizations were better localised than neutral ones when they were presented from the back and the right side. In Experiment 2, emotion recognition accuracy was increased for positive *vs.* negative and neutral vocalizations, and perceived arousal was increased for emotional *vs.* neutral vocalizations, but this was independent of spatial location. These findings indicate that emotional salience affects how we perceive the spatial location of voices. They additionally suggest that the interaction between spatial ("where") and emotional ("what") properties of the voice differs as a function of task.

ARTICLE HISTORY
Received 18 September 2018
Revised 11 February 2019
Accepted 12 February 2019

KEYWORDS
Emotion; space; voice; perception; accuracy

The human voice is plausibly the most important sound category in our social landscape. Confirming the special status of voices in the brain, there are regions of the temporal cortex that selectively respond to voices compared to non-vocal sounds (e.g. environmental sounds), located bilaterally along the mid and anterior parts of the superior temporal gyrus and sulcus (Belin, Zatorre, Lafaille, Ahad, & Pike, 2000). In addition to being the carrier of speech, voices convey socially relevant information about the speaker, namely his/her emotional state. Imagine a family reunion where people are interacting with each other. If, suddenly, somebody starts shouting from another room in the house, the tone of voice may signal a potential threat that can be promptly inferred even if the speaker's face and gestures are not visible. This illustrates the relevance of a rapid and accurate recognition of vocal emotional expressions that allows us to make inferences about the intentions of others, shaping behaviour and communication.

1.1. Emotional cues are prioritised during voice perception

Vocal emotions may be expressed through speech prosody modulations or through purely nonverbal vocalizations (e.g. laughter, crying). Most studies probing vocal emotions have focused on speech prosody. However, nonverbal vocalizations arguably represent a more primitive form of emotional expression, and may be considered an auditory analogue of facial emotions (e.g. Belin, Fillion-Bilodeau, & Gosselin, 2008).

Event-related potential (ERP) studies indicate that emotional vocalizations are rapidly differentiated from neutral ones within 50 ms of exposure (Liu et al., 2012) and capture more attention (e.g. Pinheiro,

CONTACT Ana P. Pinheiro appinheiro@psicologia.ulisboa.pt
Supplemental data for this article can be accessed at https://doi.org/10.1080/02699931.2019.1586647.

Barros, Dias, & Kotz, 2017). The rapid and accurate detection of emotionally salient cues in the environment also requires the estimation of their proximity, for instance when these signals convey threat. In the example above, listeners would be particularly interested in identifying where the voice expressing distress or danger was coming from: did it come from the nearby kitchen (maybe an accident involving the oven?) or did it come from the more distant balcony (where a scary insect was hiding)? However, we know remarkably little about the interaction between emotional (*what?*) and spatial (*where?*) properties of voices.

1.2. Spatial and emotional representations interact during sound processing

Neuroimaging evidence indicates that different neural mechanisms underpin the processing of "what" (e.g. what type of emotional meaning does a vocalisation convey?) and "where" (e.g. where does a vocal emotion come from?) information. More generally, auditory processing is known to rely on two separate cortical streams that resemble the dorsal and ventral streams of the visual system (e.g. Rauschecker & Scott, 2009). Auditory spatial cues are processed primarily in posterior-medial regions of the auditory cortex (e.g. Altmann, Henning, Döring, & Kaiser, 2008), whereas auditory identity cues are processed in anterior-lateral regions of the auditory cortex (e.g. Grandjean, Sander, Lucas, Scherer, & Vuilleumier, 2008). While auditory ventral and dorsal streams have been mostly discussed in the context of speech perception, recent evidence suggests that this dual-stream architecture also supports non-linguistic processing (Sammler, Grosbras, Anwander, Bestelmeyer, & Belin, 2015).

Converging evidence from functional magnetic resonance imaging (fMRI), lesion and ERP studies indicates that spatial cues modulate the processing of emotional sounds. For instance, neuroimaging studies show that temporal lobe activity is modulated by the spatial features of vocal sounds: enhanced activation in right voice-sensitive regions is seen for proximal as compared to distal vocal anger (Ceravolo, Frühholz, & Grandjean, 2016b). Studies with right hemisphere stroke patients presenting left auditory extinction revealed that these patients detect prosodic stimuli presented to the left ear more easily when the stimuli are emotional *vs.* neutral, demonstrating that emotion moderates auditory extinction effects (Grandjean et al., 2008). Studies using cross-modal stimuli indicate that the detection of a visual target location is facilitated by spatially congruent emotional (but not neutral) prosody preceding the target (e.g. Brosch, Grandjean, Sander, & Scherer, 2009). A similar facilitation effect was observed in the detection of sine wave tones preceded by congruent emotional (but not neutral) prosodic stimuli (e.g. Ceravolo, Frühholz, & Grandjean, 2016a). These findings suggest that spatial processing is facilitated by the emotional properties of voices, plausibly via enhanced attentional orienting to emotionally arousing (*vs.* neutral) cues that boosts early processing stages (Liu et al., 2012). While distinct neurobiological pathways might support the processing of emotion and sound location, a robust body of evidence suggests that they interact during voice perception.

Even though the processing of vocal emotions involves both hemispheres, some acoustic features, namely pitch, were found to engage the right hemisphere to a greater extent (Zatorre & Gandour, 2008). Behaviourally, the facilitated detection of vocal emotions presented to the left ear was attributed to a right-hemisphere advantage (Alba-Ferrara, de Erausquin, Hirnstein, Weis, & Hausmann, 2013). These studies extend behavioural evidence supporting an implicit association of different emotion categories with different locations in the left-right space or up–down space (e.g. Amorim & Pinheiro, 2018). To date, however, no studies have examined how our capacity to decode vocal emotions is affected by their spatial location. It also remains unknown whether the spatial location of vocal sounds is more easily detected for emotional *vs.* neutral voices, i.e. if emotional salience confers an advantage in space processing.

1.3. The current study

Using a novel within-subjects design with spatialised sounds, the current study probed the interplay between emotion and spatial properties during voice perception. Nonverbal emotional vocalizations were selected to rule out potential confounding effects associated with the concurrent processing of semantic and lexical information in speech prosody. We presented vocalizations coming from different spatial locations around the head (left *vs.* right; front *vs.* back), and tested interactions between emotion and spatial location under different attentional requirements. In the first experiment, participants identified the source locations of neutral, positive (amusement) and negative (anger) nonverbal vocalizations presented in a virtual auditory environment. The same spatialised stimuli were presented in a second experiment, in which participants explicitly identified the emotional qualities of the vocalizations. Testing how the same vocal sounds are processed under different task instructions is particularly relevant as the attentional focus toward or away from the emotional quality of the stimulus influences how voices are processed (e.g. Frühholz, Ceravolo, & Grandjean, 2012).

This study is exploratory to a certain extent, as the interactions between spatial location and emotion have not been tested before in a unimodal paradigm with voices only. However, we anticipated that the spatial location of emotional voices could be identified more accurately than the location of neutral voices (Experiment 1). This hypothesis is indirectly supported by previous research revealing interactions between emotion and space, namely facilitated detection of auditory targets (sine wave tones) preceded by emotional compared to neutral prosodic cues presented at the same location (e.g. Ceravolo et al., 2016a), and of visual targets preceded by emotional compared to neutral prosody presented at congruent locations (Brosch et al., 2009). Plausibly, the emotional quality of the voice could influence spatial processing by enhancing attention orienting (Asutay & Västfjäll, 2015), which might be related to direct feedback signals from the amygdala to cortical sensory pathways (Sander, Grafman, & Zalla, 2003).

Additionally, we hypothesised that the emotional quality of vocalizations (Experiment 2) could be more accurately decoded if presented at back locations (the space behind the head where visual cues in social interactions are not available). This was supported by anecdotal and experimental evidence suggesting enhanced perceived arousal for sounds presented at the back *vs.* front locations (Asutay & Västfjäll, 2015; Occelli, Spence, & Zampini, 2011).

2. Method

2.1. Participants

Thirty college students (15 female; $M = 21.53$, $SD = 2.63$ years, age range 18–31 years) participated in the two experiments. Inclusion criteria were: European Portuguese as native language; no history of DSM-V diagnosis of drug or alcohol abuse (American Psychiatric Association [APA], 2013); no hearing or vision impairment; no history of psychiatric disorder; no current use of medication for psychiatric disorders; right handedness. Written informed consent was obtained from all participants, who received course credit for their participation in the experiments. The study was approved by the Ethics Committee of the University of Minho.

2.2. Stimuli

Emotional stimuli were selected from the corpus of nonverbal vocalizations (amusement; anger) by Lima, Castro, and Scott (2013). As this corpus does not contain neutral vocal expressions, neutral control vocalizations were selected from the Montreal Affective Voices battery (MAV; Belin et al., 2008), considering Portuguese normative ratings (Vasconcelos, Dias, Soares, & Pinheiro, 2017). Ten vocalizations portraying positive (amusement), negative (anger) and neutral states (50% female and 50% male vocalizations) with the highest recognition accuracy rates were selected.[1] The two emotion categories were selected as they have relatively similar acoustic profiles (e.g. high intensity and variable F0; Juslin & Laukka, 2003).

The stimuli included in the two experiments consisted of binaural sounds (frequency rate = 44.1 KHz) processed in free-field and generated in Matlab® using the MIT Head Related Transfer Function (HRTF) database (http://sound.media.mit.edu/resources/KEMAR.html). Stimuli were generated at [0°, 45°] elevation degrees combined with [45°, 90°, 135°, 225°, 270°, 315°] azimuth degrees allowing the listener to perceive the sounds at different locations (Figure 1). These locations were selected after a pilot study ($n = 37$ college students, 24 females), in which we determined the locations that were associated with above-chance recognition accuracy (for example, vocalizations presented at 0° and 180° locations could not be distinguished by the participants).

2.3. Procedure

Each participant was seated comfortably at a distance of 100 cm from a computer monitor in a sound-attenuating booth. Sounds were presented via Sennheiser HD 202 II headphones at 80 dB SPL. Participants responded using a standard QWERTY keyboard, with the response keys associated with each spatial location (Figure 1). The two experiments occurred in

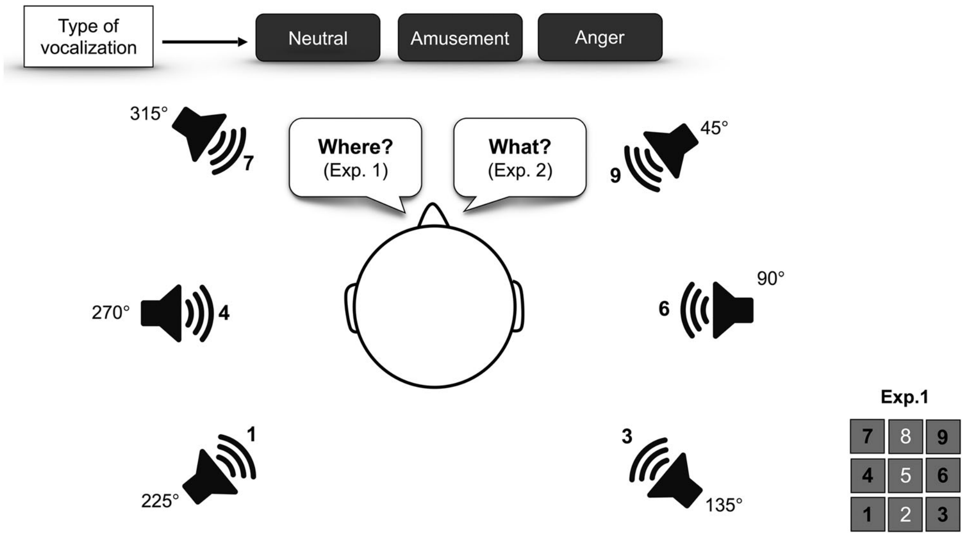

Figure 1. Illustration of the six spatial sources of the vocalizations presented in Experiment 1 and 2.

Notes: To generate these binaural sounds, impulse responses of each HRTF were convolved with the original sound samples, consequently originating a spatially located sound. The convolution of the HRTF with the sound sample added inter-aural time difference (ITD) and inter-aural level difference (ILD) cues to correspond to a new spatial location. In Experiment 1, participants assigned a key number to the spatial location of the vocal sound.

different days with at least 24 h between them. Stimulus presentation, timing of events and recording of participants' responses were controlled by Superlab 5.0 (Cedrus Corporation, San Pedro, CA, USA).

2.3.1. Experiment 1

Participants were instructed to discriminate between six different locations (Figure 1). Before the experiment, participants completed two training blocks to get familiarised with the instructions and task. Training stimuli were not used in the main experiment. During training, vocalizations could be replayed. The task was completed when participants reached 7 out of 10 consecutive correct answers in a maximum of 70 trials, otherwise they were excluded from subsequent analyses ($n = 4$; these participants were not included in the final sample described in the Methods section).

Before each sound, a fixation cross was presented centrally on the screen for 500 ms. At the end of the vocalisation, a question mark was presented during 5 s and the participant was required to indicate the location of the vocalisation. At the end of each trial (in both training and experimental blocks), feedback was provided. A total of 180 trials (60 per valence type and, for each valence type, 10 per spatial location) were presented in a randomised order. The experiment lasted approximately 40 min.

2.3.2. Experiment 2

Stimuli were the same as in Experiment 1. However, here participants were instructed to identify the emotional category associated with each vocalisation in a forced choice emotion recognition task. Additionally, participants were instructed to rate the arousal of each vocalisation (from 1 = not arousing at all, to 9 = extremely arousing).

Before the experimental block, participants completed a training session with 8 vocalizations to make sure they understood the instructions. Before each sound, a fixation cross was presented centrally on the screen for 500 ms. At the end of the vocalisation, the participant was instructed to identify the emotion category, and to subsequently rate the arousal of each vocalisation. The experiment included 216 trials (including 36 fillers, i.e. MAV vocalizations portraying fear, sadness and disgust). The experiment lasted approximately 45 min.

2.4. Statistical analyses

Unaggregated, trial-level data[2] were analysed with Bayesian mixed models (R 3.4.3). The accuracy of detecting the spatial location (Experiment 1) or

Table 1. Confusion matrix of responses in Experiment 1 by spatial location.

Target \ Response	45°	90°	135°	225°	270°	315°
45°	194	344	355	3	1	0
90°	172	527	197	0	0	0
135°	104	284	501	4	0	1
225°	0	0	2	453	289	155
270°	1	0	0	138	531	223
315°	0	1	0	266	385	245

emotional category (Experiment 2) of sounds was modelled as a binary outcome with location, emotion, and their interaction as fixed effects. We included a random intercept per stimulus to account for the variation in their intrinsic difficulty, and a random slope of location per participants. This structure of random effects was chosen based on theoretical considerations as well as comparisons of Widely Applicable Information Criterion (WAIC) values. Arousal ratings (Experiment 2) were modelled with the same random effects, but assuming the Gaussian distribution for the outcome variable. These models were fit with the brms package (Bürkner, 2017) using default, mildly normalising priors.

The location of sounds was entered as a categorical predictor with six possible values (i.e. all six angles were treated as unique). To compare left *vs.* right and front *vs.* back, the corresponding locations were then grouped without refitting the model, through averaging of fitted values. For example, the contrast between the accuracy of localising sounds coming from the front *vs.* back was calculated for each step in the MCMC chain as the difference between the predicted accuracy averaged for the two front angles (45° and 315°) and the two back angles (135° and 225°). The posterior distribution of each contrast was then summarised by its median and 95% coverage interval.

Statistical power analysis is presented in the online supplemental information file.

3. Results

3.1. Experiment 1

The overall location accuracy was 45.6% (chance level = 16.7%). The analysis of confusion patterns (Table 1) indicated that participants could nearly always guess whether the vocalisation had come from the left or right side (99.7% correct), with most errors resulting from a failure to locate the source in the sagittal plane (front *vs.* back).

As we aimed to probe the impact of emotion on the ability to detect spatial location, we explored a model predicting the accuracy of responses as a function of voice location and emotion, with an interaction. In general, vocalizations presented from the back (135° and 225°) were localised 33.5% (95% CI [23.5, 42.7]) more accurately than vocalizations presented in the front (45° and 315°; Figure 2). In contrast, there was no difference between the accuracy of localising vocalizations coming from the right (45°, 90° or 135°) *vs.* left (225°, 270° or 315°) side: − 0.9% [−4.6, 2.8]. We also observed no overall difference in the accuracy of localising either vocalizations with negative *vs.* positive valence (−0.03% [−3.27, 3.11]) or emotional *vs.* neutral vocalizations (2.1% [−0.65, 4.85]).

Notably, however, the Location × Emotion interaction significantly improved model fit as measured by a likelihood ratio test ($\chi^2 = 19.7$, $df = 10$, $p = .03$; WAIC = 6348.71 without interaction and 6348.69 with interaction, difference = −0.02, SE = 8.78). Partly supporting our hypothesis, emotional (negative or positive) vocalizations were localised more accurately than neutral ones when presented in the back (6.6% [1.5, 11.7]) or on the right side (4.4% [0.7, 8.2]). This was not the case when they were presented in the front (2.5% [−1.6, 6.8]) or on the left side (−0.25% [−4.27, 3.77]).

3.2. Experiment 2

3.2.1. Accuracy

The accuracy of vocal emotion recognition was overall high (94.8%), with no differences between right *vs.* left (−0.4% [−1.3, 0.3]) or back *vs.* front (0.1% [−0.9, 1.2]) locations (Figure 3). Averaging across all locations, recognition accuracy was 91.5% for anger, 95.7% for neutral and 98.5% for amusement. That is, accuracy was increased for positive relative to neutral vocalizations (predicted difference = 1.3% [0.6, 2.5]), and for neutral relative to negative vocalizations (predicted difference = 2.2% [1.0, 4.3]). Considering possible interactions between location and emotion, we found no evidence that the advantage of positive *vs.* neutral vocalizations in emotion recognition accuracy depended on whether the sounds were presented on the right *vs.* left (difference = 1.2% [−0.3, 3.6]) or in the front *vs.* back (−0.3% [−2.6, 1.9]). The Location x Emotion interaction was not a significant predictor of emotion recognition accuracy ($\chi^2 = 7.5$, $df = 10$, $p = .67$). The model without an interaction was strongly preferred based on WAIC (1626.93

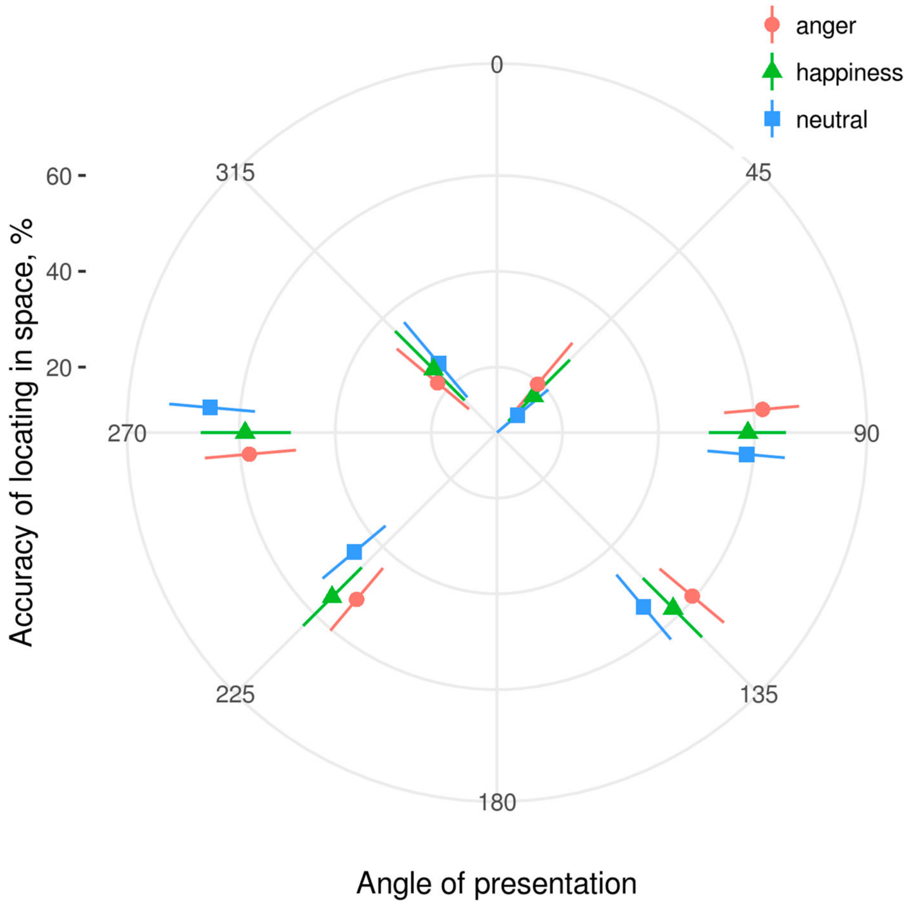

Figure 2. Accuracy in spatial location of vocalizations in Experiment 1 ("where").

without interaction and 1641.90 with interaction, difference = 14.97, SE = 6.39).

Arousal ratings were 0.6 [0.5, 0.7] points higher for negative *vs.* positive vocalizations and 2.1 [2.0, 2.2] higher for emotionally charged *vs.* neutral vocalizations. These ratings did not depend on whether vocalizations were presented on the left *vs.* right side (0.01 [−0.08, 0.09]), but vocalizations presented in front *vs.* back locations were judged to be 0.14 [0.04, 0.23] points more arousing. The advantage of negative *vs.* positive vocalizations in terms of perceived arousal was the same regardless of whether they were presented on the right *vs.* left (difference = 0.08 [−0.11, 0.27]) or in the front *vs.* back (−0.2 [−0.43, 0.03]). Likewise, the advantage of emotional *vs.* neutral vocalizations in terms of perceived arousal did not depend on whether they were presented on the right *vs.* left (−0.04 [−0.21, 0.12]) or front *vs.* back (0.08 [−0.13, 0.27]). The Location x Emotion interaction was not a significant predictor of arousal ratings ($\chi^2 = 5.76$, $df = 10$, $p = .83$), and the model without the interaction was strongly preferred based on WAIC (19,360.54 without interaction and 19,374.83 with interaction, difference = 14.29, SE = 4.66).

4. Discussion

The current study examined emotion and spatial location effects in a spatial hearing paradigm with nonverbal emotional vocalizations. Participants' performance was examined when attention was focused on the spatial location (Experiment 1) or on the emotional quality (Experiment 2) of the voice. In the first experiment, interaction effects between emotion and space were observed: spatial detection was facilitated when the voice had an emotional *vs.* neutral quality, namely when it was presented on the right side or from the back. In the second

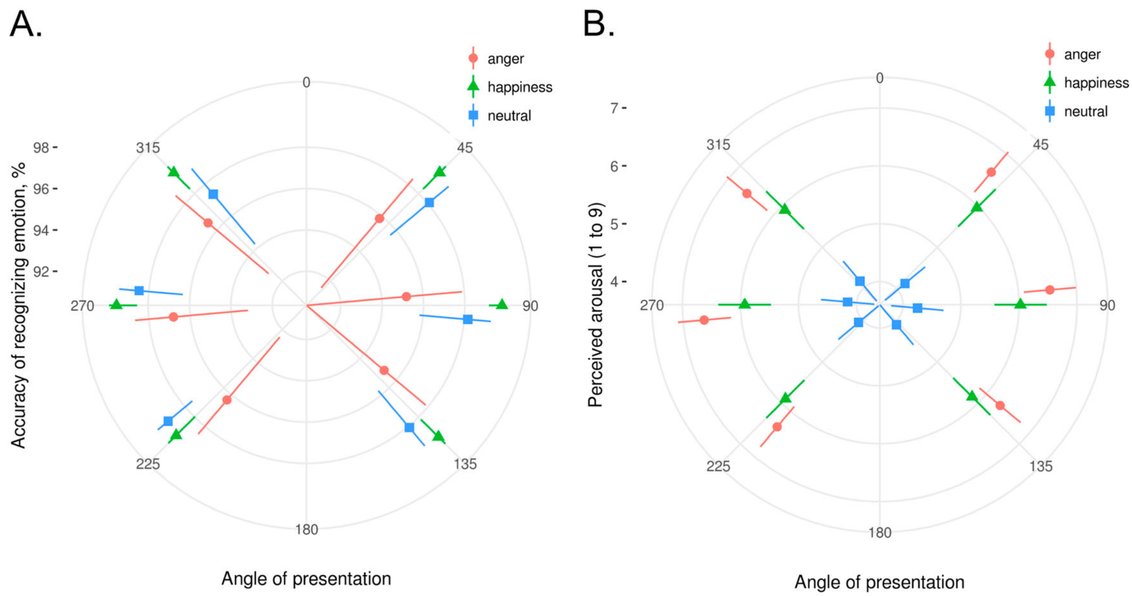

Figure 3. Accuracy in emotion recognition (Panel A) and perceived arousal (Panel B) of vocalizations in Experiment 2 ("what").

experiment, no interaction effects were observed: accuracy was enhanced for positive vocalizations; positive and negative vocalizations were rated as more arousing than neutral ones, but these effects were independent of spatial location.

4.1. Decoding the location of vocal emotions ("where")

Our first experiment revealed that, even if task-irrelevant, emotion impacted upon space detection and representation. The observed emotion-space interactions agree with previous evidence showing that emotion facilitates the discrimination of target locations in the absence of an explicit valence categorisation task (e.g. Amorim & Pinheiro, 2018).

Emotionally salient sounds have been found to modulate attention (Pinheiro et al., 2017), which could facilitate spatial detection. ERP studies provided evidence for an early automatic processing of emotional salience in the voice (Liu et al., 2012), indicating that emotional cues in the voice are prioritised in the processing stream, which has been linked to evolutionary adaptations. Nonetheless, the emotion facilitation effects in the current study were only observed for some locations, namely for vocalizations coming from the back or from the right side.

Detecting approaching objects may be particularly relevant when these objects come from the back of the head and, thereby, are out of sight, which may imply that front and rear space locations are represented by separate neural networks. Vision plays a critical role in the processing of spatial information in frontal space; hence, the location of sounds coming from behind a person's head rely almost exclusively on audition. Previous studies found that the processing of auditory spatial cues is improved in the rear compared to frontal space (Asutay & Västfjäll, 2015; reviewed in Occelli et al., 2011). Accordingly, studies with blind participants revealed faster vocal emotion discrimination as compared to sighted controls, and this behavioural advantage was paralleled by increased amygdala activation (Klinge, Röder, & Büchel, 2010). It is plausible that automatic capture of attention by emotional cues is increased in the rear space, leading to facilitated spatial detection. This could represent a compensatory mechanism in audition when visual cues are not available. The enhanced location accuracy of emotional vocalizations presented in rear space may involve similar mechanisms to those underlying auditory looming perception effects (Ghazanfar, Neuhoff, & Logothetis, 2002).

The facilitated detection of emotional vocalizations presented from the right side could be partly explained by manual motor fluency (Casasanto & Chrysikou, 2011). As participants responded with their (dominant) right hand, this could have changed space-valence mappings and lead to facilitated spatial judgements of emotional vocalizations presented at the same side of the response hand, irrespective of valence. However, this explanation is admittedly speculative and needs to be tested in future studies.

4.2. Decoding the emotional quality of vocalizations presented at different locations ("what")

When comparing emotion recognition accuracy for spatialised vocalizations, our prediction of emotion-space interactions was not confirmed. Studies with non-spatialised vocal emotional stimuli have consistently demonstrated the preferential processing of emotional vocal sounds, reflected in increased recognition accuracy (e.g. Lima et al., 2013; Vasconcelos et al., 2017). The current study revealed a specific advantage of positive vocalizations in terms of emotion recognition accuracy. The positivity bias in Experiment 2 confirms prior reports of enhanced recognition accuracy of non-spatialised positive (laughter) vocal emotions in participants from the same cultural context (Vasconcelos et al., 2017). Nonetheless, both positive and negative vocalizations were generally rated as more arousing than neutral vocalizations, irrespective of space. Previous studies revealed that explicit emotion recognition is fast and relatively impermeable to cognitive load manipulations, indicating that it relies on largely automatic mechanisms (Lima, Anikin, Monteiro, Scott, & Castro, 2018). The current findings add to this evidence by suggesting that emotion decoding is also robust to space manipulations.

4.3. Processing spatialised vocal emotions as a function of task instructions

The distinct pattern of results observed in Experiment 1 and Experiment 2 indicates that voice processing is affected by task instructions and, specifically, by attention focus. Emotion interacted with space when the focus was on the detection of spatial location, but not when the focus was on explicit emotional evaluations. This is in good agreement with prior studies showing a differential processing of vocal emotions under implicit (e.g. identifying the gender of a vocalisation) *vs.* explicit (e.g. decoding the emotion category of the voice) task instructions (Frühholz et al., 2012).

The current findings open a new avenue of inquiry regarding the interactions between "what" and "where" properties of vocal emotions. Additionally, they have implications for models of vocal emotional processing and voice perception more generally, by pointing out the need to account for – and to better understand – how emotional and spatial properties modulate implicit and explicit aspects of vocal processing. We should note, though, that our results reflect later evaluative processes. Evidence that describes how these two types of information interact through distinct processing stages of voice perception is still missing. Future studies using fMRI and EEG will clarify the neural mechanisms underpinning the interactions between emotion and sound localisation from stimulus onset until a response is made.

5. Conclusions

The current study investigated how emotion interacts with spatial location during voice perception, when attention was focused on the spatial source (Experiment 1) or on the emotional quality of the voice (Experiment 2). Emotion-space interactions were observed when the emotional quality of the vocalizations was task-irrelevant (Experiment 1), as indicated by a facilitated spatial location of emotional *vs.* neutral voices for back and right side locations. When participants had to explicitly identify the emotional qualities of vocalizations (Experiment 2), accuracy did not vary by space.

These findings demonstrate that emotional vocal sounds are prioritised also when the spatial source of the stimulus is manipulated. Together, they support the interaction between spatial ("where") and emotional ("what") properties of the voice, which differs as a function of task instructions.

Notes

1. The vocalizations selected were: Anger - Anger_C_5, Anger_C_3, Anger_C_2, Anger_C_1, Anger_C_4, Anger_C_6, Anger_T_11, Anger_T_12, Anger_M_7, Anger_M_8; Amusement - Amusement_C_3, Amusement_C_4, Amusement_MS_10, Amusement_C_2, Amusement_C_1, Amusement_T_14, Amusement_T_16, Amusement_M_5, Amusement_M_6, Amusement_T_15; Neutral - 6_neutral, 42_neutral, 45_neutral, 46_neutral, 53_neutral, 55_neutral, 58_neutral, 59_neutral, 60_neutral, 61_neutral.
2. Raw data available at https://osf.io/s6785.

Acknowledgments

The Authors gratefully acknowledge all the participants who collaborated in the study, and particularly Prof. Jorge Almeida Santos for help with stimulus generation.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by grant numbers IF/00334/2012 and PTDC/MHC-PCN/0101/2014 awarded to APP. These Grants were funded by the Science and Technology Foundation (Fundação para a Ciência e a Tecnologia - FCT, Portugal) and FEDER (European Regional Development Fund) through the European programmes QREN (National Strategic Reference Framework) and COMPETE (Operational Programme "Thematic Factors of Competitiveness"). During the preparation of the manuscript, CFL was supported by an FCT Investigator [grant number IF/00172/2015]. This study was conducted at CIPsi, University of Minho, and additionally supported by FCT [grant number PSI/01662] and the Portuguese Ministry of Science, Technology and Higher Education through national funds and partially by FEDER through COMPETE 2020 under the PT2020 Partnership Agreement [POCI-01-0145-FEDER-007653].

ORCID

Ana P. Pinheiro http://orcid.org/0000-0002-7981-3682

References

Alba-Ferrara, L., de Erausquin, G. A., Hirnstein, M., Weis, S., & Hausmann, M. (2013). Emotional prosody modulates attention in schizophrenia patients with hallucinations. *Frontiers in Human Neuroscience*, *7*, 59. doi:10.3389/fnhum.2013.00059

Altmann, C. F., Henning, M., Döring, M. K., & Kaiser, J. (2008). Effects of feature-selective attention on auditory pattern and location processing. *NeuroImage*, *41*(1), 69–79. doi:10.1016/j.neuroimage.2008.02.013

American Psychiatric Association. (2013). *Diagnostic* and statistical manual of mental disorders (5th ed.). Washington, DC: American Psychiatric Association.

Amorim, M., & Pinheiro, A. P. (2018). Is the sunny side up and the dark side down? Effects of stimulus type and valence on a spatial detection task. *Cognition and Emotion*. doi:10.1080/02699931.2018.1452718

Asutay, E., & Västfjäll, D. (2015). Attentional and emotional prioritization of the sounds occurring outside the visual field. *Emotion*, *15*(3), 281–286. doi:10.1037/emo0000045

Belin, P., Fillion-Bilodeau, S., & Gosselin, F. (2008). The montreal affective voices: A validated set of nonverbal affect bursts for research on auditory affective processing. *Behavior Research Methods*, *40*(2), 531–539. doi:10.3758/BRM.40.2.531

Belin, P., Zatorre, R., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, *403*, 309–312. doi:10.1038/35002078

Brosch, T., Grandjean, D., Sander, D., & Scherer, K. R. (2009). Cross-modal emotional attention: Emotional voices modulate early stages of visual processing. *Journal of Cognitive Neuroscience*, *21*(9), 1670–1679. doi:10.1162/jocn.2009.21110

Bürkner, P.-C. (2017). Brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. doi:10.18637/jss.v080.i01

Casasanto, D., & Chrysikou, E. G. (2011). When left is "right": motor fluency shapes abstract concepts. *Psychological Science*, *22*(4), 419–422. doi:10.1177/0956797611401755

Ceravolo, L., Frühholz, S., & Grandjean, D. (2016a). Modulation of auditory spatial attention by angry prosody: An fMRI auditory dot-probe study. *Frontiers in Neuroscience*, *10*, 216. doi:10.3389/fnins.2016.00216

Ceravolo, L., Frühholz, S., & Grandjean, D. (2016b). Proximal vocal threat recruits the right voice-sensitive auditory cortex. *Social Cognitive and Affective Neuroscience*, *11*(5), 793–802. doi:10.1093/scan/nsw004

Frühholz, S., Ceravolo, L., & Grandjean, D. (2012). Specific brain networks during explicit and implicit decoding of emotional prosody. *Cerebral Cortex*, *22*(5), 1107–1117. doi:10.1093/cercor/bhr184

Ghazanfar, A. A., Neuhoff, J. G., & Logothetis, N. K. (2002). Auditory looming perception in rhesus monkeys. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(24), 15755–15757. doi:10.1073/pnas.242469699

Grandjean, D., Sander, D., Lucas, N., Scherer, K. R., & Vuilleumier, P. (2008). Effects of emotional prosody on auditory extinction for voices in patients with spatial neglect. *Neuropsychologia*, *46*(2), 487–496. doi:10.1016/j.neuropsychologia.2007.08.025

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

Klinge, C., Röder, B., & Büchel, C. (2010). Increased amygdala activation to emotional auditory stimuli in the blind. *Brain*, *133*(Pt 6), 1729–1736. doi:10.1093/brain/awq102

Lima, C. F., Anikin, A., Monteiro, A. C., Scott, S. K., & Castro, S. L. (2018). Automaticity in the recognition of nonverbal emotional vocalizations. *Emotion*. doi:10.1037/emo0000429

Lima, C. F., Castro, S. L., & Scott, S. K. (2013). When voices get emotional: A corpus of nonverbal vocalizations for research on emotion processing. *Behavior Research Methods*, *45*(4), 1234–1245. doi:10.3758/s13428-013-0324-3

Liu, T., Pinheiro, A. P., Deng, G., Nestor, P. G., McCarley, R. W., & Niznikiewicz, M. A. (2012). Electrophysiological insights into processing nonverbal emotional vocalizations. *NeuroReport*, *23*(2), 108–112. doi:10.1097/WNR.0b013e32834ea757

Occelli, V., Spence, C., & Zampini, M. (2011). Audiotactile interactions in front and rear space. *Neuroscience and Biobehavioral Reviews*, *35*(3), 589–598. doi:10.1016/j.neubiorev.2010.07.004

Pinheiro, A. P., Barros, C., Dias, M., & Kotz, S. A. (2017). Laughter catches attention! *Biological Psychology*, *130*, 11–21. doi:10.1016/j.biopsycho.2017.09.012

Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nature Neuroscience*, *12*(6), 718–724. doi:10.1038/nn.2331

Sammler, D., Grosbras, M. H., Anwander, A., Bestelmeyer, P. E. G., & Belin, P. (2015). Dorsal and ventral pathways for prosody. *Current Biology*, *25*(23), 3079–3085. doi:10.1016/j.cub.2015.10.009

Sander, D., Grafman, J., & Zalla, T. (2003). The human amygdala: An evolved system for relevance detection. *Reviews in the Neurosciences*, *14*(4), 303–316. doi:10.1515/REVNEURO.2003.14.4.303

Vasconcelos, M., Dias, M., Soares, A. P., & Pinheiro, A. P. (2017). What is the melody of that voice? Probing unbiased recognition accuracy with the montreal affective voices. *Journal of Nonverbal Behavior*, *41*(3), 239–267. doi:10.1007/s10919-017-0253-4

Zatorre, R. J., & Gandour, J. T. (2008). Neural specializations for speech and pitch: Moving beyond the dichotomies. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1493), 1087–1104. doi:10.1098/rstb.2007.2161