Available online at www.sciencedirect.com

ScienceDirect

Journal homepage: www.elsevier.com/locate/cortex

Research Report

Attention and emotion shape self-voice prioritization in speech processing

Ana P. Pinheiro [a,b,*], *João Sarzedas* [a], *Magda S. Roberto* [a] and *Sonja A. Kotz* [b]

[a] *CICPSI, Faculdade de Psicologia, Universidade de Lisboa, Lisboa, Portugal*
[b] *Basic and Applied NeuroDynamics Lab, Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, the Netherlands*

ARTICLE INFO

Article history:
Received 25 April 2022
Reviewed 13 August 2022
Revised 27 September 2022
Accepted 6 October 2022
Action editor Veena Kumari
Published online 26 October 2022

Keywords:
Speech
Voice
Self-relevance
Emotion
Event-related potentials

ABSTRACT

Both self-voice and emotional speech are salient signals that are prioritized in perception. Surprisingly, self-voice perception has been investigated to a lesser extent than the self-face. Therefore, it remains to be clarified whether self-voice prioritization is boosted by emotion, and whether self-relevance and emotion interact differently when attention is focused on *who* is speaking *vs.* *what* is being said.

Thirty participants listened to 210 prerecorded words spoken in one's own or an unfamiliar voice and differing in emotional valence in two tasks, manipulating the attention focus on either speaker identity or speech emotion. Event-related potentials (ERP) of the electroencephalogram (EEG) informed on the temporal dynamics of self-relevance, emotion, and attention effects.

Words spoken in one's own voice elicited a larger N1 and Late Positive Potential (LPP), but smaller N400. Identity and emotion interactively modulated the P2 (self-positivity bias) and LPP (self-negativity bias). Attention to speaker identity modulated more strongly ERP responses within 600 ms post-word onset (N1, P2, N400), whereas attention to speech emotion altered the late component (LPP). However, attention did not modulate the interaction of self-relevance and emotion.

These findings suggest that the self-voice is prioritized for neural processing at early sensory stages, and that both emotion and attention shape self-voice prioritization in speech processing. They also confirm involuntary processing of salient signals (self-relevance and emotion) even in situations in which attention is deliberately directed away from those cues. These findings have important implications for a better understanding of symptoms thought to arise from aberrant self-voice monitoring such as auditory verbal hallucinations.

* *Corresponding author*. Faculdade de Psicologia, Universidade de Lisboa, Alameda da Universidade, 1649-013 Lisboa, Portugal.
E-mail address: appinheiro@psicologia.ulisboa.pt (A.P. Pinheiro).

https://doi.org/10.1016/j.cortex.2022.10.006

1. Introduction

Both self-relevant and emotionally salient signals are prioritized in perception. In the auditory domain, one's own voice is likely the most self-relevant sound and the one we hear most often: we hear it every time we speak. Compared to other voices, one's own voice may attract more attention and be more easily discriminated (Conde, Gonçalves, & Pinheiro, 2018; Pinheiro, Farinha-Fernandes, et al., 2019). The self-voice is also represented differently from other voices in the brain (Allen, Freeman, Johns, & McGuire, 2006; Kaplan, Aziz-Zadeh, Uddin, & Iacoboni, 2008; Nakamura et al., 2001). Accordingly, a self-voice-specific EEG topographic map, visible around 345 ms post-stimulus onset, was identified (Iannotti et al., 2021). The self-voice differs from other voices not only in familiarity but also in subjective emotional relevance (Conde, Gonçalves, & Pinheiro, 2015; Conde, Gonçalves, & Pinheiro, 2016). Notwithstanding, the self-prioritization effect (Sui, He, & Humphreys, 2012) does not seem to be accounted for by stimulus familiarity (Payne, Lavan, Knight, & McGettigan, 2021), reward value (Wang, Qi, Li, & Jia, 2021; Yankouskaya et al., 2017), emotional valence (Schäfer, Wentura, & Frings, 2020), or its potentially greater physical salience (Schäfer, Wentura, & Frings, 2017).

Perception is not only biased towards what is self-relevant but also towards what is emotionally salient (Kousta, Vinson, & Vigliocco, 2009). For example, emotional stimuli are often detected more rapidly than neutral stimuli even at threshold detection levels (Swanborough, Staib, & Frühholz, 2020). Stimuli tagged as emotionally significant may also receive privileged access to attention (Vuilleumier, 2005). Processing similarities were noted between self-relevant and emotional vocal stimuli: both are associated with faster and more accurate behavior (Payne et al., 2021; Pinheiro, Farinha-Fernandes, et al., 2019; Pinheiro, Lima, et al., 2019); also, both capture, hold, and bias attention (Conde et al., 2015, 2018; Pinheiro, Barros, et al., 2017; Pinheiro, Barros, & Pedrosa, 2015). Their effects are beneficial in a dynamically changing sensory environment, allowing for rapid differentiation of what is – and is not – relevant.

Notably, despite similar perceptual prioritization effects, voice identity (subserving the recognition of a voice as one's own) and emotion are processed in partially dissociated cortical regions (Belin, Fecteau, & Bédard, 2004). However, very few studies probed how self-voice perception is affected by emotion. This is particularly relevant for speech comprehension, in which multiple cues (verbal and non-verbal) need to be integrated in lightning speed (Scott, 2019; von Kriegstein, Eger, Kleinschmidt, & Giraud, 2003). Speaker-related features (e.g., age) are known to modulate how a verbal message is interpreted (Foucart et al., 2015; Tesink et al., 2009; van Berkum, van den Brink, Tesink, Kos, & Hagoort, 2008), but the emotional quality of a stimulus may also affect self-related processes such as self-other voice discrimination (Pinheiro, Farinha-Fernandes, et al., 2019; Pinheiro, Rezaii, et al., 2017).

As the two voice dimensions – identity and emotion – may operate at different timescales, this could affect their interaction. For example, whereas voice identity effects have been most consistently found in early sensory stages (Pinheiro, Schwartze, & Kotz, 2018; Pinheiro, Rezaii, Nestor, et al., 2016), emotion has been found to produce more consistent effects typically around 200 ms post-stimulus onset (Hajcak, Dunning, & Foti, 2009). Attention could also shape how self-relevance and emotion independently and interactively affect speech processing. For example, attention might boost sensory representations that have behavioral relevance in a particular task context, such as *what* is said *vs.* *who* is speaking. Surprisingly, self-voice perception has been investigated to a lesser extent than the self-face (Uddin, Kaplan, Molnar-Szakacs, Zaidel, & Iacoboni, 2005). A better understanding of the neural mechanisms underpinning self-voice perception is critical considering that phenomena such as auditory verbal hallucinations have been linked to dysfunctional monitoring of one's own voice (Pinheiro et al., 2018; Pinheiro, Rezaii, et al., 2017; Pinheiro, Schwartze, Amorim, et al., 2020; Pinheiro, Schwartze, & Kotz, 2020).

1.1. *Is self-voice prioritization boosted by emotion?*

Due to its exquisite temporal resolution, electroencephalography is a particularly powerful tool to disentangle at which stages of auditory perception self-relevance ('my voice') and emotion produce independent *vs.* interactive effects. However, EEG studies of self-voice processing have often relied on oddball paradigms (Conde et al., 2015, Conde et al., 2016; Graux et al., 2013), with only a few probing how emotion affects the discrimination of one's own voice (Pinheiro, Rezaii, et al., 2017; Pinheiro, Rezaii, Nestor, et al., 2016).

These studies revealed an early self-other voice distinction within 100 milliseconds (ms) of word onset, reflected in N1 amplitude modulations: the N1 was found to be larger in response to prerecorded words spoken in one's own compared to an unfamiliar voice (Pinheiro, Rezaii, Nestor, et al., 2016). This finding suggests increased automatic attention to self-related information (Haider, Spong, & Lindsley, 1964; Luck, Woodman, & Vogel, 2000). Emotion ERP effects consistently show in the P2, linked to the detection of emotional stimulus salience (Paulmann & Kotz, 2008; Schirmer & Kotz, 2006): the P2 amplitude is increased for emotional compared to neutral words (e.g., Kanske & Kotz, 2007) or non-verbal vocalizations (e.g., Liu et al., 2012). However, emotion and self-relevance affect late processing stages similarly (Conde et al., 2015; Gray, Ambady, Lowenthal, & Deldin, 2004; Heine, Lehman, Markus, & Kitayama, 1999; Phan et al., 2004). Significantly enhanced long lasting late ERP components, such as the late positive potential (LPP), have been reported for words spoken in one's own (*vs.* other) voice (Pinheiro, Rezaii, Nestor, et al., 2016) or with an emotional (*vs.* neutral) quality (Hatzidaki, Baus, & Costa, 2015; Herbert, Junghofer, & Kissler, 2008; Hettich et al., 2016; Schirmer & Gunter, 2017). The enhanced LPP effect is thought to reflect sustained elaborative processing for stimuli that are deemed self-relevant or emotionally salient (Hajcak, Macnamara, & Olvet, 2010).

Interactive effects of emotion and self-relevance were identified during early speech processing stages. Specifically, positive words spoken in one's own voice elicited an increased P2 response relative to the same words spoken by another speaker (Pinheiro, Rezaii, Nestor, et al., 2016). The P2 amplitude modulations suggest a self-positivity bias (Chen et al., 2014; Fields & Kuperberg, 2015; Watson, Dritschel, Obonsawin, & Jentzsch, 2007), reflected in facilitated processing of positive stimuli related to the self. Interactions of emotion and self-relevance also show in later latency windows. Emotional words were associated with larger LPP amplitudes when spoken in one's own voice compared to an unfamiliar voice, irrespective of valence (Pinheiro, Rezaii, Nestor, et al., 2016). These interactions reveal that the effects of self-relevance and emotion do not merely sum up but rather form an interdependent relationship.

1.2. Is self-voice prioritization affected by attention?

The results described so far often relate to explicit tasks, where a listener's attention is focused on either self-other voice discrimination (e.g., Pinheiro, Rezaii, Nestor, et al., 2016) or specific speech properties (e.g., Hatzidaki et al., 2015). It remains to be clarified whether self-relevance interacts with emotion differently when the listener is instructed to focus their attention on *who* is speaking *vs.* *what* is said.

In the auditory domain, attention is known to modulate the ERP components that respond sensitively to self-relevance and emotion modulations. For example, increased attention focus is reflected in enhanced N1 amplitudes (Lange, 2013). A P2 amplitude increase has been related to automatic capture of attentional resources by emotionally salient stimuli (Pinheiro, Rezaii, Nestor, et al., 2016), whereas increased LPP amplitudes have been associated with enhanced sustained attention to either emotionally relevant or self-relevant stimuli (Ferrari, Codispoti, Cardinale, & Bradley, 2008; Hajcak et al., 2009).

Attention can also change the effects of self-relevance and emotion on speech processing. For example, the attention-grabbing properties of one's own voice are dependent on task instructions: attention to the self- (*vs.* other-) voice is increased only when task-relevant (Conde et al., 2015). Accuracy in self-voice recognition may also change when comparing implicit and explicit tasks (Candini et al., 2014). Along the same lines, functional magnetic resonance imaging (MRI) studies revealed distinct brain activation patterns as a function of explicit attention to speech *vs.* speaker information: the right anterior superior temporal sulcus activated only when attention was directed to the speaker's voice, whereas the left middle temporal region activated more when attention was directed to speech content (von Kriegstein et al., 2003). Emotion is also known to modulate voice perception, even when attention is focused elsewhere (Liu et al., 2012; Pinheiro et al., 2015; Sander et al., 2005).

These findings indicate that the neurocognitive resources engaged in the processing of complex vocal signals may depend on whether attention is focused on the speaker (e.g., self-other discrimination) or on speech content (e.g., emotional categorization). This could affect how self-relevance and emotion interact, a hypothesis that remains untested.

1.3. The current study

The present study investigated whether and when self-voice prioritization is boosted by emotion. Furthermore, it probed whether the self-voice is perceived differently when attention is focused on speaker identity compared to speech emotion. Given its high temporal resolution, the EEG was used to characterize the temporal dynamics of self-relevance and emotion effects on speech processing. Consequently, two tasks were designed to dissociate non-verbal (speaker identity) from verbal (emotional content) vocal features, while including identical stimulus material. Participants completed identity recognition and emotion categorization tasks on the same stimulus set.

Hearing a recording of the self-voice differs from hearing vocal feedback in speech production (Maurer & Landis, 1990). Notwithstanding, previous studies showed that prerecorded self-voice stimuli are recognized as “self” with an accuracy rate above 89% (Hughes & Nicholson, 2010; Nakamura et al., 2001; Pinheiro, Farinha-Fernandes, et al., 2019; Pinheiro, Rezaii, Rauber, et al., 2016; Rosa, Lassonde, Pinard, Keenan, & Belin, 2008). Therefore, we expected above-chance accuracy in explicit self-other voice recognition.

Concerning the question of whether self-voice prioritization is boosted by emotion, we hypothesized an interaction between identity and emotion in the P2 and LPP components in both tasks, replicating previous findings (Pinheiro, Rezaii, Nestor, et al., 2016). According to appraisal theories of emotion (e.g., Scherer, 2001), stimuli are rapidly and automatically evaluated for their emotional relevance. Therefore, spoken words that are both self-relevant and emotional should be prioritized in speech perception. Specifically, the P2 was expected to increase in response to positive words spoken in one's own *vs.* an unfamiliar voice, whereas increased LPP amplitudes were expected in response to emotional words uttered in the self-voice (Pinheiro, Rezaii, Nestor, et al., 2016).

Concerning the second question of whether self-voice prioritization is affected by attention, hypotheses were exploratory. Auditory representations of voices could be boosted for task-relevant dimensions relatively early (within the first 200 ms). This could lead to the enhanced processing of emotional features in the emotion categorization task, and of the self-voice in the identity recognition task. However, auditory representations could be sensitive to the same stimulus information regardless of task in early time windows, with task sensitivity emerging only later on. A third possibility (not mutually exclusive) is that the self-voice may be a specifically salient class of auditory information that is prioritized for neural processing even when task-irrelevant.

2. Method

We report how we determined our sample size, all inclusion/exclusion criteria, whether inclusion/exclusion criteria were established prior to data analysis, all manipulations, and all measures in the study.

2.1. *Participants*

Thirty participants (15 females; mean age = 24.30, *SD* = 2.51, age range 18–33 years) were included in the study.[1] No data were excluded from statistical analyses. Inclusion criteria were: 1) European Portuguese as the native language; 2) right-handedness (Oldfield, 1971); 3) no history of psychiatric, neurological, or major medical illness; 4) no history of drug or alcohol abuse in the last 6 months; 5) no present medication for clinical disorders that would affect EEG morphology or have consequences on cognitive performance. The inclusion/exclusion criteria were established prior to data analysis. All participants received instructions about the procedure and read and signed an informed consent form to confirm their willingness to participate in the study. The study was approved by the ethics committee of the Faculty of Psychology of the University of Lisbon.

2.2. *Stimuli*

Stimuli included 35 words with negative valence (e.g., "cruel"), 35 words with neutral valence (e.g., "scrambled"), and 35 words with positive valence (e.g., "joyful"). The proportion of nouns and adjectives was equivalent across conditions (see Supplement). Words were controlled for frequency, number of letters, and number of syllables (Soares et al., 2018). Valence and arousal ratings were obtained from affective norms for 1043 European Portuguese words (Soares, Comesaña, Pinheiro, Simões, & Frade, 2012). To further confirm these ratings, 28 students (who did not participate in the EEG experiment) provided ratings of valence and arousal for each of the 105 words using a 9-point Likert scale (valence: 1 = extremely unpleasant to 9 = extremely pleasant; arousal: 1 = not at all aroused to 9 = extremely aroused). There were no differences between word valence categories regarding number of letters, number of syllables, and frequency ($p > .05$; Table 1). Moreover, there were no arousal differences for negative and positive words ($p > .05$), but emotional words were characterized by higher arousal ratings than neutral words. As expected, neutral words were perceived as more pleasant than negative words ($p < .001$), but less pleasant than positive words ($p < .001$).

2.3. *Procedure*

Each participant was tested in two experimental sessions. The first session involved the recording of the participant's voice, whereas the EEG recording took place in the second session.

Table 1 – Psycholinguistic and affective properties of the selected words.

	Negative *M (SD)*	Neutral *M (SD)*	Positive *M (SD)*	F
Frequency	13.40 (14.45)	15.32 (16.67)	15.63 (19.39)	.161
Number of letters	6.80 (1.64)	7.06 (1.86)	7.26 (2.17)	.552
Number of syllables	2.97 (.82)	3.14 (.77)	3.26 (1.01)	1.014
Valence	2.30 (1.38)	5.15 (1.30)	7.64 (1.32)	3390.887*
Arousal	5.86 (2.19)	4.20 (1.91)	5.73 (2.28)	34.187*

Note. *M* = Mean; *SD* = Standard Deviation. Valence and arousal ratings range between 1 and 9. * $p < .001$.

No part of the study procedures was pre-registered prior to the research being conducted.

2.3.1. *Voice recording*

Each participant was asked to read aloud a list of 105 words with neutral or emotional valence (self-voice condition). The words were displayed in the center of a computer screen, one at a time. Before seeing the word, participants were instructed to listen to that same word pronounced by a 'voice-model' (age = 24) using neutral prosody. They were then instructed to match the loudness and neutral prosody of each target word as spoken by the 'voice-model' at a constant voice intensity. The inclusion of a voice-model served to reduce variability between the participants' speech rate, voice loudness, and pitch. Recordings were made in an echoic studio with a Roland R-16 recorder, with a sampling rate of 44,100 kHz and 16-bit quantization.

For the other-voice condition, the same 105 words were recorded by a male (age = 23) or female (age = 25) native speaker of European Portuguese unknown to the participants. The words were spoken with neutral intonation and constant voice intensity, following the same procedure as described above. After the recording session, each word was segmented using Praat software (Boersma & Weenink, 2013). Then, the intensity of the voice stimuli was normalized to 70 dB, using a Praat script. The mean pitch and duration of each word are presented in Table 2.

Table 2 – Acoustic properties of spoken words in the self- and other-voice conditions.

	Negative *M (SD)*	Neutral *M (SD)*	Positive *M (SD)*
Duration (ms)	S-M: 755.75 (68.49) O-M: 836.57 S–F: 778.69 (40.76) O–F: 825.14	S-M: 755.71 (71.13) O-M: 851.71 S–F: 788.02 (42.94) O–F: 828.28	S-M: 794.04 (70.03) O-M: 879.14 S–F: 829.14 (44.07) O–F: 869.71
Mean F0 (Hz)	S-M: 111.19 (18.57) O-M: 115.44 S–F: 195.67 (18.21) O–F: 192.24	S-M: 111.80 (19.07) O-M: 114.81 S–F: 193.98 (18.86) O–F: 193.21	S-M: 111.97 (19.17) O-M: 114.92 S–F: 196.64 (19.28) O–F: 192.21

Note. *M* = Mean; SD = Standard Deviation; S-M = self-voice (male); S–F = self-voice (female); O-M = other-voice (male); O–F = other-voice (female).

[1] Sample size was not determined by power analyses because it would be difficult to extract an expected effect size from previous work for the different effects that we examined (given the scarcity of directly related studies). Instead, we aimed at testing as many participants as possible within the time we had available for data collection and reach numbers comparable to previous ERP studies of self-voice perception (e.g., Graux et al., 2013; Iannotti et al., 2021; Pinheiro, Rezaii, Nestor, et al., 2016).

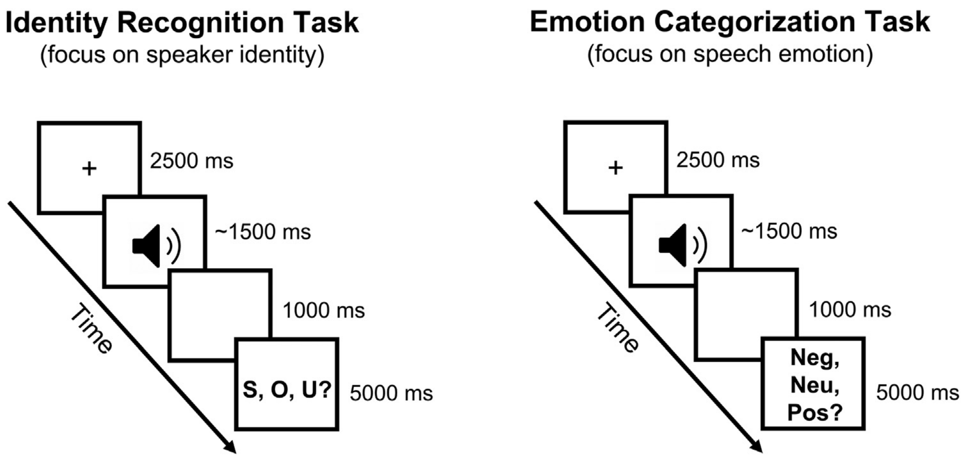

Fig. 1 – Schematic illustration of an experimental trial in the identity recognition and emotion categorization tasks. Note: S = self-voice; O = other-voice; U = unsure; Neg = negative; Neu = neutral; Pos = positive.

In a pilot study, we identified word recognition points per emotion condition (see Supplement). On average, participants needed 469.55 ms ($SD = 63.71$) to recognize the positive words, 451.59 ms ($SD = 45.48$) to recognize the neutral words, and 442.12 ms ($SD = 52.12$) to recognize the negative words. Paired *t*-tests showed that both negative ($t(32) = -3.564$, $p = .001$) and neutral words ($t(32) = -2.219$, $p = .034$) were recognized faster than positive words. No differences were found between neutral and negative words ($t(32) = -1.571$, $p = .126$). Differences in stimulus duration could have, nonetheless, contributed to these findings. Irrespective of speaker identity and sex, the duration of positive spoken words was longer compared to both negative ($t(1119) = 7.663$, $p < .001$) and neutral words ($t(1119) = 5.363$, $p < .001$), whereas there was no difference between neutral and negative words ($t(1119) = .714$, $p = .475$; positive: $M = 814.61$ ms, $SD = 183.27$ ms; negative: $M = 771.90$ ms, $SD = 158.77$ ms; neutral: $M = 776.13$ ms; $SD = 140.99$ ms).

2.3.2. *EEG task*

The EEG session took place at least three days after the voice recording session.[2] Two hundred and ten words were presented: 105 previously recorded by the participant (self-voice condition – SV), and 105 previously recorded by a speaker unknown to the participant (other-voice condition – OV; same voice for all participants – male participants listened to a male voice, whereas female participants listened to a female voice in the OV condition). There were 35 spoken words for each of the six combinations of emotion and identity: SV-neutral, SV-positive, SV-negative, OV-neutral, OV-positive, and OV-negative. The six stimulus categories were pseudorandomized and presented in two lists, with the limitation of no more than three consecutive trials of the same condition. Half of the participants received the lists in a fixed sequence, and the other half in the inverse sequence.

During EEG recordings, all participants completed two tasks. In both tasks, they chose one of three alternatives. In the first task, identity recognition was required: participants were instructed to attend to non-verbal information (focus on speaker identity; Fig. 1) and decide whether the voice they heard was their own voice, the voice of another person, or if they were unsure. In the second task (emotion categorization), they were instructed to attend to verbal information and decide whether the words were positive, negative, or neutral (focus on speech emotion; Fig. 1). The order of the button presses was counterbalanced. Stimuli were presented through headphones at a sound level comfortable for each participant and were not repeated during the experiment.

Each participant was seated at 80 cm distance from a computer monitor in a soundproof boot. Before each sound onset, a fixation cross was presented on the center of the screen for 2500 ms and was kept during sound presentation (1500 ms) to minimize eye movements. Then, a question mark appeared for 1000 ms (inter-stimulus interval) signaling the beginning of the response time (maximum duration = 5 s; Fig. 1). The experimental task was run with Presentation® software (Version 20.1, Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com; see https://osf.io/dz68e/?view_only=583d25eb5683412ba49ea934cb097ab2). Participants listened to the same stimuli in both tasks. The only difference was the attentional focus (Fig. 1).

2.4. *EEG data acquisition and analysis*

EEG data were recorded using a 64-channel BioSemi Active Two system (BioSemi, Amsterdam, The Netherlands) in a continuous mode at a digitization rate of 512 Hz and stored on hard disk for later analysis. Data were re-referenced to the average of the left and right mastoids using Letswave 6 (https://www.letswave.org/), filtered with a .1–30 Hz band-pass filter (zero phase shift Butterworth, order 4), and then

[2] The conditions of our ethics approval do not permit sharing of the data supporting the conclusions in this study with any individual outside the author team under any circumstances. These restrictions also apply to the study materials, since they include self-voice recordings. However, the script of the experimental task and the code used in data analysis are available at https://osf.io/dz68e/?view_only=583d25eb5683412ba49ea934cb097ab2.

segmented into epochs time-locked to word onset, from −200 ms to 1000 ms, with a 200 ms baseline (−200 to 0 ms). Both vertical and horizontal eye movements were removed using the method of Gratton, Coles, and Donchin (Gratton, Coles, & Donchin, 1983). After segmentation, epochs with amplitudes exceeding ±100 μV were automatically rejected from further analysis. After artifact rejection, at least 70% of the trials per condition per subject entered the analyses. Individual ERPs were averaged separately for each condition.

Following prior studies (Pinheiro, Rezaii, Nestor, et al., 2016), the mean amplitudes of the N1 and P2 were measured in the time windows of 120–190 ms (N1) and 220–290 ms (P2). We also observed a prominent negativity occurring after 400 ms post-word onset (N400; e.g., Maess, Herrmann, Hahne, Nakamura, & Friederici, 2006; Perrin & García-Larrea, 2003; Strauß, Kotz, & Obleser, 2013). The amplitude of the N400 was measured in the time window of 400–600 ms. Long lasting late positive effects (LPP) were analyzed in two post-stimulus latency windows to capture differences between the conditions more accurately (Conde et al., 2022): early LPP phase – 601–800 ms; late LPP phase – 801–1000 ms. ERP amplitude was quantified as the mean voltage within these segments. Mean amplitudes were extracted from left medial (FC3, C3, CP3), right medial (FC4, C4, CP4), and midline (FCz, Cz, CPz) electrodes, following prior studies (Pinheiro, Rezaii, Nestor, et al., 2016).

No part of the study analyses was pre-registered prior to the research being conducted.

2.5. *Statistical analysis*

Linear mixed effects models were built to fit the ERP amplitudes per participant and condition in each time-window of interest, using the lmer4 (Bates, Mächler, Bolker, & Walker, 2015) and lmerTest (Kuznetsova, Brockhoff, & Christensen, 2017) packages in the R environment (R3.4.3. GUI 1.70). The models tested the fixed effects of speaker identity (SV, OV), speech emotion (neutral, positive, negative), and task (focus on speaker identity, focus on speech emotion). To simplify the description of results, the factors "speaker identity" and "speech emotion" will be designated from now on as "identity" and "emotion" respectively. Random effects terms included intercepts varying among subjects. The Satterthwaite approximation was applied to the fitted models (Luke, 2017). The *anova* function from lmerTest was used to provide *p*-values for each factor, calculated from the *F* statistic. Package sjstats (Lüdecke, 2019) was used to obtain 95% confidence intervals (see https://osf.io/dz68e/?view_only=583d25eb5683412ba49ea934cb097ab2).

Table 3 – Mean percentage of correct and unsure responses in the identity recognition task.

Speaker Identity	Speech Emotion *M* (*SD*)	Correct Responses *M* (*SD*)	Unsure Responses *M* (*SD*)
Self	*Negative*	96.19 (6.45)	2.10 (4.42)
	Neutral	96.48 (5.41)	1.52 (2.63)
	Positive	96.00 (7.54)	1.71 (3.73)
Other	*Negative*	94.67 (14.60)	1.33 (2.63)
	Neutral	94.57 (15.14)	1.33 (3.52)
	Positive	96.10 (11.46)	.76 (1.80)

Note. M = Mean; *SD* = Standard Deviation.

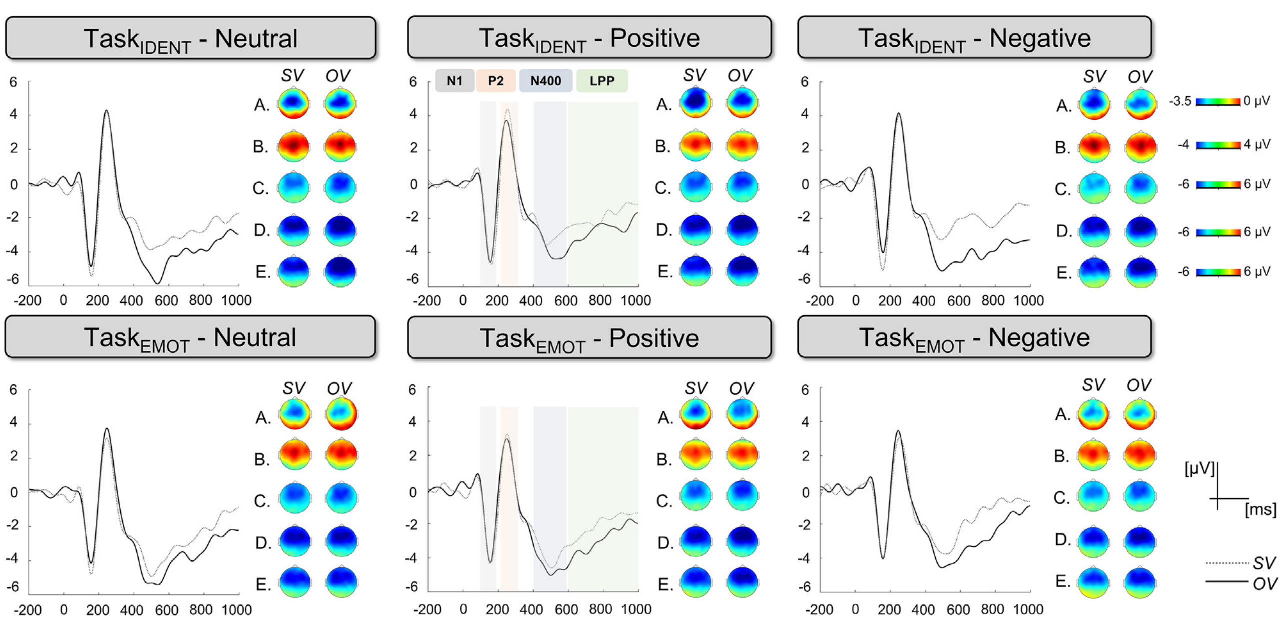

Fig. 2 – Grand average waveforms contrasting neutral, positive, and negative words spoken in one's own and other voice conditions, as a function of task instructions. Topographic maps show voltage distribution in: [A] 120–190 ms (N1); [B] 220–290 ms (P2); [C] 400–600 ms (N400); [D] 601–800 ms (early LPP); and [E] 801–1000 ms (late LPP) latency windows. Note: Task$_{IDENT}$ = focus on speaker identity; Task$_{EMOT}$ = focus on speech emotion; SV = self-voice; OV = other-voice.

3. Results

3.1. Behavioral results

Identity recognition: Self- and other voices were recognized with comparable accuracy (Table 3). Recognition accuracy was not modulated by identity ($F(1, 50) = .834$, $p = .363$) or emotion ($F(2, 150) = .100$, $p = .905$). The number of unsure responses was also not affected by identity ($F(1, 150) = 2.786$, $p = .097$) or emotion ($F(2, 150) = .529$, $p = .590$).

Emotion categorization: The evaluation of the emotional quality of spoken words was affected by emotion ($F(2, 180) = 23.237$, $p < .001$) but not by identity ($F(1, 180) < .0005$, $p > .999$): both negative ($B = .024$, $SE = .007$, $t = 3.259$, $p = .001$, 95% CI: [.010, .038]) and positive words ($B = .036$, $SE = .007$, $t = 4.899$, $p < .001$, 95% CI: [.022, .050]) were more accurately categorized than neutral words.

3.2. ERP results

Figs. 2–4 illustrate grand average waveforms for spoken words varying in speaker identity (self, other) and speech emotion (neutral, positive, and negative) as a function of task instructions (attention focus on speaker identity *vs.* speech emotion). The four ERP components under analysis are highlighted: N1 (sensory processing), P2 (rapid salience detection), N400 (semantic analysis), and LPP (sustained stimulus evaluation).

In the following sections, we first describe significant results that were independent of task manipulations and that inform on whether and when self-voice prioritization is boosted by emotion. We then focus on main effects or interactions involving the task factor to examine whether and how self-voice prioritization is affected by attention.

N1. Confirming an early self-relevance effect, the N1 amplitudes were increased (i.e., more negative) in response to SV compared to OV words ($B = -.613$, $SE = .134$, $t = -4.565$, $p < .001$, 95% CI: [−.350, −.875]), reflected in a main effect of identity ($F(1, 3210) = 37.879$, $p < .001$).

Notably, task instructions modulated auditory processing at this early sensory stage. Identity interacted with task ($F(1, 3210) = 19.055$, $p < .001$): the N1 was more negative in response to SV compared to OV words when attention was focused on

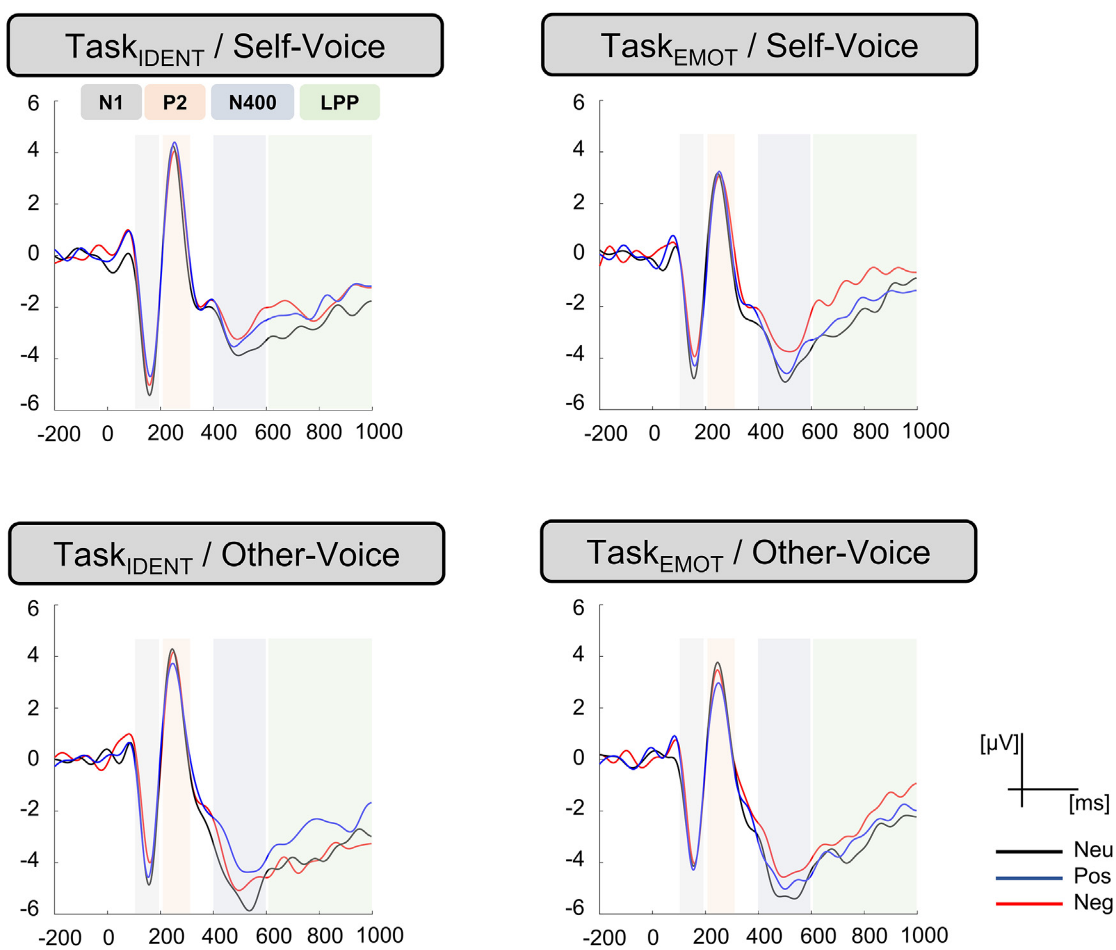

Fig. 3 – ERP responses to neutral vs. emotional spoken words in each task. Note: Task$_{IDENT}$ = focus on speaker identity; Task$_{EMOT}$ = focus on speech emotion; Neu = neutral; Pos = positive; Neg = negative.

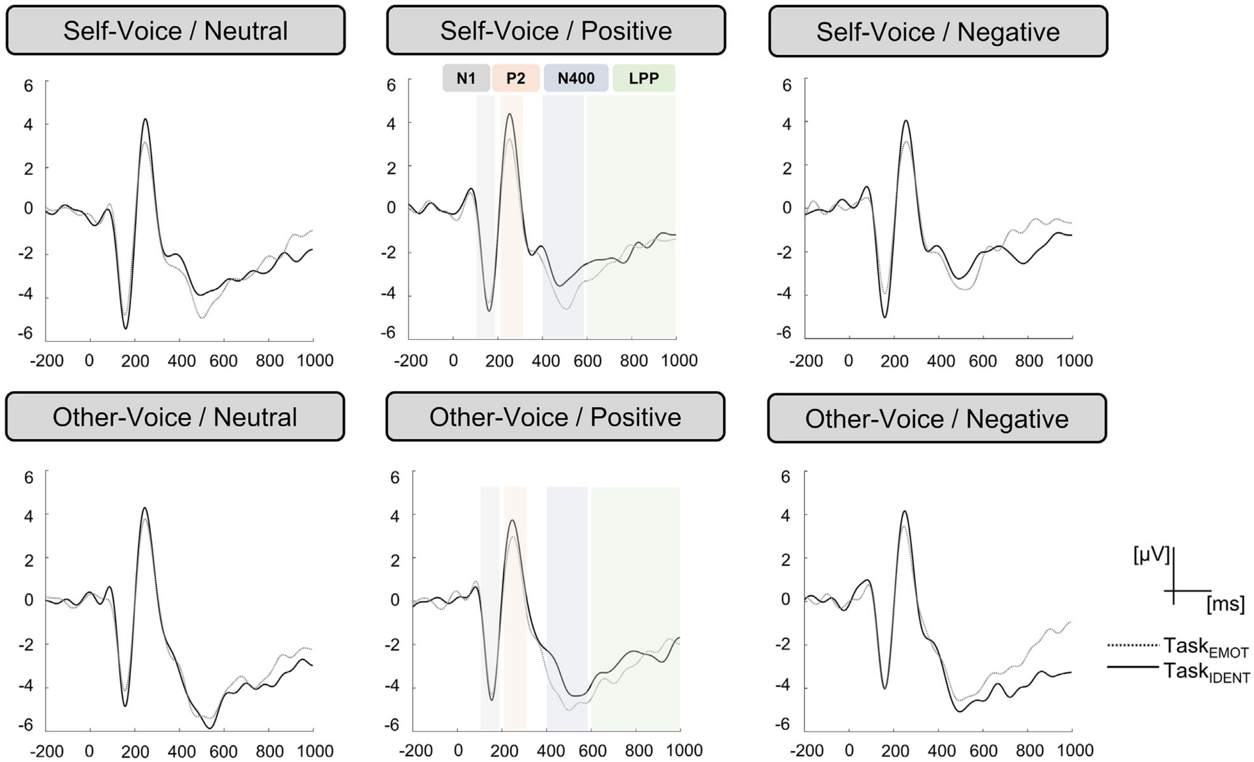

Fig. 4 – ERP responses to spoken words in the identity recognition vs. emotion categorization tasks. Note: Task$_{IDENT}$ = focus on speaker identity; Task$_{EMOT}$ = focus on speech emotion.

speaker identity ($B = -.546$, SE = .190, $t = -2.875$, $p = .004$, 95% CI: [−.917, −.174]). Task also interacted with emotion ($F(2, 3210) = 3.6701$, $p = .026$): the N1 was more negative in response to positive (*vs.* neutral) words when attention was focused on speech emotion ($B = -.403$, SE = .190, $t = -2.122$, $p = .034$, 95% CI: [−.775, −.031]).

Therefore, task modulated the sensory processing of the voice as a function of the most task-relevant acoustic features: the self-voice or emotional (positive) speech in the identity recognition and emotion categorization tasks, respectively.

P2. The P2 was sensitive to both identity ($F(1, 3210) = 6.763$, $p = .009$) and emotion ($F(2, 3210) = 3.045$, $p = .048$) manipulations. An interaction between the two factors ($F(2, 3210) = 12.261$, $p < .001$) confirmed our hypothesis: the P2 response was increased for positive (*vs.* neutral) SV words ($B = .607$, SE = .220, $t = 2.762$, $p = .006$, 95% CI: [.176, 1.038]).

The P2 response was also sensitive to task ($F(1, 3210) = 41.719$, $p < .001$): its amplitude was globally increased when attention was focused on speaker identity compared to speech emotion ($B = .334$, SE = .155, $t = 2.152$, $p = .031$, 95% CI: [.639, .030]). Task did not interact with any other factors ($p > .05$). Therefore, salience detection was facilitated when attention was focused on speaker identity and, specifically, for positive words spoken in one's own voice.

N400. A main effect of identity ($F(1, 3210) = 112.507$, $p < .001$) revealed that the semantic processing of word meaning was facilitated in the case of words spoken in one's own voice: the N400 was less negative in response to SV compared to OV words ($B = .591$, SE = .178, $t = 3.326$, $p < .001$, 95% CI: [.243, .940]).

A significant interaction between task and emotion ($F(2, 3210) = 4.639$, $p = .010$) indicated that the processing of positive words was also facilitated compared to neutral words, when attention was focused on speaker identity: the N400 was less negative for words of positive valence ($B = .690$, SE = .251, $t = 2.745$, $p = .006$, 95% CI: [.197, 1.183]). There were no other significant interaction effects ($p > .05$). In other words, semantic analysis of words was facilitated when they were uttered by the self-voice, and when their valence was positive if task instructions required attention focus on speaker identity.

3.3. *LPP*

Early phase (601–800 ms). Identity ($F(1, 3210) = 130.796$, $p < .001$) significantly modulated later stages of auditory processing, reflected in the LPP (601–800 ms). The LPP was larger (i.e., more positive) for SV compared to OV words ($B = .645$, SE = .210, $t = 3.066$, $p = .002$, 95% CI: [.233, 1.057]). A significant interaction between identity and emotion ($F(2, 3210) = 8.935$, $p < .001$) revealed that the LPP was specifically larger in response to negative compared to neutral words when spoken in one's own voice ($B = .766$, SE = .298, $t = 2.573$, $p = .010$, 95% CI: [.183, 1.349]).

Task instructions also shaped later stages of speech processing, shown by interactive effects of task and emotion ($F(2, 3210) = 10.996$, $p < .001$): the LPP was increased in response to negative compared to neutral words ($B = .713$, SE = .298,

$t = 2.395$, $p = .017$, 95% CI: [1.296, .129]) when attention was focused on speech emotion.

Late phase (801–1000 ms): We confirmed similar effects in a later LPP phase (801–1000 ms). LPP amplitudes were larger in response to SV compared to OV words ($B = .697$, $SE = .222$, $t = 3.143$, $p = .002$, 95% CI: [.262, 1.131]), reflected in a main effect of identity ($F(1, 3210) = 87.298$, $p < .001$).

Task focus also modulated the late LPP phase. A task by emotion interaction ($F(2, 3210) = 13.785$, $p < .001$) indicated that when attention was focused on speech emotion, the LPP was increased in response to negative compared to neutral words ($B = 1.159$, $SE = .314$, $t = 3.697$, $p < .001$, 95% CI: [1.774, .545]).

Therefore, words spoken in one's own voice elicited enhanced sustained elaborative processing. In an early LPP phase, this enhancement occurred specifically for negative words. Focusing attention on speech emotion led to a negativity bias in speech perception irrespective of speaker identity.

4. Discussion

The goal of the current study was to examine how self-voice processing is affected by emotion and task. Therefore, we manipulated the degree of voluntary attention directed to speaker or speech information. Overall, we confirmed that the self-voice is prioritized in early sensory processing stages but interacts with emotion after 200 ms post-stimulus onset (Pinheiro, Rezaii, Nestor, et al., 2016). Moreover, our findings revealed that the self-voice is perceived differently when attention is focused on speaker identity compared to speech emotion within ~600 ms post-stimulus onset. They also reveal that attention does not affect the interaction of self-relevance and emotion, reflected in P2 and LPP amplitude modulations. We discuss these findings in the following sections.

4.1. *Self-voice prioritization is boosted by emotion after 200 ms*

The voice is critical in representing our self (McGettigan, 2015). One's own voice therefore represents a particularly salient sound category (Conde et al., 2015; Conde et al., 2016, 2018). We confirmed that the self-voice has a privileged status in the auditory system (Pinheiro, Rezaii, Nestor, et al., 2016) and is perceptually prioritized in speech processing. We observed a rapid discrimination of self- and other voices within 100 ms post-stimulus onset, replicating prior studies (Conde et al., 2015; Graux et al., 2013; Pinheiro, Rezaii, Nestor, et al., 2016). The main effect of speaker identity indicates that self-voice prioritization is independent of emotion at least in an early processing stage (N1). Words spoken in one's own voice were also associated with facilitated semantic analysis (decreased N400) and enhanced elaborative processing (increased LPP). Notably, the ERP differences between self-generated words and words uttered by an unfamiliar speaker occurred despite similar accuracy in self- and other-voice recognition (Hughes & Nicholson, 2010; Pinheiro, Farinha-Fernandes, et al., 2019; Rosa et al., 2008). These effects cannot be explained by word differences as they were identical in the self-voice and other-voice conditions and in both tasks.

Even though the salience of a stimulus increases when it is self-relevant (e.g., Conde et al., 2015, 2018), this was not the case for all self-relevant stimuli in subsequent auditory processing stages: interactive effects of self-relevance and emotion revealed an early (P2) positivity bias (Chen et al., 2014; Fields & Kuperberg, 2015; Watson et al., 2007) and a late (LPP) negativity bias in self-voice perception. One's own voice may be ascribed greater emotional salience than an unfamiliar voice (Conde et al., 2015, 2018); however, these findings suggest that the self-voice is more than an emotionally salient signal. The self-positivity bias shown in the P2 agrees with prior studies indicating that positive (but not negative) information tends to be more easily related to personal characteristics (Fields & Kuperberg, 2015; Mezulis, Abramson, Hyde, & Hankin, 2004; Watson et al., 2007). We note though that this effect occurred before the recognition point of positive words in the pilot study (~469.55 ms; see Supplement). Notwithstanding, there is evidence that convergence on specific word candidates is largely complete within 150–200 ms of word onset, well before word offset (Klimovich-Gray et al., 2019; Marslen-Wilson & Tyler, 1975). As participants had to read the words in the voice recording session that took place before the EEG experiment, it is possible that the recognition point for these participants was shifted earlier in time. Another possibility is that positive words were better remembered and, therefore, salience detection was facilitated for these words when spoken in one's own voice. The self-negativity bias reflected in the early LPP phase implies that listeners automatically engage in sustained monitoring of negative words uttered in the self-voice. Since the LPP is often linked to higher-order cognitive processes such as sustained attention, meaning evaluation or access to memory representations (Paulmann, Bleichner, & Kotz, 2013), this effect could potentially indicate an attempt to determine how the words fit with the participant's self-concept as they are typically identified as unrelated to personal characteristics (Mezulis et al., 2004; Pahl & Eiser, 2005).

Although previous studies suggest that non-verbal cues take precedence over competing verbal information by capturing more attention (Jacob, Brück, Domin, Lotze, & Wildgruber, 2014), our findings imply a more interdependent relationship between identity and emotion in the auditory modality (Pinheiro, Rezaii, et al., 2017; Pinheiro, Rezaii, Nestor, et al., 2016): when attention is not explicitly drawn toward non-verbal (identity) cues, the P2 and LPP are still influenced by non-verbal information, and vice-versa. Despite anatomical differences between speech and speaker recognition brain networks (e.g., rostral-caudal auditory networks distinction – Jasmin, Lima, & Scott, 2019; Scott, 2019), it is critical that they interact quickly as when there is speech, there is always a talking voice (Scott, 2019). Our findings indirectly support such a claim.

4.2. *Attention enhances self-voice prioritization but does not affect the interaction of self-relevance and emotion in speech processing*

Task effects revealed that auditory representations of voices are boosted for task-relevant dimensions relatively early, reflected in N1 amplitude modulations. That is, attention focus shapes how much weight is given to specific acoustic properties. We confirmed enhanced sensory processing of the self-voice in the identity recognition task, and of emotional speech in the emotion categorization task. This finding indicates top-down attentional amplification of sensory processing as a function of task instructions (Hillyard, Vogel, & Luck, 1998). Accordingly, in previous studies the N1 was enhanced when participants were, for example, instructed to focus their attention to pain cues in the voice compared to passively listening to vocal stimuli (Meng, Shen, Li, & Peng, 2019). Whereas attention to speaker identity affected more strongly ERP responses occurring within 600 ms post-word onset (N1, P2, and N400), attention to speech emotion modulated late ERP components (LPP). Irrespective of speech valence and speaker identity, the P2 was generally increased when the task involved speaker recognition compared to emotion categorization. We also observed that the semantic analysis of positive (*vs.* neutral) spoken words, reflected in N400 amplitude modulations, was facilitated if task instructions required self-other voice discrimination. This reveals that even if the task does not explicitly require emotion categorization, spoken words are differentially processed as a function of their valence, confirming that emotion effects are automatic to an important extent.

However, in late processing stages, the LPP was larger for emotional (*vs.* neutral) spoken words if the task involved emotion categorization compared to identity recognition. Previous studies revealed that task manipulations may enhance or attenuate the LPP emotion effect (e.g., Fischler & Bradley, 2006; Holt, Lynn, & Kuperberg, 2009; Naumann, Maier, Diedrich, Becker, & Bartussek, 1997), which we confirm here. Enhancement effects in this later latency window suggest that emotionally relevant words need to be more thoroughly processed and engage sustained attention (Paulmann et al., 2013).

Notably, task instructions affected how emotion and identity were processed within the first 200 ms post-speech onset but not their interaction. The observation that identity and emotion interacted irrespective of attention suggests that the analysis of vocal features (in this case, speaker identity) is not independent of verbal analysis. As such, there may not be a complete functional dissociation between speech and speaker processing (Scott, 2019).

These findings have clinical relevance as they shed light on the neurocognitive mechanisms underlying self-voice processing and monitoring, thought to be dysfunctional in auditory verbal hallucinations (Pinheiro et al., 2018; Pinheiro, Farinha-Fernandes, et al., 2019; Pinheiro, Rezaii, et al., 2017; Pinheiro, Schwartze, Amorim, et al., 2020; Pinheiro, Schwartze, & Kotz, 2020), a cardinal symptom of psychosis (Waters et al., 2012) and whose neural correlates and origin remain elusive.

5. Conclusion

Together, the current findings suggest that the self-voice is prioritized in auditory perception over the voices of others, i.e., it is more than an emotionally salient voice. Interactive effects of self-relevance and emotion revealed an early self-positivity bias in the P2 and a later self-negativity bias in the LPP, reflecting, respectively, enhanced salience detection in positive words and increased elaborative processing of negative words spoken in one's own voice.

Task instructions shaped speech processing already at early sensory stages. Whereas attention to speaker identity affected more strongly ERP responses up until 600 ms post-word onset (N1, P2, N400), attention to the emotional valence of speech changed late ERP components (>600 ms; LPP). However, task instructions did not affect the interaction between identity and emotion, confirming involuntary processing of salient signals (i.e., self-relevance [an owned voice] and emotion) even in situations where attention is deliberately directed away from those cues.

These findings highlight the need to consider the role of self-relevance and attention in current models of voice and speech perception.

Credit author statement

APP and SK designed the study and wrote the protocol. APP, JS, and SK managed the literature searches. APP and JS collected and analyzed the data. APP, JS, and MSR undertook the statistical analyses. APP wrote the first draft of the manuscript. All authors contributed to and have approved the final manuscript.

Data availability

The authors do not have permission to share data.

Acknowledgments

This work was supported by the Portuguese Science Foundation (Fundação para a Ciência e a Tecnologia [FCT]; grant number PTDC/MHC-PCN/0101/2014, awarded to APP) and BIAL Foundation (grant number BIAL 146/2020, awarded to APP).

The Authors would like to thank all the participants who collaborated in the study, and particularly Dr. Francisco Gutierrez and Maria Amorim for collaboration in data collection, and António Farinha-Fernandes for collaboration in stimulus selection.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cortex.2022.10.006.

REFERENCES

Allen, P., Freeman, D., Johns, L., & McGuire, P. (2006). Misattribution of self-generated speech in relation to hallucinatory proneness and delusional ideation in healthy volunteers. *Schizophrenia Research, 84*(2–3), 281–288. https://doi.org/10.1016/j.schres.2006.01.021

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Belin, P., Fecteau, S., & Bédard, C. (2004). Thinking the voice: Neural correlates of voice perception. *Trends in Cognitive Sciences, 8*(3), 129–135. https://doi.org/10.1016/j.tics.2004.01.008

Boersma, P., & Weenink, D. (2013). *Praat: Doing phonetics by computer [Computer program]. Version 5.3.53* http://www.praat.org/.

Candini, M., Zamagni, E., Nuzzo, A., Ruotolo, F., Iachini, T., & Frassinetti, F. (2014). Who is speaking? Implicit and explicit self and other voice recognition. *Brain and Cognition, 92*, 112–117. https://doi.org/10.1016/j.bandc.2014.10.001

Chen, Y., Zhong, Y., Zhou, H., Zhang, S., Tan, Q., & Fan, W. (2014). Evidence for implicit self-positivity bias: An event-related brain potential study. *Experimental Brain Research, 232*, 985–994. https://doi.org/10.1007/s00221-013-3810-z

Conde, T., Correia, A. I., Roberto, M. S., Scott, S. K., Lima, C. F., & Pinheiro, A. P. (2022). The time course of emotional authenticity detection in nonverbal vocalizations. *Cortex, 151*, 116–132. https://doi.org/10.1016/j.cortex.2022.02.016

Conde, T., Gonçalves, Ó. F., & Pinheiro, A. P. (2015). Paying attention to my voice or yours: An ERP study with words. *Biological Psychology, 111*, 40–52. https://doi.org/10.1016/j.biopsycho.2015.07.014

Conde, T., Gonçalves, Ó. F., & Pinheiro, A. P. (2016). The effects of stimulus complexity on the preattentive processing of self-generated and nonself voices: An ERP study. *Cognitive, Affective, & Behavioral Neuroscience, 16*(1), 106–123. https://doi.org/10.3758/s13415-015-0376-1

Conde, T., Gonçalves, Ó. F., & Pinheiro, A. P. (2018). Stimulus complexity matters when you hear your own voice: Attention effects on self-generated voice processing. *International Journal of Psychophysiology, 133*, 66–78. https://doi.org/10.1016/j.ijpsycho.2018.08.007

Ferrari, V., Codispoti, M., Cardinale, R., & Bradley, M. M. (2008). Directed and motivated attention during processing of natural scenes. *Journal of Cognitive Neuroscience, 20*(10), 1753–1761. https://doi.org/10.1162/jocn.2008.20121

Fields, E. C., & Kuperberg, G. R. (2015). Loving yourself more than your neighbor: ERPs reveal online effects of a self-positivity bias. *Social Cognitive and Affective Neuroscience, 10*(9), 1202–1209. https://doi.org/10.1093/scan/nsv004

Fischler, I., & Bradley, M. (2006). Event-related potential studies of language and emotion: Words, phrases, and task effects. *Progress in Brain Research, 156*, 185–203. https://doi.org/10.1016/S0079-6123(06)56009-1

Foucart, A., Garcia, X., Ayguasanosa, M., Thierry, G., Martin, C., & Costa, A. (2015). Does the speaker matter? Online processing of semantic and pragmatic information in L2 speech comprehension. *Neuropsychologia, 75*, 291–303. https://doi.org/10.1016/j.neuropsychologia.2015.06.027

Gratton, G., Coles, M. G. H., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology, 55*(4), 468–484. https://doi.org/10.1016/0013-4694(83)90135-9

Graux, J., Gomot, M., Roux, S., Bonnet-Brilhault, F., Camus, V., & Bruneau, N. (2013). My voice or yours? An electrophysiological study. *Brain Topography, 26*, 72–82. https://doi.org/10.1007/s10548-012-0233-2

Gray, H. M., Ambady, N., Lowenthal, W. T., & Deldin, P. (2004). P300 as an index of attention to self-relevant stimuli. *Journal of Experimental Social Psychology, 40*(2), 216–224. https://doi.org/10.1016/S0022-1031(03)00092-1

Haider, M., Spong, P., & Lindsley, D. B. (1964). Attention, vigilance, and cortical evoked-potentials in humans. *Science, 145*(3628), 180–182. https://doi.org/10.1126/science.145.3628.180

Hajcak, G., Dunning, J. P., & Foti, D. (2009). Motivated and controlled attention to emotion: Time-course of the late positive potential. *Clinical Neurophysiology, 120*(3), 505–510. https://doi.org/10.1016/j.clinph.2008.11.028

Hajcak, G., Macnamara, A., & Olvet, D. M. (2010). Event-related potentials, emotion, and emotion regulation: An integrative review. *Developmental Neuropsychology, 35*(2), 129–155. https://doi.org/10.1080/87565640903526504

Hatzidaki, A., Baus, C., & Costa, A. (2015). The way you say it, the way I feel it: Emotional word processing in accented speech. *Frontiers in Psychology, 6*, 351. https://doi.org/10.3389/fpsyg.2015.00351

Heine, S. J., Lehman, D. R., Markus, H. R., & Kitayama, S. (1999). Is there a universal need for positive self-regard? *Psychological Review, 106*(4), 766–794. https://doi.org/10.1037/0033-295X.106.4.766

Herbert, C., Junghofer, M., & Kissler, J. (2008). Event related potentials to emotional adjectives during reading. *Psychophysiology, 45*(3), 487–498. https://doi.org/10.1111/j.1469-8986.2007.00638.x

Hettich, D. T., Bolinger, E., Matuz, T., Birbaumer, N., Rosenstiel, W., & Spüler, M. (2016). EEG responses to auditory stimuli for automatic affect recognition. *Frontiers in Neuroscience, 10*, 244. https://doi.org/10.3389/fnins.2016.00244

Hillyard, S. A., Vogel, E. K., & Luck, S. J. (1998). Sensory gain control (amplification) as a mechanism of selective attention: Electrophysiological and neuroimaging evidence. *Philosophical Transactions of the Royal Society B: Biological Sciences, 353*(1373), 1257–1270. https://doi.org/10.1098/rstb.1998.0281

Holt, D. J., Lynn, S. K., & Kuperberg, G. R. (2009). Neurophysiological correlates of comprehending emotional meaning in context. *Journal of Cognitive Neuroscience, 21*(11), 2245–2262. https://doi.org/10.1162/jocn.2008.21151

Hughes, S. M., & Nicholson, S. E. (2010). The processing of auditory and visual recognition of self-stimuli. *Consciousness and Cognition, 19*(4), 1124–1134. https://doi.org/10.1016/j.concog.2010.03.001

Iannotti, G. R., Orepic, P., Brunet, D., Koenig, T., Alcoba-Banqueri, S., Garin, D. F. A., … Michel, C. M. (2021). EEG spatiotemporal patterns underlying self-other voice discrimination. *Cerebral Cortex, 32*(9), 1978–1992. https://doi.org/10.1093/cercor/bhab329

Jacob, H., Brück, C., Domin, M., Lotze, M., & Wildgruber, D. (2014). I can't keep your face and voice out of my head: Neural correlates of an attentional bias toward nonverbal emotional cues. *Cerebral Cortex, 24*(6), 1460–1473. https://doi.org/10.1093/cercor/bhs417

Jasmin, K., Lima, C. F., & Scott, S. K. (2019). Understanding rostral–caudal auditory cortex contributions to auditory perception. *Nature Reviews Neuroscience, 20*, 425–434. https://doi.org/10.1038/s41583-019-0160-2

Kanske, P., & Kotz, S. A. (2007). Concreteness in emotional words: ERP evidence from a hemifield study. *Brain Research, 1148*, 138–148. https://doi.org/10.1016/j.brainres.2007.02.044

Kaplan, J. T., Aziz-Zadeh, L., Uddin, L. Q., & Iacoboni, M. (2008). The self across the senses: An fMRI study of self-face and self-voice recognition. *Social Cognitive and Affective Neuroscience, 3*(3), 218–223. https://doi.org/10.1093/scan/nsn014

Klimovich-Gray, A., Tyler, L. K., Randall, B., Kocagoncu, E., Devereux, B., & Marslen-Wilson, W. D. (2019). Balancing prediction and sensory input in speech comprehension: The spatiotemporal dynamics of word recognition in context. *Journal of Neuroscience, 39*(3), 519–527. https://doi.org/10.1523/JNEUROSCI.3573-17.2018

Kousta, S.-T., Vinson, D. P., & Vigliocco, G. (2009). Emotion words, regardless of polarity, have a processing advantage over neutral words. *Cognition, 112*(3), 473–481. https://doi.org/10.1016/j.cognition.2009.06.007

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models.

Journal of Statistical Software, 82(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lange, K. (2013). The ups and downs of temporal orienting: A review of auditory temporal orienting studies and a model associating the heterogeneous findings on the auditory N1 with opposite effects of attention and prediction. *Frontiers in Human Neuroscience, 7*, 263. https://doi.org/10.3389/fnhum.2013.00263

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. https://doi.org/10.1097/WNR.0b013e32834ea757

Luck, S. J., Woodman, G. F., & Vogel, E. K. (2000). Event-related potential studies of attention. *Trends in Cognitive Sciences, 4*(11), 432–440. https://doi.org/10.1016/S1364-6613(00)01545-X

Lüdecke, D. (2019). Package 'sjstats' *(0.17.3). Statistical functions for regression models*.

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods, 49*, 1494–1502. https://doi.org/10.3758/s13428-016-0809-y

Maess, B., Herrmann, C. S., Hahne, A., Nakamura, A., & Friederici, A. D. (2006). Localizing the distributed language network responsible for the N400 measured by MEG during auditory sentence processing. *Brain Research, 1096*(1), 163–172. https://doi.org/10.1016/j.brainres.2006.04.037

Marslen-Wilson, W., & Tyler, L. K. (1975). Processing structure of sentence perception. *Nature, 257*(5529), 784–786. https://doi.org/10.1038/257784a0

Maurer, D., & Landis, T. (1990). Role of bone conduction in the self-perception of speech. *Folia Phoniatrica, 42*(5), 226–229. https://doi.org/10.1159/000266070

McGettigan, C. (2015). The social life of voices: Studying the neural bases for the expression and perception of the self and others during spoken communication. *Frontiers in Human Neuroscience, 9*, 129. https://doi.org/10.3389/fnhum.2015.00129

Meng, J., Shen, L., Li, Z., & Peng, W. (2019). Top-down attention modulation on the perception of others' vocal pain: An event-related potential study. *Neuropsychologia, 133*, Article 107177. https://doi.org/10.1016/j.neuropsychologia.2019.107177

Mezulis, A. H., Abramson, L. Y., Hyde, J. S., & Hankin, B. L. (2004). Is there a universal positivity bias in attributions? A meta-analytic review of individual, developmental, and cultural differences in the self-serving attributional bias. *Psychological Bulletin, 130*(5), 711–747. https://doi.org/10.1037/0033-2909.130.5.711

Nakamura, K., Kawashima, R., Sugiura, M., Kato, T., Nakamura, A., Hatano, K., … Kojima, S. (2001). Neural substrates for recognition of familiar voices: A PET study. *Neuropsychologia, 39*(10), 1047–1054. https://doi.org/10.1016/S0028-3932(01)00037-9

Naumann, E., Maier, S., Diedrich, O., Becker, G., & Bartussek, D. (1997). Structural, semantic, and emotion-focussed processing at neutral and negative nouns: Event-related potential correlates. *Journal of Psychophysiology, 11*(2), 158–172.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh Inventory. *Neuropsychologia, 9*(1), 97–113. https://doi.org/10.1016/0028-3932(71)90067-4

Pahl, S., & Eiser, J. R. (2005). Valence, comparison focus and self-positivity biases. *Experimental Psychology, 52*(4), 303–310. https://doi.org/10.1027/1618-3169.52.4.303

Paulmann, S., Bleichner, M., & Kotz, S. A. (2013). Valence, arousal, and task effects in emotional prosody processing. *Frontiers in Psychology, 4*, 345. https://doi.org/10.3389/fpsyg.2013.00345

Paulmann, S., & Kotz, S. A. (2008). Early emotional prosody perception based on different speaker voices. *Neuroreport, 19*(2), 209–213. https://doi.org/10.1097/WNR.0b013e3282f454db

Payne, B., Lavan, N., Knight, S., & McGettigan, C. (2021). Perceptual prioritization of self-associated voices. *British Journal of Psychology, 112*(3), 585–610. https://doi.org/10.1111/bjop.12479

Perrin, F., & García-Larrea, L. (2003). Modulation of the N400 potential during auditory phonological/semantic interaction. *Cognitive Brain Research, 17*(1), 36–47. https://doi.org/10.1016/S0926-6410(03)00078-8

Phan, K. L., Taylor, S. F., Welsh, R. C., Ho, S. H., Britton, J. C., & Liberzon, I. (2004). Neural correlates of individual ratings of emotional salience: A trial-related fMRI study. *Neuroimage, 21*(2), 768–780. https://doi.org/10.1016/j.neuroimage.2003.09.072

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

Pinheiro, A. P., Barros, C., & Pedrosa, J. (2015). Salience in a social landscape: Electrophysiological effects of task-irrelevant and infrequent vocal change. *Social Cognitive and Affective Neuroscience, 11*(1), 127–139. https://doi.org/10.1093/scan/nsv103

Pinheiro, A. P., Farinha-Fernandes, A., Roberto, M. S., & Kotz, S. A. (2019). Self-voice perception and its relationship with hallucination predisposition. *Cognitive Neuropsychiatry, 24*(4), 237–255. https://doi.org/10.1080/13546805.2019.1621159

Pinheiro, A. P., Lima, D., Albuquerque, P. B., Anikin, A., & Lima, C. F. (2019). Spatial location and emotion modulate voice perception. *Cognition & Emotion, 33*(8), 1577–1586. https://doi.org/10.1080/02699931.2019.1586647

Pinheiro, A. P., Rezaii, N., Nestor, P. G., Rauber, A., Spencer, K. M., & Niznikiewicz, M. (2016). Did you or I say pretty, rude or brief? An ERP study of the effects of speaker's identity on emotional word processing. *Brain and Language, 153–154*, 38–49. https://doi.org/10.1016/j.bandl.2015.12.003

Pinheiro, A. P., Rezaii, N., Rauber, A., Nestor, P. G., Spencer, K. M., & Niznikiewicz, M. (2017). Emotional self–other voice processing in schizophrenia and its relationship with hallucinations: ERP evidence. *Psychophysiology, 54*(9), 1252–1265. https://doi.org/10.1111/psyp.12880

Pinheiro, A. P., Rezaii, N., Rauber, A., & Niznikiewicz, M. (2016b). Is this my voice or yours? The role of emotion and acoustic quality in self-other voice discrimination in schizophrenia. *Cognitive Neuropsychiatry, 21*(4), 335–353. https://doi.org/10.1080/13546805.2016.1208611

Pinheiro, A. P., Schwartze, M., Amorim, M., Coentre, R., Levy, P., & Kotz, S. A. (2020). Changes in motor preparation affect the sensory consequences of voice production in voice hearers. *Neuropsychologia, 146*, Article 107531. https://doi.org/10.1016/j.neuropsychologia.2020.107531

Pinheiro, A. P., Schwartze, M., & Kotz, S. A. (2018). Voice-selective prediction alterations in nonclinical voice hearers. *Scientific Reports, 8*(1), Article 14717. https://doi.org/10.1038/s41598-018-32614-9

Pinheiro, A. P., Schwartze, M., & Kotz, S. A. (2020). Cerebellar circuitry and auditory verbal hallucinations: An integrative synthesis and perspective. *Neuroscience and Biobehavioral Reviews, 118*, 485–503. https://doi.org/10.1016/j.neubiorev.2020.08.004

Rosa, C., Lassonde, M., Pinard, C., Keenan, J. P., & Belin, P. (2008). Investigations of hemispheric specialization of self-voice recognition. *Brain and Cognition, 68*(2), 204–214. https://doi.org/10.1016/j.bandc.2008.04.007

Sander, D., Grandjean, D., Pourtois, G., Schwartz, S., Seghier, M. L., Scherer, K. R., & Vuilleumier, P. (2005). Emotion and attention interactions in social cognition: Brain regions involved in processing anger prosody. *Neuroimage, 28*(4), 848–858. https://doi.org/10.1016/j.neuroimage.2005.06.023

Schäfer, S., Wentura, D., & Frings, C. (2017). Distinctiveness effects in self-prioritization. *Visual Cognition, 25*(1–3), 399–411. https://doi.org/10.1080/13506285.2017.1346739

Schäfer, S., Wentura, D., & Frings, C. (2020). Creating a network of importance: The particular effects of self-relevance on stimulus processing. *Attention, Perception, & Psychophysics, 82*(7), 3750–3766. https://doi.org/10.3758/s13414-020-02070-7

Scherer, K. R. (2001). Appraisal considered as a process of multilevel sequential checking. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Series in affective science. Appraisal processes in emotion: Theory, methods, research* (pp. 92–120). Oxford: Oxford University Press.

Schirmer, A., & Gunter, T. C. (2017). The right touch: Stroking of CT-innervated skin promotes vocal emotion processing. *Cognitive, Affective & Behavioral Neuroscience, 17*(6), 1129–1140. https://doi.org/10.3758/s13415-017-0537-5

Schirmer, A., & Kotz, S. A. (2006). Beyond the right hemisphere: Brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences, 10*(1), 24–30. https://doi.org/10.1016/j.tics.2005.11.009

Scott, S. K. (2019). From speech and talkers to the social world: The neural processing of human spoken language. *Science, 366*(6461), 58–62. https://doi.org/10.1126/science.aax0288

Soares, A. P., Comesaña, M., Pinheiro, A. P., Simões, A., & Frade, C. S. (2012). The adaptation of the affective norms for English words (ANEW) for European Portuguese. *Behavior Research Methods, 44*(1), 256–269. https://doi.org/10.3758/s13428-011-0131-7

Soares, A. P., Iriarte, A., de Almeida, J. J., Simões, A., Costa, A., Machado, J., … Perea, M. (2018). Procura-PALavras (P-PAL): A web-based interface for a new European Portuguese lexical database. *Behavior Research Methods, 50*(4), 1461–1481. https://doi.org/10.3758/s13428-018-1058-z

Strauß, A., Kotz, S. A., & Obleser, J. (2013). Narrowed expectancies under degraded speech: Revisiting the N400. *Journal of Cognitive Neuroscience, 25*(8), 1383–1395. https://doi.org/10.1162/jocn_a_00389

Sui, J., He, X., & Humphreys, G. W. (2012). Perceptual effects of social salience: Evidence from self-prioritization effects on perceptual matching. *Journal of Experimental Psychology: Human Perception and Performance, 38*(5), 1105–1117. https://doi.org/10.1037/a0029792

Swanborough, H., Staib, M., & Frühholz, S. (2020). Neurocognitive dynamics of near-threshold voice signal detection and affective voice evaluation. *Science Advances, 6*(50), Article eabb3884. https://doi.org/10.1126/sciadv.abb3884

Tesink, C. M. J. Y., Petersson, K. M., van Berkum, J. J. A., van den Brink, D., Buitelaar, J. K., & Hagoort, P. (2009). Unification of speaker and meaning in language comprehension: An fMRI study. *Journal of Cognitive Neuroscience, 21*(11), 2085–2099. https://doi.org/10.1162/jocn.2008.21161

Uddin, L. Q., Kaplan, J. T., Molnar-Szakacs, I., Zaidel, E., & Iacoboni, M. (2005). Self-face recognition activates a frontoparietal "mirror" network in the right hemisphere: An event-related fMRI study. *Neuroimage, 25*(3), 926–935. https://doi.org/10.1016/j.neuroimage.2004.12.018

van Berkum, J. J. A., van den Brink, D., Tesink, C. M. J. Y., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience, 20*(4), 580–591. https://doi.org/10.1162/jocn.2008.20054

Vuilleumier, P. (2005). How brains beware: Neural mechanisms of emotional attention. *Trends in Cognitive Sciences, 9*(12), 585–594. https://doi.org/10.1016/j.tics.2005.10.011

von Kriegstein, K., Eger, E., Kleinschmidt, A., & Giraud, A. L. (2003). Modulation of neural responses to speech by directing attention to voices or verbal content. *Cognitive Brain Research, 17*(1), 48–55. https://doi.org/10.1016/S0926-6410(03)00079-X

Wang, L., Qi, Y., Li, L., & Jia, F. (2021). A combined effect of self and reward: Relationship of self- and reward-bias on associative learning. *Frontiers in Psychology, 12*, 647443. https://doi.org/10.3389/fpsyg.2021.647443

Waters, F., Allen, P., Aleman, A., Fernyhough, C., Woodward, T. S., Badcock, J. C., … Larøi, F. (2012). Auditory hallucinations in schizophrenia and nonschizophrenia populations: A review and integrated model of cognitive mechanisms. *Schizophrenia Bulletin, 38*(4), 683–693. https://doi.org/10.1093/schbul/sbs045

Watson, L. A., Dritschel, B., Obonsawin, M. C., & Jentzsch, I. (2007). Seeing yourself in a positive light: Brain correlates of the self-positivity bias. *Brain Research, 1152*, 106–110. https://doi.org/10.1016/j.brainres.2007.03.049

Yankouskaya, A., Humphreys, G., Stolte, M., Stokes, M., Moradi, Z., & Sui, J. (2017). An anterior–posterior axis within the ventromedial prefrontal cortex separates self and reward. *Social Cognitive and Affective Neuroscience, 12*(12), 1859–1868. https://doi.org/10.1093/scan/nsx112