International Journal of Psychophysiology 133 (2018) 66–78

Contents lists available at ScienceDirect

International Journal of Psychophysiology

journal homepage: www.elsevier.com/locate/ijpsycho

Stimulus complexity matters when you hear your own voice: Attention effects on self-generated voice processing

Tatiana Conde[a,b], Óscar F. Gonçalves[b,c,d], Ana P. Pinheiro[a,b,e,*]

[a] Faculdade de Psicologia, Universidade de Lisboa, Lisbon, Portugal
[b] Neuropsychophysiology Lab, CIPsi, School of Psychology, University of Minho, Braga, Portugal
[c] Spaulding Center of Neuromodulation, Department of Physical Medicine & Rehabilitation, Spaulding Rehabilitation Hospital & Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA
[d] Bouvé College of Health Sciences, Northeastern University, Boston, MA, USA
[e] Cognitive Neuroscience Lab, Department of Psychiatry, Harvard Medical School, Boston, MA, USA

ARTICLE INFO

Keywords:
Self-generated voice
Non-self voice
Stimulus type
Event-related potentials
Attention
P3

ABSTRACT

The ability to discriminate self- and non-self voice cues is a fundamental aspect of self-awareness and subserves self-monitoring during verbal communication. Nonetheless, the neurofunctional underpinnings of self-voice perception and recognition are still poorly understood. Moreover, how attention and stimulus complexity influence the processing and recognition of one's own voice remains to be clarified. Using an oddball task, the current study investigated how self-relevance and stimulus type interact during selective attention to voices, and how they affect the representation of regularity during voice perception.

Event-related potentials (ERPs) were recorded from 18 right-handed males. Pre-recorded self-generated (SGV) and non-self (NSV) voices, consisting of a nonverbal vocalization (vocalization condition) or disyllabic word (word condition), were presented as either standard or target stimuli in different experimental blocks.

The results showed increased N2 amplitude to SGV relative to NSV stimuli. Stimulus type modulated later processing stages only: P3 amplitude was increased for SGV relative to NSV words, whereas no differences between SGV and NSV were observed in the case of vocalizations. Moreover, SGV standards elicited reduced N1 and P2 amplitude relative to NSV standards.

These findings revealed that the self-voice grabs more attention when listeners are exposed to words but not vocalizations. Further, they indicate that detection of regularity in an auditory stream is facilitated for one's own voice at early processing stages. Together, they demonstrate that self-relevance affects attention to voices differently as a function of stimulus type.

1. Introduction

From the first instants after birth to late adulthood, human beings are exposed to their own voice more than to any other type of sound. One's own voice is a socially relevant acoustic signal through which a wealth of critical information (e.g., sex, age, health, identity, affective state) is conveyed to social partners (Kreiman and Sidtis, 2013; Sidtis and Kreiman, 2012). Despite the role played by self-voice recognition mechanisms in successful vocal communication, research in this field has been challenged by methodological issues. These include perceptual differences in self-voice perception when producing speech (due to the presence of bone-conducted sound) vs. when passively listening to pre-recorded self-generated speech (Maurer and Landis, 1990). Notwithstanding, the accurate recognition of pre-recorded self-voice stimuli seems to occur above chance (Nakamura et al., 2001; Rosa et al., 2008a, b; Xu et al., 2013; Pinheiro et al., 2016), showing that individuals can compensate for such perceptual disparities.

1.1. How special is the self-voice?

Alterations in self-voice processing mechanisms may impair verbal communication (Lane and Webster, 1991; Moeller et al., 2007), and have been implicated in psychopathological symptoms such as auditory verbal hallucinations (e.g., Waters et al., 2012; see Conde et al., 2016a for a review). However, the neurofunctional mechanisms underpinning self-voice perception remain to be clarified. The existing studies have indicated important differences in how self- and unknown voices are perceived (Allen et al., 2005; Graux et al., 2013; Graux et al., 2015;

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

https://doi.org/10.1016/j.ijpsycho.2018.08.007
Received 8 June 2017; Received in revised form 5 June 2018; Accepted 10 August 2018
Available online 13 August 2018

Kaplan et al., 2008; Nakamura et al., 2001; Rosa et al., 2008a, b; Xu et al., 2013). For example, an improved recognition of the self-voice was demonstrated in acoustically demanding conditions, in which the voice signal only kept frequencies higher than the third formant (Xu et al., 2013). Also, an early discrimination between self- and non-self voice cues, occurring within 70–100 milliseconds (ms) post-stimulus onset, was revealed by event-related potential (ERP) studies (Graux et al., 2013). Furthermore, compared to unfamiliar voices, listening to the self-voice elicited increased activation in the left inferior frontal and right anterior cingulate gyri (Allen et al., 2005), right inferior frontal gyrus (Kaplan et al., 2008; Nakamura et al., 2001) and right parainsular brain regions (Nakamura et al., 2001). Self-related stimuli were also found to elicit prioritized processing. For instance, attentional enhancement by different categories of self-related stimuli (e.g., self-face/name/hand) has been consistently reported (e.g., Berlad and Pratt, 1995; Gray et al., 2004; Scott et al., 2005; Su et al., 2010; Sui et al., 2006; Tacikowski and Nowicka, 2010; Tacikowski et al., 2014). A memory advantage was additionally observed for information encoded in a self- (vs. non-self) referential manner (Symons and Johnson, 1997). Nonetheless, fewer studies examined how self-relevant stimuli modulate attentional resources during voice perception.

1.2. Does attention affect how the self-voice is perceived?

In dynamically changing multi-speaker contexts, vocal communication also demands a selective and flexible allocation of attentional resources to feedback generated by oneself or by others (Fritz et al., 2007a, b; Rimmele et al., 2015). Even though attention was found to modulate how one's own voice is perceived during speech production (Hu et al., 2015; Tumber et al., 2014), it remains to be clarified whether these effects are dependent on motor processes (associated with voice generation) or whether they extend to self-voice perception in general (i.e., when pre-recorded self-voice stimuli are passively presented). In a recent ERP study, we demonstrated that selective attention to voices is modulated by self-relevance, which was reflected in an increased P3 amplitude to self-compared to non-self speech (i.e., a disyllabic word) when stimuli were in the focus of attention (Conde et al., 2015). The P3 component is typically elicited by infrequent task-relevant events interspersed among frequent stimuli, and it is believed to reflect the mobilization of higher-order attentional resources after the evaluation of stimulus significance (Kok, 2001; Polich, 2007; Spencer et al., 1999, 2001). Together with previous studies focusing on other categories of self-relevant stimuli (e.g., Berlad and Pratt, 1995; Gray et al., 2004; Tacikowski and Nowicka, 2010), it is plausible that the self-voice grabs more attention due to its higher emotional salience.

Nonetheless, even when task-irrelevant (i.e., when participants are engaged in a visual distractive task), self- and unknown voices were found to distinctly modulate attention orienting in the P3 latency window (Graux et al., 2013; Graux et al., 2015). Specifically, the P3a[1] amplitude was increased to both familiar and unfamiliar vocalizations relative to self-generated vocalizations (Graux et al., 2013, 2015). As this ERP component is believed to reflect involuntary capture of attention towards an unexpected change in an otherwise regular environment (Friedman et al., 2001), these findings show that attention orienting is enhanced for (task-irrelevant) non-self voice cues. This is not surprising as, in a social context, a novel or totally unexpected voice signals the presence of a conspecific who might be approached or avoided. In this specific context, the non-self voice may become more salient. Altogether, the studies mentioned above suggest that attentional demands, i.e. ignoring (Graux et al., 2013, 2015) vs. attending (Conde et al., 2015) the voice, modulate the perceived salience of one's own voice. Nevertheless, these studies do not clarify how the self-voice is perceived when in the focus of attention (as it often happens during daily conversations), as well as whether stimulus type affects how one's own voice is discriminated.

1.3. Does stimulus complexity matter when the self-voice is perceived?

Differences in stimulus complexity should be considered when interpreting findings of self-voice perception studies. For example, the multidimensional model of voice perception (Belin et al., 2004; Belin et al., 2011) predicts that both linguistic and nonlinguistic (i.e., identity and affective) vocal cues are processed by partially dissociated but interacting brain pathways. Consistent with this model, critical bi-directional interactions between verbal and nonverbal cues were demonstrated during speaker's recognition (Fleming et al., 2014; Nygaard and Pisoni, 1998; Remez et al., 1997; Schweinberger et al., 1997; Zarate et al., 2015). Specifically, speaker recognition was found to be improved with increased stimulus duration (Cook and Wilding, 1997; Schweinberger et al., 1997), as well as with increased phonetic variability (Roebuck and Wilding, 1993).

Evidence for the effects of signal complexity on how one's own voice is processed is still scarce and limited to the realm of speech production. In one of such studies, Ventura et al. (2009) used magnetoencephalography (MEG) to investigate the effects of stimulus complexity on auditory cortical responsiveness (indexed by the M100[2] component) to self-generated sensory feedback during voice production (Ventura et al., 2009). Importantly, Ventura et al. (2009) demonstrated that the magnitude of auditory cortical suppression to one's own voice feedback (reflected in diminished M/N100 amplitude) depends on stimulus complexity: less complex vocal sounds (/a/) elicited larger M100 amplitude attenuation than more complex and dynamic self-voice stimuli (/a-a-a/ and /a-a-a-a/). Other studies revealed that when participants are presented with experimentally induced changes in voice feedback during speech production (e.g., increased voice F0), they tend to vocally compensate in the direction opposed to the changes introduced, i.e., they lower their voice F0 (e.g., Burnett et al., 1998; Chen et al., 2013). The magnitude of the compensatory vocal responses is modulated both by word content (Patel and Schell, 2008) and by language experience (participants' native language — Liu et al., 2010).

Even though the role of acoustic complexity on self-voice processing has been highlighted by the studies mentioned above, evidence is limited to experimental designs involving voice generation and short vocalizations. However, two important limitations should be noted. First, such designs are typically concerned with how one's own voice is perceived when vocalizing vs. when passively listening to a recording of the same voice (i.e., a self- vs. self-voice comparison), and not with how the self-voice is distinguished from non-self voice cues. Second, neuroscience techniques (e.g., EEG/ERP, MRI/fMRI) are highly sensitive to physiological artifact noise (e.g., muscle activity, eye movements), which limits the development of online voice production tasks with more complex speech stimuli beyond the steady vowel /a/ used in a

[1] The P3a and the P3b components are dissociable brain potentials that reflect distinct neurocognitive processes (see Polich, 2007). The P3a is thought to reflect the involuntary capture of attention by an unpredictable violation of a regular and invariant aspect of the environment. The P3b indexes the mobilization of higher-order attentional resources to a task-relevant deviant (target) event. As in the current experiment participants were asked to focus their attention on the vocal sounds and to silently count the infrequent (and task-relevant) vocal stimuli, henceforth we used the term “P3” to refer to the subcomponent elicited by the task-relevant (target) stimuli.

[2] In voice perturbation tasks, the M100 component (magnetic counterpart of the N1 ERP component) is thought to reflect the operation of an internal predictive mechanism: when sensory feedback matches the prediction, auditory cortical suppression (reflected in reduced M100/N1 amplitude to self- compared to non-self or to altered self-voice feedback) is observed; however, an error signal is generated when the incoming self-voice feedback deviates from the prediction (Behroozmand and Larson, 2011; Heinks-Maldonado et al., 2005; Sitek et al., 2013; Hickok et al., 2011).

considerable number of experiments (e.g., Ford et al., 2010; Whitford et al., 2011). Offline self-voice perception tasks with EEG, in contrast, avoid artifacts resulting from motor activity associated with speech generation, allowing the study of more complex (and thus more ecological) self-generated speech signals. It is also worth noting that our current knowledge on self-voice perception mechanisms relies mostly on studies that used a short and steady vowel (/a/) as experimental stimulus both during voice production (e.g., Ford et al., 2010; Heinks-Maldonado et al., 2005; Sitek et al., 2013) and passive listening conditions (Graux et al., 2013, 2015). This experimental scenario is far away from the diversity and complexity of voice feedback we are exposed to and produce in our daily lives. Thereby, the effects of stimulus type on attention to self- vs. non-self voice cues are not fully understood. Specifically, it is unclear whether the attentional bias previously reported for self-generated word stimuli (Conde et al., 2015) is generalizable to acoustically less complex nonverbal vocalizations used in prior experiments (e.g., Graux et al., 2013, 2015; Heinks-Maldonado et al., 2005; Sitek et al., 2013).

1.4. Does the representation of acoustic regularity differ for self- and non-self voices?

Extracting acoustic regularity representations from a dynamic auditory environment is critical for the prediction of upcoming events and for the detection of violations to these regularities (Bendixen et al., 2007; Jacobsen et al., 2005; Ranganath and Rainer, 2003; Seppänen et al., 2012). Such capacity plays a critical role in everyday social communication, since it allows detecting unexpected changes in self-voice feedback or in others' speech, and hence, a prompt adjustment to such changes (e.g., lowering the volume of the self-voice in response to an abrupt silence or rapidly responding to a sudden affective change in somebody else's voice — e.g., Behroozmand et al., 2009; Behroozmand and Larson, 2011; Burnett et al., 1998; Hu et al., 2015). Within communicational settings, the extraction of acoustic regularities (i.e., invariant aspects) from dynamic voice stimuli supports speaker recognition and guides how his/her emotions are decoded, as well as how the meaning conveyed through speech is understood. This ability should be differentiated from the process of building regularity representations based on the repetition of high-probability stimuli (i.e., abstract regularities). In the context of an oddball task design, the representation of a high-probability and invariant sound operates as a 'comparison template' against which the infrequent deviating events are contrasted with (Bendixen et al., 2007; Jacobsen et al., 2005). Of note, the representation of the self-voice relies on previously learned associations between the repeated motor experience of speaking (i.e., the vocal motor commands) and its sensory consequences, i.e., a sensorimotor representation (Hickok et al., 2011). The learned associations might be critical for vocal self-monitoring as, during voice generation, internal predictive mechanisms rely on these associations to anticipate the sensory consequences of motor commands (Hickok et al., 2011).

Regarding the neural mechanisms underlying acoustic regularity representations, it is generally accepted that the repeated presentation of a given sound leads to decreased brain responsiveness, which is thought to reflect more efficient stimulus processing associated with a stronger and more precise representation of the stimulus (Grill-Spector et al., 2006; Ranganath and Rainer, 2003; Ross and Tremblay, 2009; Seppänen et al., 2012). Reduced neural activation is believed to reflect both the activation of a smaller population of specialized neurons tuned to process the stimulus features and the deactivation of neurons that are not sensitive to such stimulus properties (Grill-Spector et al., 2006; Ranganath and Rainer, 2003; Seppänen et al., 2012). Studies using oddball tasks demonstrated that the repetition of a standard sound results in suppressed neural responsiveness to that sound (Ross and Tremblay, 2009; Seppänen et al., 2012), and that this suppression is modulated by stimulus salience (e.g., Pinheiro et al., 2017) and familiarity (Jacobsen et al., 2005). Therefore, examining the neural responses to standard stimuli might provide important information on how the brain builds abstract regularity representations from invariant and high-probability events, and on how these expectations are used to efficiently detect changes in the voice as a function of self-relevance (self- vs. non-self voice).

1.5. The current study and hypotheses

Our knowledge on the neurofunctional mechanisms subserving voice perception is still limited. Specifically, the distinctive role played by self-relevance, stimulus type and attention in voice perception remains to be clarified. Using an oddball design, our study aimed to unravel how stimulus type (vocalization vs. word) modulates selective attention to self- and unknown voices, with the focus on the N2 and P3 ERP components. Self-generated and non-self voices consisting of a vocalization (vocalization condition) or of a disyllabic word (word condition) were presented both as standard and target stimuli in four distinct blocks. Participants were instructed to detect a change in the auditory stimulation. Considering that attentional resources are more strongly engaged by self-relevant stimuli both at early (i.e., within 200 ms post-stimulus onset; Conde et al., 2015; Fan et al., 2011; Fan et al., 2013) and higher-order (after approximately 300 ms post-stimulus onset; Gray et al., 2004; Perrin et al., 2005; Scott et al., 2005; Su et al., 2010; Sui et al., 2006; Tacikowski and Nowicka, 2010; Tacikowski et al., 2014) processing stages, we predicted that the self-voice would attract more attention (reflected in increased N2 and P3 amplitude). This would dovetail with the enhanced salience of the self-related stimuli, irrespective of stimulus type. On the other hand, if stimulus complexity affects how self- and non-self voices are differentiated (e.g., Schweinberger et al., 1997; Zarate et al., 2015), then an interaction between voice identity (self- vs. non-self) and stimulus type (vocalization vs. word) should be evidenced (i.e., similar N2 and P3 amplitude to self- and non-self voices in the vocalization condition; increased N2 and P3 amplitude to the self-voice in the word condition — Conde et al., 2015). Since interactions between identity and speech dimensions were also observed in the first 200 ms after voice onset (e.g., Kaganovich et al., 2006), we expected such interaction to be observed at both early (indexed by N2) and higher-order (indexed by P3) attentional processing stages.

Additionally, we investigated the N1 and P2 responses to standard sounds to clarify whether voice identity affects regularity representations from invariant and high-probability stimuli (Ross and Tremblay, 2009; Seppänen et al., 2012). Specifically, examining ERP responses to standards might provide further insights on stimulus-driven processes related to both the sensory registration of stimulus acoustic properties and further operations related to stimulus classification, reflected in the N1 (Salisbury et al., 2010) and P2 components (Crowley and Colrain, 2004), respectively. As previous ERP studies (Roye et al., 2007; Roye, 2010) demonstrated that the brain rapidly detects salience within the first ~200 ms after stimulus onset (even in the absence of directed attention to the incoming auditory stimulation), differences in N1 and P2 responses to self- and non-self voice standards were expected. In particular, reduced neural responsiveness associated with stimulus repetition is believed to reflect the sharpening of its neural representation (Grill-Spector et al., 2006; Seppänen et al., 2012) and a reduced prediction error (Summerfield et al., 2008). Since the self-voice is likely to activate a stronger sensorimotor representation than a non-self voice (Xu et al., 2013), we hypothesized greater N1 and P2 reduction for self- relative to non-self vocal standards. If confirmed, this would indicate that the sensory analysis and categorization of the relevant physical stimulus properties (indexed by the N1 and P2 components — Crowley and Colrain, 2004; Salisbury et al., 2010) underlying the representation of acoustic regularities are enhanced for the self-voice.

Table 1
Socio-demographic and cognitive characterization of the participants.

A. Socio-demographic data	*M* (*SD*)
Age, years	30.79 (5.22)
Years of education	15.00 (2.96)
B. Cognitive data*	***M (SD)***
Full scale IQ	124.95 (12.36)
Verbal IQ	127.63 (11.53)
Performance IQ	116.32 (12.51)

Note. M = mean; SD = standard deviation.
* WAIS-III (Wechsler, 2008).

2. Material and methods

2.1. Participants

Eighteen right-handed males participated in the study (see Table 1). All participants reported normal hearing and all were native speakers of European Portuguese. Participants were screened for psychopathological symptoms with the Brief Symptom Inventory (BSI; Derogatis, 1982; Portuguese version — Canavarro, 1999). Inclusion criteria were: right-handedness (Oldfield, 1971); European Portuguese as first language; no history of electroconvulsive treatment, neurological illness, or DSM-IV diagnosis of drug or alcohol abuse; no current medication with potential impact on the electroencephalogram (EEG), or with neurological and/or cognitive functioning consequences; *Positive Symptoms Distress Index* of the BSI below 1.7 (Canavarro, 2007). Participants provided their informed consent, previously assessed by the local Institutional Review Board committee (University of Minho, Braga, Portugal).

2.2. Stimuli

At least one week before the ERP experiment, participants were asked to utter several instances of a short vocalization (using the steady vowel /a/) and of a disyllabic word (*self-generated voice condition* — SGV). The best sample of each stimulus type per participant was selected for the experiment based on the consensual rating of two judges. Voice recording took place in a sound-proof room, using a portable digital recorder Roland R-26 plus a Shure incorporated PG48 microphone (sampling rate = 44.1 kHz, 16-bit resolution). For the non-self voice condition (NSV), the voice from an unknown middle-aged male without regional accent was recorded.

The steady vowel /a/ (duration = 300 ms) was selected for the vocalization condition, following earlier experiments (e.g., Ford et al., 2010; Graux et al., 2013, 2015; Whitford et al., 2011) to allow more direct comparisons of findings across studies. The use of steady vowels in voice research has the main advantage of restricting the non-laryngeal information available in the voice signal, i.e., non-laryngeal contributions (vocal tract resonances are kept constant across utterances and speakers; Kreiman and Sidtis, 2013). The word was selected based on psycholinguistic properties derived from the P-PAL (Soares et al., 2010) set, and on affective ratings derived from the 'Affective Norms for English Words' (ANEW; Soares et al., 2012), according to the following criteria: high frequency (> 100 per million; P-PAL), grammatical class (noun; P-PAL), neutral valence (5.34; ANEW), low arousal (3.36; ANEW), and short length (two syllables and four letters; P-PAL). To control for variability in word pronunciation among participants in the voice recording session, a stable syllabic structure (i.e., consonant–vowel–consonant–vowel) also determined stimulus selection. The duration (vocalization = 300 ms; word = 483 ms) and intensity (root-mean-square amplitude — RMS = 70 dB) of both SGV and NSV were matched using Praat software (Boersma and Weenink, 2012). Background noise was removed using Audacity 2.0.2. software (http://audacity.sourceforge.net/). All voice stimuli were acoustically analyzed using Praat software (see Table 2).

Table 2
Acoustic properties of the voice stimuli.

Subject	Stimulus	Mean F0 (Hz)	Range F0 (Hz)		Formant frequencies (Hz)		
			Min	Max	F1	F2	F3
P1	V	116	110	120	849	1421	2463
	W	115	105	130	717	1016	2312
P2	V	98	93	102	928	1796	2625
	W	93	87	105	581	1090	2898
P3	V	108	97	111	945	1419	2630
	W	94	83	103	625	1185	2385
P4	V	111	108	113	746	1566	2248
	W	93	76	101	567	1382	2527
P5	V	102	100	107	744	1212	2455
	W	94	78	108	719	1046	2178
P6	V	92	88	98	858	1410	2537
	W	93	86	106	573	1138	2886
P7	V	114	112	119	809	1348	2840
	W	108	98	120	547	1002	2260
P8	V	107	99	111	799	1257	2737
	W	94	77	124	616	1012	2312
P9	V	90	89	98	882	1400	2611
	W	96	80	112	557	1032	2777
P10	V	113	112	114	746	1321	2382
	W	109	93	119	558	1071	2650
P11	V	92	89	93	698	1411	2420
	W	85	75	91	569	1239	2615
P12	V	100	98	102	881	1398	2429
	W	99	88	141	570	1223	2410
P13	V	129	124	135	875	1453	2546
	W	121	108	162	399	1273	2607
P14	V	119	116	124	741	1492	2496
	W	124	110	154	519	1143	2557
P15	V	117	115	122	816	1446	2689
	W	103	94	113	661	1377	2585
P16	V	103	101	113	890	1274	2402
	W	94	87	105	520	973	2169
P17	V	133	130	135	897	1440	2688
	W	113	106	120	374	1165	2664
P18	V	111	108	114	882	1403	2651
	W	104	98	116	527	1093	2439
M		101.78	90.50	118.33	566.66	1136.65	2512.83
(SD)		(11.00)	(11.57)	(18.53)	(88.37)	(123.74)	(221.37)
NSV	V	96	83	106	621	944	2258
	W	119	101	123	703	1428	2362

Note. M = mean; SD = standard deviation; V = vocalization; W = word.

2.3. Procedure

Participants were tested in two experimental sessions (one per stimulus type: vocalization and word), which took place in distinct days, with a minimum of 24 h separating them. Each experimental session comprised two blocks, each including 200 standards and 40 targets: in the first block, the NSV was the standard (p = .83) and the SGV was the target (p = .17), whereas in the second block the opposite was observed (see Fig. 1). Participants were seated in a comfortable chair at a distance of 100 cm from the computer screen. Voice stimuli were binaurally presented through Sennheiser CX 300-II earphones in a pseudorandom order (a minimum of two standards were presented before a target sound — Özgürdal et al., 2008). Presentation software (Neurobehavioral Systems, Albany, CA, USA) was used to control stimulus timing and presentation. Each trial comprised a fixation cross (which remained on the screen from the beginning until the end of the trial) and, after 300 ms, a voice stimulus (vocalization = 300 ms; word = 483 ms) was presented (see Fig. 1). Stimulus onset asynchrony (SOA) varied between 1100 and 1200 ms in the vocalization condition,

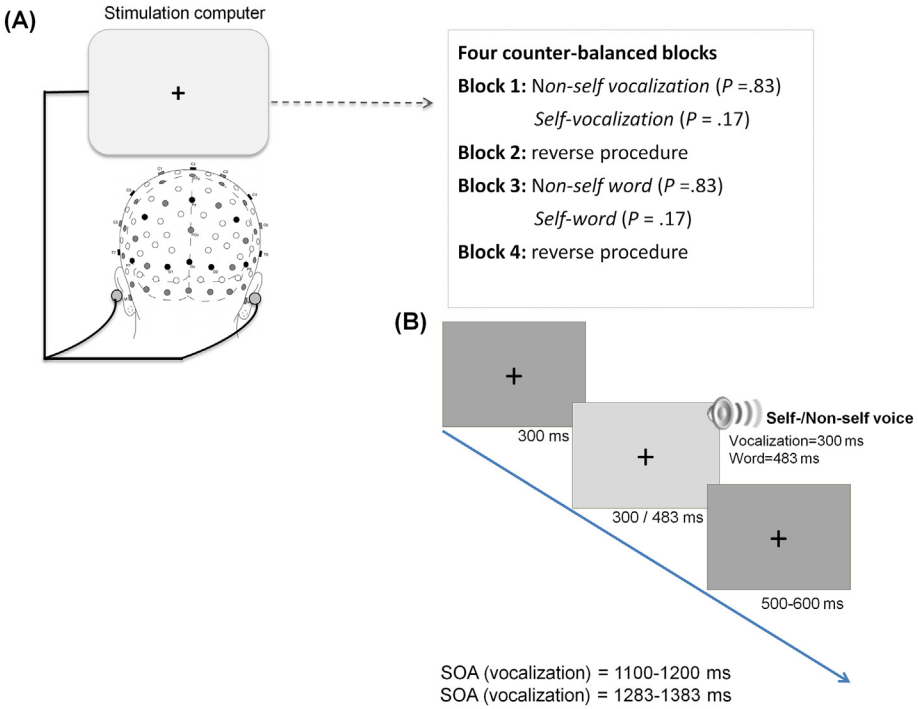

Fig. 1. Schematic illustration of the experimental task. (A) Illustration of the task; (B) illustration of a trial.

and between 1283 and 1383 ms in the word condition. Both the order of experimental sessions and the order of blocks were counterbalanced across participants, and each block lasted approximately 4 min. In each block, participants were asked to silently count the number of times a target voice was presented and to report it at the end of the block.

2.4. EEG recording and data analyses

EEG data were continuously recorded using a 64 channels BioSemi Active Two System (http://www.biosemi.com/products.htm), at a digitization rate of 512 Hz, and stored for later analysis. Eye movements were recorded by placing electrodes at left and right temples (horizontal electrooculogram — EOG) and one below the left eye (vertical EOG). Electrodes were also placed on left and right mastoids for offline referencing. Brain Vision Analyzer 2.0.4 software (www.brainproducts.com) was used for offline analysis of EEG data. A high-pass filter of .01 Hz was applied. EEG data were referenced offline to the average of the left and right mastoids. Individual ERP epochs of 1100 ms, time-locked to the onset of the voice, were created and included a −100 ms prestimulus baseline. Ocular artifacts were corrected based on Gratton et al. (1983). EEG epochs exceeding ± 100 μV were not included in individual ERP averages. At least 75% of segments in each condition were included in the individual ERP averages (vocalization condition: standard SGV = 172.00 ± 19.08, target SGV = 37.39 ± 2.57, standard NSV = 182.89 ± 12.54, target NSV = 34.83 ± 2.96; word condition: standard SGV = 179.75 ± 15.88, target SGV = 34.88 ± .20, standard NSV = 174.69 ± 14.33, target NSV = 36.50 ± 3.35).

2.4.1. Difference waveforms (target − standard)

Since the computation of target-minus-standard difference waveforms in an oddball task is considered a valuable tool to isolate specific ERP components (Luck, 2005), ERP waveforms elicited by standards were subtracted from the ERP waveforms elicited by targets. Following prior studies (e.g., Graux et al., 2013, 2015; Schirmer et al., 2005; Schirmer & Escoffier, 2010), difference waveforms were computed using a like-from-like subtraction approach, i.e., subtracting SGV standards from SGV targets, and NSV standards from NSV targets. This approach allowed us to control for the physical differences between SGV and NSV stimuli.

After visually inspecting the grand average difference waveforms and following prior ERP studies (Fan et al., 2011; Kayser et al., 2001; Perrin et al., 1999; Spencer et al., 1999, 2001), the N2 and P3 ERP components were selected for further analyses. Since maximal effects were observed at central and parietal sites (Duncan et al., 2009; Fan et al., 2011; Fan et al., 2013; Gray et al., 2004; Özgürdal et al., 2008), the N2 and P3 components were measured at central (Cz/1/2), centro-parietal (CPz/1/2) and parietal (Pz/1/2) electrodes. The N2 amplitude was measured as mean voltage in the 80-ms time window centered around the N2 peak, separately determined for each condition (vocalization condition: SGV = 238 ms; NSV = 230 ms; word condition: SGV = 253 ms; NSV = 249 ms). Likewise, the P3 amplitude was computed as mean amplitude in the 140-ms time window centered at the P3 peak (vocalization condition: SGV = 360 ms; NSV = 356 ms; word condition: SGV = 391 ms; NSV = 384 ms). Latency windows for subtraction-based N2 and P3 components were selected based on visual inspection of the grand average waveforms, in good agreement with previous ERP reports (e.g., Fan et al., 2011; Fan et al., 2013; Gray et al., 2004; O'Donnell et al., 1993; Schweinberger et al., 2011). The N2 peak latency for both SGV and NSV conditions was computed as the time corresponding to the most negative point between 180 and 280 ms. The P3 peak latency was measured as the time of the maximum positive point between 280 and 460 ms.

2.4.2. Standard-related waveforms

Inspection of the grand averages for vocal standards revealed that the N1 and P2 components had maximal effects at fronto-central and central electrode sites (e.g., O'Donnell et al., 1993; Pinheiro et al., 2015; Salisbury et al., 2010; Schweinberger et al., 2011). Following previous studies (e.g., Luck et al., 1990; Salisbury et al., 2010; Schweinberger et al., 2011), the amplitudes of N1 and P2 components were computed as the mean voltage in the 40-ms (N1) and 80-ms (P2) latency windows centered around each respective peak, separately determined for each stimulus type (vocalizations: N1 — SGV = 146 ms; NSV = 141 ms; words: N1 — SGV = 155 ms; NSV = 153 ms; vocalizations: P2 — SGV = 238 ms; NSV = 234 ms; words: P2 — SGV = 240 ms; NSV = 246 ms). Latency windows for the N1 and P2 responses to

standards were selected based on visual inspection of the grand average waveforms, keeping with previous ERP studies (e.g., Pinheiro et al., 2013, 2015; Schweinberger et al., 2011; Salisbury et al., 2010). The N1 peak latency was measured as the time of the maximum negative point between 100 and 180 ms, whereas the P2 peak latency was measured as the time of the maximum positive point between 180 and 280 ms.

2.5. Statistical analyses

All statistical analyses were conducted using IBM SPSS 22.0 software (SPSS, Corp., USA).

2.5.1. Difference waveforms (target − standard)

The mean amplitude and peak latency of the N2 and P3 components in the difference waveforms were subjected to separate repeated-measures analyses of variance (ANOVA), with stimulus type (vocalization, word), voice identity (self-, non-self), and electrode (Cz, CPz, Pz) as within-subject factors. Main effects and interactions were followed with pairwise comparisons (i.e., *t*-tests) between conditions, using the Bonferroni adjustment for multiple comparisons. Where appropriate (i.e., for main effects and interactions involving the electrode factor), analyses were corrected for non-sphericity using the Greenhouse-Geisser correction method. Partial eta squared (η_p^2) values for main effects and interactions are reported to provide an estimate of effect size (Field, 2013).

2.5.2. Standard-related waveforms

The mean amplitude and peak latency of the N1 and P2 components for vocal standards were subjected to separate repeated-measures ANOVA, with stimulus type (vocalization, word), voice identity (self, non-self), and electrode (FCz, Cz) as within-subjects factors. Main effects and interactions were followed with pairwise comparisons between conditions, using the Bonferroni adjustment for multiple comparisons. Partial eta squared values (η_p^2) for main effects and interactions are reported.

3. Results

3.1. Task performance

In the vocalization condition, accuracy in target detection was 99.13% (*range* = 95–100%) for the SGV and 99.00% (*range* = 95–100%) for the NSV, whereas in the word condition accuracy was 98.40% (range = 92.5–100%) for the SGV and 98.75% (range = 95–100%) for the NSV.

3.2. ERP data

Grand average waveforms are illustrated in Figures 2 and 4 (see also Figures 3 and 5).

3.2.1. Difference waveforms (target − standard)

3.2.1.1. N2

3.2.1.1.1. N2 amplitude. A main effect of voice identity, $F(1,17) = 5.739$, $p = .028$, $\eta_p^2 = .252$, revealed more negative N2 for the self- relative to the non-self voice ($p = .028$). No significant main effects or interactions involving stimulus type were observed ($p > .05$).

3.2.1.1.2. N2 latency. A main effect of stimulus type, $F(1,17) = 13.026$, $p = .002$, $\eta_p^2 = .434$, indicated that the N2 peaked earlier for vocalizations relative to words ($p = .002$). No significant main effects or interactions involving identity ($p > .05$) were observed.

3.2.1.2. P3

3.2.1.2.1. P3 amplitude. A significant interaction between stimulus type and voice identity, $F(1,17) = 5.230$, $p = .035$, $\eta_p^2 = .235$, was observed: for words only, the P3 was more positive for the self-voice ($p = .018$); no significant differences between self- and non-self voices emerged in the case of vocalizations ($p > .05$). No significant main effects of stimulus type or identity ($p > .05$) were found.

3.2.1.2.2. P3 latency. A main effect of stimulus type, $F(1,17) = 16.280$, $p = .001$, $\eta_p^2 = .489$, revealed that the P3 peaked earlier for vocalizations relative to words ($p = .001$). No significant main effects or interactions involving identity ($p > .05$) were observed.

3.2.2. Standard-related waveforms

3.2.2.1. N1

3.2.2.1.1. N1 amplitude. A main effect of identity, $F(1,17) = 7.805$, $p = .012$, $\eta_p^2 = .315$, revealed a more negative N1 in response to the non-self voice ($p = .012$). No significant main effects or interactions involving stimulus type were observed ($p > .05$).

3.2.2.1.2. N1 latency. A main effect of stimulus type, $F(1,17) = 20.360$, $p < .001$, $\eta_p^2 = .545$, revealed that the N1 peaked earlier for vocalizations compared to words ($p < .001$). No significant main effects or interactions involving identity were observed ($p > .05$).

3.2.2.2. P2

3.2.2.2.1. P2 amplitude. A main effect of identity, $F(1,17) = 9.567$, $p = .007$, $\eta_p^2 = .360$ indicated that the P2 was more positive in response to non-self voices ($p = .007$). No significant main effects or interactions involving stimulus type were observed ($p > .05$).

3.2.2.2.2. P2 latency. P2 latency was not affected by any of the factors tested ($p > .05$).

4. Discussion

In this study, we investigated the effects of self-relevance ('*my voice*' vs. '*someone else's voice*') and stimulus type (vocalizations vs. words) on selective attention to voices using ERPs. Participants performed an oddball task in which they were instructed to detect self- and non-self voice targets interspersed with self- and non-self vocal standards, respectively. Our results extend previous investigations (e.g., Conde et al., 2015; Graux et al., 2013, 2015) of self-voice perception, by revealing that the processing resources directed to one's own voice are modulated by stimulus type (short vocalizations vs. words). These findings are discussed below.

4.1. Self-relevance and stimulus type modulate selective attention to voices

The N2 is thought to index the early discrimination and categorization of task-relevant stimuli deviating from an invariant auditory stream (O'Donnell et al., 1993; Patel and Azzam, 2005; Salisbury et al., 1994). In our study, the self-voice elicited increased (i.e., more negative) N2 amplitude when compared with the unknown voice, irrespective of stimulus type. These findings support an early discrimination of self- and non-self voices, revealing that attention is enhanced when the voice is self-relevant. They agree with previous evidence showing increased N2 to self-related stimuli (Conde et al., 2015; Fan et al., 2013), which might be related to their putatively enhanced salience (e.g., Gray et al., 2004; Tacikowski and Nowicka, 2010). Contrary to our initial hypothesis, the N2 was not affected by stimulus type. Hence, at early attentional stages, the discrimination and categorization processes underlying voice identity perception might be less dependent on the complexity of the acoustic signal. This finding keeps with the observation of interaction effects between identity and stimulus type at the P3 latency window only, whereas earlier stages of acoustic change detection (indexed by the MMN ERP component) appear to be less sensitive to such interaction effects (Conde et al., 2016a, b).

The P3 component reflects the mobilization of higher-order

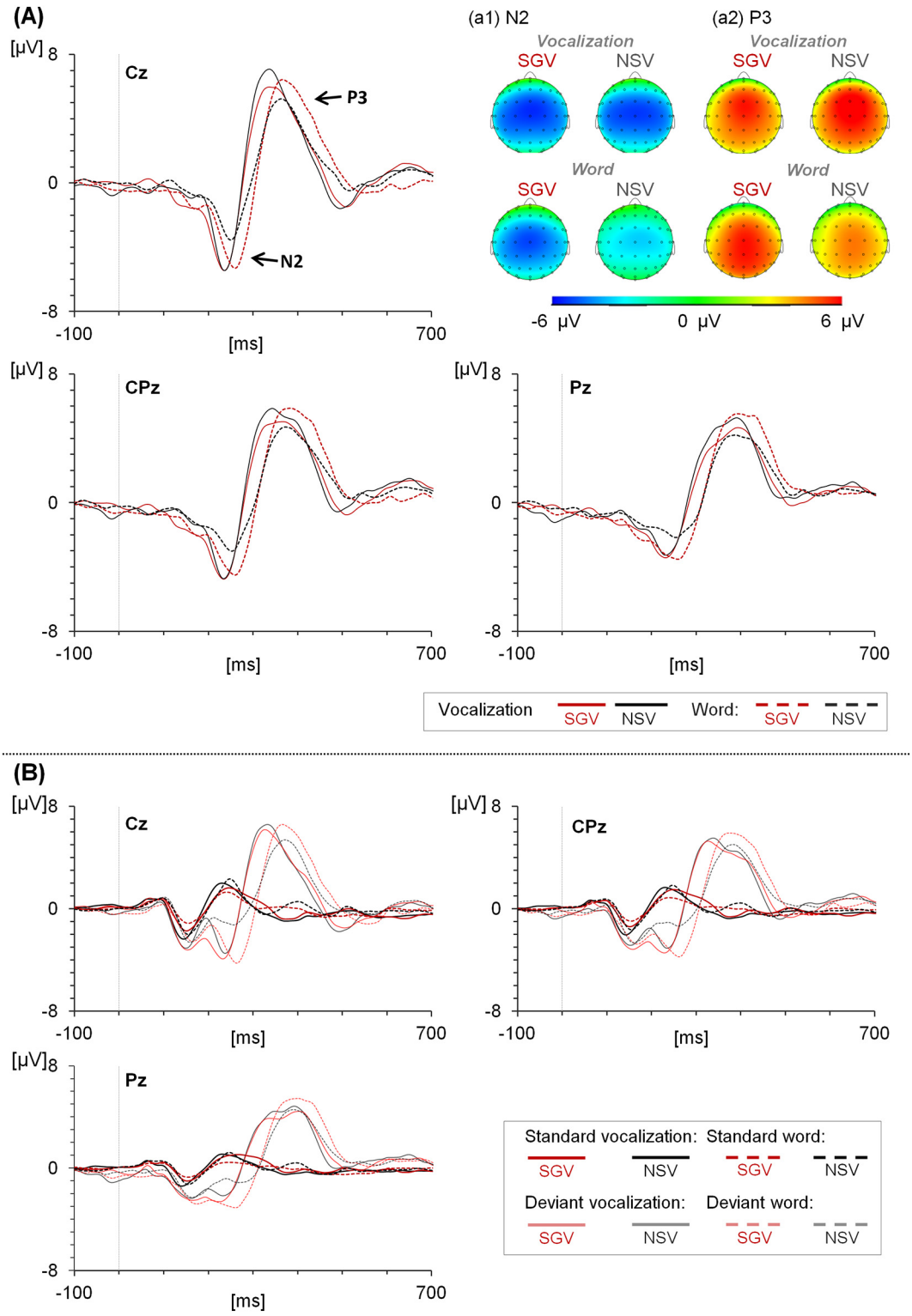

Fig. 2. (A) Illustration of the grand average difference waveforms for SGV and NSV stimuli in both vocalization and word conditions at Cz, CPz and Pz channels (a 15 Hz high cutoff filter was applied for illustration purposes only). Topographical maps representing (a1) N2 and (a2) P3 components. (B) Illustration of the grand average 'raw' waveforms for SGV and NSV standard and deviant stimuli, in both vocalization and word conditions.

attentional resources to task-relevant events, as well as the evaluation of stimulus relevance/salience (Kok, 2001; Polich, 2007; Polich and Criado, 2006; Spencer et al., 1999, 2001). In the current study, the P3 findings revealed differences in the processing of self- and non-self voices that emerged as a function of stimulus type: when words were presented, the P3 amplitude was increased in response to the self-voice, whereas both self- and non-self voices captured a similar amount of attentional resources when nonverbal vocalizations were presented. This result confirmed our initial hypothesis that stimulus type matters and determines how much attention is paid to the self-voice. Previous studies provided evidence for the modulatory effects of emotional salience (e.g., Delplanque et al., 2006) and self-relevance (e.g., Chen et al.,

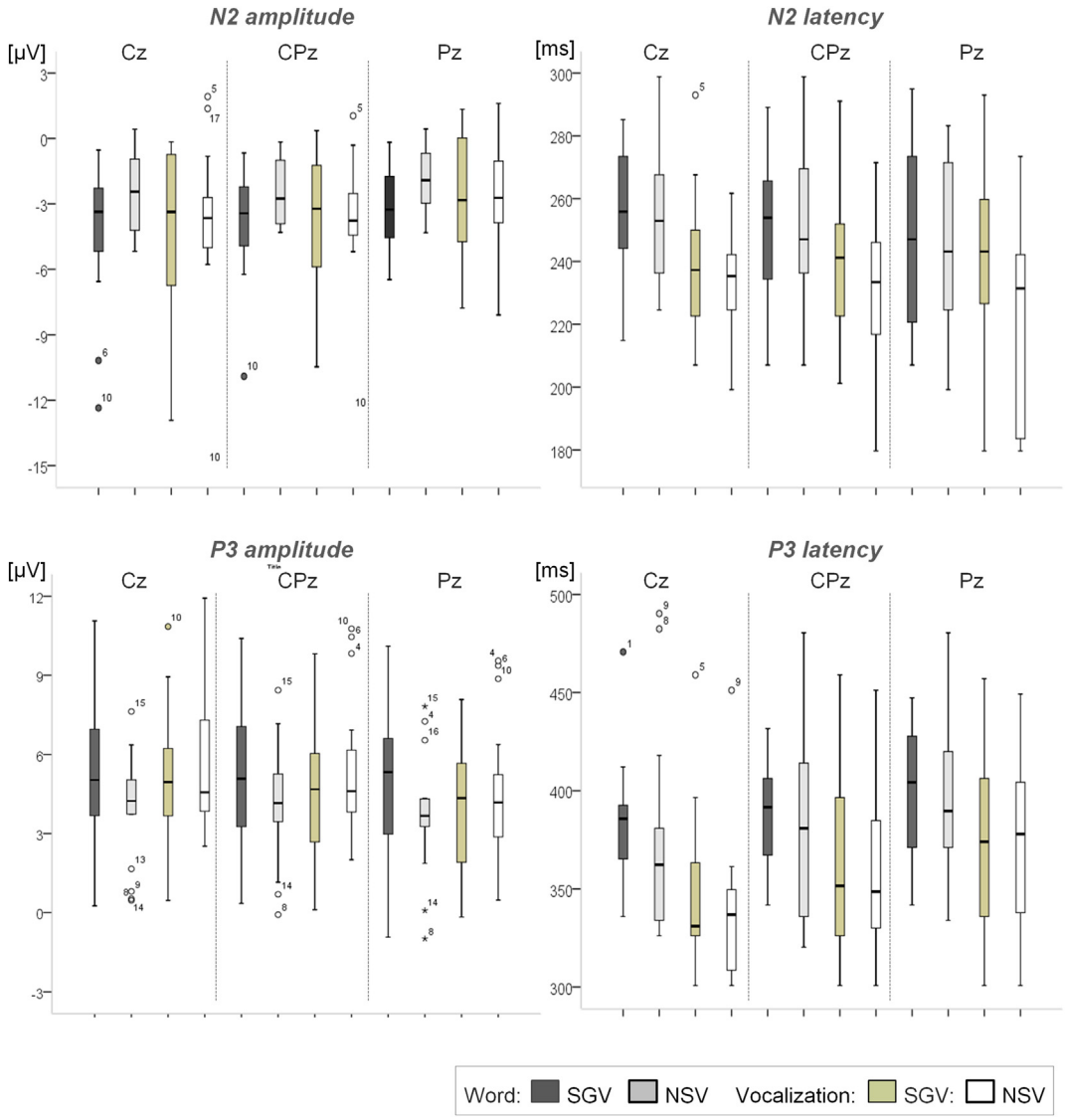

Fig. 3. Box plots representing N2 and P3 amplitudes and peak latencies, based on difference waveforms, in both vocalization and word conditions.

2011; Gray et al., 2004; Fan et al., 2013; Tacikowski and Nowicka, 2010) on the P3 component. Importantly, the emotional salience of a given event, and hence, the processing resources recruited, depend upon the context where this event occurs, as well as on its implications to the current goals/behavior of the organism (Domínguez-Borràs et al., 2009; Friedman et al., 2001; Sussman, 2007). For instance, consider the salience of hearing the exact same unknown speaker's voice at our door in the morning vs. at late hours in the night: it is likely enhanced in the latter context. In the current study, it is plausible that both types of voice identity were assigned comparable salience but only when the vocal information provided was shorter in duration and acoustically less complex.

When discussing our findings of greater P3 amplitude to self-generated word stimuli and the findings observed by Graux et al. (2013, 2015) of reduced P3a to self vs. non-self (familiar and unfamiliar) vocalizations, critical differences in the experimental task conditions and in the ERP components under study should be considered. In the studies of Graux et al. (2013, 2015), participants directed attention to a silent movie, whilst ignoring the voice stimuli. Hence, in these studies, the P3a reflects an orienting response, that is, an involuntary shift in attention elicited by an unexpected violation in an otherwise unchangeable auditory stream (Friedman et al., 2001; Kok, 2001; Spencer et al., 1999; Spencer et al., 2001). This involuntary capture of attention is thought to represent a task-irrelevant response with implications for survival (Friedman et al., 2001). On the other hand, in our study, participants were instructed to focus their attention on the target voices and to silently count the number of times a target stimulus was presented. Therefore, the P3 component (i.e., the P3b) in our study is thought to index the allocation of higher-order attentional resources to a task-relevant target stimulus and the cognitive evaluation of its significance (Kok, 2001; Polich, 2007). When participants are engaged in a concurrent task and are not paying voluntary attention to the voice background (as in the studies of Graux et al., 2013, 2015), identifying a novel and completely unexpected conspecific voice might be more relevant than the detection of one's own voice: consequently, the P3a amplitude is increased to non-self voice cues (familiar/unknown). Taken together, these findings demonstrate that attentional demands (ignoring vs. attending) affect how the brain responds to self- or to non-self voices, plausibly by influencing their perceived salience/relevance to the current goals/behavior of the listener. Of note, the current study relied on a “like-from-like” subtraction approach to maximally reduce the influence of context on the N2 and P3, our experimental design does not allow dissociating whether attentional resources are enhanced for self-generated words due to the salience of the self- (vs. non-self) voice representation, to context (the unknown voice as the high-probability stimulus) or to both. Indeed, as mentioned before, what is perceived as

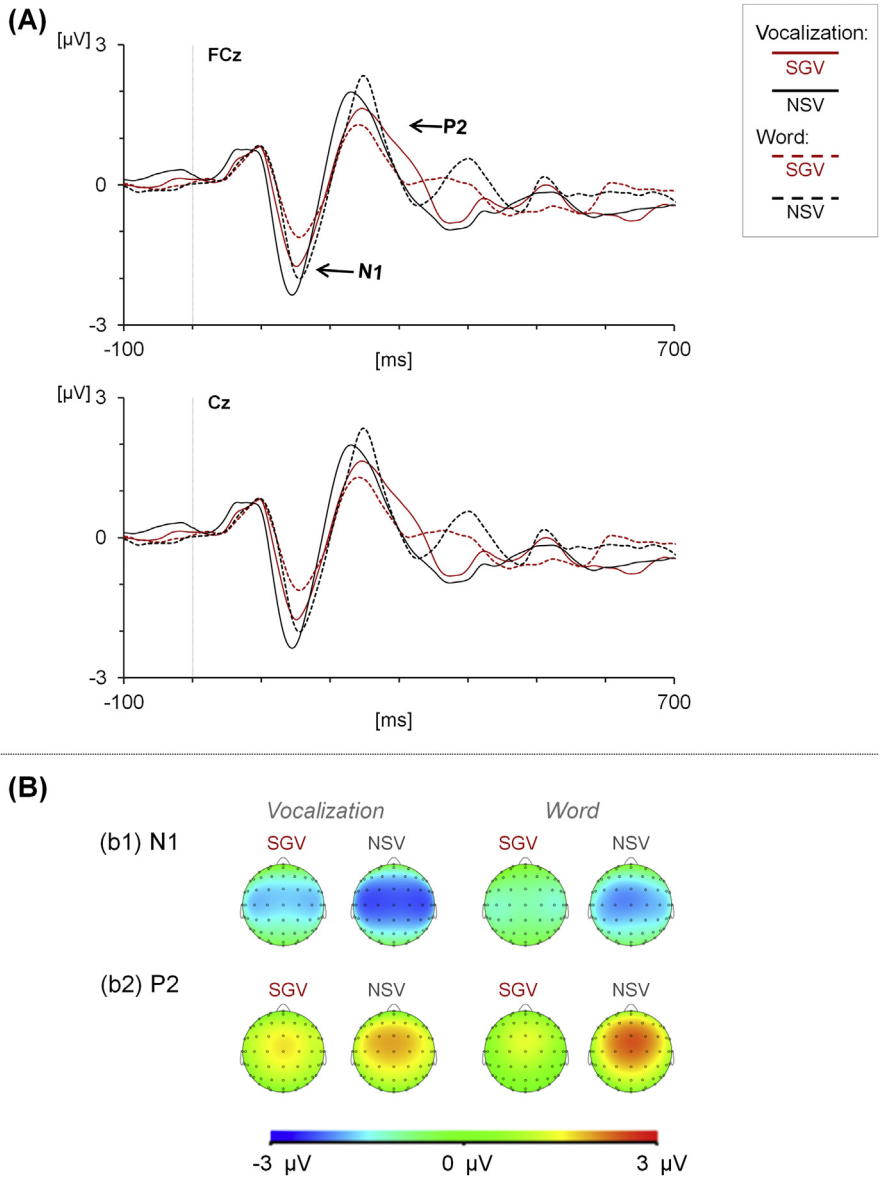

Fig. 4. Illustration of the ERP responses for vocal standards. (A) Grand average 'raw' waveforms for SGV and NSV standards in both vocalization and word conditions. (B) Topographical maps for (b1) N1 and (b2) P2 amplitudes.

salient also depends on the context (e.g., Domínguez-Borràs et al., 2009; Friedman et al., 2001; Sussman, 2007).

Modulatory effects of stimulus type on voice identity perception have been pointed out by prior studies (e.g., Nygaard and Pisoni, 1998; Roebuck and Wilding, 1993; Schweinberger et al., 1997; Zarate et al., 2015). Specifically, stimulus duration was found to have a strong impact on voice perception, with longer stimuli favoring recognition accuracy of speaker's identity (Cook and Wilding, 1997; Schweinberger et al., 1997) and vocal affect (Pell and Kotz, 2011). This may occur because the speaker's phonetic repertoire tends to be better represented by longer (vs. shorter) utterances, as they convey a greater amount of phonetic information that facilitates speaker recognition (e.g., articulatory features; variety of vocal tract resonance patterns — Kreiman and Sidtis, 2013; Remez et al., 1997; Roebuck and Wilding, 1993; Schweinberger et al., 1997). Further, increasing the duration of the voice signal enhances the range of nonlinguistic information (e.g., speaking rate, loudness, mean F0, F0 variability) that is available to listeners, which has been shown to significantly contribute to the recognition of the identity and affective state of a speaker (Belin et al., 2011; Kreiman and Sidtis, 2013; Latinus and Belin, 2011).

Considering prior work supporting the influence of both stimulus duration and phonetic variability cues on speaker recognition (e.g., Cook and Wilding, 1997; Remez et al., 1997; Roebuck and Wilding, 1993; Schweinberger et al., 1997), one might claim that the lack of differences between self- and non-self vocalizations in the current study resulted from the reduced amount of invariant identity cues available in the vocalization /a/ (vs. word). This may have increased the difficulty of participants to recognize their own voice as self-generated when vocalizations were presented. This explanation keeps with earlier evidence showing improvements in speaker recognition as a function of the amount of acoustic and phonological cues available to the listener (Zarate et al., 2015). Furthermore, the use of steady vowels in voice experiments has the main advantage of allowing the control of non-laryngeal contributions to a speaker's voice, such as vocal tract resonances, which are kept quite constant within and between speakers (Kreiman and Sidtis, 2013). Non-laryngeal factors (vocal tract resonances) are fundamental cues both in speaker recognition and speech perception (Belin et al., 2004, 2011; Kreiman and Sidtis, 2013; Latinus and Belin, 2011; Schweinberger et al., 2014). On the other hand, previous ERP studies that used a steady vowel /a/ with the exact same

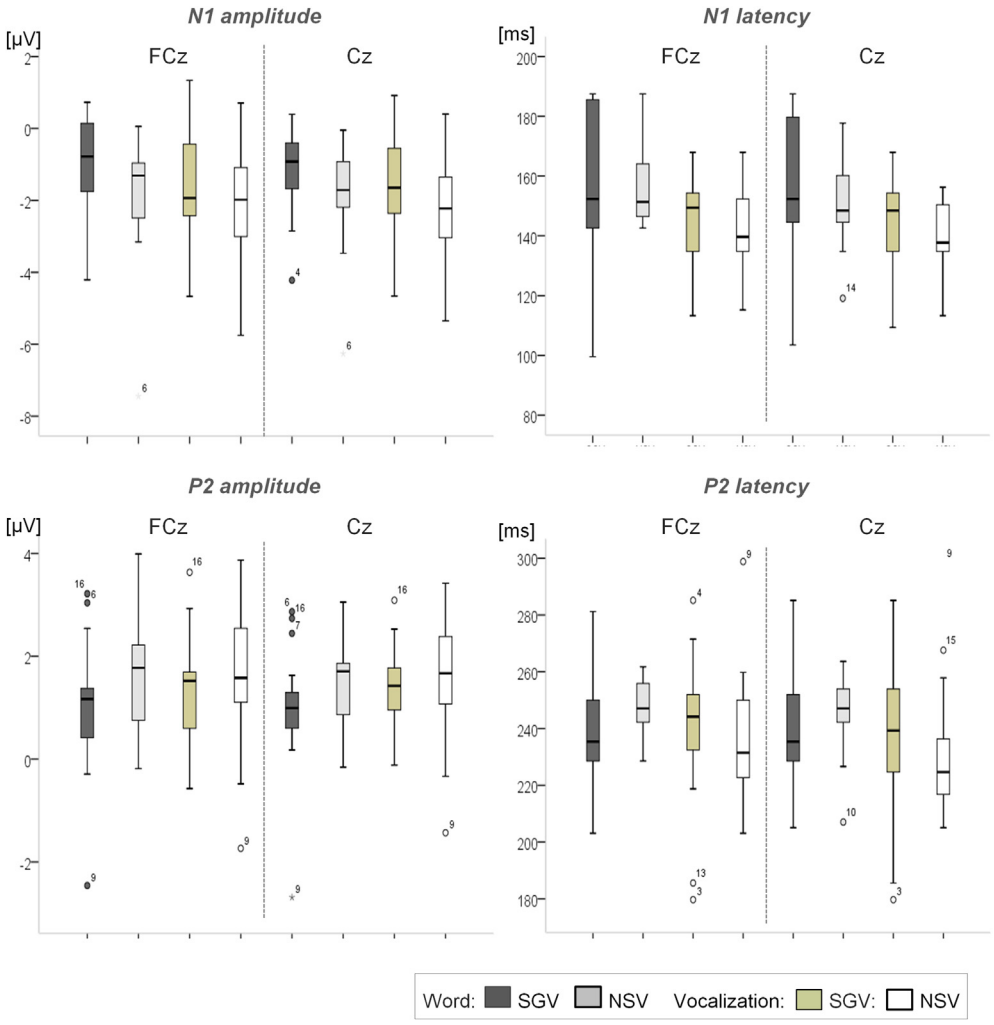

Fig. 5. Box plots representing N1 and P2 amplitudes and peak latencies, based on grand average 'raw' waveforms for SGV and NSV standards, in both vocalization and word conditions.

duration of our study revealed that, when participants' attention is not directed to the voices, the vocalization /a/ conveys sufficient identity cues to result in reduced attentional orienting to the self-voice compared to a totally unpredictable unknown voice (Conde et al., 2016a, b; Graux et al., 2013, 2015). However, in our study we did not directly probe whether a steady vowel carries enough identity cues to allow for explicit self-recognition (i.e., above chance). To the best of our knowledge, no study has tested this before, nor how differences in the amount of linguistic and nonlinguistic cues affect self-voice recognition accuracy at the behavioral level.

On the other hand, an alternative interpretation for the absence of identity-related differences in the P3 component to vocalizations is that the reduced amount of acoustic information conveyed by one's own voice might have decreased the amount of self-monitoring resources engaged in the task. This could result in a reduced perceived salience of the self-vocalization, in contrast to more complex linguistic stimuli (Conde et al., 2015). The differential responsiveness to self- vs. non-self voices as a function of stimulus type could represent an advantageous feature of the vocal self-monitoring system. This system may selectively increase processing resources to more meaningful aspects (e.g., verbal cues) of one's own voice, whilst a similar amount of resources is available for the processing of both self- and non-self voice nonverbal stimuli. This interpretation agrees with previous studies showing that the magnitude of the compensatory vocal responses during speech production is modulated by the linguistic information conveyed in the signal, particularly by word content (Patel and Schell, 2008) and language experience (Liu et al., 2010). Furthermore, it is also consistent with evidence showing that the vocal self-monitoring system is tuned to flexibly detect and correct for vocal production errors, while adapting to the challenges imposed by 'noisy' social acoustic environments (e.g., Behroozmand and Larson, 2011; Sitek et al., 2013). When considering the challenging nature of everyday spoken language (e.g., rapid turn-taking between interlocutors, overlap between hearing the current speaker and preparing one's own vocal response, deciding exactly when to speak — Pickering and Garrod, 2009), the modulatory effects of stimulus type on self-voice perception are critical for efficient real-time communication processes. Hence, neural processing resources might be selectively and differentially focused on those aspects of a conversation that are deemed relevant.

Nevertheless, it should be noted that, in the current experiment, participants were instructed to passively listen to the voice stimuli (in contrast with previous studies in which self-voice perception was associated with an action, i.e. talking). Based on ERP and neuroimaging evidence showing differences in the neurofunctional mechanisms subserving self-voice perception in talking vs. passive listening conditions (e.g., Behroozmand et al., 2009; Behroozmand et al., 2015; Sitek et al., 2013), the generalization of the current findings to the realm of speech production is limited. For instance, self-voice feedback perception during speech production was found to recruit sensory-motor brain regions which are not activated when passively listening to the same

vocal sounds (Behroozmand et al., 2015; Parkinson et al., 2012; Zheng, 2010; Zheng et al., 2013). Therefore, future studies probing stimulus complexity effects during speech production are of the upmost importance for clarifying the neurofunctional mechanisms underlying self-voice perception.

Latency data of the N2 and P3 components indicated more difficult processing of words compared to vocalizations, irrespective of voice identity. The concurrent presence of verbal information might have increased perceptual demands, thus resulting in an increased time to categorize and to direct higher-order attention to relevant target stimuli. In other words, latency differences between the vocalizations and words may be related to a 'processing cost' associated with more complex linguistic stimuli, thus suggesting that the complexity of the voice signal influence the time course of voice identity processing. Effects of stimulus complexity on latency measures of ERP components have been previously demonstrated (Kok, 2001; Shucard et al., 2004).

4.2. Self-relevance facilitates abstract regularity representations

We also probed how representations of acoustic regularity are generated as a function of both speaker's identity and stimulus type, by examining the N1 and P2 amplitude to standard (i.e., expected) stimuli. We found that the self-voice elicited reduced N1 amplitude compared to the non-self voice, regardless of stimulus type. Decreased neural responsiveness to standards is thought to reflect a more accurate stimulus representation (Freedman et al., 2006; Ross and Tremblay, 2009; Seppänen et al., 2012) or reduced prediction error (Grotheer and Kovács, 2016; Summerfield et al., 2008). In the current study, the reduced N1 to the self-voice suggests that it was more efficiently processed than an unfamiliar voice. Considering that N1 and P2 amplitude modulations have been linked to the sensory registration and categorization of stimulus acoustic properties respectively (Ford et al., 2010), these findings suggested that self-relevant voice cues are associated with a facilitated analysis and categorization. These effects might be due to the stronger sensorimotor representation associated with one's own voice (Xu et al., 2013). Indeed, the sensorimotor features of the self-voice representation are critical for monitoring purposes, as internal predictive mechanisms (which operate when we vocalize or produce speech) build upon these learned associations to generate predictions about the likely sensory consequences of vocal motor commands (Hickok et al., 2011). The facilitated analysis and categorization of the physical properties of repeated self-voice stimuli were evident both in an early sensory-driven processing stage (N1) and in a more intermediate stage of stimulus classification (P2). This finding seems to be consistent with the N1 suppression effect observed in response to expected (i.e., predicted) self-voice feedback (compared to externally presented or distorted self-voice feedback) in voice production studies (e.g., Behroozmand et al., 2009; Behroozmand and Larson, 2011; Ford et al., 2010; Heinks-Maldonado et al., 2005), which has been linked to the operations of an internal predictive mechanism.

Nonetheless, we should note that regularity representations derived from stimulus repetition represent a process that differs from building regularities based on the invariant aspects of the unfolding speech signal. Hence, conventional oddball tasks may not represent an optimal paradigm for studying the neural mechanisms underlying voice regularity representations (Bendixen et al., 2007). Future experiments using more 'ecological' tasks with dynamic voice changes (i.e., distinct voice identities) are warranted.

Given that a great amount of studies on self-voice processing have used steady short vocalizations (e.g., Ford et al., 2010; Graux et al., 2013, 2015; Whitford et al., 2011), and since our experimental design does not allow dissociating effects of duration from effects of linguistic information, future experiments should clarify the role of stimulus complexity, disentangling the selective contribution of verbal (e.g., phonemic, phonological, semantic) and nonverbal (e.g., duration, loudness, mean F0) information. Future studies that orthogonally manipulate these cues and test their independent contribution to self-voice recognition and to self/non-self voice discrimination are necessary.

Furthermore, contrary to earlier experiments (e.g., Graux et al., 2013, 2015; Beauchemin et al., 2006), the current design included a single deviant voice per block. As our study aimed to probe how self- vs. non-self voices are perceived when in the focus of attention, and as participants had to mentally count the number of times a target stimulus was presented, employing more than one deviant per block would have significantly increased working memory load. Future studies using a more varied voice stimulus pool are thus necessary. Moreover, it is worth noting the small sample size used in our study. Future studies should include larger samples to decrease the likelihood of type II error. Even though our experimental stimuli preclude the generalization of findings to the huge variety of voice information we may encounter in our daily life, we expect that replicating results with a larger set of voice stimuli with reduced variance in acoustic and psycholinguistic cues may provide the basis for such generalized conclusions. Future studies employing a larger set of voice stimuli with reduced variance in their acoustic and psycholinguistic properties are needed to enhance the generalization of conclusions regarding self-voice perception mechanisms.

5. Conclusion

The current study examined the effects of self-relevance and stimulus type on selective attention to voices. The results showed that stimulus type affects the later, higher-order stages of voice identity processing: whereas self- and non-self vocalizations received a similar amount of attentional resources, more complex self-voice stimuli (words) captured more attentional resources than an unknown voice. Furthermore, our study demonstrated that acoustic regularities are more easily extracted and categorized when listeners are exposed to their self-voice, probably due to the stronger sensorimotor representation associated with one's own voice. These findings contribute to a better understanding of self-voice perception mechanisms by demonstrating that the underlying neural underpinnings are modulated by stimulus type and task instructions. These effects should be considered in future studies.

Acknowledgements

This work was supported by Grants IF/00334/2012 and PTDC/MHC-PCN/0101/2014 funded by *Fundação para a Ciência e a Tecnologia* (FCT, Portugal) and FEDER (*Fundo Europeu de Desenvolvimento Regional*) through the European programs QREN (*Quadro de Referência Estratégico Nacional*), and COMPETE (*Programa Operacional Factores de Competitividade*), awarded to A.P.P., and a by a FCT Postdoctoral Grant (SFRH/BPD/117187/2016) awarded to T.C. This work received additional support from Grant BIAL 238/16 funded by BIAL Foundation, awarded to A.P.P.

References

Allen, P.P., Amaro, E., Fu, C.H.Y., Williams, S.C.R., Brammer, M., Johns, L.C., McGuire, P.K., 2005. Neural correlates of the misattribution of self-generated speech. Hum. Brain Mapp. 26 (1), 44–53. https://doi.org/10.1002/hbm.20120.

Beauchemin, M., De Beaumont, L., Vannasing, P., Turcotte, A., Arcand, C., Belin, P., Lassonde, M., 2006. Electrophysiological markers of voice familiarity. Eur. J. Neurosci. 23, 3081–3086. https://doi.org/10.1111/j.1460-9568.2006.04856.x.

Behroozmand, R., Larson, C.R., 2011. Error-dependent modulation of speech-induced auditory suppression for pitch-shifted voice feedback. BMC Neurosci. 12 (1), 54. https://doi.org/10.1186/1471-2202-12-54.

Behroozmand, R., Karvelis, L., Liu, H., Larson, C.R., 2009. Vocalization-induced enhancement of the auditory cortex responsiveness during voice F0 feedback perturbation. Clin. Neurophysiol. 120 (7), 1303–1312.

Behroozmand, R., Shebek, R., Hansen, D.R., Oya, H., Robin, D.A., Howard, M.A., Greenlee, J.D., 2015. Sensory-motor networks involved in speech production and motor control: an fMRI study. NeuroImage 109, 418–428. https://doi.org/10.1016/j.

neuroimage.2015.01.040.
Belin, P., Fecteau, S., Bédard, C., 2004. Thinking the voice: neural correlates of voice perception. Trends Cogn Sci. 8 (3), 129–135. https://doi.org/10.1016/j.tics.2004.01.008.
Belin, P., Bestelmeyer, P.E., Latinus, M., Watson, R., 2011. Understanding voice perception. Br. J. Psychol. 102 (4), 711–725. https://doi.org/10.1111/j.2044-8295.2011.02041.x.
Bendixen, A., Roeber, U., Schröger, E., 2007. Regularity extraction and application in dynamic auditory stimulus sequences. J. Cogn. Neurosci. 19 (10), 1664–1677. https://doi.org/10.1162/jocn.2007.19.10.1664.
Berlad, I., Pratt, H., 1995. P300 in response to the subject's own name. Electroencephalogr. Clin. Neurophysiol. 96, 472–474.
Boersma, P., Weenink, D., 2012. Praat, Version 5.3. http://www.fon.hum.uva.nl/praat/.
Burnett, T.A., Freedland, M.B., Larson, C.R., Hain, T., 1998. Voice F0 responses to manipulations in pitch feedback. J. Acoust. Soc. Am. 103, 3153–3161. https://doi.org/10.1121/1.423073.
Canavarro, M.C., 1999. Inventário de Sintomas Psicopatológicos — BSI (Psychopathological symptoms inventory — BSI). In: Simões, Mário R., Gonçalves, M., Almeida, L.S. (Eds.), Testes e Provas Psicológicas em Portugal. 2. APPORT/SHO, Braga, pp. 96–109.
Canavarro, M.C. (2007). Inventário de Sintomas Psicopatológicos: Uma Revisão crítica dos estudos realizados em Portugal [Psychopathological Symptoms Inventory: A critical revision of the Portuguese studies]. In L. Almeida, M. Simões, C. Machado e M. Gonçalves (Eds.) Avaliação psicológica. Instrumentos validados para a população Portuguesa [Psychological Assessment: Tests Validated for the Portuguese Population] (Vol. 3, pp. 305-31). Coimbra: Quarteto Editora.
Chen, J., Yuan, J., Feng, T., Chen, A., Gu, B., Li, H., 2011. Temporal features of the degree effect in self-relevance: neural correlates. Biol. Psychol. 87 (2), 290–295.
Chen, Z., Jones, J.A., Liu, P., Li, W., Huang, D., Liu, H., 2013. Dynamics of vocalization-induced modulation of auditory cortical activity at mid-utterance. PloS One 8 (3), e60039. https://doi.org/10.1371/journal.pone.0060039.
Conde, T., Goncalves, O., Pinheiro, A.P., 2015. Paying attention to my voice or yours: an ERP study with words. Biol. Psychol. 111, 40–52. https://doi.org/10.1016/j.biopsycho.2015.07.014.
Conde, T., Gonçalves, O.F., Pinheiro, A.P., 2016a. A cognitive neuroscience view of voice-processing abnormalities in schizophrenia: a window into auditory verbal hallucinations? Harv. Rev. Psychiatry 24 (2), 148–163.
Conde, T., Gonçalves, O., Pinheiro, A.P., 2016b. The effects of stimulus complexity on the preattentive processing of self-generated and non-self voices: an ERP study. Cogn. Affect. Behav. Neurosci. 16 (1), 106–123. https://doi.org/10.3758/s13415-015-0376-1.
Cook, S., Wilding, J., 1997. Earwitness testimony: never mind the variety, hear the length. Appl. Cogn. Psychol. 11, 95–111.
Crowley, K.E., Colrain, I.M., 2004. A review of the evidence for P2 being an independent component process: age, sleep and modality. Clin. Neurophysiol. 115 (4), 732–744. https://doi.org/10.1016/j.clinph.2003.11.021.
Delplanque, S., Silvert, L., Hot, P., Rigoulot, S., Sequeira, H., 2006. Arousal and valence effects on event-related P3a and P3b during emotional categorization. Int. J. Psychophysiol. 60 (3), 315–322. https://doi.org/10.1016/j.ijpsycho.2005.06.006.
Domínguez-Borràs, J., Trautmann, S., Erhard, P., Fehr, T., Herrmann, M., Escera, C., 2009. Emotional context enhances auditory novelty processing in superior temporal gyrus. Cereb. Cortex 19, 1521–1529. https://doi.org/10.1093/cercor/bhn188.
Duncan, C.C., Barry, R.J., Connolly, J.F., Fischer, C., Michie, P.T., Näätänen, R., ... Van Petten, C., 2009. Event-related potentials in clinical research: guidelines for eliciting, recording, and quantifying mismatch negativity, P300, and N400. Clin. Neurophysiol. 120 (11), 1883–1908. https://doi.org/10.1016/j.clinph.2009.07.045.
Fan, W., Zhang, Y., Wang, X., Wang, X., Zhang, X., Zhong, Y., 2011. The temporal features of self-referential processing evoked by national flag. Neurosci. Lett. 505 (3), 233–237. https://doi.org/10.1016/j.neulet.2011.10.017.
Fan, W., Chen, J., Wang, X.-Y., Cai, R., Tan, Q., Chen, Y., ... Zhong, Y., 2013. Electrophysiological correlation of the degree of self-reference effect. PloS One 8 (12), e80289. https://doi.org/10.1371/journal.pone.0080289.
Field, A.P., 2013. Discovering Statistics Using IBM SPSS Statistics, Fourth edition. Sage Publications, London.
Fleming, D., Giordano, B.L., Caldara, R., Belin, P., 2014. A language-familiarity effect for speaker discrimination without comprehension. Proc. Natl. Acad. Sci. U. S. A. 111 (38), 13795–13798.
Ford, J.M., Roach, B.J., Mathalon, D.H., 2010. How to assess the corollary discharge in humans using noninvasive neurophysiological methods. Nat. Protoc. 5 (6), 1160–1168. https://doi.org/10.1038/nprot.2010.67.
Freedman, D.J., Riesenhuber, M., Poggio, T., Miller, E.K., 2006. Experience-dependent sharpening of visual shape selectivity in inferior temporal cortex. Cerebral Cortex 16, 1631–1644.
Friedman, D., Cycowicz, Y.M., Gaeta, H., 2001. The novelty P3: an event-related brain potential (ERP) sign of the brain's evaluation of novelty. Neurosci. Biobehav. Rev. 25 (4), 355–373. https://doi.org/10.1016/S0149-7634(01)00019-7.
Fritz, J.B., Elhilali, M., David, S.V., Shamma, S.A., 2007a. Auditory attention—focusing the searchlight on sound. Curr. Opin. Neurobiol. 17 (4), 437–455.
Fritz, J.B., Elhilali, M., David, S.V., Shamma, S.A., 2007b. Does attention play a role in dynamic receptive field adaptation to changing acoustic salience in A1? Hear. Res. 229 (1), 186–203. https://doi.org/10.1016/j.heares.2007.01.009.
Gratton, G., Coles, M.G., Donchin, E., 1983. A new method for off-line removal of ocular artifact. Electroencephalogr. Clin. Neurophysiol. 55 (4), 468–484.
Graux, J., Gomot, M., Roux, S., Bonnet-Brilhault, F., Camus, V., Bruneau, N., 2013. My voice or yours? An electrophysiological study. Brain Topogr. 26 (1), 72–82.
Graux, J., Gomot, M., Roux, S., Bonnet-Brilhault, F., Bruneau, N., 2015. Is my voice just a familiar voice? An electrophysiological study. Soc. Cogn. Affect. Neurosci. 10, 101–105.
Gray, H.M., Ambady, N., Lowenthal, W.T., Deldin, P., 2004. P300 as an index of attention to self-relevant stimuli. J. Exp. Soc. Psychol. 40, 216–224. https://doi.org/10.1016/S0022-1031(03)00092-1.
Grill-Spector, K., Henson, R., Martin, A., 2006. Repetition and the brain: neural models of stimulus-specific effects. Trends Cogn. Sci. 10 (1), 14–23.
Grotheer, M., Kovács, G., 2016. Can predictive coding explain repetition suppression? Cortex 80, 113–124. https://doi.org/10.1016/j.cortex.2015.11.027.
Heinks-Maldonado, T.H., Mathalon, D.H., Gray, M., Ford, J.M., 2005. Fine-tuning of auditory cortex during speech production. Psychophysiology 42 (2), 180–190. https://doi.org/10.1111/j.1469-8986.2005.00272.x.
Hickok, G., Houde, J., Rong, F., 2011. Sensorimotor integration in speech processing: computational basis and neural organization. Neuron 69 (3), 407–422. https://doi.org/10.1016/j.neuron.2011.01.019.
Hu, H., Liu, Y., Guo, Z., Li, W., Liu, P., Chen, S., Liu, H., 2015. Attention modulates cortical processing of pitch feedback errors in voice control. Sci. Rep. 5, 7812. https://doi.org/10.1038/srep07812.
Jacobsen, T., Schröger, E., Winkler, I., Horváth, J., 2005. Familiarity affects the processing of task-irrelevant auditory deviance. J. Cogn Neurosci. 17 (11), 1704–1713. https://doi.org/10.1162/089892905774589262.
Kaganovich, N., Francis, A.L., Melara, R.D., 2006. Electrophysiological evidence for early interaction between talker and linguistic information during speech perception. Brain Res. 1114, 161–172. https://doi.org/10.1016/j.brainres.2006.07.049.
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. Soc. Cogn. Affect. Neurosci. 3, 218–223. https://doi.org/10.1093/scan/nsn014.
Kayser, J., Bruder, G.E., Tenke, C.E., Stuart, B.K., Amador, X.F., Gorman, J.M., 2001. Event-related brain potentials (ERPs) in schizophrenia for tonal and phonetic oddball tasks. Biol. Psychiatry 49 (10), 832–847. https://doi.org/10.1016/S0006-3223(00)01090-8.
Kok, A., 2001. On the utility of P3 amplitude as a measure of processing capacity. Psychophysiology 38 (3), 557–577. https://doi.org/10.1017/S0048577201990559.
Kreiman, J., Sidtis, D., 2013. Foundations of Voice Studies: An Interdisciplinary Approach to Voice Production and Perception. Wiley-Blackwell, Boston.
Lane, H., Webster, J.W., 1991. Speech deterioration in postlingually deafened adults. J. Acoust. Soc. Am. 89 (2), 859–866. https://doi.org/10.1121/1.1894647.
Latinus, M., Belin, P., 2011. Human voice perception. Curr. Biol. 21 (4), R143–R145. https://doi.org/10.1016/j.cub.2010.12.033.
Liu, H., Wang, E.Q., Chen, Z., Liu, P., Larson, C.R., Huang, D., 2010. Effect of tonal native language on voice fundamental frequency responses to pitch feedback perturbations during vocalization. J. Acoust. Soc. Am. 128 (6), 3739–3746. https://doi.org/10.1121/1.3500675.
Luck, S.J., 2005. An Introduction to the Event-related Potential Technique. The MIT Press, Cambridge, MA.
Luck, S.J., Heinze, H.J., Mangun, G.R., Hillyard, S.A., 1990. Visual event-related potentials index focused attention within bilateral stimulus arrays. II. Functional dissociation of P1 and N1 components. Electroencephalogr. Clin. Neurophysiol. 75, 528–542.
Maurer, D., Landis, T., 1990. Role of bone conduction in the self-perception of speech. Folia Phoniatr. Logop. 42 (5), 226–229.
Moeller, M.P., Hoover, B., Putman, C., Arbataitis, K., Bohnenkamp, G., Peterson, B., ... Stelmachowicz, P., 2007. Vocalizations of infants with hearing loss compared with infants with normal hearing: part II — transition to words. Ear Hear. 28 (5), 628–642. https://doi.org/10.1097/AUD.0b013e31812564c9.
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.
Nygaard, L.C., Pisoni, D.B., 1998. Talker-specific learning in speech perception. Percept. Psychophys. 60 (3), 355–376.
O'Donnell, B.F., Shenton, M.E., McCarley, R.W., Faux, S.F., Smith, R.S., Salisbury, D.F., ... Jolesz, F., 1993. The auditory N2 component in schizophrenia: relationship to MRI temporal lobe gray matter and to other ERP abnormalities. Biol. Psychiatry 34 (1-2), 26–40.
Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9, 97–113. https://doi.org/10.1016/0028-3932(71)90067-4.
Özgürdal, S., Gudlowski, Y., Witthaus, H., Kawohl, W., Uhl, I., Hauser, M., ... Juckel, G., 2008. Reduction of auditory event-related P300 amplitude in subjects with at-risk mental state for schizophrenia. Schizophrenia Research 105 (1), 272–278.
Parkinson, A.L., Flagmeier, S.G., Manes, J.L., Larson, C.R., Rogers, B., Robin, D.A., 2012. Understanding the neural mechanisms involved in sensory control of voice production. NeuroImage 61, 314–322. https://doi.org/10.1016/j.neuroimage.2012.02.068.
Patel, S.H., Azzam, P.N., 2005. Characterization of N200 and P300: selected studies of the event-related potential. Int. J. Med. Sci. 2 (4), 147–154.
Patel, R., Schell, K.W., 2008. The influence of linguistic content on the Lombard effect. J. Speech Lang. Hear. Res. 51, 209–220.
Pell, M.D., Kotz, S.A., 2011. On the time course of vocal emotion recognition. PLoS One 6 (11), e27256. https://doi.org/10.1371/journal.pone.0027256.
Perrin, F., García-Larrea, L., Mauguière, F., Bastuji, H., 1999. A differential brain response to the subject â€TM s own name persists during sleep. Clin. Neurophysiol. 110, 2153–2164. https://doi.org/10.1016/s1388-2457(99)00177-7.
Perrin, F., Maquet, P., Peigneux, P., Ruby, P., Degueldre, C., Balteau, E., Laureys, S., 2005. Neural mechanisms involved in the detection of our first name: a combined ERPs and PET study. Neuropsychologia 43 (1), 12–19.
Pickering, M.J., Garrod, S., 2009. Prediction and embodiment in dialogue. Eur. J. Soc.

Psychol. 39 (7), 1162–1168. https://doi.org/10.1002/ejsp.663.
Pinheiro, A.P., Del Re, E., Mezin, J., Nestor, P.G., Rauber, A., McCarley, R.W., ... Niznikiewicz, M.A., 2013. Sensory-based and higher-order operations contribute to abnormal emotional prosody processing in schizophrenia: an electrophysiological investigation. Psychol. Med. 43 (3), 603–618. https://doi.org/10.1017/S003329171200133X.
Pinheiro, A.P., Vasconcelos, M., Dias, M., Arrais, N., Gonçalves, O.F., 2015. The music of language: an ERP investigation of the effects of musical training on emotional prosody processing. Brain Lang. 140, 24–34. https://doi.org/10.1016/j.bandl.2014.10.009.
Pinheiro, A.P., Rezaii, N., Rauber, A., Niznikiewicz, M.A., 2016. Is this my voice or yours? The role of emotion and acoustic quality in self-other voice discrimination in schizophrenia. Cogn. Neuropsychiatry 21 (4), 335–353. https://doi.org/10.1080/13546805.2016.1208611.
Pinheiro, A.P., Barros, C., Vasconcelos, M., Kotz, S.A., 2017. Is laughter a better vocal change detector than a growl? Cortex 92, 233–248. https://doi.org/10.1016/j.cortex.2017.03.018.
Polich, J., 2007. Updating P300: an integrative theory of P3a and P3b. Clin. EEG Neurosci. 118 (10), 2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019.
Polich, J., Criado, J.R., 2006. Neuropsychology and neuropharmacology of P3a and P3b. Int. J. Psychophysiol. 60 (2), 172–185.
Ranganath, C., Rainer, G., 2003. Neural mechanisms for detecting and remembering novel events. Nat. Rev. Neurosci. 4 (3), 193–202. https://doi.org/10.1038/nrn1052.
Remez, R.E., Fellowes, J.M., Rubin, P.E., 1997. Talker identification based on phonetic information. J. Exp. Psychol. 23 (3), 651–666.
Rimmele, J.M., Golumbic, E.Z., Schröger, E., Poeppel, D., 2015. The effects of selective attention and speech acoustics on neural speech-tracking in a multi-talker scene. Cortex 68, 144–154. https://doi.org/10.1016/j.cortex.2014.12.014.
Roebuck, R., Wilding, J., 1993. Effects of vowel variety and sample length on identification of a speaker in a line-up. Appl. Cogn. Psychol. 7 (6), 475–481.
Rosa, C., Lassonde, M., Pinard, C., Keenan, J.P., Belin, P., 2008a. Investigations of hemispheric specialization of self-voice recognition. Brain Cogn. 68, 204–214.
Rosa, C., Lassonde, M., Pinard, C., Keenan, J.P., Belin, P., 2008b. Investigations of hemispheric specialization of self-voice recognition. Brain Cogn. 68, 204–214.
Ross, B., Tremblay, K., 2009. Stimulus experience modifies auditory neuromagnetic responses in young and older listeners. Hear. Res. 248 (1), 48–59. https://doi.org/10.1016/j.heares.2008.11.012.
Roye, A., 2010. Is my mobile ringing? Evidence for rapid processing of a personally significant sound in humans. J. Neurosci. 30 (21), 7310–7313.
Roye, A., Jacobsen, T., Schröger, E., 2007. Personal significance is encoded automatically by the human brain: an event-related potential study with ringtones. Eur. J. Neurosci. 26, 784–790.
Salisbury, D.F., O'Donnell, B.F., Mccarley, R.W., Shenton, M.E., Benavage, A., 1994. The N2 event-related potential reflects attention in schizophrenia deficit. Biol. Psychiatry 39, 1–13.
Salisbury, D.F., Collins, K.C., McCarley, R.W., 2010. Reductions in the N1 and P2 auditory event-related potentials in first-hospitalized and chronic schizophrenia. Schizophr. Bull. 36 (5), 991–1000. https://doi.org/10.1093/schbul/sbp003.
Schirmer, A., Escoffier, N., 2010. Emotional MMN: anxiety and heart rate correlate with the ERP signature for auditory change detection. Clin. Neurophysiol. 121 (1), 53–59. https://doi.org/10.1016/j.clinph.2009.09.029.
Schirmer, A., Striano, C.A.T., Friederici, A.D., 2005. Sex differences in the preattentive processing of vocal emotional expressions. Neuroreport 16 (6), 635–639. https://doi.org/10.1097/00001756-200504250-00024.
Schweinberger, S.R., Herholz, A., Sommer, W., 1997. Recognizing famous voices influence of stimulus duration and different types of retrieval cues. J. Speech Lang. Hear. Res. 40 (2), 453–463. https://doi.org/10.1044/jslhr.4002.453.
Schweinberger, S.R., Walther, C., Zäske, R., Kovács, G., 2011. Neural correlates of adaptation to voice identity. Br. J. Psychol. 102 (4), 748–764.
Schweinberger, S.R., Kawahara, H., Simpson, A.P., Skuk, V.G., Zäske, R., 2014. Speaker perception. Wiley Interdiscip. Rev. 5 (1), 15–25. https://doi.org/10.1002/wcs.1261.
Scott, L.S., Luciana, M., Wewerka, S., 2005. Electrophysiological correlates of facial self-recognition in adults and children. Cognit. Brain Behav. 9 (3), 211–238.
Seppänen, M., Hämäläinen, J., Pesonen, A.-K., Tervaniemi, M., 2012. Music training enhances rapid neural plasticity of N1 and P2 source activation for unattended sounds. Front. Human Neurosci. 6 (43), 1–13. https://doi.org/10.3389/fnhum.2012.00043.
Shucard, D.W., Abara, J.P., McCabe, D.C., Benedict, R.B., Shucard, J.L., 2004. The effects of covert attention and stimulus complexity on the P3 response during an auditory continuous performance task. Int. J. Psychophysiol. 54 (3), 221–230. https://doi.org/10.1016/j.ijpsycho.2004.04.007.
Sidtis, D., Kreiman, J., 2012. In the beginning was the familiar voice: personally familiar voices in the evolutionary and contemporary biology of communication. Integr. Psychol. Behav. Neurosci. 46 (2), 146–159. https://doi.org/10.1007/s12124-011-9177-4.
Sitek, K., Mathalon, D.H., Roach, B.J., Houde, J.F., Niziolek, C., Ford, J.M., 2013. Auditory cortex processes variation in our own speech. PloS One 8 (12), e82925. https://doi.org/10.1371/journal.pone.0082925.
Soares, A.P., Comesaña, M., Iriarte, A., Almeida, J.J., Simões, A., Costa, A., et al., 2010. P-PAL: uma base lexical com índices psicolinguísticos do Português Europeu (P-PAL: a European Portuguese lexical database). Linguamática 2 (3), 67–72.
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. Behav. Res. Methods 44 (1), 256–269. https://doi.org/10.3758/s13428-011-0131-7.
Spencer, K.M., Dien, J., Donchin, E., 1999. A componential analysis of the ERP elicited by novel events using a dense electrode array. Psychophysiology 36 (3), 409–414.
Spencer, K.M., Dien, J., Donchin, E., 2001. Spatiotemporal analysis of the late ERP responses to deviant stimuli. Psychophysiology 38 (2), 343–358.
Su, Y., Chen, A., Yin, H., Qiu, J., Lv, J., Wei, D., Wang, T., 2010. Spatiotemporal cortical activation underlying self-referential processing evoked by self-hand. Biol. Psychol. 85 (2), 219–225. https://doi.org/10.1016/j.biopsycho.2010.07.004.
Sui, J., Zhu, Y., Han, S., 2006. Self-face recognition in attended and unattended conditions: an event-related brain potential study. Neuroreport 17 (4), 423–427.
Summerfield, C., Trittschuh, E.H., Monti, J.M., Mesulam, M.M., Egner, T., 2008. Neural repetition suppression reflects fulfilled perceptual expectations. Nat. Neurosci. 11, 1004–1006. https://doi.org/10.1038/nn.2163.
Sussman, E.S., 2007. A new view on the MMN and attention debate: the role of context in processing auditory events. J. Psychophysiol. 21 (3), 164–175.
Symons, C.S., Johnson, B.T., 1997. The self-reference effect in memory: a meta-analysis. Psychol. Bull. 121, 371–394.
Tacikowski, P., Nowicka, A., 2010. Allocation of attention to self-name and self-face: an ERP study. Biol. Psychol. 84, 318–324. https://doi.org/10.1016/j.biopsycho.2010.03.009.
Tacikowski, P., Cygan, H.B., Nowicka, A., 2014. Neural correlates of own and close-other's name recognition: ERP evidence. Front. Human Neurosci. 8 (194), 1–10. https://doi.org/10.3389/fnhum.2014.00194.
Tumber, A.K., Scheerer, N.E., Jones, J.A., 2014. Attentional demands influence vocal compensations to pitch errors heard in auditory feedback. PloS One 9 (10), e109968.
Ventura, M.I., Nagarajan, S.S., Houde, J.F., 2009. Speech target modulates speaking induced suppression in auditory cortex. BMC Neurosci. 10 (58), 1–11. https://doi.org/10.1186/1471-2202-10-58.
Waters, F., Woodward, T., Allen, P., Aleman, A., Sommer, I., 2012. Self-recognition deficits in schizophrenia patients with auditory hallucinations: a meta-analysis of the literature. Schizophr. Bull. 38 (4), 741–750. https://doi.org/10.1093/schbul/sbq144.
Whitford, T.J., Mathalon, D.H., Shenton, M.E., Roach, B.J., Bammer, R., Adcock, R.A., ... Ford, J.M., 2011. Electrophysiological and diffusion tensor imaging evidence of delayed corollary discharges in patients with schizophrenia. Psychol. Med. 41 (5), 959–969. https://doi.org/10.1017/S0033291710001376.
Xu, M., Homae, F., Hashimoto, R., Hagiwara, H., 2013. Acoustic cues for the recognition of self-voice and other-voice. Front. Psychol. 4 (735), 1–7. https://doi.org/10.3389/fpsyg.2013.00735.
Zarate, J.M., Tian, X., Woods, K.J., Poeppel, D., 2015. Multiple levels of linguistic and paralinguistic features contribute to voice recognition. Sci. Rep. 19 (5), 11475. https://doi.org/10.1038/srep11475.
Zheng, Z.Z., 2010. Functional overlap between regions involved in speech perception and in monitoring one's own voice during speech production. J. Cogn. Neurosci. 22, 1770–1781. https://doi.org/10.1162/jocn.2009.21324.
Zheng, Z.Z., Vicente-Grabovetsky, A., MacDonald, E.N., Munhall, K.G., Cusack, R., 2013. Multivoxel patterns reveal functionally differentiated networks underlying auditory feedback processing of speech. J. Neurosci. 33, 4339–4348. https://doi.org/10.1523/JNEUROSCI.6319-11.2013.