

The effects of stimulus complexity on the preattentive processing of self-generated and nonself voices: An ERP study

Tatiana Conde¹ · Óscar F. Gonçalves^{1,2,3} · Ana P. Pinheiro^{1,4}

Published online: 28 September 2015
© Psychonomic Society, Inc. 2015

Abstract The ability to differentiate one's own voice from the voice of somebody else plays a critical role in successful verbal self-monitoring processes and in communication. However, most of the existing studies have only focused on the sensory correlates of self-generated voice processing, whereas the effects of attentional demands and stimulus complexity on self-generated voice processing remain largely unknown. In this study, we investigated the effects of stimulus complexity on the preattentive processing of self and nonself voice stimuli. Event-related potentials (ERPs) were recorded from 17 healthy males who watched a silent movie while ignoring prerecorded self-generated (SGV) and nonself (NSV) voice stimuli, consisting of a vocalization (vocalization category condition: VCC) or of a disyllabic word (word category condition: WCC). All voice stimuli were presented as standard and deviant events in four distinct oddball sequences. The mismatch negativity (MMN) ERP component peaked earlier for NSV than for SGV stimuli. Moreover, when compared with SGV stimuli, the P3a amplitude was increased for NSV stimuli in the VCC only, whereas in the WCC no

significant differences were found between the two voice types. These findings suggest differences in the time course of automatic detection of a change in voice identity. In addition, they suggest that stimulus complexity modulates the magnitude of the orienting response to SGV and NSV stimuli, extending previous findings on self-voice processing.

Keywords Self-generated voice · Nonself voice · Stimulus complexity · Event-related potentials · Mismatch negativity · P3a

Given the highly dynamic nature of communication, speakers frequently face the challenge of monitoring their vocalizations while directing attention to other aspects of the social environment (e.g., attending to the emotional reaction of a social partner in response to self-generated speech, or talking on the phone with a relative while paying attention to the TV). Even in these challenging situations, speakers automatically and effortlessly identify the voice they are hearing as their own (Keenan, Falk, & Gallup, 2003; Xu, Homae, Hashimoto, & Hagiwara, 2013).

This apparently simple ability is thought to play a critical role in the efficient identification and correction of vocal production errors that speakers have to perform in order to adjust to the challenges of the social acoustic environment (Behroozmand & Larson, 2011; Brumm & Zollinger, 2011; Burnett, Freedland, Larson, & Hain, 1998; S. H. Chen, Liu, Xu, & Larson, 2007; Z. Chen et al., 2013; Liu & Larson, 2007; Liu, Meshman, Behroozmand, & Larson, 2011; Sitek et al., 2013; Sugimori, Asai, & Tanno, 2013). Abnormalities in the detection of self-generated voice feedback during speech production disrupt verbal communication processes (Lane & Webster, 1991; Moeller et al., 2007; Oller & Eilers, 1988; Schauwers et al., 2004). Moreover, when individuals are

✉ Ana P. Pinheiro
ana.pinheiro@psi.uminho.pt

¹ Neuropsychophysiology Lab, CIPsi, School of Psychology, University of Minho, 4710-057 Braga, Portugal

² Spaulding Center of Neuromodulation, Department of Physical Medicine & Rehabilitation, Spaulding Rehabilitation Hospital & Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA

³ Bouvé College of Health Sciences, Northeastern University, Boston, MA, USA

⁴ Cognitive Neuroscience Lab, Department of Psychiatry, Harvard Medical School, Boston, MA, USA

presented with distorted feedback of their own voice during speech production (e.g., decrement of voice fundamental frequency: F0), they exhibit a subsequent compensatory vocal response in the reverse direction of the manipulation (e.g., increase of voice F0; Behroozmand, Korzyukov, Sattler, & Larson, 2012; Burnett et al., 1998; S. H. Chen et al., 2007; Larson, Altman, Liu, & Hain, 2008; Letowski, Frank, & Caravella, 1993; H. Liu & Larson, 2007; Liu, Xu, & Larson, 2009; H. Liu et al., 2011; P. Liu, Chen, Jones, Huang, & Liu, 2011; Patel & Schell, 2008; Pittman & Wiley, 2001). These findings suggest the continuous operation of a compensatory vocal mechanism dedicated to the control of voice F0. Additionally, they demonstrate the existence of regulatory mechanisms that modulate voice physical parameters during vocalization in response to any disturbance in voice F0 feedback, playing a critical role in error detection and correction during speech production. Abnormalities in self-generated voice perception have been reported in neuropsychiatric disorders such as schizophrenia and have been proposed to underlie the experience and maintenance of auditory hallucinations (Allen et al., 2004; Allen et al., 2007; Ford, Mathalon, Whitfield, Faustman, & Roth, 2002; Ford, Roach, Faustman, & Mathalon, 2007; Ilankovic et al., 2011; Waters & Badcock, 2010; Waters, Woodward, Allen, Aleman, & Sommer, 2012; see Conde, Gonçalves, & Pinheiro, *in press*, for a review). Hence, a thorough comprehension of the neurofunctional underpinnings of self-voice perception is crucial.

Evidence from event-related potentials (ERPs) and magnetoencephalography (MEG) studies, using the same experimental approach, brings further support for the relationship between self-generated voice detection and efficient vocal production. In these studies, the N1 ERP component (or its magnetic counterpart, the N100m) has been used to examine the auditory cortical responsiveness during voice production relative to the passive listening to one's own voice. This component indexes the sensory registration of auditory stimuli (Ford, Roach, & Mathalon, 2010; Whitford et al., 2011), and it is thought to be generated in the auditory cortex (Godey, Schwartz, de Graaf, Chauvel, & Liégeois-Chauvel, 2001). These studies have found an N1 amplitude suppression effect during vocal production when compared with the condition of passively listening to prerecorded self-generated voice stimuli (Behroozmand & Larson, 2011; Curio, Neuloh, Numminen, Jousmaki, & Hari, 2000; Ford & Mathalon, 2004, 2005; Ford et al., 2001; Ford et al., 2002; Ford et al., 2007; Heinks-Maldonado, Mathalon, Gray, & Ford, 2005; Heinks-Maldonado et al., 2007; Houde, Nagarajan, Sekihara, & Merzenich, 2002; H. Liu et al., 2011; Numminen, Salmelin, & Hari, 1999; Sitek et al., 2013; Timm, SanMiguel, Saupe, & Schröger, 2013; Ventura, Nagarajan, & Houde, 2009). Moreover, when participants receive altered self-generated voice feedback, the N1 attenuation effect is smaller than when they listen to intact voice feedback (Behroozmand & Larson,

2011; Ford & Mathalon, 2004, 2005; Ford et al., 2001; Ford et al., 2007; Heinks-Maldonado et al., 2005; Heinks-Maldonado et al., 2007; Sitek et al., 2013; Timm et al., 2013). This suggests that the N1 component is also a signature of an internal predictive mechanism: When the prediction of the upcoming self-generated vocal sound is violated by the mismatching voice feedback, an error signal is generated and, as a consequence, amplitude attenuation is smaller (Behroozmand & Larson, 2011; Ford & Mathalon, 2004, 2005; Ford et al., 2001; Ford et al., 2007; Heinks-Maldonado et al., 2005; Heinks-Maldonado et al., 2007; Sitek et al., 2013).

The greater N1 amplitude reduction to intact than to altered feedback of one's own voice is thought to reflect an internal feedforward system. During vocalization, the frontal brain regions responsible for speech production send an efference copy of the speech motor plan to auditory sensory regions (Behroozmand & Larson, 2011; Heinks-Maldonado et al., 2005; Heinks-Maldonado et al., 2007; Sitek et al., 2013). During vocal production, the incoming voice feedback is compared against the predicted sensory consequences of the speech motor plan: If they match closely, auditory cortical responsiveness is suppressed (Behroozmand & Larson, 2011; Heinks-Maldonado et al., 2005; Heinks-Maldonado et al., 2007; Houde et al., 2002; Ventura et al., 2009; Sitek et al., 2013). Therefore, this system allows for the differentiation between self-generated and external stimulation (Behroozmand & Larson, 2011; Ford & Mathalon, 2004; Ford et al., 2007; Heinks-Maldonado et al., 2005; Heinks-Maldonado et al., 2007). On the contrary, if the voice feedback and the predicted sensory consequences do not match, an error signal is generated, and hence auditory cortical suppression is reduced (i.e., auditory activity is increased; Behroozmand, Karvelis, Liu, & Larson, 2009; Behroozmand & Larson, 2011; Eliades & Wang, 2008; Heinks-Maldonado et al., 2005; Heinks-Maldonado et al., 2007; Sitek et al., 2013). Complementing these studies, recent evidence (Tian & Poeppel, 2013, 2015) has shown that the operation of the internal feedforward system is not an "artifact" of overt vocal production, since it operates during articulation imagery.

Yet the majority of these studies have only examined the sensory correlates (indexed by the N1 or N100m) of one's own voice processing during vocalization (e.g., Ford et al., 2001; Ford et al., 2007; Ford & Mathalon, 2004, 2005; Heinks-Maldonado et al., 2005; Heinks-Maldonado et al., 2007; Sitek et al., 2013); more recent ERP evidence shows that distinct stages of information processing are differentially sensitive to online manipulations of voice feedback during speech production (Behroozmand, Ibrahim, Korzyukov, Robin, & Larson, 2014; Behroozmand et al., 2009). For instance, experimentally induced F0 perturbations in voice feedback that are unexpectedly delivered after 500 ms of vocalization onset elicit greater P1 and P2 amplitudes than passively

listening to the same self-generated voice stimuli (Behroozmand et al., 2009). This suggests that more neural resources are devoted to the processing of F0 disturbances in one's own voice during speaking than during listening, plausibly as a consequence of motor-driven predictions (Behroozmand et al., 2009). This might be especially relevant for online error detection/correction of one's own voice during communication processes (Behroozmand et al., 2009). Highlighting the evolutionary relevance of vocal self-monitoring processes, Eliades and Wang (2008) reported similar suppressive and enhancement effects during vocal production in the auditory cortex of marmoset monkeys. By directly recording single neurons in the auditory cortex, this study showed that those neural cells that were normally suppressed by intact voice feedback during vocalization also increased their activity to online disturbances in voice feedback (i.e., both increments and decrements in F0). Importantly, the sensitivity of auditory cortex neurons to changes in voice F0 only occurred during vocalization, as compared with the passive listening to the same sounds (Eliades & Wang). This finding suggests that the vocal production system plays a critical role in self-monitoring processes.

In addition, studies probing the neural response to voice F0 perturbation in musicians suggest that musical expertise is associated with heightened sensitivity to detect and correct online deviations in voice feedback during speech production (Behroozmand et al., 2014; Parkinson et al., 2014). For instance, as compared with nonmusicians, musicians exhibited faster compensatory vocal responses to unexpected deviations in voice F0 during speech production (Behroozmand et al., 2014). Furthermore, musicians exhibit increased neural sensory responsiveness to unexpected manipulations in voice F0 during vocal production, as reflected by increased N1 and P2 amplitudes (Behroozmand et al., 2014) and by decreased functional connectivity between brain regions involved in the vocal self-monitoring system (i.e., left premotor and superior temporal gyrus [STG] regions; Parkinson et al., 2014). Together, these findings suggest that vocal monitoring mechanisms are modulated by auditory expertise underlying extensive musical training.

An important contribution to the comprehension of the neurocognitive correlates of self-generated voice processing in later information-processing stages comes from two recent ERP studies investigating both the preattentive deviance detection and the orienting response to one's own versus a non-self voice (Graux, Gomot, Roux, Bonnet-Brilhault, & Bruneau, 2015; Graux et al., 2013). In these experiments, participants were asked to pay attention to a silent movie and to ignore the concurrent acoustic stimulation, consisting of recordings of their own voices, as well as voices from unknown (Graux et al., 2013) and familiar (Graux et al., 2015) speakers. These studies have the advantage of shedding light on brain responses to voice information, irrespective of the participants' attentional focus, avoiding potential

confounds resulting from attentional biases (Garrido, Kilner, Stephan, & Friston, 2009; Näätänen, 2001; Näätänen, Paavilainen, Rinne, & Alho, 2007; Pulvermüller & Shtyrov, 2006). The authors reported no differences in the mismatch negativity (MMN) responses to self-generated versus nonself voice stimuli (Graux et al., 2015; Graux et al., 2013). However, the authors observed a reduced reorienting response to one's own voice, in comparison with both unknown (Graux et al., 2013) and familiar (Graux et al., 2015) voice stimuli, reflected by smaller P3a amplitudes to the self-generated voice. These findings were interpreted as evidence for the prioritized processing of nonself voice stimuli when compared with a self-generated voice, in a context in which individuals' attention was engaged in a concurrent visual task.

However, it is worth noting that the voice stimuli used in the studies of Graux et al. (2013; Graux et al., 2015) were short vocalizations (/a/) lasting 300 ms. Previous studies have demonstrated an important interplay between stimulus complexity and voice identity processing (Cook & Wilding, 1997; Fleming, Giordano, Caldara, & Belin, 2014; Goggin, Thompson, Strube, & Simental, 1991; Nygaard & Pisoni, 1998; Nygaard, Sommers, & Pisoni, 1994; Patel & Schell, 2008; Perrachione & Wong, 2007; Schweinberger, Herholz, & Sommer, 1997; Ventura et al., 2009). For instance, Patel and Schell found that, during vocal production, online compensatory vocal responses (after experimentally induced manipulations) were modulated by the content of word stimuli: During a 90-dB noise condition, significant increases were observed in the duration of content words referring to agents, objects, and locations, and in the F0s of agent words as compared with function words (e.g., prepositions, pronouns, and articles). Furthermore, a recent MEG study (Ventura et al., 2009) revealed evidence for a modulatory role of stimulus complexity in the auditory cortical suppression effect during vocal production, since the magnitude of the N100m amplitude reduction was greater for simpler (/a/) than for more complex and dynamic stimuli (/a-a-a/ and /a-a-a-a/). Additional support has come from a study manipulating the acoustic complexity of the feedback provided during vocal production versus passive listening to the same sounds (Behroozmand, Korzyukov, & Larson, 2011). In this study, disturbed feedback was introduced in the middle of (a) a self-generated vowel during voice production, (b) complex nonvocal sounds with the same F0 and harmonics cues as the self-generated voice, or (c) a simple tone with the same F0 as the participant's voice. Demonstrating that compensatory vocal motor mechanisms are especially tuned to correct for alterations in more complex vocal sounds, the magnitude of vocal compensation was increased to the self-generated voice, as compared with complex nonvocal and simple sounds. Relative to the passive-listening condition, N1 and P2 amplitudes were increased to F0 disturbances in self-generated voice feedback only. This suggests that during vocal production, the brain is tuned to detect and to correct F0 deviations in one's own voice. On the other hand,

in the passive-listening condition, F0 disturbances were more easily detected in the case of both complex vocal and nonvocal sounds than of simple tones, as is shown by increased P2 amplitude to the former. This study provided important insights as to how the acoustic complexity of vocal and nonvocal sounds modulates the amount of neural resources engaged during vocalization versus passive listening to self-generated vocalizations. Studies on nonself voice processing have revealed that other variables accounting for the complexity of the voice signal also interact with voice identity processing. In particular, accuracy in speaker recognition seems to strongly depend on language proficiency (Fleming et al., 2014; Goggin et al., 1991; Perrachione & Wong, 2007; Thompson, 1987) and the duration of voice stimuli (Cook & Wilding, 1997; Schweinberger et al., 1997). Taken together, these findings highlight the critical role that linguistic information available in the vocal signal might have on how voice identity is processed. However, to the best of our knowledge, no studies have yet examined the role of stimulus complexity during the preattentive processing of one's own voice versus an unfamiliar voice.

The aim of the present study was to investigate whether preattentive detection of changes in voice identity differs as a function of the type of vocal stimulus, using a high temporal resolution methodology—that is, the ERP. The MMN and P3a components were the focus of the analysis. In order to manipulate stimulus complexity, two stimulus categories—vocalization (VCC) and word (WCC) category conditions—were used that differed in terms of their durations (i.e., 300 vs. 501 ms) and phonetic variability (i.e., vowel vs. disyllabic word). Furthermore, recent studies have shown that listeners strongly rely on both perceived pitch (fundamental frequency: F0) and formant frequencies to extract voice identity information (Baumann & Belin, 2010; Belin, Bestelmeyer, Latinus, & Watson, 2011; Belin, Fecteau, & Bédard, 2004; Latinus & Belin, 2011, 2012; Latinus, McAleer, Bestelmeyer, & Belin, 2013; Schweinberger, Kawahara, Simpson, Skuk, & Zäske, 2014; Schweinberger, Walther, Zäske, & Kovács, 2011; Xu et al., 2013). The existing evidence suggests that each speaker's voice is coded within temporal voice-sensitive regions as a function of its physical acoustic deviation regarding an internal vocal prototype (Latinus & Belin, 2012; Latinus et al., 2013; Schweinberger et al., 2011): More acoustically distant voices are not only perceived as being more “distinctive” (Baumann & Belin, 2010; Latinus & Belin, 2011, 2012), but they also elicit increased activation within these temporal voice-sensitive regions (Latinus et al., 2013). Thus, considering the reported relationship between the physical features of the acoustic signal and the representation of speaker's identity, we probed whether the ERP correlates of automatic change detection and attention orienting to one's own voice were associated with the voice's acoustic properties.

Following previous evidence on self-generated voice processing (Graux et al., 2013, 2015), we predicted no differences

in automatic vocal change detection in the VCC, which would be reflected in similar MMN amplitudes to both self and nonself voice stimuli. Also, we expected a reduced orienting response to the self-generated relative to the nonself vocalization in the VCC, indexed by reduced P3a amplitudes (Graux et al., 2013, 2015). Additionally, if preattentive deviance detection and attention orienting depend upon the complexity of the vocal stimulus and its salience (self vs. nonself voice), then an identity by stimulus category interaction should be observed. In contrast, if these processes are not modulated by stimulus complexity, then the electrophysiological signatures of self-generated and nonself voice processing should be similar in the VCC and WCC conditions. Since we controlled for the physical differences between the voice stimuli (in line with earlier MMN studies: Graux et al., 2013; Graux et al., 2015; Leitman, Foxe, Sehatpour, Shpaner, & Javitt, 2009; Leitman, Sehatpour, Garidis, Gomez-Ramirez, & Javitt, 2011; X. Pang et al., 2014; Schirmer & Escoffier, 2010; Schirmer, Simpson, & Escoffier, 2007; Schirmer, Striano, & Friederici, 2005), using a “like from like” subtraction approach, we hypothesized a nonsignificant association between voice acoustic properties and ERP measures of self-generated voice processing. This approach aimed to ensure that the hypothesized differences in preattentive vocal change detection and attentional orienting toward self-generated versus nonself voice stimuli would occur independently of voice acoustic parameters, and instead would reflect the activation of a self-voice representation (Kaplan, Aziz-Zadeh, Uddin, & Iacoboni, 2008; Sugiura, 2013; Sugiura et al., 2014; Sugiura et al., 2008).

Method

Participants

A total of 17 right-handed native male speakers of European Portuguese participated in this study. Right handedness was assessed by using the Edinburgh Handedness Inventory (Oldfield, 1971). All participants met the following criteria: verbal intelligence quotient (IQ) above 90; no history of neurological illness, electroconvulsive treatment, or drug or alcohol abuse in the past year; no current medication for medical disorders that would affect electroencephalogram (EEG) morphology, as well as neurological and/or cognitive functioning consequences.

After hearing a detailed description of the study, all participants gave oral and written informed consent. The study protocol and consent form were reviewed and approved by the Institutional Review Board of the University of Minho.

Participants were screened in terms of psychopathological symptoms using the Portuguese version of the Brief Symptom Inventory (BSI; Derogatis & Spencer, 1982; Portuguese version: Canavarro, 1999).

Participants were excluded if they scored ≥ 1.7 in the Positive Symptoms Distress Index of the BSI (Canavaro, 2007) or if they reported a history of psychiatric disorder in themselves or in first-degree relatives (none were excluded). Participants' cognitive functioning was assessed with the Portuguese version of the Wechsler Adult Intelligence Scale (WAIS-III—Wechsler, 1997/2008; see Table 1).

Stimuli

At least one week prior to the ERP experiment, all participants underwent a voice recording session in a sound-attenuated room, using a portable digital recorder Roland R-26 and a Shure PG48 microphone, with a sampling rate of 44.1 kHz and 16-bit resolution. In this session, two stimulus categories—a disyllabic Portuguese word (/mesa/ [table]) and a vocalization (vowel /a/)—were recorded by each participant (*self-generated voice*: SGV). In order to reduce the variability in word pronunciation, all participants were instructed to carefully listen to each stimulus category (previously recorded by a middle-aged male voice and without regional accent) presented through a set of Sennheiser CX 300-II earphones, and then to reproduce the stimulus exactly as they had just heard. In addition, both stimulus categories were also recorded by an unknown middle-aged male (age = 42 years) who was also a native speaker of European Portuguese (*nonself voice*: NSV).

Following previous studies (e.g., Ford et al., 2007; Graux et al., 2015; Graux et al., 2013; Heinks-Maldonado, Nagarajan, & Houde, 2006; Sitek et al., 2013; Whitford et al. 2011), the vowel /a/ (duration = 300 ms) was selected for the VCC condition, thus allowing more direct comparisons with those experiments. For the WCC, a word was selected from the Affective Norms for English Words (ANEW; Soares, Comesaña, Pinheiro, Simões, & Frade, 2012) and P-PAL (Soares et al., 2010) sets, according to the following criteria: neutral valence (5.34), low arousal (3.36), high frequency (>100 per million), grammatical class (noun), and short extension (two syllables and four letters). An additional criterion for word selection was a stable syllabic structure (i.e., consonant–vowel–consonant–vowel), to reduce the variability in word pronunciation among participants.

Praat acoustic analysis software (Boersma & Weenink, 2012) was used to match the durations (300 ms for VCC, 501 ms for WCC) and intensities (root-mean-square amplitude [RMS] = 70 dB) of SGV and NSV stimuli. Acoustic noise was reduced using a Fourier-based noise reduction algorithm (noise reduction = 14 dB; frequency smoothing = 150 Hz; attack/decay time = 0.15 s) implemented in the Audacity 2.0.2 software (<http://audacity.sourceforge.net/>). The acoustic analysis of all vocal stimuli was performed using Praat software (see Table 2).

Table 1 Sociodemographic and cognitive characterization of the participants

	<i>M (SD)</i>
<i>Sociodemographic data</i>	
Age, years	29.94 (5.36)
Education, years	15.88 (2.74)
<i>Cognitive data*</i>	
Full scale IQ	125.24 (12.97)
Verbal IQ	127.94 (11.63)
Performance IQ	116.18 (13.44)

M = mean; *SD* = standard deviation. * WAIS-III (Wechsler, 1997/2008).

Procedure

Participants were tested in two experimental sessions (one per stimulus category: WCC and VCC) taking place on distinct days, with a minimum of 24 h separating them. Participants were comfortably seated at a distance of 100 cm from the computer monitor in an acoustically and electrically shielded room. Each condition comprised two experimental blocks, each one including a total of 1,050 standard ($p = .875$) and 150 deviant ($p = .125$) stimuli. In both experimental sessions, the NSV was the standard stimulus and the SGV was the deviant in one of the blocks, whereas in the other block the reverse was observed. The standard and deviant voices were binaurally presented in a pseudorandom order, with the restriction that at least two standard stimuli lay between two deviant sounds, in line with previous studies (e.g., Özgürdal et al., 2008; Schirmer & Escoffier, 2010; Schirmer et al., 2007; Shtyrov & Pulvermüller, 2002). The Presentation software (Neurobehavioral Systems, Albany, CA, USA) was used to control stimulus timing and presentation. Each trial comprised a voice stimulus (WCC duration = 501 ms, VCC duration = 300 ms), followed by a blank screen (duration = 700 ms). The stimulus onset asynchronies were 1,201 ms for the WCC and 1,000 ms for the VCC (see Fig. 1). The order of the experimental sessions and blocks was counterbalanced across participants. In all experimental blocks, participants were asked to pay attention to a silent movie and to ignore the auditory stimulation. At the end of each experimental block, they were asked to describe the movie they had just watched.

EEG data acquisition and analysis

The EEG was recorded with a 64-channel BioSemi Active Two System (www.biosemi.com/products.htm). The EEG was continuously sampled at 512 Hz, with a bandpass of 0.01–100 Hz, and stored for later analysis. Electrodes placed at the left and right temples (horizontal electrooculogram: EOG) and one below the left eye (vertical EOG) were used to

Table 2 Acoustic properties of the voice stimuli used in the event-related potential experiments

Participant	Stimulus category	Mean F0 (Hz)	Range F0 (Hz)		Formant frequencies (Hz)		
			Min	Max	F1	F2	F3
C2	WCC	108	87	119	426.59	1718.79	2327.15
	VCC	113	112	114	745.74	1321.30	2382.04
C6	WCC	94	88	103	529.07	1783.44	2507.28
	VCC	103	101	113	889.68	1273.62	2401.54
C7	WCC	84	80	89	438.11	1630.49	2615.41
	VCC	95	93	97	877.42	1303.62	2633.31
C8	WCC	113	101	124	494.88	1547.86	2634.91
	VCC	133	130	135	896.69	1440.13	2688.37
C9	WCC	90	81	101	369.76	1693.95	2621.18
	VCC	111	108	114	882.33	1402.58	2651.14
C11	WCC	94	72	104	510.86	1789.76	2595.27
	VCC	98	93	102	928.33	1795.61	2624.87
C12	WCC	96	87	110	446.14	1939.83	2762.03
	VCC	108	97	111	944.57	1419.27	2629.54
C13	WCC	97	86	109	453.66	1554.96	2352.01
	VCC	92	89	96	694.67	1389.80	2354.80
C14	WCC	97	74	105	502.96	1702.89	2559.64
	VCC	111	108	113	746.41	1565.74	2248.42
C15	WCC	102	84	117	477.38	1925.78	2463.82
	VCC	102	100	107	744.16	1211.77	2455.43
C17	WCC	110	97	123	475.84	2025.54	2717.16
	VCC	114	112	119	809.26	1347.60	2839.58
C18	WCC	90	73	99	483.48	1607.46	2257.80
	VCC	107	99	111	799.26	1256.90	2737.42
C19	WCC	120	103	143	461.62	1676.25	2627.96
	VCC	90	89	98	881.96	1399.53	2611.06
C21	WCC	94	86	107	545.13	1720.50	2636.03
	VCC	100	98	102	881.40	1398.46	2428.65
C22	WCC	153	122	178	439.10	1665.28	2386.77
	VCC	154	146	157	727.06	1381.43	2337.38
C23	WCC	139	121	163	430.67	1694.01	2742.58
	VCC	129	124	135	874.87	1453.36	2546.16
C24	WCC	123	108	150	484.01	1820.15	2616.12
	VCC	119	116	124	740.62	1491.93	2496.37
<i>M</i>		108.32	98.97	117.41	648.05	1569.11	2543.80
<i>(SD)</i>		(17.38)	(17.12)	(20.56)	(192.65)	(213.76)	(153.48)
NSV	WCC	99	87	106	531.07	1783.08	2582.09
	VCC	119	101	123	703.17	1428.34	2362.85

M = mean; *SD* = standard deviation; WCC = word category condition; VCC = vocalization category condition

monitor horizontal and vertical eye movements. In addition, electrodes were also placed on the left and right mastoids.

The EEG data were analyzed using the Brain Vision Analyzer 2.0.4 software (www.brainproducts.com). EEG data were referenced offline to the average of the left and right mastoids. Individual ERP epochs were created for each

stimulus type in each stimulus category condition, with a –100-ms prestimulus baseline and an 800-ms poststimulus duration. After subtracting the –100-ms prestimulus baseline, eye blinks were corrected using Gratton, Coles, and Donchin's (1983) method. Single epochs containing excessive eye blinks, movement artifacts, or amplifier blocking (± 100 -

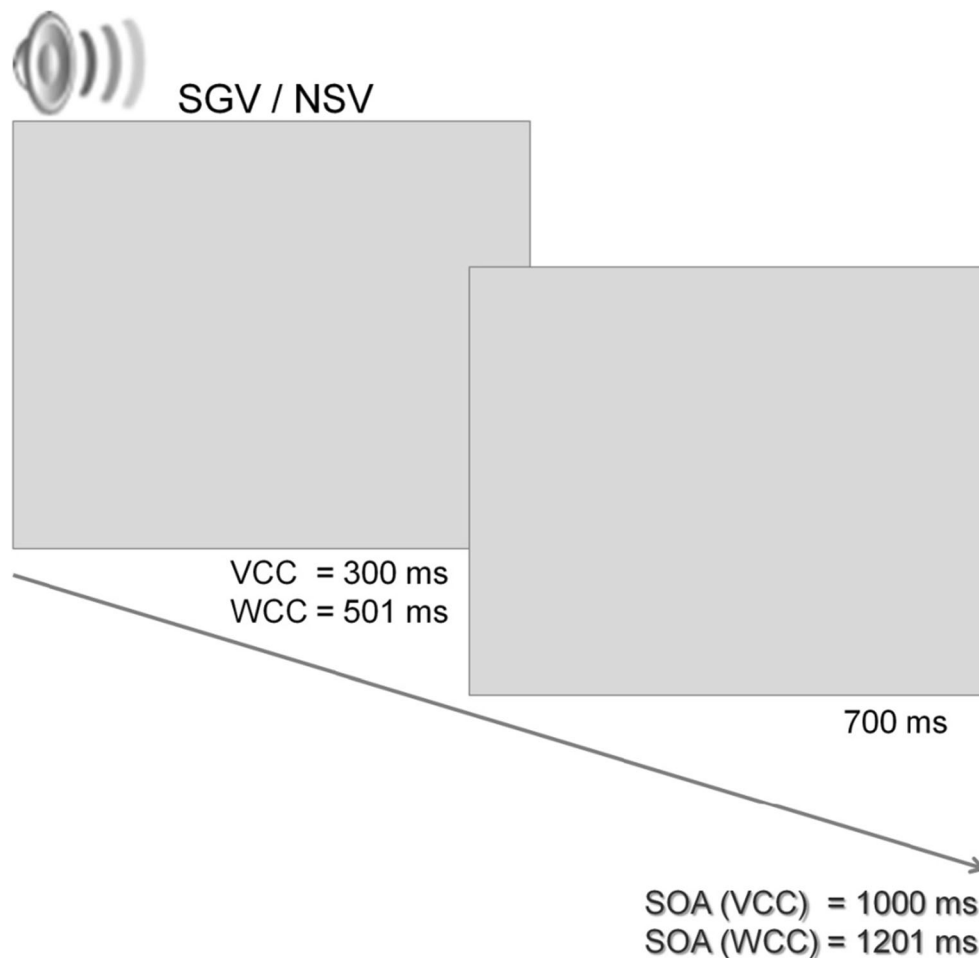


Fig. 1 Schematic illustration of a trial in the word (WCC) and vocalization (VCC) category conditions

μV criterion) were excluded from the analysis. Following artifact rejection, ERP averages were based on at least 75 % of the trials for each participant in each condition. ERP grand average difference waveforms were based on at least 75 % of the trials in each condition (standard SGV in VCC = 983.44 ± 49.50 ; deviant SGV in VCC = 140.59 ± 10.17 ; standard SGV in WCC = 954.53 ± 62.39 ; deviant SGV in WCC = 139.35 ± 10.33 ; standard NSV in VCC = 972.24 ± 66.56 ; deviant NSV in VCC = 139.50 ± 10.32 ; standard NSV in WCC = 978.12 ± 74.37 ; deviant NSV in WCC = 134.47 ± 10.99). In both the VCC and WCC, difference waveforms were calculated by subtracting SGV standards from SGV deviants, and NSV standards from NSV deviants.

After a careful visual inspection of grand average difference waveforms, two auditory ERP components with maximal effects at the FCz and Cz electrodes were identified: the MMN and P3a. For both the SGV and NSV difference waveforms in both the VCC and WCC conditions, the MMN mean amplitude was measured between 160 and 270 ms, whereas the P3a mean amplitude was measured between 280 and 410 ms. The MMN peak latency was measured as the time of the maximum negative point between 160 and 270 ms for

both the VCC and WCC, in which a local maximum value was determined separately for each electrode (FC1, FCz, FC2, C1, Cz, and C2) for every individual participant's data. Likewise, the P3a peak latency was computed as the time of the maximum positive point between 280 and 410 ms for both the VCC and the WCC.

Statistical analyses

Statistical analyses were conducted using SPSS 22.0 (SPSS, Corp., USA). All statistical analyses were based on unfiltered ERP data. The mean amplitude and latency of the MMN and P3a ERP components were analyzed using separate repeated measures analyses of variance (ANOVAs), with Stimulus Category Condition (VCC, WCC), Voice Identity (SGV, NSV), and Electrode (FCz, Cz) as within-subjects factors. Main effects and interactions were followed with pairwise comparisons between the conditions, using the Bonferroni adjustment for multiple comparisons. Analyses were corrected for nonsphericity by using the Greenhouse–Geisser correction method, when appropriate.

Additionally, we explored amplitude differences between the hemispheres for both the MMN and P3a components by running two repeated measures ANOVAs with Hemisphere (FC1/C1, FC2/C2) as an additional within-subjects factor. Main effects and interactions were followed with pairwise comparisons between the conditions, using the Bonferroni adjustment for multiple comparisons. Where appropriate, analyses were corrected for nonsphericity by using the Greenhouse–Geisser correction method.

Pearson's correlation coefficients were calculated to test whether the voices' acoustic properties from each participant (i.e., F0 and formant frequencies F1–F3) were associated with the MMN and P3a amplitudes elicited by SGV at the FCz channel. Following the procedure recommended by Field (2013) and others (Li, Chan, & Cui, 2011), we adopted the bootstrap sampling procedure (1,000 bootstrap samples) in SPSS via bias-corrected and accelerated (BCa) 95 % confidence intervals (CIs), since this allowed us to compute robust CIs of Pearson's correlation coefficients (Field, 2013; Li et al., 2011). A correlation was considered to be significant if its BCa bootstrap 95 % CI did not cross zero (Field, 2013).

Results

Figures 2 and 3 illustrate the grand average difference waveforms for the SGV and NSV conditions. A 12-Hz low-pass filter was applied to the grand average difference waveforms presented in Figs. 2 and 3, for illustration purposes only.

MMN

Bar plots of the results for both the MMN and the P3a can be found in Fig. 4. The repeated measures ANOVA revealed a main effect of category, $F(1, 16) = 16.133$, $p = .001$, $\eta_p^2 = .502$, on MMN amplitudes. Pairwise comparisons showed that the MMN was more negative for the VCC than for the WCC ($p = .001$). Furthermore, the hemispheric analysis showed no hemispheric differences in MMN amplitudes, $F(1, 16) = 0.002$, $p = .965$, $\eta_p^2 \leq .001$.

In addition, we observed a main effect of identity, $F(1, 16) = 8.865$, $p = .009$, $\eta_p^2 = .357$, on MMN latencies. Pairwise comparisons revealed that the MMN peaked earlier for the NSV than for the SGV condition ($p = .009$).

P3a

A main effect of category, $F(1, 16) = 17.589$, $p = .001$, $\eta_p^2 = .524$, was observed for P3a amplitudes. Pairwise comparisons demonstrated that the P3a was more positive in the VCC than in the WCC ($p = .001$). Moreover, the hemispheric analysis did not reveal any significant hemispheric differences in P3a amplitudes, $F(1, 16) = 1.506$, $p = .237$, $\eta_p^2 = .086$.

Importantly, we observed a significant identity by stimulus category interaction, $F(1, 16) = 6.706$, $p = .020$, $\eta_p^2 = .295$. Pairwise comparisons revealed that the P3a was more positive for the NSV than for the SGV in the VCC only ($p = .043$), whereas in the WCC no significant differences occurred between SGV and NSV ($p > .05$).

We observed a significant main effect of category, $F(1, 16) = 17.747$, $p = .001$, $\eta_p^2 = .526$, on P3a latencies. Pairwise comparisons demonstrated that the P3a peaked earlier in the VCC than in the WCC ($p = .001$).

Correlational analyses

We found no significant association between voices' acoustic properties and the MMN and P3a amplitudes elicited by SGVs. See Fig. 5 for a graphic display of these results.

Discussion

In the present study, we investigated the role of stimulus complexity on preattentive discrimination of self versus unknown voice stimuli, using ERPs. Participants in this study were asked to focus their attention on a silent movie while ignoring the presentation of voice stimuli.

MMN data shed light on the automatic detection of SGV versus NSV stimuli. We found that the MMN peaked earlier to NSV than to SGV stimuli in both the VCC and WCC. Since the MMN reflects a neural mechanism related to automatic deviance detection in response to change in a regular and invariant auditory stream (Garrido et al., 2009; Näätänen, 2001; Näätänen et al., 2007), these findings indicate that an unknown voice is detected earlier than one's own voice when attention is focused elsewhere, irrespective of the complexity of the voice stimuli (i.e., vocalization/word). In addition, considering that the MMN amplitudes were similar to both SGV and NSV stimuli in the VCC and WCC, this suggests that both voices elicit a similar amount of processing resources no matter *what* is being uttered.

The P3a elucidated how stimulus complexity modulates the orienting response to SGV versus NSV stimuli. As expected, the P3a amplitude was increased to NSV relative to SGV stimuli during the VCC, consistent with previous studies (Graux et al., 2015; Graux et al., 2013). A similar finding was also reported by Baess, Horvath, Jacobsen, and Schröger (2011), who observed that the P3a amplitude was enhanced to sounds that were externally generated compared with self-triggered sounds. Since the P3a is thought to reflect an orienting response toward an unpredictable change in an otherwise repetitive auditory background (Combs & Polich, 2006; Friedman, Cycowicz, & Gaeta, 2001; Gaeta, Friedman, & Hunt, 2003; Knight, 1996; Spencer, Dien, & Donchin, 1999, 2001), our finding suggests that the involuntary shift

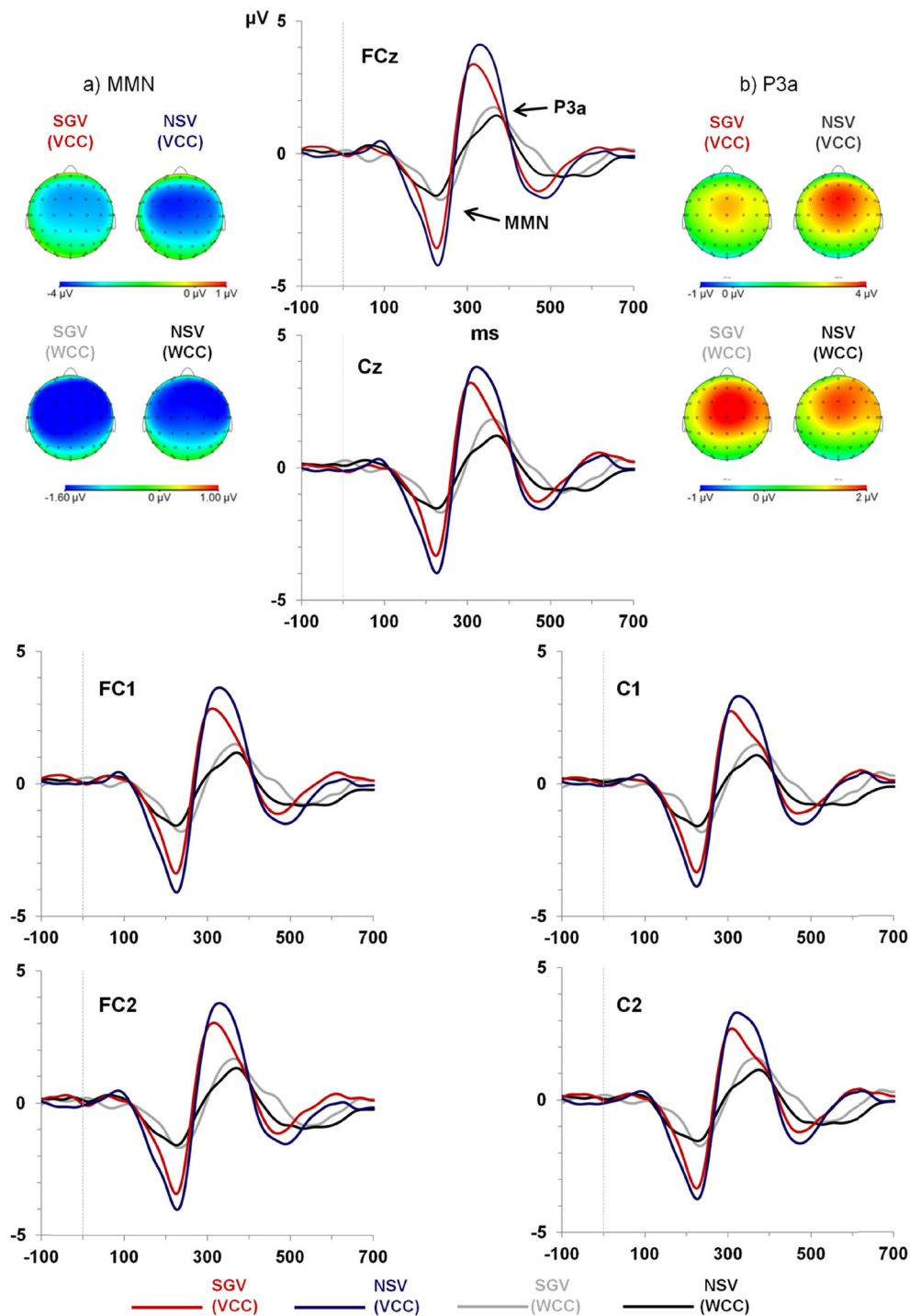


Fig. 2 Illustration of grand average difference waveforms for the self-generated (SGV) and nonself (NSV) voice stimuli in both the VCC and WCC

of attention is enhanced to nonself relative to self-generated deviant vocalizations that unexpectedly interrupt an unattended regular auditory stream. On the basis of similar results, Graux et al. (2013; Graux et al., 2015) proposed that enhanced processing resources are allocated to nonself voice stimuli when compared with one's own voice. Since the involuntary switch of attention is claimed to be a crucial biological process

for survival (Friedman et al., 2001; Friedman, Nessler, Kulik, & Hamberger, 2012), a plausible interpretation for our findings, and for those reported by Graux et al. (2013; Graux et al., 2015), is that the reorienting of attention toward an unexpected unfamiliar voice stimulus in the context of an unattended auditory environment may be more critical for survival than the detection of one's own voice.

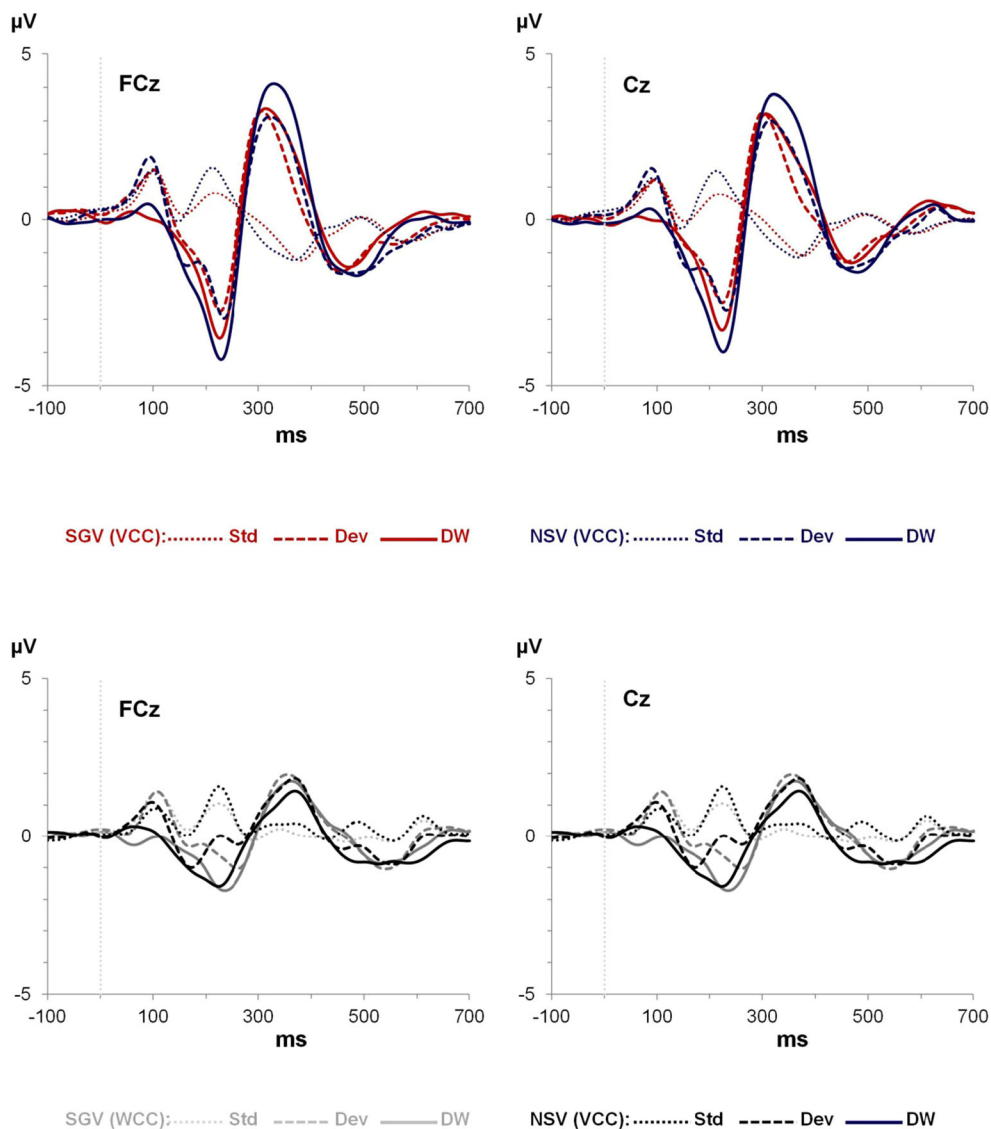


Fig. 3 Illustration of grand average “raw” waveforms for the SGV and NSV standard and deviant stimuli, in both the VCC and WCC

Surprisingly, we found no significant differences in P3a amplitudes between self-generated and nonself deviants in the WCC. This finding suggests similar engagement of involuntary attention by both SGV and NSV deviants when attention is directed away from the primary task (i.e., watching a silent movie) by the detection of an infrequent voice stimulus. These findings fit the results of Ventura et al. (2009), in which a reduced ERP differentiation between self-generated versus nonself voices was found for more complex relative to less complex voice stimuli. As such, our findings extend the studies of Graux et al. (2013; Graux et al., 2015), showing that the magnitude of the orienting response to SGV and NSV deviants depends on stimulus complexity and, more specifically, on linguistic information.

But how does stimulus complexity modulate the way that attention is reoriented to SGV versus NSV? In our study, the voice stimuli varied in terms of duration and linguistic

information: The vocalization /a/ had a shorter duration and reduced phonetic variability relative to the word /mesa/, from which more linguistic information (i.e., phonological, lexical, and semantic information) needed to be extracted. Furthermore, the word in our study comprised both consonant and vowel information, which in turn depend on distinctive acoustic parameters (Kaganovich, Francis, & Melara, 2006). According to the multidimensional model of voice perception proposed by Belin and colleagues (2004), linguistic and paralinguistic information (i.e., identity and affective cues) conveyed by the voice signal is processed in partially dissociated functional pathways that interact with each other during voice processing (Belin et al., 2011; Belin et al., 2004; Schweinberger et al., 2014). Therefore, both identity and linguistic information are simultaneously extracted from the voice signal, and they rely on some common acoustic cues, such as formant frequency and voice onset time (Allen &

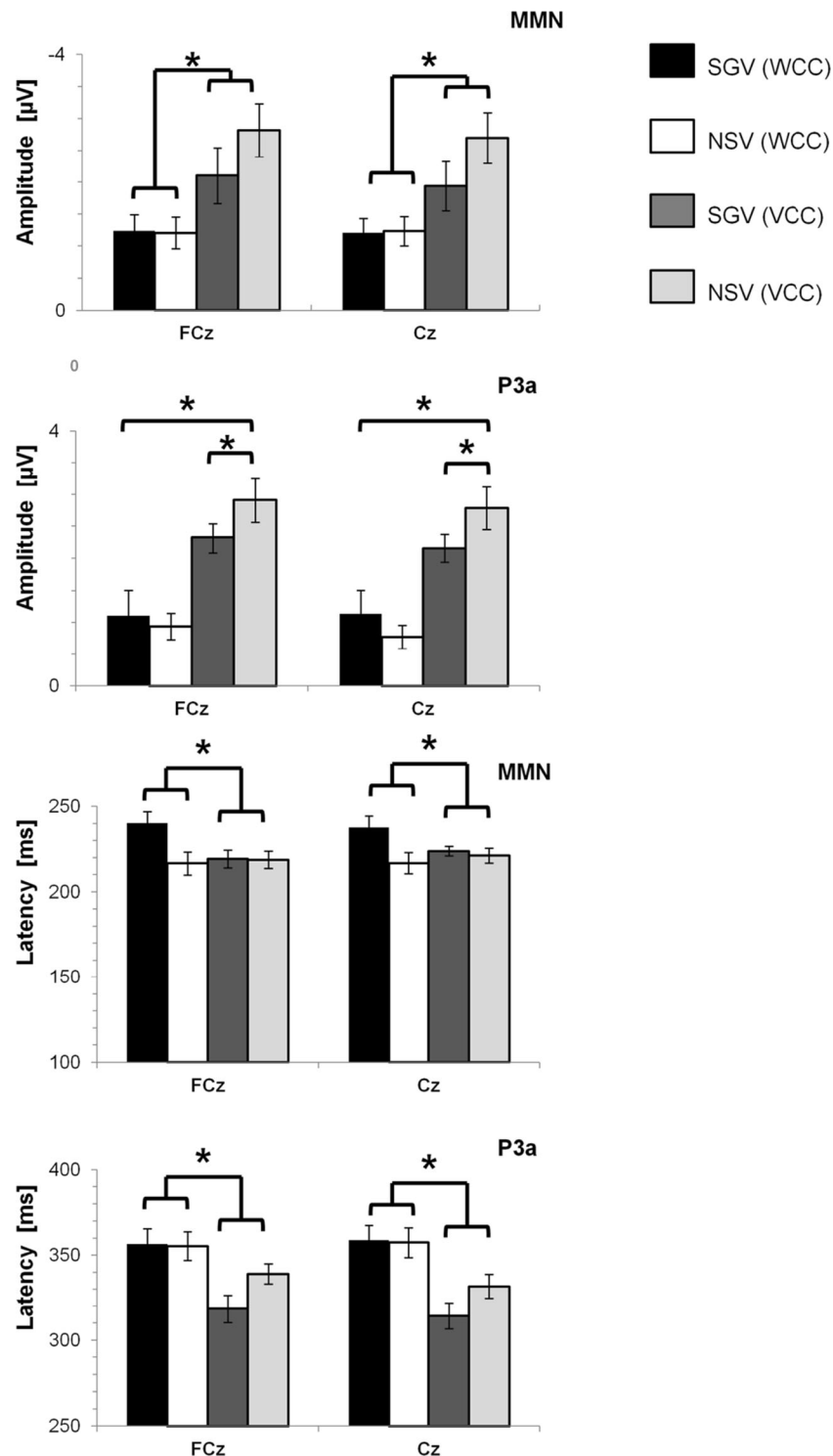


Fig. 4 Bar plots representing mismatch negativity (MMN) and P3a amplitudes and peak latencies for the SGV and NSV difference waveforms in both the VCC and WCC. * $p < .05$

Miller, 2004; Belin et al., 2011; Belin et al., 2004; Kaganovich et al., 2006; Schweinberger et al., 2014). Importantly, the concurrent processing of vocal information was found to take place in early stages of information processing—that is, within the first 200 ms after voice stimulus onset (Beauchemin

et al., 2006; Charest et al., 2009; Holeckova, Fischer, Giard, Delpuech, & Morlet, 2006; Kaganovich et al., 2006; Knösche, Lattner, Maess, Schauer, & Friederici, 2002; Titova & Näätänen, 2001). Besides the parallel processing of voice information, MMN studies demonstrated that concurrent

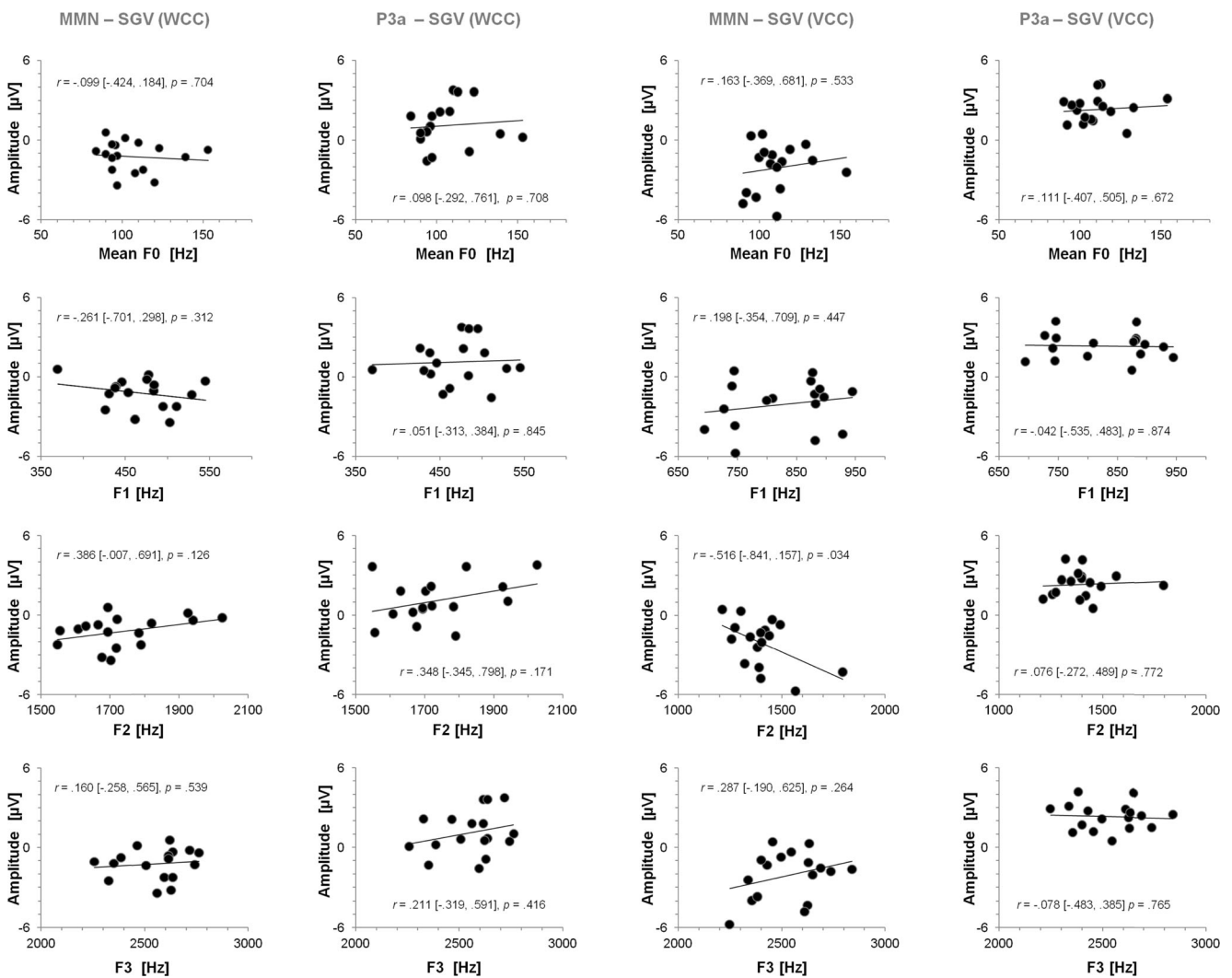


Fig. 5 Results of the correlation analysis using Pearson's coefficients for both the MMN and P3a amplitudes for SGV stimuli at the FCz electrode against the acoustic properties of the SGV stimuli. Bias-corrected and

accelerated (BCa) bootstrap 95 % confidence intervals (CIs) are reported in square brackets; correlations were considered to be significant if the BCa bootstrap 95 % CI did not cross zero

linguistic processes (e.g., phonological, lexical, semantic, grammatical, and pragmatic) occur very early in the information-processing stream within the MMN time window (Kujala et al., 2010; Kujala, Tervaniemi, & Schröger, 2007; Näätänen et al., 2007; Pakarinen et al., 2009; E. Pang et al., 1998; Pulvermüller & Shtyrov, 2006). Therefore, a plausible explanation for differences in the orienting response to SGV versus NSV based on stimulus complexity category is that the parallel processing of linguistic (e.g., phonological, lexical, and semantic) and paralinguistic (i.e., identity) information in the WCC might represent an increased “processing cost,” when compared with the VCC. Indeed, when the voice stimulus was simpler and devoid of semantic content, differences were observed in attention reorienting to self-generated versus nonself vocalizations, plausibly because more attentional resources were available to process voice identity. The hypothesis of an increased “processing cost” imposed by linguistic

information processing is consistent with our finding of an increased orienting response to the VCC relative to the WCC, as reflected by a more positive P3a amplitude in the VCC. Although the enhanced “processing cost” would be consistent with the increased MMN amplitude observed for the VCC versus the WCC, no differences were found in the time required to preattentively detect changes in the vocalization and word stimuli. This might suggest that other than language-related processes modulate SGV and NSV deviance detection.

Differences in the affective salience of voice stimuli might also have contributed to the earlier MMN to NSV than to SGV stimuli, as well as for the increased MMN and P3a amplitudes to VCC versus WCC in our study. The vocalization /a/, although devoid of semantic content, is a very frequently encountered speech sound that might be produced with subtle different emotional intonations to convey distinctive affective

meanings (e.g., agreement, anger, or amusement) in a particular social context (Ward, 2006; Ward & Tsukahara, 2000). Hence, it may be perceived as having a more ambiguous dialogue function, and thus as representing a more emotionally salient stimulus than a word conveying neutral semantic content. Since events with greater affective salience have been found to receive prioritized processing resources (e.g., Delplanque, Silvert, Hot, Rigoulot, & Sequeira, 2006; Delplanque, Silvert, Hot, & Sequeira, 2005; Vuilleumier & Huang, 2009), the hypothesized greater affective salience of the vocalization /a/ relative to the word stimulus may explain the observed differences between the VCC and WCC.

An alternative explanation for the decreased attentional orienting observed to self-generated deviant vocalizations only, in our study, is that less complex self-voice stimuli might have recruited reduced vocal self-monitoring resources that would otherwise be directed to process more complex self-generated linguistic stimuli. This would allow more efficient control of more complex linguistic messages by the vocal self-monitoring system, and in the presence of nonlinguistic self-generated vocalizations, more resources would be available to process nonself voice stimuli. Nonetheless, since recent neuroimaging evidence has shown that different neural processes are involved in how one's own voice is processed during speech production versus passive listening to the same vocal sounds (Behroozmand et al., 2015; Golfinopoulos, Tourville, & Guenther, 2010; Parkinson et al., 2012; Zheng, Munhall, & Johnsrude, 2010; Zheng et al., 2013), extending these findings to the realm of online vocal self-monitoring processes remains merely speculative. In particular, studies have demonstrated that, compared with a passive-listening condition, self-generated voice feedback during vocal production recruits additional sensory–motor brain regions, which are otherwise not activated in response to the passive-listening condition (Behroozmand et al., 2015; Parkinson et al., 2012; Zheng et al., 2010; Zheng et al., 2013). This may occur possibly due to the feedback's relevance in driving corrective mechanisms during vocal production. Therefore, since in our study participants simply had to hear previously recorded voice stimuli while attention was focused elsewhere (i.e., on watching a silent movie), the reported differences in the ERP correlates of preattentive deviance detection and attentional orienting to unexpected SGV and NSV stimuli (indexed by the MMN and P3a) should be understood in the context of a passive-listening design, rather than of a vocal production design. Interestingly, suggesting that even during passive listening motor-driven mechanisms may play a role in the distinction between “self” and “other,” earlier studies reported that, compared with an NSV, hearing a prerecorded SGV leads to increased activation within frontal regions, particularly in the left (Allen, Amaro, Fu, Williams, Brammer, Johns, & McGuire, 2005) and right (Kaplan et al., 2008; Nakamura et al., 2001) inferior frontal cortex.

Regarding the time course of attentional orienting to SGV and NSV stimuli, we found that the P3a peaked earlier in the VCC than in the WCC. Given that our voice stimuli varied in terms of duration and phonetic variability, it seems likely that the use of a more complex and meaningful word might have slowed down the processes of the orienting response to changes in voice identity. Consistent with this hypothesis, previous studies demonstrated that latency is normally increased when processing more versus less complex stimuli (Pakarinen et al., 2013; Polich, 2007; Pulvermüller & Shtyrov, 2006). Furthermore, robust evidence shows that the STG is particularly sensitive to “errors” in one's own voice, acting as a more general deviance detector, both in active vocal production or passive-listening conditions (Behroozmand et al., 2015; Parkinson et al., 2012; Zheng et al., 2010). Indeed, auditory error cells, whose activity reflects the mismatch between sensory prediction and the incoming feedback, are thought to be located in the STG (planum temporal and posterior STG; Golfinopoulos et al., 2010). Since the MMN is generated within supratemporal auditory cortical regions (Alho et al., 1996; Näätänen et al., 2007), this might suggest that the STG plays an important role in detecting deviance in SGV and NSV stimuli, even when participants' attention is focused elsewhere. Given the differential processing of one's own voice during speech production versus passive listening (Behroozmand et al., 2015; Golfinopoulos et al., 2010; Parkinson et al., 2012; Zheng et al., 2010; Zheng et al., 2013), the faster detection of a nonself than of a self-generated voice stimulus observed in our study might be specific of passive-listening contexts, wherein the reduced activation of sensory–motor mechanisms signals the biological relevance of detecting a nonself voice stimulus violating a regular and highly predictable auditory background. Moreover, the left STG has been consistently implicated in the processing of linguistic content and in the extraction of speech meaning conveyed by voice signals (e.g., Belin et al., 2011; Binder, 2000; DeWitt & Rauschecker, 2012; Obleser, Zimmermann, Van Meter, & Rauschecker, 2007). Given the increased responsivity of bilateral STG to one's own speech as compared with acoustic noise (Zheng et al., 2013), it seems plausible that this region might play an important role in decoding the stimulus complexity of SGV and NSV stimuli, as well as in allocating more resources to NSV processing consisting of less complex vocalizations, as we observed in our study.

On the other hand, we tested whether the ERP correlates of change detection and the orienting response underlying SGV processing were associated with the acoustic properties of this stimulus type (i.e., F0 and formant frequencies). We found no association between a voice's acoustic properties and the amplitudes of both the MMN and P3a to an SGV, yet earlier evidence had demonstrated that both F0 and formant frequencies are critical acoustic cues underlying successful voice identity recognition (e.g., Latinus et al., 2013; Xu et al., 2013). As expected, this lack of association was plausibly

due to the fact that our ERP analysis controlled for the physical differences between the voice stimuli by using a “like from like” subtraction approach. This finding suggests that the ERP correlates of automatic deviance detection and attention reorienting underlying SGV processing were independent of the SGV’s physical parameters. Thus, the reduced reorienting response to a self-generated relative to a nonself vocalization might be related to the increased salience of the nonself voice in the context of an unattended auditory background. In this specific context, detecting a novel and unpredictable speaker’s voice might be more advantageous for evolutionary purposes than the detection of a self-generated voice.

Since most of the studies on self-voice processing have used simple voice stimuli, such as vocalizations (e.g., Ford et al., 2007; Graux et al., 2015; Graux et al., 2013; Sitek et al., 2013; Whitford et al., 2011), future studies should test the specific contribution of stimulus complexity to these processes by including different stimuli varying in their amounts of linguistic and paralinguistic information. Furthermore, stimulus length is an important marker of linguistic complexity, and our experimental design does not allow for disentangling the effects of stimulus duration from linguistic information effects in the WCC. Thus, we hope that future research can dissociate both effects and their independent contributions to self and nonself voice discrimination. Also, it is relevant to investigate self versus nonself voice processing during vocal production while attentional demands and self-generated voice feedback are dynamically changing due to alterations in the acoustic environment (such as in a social conversational setting). This may contribute to a more ecological understanding of how one’s own voice is processed during daily social communication processes.

Conclusion

In the present study, we investigated the role of stimulus complexity in the ERP correlates of preattentive change detection and attention orienting to a self-generated in comparison with a nonself voice. Our study provided evidence for earlier detection of a nonself relative to a self-generated voice, which was not associated with the voice’s acoustic properties. Differences in attention orienting to self and nonself voices were observed only when a less complex stimulus was uttered, suggesting that the magnitude of the orienting response to changes in voice identity depends on the complexity of the voice signal.

Author note This work was supported by Grant Numbers IF/00334/2012, PTDC/PSI-PCL/116626/2010, and PTDC/MHN-PCN/3606/2012, funded by the Fundação para a Ciência e a Tecnologia (FCT, Portugal) and the Fundo Europeu de Desenvolvimento Regional through the European programs Quadro de Referência Estratégico Nacional and Programa Operacional Factores de Competitividade, awarded to A.P.P., and by FCT Doctoral Grant Number SFRH/BD/77681/2011, awarded to T.C.

References

- Alho, K., Tervaniemi, M., Huotilainen, M., Lavikainen, J., Tiitinen, H., Ilmoniemi, R. J., . . . Näätänen, R. (1996). Processing of complex sounds in the human auditory cortex as revealed by magnetic brain responses. *Psychophysiology*, *33*, 369–375. doi:10.1111/j.1469-8986.1996.tb01061.x
- 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. *Human Brain Mapping*, *26*(1), 44–53. doi:10.1002/hbm.20120
- Allen, P. P., Amaro, E., Fu, C. H., Williams, S. C. R., Brammer, M. J., Johns, L. C., & McGuire, P. K. (2007). Neural correlates of the misattribution of speech in schizophrenia. *British Journal of Psychiatry*, *190*, 162–169. doi:10.1192/bjp.bp.106.025700
- Allen, P. P., Johns, L. C., Fu, C. H., Broome, M. R., Vythelingum, G. N., & McGuire, P. K. (2004). Misattribution of external speech in patients with hallucinations and delusions. *Schizophrenia Research*, *69*, 277–287. doi:10.1016/j.schres.2003.09.008
- Allen, J. S., & Miller, J. L. (2004). Listener sensitivity to individual talker differences in voice-onset-time. *Journal of the Acoustical Society of America*, *115*, 3171–3183. doi:10.1121/1.1701898
- Baess, P., Horvath, J., Jacobsen, T., & Schröger, E. (2011). Selective suppression of self-initiated sounds in an auditory stream: An ERP study. *Psychophysiology*, *48*, 1276–1283. doi:10.1111/j.1469-8986.2011.01196.x
- Baumann, O., & Belin, P. (2010). Perceptual scaling of voice identity: Common dimensions for different vowels and speakers. *Psychological Research*, *74*, 110–120. doi:10.1007/s00426-008-0185-z
- Beauchemin, M., De Beaumont, L., Vannasing, P., Turcotte, A., Arcand, C., Belin, P., & Lassonde, M. (2006). Electrophysiological markers of voice familiarity. *European Journal of Neuroscience*, *23*, 3081–3086. doi:10.1111/j.1460-9568.2006.04856.x
- Behroozmand, R., Ibrahim, N., Korzyukov, O., Robin, D. A., & Larson, C. R. (2014). Left-hemisphere activation is associated with enhanced vocal pitch error detection in musicians with absolute pitch. *Brain and Cognition*, *84*, 97–108. doi:10.1016/j.bandc.2013.11.007
- Behroozmand, R., Karvelis, L., Liu, H., & Larson, C. R. (2009). Vocalization-induced enhancement of the auditory cortex responsiveness during voice F 0 feedback perturbation. *Clinical Neurophysiology*, *120*, 1303–1312. doi:10.1016/j.clinph.2009.04.022
- Behroozmand, R., Korzyukov, O., Sattler, L., & Larson, C. R. (2012). Opposing and following vocal responses to pitch-shifted auditory feedback: Evidence for different mechanisms of voice pitch control. *Journal of the Acoustical Society of America*, *132*, 2468–2477. doi:10.1121/1.4746984
- Behroozmand, R., Korzyukov, O., & Larson, C. R. (2011). Effects of voice harmonic complexity on ERP responses to pitch-shifted auditory feedback. *Clinical Neurophysiology*, *122*, 2408–2417. doi:10.1016/j.clinph.2011.04.019
- Behroozmand, R., & Larson, C. R. (2011). Error-dependent modulation of speech-induced auditory suppression for pitch-shifted voice feedback. *BMC Neuroscience*, *12*, 54. doi:10.1186/1471-2202-12-54
- 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. doi:10.1016/j.neuroimage.2015.01.040
- Belin, P., Bestelmeyer, P. E. G., Latinus, M., & Watson, R. (2011). Understanding voice perception. *British Journal of Psychology*, *102*, 711–725. doi:10.1111/j.2044-8295.2011.02041.x
- Belin, P., Fecteau, S., & Bédard, C. (2004). Thinking the voice: Neural correlates of voice perception. *Trends in Cognitive Sciences*, *8*, 129–135. doi:10.1016/j.tics.2004.01.008

- Binder, J. (2000). The new neuroanatomy of speech perception. *Brain*, *123*, 2371–2372. doi:10.1093/brain/123.12.2371
- Boersma, P., & Weenink, D. (2012). Praat, version 5.3. Retrieved from www.fon.hum.uva.nl/praat/
- Brumm, H., & Zollinger, S. A. (2011). The evolution of the Lombard effect: 100 years of psychoacoustic research. *Behaviour*, *148*, 1173–1198. doi:10.1163/000579511X605759
- Burnett, T. A., Freedland, M. B., Larson, C. R., & Hain, T. C. (1998). Voice F0 responses to manipulations in pitch feedback. *Journal of the Acoustical Society of America*, *103*, 3153–3161. doi:10.1121/1.423073
- Canavaro, M. C. (1999). Inventário de Sintomas Psicopatológicos—BSI [Psychopathological Symptoms Inventory—BSI]. In M. R. Simões, M. Gonçalves, & L. S. Almeida (Eds.), *Testes e provas psicológicas em Portugal* (Vol. 2, pp. 96–109). Braga, Portugal: APPORT/SHO.
- Canavaro, 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, & M. Gonçalves (Eds.), *Avaliação psicológica: Instrumentos validados para a população Portuguesa [Psychological assessment: Tests validated for the Portuguese population]* (Vol. 3, pp. 305–331). Coimbra, Portugal: Quarteto Editora.
- Charest, I., Pernet, C. R., Rousselet, G. A., Quiñones, I., Latinus, M., Fillion-Bilodeau, S., . . . Belin, P. (2009). Electrophysiological evidence for an early processing of human voices. *BMC Neuroscience*, *10*, 127. doi:10.1186/1471-2202-10-127
- 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*, e60039. doi:10.1371/journal.pone.0060039
- Chen, S. H., Liu, H., Xu, Y., & Larson, C. R. (2007). Voice F0 responses to pitch-shifted voice feedback during English speech. *Journal of the Acoustical Society of America*, *121*, 1157–1163. doi:10.1121/1.2404624
- Combs, L. A., & Polich, J. (2006). P3a from auditory white noise stimuli. *Clinical Neurophysiology*, *117*, 1106–1112. doi:10.1016/j.clinph.2006.01.023
- Conde, T., Gonçalves, Ó. F., & Pinheiro, A. P. (in press). A cognitive neuroscience view on voice processing abnormalities in schizophrenia: A window into auditory verbal hallucinations? *Harvard Review of Psychiatry*.
- Cook, S., & Wilding, J. (1997). Earwitness testimony: Never mind the variety, hear the length. *Applied Cognitive Psychology*, *11*, 95–111. doi:10.1002/(SICI)1099-0720(199704)11:2<95::AID-ACP429>3.0.CO;2-O
- Curio, G., Neuloh, G., Numminen, J., Jousmaki, V., & Hari, R. (2000). Speaking modifies voice-evoked activity in the human auditory cortex. *Human Brain Mapping*, *9*, 183–191. doi:10.1002/(SICI)1097-0193(200004)9:43.0.CO;2-Z
- Delplanque, S., Silvert, L., Hot, P., Rigoulot, S., & Sequeira, H. (2006). Arousal and valence effects on event-related P3a and P3b during emotional categorization. *International Journal of Psychophysiology*, *60*, 315–322. doi:10.1016/j.ijpsycho.2005.06.006
- Delplanque, S., Silvert, L., Hot, P., & Sequeira, H. (2005). Event-related P3a and P3b in response to unpredictable emotional stimuli. *Biological Psychology*, *68*, 107–120. doi:10.1016/j.biopsycho.2004.04.006
- Derogatis, L. R., & Spencer, M. S. (1982). *The Brief Symptom Inventory (BSI): Administration, scoring, and procedures manual—I*. Baltimore, MD: Johns Hopkins University School of Medicine, Clinical Psychometrics Research Unit.
- DeWitt, I., & Rauschecker, J. P. (2012). Phoneme and word recognition in the auditory ventral stream. *Proceedings of the National Academy of Sciences*, *109*, E505–E514. doi:10.1073/pnas.1113427109
- Eliades, S. J., & Wang, X. (2008). Neural substrates of vocalization feedback monitoring in primate auditory cortex. *Nature*, *453*, 1102–1106. doi:10.1038/nature06910
- Field, A. P. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). London, UK: Sage.
- Fleming, D., Giordano, B. L., Caldara, R., & Belin, P. (2014). A language-familiarity effect for speaker discrimination without comprehension. *Proceedings of the National Academy of Sciences*, *111*, 13795–13798. doi:10.1073/pnas.1401383111
- Ford, J. M., & Mathalon, D. H. (2004). Electrophysiological evidence of corollary discharge dysfunction in schizophrenia during talking and thinking. *Journal of Psychiatric Research*, *38*, 37–46. doi:10.1016/S0022-3956(03)00095-5
- Ford, J. M., & Mathalon, D. H. (2005). Corollary discharge dysfunction in schizophrenia: Can it explain auditory hallucinations? *International Journal of Psychophysiology*, *58*, 179–189. doi:10.1016/j.ijpsycho.2005.01.014
- Ford, J. M., Mathalon, D. H., Kalba, S., Whitfield, S., Faustman, W. O., & Roth, W. T. (2001). Cortical responsiveness during talking and listening in schizophrenia: An event-related potential study. *Biological Psychiatry*, *50*, 540–549.
- Ford, J. M., Mathalon, D. H., Whitfield, S., Faustman, W. O., & Roth, W. T. (2002). Reduced communication between frontal and temporal lobes during talking in schizophrenia. *Biological Psychiatry*, *51*, 485–492. doi:10.1016/s0006-3223(01)01335-x
- Ford, J. M., Roach, B. J., Faustman, W. O., & Mathalon, D. H. (2007). Synch before you speak: Auditory hallucinations in schizophrenia. *American Journal of Psychiatry*, *164*, 458–466. doi:10.1176/appi.ajp.164.3.458
- Ford, J. M., Roach, B. J., & Mathalon, D. H. (2010). Assessing corollary discharge in humans using noninvasive neurophysiological methods. *Nature Protocols*, *5*, 1160–1168. doi:10.1038/nprot.2010.67
- 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. *Neuroscience & Biobehavioral Reviews*, *25*, 355–373. Retrieved from www.ncbi.nlm.nih.gov/pubmed/11445140
- Friedman, D., Nessler, D., Kulik, J., & Hamberger, M. (2012). The brain's orienting response (novelty P3) in patients with unilateral temporal lobe resections. *Neuropsychologia*, *49*, 3474–3483. doi:10.1016/j.neuropsychologia.2011.08.023
- Gaeta, H., Friedman, D., & Hunt, G. (2003). Stimulus characteristics and task category dissociate the anterior and posterior aspects of novelty P3. *Psychophysiology*, *40*, 198–208. doi:10.1111/1469-8986.00022
- Garrido, M. I., Kilner, J. M., Stephan, K. E., & Friston, K. J. (2009). The mismatch negativity: A review of underlying mechanisms. *Clinical Neurophysiology*, *120*, 453–463. doi:10.1016/j.clinph.2008.11.029
- Godey, B., Schwartz, D., de Graaf, J. B., Chauvel, P., & Liégeois-Chauvel, C. (2001). Neuromagnetic source localization of auditory evoked fields and intracerebral evoked potentials: A comparison of data in the same patients. *Clinical Neurophysiology*, *112*, 1850–1859. doi:10.1016/S1388-2457(01)00636-8
- Goggin, J. P., Thompson, C. P., Strube, G., & Simental, L. R. (1991). The role of language familiarity in voice identification. *Memory Cognition*, *19*, 448–458. doi:10.3758/BF03199567
- Golfinopoulos, E., Tourville, J. A., & Guenther, F. H. (2010). The integration of large-scale neural network modeling and functional brain imaging in speech motor control. *NeuroImage*, *52*, 862–874. doi:10.1016/j.neuroimage.2009.10.023
- Gratton, G., Coles, M. G. H., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology*, *55*, 468–484. doi: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. doi:10.1007/s10548-012-0233-2
- Graux, J., Gomot, M., Roux, S., Bonnet-Brilhault, F., & Bruneau, N. (2015). Is my voice just a familiar voice? An electrophysiological study. *Social Cognitive and Affective Neuroscience*, 10, 101–105. doi:10.1093/scan/nsu031
- Heinks-Maldonado, T. H., Mathalon, D. H., Gray, M., & Ford, J. (2005). Fine-tuning of auditory cortex during speech production. *Psychophysiology*, 42, 180–190. doi:10.1111/j.1469-8986.2005.00272.x
- Heinks-Maldonado, T. H., Mathalon, D. H., Houde, J. F., Gray, M., Faustman, W. O., & Ford, J. M. (2007). Relationship of imprecise corollary discharge in schizophrenia to auditory hallucinations. *Archives of General Psychiatry*, 64, 286–296. doi:10.1001/archpsyc.64.3.286
- Heinks-Maldonado, T. H., Nagarajan, S. S., & Houde, J. F. (2006). Magnetoencephalographic evidence for a precise forward model in speech production. *NeuroReport*, 17, 1375–1379. doi:10.1097/01.wnr.0000233102.43526.e9
- Holeckova, I., Fischer, C., Giard, M. H., Delpuech, C., & Morlet, D. (2006). Brain responses to a subject's own name uttered by a familiar voice. *Brain Research*, 1082, 142–152. doi:10.1016/j.brainres.2006.01.089
- Houde, J. F., Nagarajan, S. S., Sekihara, K., & Merzenich, M. M. (2002). Modulation of the auditory cortex during speech: An MEG Study. *Journal of Cognitive Neuroscience*, 14, 1125–1138. doi:10.1162/089892902760807140
- Illankovic, L. M., Allen, P. P., Engel, R., Kambeitz, J., Riedel, M., Müller, N., & Hennig-Fast, K. (2011). Attentional modulation of external speech attribution in patients with hallucinations and delusions. *Neuropsychologia*, 49, 805–812. doi:10.1016/j.neuropsychologia.2011.01.016
- Kaganovich, N., Francis, A. L., & Melara, R. D. (2006). Electrophysiological evidence for early interaction between talker and linguistic information during speech perception. *Brain Research*, 1114, 161–172. doi: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. *Social Cognitive and Affective Neuroscience*, 3, 218–223. doi:10.1093/scan/nsn014
- Keenan, J. P., Falk, D., & Gallup, G. G., Jr. (2003). *The face in the mirror: The search for the origins of consciousness*. New York, NY: HarperCollins.
- Knight, R. T. (1996). Contribution of human hippocampal region to novelty detection. *Nature*, 383, 256–259. doi:10.1038/383256a0
- Knösche, T. R., Lattner, S., Maess, B., Schauer, M., & Friederici, A. D. (2002). Early parallel processing of auditory word and voice information. *NeuroImage*, 17, 1493–1503. doi:10.1006/nimg.2002.1262
- Kujala, T., Kuuluvainen, S., Saalasti, S., Jansson-Verkasalo, E., von Wendt, L., & Lepistö, T. (2010). Speech-feature discrimination in children with Asperger syndrome as determined with the multi-feature mismatch negativity paradigm. *Clinical Neurophysiology*, 121, 1410–1419. doi:10.1016/j.clinph.2010.03.017
- Kujala, T., Tervaniemi, M., & Schröger, E. (2007). The mismatch negativity in cognitive and clinical neuroscience: Theoretical and methodological considerations. *Biological Psychology*, 74, 1–19. doi:10.1016/j.biopsycho.2006.06.001
- Lane, H., & Webster, J. W. (1991). Speech deterioration in postlingually deafened adults. *Journal of the Acoustical Society of America*, 89, 859–866. doi:10.1121/1.1894647
- Larson, C. R., Altman, K. W., Liu, H., & Hain, T. C. (2008). Interactions between auditory and somatosensory feedback for voice F0 control. *Experimental Brain Research*, 187(4), 613–621.
- Latinus, M., & Belin, P. (2011). Anti-voice adaptation suggests prototype-based coding of voice identity. *Frontiers in Psychology*, 2(175), 1–12. doi:10.3389/fpsyg.2011.00175
- Latinus, M., & Belin, P. (2012). Perceptual auditory aftereffects on voice identity using brief vowel stimuli. *PLoS One*, 7, e41384. doi:10.1371/journal.pone.0041384
- Latinus, M., McAleer, P., Bestelmeyer, P. E. G., & Belin, P. (2013). Norm-based coding of voice identity in human auditory cortex. *Current Biology*, 23, 1075–1080. doi:10.1016/j.cub.2013.04.055
- Leitman, D. I., Foxe, J. J., Sehatpour, P., Shpaner, M., & Javitt, D. C. (2009). Mismatch negativity to tonal contours suggests preattentive perception of prosodic content. *Brain Imaging Behavior*, 3, 284–291. doi:10.1007/s11682-009-9070-7
- Leitman, D. I., Sehatpour, P., Garidis, C., Gomez-Ramirez, M., & Javitt, D. C. (2011). Preliminary evidence of pre-attentive distinctions of frequency-modulated tones that convey affect. *Frontiers in Human Neuroscience*, 5(96), 1–8. doi:10.3389/fnhum.2011.00096
- Letowski, T., Frank, T., & Caravella, J. (1993). Acoustical properties of speech produced in noise through supra-aural headphones. *Ear and Hearing*, 14, 332–338.
- Li, J. C.-H., Chan, W., & Cui, Y. (2011). Bootstrap standard error and confidence intervals for the correlations corrected for indirect range restriction. *British Journal of Mathematical and Statistical Psychology*, 64, 367–387. doi:10.1348/2044-8317.002007
- Liu, P., Chen, Z., Jones, J. A., Huang, D., & Liu, H. (2011a). Auditory feedback control of vocal pitch during sustained vocalization: A cross-sectional study of adult aging. *PLoS One*, 6, e22791. doi:10.1371/journal.pone.0022791
- Liu, H., & Larson, C. R. (2007). Effects of perturbation magnitude and voice F0 level on the pitch shift reflex. *Journal of the Acoustical Society of America*, 122, 3671–3677. doi:10.1121/1.2800254
- Liu, H., Meshman, M., Behroozmand, R., & Larson, C. R. (2011b). Differential effects of perturbation direction and magnitude on the neural processing of voice pitch feedback. *Clinical Neurophysiology*, 122, 951–957. doi:10.1016/j.clinph.2010.08.010
- 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 and Hearing*, 28, 628–642. doi:10.1097/AUD.0b013e31812564c9
- Näätänen, R. (2001). The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology*, 38, 1–21. doi:10.1111/1469-8986.3810001
- Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*, 118, 2544–2590. doi:10.1016/j.clinph.2007.04.026
- 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, 1047–1054.
- Numminen, J., Salmelin, R., & Hari, R. (1999). Subject's own speech reduces reactivity of the human auditory cortex. *Neuroscience Letters*, 265, 119–122. doi:10.1016/S0304-3940(99)00218-9
- Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics*, 60, 355–376. doi:10.3758/BF03206860
- Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science*, 5, 42–46. doi:10.1111/j.1467-9280.1994.tb00612.x
- Obleser, J., Zimmermann, J., Van Meter, J., & Rauschecker, J. P. (2007). Multiple stages of auditory speech perception reflected in event-related FMRI. *Cerebral Cortex*, 17, 2251–2257. doi:10.1093/cercor/bhl133
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, 9, 97–113. doi:10.1016/0028-3932(71)90067-4

- Oller, D. K., & Eilers, R. E. (1988). The role of audition in infant babbling. *Child Development*, 59, 441–449.
- Özgül, S., Gudlowski, Y., Withaus, 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, 272–278. doi:10.1016/j.schres.2008.05.017
- Pakarinen, S., Lovio, R., Huotilainen, M., Alku, P., Näätänen, R., & Kujala, T. (2009). Fast multi-feature paradigm for recording several mismatch negativities (MMNs) to phonetic and acoustic changes in speech sounds. *Biological Psychology*, 82, 219–226. doi:10.1016/j.biopsycho.2009.07.008
- Pakarinen, S., Teinonen, T., Shestakova, A., Kwon, M. S., Kujala, T., Hämäläinen, H., . . . Huotilainen, M. (2013). Fast parametric evaluation of central speech-sound processing with mismatch negativity (MMN). *International Journal of Psychophysiology*, 87, 103–110. doi:10.1016/j.ijpsycho.2012.11.010
- Pang, E., Edmonds, G., Desjardins, R., Khan, S., Trainor, L., & Taylor, M. (1998). Mismatch negativity to speech stimuli in 8-month-old infant and adults. *International Journal of Psychophysiology*, 29, 227–236. doi:10.1016/S0167-8760(98)00018-X
- Pang, X., Xu, J., Chang, Y., Tang, D., Zheng, Y., Liu, Y., & Sun, Y. (2014). Mismatch negativity of sad syllables is absent in patients with major depressive disorder. *PLoS One*, 9, e91995. doi:10.1371/journal.pone.0091995
- Parkinson, A. L., Behroozmand, R., Ibrahim, N., Korzyukov, O., Larson, C. R., & Robin, D. A. (2014). Effective connectivity associated with auditory error detection in musicians with absolute pitch. *Frontiers in Neuroscience*, 8, 46. doi:10.3389/fnins.2014.00046
- 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. doi:10.1016/j.neuroimage.2012.02.068
- Patel, R., & Schell, K. W. (2008). The influence of linguistic content on the Lombard effect. *Journal of Speech, Language, and Hearing Research*, 51, 209–220. doi:10.1044/1092-4388(2008/016)
- Perrachione, T. K., & Wong, P. C. M. (2007). Learning to recognize speakers of a non-native language: Implications for the functional organization of human auditory cortex. *Neuropsychologia*, 45, 1899–1910. doi:10.1016/j.neuropsychologia.2006.11.015
- Pittman, A. L., & Wiley, T. L. (2001). Recognition of speech produced in noise. *Journal of Speech, Language, and Hearing Research*, 44, 487–496. doi:10.1044/1092-4388(2001/038)
- Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical EEG and Neuroscience*, 118, 2128–2148. doi:10.1016/j.clinph.2007.04.019
- Pulvermüller, F., & Shtyrov, Y. (2006). Language outside the focus of attention: The mismatch negativity as a tool for studying higher cognitive processes. *Progress in Neurobiology*, 79, 49–71. doi:10.1016/j.pneurobio.2006.04.004
- Schauwers, K., Gillis, S., Daemers, K., De Beukelaer, C., De Ceulaer, G., Yperman, M., & Govaerts, P. J. (2004). Normal hearing and language development in a deaf-born child. *Otology and Neurology*, 25, 924–929. doi:10.1097/00129492-200411000-00011
- Schirmer, A., & Escoffier, N. (2010). Emotional MMN: Anxiety and heart rate correlate with the ERP signature for auditory change detection. *Clinical Neurophysiology*, 121, 53–59. doi:10.1016/j.clinph.2009.09.029
- Schirmer, A., Simpson, E., & Escoffier, N. (2007). Listen up! Processing of intensity change differs for vocal and nonvocal sounds. *Brain Research*, 1176, 103–112. doi:10.1016/j.brainres.2007.08.008
- Schirmer, A., Striano, C. A. T., & Friederici, A. D. (2005). Sex differences in the preattentive processing of vocal emotional expressions. *NeuroReport*, 16, 635–639. doi: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. *Journal of Speech, Language, and Hearing Research*, 40, 453–463. doi:10.1044/jslhr.4002.453
- Schweinberger, S. R., Kawahara, H., Simpson, A. P., Skuk, V. G., & Zäske, R. (2014). Speaker perception. *Wiley Interdisciplinary Reviews: Cognitive Science*, 5, 15–25. doi:10.1002/wcs.1261
- Schweinberger, S. R., Walther, C., Zäske, R., & Kovács, G. (2011). Neural correlates of adaptation to voice identity. *British Journal of Psychology*, 102, 748–764. doi:10.1111/j.2044-8295.2011.02048.x
- Shtyrov, Y., & Pulvermüller, F. (2002). Memory traces for inflectional affixes as shown by mismatch negativity. *European Journal of Neuroscience*, 15, 1085–1091.
- Sitek, K. R., Mathalon, D. H., Roach, B. J., Houde, J. F., Niziolek, C. A., & Ford, J. M. (2013). Auditory cortex processes variation in our own speech. *PLoS One*, 8, e82925. doi:10.1371/journal.pone.0082925
- Soares, A. P., Comesaña, M., Iriarte, A., Almeida, J. J., Simões, A., Costa, A., . . . Machado, J. (2010). P-PAL: Uma base lexical com índices psicolinguísticos do Português Europeu [P-PAL: A European Portuguese lexical database]. *Linguamática*, 2, 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. *Behavior Research Methods*, 44, 256–269. doi: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, 409–414. doi:10.1017/S0048577299981180
- Spencer, K. M., Dien, J., & Donchin, E. (2001). Spatiotemporal analysis of the late ERP responses to deviant stimuli. *Psychophysiology*, 38, 343–358. doi:10.1017/S0048577201000324
- Sugimori, E., Asai, T., & Tanno, Y. (2013). The potential link between sense of agency and output monitoring over speech. *Consciousness and Cognition*, 22, 360–374. doi:10.1016/j.concog.2012.07.010
- Sugiura, M. (2013). Associative account of self-cognition: Extended forward model and multi-layer structure. *Frontiers in Human Neuroscience*, 7(535), 1–16. doi:10.3389/fnhum.2013.00535
- Sugiura, M., Miyauchi, C. M., Kotozaki, Y., Akimoto, Y., Nozawa, T., Yomogida, Y., . . . Kawashima, R. (2014). Neural mechanism for mirrored self-face Recognition. *Cerebral Cortex*, 25, 1–9. doi:10.1093/cercor/bhu077
- Sugiura, M., Sassa, Y., Jeong, H., Horie, K., Sato, S., & Kawashima, R. (2008). Face specific and domain-general characteristics of cortical responses during self-recognition. *NeuroImage*, 42, 414–422. doi:10.1016/j.neuroimage.2008.03.054
- Thompson, C. P. (1987). A language effect in voice identification. *Applied Cognitive Psychology*, 1(2), 121–131. doi:10.1002/acp.2350010205
- Tian, X., & Poeppel, D. (2013). The effect of imagination on stimulation: The functional specificity of efference copies in speech processing. *Journal of Cognitive Neuroscience*, 25, 1020–1036. doi:10.1162/jocn_a_00381
- Tian, X., & Poeppel, D. (2015). Dynamics of self-monitoring and error detection in speech production: Evidence from mental imagery and MEG. *Journal of Cognitive Neuroscience*, 27, 352–364. doi:10.1162/jocn_a_00692
- Timm, J., SanMiguel, I., Saupé, K., & Schröger, E. (2013). The N1-suppression effect for self-initiated sounds is independent of attention. *BMC Neuroscience*, 14(2), 1–11. doi:10.1186/1471-2202-14-2
- Titova, N., & Näätänen, R. (2001). Preattentive voice discrimination by the human brain as indexed by the mismatch negativity. *Neuroscience Letters*, 308, 63–65. doi:10.1016/S0304-3940(01)01970-X
- Ventura, M. I., Nagarajan, S. S., & Houde, J. F. (2009). Speech target modulates speaking induced suppression in auditory cortex. *BMC Neuroscience*, 10(58), 1–11. doi:10.1186/1471-2202-10-58

- Vuilleumier, P., & Huang, Y.-M. (2009). Emotional attention: Uncovering the mechanisms of affective biases in perception. *Current Directions in Psychological Science*, 18, 148–152. doi:10.1111/j.1467-8721.2009.01626.x
- Ward, N. (2006). Non-lexical conversational sounds in American English. *Pragmatics and Cognition*, 14, 131–184. doi:10.1075/pc.14.1.08wa
- Ward, N., & Tsukahara, W. (2000). Prosodic features which cue back-channel responses in English and Japanese. *Journal of Pragmatics*, 32, 1177–1207. doi:10.1016/S0378-2166(99)00109-5
- Waters, F., & Badcock, J. C. (2010). First-rank symptoms in schizophrenia: Reexamining mechanisms of self-recognition. *Schizophrenia Bulletin*, 36, 510–517. doi:10.1093/schbul/sbn112
- 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. *Schizophrenia Bulletin*, 38, 741–750. doi:10.1093/schbul/sbq144
- Wechsler, D. (1997). *Escala de inteligência de Wechsler para adultos – 3ª Edição: Instruções para a administração e cotação* [Wechsler Adult Intelligence Scale – Third Edition: Administration and scoring manual]. (trad. e adapt. Portuguesa Cegoc-Tea, 2008). Lisboa: Cegoc-Tea.
- 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. *Psychological Medicine*, 41, 959–969. doi:10.1017/S0033291710001376
- Xu, M., Homae, F., Hashimoto, R., & Hagiwara, H. (2013). Acoustic cues for the recognition of self-voice and other-voice. *Frontiers in Psychology*, 4(735), 1–7. doi:10.3389/fpsyg.2013.00735
- Zheng, Z. Z., Munhall, K. G., & Johnsrude, I. S. (2010). Functional overlap between regions involved in speech perception and in monitoring one's own voice during speech production. *Journal of Cognitive Neuroscience*, 22, 1770–1781. doi:10.1162/jocn.2009.21324
- Zheng, Z. Z., Vicente-Grabovetsky, A., MacDonald, E. N., Munhall, K. G., Cusack, R., & Johnsrude, I. S. (2013). Multivoxel patterns reveal functionally differentiated networks underlying auditory feedback processing of speech. *Journal of Neuroscience*, 33, 4339–4348. doi:10.1523/JNEUROSCI.6319-11.2013