Brain & Language 153–154 (2016) 38–49

Contents lists available at ScienceDirect

Brain & Language

journal homepage: www.elsevier.com/locate/b&l

Did you or I say pretty, rude or brief? An ERP study of the effects of speaker's identity on emotional word processing

Ana P. Pinheiro [a,b,c,*], Neguine Rezaii [b], Paul G. Nestor [b,d], Andréia Rauber [e], Kevin M. Spencer [f], Margaret Niznikiewicz [b]

[a] Neuropsychophysiology Laboratory, Psychology Research Center (CIPsi), School of Psychology, University of Minho, Braga, Portugal
[b] Clinical Neuroscience Division, Laboratory of Neuroscience, VA Boston Healthcare System-Brockton Division, Department of Psychiatry, Harvard Medical School, Brockton, MA, United States
[c] Faculty of Psychology, University of Lisbon, Lisbon, Portugal
[d] Department of Psychology, University of Massachusetts, Boston, MA, United States
[e] International Studies in Computational Linguistics, University of Tübingen, Tübingen, Germany
[f] Neural Dynamics Laboratory, Research Service, VA Boston Healthcare System, and Department of Psychiatry, Harvard Medical School, Boston, MA, United States

ARTICLE INFO

Article history:
Received 24 June 2015
Revised 19 November 2015
Accepted 10 December 2015
Available online 16 February 2016

ABSTRACT

During speech comprehension, multiple cues need to be integrated at a millisecond speed, including semantic information, as well as voice identity and affect cues. A processing advantage has been demonstrated for self-related stimuli when compared with non-self stimuli, and for emotional relative to neutral stimuli. However, very few studies investigated self-other speech discrimination and, in particular, how emotional valence and voice identity interactively modulate speech processing. In the present study we probed how the processing of words' semantic valence is modulated by speaker's identity (self *vs.* non-self voice).

Sixteen healthy subjects listened to 420 prerecorded adjectives differing in voice identity (self *vs.* non-self) and semantic valence (neutral, positive and negative), while electroencephalographic data were recorded. Participants were instructed to decide whether the speech they heard was their own (self-speech condition), someone else's (non-self speech), or if they were unsure.

The ERP results demonstrated interactive effects of speaker's identity and emotional valence on both early (N1, P2) and late (Late Positive Potential – LPP) processing stages: compared with non-self speech, self-speech with neutral valence elicited more negative N1 amplitude, self-speech with positive valence elicited more positive P2 amplitude, and self-speech with both positive and negative valence elicited more positive LPP. ERP differences between self and non-self speech occurred in spite of similar accuracy in the recognition of both types of stimuli.

Together, these findings suggest that emotion and speaker's identity interact during speech processing, in line with observations of partially dependent processing of speech and speaker information.

1. Introduction

In the course of a conversation, people have to rapidly detect and integrate multiple signals in order to make sense of speech information. These signals include not only linguistic but also paralinguistic cues (Belin, Bestelmeyer, Latinus, & Watson, 2011; Schirmer & Kotz, 2006). From the perspective of a listener, it is not only important to understand "what" is being said and "how", but also to relate that information to "who" is saying it (Belin, Fecteau, & Bédard, 2004; Belin et al., 2011; Formisano, De Martino, Bonte, & Goebel, 2008). From the perspective of a speaker, it is important to distinguish between self-generated and non-self generated voices, i.e. to recognize speech as one's own.

Studies in the last two decades lend support to the idea that distinct types of information – speech, affect and identity – conveyed by the voice are processed by functionally dissociable neural pathways: the analysis of *speech* information recruits temporal (anterior and posterior superior temporal sulcus) and inferior prefrontal regions, particularly in the left hemisphere; the analysis of *vocal affect* recruits temporo-medial regions, the anterior insula, the amygdala and inferior prefrontal regions, particularly in the right hemisphere; the analysis of *vocal identity* recruits regions of the right anterior superior temporal sulcus (e.g., Belin et al., 2004, 2011). Nonetheless, the interactions between these distinct

* Corresponding author at: Neuropsychophysiology Lab, CIPsi, School of Psychology, University of Minho, Campus de Gualtar, 4710-957 Braga, Portugal.
E-mail address: ana.pinheiro@psi.uminho.pt (A.P. Pinheiro).

http://dx.doi.org/10.1016/j.bandl.2015.12.003
0093-934X/© 2016 Published by Elsevier Inc.

types of information remain largely unexplored. In spite of the limited number of studies probing the neuro-functional correlates of speaker's identity processing, there is evidence indicating that identity-related information is extracted and used from early stages of speech perception, within the first 200–300 milliseconds (ms) after a spoken word onset (Van Berkum, van den Brink, Tesink, Kos, & Hagoort, 2008). However, few studies have investigated the interaction between speaker's identity and speech semantic valence and, in particular, how the assessment of self-relevance (self *vs.* non-self voice) influences emotional language comprehension at the neural level.

1.1. The role of voice identity during speech perception

A considerable number of studies have demonstrated a processing advantage for self-related stimuli when compared with non-self stimuli, such as one's own name (Gray, Ambady, Lowenthal, & Deldin, 2004; Tacikowski & Nowicka, 2010; Zhao et al., 2009), self-related pronouns (Herbert, Herbert, Ethofer, & Pauli, 2011; Zhou et al., 2010), self-face (Keyes, Brady, Reilly, & Foxe, 2010; Sui, Zhu, & Han, 2006; Tacikowski & Nowicka, 2010), self-relevant objects (Miyakoshi, Nomura, & Ohira, 2007), or self-hand (Su et al., 2010).

Few studies probed the neural correlates of self-voice perception and, in particular, the time course of self-generated speech processing (Conde, Gonçalves, & Pinheiro, 2015, in press; Graux, Gomot, Roux, Bonnet-Brilhault, & Bruneau, 2015; Graux et al., 2013). More recent investigations suggest that self and non-self voices are discriminated within the first 100 ms after voice onset (e.g., Ford et al., 2001), and indicate high self-voice recognition rates (94–96% – Rosa, Lassonde, Pinard, Keenan, & Belin, 2008). ERP studies testing the corollary discharge mechanism (i.e., the expected sensory feedback resulting from one's own action) demonstrated effects on the N1 component, a negativity peaking between 80 and 120 ms after stimulus onset, and maximal at frontocentral electrodes locations (e.g., Rosburg, Boutros, & Ford, 2008). Generally, the N1 is sensitive to the physical properties of the stimuli, and has been proposed as a neurophysiological signature of automatic attention allocation to salient stimuli (Rosburg et al., 2008). Specifically, the studies that probed the corollary discharge showed that the N1 amplitude is reduced in response to the onset of unaltered self-voice auditory feedback during speech production in comparison with the passive listening to the recording of one's own voice (e.g., Baess, Widmann, Roye, Schröger, & Jacobsen, 2009; Behroozmand & Larson, 2011; Ford et al., 2001); or in response to self-triggered (i.e., button press) relative to externally-triggered tones (Behroozmand & Larson, 2011; Knolle, Schröger, & Kotz, 2013a, 2013b). As such, these studies expanded the functional significance of the N1, by demonstrating that this component is also sensitive to voice identity, and that it is specifically modulated by stimulus predictability or agency (e.g., being the author of the action of pressing a button to elicit a sound).

These experiments also reported effects on the P2 component, a positivity that is typically observed around 150–300 ms after stimulus onset, and has been associated with early stimulus categorization and attention processes (e.g., Crowley & Colrain, 2004), and more recently with the detection of the emotional salience of a stimulus (e.g., Paulmann & Kotz, 2008; Pinheiro et al., 2013, 2014). Specifically, these experiments demonstrated reduced P2 amplitude in response to self-generated sounds (Knolle et al., 2013b). These effects were interpreted as a neurophysiological signature of the conscious detection of a self-initiated auditory stimulus (Knolle et al., 2013b). However, a particular methodological feature of this type of paradigms is that they either involve active speech production or a motor condition, in which participants are asked to press a button in order to elicit a given sound (e.g., Knolle et al., 2013a, 2013b), with its disadvantages in terms of artifacts during EEG data acquisition (e.g., Ford et al., 2001).

Of relevance to the current study, evidence from experiments consisting of the passive listening to pre-recorded self-voice and non-self voice stimuli indicated that voice stimuli can be discriminated as a function of their identity in early stages of information processing. The differential sensitivity of the ERP response to self-relevance was corroborated by studies reporting effects around 100 ms. For example, Graux et al. (2013) observed that self and non-self pre-recorded voices were discriminated within the first 100 ms after voice onset: the self-voice was characterized by greater negative amplitude within this time window compared with the unfamiliar voice. Together, these studies reveal important effects of voice identity in early processing stages, i.e. within the first 200 ms after voice onset (N1, P2).

Moreover, in a recent study, we probed how self-relevance (pre-recorded self *vs.* non-self voice) modulates selective attention (Conde et al., 2015). We found that selective attention to voice stimuli was enhanced in the case of self-relevant ('my voice') compared to non-self ('someone else's voice) stimuli. This finding suggests that in experiments consisting of listening to task-relevant pre-recorded self *vs.* non-self voices, N1/P2 amplitude for the self-voice might be increased, rather than decreased as reported in experiments requiring speech production (e.g., Ford et al., 2001) or a button-press eliciting a sound (e.g., Knolle et al., 2013b). In this case, self-relevance may engender increased neural activity and larger ERP amplitude through an increase of attention (e.g., self-voice – Conde et al., 2015; self-name and self-face – Tacikowski & Nowicka, 2010). The finding of Conde et al. (2015) additionally indicated that a self-voice represents a particularly salient stimulus, suggesting that the self-relevance and the emotionality of a stimulus may produce similar effects.

1.2. The role of semantic valence in speech processing

A close relationship between the processing of self-related information and the processing of emotional valence (i.e., the overall unpleasantness/displeasure relative to pleasantness/attraction of a stimulus – Bradley & Lang, 1994) has been reported (e.g., Fossati et al., 2003). The existing evidence converges in showing that we quickly discriminate between emotionally salient and neutral stimuli, and that this differentiation occurs already at early processing stages, in the first 200 ms after speech onset (e.g., Paulmann & Kotz, 2008; Pinheiro et al., 2013, 2014). Specifically, larger P2 amplitude was observed for positive relative to neutral words (e.g., Kanske & Kotz, 2007; Kissler, Assadollahi, & Herbert, 2006), or for emotional (negative and positive) relative to neutral words (Bernat, Bunce, & Shevrin, 2001; Schapkin, Gusev, & Kühl, 2000), even when presented subliminally (Bernat et al., 2001). A putative explanation for these P2 effects is that they reflect increased automatic attentional capture by emotional stimuli (Kanske, Plitschka, & Kotz, 2011). Emotion effects were also observed in a later positive component, observed after 500 ms post-stimulus onset – the Late Positive Component (LPP). Some studies found a processing advantage for pleasant words, suggesting that they lead to increased sustained attention and deeper stimulus encoding relative to both negative and neutral words, which is reflected in increased amplitude of the LPP (e.g., Ferré, 2003; Herbert, Junghofer, & Kissler, 2008). Together, these studies show that the effects of the emotionality of verbal stimuli may be observed in both early (P2) and later (LPP) components. Thus, in addition to P2 indexing identity as described above (e.g., Knolle et al., 2013b), these findings highlight the relevance of the P2 to the investigation of both voice identity and emotion processing. We note that in case of self-relevant and emotional speech stimuli, these effects should be additive.