Neuropsychologia 114 (2018) 158–167

Contents lists available at ScienceDirect

Neuropsychologia

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

Detecting our own vocal errors: An event-related study of the thresholds for perceiving and compensating for vocal pitch errors

Nichole E. Scheerer, Jeffery A. Jones[*]

Psychology Department and Laurier Centre for Cognitive Neuroscience, Wilfrid Laurier University, Waterloo, Ontario, Canada

ARTICLE INFO

Keywords:
Dual-stream theory
Auditory feedback
Speech production
Speech Perception
Frequency-altered feedback
Pitch-shift reflex
Response thresholds

ABSTRACT

Previous studies suggest that a perception-action dissociation exists for the cortical processing of vocal pitch, because speakers compensate for small vocal errors without awareness. In this event-related potential (ERP) study, participants vocalized while hearing their productions either altered or unaltered in pitch, and reported whether their auditory feedback was altered. Pitch alterations as small as 10 cents resulted in compensatory vocal responses, while participants reported hearing perturbations that were 15 cents and larger. Similarly, P1 ERP responses were elicited by perturbations 15 cents and larger, while N1 responses followed a linear trend with increasing perturbation magnitudes, and P2 responses were elicited by perturbations 30 cents and larger. Although their thresholds differed, both motor and perceptual responses were elicited by small frequency altered feedback (FAF) perturbations. Previous reports of a perception-action dissociation may reflect differences in the magnitude of vocal error required to elicit a motor response, and for an individual to report a pitch change, rather than to detect a pitch change (as reflected by ERP responses).

1. Introduction

Ungerleider and Mishkin (1982) theorized that visual information processing is subserved by separate dorsally and ventrally located processing streams. These anatomically divergent streams were proposed to support the processing of object locations and identities, respectively. Later, Goodale and Milner (1992) refined this dual stream notion by suggesting that the dorsal stream is in fact responsible for visual-motor transformations that support vision for action, while the ventral stream supports object recognition for perception. More recently, it has been postulated that a similar motor and perceptual segregation may exist in the auditory domain. According to this theory, a dorsal stream projecting from the auditory cortex to the parietal-temporal boundary of the Sylvian fissure (area Spt) plays a role in auditory-motor transformations, while a ventral stream projecting from the auditory cortex towards the inferior posterior temporal cortex plays a role in linking auditory information to conceptual representations (Hickok, 2012; Hickok et al., 2011; Hickok and Poeppel, 2000, 2004, 2007). Although this dual stream theory of auditory processing makes clear predictions about the processing of lexical aspects of speech, it is currently unclear how sub-lexical aspects of speech, such as pitch, fit into this dual stream framework.

One method commonly used to assess both perceptual and motor responses to changes in pitch during the processing of speech is the frequency altered feedback (FAF) paradigm (Elman, 1981; Burnett et al., 1997). As part of this paradigm, participants produce vocalizations while the fundamental frequency (F0) of their auditory feedback is shifted upwards or downwards. Upon exposure to this FAF, speakers tend to compensate for the FAF by adjusting the pitch of their voice in the opposite direction of the feedback alteration. These compensatory responses are often only a fraction of the size of the manipulation, suggesting that these responses are optimal for correcting for small feedback alterations, as larger feedback alterations are only partially compensated for (Burnett et al., 1997, 1998; Korzyukov et al., 2012b; Liu et al., 2011; Scheerer et al., 2013a, 2013b; Scheerer and Jones, 2014). For this reason, it has been suggested that the function of this response is to stabilize pitch around a desired target (Hain et al., 2000; Hawco and Jones, 2009; Natke et al., 2003). Importantly, this research has shown that individuals compensate for FAF alterations as small as 25 cents (100 cents = 1 semitone; Burnett et al., 1998; Liu and Larson, 2007), which indicates that the speech motor control system is highly sensitive to even small feedback alterations.

This FAF paradigm was utilized by Hafke (2008) in an attempt to illustrate a 'perception-action' dissociation, and to provide support for the notion that pitch is processed by dual processing streams. During this experiment, trained singers produced vocalizations while exposed to unaltered feedback and FAF, which was either 9, 19, 50, or 99 cents in magnitude. After each vocalization, participants reported whether or

---

[*] Correspondence to: Psychology Department, Wilfrid Laurier University, Waterloo, Ontario, Canada N2L 4A6.
E-mail address: jjones@wlu.ca (J.A. Jones).

https://doi.org/10.1016/j.neuropsychologia.2017.12.007
Received 6 August 2015; Received in revised form 1 December 2017; Accepted 3 December 2017

not they perceived a change in their pitch during the vocalization. These perceptual judgements were used to determine the magnitude at which participants became aware of changes in their auditory feedback, or their 'perceptual threshold.' In addition, the participants' compensatory responses to the FAF were assessed in order to determine the magnitude of FAF alteration required to elicit a motor response. The results of this study suggest that on average, the participants' perceptual thresholds for detecting FAF alterations were around 21 cents. On the other hand, participants compensated for FAF alterations as small as 9 cents, which suggests that the motor system is more sensitive to FAF than the perceptual system. In a follow up study, the authors had participants perform a psychophysical task, where the participants were required to make perceptual judgements while listening to recordings of the vowel sound /u/ that varied in frequency (Wrzosek et al., 2013). Following the psychophysical experiment, participants also took part in an electrophysiological experiment where the participants' electroencephalographic (EEG) activity was recorded while they listened to recordings of the vowel /u/ that also varied in frequency. The EEG responses recorded were used to calculate the mismatch negativity (MMN), an event-related potential (ERP) that is often used to index the detection of change in a stimulus. Much like the study reported by Hafke (2008), the results of this study indicated that participants were able to detect changes in pitch as small as 27 cents during the psychophysical experiment. In contrast, the MMN was only elicited by pitch changes 50 cents and larger. Based on the results of these two studies, the authors concluded that the different magnitudes of FAF alterations required to elicit perceptual awareness, whether indexed by perceptual judgements or EEG, and compensatory motor responses, provides support for a perception-action dissociation and dual stream processing of auditory information.

Similarly, Loui et al. (2008) demonstrated that individuals with congenital amusia, or impaired pitch perception, were able to sing pitch intervals in a specified direction, despite being unable to correctly categorize the direction of the same intervals. Hutchins and Peretz (2013) also worked with a group of individuals with congenital amusia, and found that although they produced compensatory motor responses to FAF, these individuals were unable to consciously detect the same pitch changes. Together, these results provide additional support for the notion that information from auditory feedback is processed differently for perception and action. That being said, more recently Williamson et al. (2012) have reported evidence for a bidirectional dissociation of perception and production in individuals with amusia. Based on these reports of both impaired perception and production in individuals with amusia, Williamson et al. (2012) argue that the auditory deficits witnessed in individuals with amusia provides poor support for a functional dissociation between the processing of auditory information for perception and action.

Although compensatory motor responses to FAF, and perceptual judgments of changes in one's auditory feedback, are useful for assessing the processing of pitch during ongoing speech, EEG and magnetoencephalographic (MEG) responses recorded during ongoing speech can provide useful information about the auditory-cortical processing of vocal pitch. As demonstrated by Wrzosek et al. (2013), the MMN can be used to index the detection of change in an auditory stimulus (see also Hawco et al., 2009). However, the P1-N1-P2 event-related potentials are more commonly used to assess changes in auditory-cortical processing following exposure to FAF (Behroozmand et al., 2014, 2009, 2011; Heinks-Maldonado et al., 2005; Korzyukov et al., 2012a, 2012b; Liu et al., 2011; Scheerer et al., 2013a, 2013b; Scheerer and Jones, 2014; Tumber et al., 2014). Previous research has suggested that the P1 component reflects the basic detection of deviant auditory feedback, as FAF perturbations of varying magnitudes elicit P1 responses that do not differ significantly in magnitude (Scheerer et al., 2013a). For this reason, it has been suggested that the P1 component is not sensitive to the magnitude of deviant auditory feedback, rather the P1 component increases in an all-or-nothing manner to deviant auditory feedback (Scheerer et al., 2013a). On the other hand, the N1 component, and its magnetic equivalent the M1, have been shown to be maximally attenuated during the perception of one's own unaltered speech, relative to FAF (Heinks-Maldonado et al., 2005, 2006) and passive listening (Ford et al., 2001; Heinks-Maldonado et al., 2005, 2006; Houde et al., 2002). It has been suggested that this neural modulation reflects suppression in the auditory cortex during the perception of one's own unaltered speech, as a result of a match between the perceived auditory feedback, and a prediction of the expected sensory feedback issued by the motor system during the execution of the motor commands for speech. Although the N1 component is suppressed during the processing of one's own unaltered speech, it has been shown to increase in magnitude in response to FAF (Heinks-Maldonado et al., 2005; Scheerer and Jones, 2014; Scheerer et al., 2013a, 2013b). Similarly, previous FAF studies have shown that the amplitude of the P2 ERP component increases linearly with increasing feedback perturbation magnitudes (Behroozmand et al., 2009; Scheerer et al., 2013a). Together, these results suggest that the P1-N1-P2 ERP components, coupled with the FAF paradigm, may provide a useful means for objectively assessing both perceptual and motor responses to changes in pitch during ongoing speech, respectively.

Although previous studies have highlighted differences in the magnitude of FAF alterations required for participants to report the detection of FAF, and produce compensatory responses to FAF (Hafke, 2008; Wrzosek et al., 2013), these studies determined perceptual thresholds by interpolating responses based on measured responses to feedback alterations of a small number of magnitudes. The aim of the current study was to obtain a more precise estimate of the magnitude at which participants produce compensatory responses to FAF, by presenting participants with brief pitch perturbations during active vocalization that ranged in magnitude from 0 to 40 cents. Following each vocalization, participants were required to report whether or not they detected a change in their pitch, which allowed the threshold at which participants detected errors in their auditory feedback to be determined. In addition to these perceptual judgements, the magnitudes of the P1-N1-P2 ERP responses elicited by the FAF were also measured, as previous research has shown that these components are readily elicited by FAF, with changes in the amplitudes of these components reflecting the neural processing involved in the detection and correction of vocal errors (Behroozmand et al., 2009, 2011; Scheerer et al., 2013a, 2013b). While the primary aim of this study was to further the investigation of a potential perception-action dissociation in the processing of vocal pitch by comparing P1-N1-P2 ERP responses to vocal responses elicited by FAF, the current FAF literature is limited in its reports of P1-N1-P2 ERP responses to FAF perturbation magnitudes smaller than 50 cents. For this reason, the results of this study also provide valuable information to aid in our understanding of how these ERP components are modulated by small, ecologically valid changes in auditory feedback. Since previous research has shown that vocal responses are elicited by FAF alterations as small as 9 cents (Hafke, 2008), we expected that all FAF perturbation magnitudes utilized in this study (5, 10, 15, 20, 25, 30, & 40 cents) would elicit compensatory vocal responses. On the other hand, as previous research suggests that larger magnitude FAF alterations are required for the conscious detection of pitch changes (Hafke, 2008; Wrzosek et al., 2013), we expected that reports of a change in pitch would occur at a larger magnitude than those required to elicit a compensatory vocal response. Although to our knowledge this is the first study to investigate P1-N1-P2 ERP responses to FAF perturbations smaller than 50 cents, previous research suggests that the P1 ERP component increases in an all-or-nothing manner to FAF (Scheerer et al., 2013a), thus we expected that all magnitudes of FAF perturbations would elicit a P1 response. On the other hand, the N1 component is often only sensitive to larger FAF alterations (Scheerer et al., 2013a), thus we only expected to find N1 responses following the larger magnitude FAF perturbations, if at all. Lastly, the P2 ERP component has been found to increase in a linear fashion as the size of the