Brain & Language 165 (2017) 33–44

Contents lists available at ScienceDirect

Brain & Language

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

Voice-sensitive brain networks encode talker-specific phonetic detail

Emily B. Myers [a,b,c,d,*], Rachel M. Theodore [a,c,d]

[a] University of Connecticut, Department of Speech, Language, and Hearing Sciences, 850 Bolton Road, Unit 1085, Storrs, CT 06269-1085, United States
[b] University of Connecticut, Department of Psychological Sciences, 406 Babbidge Road, Unit 1020, Storrs, CT 06269-1020, United States
[c] Haskins Laboratories, 300 George Street, Suite 900, New Haven, CT 06511, United States
[d] Connecticut Institute for the Brain and Cognitive Sciences, 337 Mansfield Road, Unit 1272, Storrs, CT 06269-1085, United States

ARTICLE INFO

Article history:
Received 5 April 2016
Revised 13 September 2016
Accepted 4 November 2016
Available online 27 November 2016

Keywords:
Voice recognition
Phonetic processing
Superior temporal gyrus
VOT

ABSTRACT

The speech stream simultaneously carries information about talker identity and linguistic content, and the same acoustic property (e.g., voice-onset-time, or VOT) may be used for both purposes. Separable neural networks for processing talker identity and phonetic content have been identified, but it is unclear how a singular acoustic property is parsed by the neural system for talker identification versus phonetic processing. In the current study, listeners were exposed to two talkers with characteristically different VOTs. Subsequently, brain activation was measured using fMRI as listeners performed a phonetic categorization task on these stimuli. Right temporoparietal regions previously implicated in talker identification showed sensitivity to the match between VOT variant and talker, whereas left posterior temporal regions showed sensitivity to the typicality of phonetic exemplars, regardless of talker typicality. Taken together, these results suggest that neural systems for voice recognition capture talker-specific phonetic variation.

1. Introduction

Individual talkers differ in how they implement phonetic properties of speech (e.g., Allen, Miller, & DeSteno, 2003; Hillenbrand, Getty, Clark, & Wheeler, 1995; Newman, Clouse, & Burnham, 2001; Peterson & Barney, 1952; Theodore, Miller, & DeSteno, 2009). This kind of systematic talker-specific, within-category phonetic variation contributes to one's *idiolect*, that is, one's vocal identity. It has long been known that within-category phonetic variation is not discarded by the perceptual system; rather, it is used probabilistically to constrain and facilitate linguistic processing at both prelexical and lexical levels of representation (Andruski, Blumstein, & Burton, 1994; McMurray, Tanenhaus, Aslin, & Spivey, 2003; Myers, 2007; Pisoni & Tash, 1974; Utman, Blumstein, & Sullivan, 2001). Indeed, behavioral evidence suggests that listeners can perceptually track talker-specific phonetic variation and simultaneously use this information to identify *who* is doing the talking (e.g., Theodore & Miller, 2010) and to facilitate processing of *what* is being said (e.g., Nygaard, Sommers, & Pisoni, 1994). Of interest for the current study, the same acoustic cues (e.g., voice-onset-time values specifying stop consonants, formant patterns specifying vowels) are useful for both the *who* and *what* purposes (e.g., Theodore, Myers, & Lomibao, 2015). While much is known about the neural systems that underlie the processing of talker identity and those involved in processing the phonetic details of speech (e.g. Blumstein & Myers, 2014), it is unclear to what extent the neural systems that process talker information and phonetic information are dissociable or mutually interactive, particularly in the context of speech variants that contribute to a talker's idiolect. Below we review evidence from behavioral and neuroimaging paradigms that inform this question.

1.1. Interactive processing of phonetic and talker information

Behavioral examinations have revealed a tight link between the processing of phonetic and talker information (e.g., Theodore & Miller, 2010; Theodore et al., 2015). With respect to the processing of phonetic information, listeners receive comprehension benefits for familiar compared to unfamiliar talkers including heightened word recognition in degraded listening environments (e.g., Nygaard et al., 1994) and faster processing times (e.g., Clarke & Garrett, 2004). Research suggests that the processing benefits observed at higher levels of linguistic processing (e.g., word recognition) reflect adjustments that listeners make earlier in the perceptual stream (e.g., Nygaard & Pisoni, 1998). Listeners can learn a talker's characteristic VOT production for word-initial voiceless stops, indicating sensitivity to talker differences in individual pho-

* Corresponding author at: Department of Speech, Language, and Hearing Sciences, University of Connecticut, 850 Bolton Rd, Unit 1085, Storrs, CT 06269-1085, United States.
E-mail address: emily.myers@uconn.edu (E.B. Myers).

http://dx.doi.org/10.1016/j.bandl.2016.11.001
0093-934X/

netic properties of speech (Allen & Miller, 2004; Theodore & Miller, 2010). Moreover, exposure to a talker's characteristic productions promotes a comprehensive reorganization of phonetic category structure such that behavioral judgments of phonetic category prototypicality reflect experience with individual talkers' characteristic productions of those phonetic categories (Theodore et al., 2015). With respect to the processing of talker information, research has shown that voice recognition is heightened in the native compared to a non-native language (Goggin, Thompson, Strube, & Simental, 1991; Xie & Myers, 2015). The language familiarity benefit for voice recognition has been linked to experience and expertise with phonetic variation associated with linguistic sound structure (Johnson, Westrek, Nazzi, & Cutler, 2011; Orena, Theodore, & Polka, 2015). As argued in Perrachione and Wong (2007), the reliance of talker identification processes on phonetic qualities of the input points to a shared neural substrate for perception of talker identity and phonetic characteristics. Collectively, these behavioral findings demonstrate that the processing of phonetic and talker information are fundamentally linked with respect to spoken language processing, and further suggest that their linkage emerges at a sublexical level of representation.

1.2. Neural systems for processing phonetic variation and talker information

The neural systems that extract phonetic content from the speech signal and those that extract information about vocal identity are partially overlapping, but can be argued to recruit different circuits (Blumstein & Myers, 2014; Van Lancker, Kreiman, & Cummings, 1989). For the purposes of phonetic processing, within-category phonetic variability is processed in the bilateral posterior superior temporal gyrus (STG) (see Blumstein & Myers, 2014 for review; Chang et al., 2010; Liebenthal, Binder, Spitzer, Possing, & Medler, 2005; Myers, 2007). Of interest, regions in the bilateral superior temporal lobes show tuning to the best exemplars of one's native language phonetic category, with greater activation observed for phonetic tokens that are less typical as a member of the category (e.g., a /t/ with an extremely long VOT) compared to more standard productions (Myers, 2007). These core phonetic processing regions are permeable to at least some top-down influences. For example, these regions show differences in sensitivity to tokens along a phonetic continuum when the phonetic category boundary has been shifted by embedding a token in a biasing lexical or sentential context (Gow, Segawa, Ahlfors, & Lin, 2008; Guediche, Salvata, & Blumstein, 2013; Myers & Blumstein, 2008). This said, it is unclear whether every source of information—and in particular, whether a given token is typical of a *talker's* voice—modulates the tuning of the STG to native language typicality.

While phonetic processing is thought to be bilateral, with some preference for leftwards laterality, processing of vocal identity has largely been attributed to right hemisphere regions (e.g. Van Lancker et al., 1989; von Kriegstein, Eger, Kleinschmidt, & Giraud, 2003). A classic study by Van Lancker et al. (1989) tested a group of individuals with left and right hemisphere lesions on identification of familiar voices as well as voice discrimination. Voice discrimination was impaired in individuals with both left and right hemisphere temporal lesions, but identification of familiar voices was impaired in individuals with right inferior parietal lesions (see also Van Lancker, Cummings, Kreiman, & Dobkin, 1988). Imaging studies have further corroborated the separation between regions that are sensitive to the acoustics of the voice—and thus could be used for discriminating between talkers—and those responsible for mapping voice acoustics to an individual identity which can be used for talker identification (von Kriegstein et al., 2003). In particular, while voice acoustics may be processed in bilateral temporal regions (specifically the superior temporal sulcus or STS), imaging studies have sited vocal identity processing (or access to familiar voices) in the anterior right temporal lobe rather than the right posterior region implicated in lesion studies (Andics, McQueen, & Petersson, 2013; Andics et al., 2010; Belin & Zatorre, 2003; Campanella & Belin, 2007).

Other evidence corroborates the role of either right anterior temporal or posterior temporoparietal regions for processing vocal identity (Belin & Zatorre, 2003; Stevens, 2004; von Kriegstein et al., 2003). For instance, von Kriegstein et al. (2003) showed that shifts in attention to vocal identity resulted in shifts in activation to right STS (see also, Belin & Zatorre, 2003), and revealed a gradient of processing such that anterior regions did not differentiate between familiar and unfamiliar voices, whereas posterior regions responded more to unfamiliar than familiar voices (Kriegstein & Giraud, 2004). A role for the right hemisphere in processing vocal identity is also supported by evidence that right frontal (middle frontal gyrus, MFG) and right inferior parietal regions (angular gyrus, AG) respond to short-term memory for talker identity (Stevens, 2004). Other studies have implicated bilateral temporal structures in processing changes in vocal identity when the linguistic message was held constant (Salvata, Blumstein, & Myers, 2012; Wong, Nusbaum, & Small, 2004). Of note, these latter studies cannot determine whether regions that respond to changes in vocal identity are those responsible for processing that identity itself, or whether they instead respond to other characteristics of the stimuli (e.g., differences in sensitivity to low-level acoustic properties that happen to differ across talkers). Notably, regions that respond to voice acoustics are more likely to be shared with the linguistic system, simply because the same properties of the acoustic signal can carry information about talker identity as well as linguistic content. The co-dependence of shared acoustic cues for phonetic processing and talker identity is highlighted in a study by von Kriegstein, Smith, Patterson, Kiebel, and Griffiths (2010). In this study, the perception of vocal tract length was manipulated by shifting the formant structure of utterances—crucially, this manipulation results in a change in the percept of talker identity that is signaled by an acoustic cue (formant structure) that is also used for vowel identity. In this study, the right STS/STG response to perceived vocal tract length (or talker identity) was amplified when participants were engaged in a task that focused attention at the phonetic level. Taken together, these studies suggest that neural systems arrayed along the right temporal lobe process talker-level information.

Andics et al. (2010) proposed that not only are the processing stages that process voice acoustics and voice identity separable, but that there are two sets of neural coding spaces, a 'voice-acoustics' space and 'voice-identity' space, each of which codes prototypical members of that space more sparsely then items distant from the prototype. This research group showed that the same core temporal lobe regions found to be sensitive to phonetic processing in other studies (see Myers, 2007; Myers, Blumstein, Walsh, & Eliassen, 2009) were sensitive to the internal structure of a learned voice identity space, showing less activation for stimuli that were more prototypical of a learned voice and greater activation for stimuli that were less typical of the talker's voice. Notably, in this study, the stimulus space was defined as a morph between two talker's voices, and therefore the continuum was likely to vary in voice-diagnostic features such as pitch contour and timbre, but potentially also in phonetic variation between the two talkers, although these details were not specified in the report. Nonetheless, a separate set of regions in anterior temporal areas was found to be sensitive to talker identity when controlling for the acoustic distinctions along the continuum, and only these anterior temporal regions correlated with identification performance. The authors take this finding as evidence that regions that