Journal of Memory and Language 76 (2014) 80–93

Contents lists available at ScienceDirect

Journal of Memory and Language

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

Neural systems underlying perceptual adjustment to non-standard speech tokens

Emily B. Myers [a,b,c,d,*], Laura M. Mesite [c,d]

[a] University of Connecticut, Department of Speech, Language, and Hearing Sciences, 850 Bolton Road, Storrs, CT 06269, United States
[b] University of Connecticut, Department of Psychology, 406 Babbidge Road, Storrs, CT 06269, United States
[c] Brown University, Department of Cognitive, Linguistic, and Psychological Sciences, 190 Thayer Street, Providence, RI 02912, United States
[d] Haskins Laboratories, 300 George Street #900, New Haven, CT 06511, United States

ARTICLE INFO

Article history:
Received 2 December 2013
revision received 14 June 2014

Keywords:
Speech perception
Neuroimaging
fMRI
Talker perception

ABSTRACT

Evidence suggests that listeners use top-down information to guide perception of speech sounds. A phenomenon termed 'perceptual learning for speech' shows that listeners also use top-down information to adjust perceptual boundaries in subsequent processing of speech from the same talker. The neural mechanisms that underlie this process are not well understood. Of interest is whether boundary shifts arise because of a retuning of phonetic sensitivities early in the neural processing stream or whether they result from decision-related or attentional mechanisms further downstream. In the current study, activation was measured using fMRI as participants underwent a behavioral replication of this paradigm. Sensitivity to boundary shifts emerged in right frontal and middle temporal regions, implicating adjustments of perceptual criteria downstream from primary auditory cortex. Later in the session, this same sensitivity emerged in left superior temporal areas, implicating a slowly-adapting system in regions of the brain related to phonetic processing.

Introduction

The speech we listen to is full of perceptual problems to be solved. Slips of the tongue produce imperfect or ambiguous speech tokens, environmental noise may occlude the speech signal, and different talkers, for reasons of idiolect or dialect, implement speech sounds in ways that consistently deviate from 'standard' pronunciations (e.g. Allen, Miller, & DeSteno, 2003). An important question in speech perception is how listeners resolve such indeterminacies in order to map the speech signal to meaning. Solving this problem requires a degree of functional plasticity in our speech perception system.

A line of research suggests that listeners can capitalize upon regularities in the speech signal in order to adjust phonetic category boundaries. One source of information that may help listeners resolve ambiguities in the speech signal is top-down context. Evidence suggests that listeners use the local linguistic context to adjust their responses to phonetic input, particularly when that input is ambiguous. For example, Ganong (1980) observed that lexical context shifted participants' phonetic category boundaries in a categorization task, such that an ambiguous token midway between/g/ and /k/ was likely to be perceived as a /g/ sound in the context of a gift—kift continuum, and as a /k/ sound in the context of a giss—kiss continuum. These types of boundary adjustments have been shown as a

* Corresponding author at: University of Connecticut, Department of Speech, Language, and Hearing Sciences, 850 Bolton Road, Storrs, CT 06269, United States.

E-mail address: Emily.myers@uconn.edu (E.B. Myers).

http://dx.doi.org/10.1016/j.jml.2014.06.007
0749-596X/

function of other lexical phenomena, such as neighborhood density and lexical frequency, and from semantic and syntactic expectation (Borsky, Tuller, & Shapiro, 1998; Connine, Titone, & Wang, 1993; Newman, Sawusch, & Luce, 1997). Taken together, these results suggest that disambiguating lexical, sentential, or semantic information in the local linguistic context plays a role in guiding the resolution of phonetic ambiguity.

At the same time, top-down information not only plays a role in shaping responses to local perceptual indeterminacies, but also in guiding longer-term adjustments to talker idiolect. A line of research in recent years has extended Ganong-type effects to investigate *persisting* effects of lexically-conditioned shifts in phonetic perception, termed 'perceptual learning in speech' (Norris, McQueen, & Cutler, 2003). In a series of related studies, researchers demonstrated that listeners who are exposed to multiple exemplars of an ambiguous phoneme blend (e.g. one between 's'-/s/ and 'sh'-/ʃ/) in an unambiguous lexical context (e.g. in the place of the 's' in 'Tennessee'), show subsequent shifts in the categorization function for items along an unbiased non-word to non-word continuum (e.g. /asi/ to /aʃi/) (see among others Clarke-Davidson, Luce, & Sawusch, 2008; Kraljic & Samuel, 2005, 2007; Norris et al., 2003).

One interpretation of these studies is that phonetic processing must be permeable to the effects of top-down information. Crucially, the phonetic processing system must adjust to top-down expectation both to resolve local perceptual indeterminacies (see the Ganong effect, above) but also to accommodate phonetic detail as it varies between talkers. However, this view has not gone unchallenged. In particular, it has been argued that lexically-guided phonetic category adjustments result not from an influence of top-down information on a phonetic level of processing, but instead from post-perceptual processing (McClelland, Mirman, & Holt, 2006; Norris et al., 2003).

Neuroimaging studies have informed this debate. Two investigations of the classic lexical effect phenomenon reported that activation in the left superior temporal gyrus (STG) was modulated by the top-down effects of lexical context (Gow, Segawa, Ahlfors, & Lin, 2008; Myers & Blumstein, 2008). The lexically-biased shift in the location of the phonetic category boundary was reflected by modulation of activity in the left STG as well as left inferior frontal areas (Myers & Blumstein, 2008). Data from a multi-modal imaging study (Gow et al., 2008) suggested that lexically-modulated shifts in perception may arise from the early interaction of information in the supramarginal gyrus (SMG) with fine-grained phonetic processing in the STG. The fact that lexically-biased shifts in perception are reflected in regions relatively early in the *neural* processing stream gives rise to the suggestion that lexical information contacts phonetic stages of processing quite early in the *language* processing stream.

What is unclear, however, is how neural systems underlying phonetic processing respond when lexical information guides *longer-term* adjustment of the phonetic category boundary, as in the case of a listener adapting to a new talker's accent. This question differs from that investigated in Ganong effect studies in that rather than focusing on the integration of top-down information during phonetic processing (i.e. during lexical decision), the question is how this top-down information is stored and used for subsequent processing. When listeners use lexical information to adjust to a talker idiosyncrasy, either online or offline use of lexical information disambiguates the ambiguous sounds during the lexical decision task. As listeners are exposed to increasing numbers of these resolved ambiguities, shifts in categorization for speech sounds for that talker begin to emerge. At least two non-exclusive mechanisms for these shifts are possible. First, it may be the case that exposure to disambiguated non-standard speech variants over time will fundamentally re-tune the responsiveness of populations of neurons that respond to phonetic category identity and phonetic category structure in the superior temporal lobes (Chang et al., 2010; Desai, Liebenthal, Waldron, & Binder, 2008; Guenther, Nieto-Castanon, Ghosh, & Tourville, 2004; Myers, Blumstein, Walsh, & Eliassen, 2009). On the other hand, one might imagine that dynamically adjusting the perceptual tuning for non-standard sounds for each talker that we encounter might introduce instability into the phonetic processing system. Given that these non-standard variants deviate significantly from the expected phonetic category shape, it may instead be the case that the processing of such tokens is guided by decision-related or attentional mechanisms. Specifically, a new decision criterion may be generated for non-standard tokens such that they are processed according to the talker's novel speech pattern.

The goal of the current study was to investigate the neural systems that underlie lexically-mediated perceptual learning for speech using fMRI. While in the scanner, participants listened to speech from a talker in which an ambiguous phonetic token between /s/ and /ʃ/ substituted for either an /s/ sound (S Group) or a /ʃ/ sound (SH Group). Subsequent shifts in the location of the phonetic category boundary were assessed by examining categorization functions for stimuli sampled from a non-word to non-word, /asi/ to /aʃi/ continuum. Behavioral responses as well as activation to the ambiguous tokens were measured, and data from a separate control group who did not participate in scanning was used to estimate the average location of the 'unshifted' phonetic category boundary in the population. Behaviorally, we predicted that there would be a difference in the location of the phonetic category boundary across groups, such that the S Group would show more 's' categorizations for tokens near the category boundary, and the SH Group would show the opposite pattern, thus replicating previous studies using this paradigm (e.g. Kraljic & Samuel, 2005).

Previous imaging work has demonstrated increased activation in both frontal and temporal areas for phonetic tokens which fall on the phonetic category boundary compared to those further from the boundary (Binder, Liebenthal, Possing, Medler, & Ward, 2004; Myers & Blumstein, 2008; Myers, 2007). This increase in activation for near-boundary tokens may reflect increased perceptual and executive resources necessary for categorization of ambiguous tokens. While many regions are likely to show greater activation to boundary value tokens, we predicted that only a subset would show sensitivity to the shifted