



Learning a talker or learning an accent: Acoustic similarity constrains generalization of foreign accent adaptation to new talkers



Xin Xie ^{a,*}, Emily B. Myers ^b

^a University of Rochester, United States

^b University of Connecticut, United States

ARTICLE INFO

Article history:

Received 11 August 2016

Revision received 9 July 2017

Available online 24 July 2017

Keywords:

Perceptual learning

Foreign-accented speech

Generalization

Spoken word recognition

Talker specificity

Adaptation

ABSTRACT

Past research has revealed that native listeners use top-down information to adjust the mapping from speech sounds to phonetic categories. Such phonetic adjustments help listeners adapt to foreign-accented speech. However, the mechanism by which talker-specific adaptation generalizes to other talkers is poorly understood. Here we asked what conditions induce cross-talker generalization in talker accent adaptation. Native-English listeners were exposed to Mandarin-accented words, produced by a single talker or multiple talkers. Following exposure, adaptation to the accent was tested by recognition of novel words in a task that assesses online lexical access. Crucially, test words were novel words and were produced by a novel Mandarin-accented talker. Results indicated that regardless of exposure condition (single or multiple talker exposure), generalization was greatest when the talkers were acoustically similar to one another, suggesting that listeners were not developing an accent-wide schema for Mandarin talkers, but rather attuning to the specific acoustic-phonetic properties of the talkers. Implications for general mechanisms of talker generalization in speech adaptation are discussed.

© 2017 Elsevier Inc. All rights reserved.

Introduction

Speech perception requires listeners to extract a meaningful message out of a highly variable and sometimes ambiguous signal. Dominant among many sources of variability are talker differences. Each speaker represents a unique combination of age, gender, vocal tract anatomy, idiosyncratic speaking style, and long-term language experience (e.g., regional dialect, native or non-native, bilingual or monolingual). Talker variability is manifested as a very wide variety of audible acoustic-phonetic variation in speech production, which further leads to differences in perceptual tasks (e.g., Peterson & Barney, 1952; Allen & Miller, 2004). Despite this variation, listeners efficiently identify spoken words across novel talkers, at least in most scenarios of native communication.

In order to understand how listeners accommodate talker variability, a large body of work has investigated how speech perception can be brought back to 'normal' (or at least, can progress in this direction) in *atypical* communication scenarios. For instance, in initial encounters with acoustically-distorted speech or non-standard speakers (e.g., foreign-accented), listeners typically experience greater perceptual difficulty (e.g., Dupoux & Green, 1997;

Munro & Derwing, 1995; Clarke & Garrett, 2004). However, as listeners gain more experience with the particular type of speech variation, comprehension improves, sometimes within a few minutes (e.g., Dahan & Mead, 2010; Norris, McQueen, & Cutler, 2003; Maye, Aslin, & Tanenhaus, 2008). In cases where the phonetic deviation is associated with a particular talker (e.g., an unfamiliar accent), listeners are remarkably adept in learning the idiosyncratic acoustic details of *specific* talkers and thereby demonstrating experience-dependent adaptation (e.g., Bradlow & Bent, 2008; Dahan & Mead, 2010; Norris et al., 2003; Kraljic & Samuel, 2005, 2006, 2007). Critically, evidence suggests that as listeners adapt to non-standard speech, they modify existing phonetic representations used to evaluate standard speech and form a separate sound-to-category mapping for the adapted (nonstandard) talker (e.g., Dahan, Drucker, & Scarborough, 2008; Xie, Theodore, & Myers, 2017).

What remains unclear is how listeners draw on these recent individual-based learning experiences in perceiving novel talkers to whom they have no direct exposure. As we review below, despite much progress in documenting talker-specific perceptual improvements, results are ambiguous concerning the necessary conditions required for successful generalization across talkers (Bradlow & Bent, 2008; Sidaras, Alexander, & Nygaard, 2009; Kraljic & Samuel, 2007; Reinisch & Holt, 2013). At the core of this question is whether listeners represent speech episodically, that

* Corresponding author.

E-mail address: xxie13@ur.rochester.edu (X. Xie).

is, packaging talker-specific acoustic detail together with linguistic information in memory, or whether listeners abstract away from talker-specific acoustic phonetic detail (Goldinger, 1998; Johnson, 2006; Pierrehumbert, 2006). This debate has led to recent hybrid accounts that allow for intermediate levels of talker knowledge (i.e., either by grouping talkers into higher-order categories, or by forming generative speaker models; Johnson, 2013; Kleinschmidt & Jaeger, 2015). As we shall elaborate in greater detail below, even these hybrid accounts have relatively little to say about how those “higher-order categories” are formed for talker representation or what factors aid the selection of “speaker models” during adaptation. Here, we present three experiments using a phonetic adaptation paradigm to explore the processes by which listeners generalize experience of particular foreign-accented talkers to novel talkers. We begin by briefly noting why foreign accent adaptation is a good place to look for evidence of cross-talkers generalization. In relation to that, we describe evidence of talker-specific adaptation, either in the context of foreign-accented speech or native speech. Then we consider some empirical gaps and discuss the theoretical implications of closing these gaps, before laying out the specific goals of the paper and the general methods used to achieve these goals.

Perceiving foreign-accented speech is a particularly challenging task. Foreign-accented speech not only contains idiolectal differences seen in native-accented speech (for instance, a talker might have a personal tendency to raise pitch at the end of a phrase), but additionally presents global deviations from native language categories. These deviations are manifested as differences in the acoustic distributions of speech tokens along multiple dimensions for multiple categories (e.g., Flege, Munro, & Skelton, 1992), making recognition of non-native speech effortful and often times, inaccurate (e.g., Munro & Derwing, 1995). A classic example of this phenomenon is vowel assimilation for Spanish-accented speakers of English. Because Spanish does not have the vowel /i/ as in ‘pick’, native speakers of Spanish will often produce this word closer to the nearby vowel /i/, as in ‘peek’, which exists in both Spanish and English. Needless to say, speakers differ in their second language (L2) proficiency; speaker intelligibility can vary considerably across L2 speakers of the same accent (e.g., Flege & Schmidt, 1995; Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1999). At the same time, exactly due to systematic influences from their first language (L1), speakers with the same L1 do share some accent regularities in their L2 speech, for instance, they may contrast vowels by duration instead of spectral quality (e.g., Flege, Bohn, & Jang, 1997; Flege & Schmidt, 1995). In other words, talker variability in foreign accents is expressed in a hierarchical structure that can benefit perception if successfully learned, such that applying the acoustic-phonetic mappings from one accented talker to a new talker with the same non-native accent should yield faster comprehension benefits than simply learning the accent of the novel talker in a talker-specific (that is, accent-agnostic) way. Given this, there are potentially strong motivations to generalize across non-native speakers of the same accent, whereas generalizing across idiolectal differences in one’s native speech has less utility.

Talker-specific adaptation

A productive line of research has demonstrated that phonetic representations can be altered to reflect the properties of the current talker. As native listeners encounter unfamiliar pronunciations that cause perceptual ambiguity, they use top-down lexical information to constrain the interpretation of the ambiguous sound and alter the sound-to-category mapping accordingly (Norris et al., 2003). For example, if listeners hear a speaker pronouncing a sound ambiguous between /s/ and /f/ (denoted here as /ʔ/), then hearing the sound in a carrier word such as ‘belie?’

(‘belief’) biases its interpretation as /f/. This exposure also affects subsequent interpretation of other similar ambiguous sounds in a way consistent with prior exposure. These findings, often referred to as ‘lexically-guided phonetic retuning’, reveal a specific mechanism by which the phonetic processing system might adjust to nonstandard *talker-specific* pronunciation variants by interfacing with the mental lexicon (e.g., Kraljic & Samuel, 2005; McQueen, Cutler, & Norris, 2006; Dahan et al., 2008).

How can phonetic representations be updated to reflect the properties of a foreign-accented talker? Using a similar paradigm to Norris et al. (2003), Xie et al. (2017) investigated how native-English listeners’ adapt to Mandarin-accented English. Word-final voiced stop consonants (e.g., the /d/ in ‘seed’) were selected as the focus of investigation because they are perceptually confusable with voiceless tokens (e.g., ‘seed’ may sound like ‘seat’) in Mandarin-accented English and they differ from native-English tokens. In English, vowels are generally lengthened before voiced consonants, and native-English listeners rely primarily on vowel length as an informative cue to voicing contrasts (e.g., Flege et al., 1992). In contrast, Mandarin-accented /d/ tokens sound /t/-like to native-English listeners because vowels are shorter before /d/ in Mandarin-accented English than in native-accented speech, and vowel length tends not to be a useful cue to the identity of the following consonant (/d/ and /t/) in Mandarin-accented English. Consequently, native-English listeners, who tend to rely primarily on vowel length, often find Mandarin-accented /d/s perceptually ambiguous (Xie & Fowler, 2013). However, it is important to note that Mandarin-accented /d/ and /t/ tokens are in fact acoustically distinguishable if listeners attend to a different cue, namely the length of burst release (i.e., word-final /d/ tokens usually have noticeably shorter bursts than /t/ tokens). For this reason, Mandarin-accented /d/ and /t/ tokens in word-final position are easy to tell apart by Mandarin listeners, but not by English listeners.

Results of Xie et al. (2017) provided support for adaptation-elicited changes in lexical access. In this study, a cross-modal priming task probed changes in online processing of the accent. Following adaptation to a Mandarin-accented speaker, listeners showed more efficient processing of accented ‘seed’ (sounding like ‘seat’ to native-English listeners) and more easily disambiguated ‘seed’ from the phonetically similar ‘seat’. Therefore, a brief exposure to a foreign-accented speaker (see also Eisner, Melinger, & Weber, 2013) created similar effects as those induced by exposure to an idiosyncratic speaker (McQueen et al., 2006) or long-term familiarity with a regional dialect (Sumner & Samuel, 2009). Taken together, this body of work suggests that listeners are capable of dynamically adjusting phonetic representations in adapting to *specific* talkers, non-native and native talkers alike. For an adapted talker, perceptual benefits manifest in both fewer offline confusions and more efficient online lexical disambiguation.

Generalization across talkers

Presumably, the newly formed phonetic representations, which differ from those used in perceiving typical native speech, could potentially render listeners an advantage when applied in appropriate contexts. In reality, whether listeners apply learning to new talkers is affected by a number of factors. First, generalization is sensitive to phonetic classes, as probed by phonetic categorization tasks (Norris et al., 2003; Kraljic & Samuel, 2005; Kraljic & Samuel, 2006; Kraljic & Samuel, 2007). Namely, listeners do not generalize across talkers for fricatives (e.g., /s/ vs. /f/) (e.g., Kraljic & Samuel, 2005; Kraljic & Samuel, 2007; Eisner & McQueen, 2005), but do generalize across talkers for stop categories (e.g., /d/ and /t/) (Kraljic & Samuel, 2006, see also Kraljic & Samuel, 2007). Second, generalization seems to occur between some talker pairs but not others. Reinisch and Holt (2014) examined native-

Download English Version:

<https://daneshyari.com/en/article/5042574>

Download Persian Version:

<https://daneshyari.com/article/5042574>

[Daneshyari.com](https://daneshyari.com)