Journal of Memory and Language 89 (2016) 23–36

Contents lists available at ScienceDirect

Journal of Memory and Language

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

Listeners beware: Speech production may be bad for learning speech sounds

Melissa M. Baese-Berk [a,b,*], Arthur G. Samuel [b,c,d]

[a] Department of Linguistics, 1290 University of Oregon, Eugene, OR 97403, United States
[b] Basque Center on Cognition Brain and Language, Paseo Mikeletegi 69, 2nd Floor, Donostia-San Sebastian 20009, Spain
[c] IKERBASQUE, Basque Foundation for Science, Bilbao 48011, Spain
[d] Department of Psychology, Stony Brook University, Stony Brook, NY 11794-2500, United States

ARTICLE INFO

Article history:
Received 4 November 2014
revision received 15 October 2015
Available online 14 November 2015

Keywords:
Language production
Language comprehension

ABSTRACT

Spoken language requires individuals to both perceive and produce speech. Because both processes access lexical and sublexical representations, it is commonly assumed that perception and production involve cooperative processes. However, few studies have directly examined the nature of the relationship between the two modalities, particularly how producing speech influences speech perception. In a series of experiments, we examine the counter-intuitive finding that learning perceptual representations can be disrupted by producing tokens during training. We investigate whether this disruption can be alleviated by prior experience with the speech sounds, and whether the cause of the disruption is production of the particular sound being learned, or is a more general conflict between the production system and the system that develops new perceptual representations. Our results paint a more competitive relationship between perception and production than might be assumed and suggest that both demands inherent to production and cognitive demands modulate this relationship.

Introduction

Spoken language is a communication system that involves the interaction of production and perception: Each person produces a series of words, and perceives those produced by another person. Because the intention is to transfer a message from one person to another, and the medium is a series of words that are each made up of a series of sublexical sounds, it is natural to assume that perception and production are cooperative processes using common elements. This assumption may be natural, but it may also be wrong. In the current study, we examine whether speech perception and speech production are in fact cooperative processes when a listener is learning a new phonemic contrast.

The existing literature indicates that the relationship between speech perception and speech production is more complex than one might assume, particularly during learning. Studies have demonstrated that although perception and production are usually correlated during novel speech sound learning, individual performance differs greatly (Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Wang, Jongman, & Sereno, 2003). In learning novel speech sounds, perception frequently improves without production learning, and vice versa (Bradlow et al., 1997; de Jong, Hao, & Park, 2009; Flege, 1993; Sheldon & Strange, 1982). In the present paper, we report the results of a series of studies examining how producing speech sounds while learning those sounds can influence how well the sounds are learned. A rather counter-intuitive finding has

* Corresponding author at: Department of Linguistics, 1290 University of Oregon, Eugene, OR 97403, United States.
E-mail address: mbaesebe@uoregon.edu (M.M. Baese-Berk).

http://dx.doi.org/10.1016/j.jml.2015.10.008
0749-596X/

emerged: When training on perception alone, the improvement in perception of the sounds is rather robust. However, when training includes both a perception and a production component, the perceptual improvement is disrupted.

Laboratory training of speech sounds

One of the classic hallmarks of speech perception in a listener's native language is categorical perception. In many early studies of speech perception (see Libermann, Cooper, Shankweiler, & Studdert-Kennedy, 1967 for examples) researchers created sets of syllables in which an acoustic parameter was varied in such a way that the syllable at one end of the continuum was heard in one way (e.g., /ba/), and the syllable at the other end in a different way (e.g., /pa/). For many speech continua, perception seemed categorical: Listeners heard a few items as one category (e.g., /ba/) and then things abruptly changed, with the remaining items heard as the other category (e.g., /pa/). Moreover, the ability to discriminate different tokens on the continuum depended on where the tokens were: Two tokens from within the same category were discriminated at near-chance levels, whereas tokens that crossed the category boundary were well discriminated. The categorical tendency in perception was strongest for stop consonants, somewhat weaker for other consonants (e.g., fricatives), and weaker still for vowels (see Repp, 1984 for an excellent review of the categorical perception literature).[1]

When listening to speech sounds from a non-native language, however, the picture can be quite different. It is not uncommon for listeners to be unable to correctly divide tokens from the continuum into two categories. These listeners are typically unable to discriminate among tokens at any point on the continuum, even if these tokens do cross a category boundary in the non-native language. The best known example of such non-native perception comes from Japanese learners of English, who have a notoriously difficult time categorizing and discriminating between tokens on an /r/–/l/ continuum (e.g., Goto, 1971; Strange & Dittman, 1984), tasks which are not difficult for English listeners. It is hypothesized that this is because Japanese does not have two distinct categories for /r/ and /l/. Flege (1995) and Best (1995) have argued that the perception of non-native phonemes is reliant on the category structure of the listeners' native language. That is, when learning their native language, a listener not only learns what variability is important, but also what variability is not. Therefore, listeners must not only learn the categories of their language, but also learn to not use irrelevant information in categorizing speech stimuli. The task of the non-native learner, then, is to learn to attend to variability which is not important for categorization in their native language.

[1] Recent work on categorical perception (e.g., Gerrits & Schouten, 2004; Schouten, Gerrits, & van Hessen, 2003) has suggested that it is more complex than some of the classic work portrayed it. Specifically, categorical perception can be triggered by certain tasks. In the current study, we are examining (and reporting) "classic" categorical perception data using tasks that have been shown to trigger such perception.

Many previous studies have demonstrated that even within a listener's native language category boundaries can be shifted with experience (Kraljic & Samuel, 2005, 2006, 2007; Maye, Aslin, & Tanenhaus, 2008; Norris, McQueen, & Cutler, 2003), suggesting substantial flexibility in the perceptual system and its representations. Within a non-native language, dozens of previous studies have demonstrated that individuals are capable of learning novel phonological categories in the laboratory (Strange & Dittman, 1984). A number of language backgrounds and many different segmental and suprasegmental contrasts have been examined including English listeners' perception of a three-way voicing continuum (McClaskey, Pisoni, & Carrell, 1983; Tremblay, Kraus, & McGee, 1998), Spanish and German listeners' perception of English vowels (Iverson & Evans, 2009), English listeners' perception of German vowels (Kingston, 2003), French listeners' perception of English interdental fricatives (Jamieson & Morosan, 1986, 1989), Japanese listeners' perception of English /r/ and /l/ (Logan, Lively, & Pisoni, 1991), and English listeners' perception of Mandarin tone (Wang, Spence, Jongman, & Sereno, 1999). In many cases, listeners are able to learn to perceive contrasts that do not exist in their native language with a relatively small amount of laboratory-based training. Taken with the results from perceptual learning within a native language, this suggests that listeners' perception is relatively flexible and can be changed through experience.

The relationship between perception and production during learning

Most models of speech perception and production suggest a close tie between the two modalities. The two modalities are frequently discussed as being two ends of a single process, as in Denes and Pinson's speech chain (Denes & Pinson, 1993). Researchers have frequently cited perception-oriented changes in production such as the Lombard effect (Lane & Tranel, 1971; Lombard, 1911) as evidence for a necessarily tight connection between the two modalities. Other such effects include shadowing of various phonetic properties such as voice onset time (Goldinger, 1998), shifts in vowel production as a result of shifted perceptual input (Houde & Jordan, 1998, 2002), and phonetic accommodation to a conversation partner's speech (Kim, Horton, & Bradlow, 2011; Pardo, 2006). These findings have been interpreted as demonstrating that perception and production must be tightly coupled, or as Casserly and Pisoni (2010) state "...the two processes must at some point even deal in the same linguistic currency." These results all conform to the basic fact, stated at the beginning of this article, that production and perception must be compatible enough to allow communication to occur: Speakers produce in order that perceivers will understand a message.

Most prior work showing this naturally cooperative relationship has looked at perception and production within a relatively steady-state native language system. How the two modalities relate during learning of non-native contrasts is less clear. Most of the previous studies examining learning of novel phonemes have focused on