Available online at www.sciencedirect.com

ScienceDirect

Speech Communication 77 (2016) 84–100

www.elsevier.com/locate/specom

Effects of native language on compensation for coarticulation☆

Shinae Kang[1], Keith Johnson*, Gregory Finley[2]

Department of Linguistics, 1203 Dwinelle Hall, University of California, Berkeley CA 94720-2650, United States

Received 20 May 2015; received in revised form 4 December 2015; accepted 18 December 2015
Available online 29 December 2015

Abstract

This paper investigates whether compensation for coarticulation in speech perception can be mediated by native language. Substantial work has studied compensation as a consequence of aspects of general auditory processing or as a consequence of a perceptual gestural recovery processes. The role of linguistic experience in compensation for coarticulation potentially cross-cuts this controversy and may shed light on the phonetic basis of compensation. In Experiment 1, French and English native listeners identified an initial sound from a set of fricative-vowel syllables on a continuum from [s] to [ʃ] with the vowels [a,u,y]. French speakers are familiar with the round vowel [y], while it is unfamiliar to English speakers. Both groups showed compensation (a shifted ‘s’/‘sh’ boundary compared with [a]) for the vowel [u], but only the French-speaking listeners reliably compensated for the vowel [y]. In Experiment 2, 39 American English listeners judged videos in which the audio stimuli of Experiment 1 were used as soundtracks of a face saying [s]V, [ʃ]V, or a visual-blend of the two fricatives. The study found that videos with [ʃ] visual information induced significantly more “ʃ” responses than did those made from visual [s] tokens. However, as in Experiment 1, English-speaking listeners reliably compensated for [u], but not for the unfamiliar vowel [y]. The listeners used visual consonant information for categorization, but did not use visual vowel information for compensation for coarticulation. The results indicate that perceptual compensation for coarticulation is a language specific effect tied to the listener’s experience with the conditioning phonetic environment.

Keywords: Speech perception; Compensation for coarticulation; Linguistic experience; Direct realism; Audiovisual perception.

1. Introduction

1.1. Three modes of speech perception

General properties of the auditory system determine what can and cannot be heard, what speech cues will be recoverable in particular segmental contexts, and to at least some extent how adjacent sounds will influence each other. For example, the cochlea’s nonlinear frequency scale probably underlies the fact that no language distinguishes fricatives on the basis of frequency components above 6000 Hz (Johnson, 2012). Similarly, limitations on the auditory system’s ability to detect the simultaneous onset of tones at different frequencies probably underlies the fact that the most common VOT boundary across languages is at about ± 30 ms (Pastore and Farrington, 1996).

In addition to these general auditory factors, speech perception may also be shaped by phonetic knowledge. Because language users are both speakers and listeners, we come to the task of speech perception with a base of knowledge that makes available a “phonetic mode” of listening (or “speech mode”; Liberman and Mattingly, 1985). [Strictly speaking, Liberman and Mattingly, 1985 “speech mode” is not completely synonymous with our concept of the “phonetic mode” because we use the term “phonetic mode” in a more general sense to contrast knowledge-based phonetic processing with general auditory processing.] By hypothesis, the phonetic mode elaborates and reinterprets the auditory image of speech. Thus, the phonetic mode may underlie the tendency for multimodal information to be combined into a phonetic

☆ The original version of this paper was selected as one of the best papers from Interspeech 2010. It is presented here in revised form following additional peer review.

* Corresponding author. Tel.: +1 510 642 2757; fax: +510 643 5688.
E-mail addresses: sakang2@berkeley.edu (S. Kang), keithjohnson@berkeley.edu (K. Johnson), finley@berkeley.edu (G. Finley).
[1] Present address: Georgetown University, Washington, DC.
[2] Present address: University of Minnesota, Minneapolis, MN.

http://dx.doi.org/10.1016/j.specom.2015.12.005
0167-6393/

percept (McGurk and Macdonald, 1976), and may explain how the perception of sine wave analogs of speech may suddenly shift from nonphonetic to phonetic (e.g. Remez et al., 1981). Additionally, the phonetic mode of speech perception is probably also involved in the perceptual coherence of signal components that might not ordinarily be grouped with each other in the phenomenon of duplex perception (Bregman, 1990; Whalen and Liberman, 1987) or in the integration of asynchronous audio signals (Nygaard and Eimas, 1990).

Scholars differ in their view of whether the phonetic mode of perception is innate or learned. For example, according to Liberman and Mattingly (1985), the speech mode of listening is innate and does not require experience as a speaker. On the other hand, Best (1995) assumed that the phonetic mode is grounded in experience-based perceptual learning and this underlies the strong tendency to hear foreign speech in terms of native segments. Fowler (1986, 1996) places much less emphasis on learning and in this respect is similar to Liberman and Mattingly's view.

Speech perception is also shaped by lexical knowledge. The fact that the listener's ultimate aim in speech communication is to figure out what words the speaker is saying underlies lexical effects in speech perception. For example, perceptual errors ("slips of the ear") overwhelmingly result in words (Bond, 2005). Similarly, Ganong (1980) showed a lexical effect on phoneme identification. In a "tash-dash" VOT continuum there are more "d"-responses, consistent with the word "dash", than in a "task-dask" continuum. Similarly, a missing or obliterated phoneme can be perceptually restored (Pitt and Samuel, 1995), and the restored phones interact with phonetic mode processes like compensation for coarticulation (Elman and McClelland, 1988; but see McQueen et al., 2009).

Researchers who primarily focus on one or the other of these three aspects of speech perception (auditory, phonetic, or lexical) are often critical of the others (e.g. Fowler, 2006 against the exclusive effects of auditory spectral contrast on compensation for coarticulation; McQueen (2006) against direct lexical involvement in speech perception; and Diehl and Walsh, 1989; Lotto and Kluender, 1998 against a specifically phonetic mode of processing). Our view is that it is more plausible to assume that all three factors are simultaneously involved in speech perception. Indeed, recent findings from neuroscience (cf. Hickok and Poeppel, 2004) indicate that all three are simultaneously involved in speech perception. Ultimately, a successful theory of speech perception has to predict which listening circumstances will engage greater or lesser reliance on phonetic processing, or lexical processing, and what aspects of speech perception ultimately derive more from auditory processing than from specifically linguistic processing.

1.2. Compensation for coarticulation

In this paper, we explore how the phonetic mode of listening may be shaped by linguistic experience in a compensation for coarticulation task. Our experiments on compensation do not test for auditory contrast or lexical activation effects, but we are aware of the literature in these areas. For example, in the literature on whether a lexically biased percept can induce compensation for coarticulation (Elman and McClelland, 1988; Pitt and McQueen, 1998), compensation is assumed to exist as a separate, phonetic mode, phenomenon that can be used as a diagnostic to determine whether the restored phoneme is truly restored. We do not go further in lexically induced compensation for it is beyond the scope of this study.

Compensation for coarticulation (Mann, 1980; Mann and Repp, 1981) is a listener's perceptual "demodulation" of coarticulatory information during speech perception. For example, Mann and Repp (1981) found that the lower fricative pole induced by adjacent vowel lip rounding in [s] did not induce the percept of a more alveopalatal fricative [ʃ], while the same fricative noise paired with the unrounded vowel [a] does sound more like [ʃ]. This phenomenon of attributing one aspect of the acoustic signal (lower pole frequency) to coarticulation with a neighboring vowel, and thus not only an inherent property of the fricative itself, is a prototypical case of compensation for coarticulation. Compensation has been investigated in many studies of consonant–vowel interactions in consonant place perception (e.g. Mann and Repp, 1981; Mitterer, 2006; Smits, 2001; Whalen, 1981), vowel perception (Holt et al., 2000), and consonant voicing perception (Diehl and Walsh, 1989), as well as in vowel–vowel interactions (Fowler and Smith, 1986; Bradlow and Bent, 2002), and in consonant–consonant interactions (Fowler, 2006; Lotto and Kluender, 1998; Mann and Repp, 1981; Pitt and McQueen, 1998).

Much of this literature is steeped in controversy regarding the basis of the compensation mechanism—whether it is due to the auditory interaction between adjacent segments, or due to a phonetic mode of processing "undoing" the gestural interactions inherent in speaking and thus an indication that speech is perceived in terms of phonetic gestures. While some researchers have suggested that auditory spectral contrast plays a primary role in the phenomenon of compensation for coarticulation (Johnson, 2011; Lotto and Kluender, 1998), several studies have provided evidence showing that spectral contrast alone cannot capture the whole phenomenon (e.g. Fowler, 2006).

For example, Mitterer (2006) found an effect of visible lip rounding by Dutch listeners and concluded that compensation for coarticulation has a phonological basis. He studied perception of a [si]-[sy] fricative continuum, first testing whether compensation for vowel rounding can be replicated with non-speech audio that imitates critical acoustic characteristics (spectra contrast etc.), and second testing whether compensation for vowel rounding (in natural speech tokens) increased when the participants saw audio/visual stimuli with lip rounding during the vowel. The participants showed no compensation effect for the non-speech audio, and an increased effect for AV stimuli. Based on these results, he concluded that the basis of compensation for coarticulation is not solely auditory.