Journal of Memory and Language 77 (2014) 1–24

Contents lists available at ScienceDirect

Journal of Memory and Language

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

Effects of acoustic–phonetic detail on cross-language speech production

Colin Wilson [a,*], Lisa Davidson [b], Sean Martin [b]

[a] Johns Hopkins University, 3400 N. Charles Street, Department of Cognitive Science, Baltimore, MD 21218, United States
[b] Department of Linguistics, 10 Washington Place, New York, NY 10003, United States

ARTICLE INFO

Article history:
Received 4 October 2013
revision received 8 August 2014
Available online 6 September 2014

Keywords:
Phonotactics
Cross-language speech production
Phonetic detail
Phonetic decoding

ABSTRACT

Nonnative sounds and sequences are systematically adapted in both perception and production. For example, American English speakers often modify illegal word-initial clusters by inserting a vocalic transition between the two consonants (e.g., (/bdagu/ → [bᵊdagu]). Previous work on such modifications has for the most part focused on relatively abstract properties of the nonnative structures, such as their phonemic content and whether they conform to sonority sequencing principles. The current study finds that fine-grained phonetic details of the stimulus can be equally important for predicting cross-language production patterns. Several acoustic–phonetic properties were manipulated to create stimulus variants that are phonemically identical (i.e., exhibit non-contrastive variation) in the target language (Russian). In a shadowing experiment, English speakers' correct productions and detailed error patterns were significantly modulated by the acoustic manipulations. The results highlight the role of perception in accounting for cross-language production, and establish limits on the perceptual repair of nonnative sound sequences by phonetic decoding.

Introduction

Research on cross-language speech perception and production has shown that nonnative sound patterns can be misperceived and modified in systematic ways. Perhaps best known are cases in which listeners fail to reliably distinguish individual sounds that do not contrast in their native language. For example, Japanese listeners have difficulty discriminating English word-initial /l/ and /ɹ/ (e.g., Guion, Flege, Akahane-Yamada, & Pruitt, 2000), and English listeners cannot reliably categorize the Hindi dental and retroflex stops (Pruitt, Jenkins, & Strange, 2006; Werker & Tees, 1984). There has also been considerable research on the perception of sounds in particular positions and combinations that do not occur natively, especially consonant clusters and word-final consonants. Dupoux, Kakehi, Hirose, Pallier, and Mehler (1999) provide evidence that Japanese listeners often 'perceptually epenthesize' a vowel between word-medial French consonants (e.g., /ebzo/ → [ebuzo]). Related cases of perceptual epenthesis and other types of perceptual 'repair' have been reported for a wide range of nonnative clusters (Berent, Steriade, Lennertz, & Vaknin, 2007; de Jong & Park, 2012; Hallé, Dominguez, Cuetos, & Segui, 2008; Kabak & Idsardi, 2007).

Psycholinguistic theories of cross-language speech perception, like related second language (L2) models (e.g., Best, 1995; Flege, 1995), have focused on the role of *phonetic decoding*. While details vary across accounts, the following description by Peperkamp and Dupoux (2003) is representative of how phonetic decoding is thought to apply to nonnative inputs:

* Corresponding author. Fax: +1 410 516 8020.
E-mail addresses: colin@cogsci.jhu.edu (C. Wilson), lisa.davidson@nyu.edu (L. Davidson).

http://dx.doi.org/10.1016/j.jml.2014.08.001
0749-596X/

"During phonetic decoding, a given input sound will be mapped onto the closest available phonetic category . . .. With respect to nonnative sounds, this mapping is of course massively unfaithful, since the phonetic categories to which these sounds are mapped in the foreign language can simply be absent from the native one."

Unfaithful decoding also applies to nonnative sequences, as demonstrated by perceptual epenthesis (e.g., Dupoux, Parlato, Frota, Hirose, & Peperkamp, 2011), and to suprasegmental structures such as stress (e.g., Dupoux, Pallier, Sebastián-Gallés, & Mehler, 1997). In all cases, it is plausible that unfaithful decoding maps nonnative inputs to the most phonetically similar native sound structures (e.g., Best, 1995; Escudero, Simon, & Mitterer, 2012; Flege, 1995).

While phonetic decoding has been extensively investigated with perceptual tasks, a number of basic questions about the process and its connection to other components of the language system remain open. Does phonetic decoding consistently map incoming speech signals to phonetic/phonological representations that are legal in the native language, or are illegal representations sometimes formed? If the latter, what factors determine the relative probability with which phonetic decoding 'repairs' or leaves intact a given nonnative structure? Finally, can nonnative structures that are faithfully represented by phonetic decoding be preserved by subsequent task-dependent processes?

We address these questions by investigating how the acoustic–phonetic details of nonnative inputs affect *speech production*. Specifically, we focus on how English speakers with no prior knowledge of Russian produce consonant clusters such as those at the beginning of words like /knʲigə/ 'book' and /zdarov/ 'healthy'. Adopting a method that has been widely used in perception studies, but which has only rarely been applied to production, we systematically manipulate acoustic properties—including the presence of voicing before the beginning of an obstruent, and the amplitude and duration of stop bursts—to create phonetic variants of the clusters. These properties are part of the non-contrastive system of phonetic realization in Russian speech. Our main focus here is the relation between the (manipulated) acoustics of stimulus clusters and the detailed production patterns of English speakers.

Previous production studies have found that nonnative consonant clusters are often modified by epenthesis and a wide range of other 'repairs', including consonant deletion and change of one or more distinctive features (Broselow, 1992; Broselow & Finer, 1991; Davidson, 2006a, 2010; Hancin-Bhatt & Bhatt, 1997). It is also known that English speakers can produce such clusters correctly—matching the phonetic realization of Russian speakers—a certain proportion of the time. However, the *rates* and *types* of modification and correct production vary across clusters in a way that has not been satisfactorily explained. If detailed modification patterns can be demonstrated to be sensitive to fine-grained phonetic details of the stimulus, this will simultaneously shed light on the phonetic decoding process and provide novel insights about a rich body of cross-language production data.

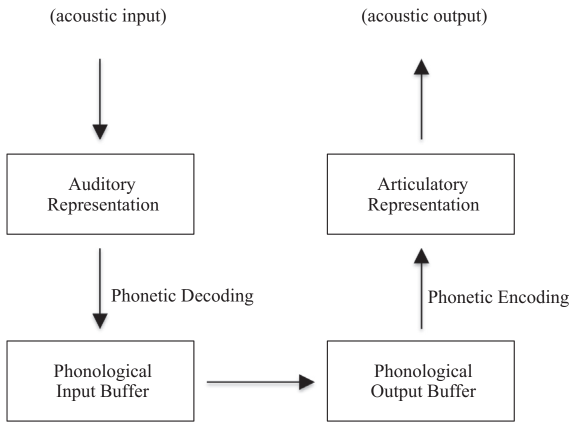

Fig. 1. Processing architecture for nonword repetition (see text for details).

Our discussion is framed within a cognitive processing architecture that has been developed for native word perception and production (Ellis & Young, 1988; Goldrick & Rapp, 2007; Patterson & Shewell, 1987; Ramus et al., 2010). In this architecture, illustrated in Fig. 1, our task (nonword repetition) is decomposed into phonetic decoding, which maps an auditory form to a representation in the phonological input buffer, and a production process, in which phonological encoding creates a representation in the phonological output buffer that is implemented by vocal tract movements. While some modifications of nonnative structures may originate in phonological encoding or articulatory execution (Davidson, 2006a et seq., see General Discussion), our experimental manipulations target phonetic decoding. We begin by considering how patterns of nonnative perception and production bear on the nature of this process, and then turn to the motivation and design of our production experiment.

Phonetic decoding in nonnative perception and production

The input to phonetic decoding is an auditory representation of the incoming acoustic signal. Evidence for language-specific shaping of the auditory system is presently mixed (Breen, Kingston, & Sanders, 2013; Dehaene-Lambertz, Dupoux, & Gout, 2000; Jacquemot, Pallier, LiBihan, Dehaene, & Dupoux, 2003), so we take auditory representations to be largely language-independent (Kingston, 2005). These representations contain measurements of acoustic–phonetic properties—such as formants, durations, and intensities—that are commonly referred to as *cues* in the speech perception literature (e.g., Lisker, 1986; Wright, 2004). Phonetic decoding interprets the cues from the stimulus as phonetic/phonological structures consisting of segments, syllables, etc. Language-specific sound structures begin to influence processing at the level of phonetic decoding, but the nature and extent of the influence there (and at later levels) is not fully understood.

As indicated by the quote from Peperkamp and Dupoux (2003) above, previous perceptual investigations of