Journal of Phonetics 54 (2016) 15–34

Contents lists available at ScienceDirect

Journal of Phonetics

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

Research Article

The role of phrasal phonology in speech perception: What perceptual epenthesis shows us

Karthik Durvasula [a,*], Jimin Kahng [b]

[a] Department of Linguistics and Languages, Michigan State University, B330 Wells Hall, 619 Red Cedar Road, East Lansing, MI 48824, USA
[b] Teaching English to Speakers of Other Languages, Northeastern Illinois University, 3069 LWH, 5500N St. Louis Avenue, Chicago, IL 60625, USA

ARTICLE INFO

Article history:
Received 6 April 2014
Received in revised form
27 July 2015
Accepted 24 August 2015

Keywords:
Speech perception
Illusory vowels
Prosodic domains
Korean
American English

ABSTRACT

Recent research using the phenomenon of illusory vowels has raised our awareness of the extent to which speech perception is modulated by the listener's native-language phonological knowledge. However, most of the focus has been limited to word-level phonological knowledge. In this article, we suggest that the perceptual system recruits segmental phonological knowledge that makes crucial reference to prosodic domains far beyond the word-level. We report the results from three identification experiments on Korean and American English participants. In accordance with their native-language phonotactic constraints at the level of the Intonational Phrase, Korean listeners unlike American English listeners hear more illusory vowels in stimuli containing the sequence of voiced stops followed by nasal consonants (e.g. [egma]) than those containing voiceless stops followed by nasal consonants (e.g. [ekma]). The results are interpreted as support for the view that speech perception makes crucial reference to the concept of reverse inference.

1. Introduction

There has been growing awareness in the last few decades that speech perception is modulated by the listener's phonological knowledge. Listeners are more attuned to phonological categories that are contrastive in their native-language (Best, 1994; Dupoux, Sebastián-Gallés, Navarete, & Peperkamp, 2008). Listeners also seem to employ their knowledge of the phonotactic constraints in their native-language (Dehaene-Lambertz, Dupoux, & Gout, 2000; Dupoux, Kakehi, Hirose, Pallier, & Mehler, 1999; Kabak & Idsardi, 2007). However, it is important to note that the evidence provided in support of phonotactic knowledge being recruited during speech perception is largely focused on *word-level* phonotactics. In this article, we suggest through the phenomenon of *illusory vowels* that, along with knowledge of word-level phonotactic constraints and phonological categories in the native language, the perceptual system recruits segmental phonological knowledge that makes crucial reference to prosodic domains far beyond the word-level.

In the past decade and a half, the phenomenon of illusory vowels has been especially useful in allowing researchers to develop an enriched understanding of the nature of the phonological knowledge that is implicated during speech perception (Berent, Lennertz, Smolensky, & Vaknin-Nusbaum, 2009; Berent, Steriade, Lennertz, & Vaknin, 2007; Dehaene-Lambertz et al., 2000; Dupoux et al., 1999; Dupoux, Parlato, Frota, Hirose, & Peperkamp, 2011; Kabak & Idsardi, 2007; Monahan, Takahashi, Nakao, & Idsardi, 2009; inter alia). The general finding with studies on illusory vowels is that when a native speaker is presented with a nonsense word that contains a consonant sequence that violates the phonotactic constraints in their language, an illusory vowel is often perceptually induced in between such sequences, thereby creating an illusory sequence that respects the phonotactic constraints of the language. For example, when a Japanese listener is auditorily presented with stimuli such as [ebzo], where the consonant sequence is either naturally produced or created by splicing out the medial vowel completely from productions such as [ebizo] or [ebazo], they may actually perceive [ebuzo] given that [bz] is an illicit consonant sequence in Japanese, as shown originally by Dupoux et al. (1999). In contrast, French speakers correctly identified far fewer (if any) illusory vowels for the relevant stimuli in the same experiment. Thereby suggesting that the illusory vowels heard by the Japanese speakers were a language specific effect and not simply due to fine

* Corresponding author. Tel.: +1 517 432 0194; fax: +1 517 432 2736.
E-mail addresses: durvasul@msu.edu (K. Durvasula), j-kahng@neiu.edu (J. Kahng).

0095-4470/$ - see front matter © 2015 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.wocn.2015.08.002

phonetic detail. The characteristic that is most crucial to such studies is that the test stimuli have sequences of sounds that are phonotactically illicit, and therefore contain novel combinations of segments (and phonological features). This allows one to probe the types of generalizations that are crucially implicated in speech perception.

Previous research on the perception of illegal phonotactic sequences has led to a variety of new proposals and insights related to speech perception: (1) Listeners make use of abstract featural co-occurrence constraints (Moreton, 2002). (2) Abstract principles such as the Sonority Sequencing Principle modulate speech perception (Berent, Lennertz, Jun, Moreno, & Smolensky, 2008; Berent et al., 2007, 2009). (3) Syllable-structure phonotactics play a more important role than linear phonotactics[1] (Kabak & Idsardi, 2007). (4) More generally, phonotactically illegal consonant sequences are perceptually "repaired" in a variety of ways other than through illusory vowels (Davidson, 2006, 2007; Davidson & Shaw, 2012; Hallé, Segui, Frauenfelder, & Meunier, 1998; Wilson & Davidson, 2013).

It is important to note that a lot of the above work probing the perception of illegal phonotactics is dependent on the phenomenon of perceptual epenthesis. First, Berent et al. (2008, et seq) show that both English and Korean listeners perceive more illusory vowel in onsets with decreasing sonority (e.g., [md] and [nb]) than in onsets with rising sonority (e.g., [ml] and [nw]). Second, Kabak and Idsardi (2007) argue that for Korean listeners, there are more illusory vowels perceived in stimuli that violate syllable-structure phonotactic restrictions (e.g., [ecma], where [c] is not possible in coda positions in Korean), than those that violate just linear phonotactic restrictions (e.g., [elma], where the sequence [lm] is not a possible sequence of sounds in Korean within a word, though [l] is an acceptable coda, and [m] is an acceptable onset). Third, the phenomenon of illusory vowels is also modulated by the low-level acoustic differences of the stimuli containing illegal phonotactic sequences (Davidson, 2006; Davidson & Shaw, 2012; Dupoux et al., 2011). For example, Dupoux et al. (2011) show that illegal phonotactic sequences created by splicing out a medial vowel from stimuli trigger more perceptual illusions for the spliced out vowel than those that are naturally produced with the same phonotactic violation. Finally, there have also been reports of illusory vowels being modulated by the place of articulation of the preceding consonant (De Jong & Park, 2012). This last set of results might perhaps be attributed both to language specific (stochastic) phonotactic patterns or to language-specific phonetic patterns such as differences in stop burst releases for different places of articulation[2] (Byrd, 1992; Crystal & House, 1988; Kang, 2003; Rositzke, 1943).

While there is extensive work on the perception of stimuli containing illegal phonotactics and how such violations along with other factors influence perceptual illusions, the bulk of the research has focused on the perception of nonsense words containing word-level phonotactic violations (Berent et al., 2007, 2009; Davidson & Shaw, 2012; Davidson, 2007; Dehaene-Lambertz et al., 2000; Dupoux et al., 1999, 2011; Kabak & Idsardi, 2007; Monahan et al., 2009; inter alia).

In contrast, there is little research on whether phonological generalizations beyond the word-level are utilized in speech perception. In a study that probed the use of language specific prosodic knowledge by listeners, French participants were presented with identical segmental sequences of /selafiʃ/ *C'est la fiche/l'affiche* 'It's the sheet/poster' that differed only in the *f*0 of the vowel /a/ (Spinelli, Grimault, Meunier, & Welby, 2010). In a two-way forced choice task, increasing the *f*0 of the vowel /a/ increased the identification rates for the word *affiche* 'poster' in line with the intonational patterns in French across the *Accentual Phrase*, a phonological domain beyond the word level (Jun & Fougeron, 2002). While the results were consistent with listeners using language specific prosodic knowledge, a potential confound stems from the fact that the study lacked a control linguistic group that did not pattern like the French participants. One might argue that this control linguistic group is important to establish that language-general auditory processes to do with *f*0 processing cannot explain the results.

Additionally, in very recent work probing the effect of prosodic boundaries on phonetic categorization, Kim and Cho (2013) show that native Korean and American English listeners needed a longer VOT to identify a sound as /p/ after an *Intonational Phrase* boundary than after a word boundary. They interpreted this as evidence of the use of general prosodic knowledge by both native Korean and American English listeners. However, as the authors acknowledge, it is not clear that the participants used language-specific prosodic knowledge during the task, given both language groups showed the same pattern. Furthermore, an alternative explanation for their results is the possibility that the VOT boundary of the listeners was modulated by the durational differences in the pre-stimulus carrier phrase. For example, for the English listeners, the pre-stimulus carrier phrase "*Let's hear*" was 355 ms in the Intonational Phrase condition, and 147 ms in the word boundary condition. It is possible that this durational difference in the two conditions was interpreted by the participants as a speech rate difference, thereby resulting in a shift of the listeners' VOT boundary (Nagao & de Jong, 2007). More generally, given that listeners are able to use distal and proximal speech rate cues available in the acoustic input during speech perception (Dilley & McAuley, 2008; Heffner, Dilley, McAuley, & Pitt, 2013), to show that listeners are necessarily using their knowledge of prosodic patterns in their language during speech perception, it is important to ensure that such durational/speech rate cues be absent.

There is of course related work on speech perception that argues that listeners are able to compensate for assimilatory changes at the edges of words (Coenen, Zwitserlood, & Bölte, 2001; Gaskell & Marslen-Wilson, 1996; Gow, 2003; Mitterer, Kim, & Cho, 2013). For example, the phrase "garden bench" /gɑːdn bɛntʃ/ is often pronounced as [gɑːdm bɛntʃ],[3] where the word-final nasal /n/ has assimilated to the place of articulation of the following segment. It has been shown that listeners are able to compensate for such

[1] Following Kabak and Idsardi (2007), 'linear phonotatics' refers to phonotactic knowledge about word-internal segmental sequences, without reference to any prosodic information. For example, in the nonsense word [elna], the sequence [ln] violates linear phonotactic constraints in Korean, since the sequence of segments is not possible within words. However, the sequence [ln] does not violate syllable-structure phonotactics because [l] is a perfectly acceptable coda segment and [n] is a perfectly acceptable onset segment in Korean, i.e., since the syllable [el] and [na] are both acceptable syllables in Korean, the nonsense word [elna] does not violate any syllable-structure phonotactics.

[2] This latter connection between the modulation of perceptual epenthesis by the place of articulation of the preceding consonant was first explored to our knowledge in the context of accounting for vowel epenthesis in loanword patterns by Kang (2003).

[3] We use /…/ to represent phonemic or "underlying" representations, and […] to represent pronunciations or "surface" representations.