Journal of Phonetics 56 (2016) 110–123

Contents lists available at ScienceDirect

Journal of Phonetics

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

Research Article

What are the letters of speech? Testing the role of phonological specification and phonetic similarity in perceptual learning

Holger Mitterer [a], Taehong Cho [b,*], Sahyang Kim [c]

[a] Department of Cognitive Science, University of Malta, Msida, Malta
[b] Hanyang Phonetics and Psycholinguistics Lab, Department of English Language and Literature, Hanyang University, Seoul, Republic of Korea
[c] Department of English Education, Hongik University, Seoul, Republic of Korea

ARTICLE INFO

Article history:
Received 14 December 2015
Received in revised form
23 February 2016
Accepted 11 March 2016
Available online 2 April 2016

Keywords:
Speech perception
Perceptual learning
Prelexical processing
Phonetic similarity
Korean
Underlying representation

ABSTRACT

Recent studies on perceptual learning have indicated that listeners use some form of pre-lexical abstraction (an intermediate unit) between the acoustic input and lexical representations of words. Patterns of generalization of learning that can be observed with the perceptual learning paradigm have also been effectively examined for exploring the nature of these intermediate pre-lexical units. We here test whether perceptual learning generalizes to other sounds that share an underlying or a phonetic representation with the sounds based on which learning has taken place. This was achieved by exposing listeners to phonologically altered (tensified) plain (lax) stops in Korean (i.e., underlyingly plain stops are produced as tense due to a phonological process in Korean) with which listeners learned to recalibrate place of articulation in tensified plain stops. After the recalibration with tensified plain stops, Korean listeners generalized perceptual learning (1) to phonetically similar but underlyingly (phonemically) different stops (i.e., from tensified plain stops to underlyingly tense stops) and (2) to phonetically dissimilar but underlyingly (phonemically) same stops (i.e., from tensified plain stops to non-tensified ones) while generalization failed to phonetically dissimilar *and* underlyingly different consonants (aspirated stops and nasals) even though they share the same [place] feature. The results imply that pre-lexical units can be better understood in terms of phonetically-definable segments of granular size rather than phonological features, although perceptual learning appears to make some reference to the underlying (phonemic) representation of speech sounds based on which learning takes place.

1. Introduction

One of the most salient differences in the research on visual- and spoken-word recognition may be found in the issue regarding pre-lexical units—i.e., whether listeners use some form of abstraction in pre-lexical processing, and, if they do, what the nature of the units would be. In visual-word recognition, there is no controversy that letters are important units that help listeners to recognize words. Models of visual-word recognition usually involve some kind of letter units (Rastle, 2007), so that the difference between *lice* and *dice* at the lexical level, for example, is not the presence versus absence of a semicircle at the left edge with 'l', but the difference between the letter 'l' and the letter 'd' in the first position. However, it is precisely this kind of question that is controversial in the field of spoken-word recognition.

On the one end, it has been argued that there are no units at all in pre-lexical processing, as listeners store multiple "grainy spectrograms" for a word (Pierrehumbert, 2002) and the incoming input is compared to the multiple versions or the "episodes" for each word (Goldinger, 1998). This view provided a useful challenge for the existing models of spoken-word recognition, which tended to assume some kind of an intermediate unit between the acoustic input and the stored abstract mental representation in the mental lexicon (McClelland & Elman, 1986; Norris, 1994). As a consequence, subsequent studies explored this issue and provided some evidence that listeners indeed make active use of pre-lexical units in generalization of perceptual learning. For example, in a phonetic categorization study, confronted with a word that ends on /s/ but is produced with an ambiguous phonetic form between /s/ and /f/ (e.g., maʊ[s/f]), listeners not only recognize the ambiguous sound as /s/ (due to the lexical bias, since *mouse* is a word but *mouf* is

* Correspondence to: Hanyang Phonetics and Psycholinguistics Lab, Department of English Language and Literature, Hanyang University, 222 Wangsimni-ro, Seongdong-gu Seoul (133-791), Korea. Tel.: +82 2 2220 0746; fax: +82 2 2220 0741.
E-mail addresses: holger.mitterer@um.edu.mt (H. Mitterer), tcho@hanyang.ac.kr (T. Cho), sahyang@hongik.ac.kr (S. Kim).

http://dx.doi.org/10.1016/j.wocn.2016.03.001

not) but also learn, through multiple exposures, the speaker's idiosyncratic way of producing /s/. They are then able to generalize this learning to any other word containing /s/, showing that they have learned something about fine-grained pre-lexical information across words. This learning paradigm has hence led to the general acceptance of the assumption that some form of pre-lexical abstraction occurs between the acoustic output and the lexical representation in speech perception (Goldinger, 2007).

Part of the appeal of an episodic model may lie in the fact that it does not need to deal with the problems in delineating the fine-grained size of these units. At one point in the past, the selective-adaptation paradigm was thought to reveal pre-lexical units (Samuel, 1982), but it turned out that selective adaptation did not exclusively target pre-lexical representations and any perceptual decision on different levels could be influenced by selective adaptation (Remez, 1987). Similar problems arose with other paradigms, like the one using sub-categorical mismatches (Marslen-Wilson & Warren, 1994; McQueen, Norris, & Cutler, 1999). Proponents of episodic models also pointed out that one could find evidence for any kind of unit in a selective-adaptation paradigm (Goldinger & Azuma, 2003). In summary, there never was any consensus on how the evidence from different types of procedures should be weighted with regard to the question of the grain size of pre-lexical units.

However, if the perceptual-learning paradigm is able to show the existence of some kind of pre-lexical units, it may also be useful to indicate what form these units have. This is especially so, because this paradigm reveals a unit that listeners are actively using in adapting to variability in speech, and as such, has some ecological validity. In an effort to explore the form of pre-lexical units with this paradigm, Mitterer, Scharenborg, and McQueen (2013) tested whether learning generalizes from one allophone of Dutch /r/ to another, with the reasoning that, if pre-lexical units are phonemic, learning should generalize over phonetic variants of a given phoneme, that is, over allophones. The results, however, showed no generalization from an approximant variant to a trilled variant of /r/, suggesting that learning does not generalize on a phonemic level. Based on this, Mitterer et al. argued that the pre-lexical units are not phonemic but are likely to be sub-phonemic in nature.

Generalization in the perceptual-learning paradigm was found, however, for the voicing contrast across place of articulation (Kraljic & Samuel, 2006). Participants, who learned that an ambiguous voice onset time (in terms of voiced versus voiceless) was indicative of a voiced /d/ rather than a voiceless /t/, generalized the learning to the voicing contrast in another place of articulation (e.g., between /g/ and /k/). This finding would indicate that the pre-lexical processing makes use of features (in this case, the [voice] feature) rather than phonemes. It is then useful to consider how a featural account might deal with the failure of generalization over allophones of /r/, because this may depend on what kind of features one would assume. If one assumes that the features refer to articulatory gestures, as assumed by Articulatory Phonology (Goldstein & Fowler, 2003), the failure of generalization is predicted, because completely different articulatory gestures are involved for a trilled variation and an approximant variation of /r/. If one, however, assumes the role of abstract phonological features (Lahiri & Reetz, 2010), and perceptual learning takes place on the [+rhotic] feature, generalization would be expected because both allophones share the feature.[1]

In fact, it appears that the two disciplines, psychology and linguistics that deal with spoken-word recognition have quite different ideas about the nature of pre-lexical units. Embick and Poeppel (2014) note that psychologists tend to think in terms of segments of some sort while linguistic accounts tend to argue for the role of phonological features. A recent study by Reinisch, Wozny, Mitterer, and Holt (2014), however, questioned the role of features. Reinisch et al. (2014) used a visually-guided learning paradigm (rather than a more commonly employed lexically-guided learning paradigm), in which an ambiguous sound is accompanied by an unambiguous visual signal. Whereas an ambiguous sound (e.g., a sound between /s/ and /f/ after *gira*…) may be disambiguated by lexical knowledge (because *giraffe* is a word and *girasse* is not), the paradigm makes use of a kind of the McGurk effect, so that an ambiguous syllable between /ba/ and /da/ is perceived as /ba/ if it is accompanied by a visual lip-closing gesture, the visual cue to /b/. If the same syllable is subsequently heard in an audio-only condition, it is more likely to be labeled as /ba/ than when the same syllable was previously accompanied by a visual gesture for /da/ without a lip-closing gesture (Bertelson, Vroomen, & de Gelder, 2003). Reinisch et al. (2014) tested how such learning or recalibration of phonetic categories may generalize, and found that it is context-specific. In contrast with the assumption that pre-lexical units are phonemic, they found that listeners did not generalize the visually-guided learning for /b/ in one vowel context (i.e., /aba/) to perception of the same phoneme /b/ in another vowel context (i.e., /ibi/). Furthermore, exploring the role of features, they also tested generalization from /aba/ to /ama/—i.e., whether the perceptual recalibration about place of articulation (the [place] feature for /b/ vs. /d/) may be generalizable to the /m/-/n/ contrast with the same [place] feature in question but across the manner of articulation (stops vs. nasals). Again they found no generalization. That is, the category boundaries regarding the [place] feature between /ama/ and /ana/, and between /aba/ and /ada/ were both amenable to be recalibrated if the visually-guided exposure contained the same manner of articulation (stops or nasals), but no generalization of the learning on the [place] feature occurred across different manners of articulation (from stops to nasals and vice versa).

Reinisch et al.'s results provide a useful challenge for theories that assume that speech processing is based on phonological features, independent of whether features are articulatory or acoustic. In a featural account, the input is generally analyzed as being decomposable into independent features, so that learning about one feature (e.g., place of articulation) should generalize to a new situation that involves the feature, independent of whether or not the new situation involves other features (such as manner features) which are not included in the situation in which learning has taken place. The aforementioned studies on perceptual learning, however, suggest that this is not the case. Instead, they show that generalization does not easily occur when there are phonetic differences in the surface (phonetic) representation between the segment with which learning has taken place (the learning condition) and the new segment to which learning may be generalizable (the generalization condition). This implies that perceptual recalibration

[1] Alternatively, one might argue that no learning should occur, since contextually specified features are not coded in the lexicon, and hence, lexically induced recalibration should not occur. But since learning was observed in the baseline condition by Mitterer et al. (2013), this alternative explanation can be ruled out as well.