Journal of Phonetics 55 (2016) 96–108

Contents lists available at ScienceDirect

Journal of Phonetics

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

Research Article

Exposure modality, input variability and the categories of perceptual recalibration

Eva Reinisch [a,*], Holger Mitterer [b]

[a] Institute of Phonetics and Speech Processing, Ludwig Maximilian University Munich, Germany
[b] Department of Cognitive Science, University of Malta, Malta

ARTICLE INFO

Article history:
Received 1 July 2015
Received in revised form
12 December 2015
Accepted 22 December 2015
Available online 19 January 2016

Keywords:
Speech perception
Perceptual learning
Audiovisual processing
Prelexical processing

ABSTRACT

Recent evidence shows that studies on perceptual recalibration and its generalization can inform us about the presence and nature of prelexical units used for speech perception. Listeners recalibrate perception when hearing an ambiguous auditory stimulus between, for example, /p/ and /t/ in unambiguous lexical context (*kee[p/t]->*/p/, *mee[p/t]->*/t/) or visual context (presence vs. absence of lip closure). A later encountered ambiguous auditory-only stimulus is then perceived in line with the previously experienced context. Unlike studies using lexical context to guide learning, experiments with the visual paradigm suggested that prelexical units are rather specific and context-dependent. However, these experiments raised doubts whether lexically-guided and visually-guided recalibration are targeting the same type of units, or whether learning in the visually-guided paradigm—with limited variability during exposure—is task-specific. The present study shows successful visually-guided learning following exposure to a variety of different learning trials. We also show that patterns of generalization found with the visually-guided paradigm can be replicated with a lexically-guided paradigm: listeners do not generalize a recalibrated stop contrast across manner of articulation. This supports suggestions that the units of perception depend on the distribution of relevant cues in the speech signal.

1. Introduction

The nature of mental representations has been a long-standing issue in cognitive science. In the study of language, this leads to the question how we store our knowledge about the words of our language(s), and through which processes they can be accessed during listening. In recent years, the field converged on the conclusion that lexical access involves at least to some extent abstraction from the acoustic form of the input and the use of some kinds of units mediating between acoustic input and lexical representations (see e.g., Goldinger & Azuma, 2003, for an overview). In order to investigate the nature of these intermediate units, a number of recent studies made use of a perceptual learning paradigm ("phonetic recalibration"), based on the assumption that patterns of generalization are indicative of the shape of these intermediate units (e.g., Mitterer, Scharenborg, & McQueen, 2013; Reinisch, Wozny, Mitterer, & Holt, 2014). Since, however, the list of proposed units is long, and previous evidence for generalization or specificity of perceptual learning is mixed, the present study addressed two related questions: First, what role does variability in the input play in the occurrence of recalibration in different perceptual learning paradigms? This is to disentangle previous conflicting results that may be attributed to the use of visual vs. lexical context information to guide recalibration. Second, what role does variability in the input play in the occurrence of *generalization* of perceptual learning? Establishing conditions in which perceptual learning does or does not occur and generalize will allow us to assess what kinds of units listeners use for speech perception.

The perceptual learning paradigms discussed here were initially used to investigate how listeners use various types of context information to adjust or "recalibrate" speech perception in adapting to unusual speakers. That is, if listeners hear a speaker produce a sound that is ambiguous between /f/ and /s/ in lexical context where the interpretation as /f/ leads to the perception of a real word but /s/ does not (e.g., *giraffe* where *gira[s]* is not a word in English), listeners tend to resolve the acoustic ambiguity via the lexical context (i.e., they interpret the ambiguous sounds as /f/; Ganong, 1980). After repeated exposure to ambiguous sounds in unambiguous

* Correspondence to: Institute of Phonetics and Speech Processing, Ludwig Maximilian University Munich, Schellingstr. 3, 80799 Munich, Germany. Tel.: +49 89 2180 5752.
E-mail addresses: evarei@phonetik.uni-muenchen.de (E. Reinisch), holger.mitterer@um.edu.mt (H. Mitterer).

0095-4470/$ - see front matter © 2015 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.wocn.2015.12.004

lexical context by that speaker, listeners interpret ambiguous sounds in line with the previously experienced context even if the current context is ambiguous. That is, listeners who had heard ambiguous sounds in words like *giraffe* before would perceive *knife* more often than *nice* (Norris, McQueen, & Culter, 2003; Samuel & Kraljic, 2009, for an overview over similar types of studies). Similar effects have been shown with visual (lipread) context. After hearing an ambiguous auditory stimulus between /b/ and /d/ paired with the visual image of the speaker closing his/her lips (indicating the labial) or not (here: indicating the alveolar sound) an ambiguous auditory-only stimulus is perceived in line with the previously experienced visual stimulus (Bertelson, Vroomen, & deGelder, 2003). These types of perceptual learning have been termed lexically-guided and visually-guided recalibration.

1.1. The units of perceptual recalibration

As suggested above, such studies on perceptual learning not only contributed to our understanding of adaptation to unusual pronunciation variants but also to the debate about the existence of pre-lexical units. Goldinger (1998) started off this debate by suggesting that the mental lexicon contains the combined acoustic traces of the words a listener has heard in a lifetime. This questioned the existence of pre-lexical units altogether. This extreme point of view, however, turned out to be difficult to maintain, among others, in the light of experiments on perceptual learning (e.g. McQueen, Cutler & Norris, 2006; see also Mitterer, Chen, & Zhou, 2011; Sjerps & McQueen, 2010). With the current state of research there seems to be a general consensus that listeners use some form of abstract units, but are also able to store exemplars (Cutler, Eisner, McQueen, & Norris, 2010; Goldinger, 2007). The present study takes the reasoning from perceptual-learning studies one step further: if perceptual learning studies can show the existence of some type of pre-lexical units, they might also reveal the shape of these units (see Mitterer et al., 2013; Reinisch et al., 2014 for similar arguments).

Following this line of thought, recent studies on perceptual learning specifically investigated the grain-size of prelexical units that are used in speech perception. Suggestions for candidate units range from abstract phonemes (McClelland & Elman, 1986), via articulatory features that are directly perceived from the speech input (D'Ausilio et al., 2009; Galantucci, Fowler & Turvey, 2006) to context-invariant phonological features (e.g., Chomsky & Halle, 1968; Lahiri & Reetz, 2002; Marslen-Wilson & Warren, 1994). Notably, all of these units have in common that they assume independence from the surrounding context, which should allow for some generalization. Indeed, Jesse and McQueen (2011) showed that perceptual recalibration of a fricative contrast can generalize from syllable-coda to –onset position. Mitterer et al. (2011) found generalization across phonetic contexts. But the phoneme may be too big a unit for recalibration.

The "next step down" in the size of abstract units is to assume phonological or articulatory features as the units of perception. On the one hand, abstract phonological features could simply be seen as a construct from linguistics that conveniently serves the description of language. As such, they may fall short of psychological reality. On the other hand, there seems to be (at least some) evidence in line with claims that abstract phonological features are not only useful for linguistic description but are also primary in language processing (Cornell, Lahiri, & Eulitz, 2013; Embick & Poeppel, 2014; Roon & Gafos, 2014; Scharinger, Merickel, Riley, & Idsardi, 2011). Moreover, Embick and Poeppel (2014) argue for using phonological features as the base for neurobiological investigations of language (see also e.g., Poeppel, Idsardi & van Wassenhove, 2008). Importantly, the idea of abstract features is also supported by findings on generalization of perceptual learning, for example, the generalization of a recalibrated voicing contrast from an alveolar place of articulation (i.e., /d/–/t/) to labial place of articulation (/b/–/p/; Kraljic & Samuel, 2006). That is, recalibration may not be specific to a certain exposure contrast but rather affect features or gestures, for example, for "voicing". However, in all these cases, the acoustic variation in the implementation of the phonological contrast over contexts was rather small, so that the finding of generalization does not contradict a feature account but cannot be used to argue *for* abstract features. Abstract features, like phonemes, are supposedly independent of context and hence predict generalization over acoustically dissimilar cues.

Recent studies then focused on cases in which the acoustic cues to a given distinction varied over contexts and these studies support some kind of context-sensitivity (e.g., Mitterer et al., 2013, Reinisch et al., 2014). Mitterer et al. (2013) showed that Dutch listeners recalibrate an /l/–/r/ contrast but fail to show recalibration when tested on one of the articulatorily and acoustically different allophones of /r/ (i.e., alveolar approximant vs. apical trill). The authors suggest that (context-sensitive) allophones may be the units of recalibration. Interestingly, this example also shows how the assumption of different types of features (i.e., abstract phonological features vs. gestures) results in different predictions regarding generalization. Assuming abstract features, one would have to predict that learning generalizes from one allophone to another, because all allophones of a given phoneme (in a given language) share the same feature specification (Lahiri & Reetz, 2002). If, in contrast, features were defined in terms of articulatory gestures (Galantucci et al., 2006) then the results could still be explained since the two allophones used in Mitterer et al. make use of different articulatory gestures. Finally, an account of acoustic similarity as requirement for generalization would also support the lack of generalization (see e.g., Kraljic & Samuel, 2006; Reinisch & Holt, 2014).

In order to adjudicate between abstract features, articulatory features, or (context-sensitive) allophones as the prelexical units that are relevant to perceptual recalibration, Reinisch et al. (2014) tested conditions of generalization in a tightly controlled experimental setup. They tested recalibration of a place of articulation contrast (/b/–/d/) and its generalization. A first experiment tested the generalization from "a[$^{d}/_{b}$]a" to "i[$^{d}/_{b}$]i" (where [$^{d}/_{b}$] indicates a perceptually ambiguous sound between /b/ and /d/). Via additional cue-trading experiments, Reinisch et al. established that the labial–alveolar distinction was cued mainly by formant transitions in the "a[$^{d}/_{b}$]a" case but mainly by burst spectra in the "i[$^{d}/_{b}$]i" case. The results of perceptual recalibration then showed that listeners did not generalize learning across these contexts (in either direction). While this first experiment tested the generalization to other acoustic implementations of the same phonemic contrast (i.e., comparable to allophones), the second experiment tested whether