Journal of Memory and Language 98 (2018) 77–92

Contents lists available at ScienceDirect

Journal of Memory and Language

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

Allophones, not phonemes in spoken-word recognition ☆

Holger Mitterer [a,*], Eva Reinisch [b], James M. McQueen [c,d]

[a] Department of Cognitive Science, Faculty of Media and Knowledge Sciences, University of Malta, Msida, Malta
[b] Institute of Phonetics and Speech Processing, Ludwig Maximilian University Munich, Germany
[c] Donders Institute for Brain, Cognition, and Behaviour, Radboud University, Nijmegen, The Netherlands
[d] Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

ARTICLE INFO

Article history:
Received 20 July 2016
Revision received 1 September 2017
Available online 28 September 2017

Keywords:
Spoken-word recognition
Phonemes
Allophones
Pre-lexical representations
Selective adaptation

ABSTRACT

What are the phonological representations that listeners use to map information about the segmental content of speech onto the mental lexicon during spoken-word recognition? Recent evidence from perceptual-learning paradigms seems to support (context-dependent) allophones as the basic representational units in spoken-word recognition. But recent evidence from a selective-adaptation paradigm seems to suggest that context-independent phonemes also play a role. We present three experiments using selective adaptation that constitute strong tests of these representational hypotheses. In Experiment 1, we tested generalization of selective adaptation using different allophones of Dutch /r/ and /l/ – a case where generalization has not been found with perceptual learning. In Experiments 2 and 3, we tested generalization of selective adaptation using German back fricatives in which allophonic and phonemic identity were varied orthogonally. In all three experiments, selective adaptation was observed only if adaptors and test stimuli shared allophones. Phonemic identity, in contrast, was neither necessary nor sufficient for generalization of selective adaptation to occur. These findings and other recent data using the perceptual-learning paradigm suggest that pre-lexical processing during spoken-word recognition is based on allophones, and not on context-independent phonemes.

Introduction

One of the fundamental questions in cognitive science regards the nature of the mental representations that underlie cognitive functioning. In spoken-word recognition, the question is which code we use to map the highly variable speech signal onto knowledge stored in the mental lexicon – knowledge about the phonological form of words. What, in short, are the pre-lexical units of speech perception?

Theories answer this question in many different ways. Some theories claim that there are no phonologically abstract pre-lexical representations (Goldinger, 1998) and others that there are, but disagree about the grain-size of the units, which could be abstract phonological features (Lahiri & Reetz, 2010), context-dependent allophones (Luce, Goldinger, Auer, & Vitevitch, 2000), context-independent phonemes (McClelland & Elman, 1986; Norris, 1994), or syllables (Mehler, Dommergues, Frauenfelder, & Segui, 1981), or could be a combination of units of different size (Wickelgren, 1969). One recurring issue in this long-running debate has been that evidence in favour of one or the other type of unit often turned out to be paradigm-specific. Evidence for many different units can therefore be found (for a review, see Goldinger & Azuma, 2003).

For instance, evidence in favour of syllables stems from monitoring paradigms (Mehler et al., 1981) and illusory conjunctions in dichotic listening (Kolinsky, Morais, & Cluytens, 1995). However, Dumay and Content (2012) were not able to find converging evidence for syllables with auditory priming of shadowing responses. In other cases, evidence from subcategorical mismatches (i.e., where a secondary cue for a phonetic distinction mismatches the primary cue that determines the percept, e.g., *jo$_g$b*) supposedly favoured a featural account (Marslen-Wilson & Warren, 1994), but it was later shown that these data are also in line with an account assuming segments (McQueen, Norris, & Cutler, 1999). A general problem in this area of research has been the long chain of auxiliary assumptions that linked theoretical claims about units of speech perception to the data. Results are thus open to multiple interpretations. For example, evidence of phonetic priming with no phonemic overlap (e.g., from bull to veer; Goldinger, Luce, & Pisoni,

☆ This work was supported by a University of Malta Research Grant to the first author. The second author was supported by an Emmy-Noether Fellowship by the German Research Council (DFG, grant nr. RE 3047/1-1).

* Corresponding author at: Department of Cognitive Science, Faculty of Media and Knowledge Sciences, University of Malta, Msida MSD 2080, Malta.
E-mail address: holger.mitterer@um.edu.mt (H. Mitterer).

https://doi.org/10.1016/j.jml.2017.09.005
0749-596X/

1989) could be taken as evidence for units smaller than the phoneme (e.g. phonological features), but are also consistent with accounts with no abstract phonological units that instead capture phonetic similarity in terms of acoustic similarity (e.g., Goldinger, 1998). Many classic paradigms depend on meta-linguistic judgements (e.g., about syllables, Mehler et al., 1981) and may thus reflect the conscious products of speech processing and/or task-specific processing rather than the units that are extracted during pre-lexical perceptual processing (McQueen, 2005).

Recent evidence from learning and adaptation paradigms has breathed new life into this debate. This is because such paradigms offer the possibility of establishing which units play a role in speech perception by asking which units are learned about, and thus offer a more direct measure than the classic paradigms. Importantly, data from a perceptual-learning paradigm showed that some form of prelexical unit has to be assumed to allow learning to generalize from one set of words to another (McQueen, Cutler, & Norris, 2006; Mitterer, Chen, & Zhou, 2011; Sjerps & McQueen, 2010). Regarding the size of the units, data using this perceptual-learning paradigm supports the hypothesis that there are allophonic units (Mitterer, Scharenborg, & McQueen, 2013), while data using a selective-adaptation paradigm supports the additional hypothesis that there are also phonemic units (Bowers, Kazanina, & Andermane, 2016). The present study tests these two representational hypotheses. We define an "allophone" as a speech segment with a distinct acoustic realization that can be context dependent and position specific, but not necessarily so (e.g., English /l/ has "light" and "dark" allophones, [l] and [ɫ],[1] which are position specific; but English /f/ has only one allophone, [f], which appears in different positions). We define a "phoneme" as a context-independent and position-nonspecific representation of a speech segment (e.g., /l/ and /f/).

There is an important a priori reason to favour the allophone as the pre-lexical representation in speech recognition. The primary function of pre-lexical processing is to help the listener solve the invariance problem. The invariance problem is arguably the central problem of speech perception, that there are no physically invariant cues that go along with any given unit of speech. The speech signal varies enormously (as a function of talker and style differences, phonological context effects, background noise and so on) and yet the listener needs to be able to recognize the words the talker intends despite this variability. Pre-lexical representations of the segmental content of the incoming speech signal provide a means for phonological abstraction, linking between the variable input and the (phonologically abstract) mental lexicon. On this view, context-dependent allophonic units are more plausible than context-independent phonemic units precisely because speech segments are not context independent. As noted above, English /l/, for example, has light (syllable-initial) and dark (syllable-final) allophonic variants. Variability about light [l] may be irrelevant and potentially even misleading for the recognition of dark [ɫ], and vice versa. If listeners have allophonic units, they could optimize the mapping of the input onto the lexicon for each allophone separately. This would be harder to achieve with phonemic units. In short, the listener needs to track the acoustic variability relevant for word recognition, and those acoustics are not always position-invariant.

Evidence from perceptual learning supports the allophonic account (Mitterer et al., 2013). As Mitterer et al. argued, perceptual-learning paradigms can be used to address this issue because these paradigms reveal the units that are functional in solving the invariance problem. In the paradigm as first used by Norris, McQueen and Cutler (2003), participants learn about an unusual pronunciation of a given segment. In the original study this was a fricative that was perceptually ambiguous between /f/ and /s/ (henceforth [$^{s}/_{f}$] and analogously for other segments). Participants heard this segment either replacing /s/ in /s/-final words (e.g., [maʊ $^{s}/_{f}$] for *mouse*) or replacing /f/ in f-final words (e.g., [ʃɛɹɪ $^{s}/_{f}$] for *sheriff*). This was implemented as a between-participant factor, and, after exposure, both groups categorized sounds along an /f/-/s/ continuum. Participants who heard [$^{s}/_{f}$] replace /s/ categorized members of this continuum more often as /s/ than participants who heard [$^{s}/_{f}$] replace /f/. Importantly, this was not a simple perceptual adaptation, as no such effect occurred if the ambiguous sound occurred in nonwords. This suggests that the participants had used the lexical contexts during exposure to learn about the intended identity of the ambiguous sound.

This paradigm is well-suited to investigate the nature of pre-lexical representations for two reasons. First, learning has been shown to generalize from one set of words to other words (McQueen et al., 2006; Mitterer et al., 2011; Sjerps & McQueen, 2010), even if the other words come from a different language than those heard during exposure (Reinisch, Weber, & Mitterer, 2013). Perceptual learning therefore appears to target representations that are functional in spoken-word recognition. Once listeners have learned about a given talker's way of speaking, they can apply what they have learned to other words containing the same sound, helping them to understand the talker. Second, Mitterer and Reinisch (2013) used eye-tracking to show that perceptual learning influences the processing of speech at the same point in time as the phonetic differences in the signal itself. Visual-world eye-tracking has been shown to reveal the processing of possible referents to the speech signal at a constant delay of about 150–200 ms. This delay is caused by the planning of eye movements (Salverda, Kleinschmidt, & Tanenhaus, 2014). Mitterer and Reinisch (2013) showed that effects of perceptual learning could also be detected at this point in time. That is, perceptual learning influences processing at a pre-lexical level, at the same time as acoustic input is being analysed phonetically.

Given that the perceptual-learning paradigm shows generalization of learning across words and early effects on processing, the extent of generalization across sounds may be used to gauge the grain-size of the pre-lexical representations involved. If learning were entirely position- and context-independent (i.e., if generalization would occur across the board), this would argue for the use of phonemes, which are defined as being context- and position-independent. Mitterer et al. (2013) showed however that learning about the /r/-/l/ boundary in Dutch based on the allophones [ɹ] and [ɫ] does not generalize to acoustically and articulatorily different implementations of the phonemes /r/ and /l/. These findings suggest that the units of speech perception are allophonic.

Even more specific learning has been reported by Reinisch, Wozny, Mitterer, and Holt (2014), who tested learning for /b/ versus /d/, and found that learning is specific to vowel context, so that learning for [aba] versus [ada] did not generalize to [ibi] versus [idi]. This again argues for allophonic representations, but with even more specificity than allophones are typically associated with. That is, the term allophone is usually used to describe two quite distinct versions of the same phoneme with clearly different articulations. The data of Reinisch et al. (2014) suggest that even small acoustic differences can give rise to independent representations of the same phoneme in spoken-word recognition.

The studies of Mitterer et al. (2013) and Reinisch et al. (2014) indicate that perceptual learning can be used to delineate the nature of pre-lexical representations and suggest that those representations are allophonic. As we have already argued, it makes sense that learning about one allophone does not generalize to the processing of another allophone, since the two allophones are

[1] Throughout this paper, we follow the linguistic convention that forward slashes indicate phonological forms, which do not distinguish between different allophones of the same phoneme, while square brackets indicate phonetic forms.