Allophones, not phonemes in spoken-word recognition

Holger Mitterer — University of Malta
Eva Reinisch — Ludwig-Maximilians-Universität München
James M. McQueen — Radboud University Nijmegen; Max Planck Institute for Psycholinguistics

- Pre-lexical representations in speech perception were probed with selective adaptation
- Allophonic and phoneme overlap between adaptors and test stimuli were varied
- Only allophonic overlap led to selective adaptation
- Results argue for the use of allophones and not phonemes in speech perception

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.

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?

Holger Mitterer, Department of Cognitive Science, Faculty of Media and Knowledge Sciences, University of Malta, Msida Malta. Eva Reinisch, Institute of Phonetics and Speech Processing, Ludwig Maximilian University Munich, Germany. James M. McQueen, Donders Institute for Brain, Cognition, and Behaviour, Radboud University, and Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands.

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).

Correspondence concerning this article should be sent to: Holger Mitterer, Department of Cognitive Science, Faculty of Media and Knowledge Sciences, University of Malta, Msida MSD 2080, Malta or electronically to holger.mitterer@um.edu.mt

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, 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 [ɬ][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 recognise 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 [ɬ], 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 et al. (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

[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.

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 [ɬ] 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 acoustically distinct. Bowers at al. (2016) offer an analogy from visual-word recognition to make the same point. Visual-word recognition is often assumed to make use of abstract letters that are independent of case, but learning to recognize an uppercase "A", for example, should have little impact on how a small, lowercase letter "a" should be recognized. But Bowers et al. also use this analogy to argue that the perceptual-learning evidence in favour of allophones does not rule out the possibility that phonemes also play a role in pre-lexical speech processing. It is important to note that assuming there are phonemes implies that there are also allophones, just as assuming abstract letter codes requires that there are representations for small and capital letters. On this view, phonemes and allophones would naturally be in a processing hierarchy, as exemplified in the left panel in Figure 1 for the Dutch word for *heavy,* /zʊar/. The last segment /r/ can be produced with an approximant /r/, which we transcribe here as [ɹ].[2] On this account, allophonic units are used to abstract from the acoustic input and to adapt to novel pronunciations, but a further level of abstraction leads to phonemes, which in turn lead to words.

Interestingly, there have been further attempts to show with the perceptual-learning paradigm that more abstract units may play a role. This has been achieved by testing whether learning might be reduced if there is allophonic overlap but a phonemic difference. This can be achieved by exploiting phonological neutralization, such as final-obstruent devoicing in German, where /hund/ (Engl., *dog*) surfaces as [hunt].[3] Mitterer, Kim, and Cho (2016) used tensification in Korean where a lax stop surfaces as tense if preceded by an obstruent. They found that learning about place of articulation on a tensified lax stop generalizes to both underlying tense and lax stops. The first result indicates that a difference in phonemic identity does not impede generalization, hence complementing the finding that phonemic

[2] There is disagreement about the exact place of articulation of this approximant (cf. Van Bezooijen, 2005), but the exact place of articulation is irrelevant, as all approximants are acoustically and articulatory highly distinct from trilled variants of Dutch /r/. These variants can be alveolar or uvular trills.

[3] Importantly, underlying voicing matters in other forms of the same word, such as the plural /hund/ + plural →[hundə], where devoicing would be ill-formed (*[huntə]).

identity is not sufficient to lead to generalization (Mitterer et al., 2013). Generalization to lax stops, however, could be interpreted as showing that sharing phonemic identity fosters generalization. However, an alternative explanation was that the generalization arose because of the acoustic similarity of cues to place of articulation in tense and lax stops in Korean. Therefore, Mitterer and Reinisch (in press) tested generalization to voiced stops from devoiced stops in German, where the acoustic differences are larger. They found no generalization to phonologically identical voiced stops (that were acoustically dissimilar), but again generalization to phonologically unvoiced stops (that were acoustically similar), again showing that phonemic identity has little role to play in functional adaptations in speech perception.

Bowers et al. (2016) used a different paradigm, selective adaptation, to test for pre-lexical phonemic units. Selective adaptation occurs if participants are repeatedly exposed to one stimulus (e.g., /ba/ or /da/) and are later asked to categorize ambiguous syllables that could be categorized as one of the adaptors (e.g., [$^{b}/_{d}$a]). It is typically observed that the ambiguous sound is perceived as contrasting with the adaptor, that is after hearing [ba], [$^{b}/_{d}$a] is perceived as /da/ more often than after adaptation to [da] (for a critical review, see Remez, 1987). Bowers et al. tested whether selective adaptation generalizes from word-medial and word-final position to the onset position for two contrasts: /b/ versus /d/ and /s/ versus /f/. Also, in contrast to most studies on selective adaptation, they used lists of words that contained a given sound in a given position rather than repeating the same adaptor stimulus. That is, while most selective adaptation studies used the same token of [ba] repeatedly to generate adaptation to /b/, Bowers et al. (2016) used a list of /b/-initial (or /b/-medial or /b/-final) words (e.g., *bail, balance, bank,* etc.). For both types of segment contrasts tested (i.e., /b/-/d/ and /s/-/f/), selective adaptation generalized over syllable position. However, generalization was strongly reduced relative to a control condition with onset adaptors and onset test syllables. The effect of generalization across position was only one third of the size of the effect within position. The model on the left in Figure 1 captures the data of Bowers et al. (2016) well. They found that adaptation to word-medial and word-final stops influences perception of initial stops, but not as strongly as adaptation to word-initial stops. Adaptation to initial segments appears to lead to adaptation on allophonic and phonemic levels, while adaptation of stops in other positions appears to lead only to adaptation at the phonemic level.

Bowers et al. (2016) highlight the issue that competent speakers of a language must know about phonemes. In the case of the Dutch word *zwaar*, a speaker needs to know that the underlying phoneme is /r/, because the approximant cannot be used for the inflected form *zware* (*[zʊa.ɹə]; the approximant can only be used in coda position, in syllable onset position a trill or tap has to be used (e.g., [zʊa.rə]). Such examples show that the need for speakers to know about phonemes cannot be disputed. But it is

not the case that this knowledge must reside in the pre-lexical perceptual system. The knowledge is required mainly in speech production, where speakers must choose the correct allophone for an inflected form. Since speech perception and speech production are not necessarily tightly linked (Lotto, Hickok, & Holt, 2009; Mitterer & Ernestus, 2008; Ohala, 1996), knowledge about phonemic identity does not need to reside in the speech perception system for it still to be able to influence speech production. As shown in the right panel of Figure 1, retrieval of phonemic representations need not be a pre-requisite for lexical access in perception. While the phoneme representations in the model in the left panel are used in both perception and production, the phonemes in the model in the right panel are used only in production. According to this latter model, allophones are used to access words in the mental lexicon (as suggested by the data from Mitterer et al., 2013, and Reinisch et al., 2014). Lexical access, however, makes it possible to retrieve knowledge about the underlying phonemes – knowledge that can then be used in production.

In summary, the current data from perceptual learning and selective adaptation suggest that allophones are pre-lexical representations. But these data are ambiguous about the status of phonemes. The present study sought to resolve this ambiguity by adjudicating between the two models in Figure 1.

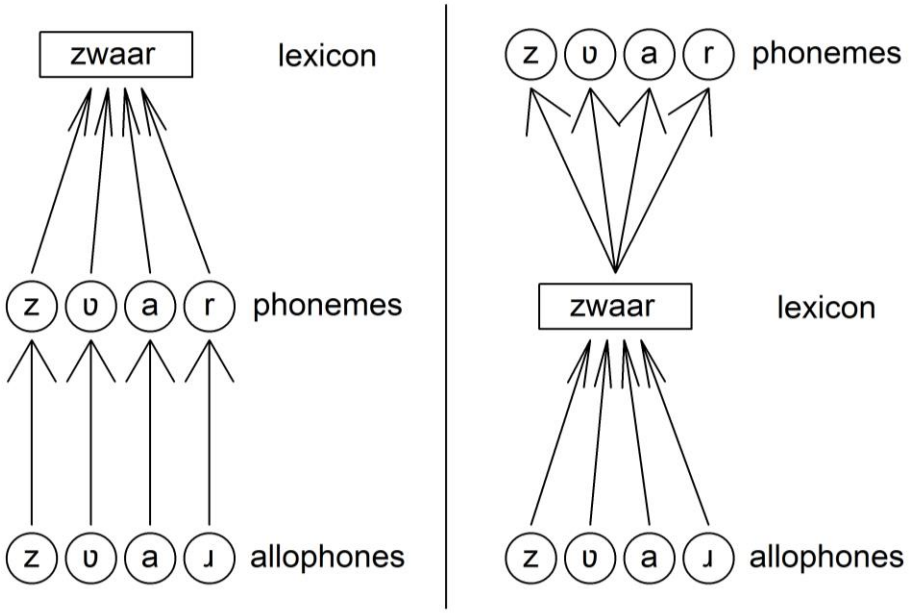

Figure 1. Possible architectures that explain how listeners represent allophonic variation and are still able to produce context-appropriate versions of a given phoneme.

We argue that testing between these hypotheses depends critically on the choice of segments. The previous data with voiceless fricatives, both from perceptual learning and selective adaptation (Bowers et al., 2016; Jesse & McQueen, 2011) are consistent with both hypotheses. Even in the perceptual-learning paradigm, there is generalization of learning across position for fricatives (Jesse & McQueen, 2011), and there is no moderation in terms of effect size in the amount of learning from coda-to-coda position to coda-to-onset position. This is presumably because voiceless fricatives are physically very similar across positions (Bowers et al., 2016; Jesse & McQueen, 2011; Mitterer et al., 2013). If voiceless fricatives therefore do not have different allophones for different positions, generalization of learning or adaptation across position with these segments can be explained at the allophonic level in both models.

The critical finding apparently arguing against the model without phonemes in perception (i.e., the model in the right panel in Figure 1) is the generalization of selective adaptation across position for place of articulation in stop consonants (the /b/-/d/ contrast tested by Bowers et al., 2016). The argument hinges on showing that there is adaptation in a case where there is no acoustic overlap. This would be inconsistent with the model without phonemes in perception and offer support for the model with phonemes (with the adaptation arising at the phonemic level). Stop consonants could provide this test. Especially in English, stops can surface in quite different acoustic forms in onset and offset position. For instance, only stops in offset position can be unreleased (i.e., not contain a stop burst) and they can be cut short by glottalization (i.e., stopping of the air stream by closing the vocal folds). However, while stops *can* be implemented quite differently across positions in English, they need not be that different. Bowers et al. (2016, see footnote 2) note that most of their word-final stops contained release bursts, as did their exposure and test stimuli with stops in onset position. It has been argued that release bursts contain invariant cues to place of articulation (Stevens & Blumstein, 1978). There was therefore acoustic overlap between the exposure material and the test stimuli in the critical generalization condition. The spectral characteristics of release bursts are different for different places of articulation, but for the same place of articulation they are the same across word-final and word-initial positions (see, e.g., Liberman, 1996, for an extensive discussion of cues to place of articulation in stops; Reinisch et al., 2014 provide in addition an empirical demonstration that detailed stop burst properties can be crucial in determining the perceived place of articulation).

Furthermore, for cues to stop place of articulation in the formant transitions into and out of neighbouring vowels, it has been suggested that listeners extract complex auditory patterns, so-called formant loci (Sussman, Fruchter, Hilbert, & Sirosh, 1998) from the signal. If that is the case, adaptation

could be on formant loci rather than on phonemes, and this would also lead to adaptation across position. It has been shown that acoustic overlap gives rise to generalization in perceptual-learning paradigms (Jesse & McQueen, 2011; Kraljic & Samuel, 2006; Mitterer et al., 2016; Reinisch & Holt, 2014).

The stop data from Bowers et al. (2016) thus fail to provide a strong case against the model on the right in Figure 1. The adaptation (due to overlap of release bursts and/or formant loci) could reflect processes at the allophonic level or an even earlier stage of processing and are thus consistent with both models. The Bowers et al. stop data would have been much stronger if the exposure materials had contained only unreleased or even glottalized stops, with cues to place of articulation primarily in the formant transitions, while the test material contained stops with strong cues to place of articulation in the burst. Bowers et al. also discussed another case which again would have provided a strong test: the case of voicing in English stops. Stop voicing is cued by voice-onset time (i.e., the time difference between the stop release and the onset of vocal fold vibration) in onset position, but the dominant cue in offset position usually is vowel duration, especially when the stops are unreleased. In such cases, perceiving a word as, for instance, either *bag* and *back* can be determined by the duration of the vowel alone. This would have been a good testing ground for the assumption of position-independent phonemes, especially with an acoustic manipulation that made sure that vowel duration cue was minimal for stops in onset position. However, Bowers at al. chose the much more ambiguous case of testing generalization over position of place of articulation in stops.

It remains possible, however, that the difference in results between two paradigms – apparent evidence for phonemic units with selective adaptation (Bowers et al., 2013) but not with perceptual learning (Mitterer et al., 2013) – may not reflect effects of acoustic overlap but instead arise because the two paradigms test different levels of pre-lexical processing. Perhaps perceptual learning reveals earlier allophonic units while selective adaptation reflects the use of later phonemic units. A good way to test this possibility is to take advantage of the allophonic variation in the Dutch liquids /l/ and /r/. Mitterer et al. showed that perceptual learning about Dutch liquids targets allophonic units. Perceptual learning did not generalize from offset to onset position. Furthermore, there were clearly defined articulatory and acoustic differences between the segments used in offset and onset position. Exposure was based on the word-final approximant [ɹ] for /r/ and the “dark” [ɬ], that is an /l/ with an additional velar gesture next to the alveolar contact. At test the alveolar trill [r] and the “light” [l], with only an alveolar contact, were used in onset position. Importantly, these examples satisfy the condition that listeners must know that these allophones are related in order to be proficient speakers of Dutch. While the use of the approximant varies between speakers (Van Bezooijen, 2005), all speakers—apart from speakers of a few local variants in the

Leiden area—use a trill in onset position. This is also the case for the example provided in Figure 1, the Dutch word /zʊar/ (Engl., *heavy*), which has to be produced with a trill if used in a definite noun phrase, in which morphological inflection necessitates the form *zware* /zʊarə/ (e.g., *de zware tas* [də.zʊa.rə.tɑs], Engl., *the heavy bag*). As noted above, Bowers et al. (2016) argue that this is an important reason to represent phonemes. If indeed selective-adaptation reveals phonemic units in spoken-word recognition, we should find selective adaptation for word-final /l/ versus /r/ based on word-initial /l/ and /r/ as adaptors. We tested this prediction in Experiment 1.

Experiment 1

We tested what has previously been shown with perceptual learning (Mitterer et al., 2013) using the selective-adaptation paradigm of Bowers et al. (2016). Specifically, we tested selective adaptation as triggered by word lists in Dutch with different allophones of /l/ and /r/. As adaptors we used lists with "onset adaptors" and "offset adaptors". Onset adaptors contained word-initial /l/, produced as "light" [l] (e.g., *lessen, leiding, lijstje,* etc.), and word-initial /r/ produced as alveolar trill [r] (e.g., *restje, reiken, rente, …*). Offset adaptors contained word-final /l/ produced as "dark" velarized [ɬ] (e.g., *appel, bijbel, ezel, …*), and word-final /r/ produced as approximant [ɹ] (e.g., *bakker, emmer, puber, …*). For the test we used the Dutch minimal word pair /wɪmpəl/-/wɪmpər/, Engl., *pennant – eye lash*, produced as [wɪmpəɬ]-[wɪmpəɹ] and hence matching the offset adaptors. According to both theoretical positions, we should find a selective-adaptation effect triggered by the offset adaptors since these share allophonic and phonemic identity with the test stimuli. The critical question was whether selective adaptation would also be triggered by the onset adaptors that share phonemic identity but differ in allophonic identity.

We followed the design used by Bowers et al. (2016), with a few exceptions. First, we presented a small range of acoustic stimuli during the test phase rather than only one. This ensured that participants were really performing a phonetic-identification task (because they heard stimuli varying along a continuum) and that we had a better chance that each participant would hear tokens of both /l/ and /r/ (participants vary substantially in where they judge the cross-over point to lie on any given phonetic continuum). Second, we did not use letters but images as response options. A recent paper (Krieger-Redwood, Gaskell, Lindsay, & Jefferies, 2013) has shown that presenting letters as response options can lead to engagement of motor representations which are not routinely used in speech perception. Additional differences with the Bowers et al. (2016) study are that position was varied within participants rather than between participants (as in the earlier study) and the control condition used stimuli in coda position rather than in onset position.

The predictions of the different accounts are as follows. The phoneme-plus-allophones account predicts that both onset- and offset-adaptors should lead to selective adaptation, that is, less /r/-responses after adaptation to /r/ than after adaptation to /l/. Given the earlier results, the effect may be stronger effect for the offset adaptors, which overlap with the target items both in allophones and phonemes, than for the onset adaptors, which only overlap in terms of phonemes. The allophones-only account predicts adaptation (i.e., less /r/ responses after /r/ adaptation) only for the offset adaptors.

Methods

Participants

Twenty-eight native speakers of Dutch participated in the experiment. They were students at Radboud University, Nijmegen. Fifteen were female and they were all aged between 18 and 24 years. They reported no language, speech or hearing impairment. The study was approved by the Ethics Committee of the Faculty of Social Sciences of Radboud University. Participants gave written informed consent and were paid a small monetary compensation for an experimental session of approximately 50min.

Materials

We selected 25 words for each adaptor condition varying position (onset – offset) and phoneme (/r/ - /l/, see the Appendix for a full list). In line with the perceptual adaptation study by Mitterer et al. (2013) all offset pairs were disyllabic with the reduced vowel [ə] in the final syllable. For the onset condition, words were also disyllabic and the adjacent vowel was a high- or mid-front vowel. These 100 words plus several renditions of the critical minimal pair used for test (*wimpel-wimper)* were recorded by a female native speaker of Dutch, who was the same speaker who recorded the stimuli for Mitterer et al. (2013). She naturally used an alveolar trill in onset position and an approximant in offset position.

To create a continuum between the words of the minimal pair for the test-phase stimuli, we selected two recordings of the two endpoints with a similar duration and pitch contour. A 21-step continuum was created using the STRAIGHT morphing algorithm (Kawahara, Masuda-Katsuse, & de Cheveigné, 1999) using 5% steps. Based on the impression of the potentially ambiguous range of the continuum, a short pretest with seven native speakers of Dutch was performed with the steps 2 to 10 of the continuum. We used a Google image search to find two pictures that represented the response options *wimpel* and *wimper*, so that we did not have to rely on presenting orthographic code during the experiment. Participants were asked to press the “1” key if the word better matched the picture on the left and the “0” key if the word better matched the picture on the right, analogue to the spatial alignment

of the response-keys on a standard keyboard. The allocation of response options to the left or right side of the screen was constant during the pretest. After pressing a button, the chosen picture moved slightly to the upper left or right while the other picture disappeared. This feedback indicated to participants that their response had been recorded. Based on the result of the pretest, we selected steps three, five, and six (perceived as [wimpəɬ] in 85.7%, 32.9%, and 17.1% of the cases during the pretest) to use as the test stimuli in the main experiment. Figure 2 shows the two endpoints of the test-stimuli continuum and one example for each adaptor condition, which overlap in abstract phonemes (see figure caption), but do or do not overlap in terms of allophones, as indicated by the IPA transcription.

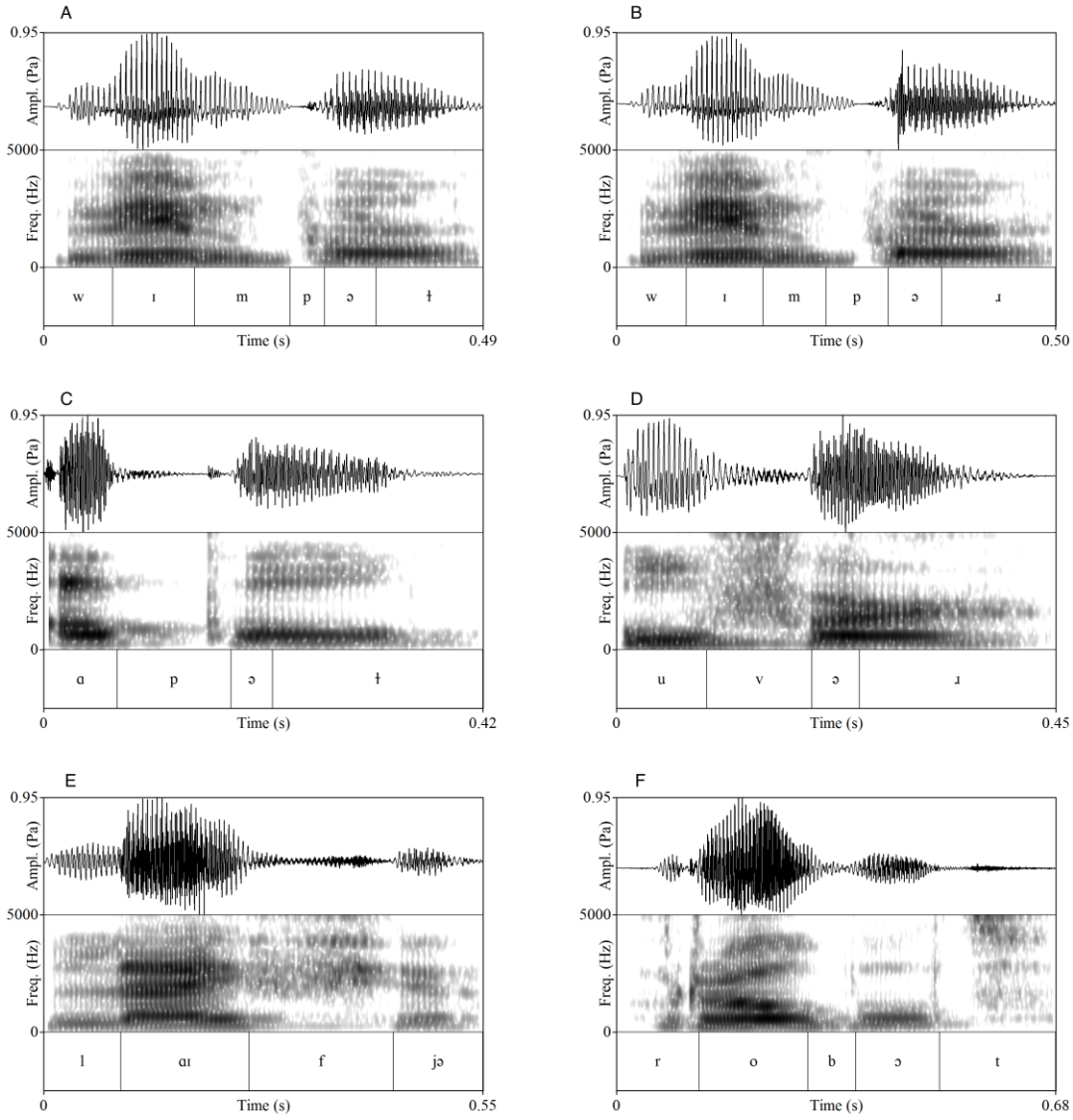

Figure 2. Test stimuli (panels A and B, the Dutch words *wimpel* /wɪmpəl/ and *wimper* /wɪmpər/) and examples of adaptor stimuli for the offset condition with allophonic overlap (panels C and D, the Dutch words *appel* /ɑpəl/ and *oever* /uvər/) and the onset condition without allophonic overlap (panels E and F, the Dutch words *lijfje* [lɑɪfjə] and *robot* /robɔt/).

Procedure

Participants were informed that the experiment would require them to listen passively to a stream of speech stimuli over headphones and then on nine trials indicate by button press which of two words they heard. They were also told that this procedure would be repeated several times in four different blocks. The experiment was divided into four blocks, one block for each adaptor. One block consisted of eight sub-blocks in which participants were presented with the 25 adaptor words twice for that block in a random order, presented with an inter-stimulus interval (ISI) of 600ms. One block lasted about one minute and was followed by nine phonetic-identification trials, in which each of the three selected stimuli was presented three times. All auditory stimuli were presented binaurally. Experiments were run with PsychoPy2 (Peirce, 2007), and random orders of adaptors were generated online and for each (sub-)block anew. The phonetic-identification trials had the same structure and format as the pretest. We used all 24 possible orders of the four blocks over the 24 participants, so that each adaptor occurred equally often in the same position and preceded and/or followed all other adaptor types.

Results and Discussion

The data from four participants were excluded because they identified all test stimuli as /r/. When it was noticed that a participant responded in that way, another participant was assigned to the same order of adaptor blocks. In the remaining sample of 24 participants, each possible order of the four adaptor blocks was used once. The mean proportion of /r/-responses for each adaptor condition is shown in Figure 3, both in log odds and in raw proportions. We present both here because proportions may be more familiar and useful to gauge the results, but the analysis was performed on log odds (Dixon, 2008). The data show a clear selective-adaptation effect for the adaptors in offset position—which is the same position as in the test words—but only a small difference between the /r/ and /l/ adaptors in onset position.

We performed an ANOVA[4] on the log odds of /r/ responses with the predictors *Position, Step* and *Adaptor.* Analyses were performed in R using the function aov and revealed various significant interactions, including a two-way interaction of adaptor x position ($F(1, 23) = 31.092$, $p < 0.001$) and a three-way interaction of adaptor x position x step ($F(2, 46) = 4.815$, Greenhouse-Geisser $\varepsilon = 0.630$, $p < 0.05$). Therefore, we ran separate analysis for the onset and offset adaptor conditions.

[4] We use ANOVAs and t-test here rather than generalized linear mixed effects models with a logistic linking function, because the former are more easily linked to the following Bayesian statistics, which are important to argue for the null hypothesis in the current case.

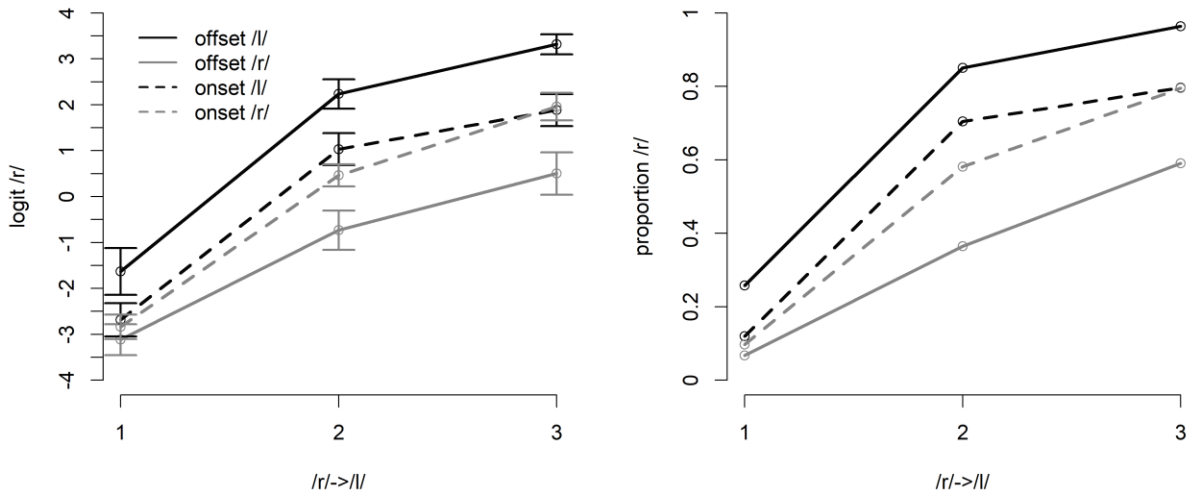

Figure 3. Mean and standard error of log odds and proportion of offset /r/ responses to the three steps on the /r/-/l/ continuum in Experiment 1. There is a clear effect of adaptor in offset position, but not in onset position. Note that the log odds are the mean log odds of all participants and not the log odds of the mean of all participants. Error bars represent standard errors based on Morey's (2008) method for normalization of repeated-measure data.

For the offset-adaptor condition (i.e., with allophonic overlap), the ANOVA revealed an effect of adaptor $F(1, 23) = 45.70$, $p < 0.001$), an effect of step ($F(2, 46) = 116.84$, $\varepsilon = 0.646$, $p < 0.001$) and an interaction of the two factors ($F(2, 46) = 7.779$, $\varepsilon = 0.834$, $p = 0.002$). Further testing of the adaptor effect at all steps showed that there was a clear adaptation effect with fewer /r/-responses after /r/ adaptation for all steps (min(logit difference) = 1.4 logit units, min(t(23)) = 3.722, max(p) = 0.001). That is, the interaction only indicates a weakening of the effect at the first step, but no absence of adaptation at any of the steps in the offset-adaptor condition.

For the onset-adaptor condition (i.e., with phonemic overlap but without allophonic overlap), there was an effect of step ($F(2, 26) = 152.982$, $\varepsilon = 0.402$, $p < 0.001$) but neither an effect of adaptor ($F(1, 23) = 1.27$, $p = 0.271$) nor an interaction ($F(2, 26) = 1.969$, $\varepsilon = 0.402$, $p = 0.163$). Given that the adaptor effect (and its interaction with step) is not significant, we also performed a Bayesian ANOVA using the default priors implemented in the R package BayesFactor (version 0.9.12-2). This analysis leads to a Bayes Factor (BF) as test statistic (Rouder, Speckman, Sun, Morey, & Iverson, 2009), which allows an estimation of how well the null hypothesis is supported by the data in comparison to the alternative hypothesis that the effect is not null. The BF provides an estimate of the likelihood of the alternative hypothesis over the null hypothesis, so that a BF of three means that the alternative hypothesis is three times more likely than

the null hypothesis. Moreover, a BF of three or larger is considered substantial evidence for the alternative; a criterion which is comparable to the 5% threshold in null-hypothesis significance testing in terms of false-positives and false-alarm rates (Dienes, 2014). Conversely, a BF smaller than one third can be considered evidence for the null hypothesis. The Bayesian ANOVA for the data with onset adaptors with the predictors *step* and *adaptor* (/r/ vs. /l/) revealed evidence for a main effect of step (BF > 1000) and evidence for a null effect of adaptor (BF = 0.207), both compared to the model with only a participant effect. For the interaction, we compared the model with both main effects with the model with the two-way interaction (Rouder, Morey, Verhagen, Swagman, & Wagenmakers, 2017), which also supported the hypothesis that the null effect of adaptor is stable across the continuum (BF = 0.248).[5]

Given that Bowers et al. (2016) indicated that phonemic effects may be one third of allophonic effects, we can make an even more specific test using the Bayesian approach proposed by Dienes (2014). This approach allows for a more explicit specification of the alternative hypothesis in cases where there is an expected effect size. In the current case, the expected effect in the onset adaptor condition with only phonemic overlap is one third of the effect size observed in the offset-adaptor condition with phonemic and allophonic overlap. In the case of such a prior, Dienes (2014) suggests the following alternative hypothesis. The population mean is to be found in a normal distribution with a mean of the expected effect size (in our case 0.614 logit units, 1/3 of the allophonic effect of 1.89 logit units averaged over the three steps) and a standard deviation of half that mean. With this prior, population means are most likely at the expected mean and unlikely around zero. We used these priors with the Bayes calculator provided by Dienes in its R implementation (Baguley & Kaye, 2010). The observed effect in the onset-adaptor condition was 0.07 logit units (SE = 0.217), which leads to a BF of 0.208, and hence is evidence for the null hypothesis.

We also used the power.t.test function (in R) to estimate the likelihood of finding a significant effect in the onset condition. The function was fed the expected mean difference (1/3 of the effect in the offset condition) and the observed standard deviation in the onset condition. This yields a power estimate of 0.87, that is, that there is an 87% chance of finding a significant effect if there is one. For power, usually a threshold of 0.8 is deemed acceptable (Cohen, 1992).

[5] Despite the fact that the interaction between adaptor and continuum was not significant and that the null effect for the interaction was supported by the Bayesian analysis, one could still wonder whether there was a phonemic effect for only the middle step of the continuum. A t-test on the phonemic contrast for this step alone was not significant ($t(23) = 1.54$, $p = .142$) and a Bayesian t-test indicated no preference for the alternative over the null hypothesis (BF = 0.590). Note that we report this, at the request of the editor and one reviewer, as an exploratory analysis; it is not part of our confirmatory hypothesis testing.

The results hence show a clear selective-adaptation effect for offset adaptors that share allophones with the test stimuli, but no effect for the onset adaptors that overlap only in phonemic identity. These results contrast with those of Bowers et al. (2016), who found that selective adaptation occurred across syllable positions if there was phonemic overlap. It is not difficult to explain this apparent contrast; there is acoustic overlap across positions in the stimuli of Bowers et al. but not in the stimuli in the current experiment.

In Experiment 1, we were able to singly dissociate allophonic and phonemic overlap between adaptors and test stimuli. The case of Dutch /r/ and /l/ allowed us to use adaptors that overlapped in phonemic identity and allophonic identity or in phonemic identity only. In Experiment 2, we made use of a case where we could doubly dissociate phonemic and allophonic overlap.

Experiment 2

Experiment 2 was designed to provide another, even stronger test of the potential role of allophones and phonemes is speech recognition by manipulating phonemic and allophonic overlap between adaptors and test stimuli independently. In Experiment 1, we had one condition in which there was phonemic and allophonic overlap and another in which there was phonemic but no allophonic overlap. For a completely crossed design, two conditions are missing; one with allophonic but no phonemic overlap and one with no allophonic and no phonemic overlap. The last condition is easy to implement; any list of adaptors that have no overlap with the critical segments in the test stimuli will do. The more interesting case is the one where adaptors have allophonic overlap with the test stimuli, but no phonemic overlap.

This can be achieved using German back fricatives, which have a well-defined allophony in Standard German[6]. Standard German has five voiceless fricatives: A labiodental /f/, an alveolar /s/, a postalveolar /ʃ/, a glottal /h/, and a back fricative which surfaces as palatal [ç] after front vowels but as velar [x] after back vowels (Weber, 2001). Their allophonic status is reinforced by morphological alterations. The German word for *book*, *Buch*, surfaces with the velar fricative in the singular [bu:x] but with the palatal fricative in the diminutive *Büchlein* [by:çlɑin] and the plural *Bücher* [by:çəɐ]. As noted in the introduction, these kinds of morphological alterations are part of the raison d'etre for phonemes as proposed by Bowers et al. (2016).

[6] This allophony in fact occurs in most varieties of German. The exceptions are Allemanic varieties, for example, most varieties spoken in Switzerland. They only use [x].

Additionally, the phone [ç] can arise not only as an allophone of the back fricative but also as an allophone of the phoneme /g/, the voiced velar stop. This is most common in words ending in *-ig* (e.g. /køːnɪg/ Engl., *king*). There is a regional tendency that such words are produced with word-final devoicing, leading to a [k] (e.g., [køːnɪk]) by southern German speakers but with a palatal fricative [ç] (e.g., [køːnɪç]) by northern and standard German speakers (see, e.g., Schuppler, Adda-Decker, & Morales-Cordovilla, 2014). The palatal variant is also the one most frequently used in German TV newscasts.[7] This variation is relatively unmarked (see, Mitterer & Müsseler, 2013, for some examples of processing consequences of this variation). Importantly, competent native speakers need to treat the critical sound as the underlying phoneme /g/, as the /g/ surfaces in morphologically related forms such as the plural. To provide an example, *König* is produced by some speakers as [køːnik] and by others as [køːniç]. Nevertheless, the plural has to be [køːnigə] for all speakers. This shows that the underlying phoneme is /g/.

Our test stimuli contained the palatal allophone /ç/ arising out of the back fricative, which was contrasted with the postalveolar fricative /ʃ/ in the minimal pair *Kirche-Kirsche* ([kɪɐçə] - [kɪɐʃə]; Engl., *church-cherry*). Note that the palatal fricative /ç/ is the allophone of the back fricative that is articulatory and acoustically closer to the postalveolar /ʃ/, so that a continuum between the two fricatives did not give rise to stimuli that would resemble the velar allophone [x]. The different adaptor conditions only overlapped with the word [kɪɐçə], containing the palatal fricative, and never overlapped with the other endpoint of the test continuum [kɪɐʃə]. That is, in contrast to Experiment 1, only one endpoint of the test continuum was involved in potential selective adaptation.

With these test stimuli, four types of adaptor condition were created: First, adaptors with the back fricative after the front vowel [ɪ] (e.g., *Gericht, dicht, friedlich*, … ; [gəʀɪçt], [dɪçt], [fridlɪç]; Engl., *court, closed, peaceful*, ...), lead to a condition with phonemic and allophonic overlap with the test stimulus [kɪɐ^{ç}/$_{ʃ}$ə]. Second, for a condition with phonemic but no allophonic overlap, adaptors containing the back fricative following the vowel /a/ were selected (e.g., *Fracht*, *einfach*, *Verdacht,* …; [fraxt], [aɪnfax], [fəɐdaxt], …; Engl., *freight, simple, suspicion, …*). Third, a condition was created that could not be tested in Experiment 1: a condition with allophonic but not phonemic overlap. In German, words ending in *–ig* can be either produced as [ɪk] (including the typical German final obstruent devoicing) or as [ɪç] (e.g., *König, wenig*, …, Engl. *king, little*).

[7] We verified this by sampling the pronunciation by the seven German newscasters of the main edition of the *Tagesschau*, the most recognized German news show, who all used [ɪç].

In this way, we created four conditions that arise by crossing ±phonemic overlap and ±allophonic overlap with the fricative in [kɪɐçə]. All lists of adaptors also did not contain words with the postalveolar fricative [ʃ], which is hence not subject to any selective adaptation in any of the four conditions. Table 1 provides an overview of the conditions in orthographic form, underlying form in IPA, and surface form in IPA.

The predictions of the two theoretical accounts are as follows. The phoneme-plus-allophones model, as illustrated in the left panel of Figure 1, predicts that allophonic and phonemic overlap should both lead to selective adaptation, that is, fewer /ç/-responses after adaptors that contain /ç/ either as an allophone or a phoneme. The allophones-only model (cf. the right panel of Figure 1) predicts adaptation (i.e., fewer /ç/ responses after /ç/ adaptation) only for the allophonic-overlap factor.

Table 1.

Adaptor overlap conditions in Experiment 2

Adaptor Overlap	Orthography	Underlying form	Surface form
+ phonemic + allophonic	friedlich (peaceful)	/fridlɪç/	[fridlɪç]
+ phonemic - allophonic	flach (shallow)	/flaç/	[flax]
- phonemic + allophonic	König (King)	/kø:nɪg/	[kø:nɪç]
- phonemic - allophonic	Auge (eye)	/aʊgə/	[aʊgə]

Note: Adaptors did or did not have phonemic and allophonic overlap with the palatal fricative in the test stimulus continuum [kiɐçə]- [kiɐʃə] *Kirche-Kirsche*, Engl., *church - cherry*), see also Figure 4.

Method

Participants

Twenty-eight native speakers of German participated in the study. They were students at the University of Munich. There were aged between 19 and 28 (18 female) and reported no language, speech or hearing impairments. Participants received a small monetary compensation for their participation.

Materials

We selected 22 words for each adaptor condition. Six words in each condition were monosyllabic and the others were disyllabic. They are listed in the Appendix. These 88 words plus the critical minimal pair *Kirche-Kirsche* were recorded by a female native speaker of German, who naturally produced words ending on *-ig* as [ɪç]. Note that although there was positional variation of [ç] across the adaptors (i.e., some were at the onset of coda clusters, as in *Licht*, and some were the entire coda, as in *üblich*) and it was in onset position in the test stimulus *Kirche*, the frication noise was acoustically matched across positions (see example in Figure 4). That is, across positions, it was always the same allophone. There was likewise no allophonic variation over position in the adaptors with [x].

For the critical minimal word pair, we selected two recordings with a similar duration and f0 contour and generated an 11-step continuum using the STRAIGHT morphing algorithm (Kawahara et al., 1999) using 10% steps. Based on the impression of the potentially ambiguous range of the continuum, a short pretest with steps one to five of the continuum (stimulus zero being based on the original *Kirche* recording) was performed with five native speakers of German. As in Experiment 1, pictures were used as response options, and participants were asked to press “0” or “1” depending on whether the utterance better matched the right or left utterance respectively. The procedure was otherwise the same as in Experiment 1. Based on the result of the pretest, we used steps one, two, and three (65.%7, 19.2%, and 6.8% *Kirche* responses, respectively) of the eleven-step continuum. Figure 4 shows the [ç] endpoint of the test continuum and the three adaptors with some form of overlap (see the figure caption for further explanation).

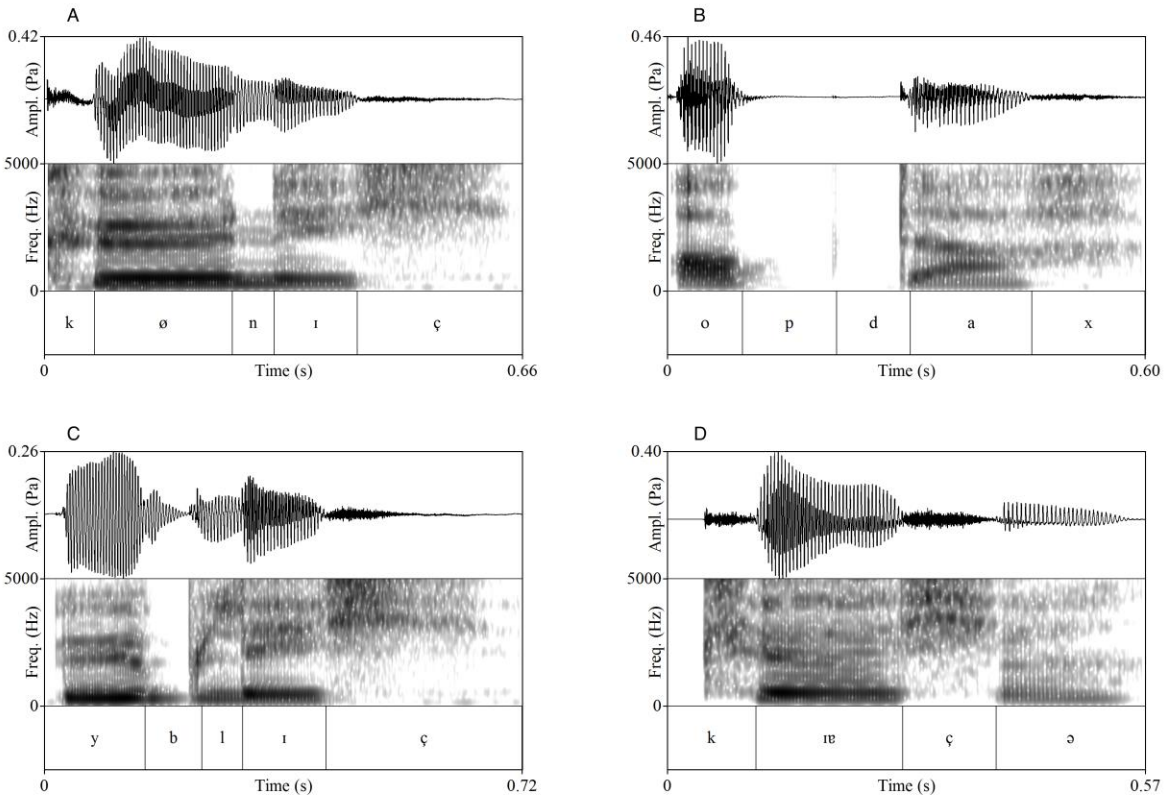

Figure 4. Test word (Panel D: the German word *Kirche* with the phoneme /ç/ and allophone [ç]) and adaptors from Experiment 2. Panel A shows the German word *König* /kø:nɪg/ → [kø:nɪç] that has no phonemic overlap but allophonic overlap with the critical phone [ç] in *Kirche.* Panel B shows the German word *Obdach* /opdaç/ → [opdax], which has phonemic overlap but no allophonic overlap with the critical phone [ç] in *Kirche*. Panel C shows the German word *üblich* /ybliç/ → [ybliç], which has both phonemic and allophonic overlap with the critical phone [ç] in *Kirche*.

Procedure

The experiment started with a short production task implemented in Speechrecorder (Draxler & Jänsch, 2004), in which participants were presented with image and sentence prompts to generate words ending on the fricatives [ç] and [ʃ]. The main reason for this production task was to potentially exclude speakers that do not distinguish [ç] and [ʃ], which is the case for some German speakers who speak certain varieties from central Germany. Additionally, we also included some items ending on /ɪg/ to see whether our participants produced those as [ɪç] or [ɪk]. Note that this was to simply record these data rather than use it as an exclusion criterion. After this short production task with nine items, the main experiment started.

The procedure of the main experiment was similar to Experiment 1. The experiment was again divided into four blocks, one block for each adaptor. One block consisted of eight sub-blocks in which

participants were presented with the 22 adaptor words for that block twice in a random order, presented with a 600ms inter-stimulus interval (ISI). This lasted about 50 seconds and was followed by nine phonetic-identification trials, in which each of the three stimuli was presented three times. The phonetic-identification trials had the same structure and format as in the pretest (see Materials section). We used all 24 possible orders of the four blocks over 24 participants. That is, each adaptor type occurred equally often in the same position and was preceded and/or followed by all other adaptor types.

Results and Discussion

In the production task, all speakers produced a clear contrast between /ç/ and /ʃ/. Given the location of testing in Southern Germany, there was a preference to produce words with an underlying /ɪg/ as [ɪk]. Only one participant out of twenty-four used [iç] throughout, and a few produced one of the three items in the production task that ended underlyingly with /ɪg/ with [ɪç].

The data from four participants were rejected because they failed to hear a difference between the stimuli and heard all of them as [kɪɐʃə]. When this was noted, another participant was assigned that particular order of adaptation blocks. For the 24 participants who entered the analysis, we calculated the mean proportion of [kɪɐçə]-responses for each of the four adaptor conditions, which are presented in Figure 5. The figure shows a clear effect of allophonic identity, with more [kɪɐçə] responses if the adaptor did not contain the allophone [ç] and no effect of phonemic identity.

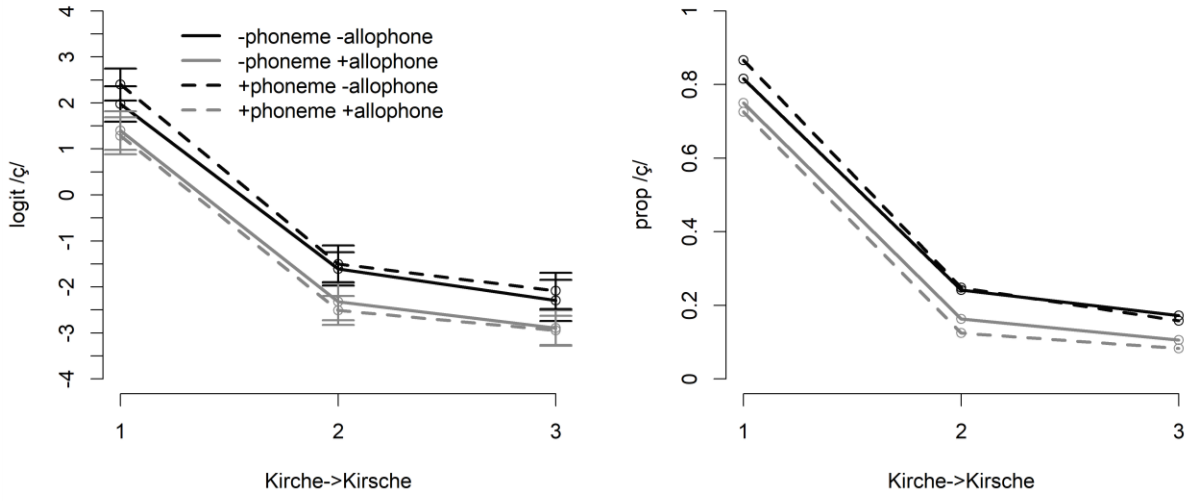

Figure 5. Mean log odds and proportions of [kɪɐçə] responses to the three steps on the *Kirche-Kirsche* continuum in Experiment 2 in the four conditions that arise by crossing ±phonemic overlap and ±allophonic overlap. Error bars show standard errors based on Morey's (2008) method for repeated-measures designs.

The log odds data were analysed with a repeated-measures ANOVA using the aov function in R. The results are displayed in Table 2, alongside a table for a Bayesian ANOVA based on the R package BayesFactor (Morey, Rouder, & Jamil, 2015) using its default priors for the function anovaBF, which provides BFs for model comparisons. The repeated-measures ANOVA shows a significant effect of allophonic overlap but no effect for phonemic overlap and no interaction between allophonic and phonemic overlap. The Bayesian analysis shows that adding the predictor *Phoneme* to a model with only participant as random factor does not lead to a better prediction, in fact the BF of 0.134 suggests evidence for the model without *Phoneme*. Conversely, there is evidence for an effect of *Allophone*, with a BF of 6.174. For the interactions, the BFs are "sequential", comparing the model with a given interaction against the model with all "lower" effects (Rouder et al., 2017). These model comparisons show that there is substantial evidence that allophonic overlap influences selective adaptation independent of all other factors and also suggests that the overall null effect of phonemic overlap is stable across the continuum. That is, the Bayesian analysis converges with the ANOVA, showing evidence for the alternative hypothesis where the ANOVA showed a significant effect and reinforcing the null effect for effects that were not significant in the ANOVA.

Table 2:

Outcome of the statistical analyses of the data obtained in Experiment 2.

	Type of Analysis				
	ANOVA				BayesFactor
Predictor	df	ε	F	p	
phonemic Overlap	(1,23)	na	0.113	0.740	0.134
allophonic Overlap	(1,23)	na	12.059	0.002	6.174
step	(2,46)	0.653	134.072	<0.001	> 1000
step x phonemic Overlap	(2,46)	0.732	0.280	0.757	0.072
step x allophonic Overlap	(2,46)	0.670	0.091	0.913	0.072
phonemic Overlap x allophonic Overlap	(1,23)	na	0.727	0.403	0.316
step x phonemic Overlap x allophonic Overlap	(2,46)	0.693	0.164	0.849	0.114

Note. The column labelled ε showed the Greenhouse Geisser correction for repeated measure ANOVAs. The Bayes Factors represent model comparisons with the model containing all "lower" effects, that is, the main effects are compared to a model with just a participant effect, two-way interactions are compared with models comparing all main effects, and the model with three-way interaction is compared to the model containing all main effects and two-way interactions.

As in Experiment 1, we also tested the specific hypothesis that the effect of phonemic overlap is one third of the effect caused by allophonic overlap using the approach proposed by Dienes (2014) using the R implementation provided by Baguley and Kaye (2010). We calculated the mean allophonic effect (0.81 logit units) and then tested the null hypothesis against the hypothesis that the phoneme effect (observed = -0.0645, SE = 0.192) is one third of the allophone effect. The alternative hypothesis was hence specified as follows: The allophonic effect is to be found in a normal distribution with a mean of the expected effect size (=0.81/3 = 0.27 logit units) and a standard deviation of one half of the expected effect. This led to a Bayes Factor of 0.313, again providing evidence for the null hypothesis. We also did a power analysis analogous to the one performed for Experiment 1, which yielded a power of 0.39.

The reason that the Bayesian analysis and the power analysis seem to diverge is that the Bayesian analysis takes into account that the observed phoneme effect was slightly in the opposite direction, but this information is not used in the power analysis. Even though the experiment by itself is hence underpowered, it still lowers the overall likelihood of having missed an effect. After Experiment 1, the likelihood of having missed an effect was 13%, over the two experiments, this likelihood is now below 8% (i.e., (1 – Power[exp1]) * (1- Power[exp2])).

It should be noted that the overlap was not maximal between exposure and test, since the fricatives in the test blocks were syllable-initial but those in the exposure blocks were syllable-final. Previous work indicated that syllable position may not matter for voiceless fricatives (Jesse & McQueen, 2011). Even so, vowel context may still play a role, especially in the current case. German /ʃ/ requires lip rounding, and lip rounding is known to lead to strong contextual effects (Smits, 2001). It is hence possible that the decision whether *Kirche* or *Kirsche* is heard may depend on the relative activation of the fricatives [ç] and [ʃ] as well as on the relative activation of the potential allophones for the vowels, $[i]_{roundedContext}$ versus $[i]_{unroundedContext}$ and $[ə]_{roundedContext}$ versus $[ə]_{unroundedContex}$. Our adaptors only contained the allophones for the preceding vowel and the fricative itself, but not the following schwa. Cues carried by the schwa in the test stimuli therefore could not be adapted during exposure. Given this absence of fully-overlapping conditions, the estimate that the phoneme effect should be one third of the allophone effect is a rather conservative one, because the phonemic effect should be one third of the maximal possible effect in the allophone condition. But this maximal effect was not achieved here. We therefore replicated Experiment 2 with some changes to increase the size and reliability of the allophonic selective-adaptation effect and to maximize the chances of observing a phonemic selective-adaptation effect.

Experiment 3

In this experiment, there were three changes in comparison to Experiment 2. In Experiment 2, participants heard three repetitions of the three test stimuli after a block of adaptors. We examined the strength of the adaptation effect over the three repetitions of the three test stimuli and found that the adaptation effect dissipated over repetitions and was nearly absent in the last repetition. Hence, in Experiment 3, we reduced the number of test stimulus repetitions after an adaptor block from three to two. To compensate for the loss of overall number of observations per stimulus, the number of adaptor-test repetitions in each bock was increased from eight to ten.

Second, to increase the overlap between test and adaptor stimuli, we recorded a new pair of test stimuli, in which the morphemes *Kirsche* (Engl., 'cherry') and *Kirche* (Engl.,'church') appeared in compounds (*Kirschbaum* Engl., 'cherry tree' and *Kirchplatz* Engl., 'church square'). In these compounds, the fricatives /ç/ and /ʃ/ appear in syllable-final position, just as in the adaptor stimuli. During the test phase, participants heard only the first syllable and were asked which compound they thought the speaker intended.

Finally, we also included a pretest for two reasons. In Experiment 2, participants did not perceive any of the stimuli near the 50% mark. At the margins of a proportional scale, perceptual differences are more difficult to measure; they require more repetitions to be measured accurately and are more susceptible to lapses of attention (cf. Macmillan & Creelman, 1991). The pre-test allowed us to select a range of stimuli that on a by-participant basis were not perceived near floor or ceiling, and hence increased the probability of being able to detect a small effect. Secondly, we also observed in the previous experiments that participants sometimes showed large differences between the first and second block of testing, probably because they had not yet "homed in" on the test continuum (see Repp & Liberman, 1987). By using a pre-test, we already familiarized the participants with the continuum.

Participants

Thirty-five native speakers of German participated in the study. They were students at the University of Munich. There were aged between 19 and 28 (18 female) and reported no language, speech or hearing impairments. Participants received a small monetary compensation for their participation.

Materials

The same adaptor stimuli were used as in Experiment 2. For the test stimuli, the same speaker was recorded saying *Kirchplatz*/kiɐçplats/ and *Kirschbaum* /kiɐʃbaʊm/. The first syllables from both utterances were spliced out and morphed using Straight (Kawahara et al., 1999) using steps of 4%.

Informal listening tests suggested that the 50% point was around step 10, so the stimuli 6, 8, 9, 10, 11, 12, and 14 (with morphing ratios from 24% [ç]- 76% [ʃ] to from 56% [ç]- 44% [ʃ]) were used for the pre-test.

Procedure

Each participant first performed a pre-test in which the nine selected stimuli from the continuum were presented ten times each. After that, a logistic-regression model was applied to the data to determine which step was closest to 50% for that participant and what step size was required so that the participant would hear stimuli that would lead to at least a 25% difference in identification scores. That is, if the step closest to 50% for a given participant was step 11, the predicted responses for steps 10 and 12 (= step size 1) were determined and it was evaluated whether there was at least a 25% difference in [ç] identifications predicted between these two. If that was the case, these three steps were used for the test phase. If not, it was iteratively tested which larger step size was sufficient to achieve that (see Results for details). The 25% criterion was used so that participants would be able to hear a difference between the stimuli in the test phases. The three stimuli determined in this way for each participant separately were then used during the test phase in the main experiment for that participant.

After the pre-test, the main experiment was started with the four adaptor conditions used in Experiment 2 (*-ig*, *-ich*, *-ach*, *control*; see Table 1). As previously, the order of these blocks was counterbalanced over participants so that each set of consecutive four participants would be in a Latin Square with each adaptor block occurring in each position. All possible permutations occurred over 24 participants. Within each block, there were 10 repetitions of adaptor stimuli followed by six test trials (two repetitions of the three test stimuli determined by the pre-test).

Results

One participant showed a perfect separation in the pre-test data, with each stimulus being perceived consistently as either [kiɐç] or [kiɐʃ]. Consequently, the logistic regression model to determine the steps for the main experiment did not converge. This participant was replaced. For the remaining thirty-four participants, the most ambiguous step ranged from step six to step thirteen (M = 10.09, SD = 1.54). For 31 participants, a step size of one sufficed to get a predicted .25 difference in identification proportions between the two outermost stimuli; two participants required a step size of 2, and one required a step size of 3.

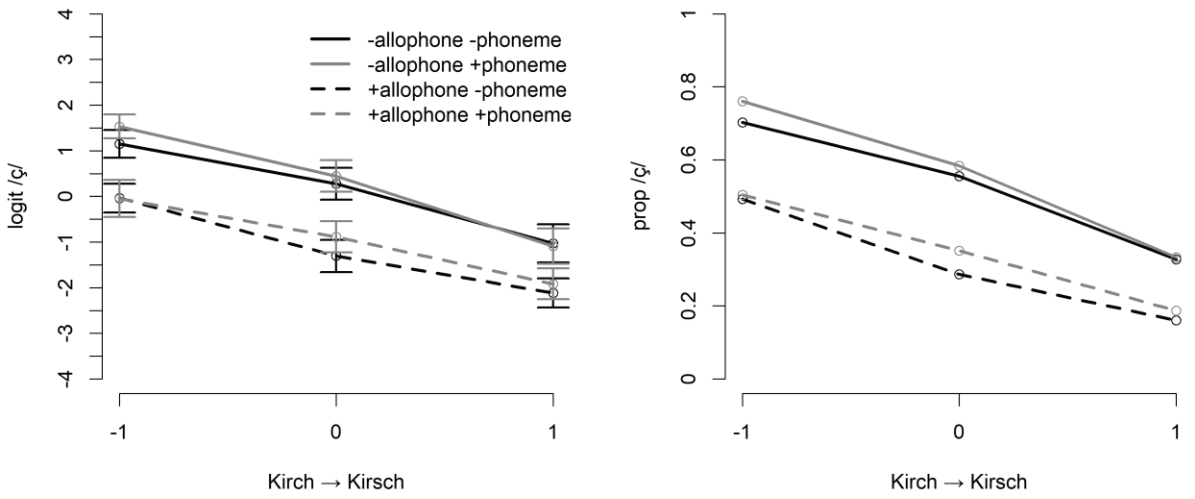

Figure 6. Mean log odds and proportions of [kɪɐç] responses to the three steps on the *Kirch-Kirsch* continuum in Experiment 3 in the four conditions that arise by crossing ±phonemic overlap and ±allophonic overlap. Note that the continuum varied across participants: Step 0 was selected in the pre-test as the most ambiguous step for each participant (see main text for details). Error bars show standard errors based on Morey's (2008) method for repeated-measures designs.

Figure 6 shows the mean proportions and logOdds of [ç] identifications for the three stimuli. As in Experiment 2, there is a clear allophonic selective adaptation effect with fewer [ç] responses if the stimuli contain the allophone [ç], while the conditions with a /ç/ as phoneme led, if anything, to more [ç] responses than the respective comparison conditions without a /ç/-phoneme. Moreover, the allophonic selective adaptation effect is stronger here than in Experiment 2.

These observations are borne out by the statistical analysis, which shows a clear effect of step and allophonic overlap and a weak significant effect of the step by allophonic overlap interaction and a trend towards an effect of phonemic overlap (see Table 3). Interestingly, the Bayesian analysis prefers the null hypothesis for all these latter effects, because the effects are numerically rather small. Note also that the trend towards a phonemic effect is in the direction opposite to that predicted by a model with phonemic representations. As shown in Figure 6, participants labelled more stimuli as [kɪɐç] in the +phoneme than in the –phoneme adaptation conditions; the opposite of what one would expect if there was selective adaptation of a phonemic representation.

Table 3

Outcome of the statistical analyses of the data obtained in Experiment 3.

	Type of Analysis				
	ANOVA				BayesFactor
Predictor	df	ε	F	p	
phonemic Overlap	(1,31)	na	3.043	0.090	0.174
allophonic Overlap	(1,31)	na	46.869	<0.001	> 1000
step	(2,62)	0.653	19.257	<0.001	> 1000
step x phonemic Overlap	(2,62)	0.904	0.858	0.426	0.061
step x allophonic Overlap	(2,62)	0.918	4.029	0.025	0.138
phonemic Overlap x allophonic Overlap	(1,31)	na	0.031	0.957	0.153
step x phonemic Overlap x allophonic Overlap	(2,62)	0.973	2.191	0.121	0.151

Note. The column labelled ε showed the Greenhouse Geisser correction for repeated measure ANOVAs. The Bayes Factors represent model comparisons with the model containing all "lower" effects, that is, the main effects are compared to a model with just a participant effect, two-way interactions are compared with models comparing all main effects, and the model with the three-way interaction is compared to the model containing all main effects and two-way interactions.

As in Experiment 1 and 2, we also tested the specific hypothesis that the effect of phonemic overlap is one third of the effect caused by allophonic overlap using the approach proposed by Dienes (2014) using the R implementation provided by Baguley and Kaye (2010). We calculated the mean allophonic effect (1.26 logit units) and then tested the null hypothesis against the hypothesis that the phoneme effect (observed = -0.184, SE = 0.105) is one third of the allophone effect. The alternative hypothesis was hence specified as follows: The allophonic effect is to be found in a normal distribution with a mean of the expected effect size (0.44 logit units) and a standard deviation of one half of the expected effect. This led to a Bayes Factor of 0.075, again providing evidence for the null hypothesis, however, this time with a value that is considered not only substantial but strong evidence for the null. Overall, we observed Bayes Factors of 0.21 in Experiment 1, 0.31 in Experiment 2, and 0.075 in Experiment 3. This means that the null hypothesis given this series of experiments is 204 times more likely than the hypothesis that there is a phonemic effect that is a third of the allophonic effect.

We also performed a power analysis for finding a phonemic effect based on the observed data, which revealed a power of .989. This large difference in power between Experiments 2 and 3 is due to a larger expected effect size in Experiment 3 (because the allophonic effect is larger), a smaller standard deviation (despite fewer trials per step, possibly because the pretest "homed in" participants on the

continuum), and a larger sample size (34 versus 24 participants). Over the three experiments, the likelihood of having missed an effect is therefore $8*10^{-4}$ (based on multiplying [1-power] over the three experiments), that is, a likelihood below one tenth of a percent. Note that these estimates are still conservative, since the overlap in the allophone conditions in Experiment 2 was not perfect.

General Discussion

Three experiments investigated the apparent conflict between results from the selective-adaptation paradigm and the perceptual-learning paradigm regarding the nature of pre-lexical representations in spoken-word recognition. It has been argued that perceptual learning only reveals allophonic units (Mitterer et al, 2013) while selective adaptation reveals additional phonemic units (Bowers et al., 2016). The evidence for phonemic units, however, was ambiguous due to the use of stimuli with acoustic overlap between adaptors and test stimuli. With better control over the acoustic differences between adaptors and test stimuli, we found no evidence for the involvement of phonemic units in spoken-word recognition.

Previous work using the perceptual-learning paradigm had already suggested that phonemes do not play a role in pre-lexical processing. The first efforts with this paradigm (Mitterer, Scharenborg, & McQueen, 2013; Reinisch et al., 2014) asked whether sharing phonemic identity is sufficient for learning to generalize, generally with negative results. A more recent paper (Mitterer et al., 2016) asked whether a phonemic difference leads to inhibition of generalization by testing generalization to the same allophone with a different underlying specification. This was possible by testing whether learning generalizes from Korean tensified stops, which are underlyingly lax, to underlying tense stops. Learning generalized and was not reduced in comparison to tensified stops, which shared both phonemic and allophonic identity with the exposure items. These previous studies, together with the present results from selective adaptation hence weaken the assumption that context-invariant phonemes have a role to play in spoken-word recognition.

Perceptual learning versus selective adaptation

The current data raise the question which method may be better suited to reveal the units used in pre-lexical processing: selective adaptation or perceptual learning. One possible answer to this may be that they are equally good. Empirically, the combined data show that with clearly defined differences between exposure and test items (as in the case of Dutch /l/ and /r/), no generalization across allophones is found.

The selective-adaptation paradigm was once considered to be pivotal in identifying the units that are involved in pre-lexical processing (Eimas & Corbit, 1973). But its popularity declined quickly due to findings that selective adaptation does not exclusively reveal feature detectors (Kleinschmidt & Jaeger, 2015; Remez, 1987). In this context, it is important to note that Samuel and Kat (1996) argued that selective adaptation arises at three levels: recognition of simple acoustic patterns (see also Holt, 2005, 2006), recognition of complex acoustic patterns, and recognition of phonetic categories. The apparently phonemic adaptation with stop consonants observed by Bowers et al. (2016) may hence occur at the level of formant tracking or at a level at which formant loci (Sussman et al., 1998) are calculated. Formant loci are assumed to be an important acoustic abstraction for the recognition of place in stop consonants. They are assumed to be invariant over vowel contexts, and hence may be subject to selective adaptation. That is, selective adaptation may occur at auditory levels of processing that precede phonetic categories. It follows that selective adaptation from one set of words to another does not necessarily imply that the two share a phonemic representation; the adaptation may be due to shared auditory patterns across the two sets of stimuli. There are in fact two alternatives that do not make reference to abstract phonemes that provide an explanation why Bowers et al. (2016) found selective adaptation across positions with stop consonants. First, on the level of recognizing simple acoustic patterns, shared burst spectra of the different types of stimuli may lead to adaptation effects. Second, on the level of extracting complex acoustic patterns, formant loci may provide a basis to explain why adaptation can occur across positions.

Clearly, the same concerns about the locus of the adaptation effect also apply to the current data. It is possible that the effects we have attributed to allophonic overlap could reflect overlap of auditory patterns. This suggests that, at a methodological level, the selective adaptation paradigm is not the ideal means to explore the units of speech perception. At a theoretical level, it is important to emphasize that, irrespective of whether the adaptation process itself concerns changes in the representations of allophones, changes of the auditory patterns that define those allophones, or changes in the mapping between these two types of representations, the present data indicate that it is knowledge about allophones, not about phonemes, that is involved.

Selective adaptation may also occur at levels beyond phonetic processing, that is, at a response level. Remez (1979) showed that selective adaptation occurs for a speech versus non-speech decision. Given that it is unlikely that this is based on a “nonspeech” unit in perceptual processing, it suggests that there may be selective adaptation at the response level as well. That is, the adaptors may well be analysed in terms of the possible response options.

Given these issues with the selective adaptation paradigm, one can ask whether the perceptual-learning paradigm fares any better. The paradigm has two merits. First of all, it is based on processes which are involved in solving the critical problem in spoken-word recognition, the invariance problem. The paradigm thus reveals units that are functional in speech perception. That is, it reveals units that are involved in active adaptation to variance in the input and that hence help the listener decode the highly variable speech signal. Secondly, the eye-tracking evidence (Mitterer & Reinisch, 2013) shows that perceptual learning influences processing at an early, pre-lexical level.

As argued by Bowers et al. (2016), however, one recent result suggests that the perceptual-learning paradigm may sometimes be very specific. Keetels, Pecoraro and Vroomen (2015) showed that perceptual-learning is "ear-specific". However, a recent paper (Keetels, Stekelenburg, & Vroomen, 2016) has nuanced that conclusion by showing that learning effects are not necessarily ear-specific but instead are moderated by spatial location. Keetels et al. (2016) tested generalization of perceptual learning across spatial location with loudspeakers and ear of presentation with headphones and found that learning generalizes to a new location, and is only reduced in its effect size. Moreover, these moderating effects were obtained with a visual recalibration paradigm, which may invoke processing that is more stimulus-specific than the lexically-driven perceptual learning paradigm (i.e., that of Norris et al., 2003). Learning from visual context seems to dissipate quite quickly (Reinisch & Mitterer, 2016), while learning based on lexical knowledge is quite long lived (Eisner & McQueen, 2006; Kraljic & Samuel, 2005). Learning has been observed to persist over twelve hours, independent of whether this involved sleep, and it seems highly unlikely that spatial constraints would survive such a long time-lag. Nevertheless, this indicates that findings with the perceptual-learning paradigm should also be taken with a grain of salt, with the critical question being whether learning is specific to the experimental situation.

With respect to the two paradigms, we can hence conclude that while neither is the perfect method, the perceptual-learning paradigm may have greater utility in the study of the units of speech perception. In particular, it appears to more precisely target pre-lexical processing. Nevertheless, the current selective adaptation results do converge with the perceptual learning results in suggesting that phonemes are not pre-lexical units in speech perception.

Why or why not phonemes

If one rejects phonemes as categories in pre-lexical processing, this begs the question of how to deal with the evidence and arguments in favour of phonemes (e.g., Bowers et al., 2016). One recent experiment (Toscano, Anderson, & McMurray, 2013) showed that phonological anadromes (e.g., *bus* vs. *sub*) seem to compete more with each other than words that only overlap in the vowel. This might indicate

that position-invariant phonemes are extracted from the input. However, on closer inspection, this result is not that strong, as it is only significant in a linear-mixed effects model with a random-intercept structure, which is anti-conservative (Barr, Levy, Scheepers, & Tily, 2013). The fact that the effect disappears when a random slope over items is included means that the effect is highly variable over items, which makes sense if some segments are position-independent (such as voiceless fricatives) while others are highly position-dependent (such as liquids and some aspects of stops). This particular result hence does not provide clear support for context- and position-independent phonemes.

As already touched upon in the introduction, Bowers et al. (2016) argue that there are strong linguistic arguments for phonemes, because morphological alterations require knowledge about the underlying phonemes. Abstract phonemes are then considered useful, so that it is easier to recognize the morpheme *zwaar* in different forms with different allophones (e.g., [zʊarə] and [zʊaɹ]). Abstracting to phonemes may be especially useful when listeners encounter a relatively low-frequency form of a morpheme that is unlikely to be stored in the mental lexicon (e.g., *a reddish hue*[8], in which the morpheme-final /d/ from *red* is likely to be flapped in American English). If the mental lexicon only stored the form [ɹɛd], the listener would be in a difficult position to retrieve it from the input [ɹɛɾɪʃ]. Such an argument is supported by evidence that listeners analyse the input in terms of the morphemes that it carries (Wurm, 1997). This argument for abstract phonemes in spoken-word recognition is less forceful, however, if multiple lexical representations exist for a given morpheme (Bürki & Gaskell, 2012; Connine, Ranbom, & Patterson, 2008). If the mental lexicon contains [ɹɛr] and [ɹɛd], words newly created from existing morphemes can be recognized easily, without the need for abstract phonemes in pre-lexical processing.

Another argument that has been brought up in support of phonemes is that phonemes may be important in production. Again, this is evident in morphological alterations, in which the correct allophones must be chosen for a given form, and speakers need to know the correct underlying form. The assumption that perception and production must be similar is questionable, however. Models such as functional phonology (Boersma, 1998) assume that the two speak two completely different languages[9]. Perception also needs to deal with many variants that are not used in production. Dutch speakers, for instance, do not only differ in the extent to which postvocalic /r/ is produced as an approximant, they also differ in what form of trill they are using for /r/ in onset position. A given speaker hence needs to recognize

[8] We do not make the claim here that the specific form *reddish* is not stored in the mental lexicon. However, we assume that listeners will sometimes be presented with inflected forms they have never heard before, especially as speakers tend to be productive and generate new words (see, e.g., Wurm, 1997, and examples such as *mid-song*).

[9] In this particular constraint-based model, perception constraints look like "349 Hz not /i/" while production constraints look like "Gesture: (jaw half-open)".

both [rot] and [ʁot] as variant of the Dutch word for *red*, even though a given speaker is unlikely to use both variants (Mitterer & Ernestus, 2008). This suggests that a tight coupling of perception and production is not particularly useful, and hence that the segments used for spoken-word recognition may be different from those used in production.

Other potential units in pre-lexical processing

We have noted that Bowers et al. (2016) argued that there are linguistic arguments for the phoneme. This turns out to be a controversial claim by itself. Most linguistic theories follow Chomsky and Halle (1968), who argued against phonemes, and do not assume that phonemes play any significant role other than being a useful device for language description. Indeed, Embick and Poeppel (2014) argued that psychology at its own peril is overlooking the mounting linguistic evidence against phonemes and for another type of pre-lexical representation: phonological features. Featural theories assume that listeners analyse the signal in terms of bundles of articulatory or phonological features; that is, the speech signal is not only decomposed along the time axis, but also into different layers that simultaneously determine the speech signal at a given point in time (e.g., voice, manner of articulation, place of articulation). Such features are abstract, linguistic representations and thus distinct from the auditory patterns discussed earlier.

The question of segments versus phonological features has already been investigated using the perceptual-learning paradigm. In general, featural accounts predict that learning will generalize to other words containing the same feature. Successful generalization along these lines was reported by Kraljic and Samuel (2006), who tested whether learning about the voiced/voiceless boundary in stops can generalize over place of articulation. In their experiment, learning about voicing in stops generalized from alveolar stops during exposure (/d/ vs. /t/) to labial stops (/b/ vs. /p/) at test, in line with the prediction of featural decomposition. However, just as in the case of position-independence of learning for voiceless fricatives (Jesse & McQueen, 2011), this does not provide a strong test of the featural hypotheses, because the presence versus absence of aspiration is an acoustically similar cue that is present in both cases (/b/-/p/ and /d/-/t/) and that is used to signal the feature [VOICE]. This mirrors the case from selective adaptation, where seminal work showed that adaptation of voicing can also generalize across place of articulation (Eimas & Corbit, 1973). Later work, however, showed that this was due to auditory similarity, and not abstract featural similarity (Sawusch & Jusczyk, 1981).

This is also true of another case of successful generalization with perceptual learning. Learning about place of articulation generalized from tense to plain stops in Korean (Mitterer et al., 2016), which can be explained by the fact that both contained a release burst. Reinisch et al. (2014) tested a more

diagnostic case: generalization of learning of place of articulation (i.e., labial /b/ versus coronal /d/) over manner (i.e., to /m/ versus /n/) and found no generalization. This result has now been replicated twice (Mitterer et al., 2016; Reinisch & Mitterer, 2016). These results thus suggest once again that what matters in perceptual learning is auditory overlap rather than abstract featural overlap. They hence indicate that abstract phonological features are unlikely to be an important aspect of pre-lexical processing (see also Ettlinger & Johnson, 2009; Kang, Johnson, & Finley, 2016).

A question that remains is whether the units of perception are auditory patterns rather than allophones. Might it be possible to explain the patterns of selective adaptation and perceptual learning without postulating allophone-sized representations? We suggest that this is not the case. With respect to selective adaptation, we have already argued that the adaptation process might reflect effects at different levels of processing. It is thus possible that the adaptation could reflect changes to representations of auditory patterns, to representations of allophones, to the mapping between the two, or to a combination of these changes. It remains to be seen what the exact locus of the effect is. With respect to perceptual learning, the recalibration most likely reflects a change in the mapping between representations of auditory patterns and allophonic representations. A change to the auditory –pattern representation ought to lead to generalization of learning across allophones that share that pattern, contrary to what is observed (Mitterer et al., 2016; Reinisch et al., 2014; Reinisch & Mitterer, 2016). A change to allophonic representations alone is unlikely, since the recalibration concerns the acoustic patterns associated with particular allophones.

In both paradigms, however, the adjustments concern allophones, not phonemes. What seems to matter for the listener is that allophones with particular acoustic properties recur in the speech stream, and that recognising those abstract allophones, in whatever words they appear, helps in mapping the input onto the lexicon. Learning about the auditory properties of those abstractions in a given listening situation helps to optimize the word-recognition process (e.g. by applying what has been learned from some words to the recognition of other words). In this sense, it is allophones that function in spoken-word recognition irrespective of whether the adjustments in learning and adaptation paradigms are applied at the level of the allophones as a whole or (possibly in the case of selective adaptation alone) at the level of their constituent auditory patterns or (particularly for perceptual learning) in the mapping between the auditory patterns and the allophonic representations.

The fact that learning nevertheless sometimes generalizes from one allophone to another (Kraljic & Samuel, 2006; Mitterer et al., 2016) indicates that allophones are not the only type of abstraction that supports generalization of learning. It is conceivable that the grain size over which learning operates may

differ according to what constitutes replicable structures in speech production. Learning about these repeating structures is what could help the listener solve the invariance problem. Some structures may be smaller than segments, such as aspiration or the release bursts as parts of voiceless stops. This explains how learning about voicing or place can generalize to segments with other place or voicing specifications. Other structures may be larger than a segment. Poellmann, Bosker, McQueen, and Mitterer (2014) tested whether learning can also take place for syllables, by presenting listeners with a speaker who routinely reduced the Dutch prefix *ver-* [fər] to [f]. Repeated exposure to words with this reduced prefix allowed listeners to recognize new words with the same reduction more efficiently than listeners from a control group who heard the same words in an unreduced form. Together with the other findings, this shows that listeners may also learn about structures that are smaller than a segment (such as aspiration patterns), segment-sized allophones, or even larger structures, such as affixes or highly frequent word combinations such as "I do not know" (Hawkins, 2003). What may matter for perceptual learning is not the grain-size of the structure per se, but rather whether the structure is a consistent production pattern in the interlocutor's speech.

Conclusion

We have provided evidence that context-insensitive phonemes are not a part of spoken-word recognition. Three experiments showed no selective adaptation if there are clearly defined acoustic-phonetic differences between adaptors and test stimuli. We also argued that there are no compelling arguments that phonemes are particularly useful in spoken-word recognition. This, however, does not mean that competent speakers of a language can do without abstract phonemes. Abstract phonemes are necessary to generate novel word forms, but spoken-word recognition does not require phonemes pre-lexically. The present findings, along with other recent data using the perceptual-learning paradigm, suggest instead that pre-lexical processing is based on allophones. This proposal has clear implications for models of speech recognition. As summarised in the Introduction, most models assume the pre-lexical units of speech perception are not allophones; in this regard, these models may all be incorrect.

Acknowledgements

We thank Nadia Klijn for helping to prepare and test participants in Experiment 1 and Rosa Franzke for help with Experiments 2 and 3. The second author is funded by an Emmy-Noether grant (nr. RE 3047/1-1) from the German Research Council (DFG). This work was also supported by a University of Malta Research Grant to the first author.

References

Baguley, T., & Kaye, D. (2010). Book review: Understanding Psychology as a Science. *British Journal of Mathematical and Statistical Psychology*, *63*(3), 695–698.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Boersma, P. (1998). *Functional Phonology. Formalizing the interactions between articulatory and perceptual drives*. The Hague: Holland Academic Graphics.

Bowers, J. S., Kazanina, N., & Andermane, N. (2016). Spoken word identification involves accessing position invariant phoneme representations. *Journal of Memory and Language*, *87*, 71–83. https://doi.org/10.1016/j.jml.2015.11.002

Bürki, A., & Gaskell, M. G. (2012). Lexical representation of schwa words: two mackerels, but only one salami. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *38*(3), 617–631. https://doi.org/10.1037/a0026167

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York, NJ: Harper & Row.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1). Retrieved from http://psycnet.apa.org/journals/bul/112/1/155/

Connine, C. M., Ranbom, L. J., & Patterson, D. J. (2008). Processing variant forms in spoken word recognition: The role of variant frequency. *Perception & Psychophysics*, *70*, 403–411. https://doi.org/10.3758/PP.70.3.403

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Quantitative Psychology and Measurement*, *5*, 781. https://doi.org/10.3389/fpsyg.2014.00781

Dixon, P. (2008). Models of accuracy in repeated-measures design. *Journal of Memory and Language*, *59*, 447–456.

Draxler, C., & Jänsch, K. (2004). Speechrecorder - a universal platform independent multi-channel audio recording software. In *Proceedings of LREC* (pp. 559–562). Lisbon.

Dumay, N., & Content, A. (2012). Searching for syllabic coding units in speech perception. *Journal of Memory and Language*, *66*(4), 680–694. https://doi.org/10.1016/j.jml.2012.03.001

Eimas, P. D., & Corbit, J. D. (1973). Selective adaptation of linguistic feature detectors. *Perception & Psychophysics*, *4*, 99–109.

Eisner, F., & McQueen, J. M. (2006). Perceptual learning in speech: Stability over time. *The Journal of the Acoustical Society of America*, *119*(4), 1950. https://doi.org/10.1121/1.2178721

Embick, D., & Poeppel, D. (2014). Towards a computational(ist) neurobiology of language: correlational, integrated and explanatory neurolinguistics. *Language, Cognition and Neuroscience*, *0*(0), 1–10. https://doi.org/10.1080/23273798.2014.980750

Ettlinger, M., & Johnson, K. (2009). Vowel discrimination by English, French and Turkish speakers: evidence for an exemplar-based approach to speech perception. *Phonetica*, *66*(4), 222–242. https://doi.org/10.1159/000298584

Goldinger, S. D., & Azuma, T. (2003). Puzzle-solving science: the quixotic quest for units in speech perception. *Journal of Phonetics*, *31*, 305–320. https://doi.org/10.1016/S0095-4470(03)00030-5

Goldinger, S. D., Luce, P. A., & Pisoni, D. B. (1989). Priming Lexical Neighbors of Spoken Words: Effects of Competition and Inhibition. *Journal of Memory and Language*, *28*(5), 501–518.

Hawkins, S. (2003). Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of Phonetics*, *31*, 373–405. https://doi.org/10.1016/j.wocn.2003.09.006

Holt, L. L. (2005). Temporally nonadjacent nonlinguistic sounds affect speech categorization. *Psychological Science*, *16*(4), 305–312.

Holt, L. L. (2006). The mean matters: Effects of statistically defined nonspeech spectral distributions on speech categorization. *Journal of the Acoustical Society of America*, *120*(5), 2801–2817. https://doi.org/10.1121/1.2354071

Jesse, A., & McQueen, J. M. (2011). Positional effects in the lexical retuning of speech perception. *Psychonomic Bulletin & Review*, *18*, 943–950. https://doi.org/10.3758/s13423-011-0129-2

Kang, S., Johnson, K., & Finley, G. (2016). Effects of native language on compensation for coarticulation. *Speech Communication*, *77*, 84–100. https://doi.org/10.1016/j.specom.2015.12.005

Kawahara, H., Masuda-Katsuse, I., & de Cheveigné, A. (1999). Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency based F0 extraction. *Speech Communication*, *27*, 187–207. https://doi.org/10.1016/S0167-6393(98)00085-5

Keetels, M., Pecoraro, M., & Vroomen, J. (2015). Recalibration of auditory phonemes by lipread speech is ear-specific. *Cognition*, *141*, 121–126. https://doi.org/10.1016/j.cognition.2015.04.019

Keetels, M., Stekelenburg, J. J., & Vroomen, J. (2016). A spatial gradient in phonetic recalibration by lipread speech. *Journal of Phonetics*, *56*, 124–130. https://doi.org/10.1016/j.wocn.2016.02.005

Kleinschmidt, D. F., & Jaeger, T. F. (2015). Re-examining selective adaptation: Fatiguing feature detectors, or distributional learning? *Psychonomic Bulletin & Review*, *23*(3), 678–691. https://doi.org/10.3758/s13423-015-0943-z

Kolinsky, R., Morais, J., & Cluytens, M. (1995). Intermediate representations in spoken word recognition: Evidence from word illusions. *Journal of Memory and Language*, *34*(1), 19–40. https://doi.org/10.1006/jmla.1995.1002

Kraljic, T., & Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology*, *51*(2), 141–178. https://doi.org/10.1016/j.cogpsych.2005.05.001

Kraljic, T., & Samuel, A. G. (2006). Generalization in perceptual learning for speech. *Psychonomic Bulletin and Review*, *13*, 262–268. https://doi.org/10.3758/BF03193841

Krieger-Redwood, K., Gaskell, M. G., Lindsay, S., & Jefferies, E. (2013). The selective role of premotor cortex in speech perception: a contribution to phoneme judgements but not speech comprehension. *Journal of Cognitive Neuroscience*, *25*(12), 2179–2188. https://doi.org/10.1162/jocn_a_00463

Lahiri, A., & Reetz, H. (2010). Distinctive features: Phonological underspecification in representation and processing. *Journal of Phonetics*, *38*, 44–59.

Liberman, A. M. (1996). *Speech: a special code*. Cambridge, Mass: MIT Press.

Lotto, A. J., Hickok, G. S., & Holt, L. L. (2009). Reflections on mirror neurons and speech perception. *Trends in Cognitive Sciences*, *13*(3), 110–114. https://doi.org/10.1016/j.tics.2008.11.008

Luce, P. A., Goldinger, S. D., Auer, E. T., & Vitevitch, M. S. (2000). Phonetic priming, neighborhood activation, and PARSYN. *Perception & Psychophysics*, *62*(3), 615–625.

Macmillan, N. A., & Creelman, D. (1991). *Detection theory: A user's guide*. Oxford: Blackwell Publishers.

Marslen-Wilson, W., & Warren, P. (1994). Levels of perceptual representation and process in lexical access: words, phonemes, and features. *Psychological Review*, *101*(4), 653–675. https://doi.org/10.1037/0033-295X.101.4.653

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*, 1–86. https://doi.org/10.1016/0010-0285(86)90015-0

McQueen, J. M. (2005). Speech perception. In K. Lamberts & R. L. Goldstone (Eds.), *The handbook of cognition* (pp. 255–275). London: Sage Publications.

McQueen, J. M., Cutler, A., & Norris, D. (2006). Phonological abstraction in the mental lexicon. *Cognitive Science*, *30*, 1113–1126. https://doi.org/10.1207/s15516709cog0000_79

McQueen, J. M., Norris, D., & Cutler, A. (1999). Lexical influence in phonetic decision-making: Evidence from subcategorical mismatches. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 1363–1389. https://doi.org/10.1037/0096-1523.25.5.1363

Mehler, J., Dommergues, J. Y., Frauenfelder, U., & Segui, J. (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning & Verbal Behavior*, *20*(3), 298–305. https://doi.org/10.1016/S0022-5371(81)90450-3

Mitterer, H., Chen, Y., & Zhou, X. (2011). Phonological abstraction in processing lexical-tone variation: Evidence from a learning paradigm. *Cognitive Science*, *35*, 184–197. https://doi.org/10.1111/j.1551-6709.2010.01140.x

Mitterer, H., Cho, T., & Kim, S. (2016). What are the letters of speech? Testing the role of phonological specification and phonetic similarity in perceptual learning. *Journal of Phonetics*, *56*, 110–123. https://doi.org/10.1016/j.wocn.2016.03.001

Mitterer, H., & Ernestus, M. (2008). The link between speech perception and production is phonological and abstract: Evidence from the shadowing task. *Cognition*, *109*(1), 168–173. https://doi.org/10.1016/j.cognition.2008.08.002

Mitterer, H., & Müsseler, J. (2013). Regional accent variation in the shadowing task: Evidence for a loose perception-action coupling in speech. *Attention, Perception & Psychophysics*. https://doi.org/10.3758/s13414-012-0407-8

Mitterer, H., & Reinisch, E. (in press). Surface forms trump underlying representations in functional generalisations in speech perception: the case of German devoiced stops. *Language, Cognition and Neuroscience*. https://doi.org/10.1080/23273798.2017.1286361

Mitterer, H., & Reinisch, E. (2013). No delays in application of perceptual learning in speech recognition: Evidence from eye tracking. *Journal of Memory and Language*, *69*, 527–545. https://doi.org/10.1016/j.jml.2013.07.002

Mitterer, H., Scharenborg, O., & McQueen, J. M. (2013). Phonological abstraction without phonemes in speech perception. *Cognition*, *129*, 356–361. https://doi.org/10.1016/j.cognition.2013.07.011

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau. In *Tutorials in Quantitative Methods for Psychology , 4* (pp. 61–64).

Morey, R. D., Rouder, J. N., & Jamil, T. (2015). BayesFactor: Computation of Bayes Factors for Common Designs (Version 0.9.12-2). Retrieved from http://cran.us.r-project.org/web/packages/BayesFactor/index.html

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, *52*, 189–234. https://doi.org/10.1016/0010-0277(94)90043-4

Ohala, J. J. (1996). Speech perception is hearing sounds, not tongues. *Journal of the Acoustical Society of America*, *9*, 1718–1725. https://doi.org/10.1121/1.414696

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1–2), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Poellmann, K., Bosker, H. R., McQueen, J. M., & Mitterer, H. (2014). Perceptual adaptation to segmental and syllabic reductions in continuous spoken Dutch. *Journal of Phonetics*, *46*, 101–127. https://doi.org/10.1016/j.wocn.2014.06.004

Reinisch, E., & Holt, L. L. (2014). Lexically guided phonetic retuning of foreign-accented speech and its generalization. *Journal of Experimental Psychology: Human Perception and Performance*, *40*, 539–555. https://doi.org/10.1037/a0034409

Reinisch, E., & Mitterer, H. (2016). Exposure modality, input variability and the categories of perceptual recalibration. *Journal of Phonetics*, *55*, 96–108. https://doi.org/10.1016/j.wocn.2015.12.004

Reinisch, E., Weber, A., & Mitterer, H. (2013). Listeners retune phoneme categories across languages. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(1), 75–86. https://doi.org/10.1037/a0027979

Reinisch, E., Wozny, D. R., Mitterer, H., & Holt, L. L. (2014). Phonetic category recalibration: What are the categories? *Journal of Phonetics*, *45*, 91–105. https://doi.org/10.1016/j.wocn.2014.04.002

Remez, R. E. (1979). Adaptation of the category boundary between speech and nonspeech: A case against feature detectors. *Cognitive Psychology*, *11*(1), 38–57. https://doi.org/10.1016/0010-0285(79)90003-3

Remez, R. E. (1987). Neural models of speech perception: a case history. In S. Harnad (Ed.), *Categorical Perception: The groundwork of cognition* (pp. 199–225). Cambridge, Mass.: Cambridge University Press.

Repp, B. H., & Liberman, A. M. (1987). Phonetic categories are flexible. In S. Harnad (Ed.), *Categorical Perception: The groundwork of cognition* (pp. 89–112). Cambridge, Mass.: Cambridge University Press.

Rouder, J. N., Morey, R. D., Verhagen, J., Swagman, A. R., & Wagenmakers, E.-J. (2017). Bayesian analysis of factorial designs. *Psychological Methods*, *22*(2), 304–321. https://doi.org/10.1037/met0000057

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*, 225–237. https://doi.org/10.3758/PBR.16.2.225

Samuel, A. G., & Kat, D. (1996). Early levels of analysis of speech. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(3), 676–694. https://doi.org/10.1037/0096-1523.22.3.676

Sawusch, J. R., & Jusczyk, P. (1981). Adaptation and contrast in the perception of voicing. *Journal of Experimental Psychology. Human Perception and Performance*, *7*(2), 408–421.

Schuppler, B., Adda-Decker, M., & Morales-Cordovilla, J. A. (2014). Pronunciation variation in read and conversational austrian German. In *INTERSPEECH* (pp. 1453–1457). Retrieved from https://mazsola.iit.uni-miskolc.hu/~czap/letoltes/IS14/IS2014/PDF/AUTHOR/IS140826.PDF

Sjerps, M. J., & McQueen, J. M. (2010). The bounds on flexibility in speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, *36*, 195–211. https://doi.org/10.1037/a0016803

Smits, R. (2001). Evidence for hierarchical categorization of coarticulated phonemes. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(5), 1145–1162. https://doi.org/10.1037//0096-1523.27.5.1145

Stevens, K. N., & Blumstein, S. E. (1978). Invariant cues for place of articulation in stop consonants. *The Journal of the Acoustical Society of America*, *64*(5), 1358–1368. https://doi.org/10.1121/1.382102

Sussman, H. M., Fruchter, D., Hilbert, J., & Sirosh, J. (1998). Linear correlates in the speech signal: The orderly output constraint. *Behavioral and Brain-Sciences*, *21*, 241–299.

Toscano, J. C., Anderson, N. D., & McMurray, B. (2013). Reconsidering the role of temporal order in spoken word recognition. *Psychonomic Bulletin & Review*, *20*(5), 981–987. https://doi.org/10.3758/s13423-013-0417-0

Van Bezooijen, R. (2005). Approximant /r/ in Dutch: routes and feelings. *Speech Communication*, *47*, 15–31. https://doi.org/10.1016/j.specom.2005.04.010

Weber, A. (2001). Help or hindrance: How violation of different assimilation rules affects spoken-language processing. *Language and Speech*, *44*, 95–118. https://doi.org/10.1177/00238309010440010401

Wickelgren, W. A. (1969). Context-sensitive coding, associative memory, and serial order in (speech) behavior. *Psychological Review*, *76*, 1–15.

Wurm, L. H. (1997). Auditory processing of prefixed English words is both continuous and decompositional. *Journal of Memory and Language*, *37*(3), 438–461.

Appendix

Table A1. Adaptor words in the four conditions of Experiment 1

Onset /r/	**Onset /l/**	**Offset /r/**	**Offset /l/**
reiken	lekken	bakker	appel
reinheid	leggen	cijfer	bijbel
reistas	lessen	deksel	danser
rijbaan	letten	emmer	ezel
rijen	lemma	fakkel	fietser
rijkdom	lengte	handel	honger
rijmen	lengen	hemel	kapper
rijping	lenzen	heuvel	kijker
rijtje	lijstje	kachel	kikker
rijzen	lende	kapsel	masker
rijgen	lente	knuppel	meter
rijkswacht	legging	kogel	moeder
rijstbouw	leiding	mantel	oever
rechten	leien	meubel	pepper
racket	lijden	mossel	puber
rechthoek	lijfje	nagel	spijker
remmen	lijmen	schotel	steiger
rente	lijdzaam	snavel	suiker
reppen	lijvig	spiegel	tijger
redden	lijzen	stempel	venster
rennen	lijsten	stengel	veter
renbaan	lijnbaan	tegel	vijver
restje	lijnen	tunnel	wekker
rekken	lijkant	winkel	winter
rekje	lijken	zadel	zender

Table A2. Adaptor words in the four conditions of Experiment 2

+ phonemic + allophonic	**+ phonemic - allophonic**	**- phonemic + allophonic**	**- phonemic - allophonic**
ähnlich	Andacht	billig	Abend
Bericht	Bedacht	eilig	Auge
deutlich	danach	Essig	Bummel
dicht	demnach	fähig	Dame
endlich	dreifach	fertig	dringend
friedlich	einfach	feurig	eben
Gericht	Eintracht	gierig	Farbe
Gewicht	flach	gültig	Feder
Gicht	Fracht	häufig	Frage
glücklich	Gemach	Honig	ganz
kindlich	Krach	Käfig	Gefahr
Kranich	mehrfach	König	Gewand
leicht	Mühlbach	mickrig	greifbar
Licht	Obdach	mollig	Hilfe
lieblich	Ohnmacht	niedrig	Himmel
möglich	Pracht	traurig	Idee
peinlich	Tracht	übrig	Jugend
Pflicht	Verdacht	völlig	Klavier
Rettich	Vordach	wellig	Lärm
täglich	wach	wendig	Leben
Teppich	Wahlfach	wenig	mager
üblich	wonach	wichtig	Mond