accepted for publication in *Journal of Phonetics*, 8 December 2017

Acoustic cues, not phonological features, drive vowel perception: Evidence from height, position and tenseness contrasts in German vowels

Miquel Llompart [a]*, Eva Reinisch [a]

*Corresponding author: M.Llompart@phonetik.uni-muenchen.de
[a] Ludwig Maximilian University Munich, Schellingstraße 3, 80799 Munich, Germany

Abstract
Phonological features have frequently been singled out as the units of perception, especially for vowels. Evidence of the use of features has been provided for vowel height and vowel position, which have one acoustic correlate only. However, findings on acoustically complex features such as tenseness are less clear. The present study assessed the role of phonological features in perception using the selective adaptation paradigm. Selective adaptation effects on German vowel contrasts differing in vowel height (Experiment 1), position (Experiment 2) and tenseness (Experiment 3) were examined. We tested how the categorization of each vowel contrast was affected by adaptation to words containing vowels that differently resembled or diverged from the vowels in the critical contrast acoustically and in terms of their phonological feature specifications. Results showed that selective adaptation patterns could be predicted by the vowels' phonological features for the height and position contrasts, but not for the tenseness contrast. However, adaptation patterns for the latter can be explained by the relationship between adaptors and continuum endpoints in each of the relevant acoustic cues to the contrast. This suggests that vowel perception may be dependent on these acoustic cues rather than phonological features.

(193 words)

Keywords
Vowels, pre-lexical units, phonological features, acoustic cues, speech perception, selective adaptation

Highlights

- Selective adaptation was used to test the pre-lexical units of vowel perception.
- Phonological features can account for the perception of vowel height and position.
- Phonological features cannot account for the perception of tenseness contrasts.
- Units driving vowel perception are linked to individual acoustic cues.

Acoustic cues, not phonological features, drive vowel perception: Evidence from height, position and tenseness contrasts in German vowels

1.0 Introduction

In order to understand spoken language, listeners have to deal with substantial variability in the speech signal, that is, they have to recognize the words that have been said despite a large range of possible acoustic implementations for those words. For instance, there are substantial between-speaker differences related to vocal tract size, which are most conspicuous in the formant frequencies of vowels (Ladefoged & Broadbent, 1957). In addition, listeners may be exposed to extensive differences caused by dialectal variation (Clopper & Tamati, 2010, 2014; Llompart & Simonet, 2017), foreign accents (Bradlow & Bent, 2008; Clarke & Garret, 2004), and even personal idiosyncrasies (Kraljic, Brennan & Samuel, 2008). Nonetheless, this is challenge that is generally overcome with astounding success.

In order to decode the intended message, listeners need to extract information from the speech signal and map it onto representations stored in the lexicon. There have been different proposals regarding how this is accomplished. It has been argued, for instance, that listeners store all heard forms and directly map newly heard instances onto these stored forms (Goldinger, 1998). However, accumulating evidence suggests that speech processing involves at least some form of pre-lexical abstraction. Pre-lexical abstraction is the mapping of the acoustic-phonetic information in the speech signal into a set of abstract units that serve as the building blocks of the lexical representations that are accessed during perception (McQueen, 2005: McQueen, Cutler & Norris, 2006; Obleser & Eisner, 2009). However, the nature and size of these pre-lexical units remains under dispute. Proposals include, for instance, syllables (Massaro & Chen, 2008; Mehler, Dommergues, Frauenfelder & Segui, 1981), diphones or triphones (Wickelgren, 1969), context-independent phonemes (McClelland & Elman, 1986; Bowers, Kazanina & Andermane, 2016), context-dependent allophones (Mitterer, Reinisch & McQueen, 2018; Mitterer, Scharenborg & McQueen, 2013; Reinisch, Wozny, Mitterer & Holt, 2014; Reinisch & Mitterer, 2016), articulatory gestures (Browman & Goldstein, 1989; Galantucci, Fowler & Turvey, 2006) and phonological features (Boerma & Chladkova, 2011; Lahiri & Reetz; 2010).

Vowel sounds, in particular, have frequently been characterized as being perceived in terms of bundles of phonological features (Boersma & Chladkova, 2011; Chladkova, Boersma & Benders, 2015; Chladkova, Podlipsky & Chionidou, 2017; Eulitz & Lahiri, 2004; Lahiri & Reetz; 2010; Obleser, Lahiri & Eulitz, 2004; Scharinger, Idsardi & Poe, 2011; Scharinger, Monahan & Idsardi, 2012). Phonological features are used to describe vowel systems by referring to abstract combinations of acoustic or articulatory settings that certain vowels share, for example, vowel length, (tongue) height, with its acoustic manifestation in first formant (F1) values, (tongue) position, mainly related to second formant (F2) values, lip rounding (F2, F3), or tenseness (F1, F2, duration).

Evidence of the relevance of phonological features for speech perception comes from studies showing that listeners rely on these features, or the acoustic correlates thereof, during categorization. For instance, Boersma and Chladkova (2011) ran simulations to assess whether the units that listeners used for the categorization of vowels are phonemes (with specific F1 and F2 values for each category) or sets of independent phonological features (*high*, *mid* or *low* with their associated F1 values, and *front*, *central* and *back*, with their associated F2 values). Particularly, they focused on vowel height in languages with five-vowel systems. Previous studies showed that the categorization boundary between high and mid vowels (i.e., /i/-/e/ and /u/-/o/) for such languages is exclusively determined by F1, with the F2 dimension being completely ignored (Chistovich, Fant, de Serpa-Leitao & Tjernlund, 1966; Savela, 2009). Based on this finding, Boersma and Chladkova (2011) taught virtual language learners to either perceive vowels as phonemes or as features and demonstrated that the human categorization patterns reported above were comparable to those of the feature learners in the simulation. This led the authors to conclude that vowel perception is driven by phonetically-based phonological features and not by phonemic representations. Parallel conclusions were drawn from a discrimination task in Moravian Czech, another language with a five-vowel system (Chladkova et al., 2015).

Neurophysiological evidence has also been put forward in support of phonological features (Obleser et al., 2004; Scharinger et al., 2011). Obleser et al. (2004) used magnetoencephalography (MEG) to examine the N100m responses (an auditory evoked potential peaking at around 100 ms) of German listeners to 7 vowels of their native language. Vowels critically differed in vowel height and position. Results showed that vowels differing in only one phonological feature were situated more closely in cortical space than vowels differing in more than one feature. Specifically, source location differences in the N100m were driven by vowel position differences, with front (coronal) vowels and back (dorsal) vowels having different locations. Moreover, back vowels showed longer peak latencies than front vowels. These findings were taken as evidence that abstract phonological features such as coronal vs. dorsal (or front vs. back) determine the processing of vowels at the cortical level.

Obleser et al. (2004) concluded that "the vowel inventory of a given language can be classified on the basis of phonological features which are closely linked to acoustic properties" (Obleser et al., 2004, p. 31). Note that a similar relationship is also outlined in Boersma and Chladkova (2011) and Chladkova et al. (2015). However, even though features in these studies are consistently referred to as phonological, they are not easily distinguishable from other proposed feature-sized units with a purely acoustic basis, such as the acoustic-phonetic features put forward, for instance, by Mesgarani, Cheung, Johnson and Cheng (2014). Mesgarani et al. (2014) used high-density direct cortical recordings in the Superior Temporal Gyrus (STG) during the perception of continuous speech to assess the encoding of phonetic information in speech perception. They found that cortical responses were selective to the phonetic features of sounds, which were directly linked to spectrotemporal cues to the sound's identity. They concluded that the STG encodes speech into a multidimensional acoustic-phonetic feature space.

In addition, all previously mentioned studies present evidence in favor of phonological features coming from the examination of vowel height and vowel position only. However, vowel height and vowel position, the two phonological features that are most frequently studied –possibly because they are most extensively used to distinguish vowels in the world's languages (Maddieson & Precoda, 1992)– have one acoustic correlate each: height distinctions linked to F1 and position distinctions are linked to F2. With height and position only, it is therefore not possible to disentangle whether the outcomes of the studies above stem from the actual abstraction of phonological feature specifications in height or position (e.g., Lahiri & Reetz, 2010) or from the encoding of their unidimensional acoustic correlates into acoustic-phonetic features (Mesgarani et al., 2014). In fact, for vowel height and position, predictions based on the abstraction of phonological features would also not differ from accounts of perception assuming a purely acoustic-auditory basis (Diehl & Kluender, 1989; Holt & Lotto, 2008).

Importantly, the phonological features that have been used to characterize vowels do not necessarily have one acoustic correlate only. Especially in languages with large vowel inventories, additional phonological features have been proposed (Chomsky & Halle, 1968; Jakobson, Fant & Halle, 1952). German is a case in point, since in addition to contrasts in height and position, it has vowels that contrast in roundedness (e.g., /i/ vs. /y/) and tenseness (e.g., /i/ vs. /ɪ/)[1] only. Crucially, roundedness and tenseness are features that are acoustically complex, that is, they have multiple acoustic correlates. Differences in roundedness are related to both F2

[1] Roundedness-only and tenseness-only vowel contrasts are less frequent than height-only and position-only vowel contrasts in the languages of the world (UPSID database; Maddieson & Precoda, 1992)

and F3: round vowels like /y/ have lowered F2 and F3 relative to their unrounded counterparts (here: /i/). The correlates of tenseness are even more complex: the tense vowel /i/, for instance, differs from its lax counterpart /ɪ/ in that it has a lower F1, a higher F2 and a longer duration. Note, however, that other tenseness contrasts may differ in their acoustic relation. In German, the tense back vowel /u/ has a lower F2 than the lax back vowel /ʊ/, in clear opposition to the front vowel contrast described above (/i/-/ɪ/), where the tense vowel has a higher F2.

Examining the perception of contrasts in roundedness and/or tenseness is hence key to test accounts of vowel perception relying on phonological features. This is because these two types of vowel contrasts allow for diverging predictions if one assumes phonological features or individual acoustic cues[2] (aligning with the acoustic-phonetic features discussed in Mesgarani et al. 2014) to drive vowel perception. If perception was determined by the phonological feature specifications of vowels, then one would expect the perception of contrasts differing in acoustically complex phonological features such as tenseness (tense vs. lax vowels, e.g., /i/-/ɪ/) to closely resemble that of contrasts in simple features like vowel height and position (e.g., /u/-/o/, or /u/-/y/). That is, vowels with the same phonological feature specification should pattern together. By contrast, if vowel perception was more closely tied to the individual acoustic cues to the vowels' identity then differences between acoustically simple vs. complex phonological features could be expected.

Crucially, studies that include vowel contrasts in tenseness in addition to height and position (Kingston, 2003; Ettlinger & Johnson, 2009) suggest that there are indeed substantial differences between the perception of simple features such as height and position and complex features such as tenseness. For example, Kingston (2003) investigated the effects of perceptual training on the discrimination abilities of American English listeners for vowel contrasts of German. Although the contrasts always differed in just one phonological feature, some contrasts included acoustically simple features (i.e., height and position) while others included an acoustically complex feature (i.e., tenseness). Results suggested that, for height (e.g., /i/-/e/; /y/-/ø/) and position contrasts (e.g., /y/-/u/; /ø/-/o/), listeners abstracted the critical information related to the phonological feature and used it to discriminate all contrasts differing along the same acoustic dimension. That is, height and position contrasts were discriminated with similar accuracy regardless of their relative position in the vowel space, and most importantly, their discrimination was hardly affected by the introduction of substantial variation in speaker and phonetic context. Discrimination of tenseness contrasts, conversely, was found to differ greatly depending on the height and position of the vowels to be discriminated. For example, for front rounded vowels, the mid vowel contrast /ø/-/œ/ was easier to discriminate than the high /y/-/ʏ/ contrast, even though all four sounds were foreign to the native English-speaking listeners. In addition, discrimination of tenseness contrasts was found to be strongly influenced by variation in speaker and context. This suggests that the perception of these contrasts may be, as the author states, strongly phonetically (acoustically) grounded. If vowels were decomposed into abstract phonological features in perception, then the phonetic complexity underlying these features should not play a role in how they are perceived. A similar conclusion was drawn by Ettlinger and Johnson (2009), who also failed to find supporting evidence for an abstract phonological feature for tenseness in naïve discrimination of foreign contrasts.

---

[2] We use the term *acoustic correlate* when referring to any acoustic dimension determining a vowel's identity and *acoustic cue* to characterize those correlates that are effectively used by listeners for categorization of vowel contrasts. *Acoustic cues* will henceforth be the term used to refer to the driving force of vowel perception according to the different acoustically-based accounts considered. Finally, we use *acoustic-phonetic features* to refer to the feature-sized units proposed to be abstracted, e.g., by Mesgarani et al. (2014), which are discussed as a possible alternative to phonological features in the present study.

In sum, findings to date suggest that phonological feature accounts (Boersma & Chladkova, 2011; Chladkova et al., 2015, 2017; Eulitz & Lahiri, 2004; Lahiri & Reetz; 2010; Obleser et al., 2004; Scharinger et al., 2011, 2012) can provide an explanation on how vowels are perceived in terms of vowel height and position but cannot easily incorporate findings on acoustically complex features such as tenseness. Note, however, that the evidence challenging phonological feature accounts comes from studies on the perception of foreign-language contrasts. This evidence must thus be treated with caution because there are many additional factors that may influence the perception of languages other than one's native language (e.g., Flege, 1995; Bohn & Flege, 1990). Considering this, the present study set out to test whether phonological feature accounts stand scrutiny in native-language perception when both acoustically simple and complex category distinctions are examined. We selected German as the target language because it has a large vowel inventory which includes vowel height, position and tenseness distinctions. We used the selective adaptation paradigm to assess vowel perception (Ades, 1974; Eimas & Corbit, 1973; Morse, Kass & Turkienicz, 1976; Samuel 1986). This paradigm is suitable for our purposes because it has a strong low-level acoustic component, just like the discrimination tasks discussed above (Ettlinger & Johnson, 2009; Kingston, 2003) but is still found to impact the phonetic category level of perception (Cooper, 1974; Diehl, 1975; Samuel, 1986; Samuel & Kat, 1996; Sawusch & Jusczyk, 1981). Selective adaptation can therefore provide insights on the processing of phonetic categories (but see Remez, 1987).

The selective adaptation paradigm consists of a series of adaptation-test sequences. During the adaptation phases, listeners are repeatedly exposed to auditory stimuli that unambiguously belong to a specific sound category. In the subsequent test phases, they are asked to categorize (the ambiguous region of) a continuum between the exposure category and another category. The expected outcome, which is known as selective adaptation (henceforth SA), is that the extensive repetition of one category during adaptation reduces the likelihood of subsequently perceiving the ambiguous stimuli of the continuum as members of that category. For instance, after exposure to many instances of /pa/, the ambiguous region of a /pa/ to /ba/ continuum is more likely to be perceived as /ba/, and conversely exposure to /ba/ results in more /pa/ responses.

Important for the present study, prior findings suggest that SA effects can be expected not only for adaptors with the endpoint categories of the test continuum, but also for adaptors with other categories that resemble one of the endpoints in the relevant dimension (Cooper & Blumstein, 1974; Eimas & Corbit, 1973; Godfrey, 1980; but see Foreit, 1977). For example, Cooper and Blumstein (1974) showed that the categorization of a /ba/-/da/ continuum was shifted towards the /ba/ endpoint (i.e., more /da/ responses) not only after exposure to /ba/, but also –though to a smaller extent– after adaptation with /ma/ and /va/, which are labial sounds like the endpoint /ba/ but differ in manner of articulation. Similarly, Godfrey (1980) reported that repeated exposure to /aɪ/ affected responses to an /ɛ/-/ɛɪ/ continuum in a similar way as the /ɛɪ/ endpoint, triggering more /ɛ/ responses.

Building on this property of SA, in the present study we assessed SA effects on vowel contrasts differing in height (Experiment 1), position (Experiment 2) and tenseness (Experiment 3). We tested how the categorization of such vowel contrasts is affected by repeated exposure to adaptors containing vowels that resemble the continuum endpoints to different extents, both acoustically and in terms of their phonological feature specifications. We made use of the classical adaptation-test sequencing of SA studies and presented adaptors that were naturally spoken mono- or disyllabic German words with the critical vowels in stressed position. After each adaptation phase, the subsequent categorization task was performed on each listeners' most ambiguous region of the test continuum. Importantly, in each experiment, five sets of adaptors were presented in separate blocks, each of them containing words with a specific vowel, while the categorization continuum was the same for all five blocks (see Methods for details).

In Experiment 1 we tested the effects of SA on a vowel height continuum from /u/ to /o/ with adaptor words containing high (/i/, /y/, /u/) and mid vowels (/e/, /o/). This initial experiment was designed to replicate previous findings on the perception of vowel height contrasts, where F1 has been found to be the only acoustic cue effectively determining such distinctions (Boersma & Chladkova, 2011; Chladkova et al., 2015). Experiment 2 focused on vowel position, examining the effects of repeated exposure to back (/u/, /o/) and front vowel adaptors (/i/, /y/, /ø/) on a back-to-front /u/-/y/ continuum. Perception of vowel position is mostly dependent on one acoustic cue, F2, just like vowel height depends on F1. Therefore, we hypothesized that the patterning of SA effects for this experiment should be similar to that of the previous experiment but determined by a different acoustic cue. Finally, Experiment 3 was the strongest test for phonological feature accounts, as it was concerned with tenseness. We investigated the SA effects caused by tense (/i/, /e/, /o/) and lax adaptor vowels (/ɪ/, /ɔ/) on a tense-to-lax /i/-/ɪ/ continuum. The acoustic differences between the two continuum endpoints are manifested in up to three acoustic dimensions: F1, F2 and duration (Bohn & Flege, 1990; Jongman, Fourakis & Sereno, 1989; Kingston, 2003), even though they differ in only one phonological feature, tenseness[3]. Experiment 3 hence allowed for diverging predictions depending on whether phonological features or the individual acoustic cues to the contrast (e.g., Mesgarani et al., 2014) are assumed as the basic units of vowel perception. Phonological feature accounts would predict that all vowels sharing the same specification for tenseness should behave similarly regardless of their absolute acoustic properties. In contrast, acoustic-phonetically rooted accounts would predict that SA effects should be driven by the relations between adaptors and continuum endpoints in all acoustic cues relevant to the contrast.

2.0 Experiment 1

Experiment 1 tested SA effects on a vowel height continuum (/u/-/o/) with adaptor words containing high (/i/, /y/, /u/) vs. mid vowels (/e/ and /o/). The main aim of this experiment was to use the selective adaptation paradigm to replicate effects of the feature "vowel height" that had previously been shown to be crucial for vowel categorization and discrimination (Boersma & Chladkova, 2011; Chladkova et al., 2015). Note that, in order to distinguish contrasts in vowel height, listeners have been shown to solely rely on the acoustic dimension of F1 while ignoring F2. Therefore, Experiment 1 was designed to confirm that other acoustic dimensions (namely F2 and F3) did not impact perception of vowel height contrasts in a selective adaptation paradigm. For this reason, we presented listeners with adaptors whose stressed vowels shared F1 values with one of the test-continuum endpoints but differed from it in F2 (/u/ vs. /y/), and sometimes also F3 (/u/ vs. /i/) (see Figure 1).

[3] Some phonological accounts use Advanced Tongue Root (ATR) as the feature distinguishing tense and lax vowels (Bakovic, 2000; Calabrese, 2000; Halle & Stevens, 1969; Halle, 1977; Lindau, Jacobson & Ladefoged, 1972), however, for the present purposes we stay with the term tenseness, since our predictions for ATR would be the same.

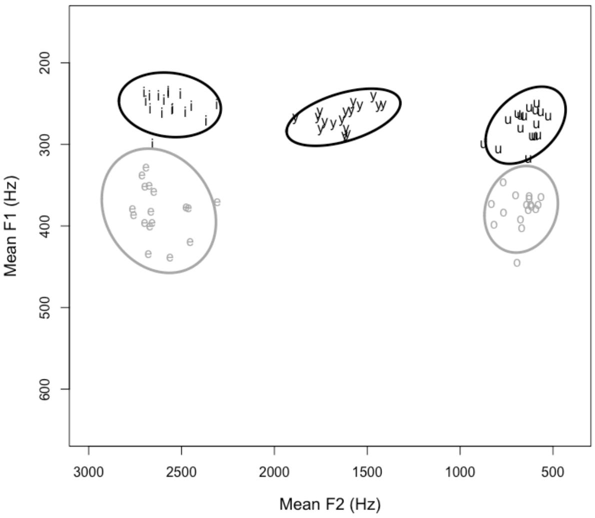

Figure 1. Mean F1 and F2 values for stressed vowels of adaptors used in Experiment 1. High vowels are in black, mid vowels are in grey.

The need for a test of the paradigm stems from the fact that, in principle, it could be possible that only adaptors with the same vowel as the continuum endpoints triggered SA effects, and therefore no SA effects with adaptors other than those with the vowels of the test continuum would be observed (Foreit, 1977). Moreover, generalization of SA effects with consonants has been found to be strongly modulated by the acoustic similarity between adaptors and categorization endpoints (Cooper & Blumstein, 1974). It could thus be that adaptors containing vowels differing from those of the test continuum in some acoustic dimension showed relatively weaker effects. This would suggest that that the vowels' specification for height, or its implementation as F1 values, is not sufficient to account for the observed categorization patterns. Nevertheless, if SA effects with non-endpoint adaptors (this term will be henceforth used to refer to adaptor sets containing vowels other than the endpoints of the categorization continuum) were indeed found, and in addition were of the same magnitude as those of the endpoint adaptors, previous findings on vowel height contrasts would be replicated. These results would then suggest that F1 is the sole determiner of the magnitude of the SA effects, in accordance with previous results using other paradigms (Boersma & Chladkova, 2011; Chladkova et al., 2015).

2.1 Method

2.1.1 Participants

Twenty native speakers of Standard German (18 females; age = 24.06, sd = 2.84) who grew up in German monolingual homes in various (mostly southern) regions of Germany participated in the experiment for a small compensation. All participants reported that they had not learned any other language during their childhood. At the time of testing they were students at the University of Munich. None reported speech or hearing problems.

2.1.2 Materials

The word materials consisted of 87 German words. Two of the words, *bog* (/bog/; preterite form of *biegen* 'to bend') and *Bug* (/bug/ 'bow of a ship') were used to create the continuum for the categorization task. The other 85 words served as adaptors, 17 in each adaptation condition (see Appendix A). Two renditions of each adaptor word plus four renditions of the test minimal pair were recorded by a 27-year-old female speaker of Standard German. She grew up in a German monolingual home in Baden-Württemberg, in the south-west of Germany, and moved to Munich at the age of 19. She is a trained phonetician and speech therapist. Recordings were made in a sound-attenuated booth at the Institute of Phonetics by means of a Neumann TLM 103 condenser microphone (Georg Neumann GmbH, Berlin, Germany) and a

Steinberg/Yamaha UR824 USB audio interface (Yamaha, Hamamatsu, Japan) connected to a Fujitsu PC P920. The *SpeechRecorder* software was used for stimulus presentation and audio recordings (Draxler and Jänsch, 2004). Data were sampled at 44.1 kHz with 16-bit quantization.

All adaptor stimuli were equalized in amplitude, and the critical vowels were manually annotated. Mean first and second formants (F1 and F2) were measured over a time window spanning the mid 50% of the vowel (LPC 25 ms Gaussian window as implemented by Praat; Boersma & Weenink, 2010). One token of each adaptor word was selected to be used in the experiment. The selection was based on recording quality and ensured that each set of adaptors was similar in perceived speech rate, and had consistent F0 contours and formant values in the critical vowels of each set (see Appendix B).

To create the categorization continuum, we selected the recordings of *bog* and *Bug* that best matched in duration and F0 contour. The minimal pair was then morphed in a 25-step continuum (from /100% *bog* – 0% *Bug*/ to /0% *bog* – 100% *Bug*/ in 4% steps) using the STRAIGHT morphing algorithm (Kawahara, Masuda-Katsuse & de Cheveigné, 1999) in Matlab (The MathWorks Inc.). The morphing algorithm decomposes the speech signal into a noise source, a voice source, and a dynamic spectral filter with 10-ms time windows. Interpolation is achieved by first mixing each of the parameters from the two words, and then a new signal is generated from these mixtures. Three time anchors, corresponding to the beginning of the vowel, end of the vowel and beginning of the final stop's release, were used during the morphing procedure in order to ensure that the same types of segments were morphed (i.e., the vowel of one word would only be mixed with the vocalic portion of the other word and not parts of the adjacent consonants). Based on two categorization pretests with six and eight native speakers of German, respectively, 17 steps of the continuum were selected such that the continuum ranged from a clear /u/ to a clear /o/ with the 50% crossover point being centered for most participants.

2.1.3 Procedure

The experiment consisted of a pretest and the main SA part. The aim of the pretest was to identify the most ambiguous step of the *bog-Bug* continuum for each participant. In the pretest, each of the 17 steps of the continuum was presented 10 times, for a total of 170 trials. As response options, two pictures retrieved from a Google image search were used. The picture for *bog* always appeared on the right side of the screen and that for *Bug* on the left side of the screen. Given that *bog* and *Bug* are rather infrequent words in German, the two pictures were accompanied by the orthographic representations of the words, which appeared just below them. Orthography was presented only during the pretest, so that participants had no doubts about which picture corresponded to which word once the SA part started. On each trial of the pretest, participants were asked to press "1" if the word they heard better matched the picture on the left of the screen and "0" if the word better matched the picture on the right. Note that the keyboard response options were spatially in line with picture placement. After pressing a button, the selected picture moved slightly upwards and the other picture disappeared. This feedback was used to indicate participants that their response had been stored. Right after the pretest, a linear regression model was fitted with a logistic linking function (lme4 package 1.1–10 in RStudio version 0.99.486), with Response (*bog*/*Bug*) as categorical dependent variable and continuum step (1-17) as predictor. The resulting curve was plotted and examined by the first author. The step that was closest to the 50% crossover point, and therefore the most ambiguous step for the participant, was selected as reference for the test phase in the SA part. In the unlikely event that two steps were equally distant from the 50% crossover point in the fitted model, the raw-data were consulted.

The SA part consisted of 5 blocks, one for each adaptor condition (i.e., /i/, /y/, /u/, /e/, /o/). Within each block, there were 8 adaptation phases and 8 categorization phases. Each adaptation phase lasted approximately 45-50 seconds. Participants passively listened to the 17 adaptors twice in random order (34 words in total; see Appendix A). Words were separated by 300 ms of silence. Each adaptation phase was followed by 9 categorization trials where listeners responded to 3 repetitions of 3 stimuli selected from the full 17-step continuum: the most ambiguous stimulus for a given participant as determined in the pretest (henceforth step 0), as well as the stimuli preceding and following the most ambiguous step by 2 steps (-2 and 2, respectively). To illustrate, if the most ambiguous step for a given participant was number 9, -2 corresponded to step 7 and 2 corresponded to step number 11. The categorization trials had the same structure and format as the pretest trials except that no orthographic representation of the two words was presented.

Participants were tested individually in a sound-attenuated booth. Each participant was presented with the five blocks in a different order. Block orders were established using Latin Square designs, that is, over sets of 5 participants each type of adaptor block was presented once in each of the five positions. The experiment was conducted on a MacBook Pro 13" running Psychopy2 (v.1.83.01; Peirce, 2007) and auditory stimuli were presented over headphones at a comfortable listening level. The experiment lasted approximately 50 minutes.

2.2 Results

Data were submitted to two generalized linear mixed-effects models with a logistic linking function with response (*bog*/*Bug*) as categorical dependent variable. The within-participant fixed factors were Step (-2, 0, 2) and Adaptor condition (/i/, /y/, /u/, /e/, /o/), as well as their interaction. Step was entered as a numeric variable centered on zero, thus the intercept indicated the overall effect of Step on the factor Adaptor condition that was set as a baseline. The adaptor condition /u/ was mapped onto the intercept in the first model, and /o/ in the second model. That is, the intercepts corresponded to the continuum endpoints. *P*-values are reported as rendered by the models but significance is interpreted based on a Bonferroni-corrected *alpha* criterion to adjust for the fact that each factor level of Adaptor condition was involved in two comparisons across the two models (p is considered significant if $p < .05/2 = .025$). The models were fit with a full random-effects structure, that is, including all slopes and interactions for within-participant factors.

Figure 2 shows the fitted curves across the three categorization steps for each of the 5 adaptation conditions. Point symbols indicate the raw means by step by condition. An inspection of Figure 2 suggests that adaptation to the three high vowels (/i/, /y/, /u/) resulted in considerably fewer *Bug* (/u/) responses than adaptation to the two mid vowels (/e/, /o/) for all three categorization steps.

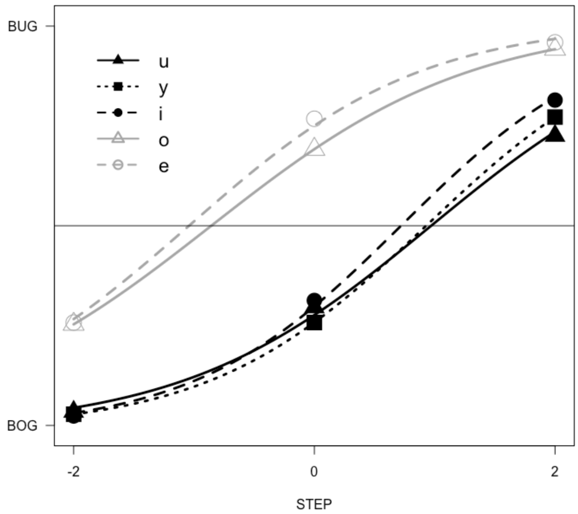

Figure 2. Predicted probability of responding "*Bug*" as a function of Adaptor condition (/u/, /y/ and /i/ in black; /o/ and /e/ in gray). Point symbols indicate the raw means by step by condition and the horizontal bar signals the 50% crossover point.

These observations were confirmed by statistical analyses. The first model had the adaptor condition /u/ mapped onto the intercept. As Figure 2 suggests, responses to the categorization stimuli in the /u/ condition significantly differed from those in the /e/ condition (b = 2.78; z = 6.79; p < .001) and the /o/ condition (b = 2.20; z = 7.64; p < .001). There was no difference in responses between the /u/ condition and the two conditions with adaptors containing high vowels (/i/: b = 0.10; z = 0.26; p = .80; /y/: b = -0.01; z = -0.03; p = .98). The model also revealed an effect of Step (b = 1.29; z = 11.79; p < .001) suggesting that even within the ambiguous region of the test continuum participants were able to perform the phonetic categorization task by giving more /u/-responses the more /u/-like the continuum step was and more /o/-responses the more /o/-like continuum step was. None of the interactions between Adaptor condition and Step were significant (all p > .09). This suggests that the effect of Adaptor condition was similar across the three steps of the test continuum.

In the second model, the /o/ adaptor condition was mapped onto the intercept. Responses in the /o/ condition were found to be significantly different from responses in the /i/ (b = -2.09; z = -6.24; p < .001), /u/ (b = -2.20; z = -7.64; p < .001) and /y/ conditions (b = -2.20; z = -7.22; p < .001). Crucially, however, the difference between the /o/ and /e/ conditions was not significant (b = 0.59; z = 1.80; p = .07). Again, there was a significant effect of Step (b = 1.26; z = 9.70; p < .001) that did not interact with any of the Adaptor conditions (all p > .15). Summarizing, adaptors with high vowels grouped together in the direction and magnitude of their effects and so did adaptors with mid vowels.

2.3 Discussion

Experiment 1 investigated whether SA effects on a vowel height continuum would be found with vowels other than the vowels that formed the test continuum (i.e., non-endpoint vowels) and if so, whether they would be modulated by adaptor vowels' height. Indeed, exposure to all three high-vowel adaptor sets (words with /i/, /y/ and /u/ presented in separate blocks) led to the categorization of *fewer* stimuli along the /u/-/o/ continuum as /u/ compared to the two mid vowels included in the experiment (/e/ and /o/). These mid vowels showed categorization patterns similar to each other with comparatively *more* /u/ responses. Therefore, Experiment 1 showed that non-endpoint adaptors with comparable F1 values to one of the continuum endpoints (see Appendices B and C) triggered SA effects despite their difference in F2 (/y/ vs. /u/) and even F2 and F3 (i.e., /i/ vs. /u/, /e/ vs. /o/).

Most importantly, effects triggered by adaptors with non-endpoint vowels were just as big as those resulting from adaptors with vowels matching one of the continuum endpoints. This indicates that the effects were not modulated by phoneme identity or the overall acoustic similarity between adaptor and endpoint vowels, but only by their relationship in the critical dimension (height; F1), with the other acoustic specifications of F2 and F3 being irrelevant. Interestingly, such a finding stands in contrast with previous results on SA with consonants, where non-endpoint adaptors were found to trigger smaller shifts than the repeated presentation of the endpoint itself (Cooper, 1974; Cooper & Blumstein, 1974; Pisoni & Tash, 1975).

In sum, the pattern of results of Experiment 1 clearly indicates that SA effects generalize across vowel height. These results replicate previous findings and are in agreement with predictions of models that consider phonological features as the basic pre-lexical units of perception (Boersma & Chladkova, 2011; Chladkova et al., 2015, Lahiri & Reetz, 2010; Obleser et al., 2004; Scharinger et al., 2011). Nevertheless, complete generalization by height does not unequivocally mean that listeners abstracted the specification for the phonological feature *height* from the adaptors and used it in the subsequent categorization task. It could also be the case that listeners resorted to acoustic feature units exclusively linked to the F1 values of the vowels presented as adaptors (Mesgarani et al., 2014), or that they simply contrasted the F1 of the adaptor vowels with the F1 values of the categorization stimuli. By acoustic contrast, we mean a change in the response criterion due to the influence of the acoustic context prior to categorization (Diehl, Kluender & Parker, 1985; Diehl, Lang & Parker, 1980; Holt, 2006). Acoustic contrast would have also resulted in adaptors with vowels with a low F1 (high vowels) pairing together and adaptors with vowels with a high F1 (mid vowels) pairing together (but see Mitterer, 2006, who did not find generalization across vowel contexts). In fact, there is no definitive way to disentangle these possibilities examining vowel height alone, since height is an acoustically simple feature with only one acoustic correlate, F1. Therefore, the results of Experiment 1 need to be considered in conjunction with those of Experiment 2 (position), and especially Experiment 3 (tenseness), so as to be able make claims on the units underlying vowel perception.

3.0 Experiment 2

In Experiment 2 we tested selective adaptation of a vowel contrast in which the sounds share vowel height but differ in position (/u/-/y/). Acoustically, they share F1 values and differ in F2 (see Methods). This second experiment sought to test whether SA effects would also generalize as a function of vowel position –similarly to vowel height in Experiment 1. Therefore, in addition to adaptor words with the two categorization vowels, adaptors with /o/, a back vowel, and with /i/ and /ø/, both front vowels, were presented (see Figure 3 for an acoustic characterization of the critical vowels). We now expected that adaptors with back vowels would group together in their effects, triggering more front vowel (/y/) responses and that front vowels would show effects in the opposite direction (more /u/ responses). This prediction was based on the results of Experiment 1, where it was shown that SA effects generalized by vowel height. Vowel position distinctions are similar to vowel height distinctions in that both are directly dependent on one acoustic cue. Moreover, some parallels between height and position have already been shown in previous studies (Chladkova, 2014; Kingston, 2003; Scharinger et al., 2012).

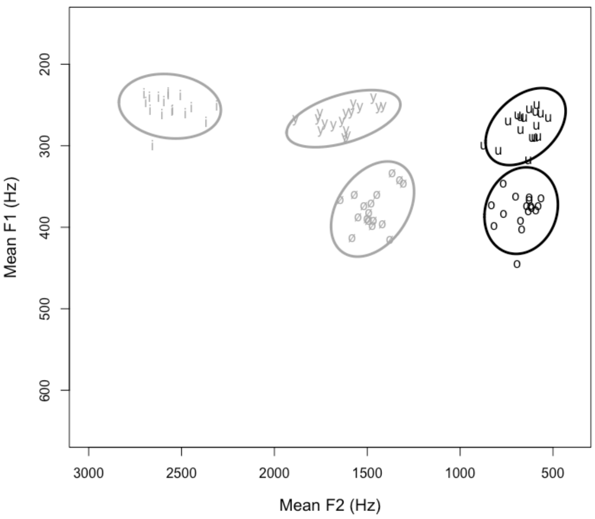

Figure 3. Mean F1 and F2 values for stressed vowels of adaptors used in Experiment 2. Back vowels are in black, front vowels are in grey.

Additionally, including /i/ adaptors made it possible to test how similar adaptors and endpoints need to be acoustically for the vowels to pattern together in their effects. Note that in Experiment 1 each of the five groups of adaptors had an exact match in the relevant dimension, F1, to one of the continuum endpoints (see Figure 1). In Experiment 2, /i/, /ø/ and /y/ have the same phonological specification for position (i.e., front) and F2 is higher than for the other continuum endpoint /u/. However, /i/ differs from /ø/ and /y/ in that it lacks lip rounding, which results in even higher F2 and F3 values. Therefore, if /ø/ and /y/, which are similar in terms of F2 values, patterned together but they differed from /i/, this would suggest that an exact match in the relevant acoustic dimension is needed for SA effects to be similar. Such a finding would speak against phonological features, since /i/, /ø/ and /y/ are equally specified as front, regardless of the diverging absolute F2 values. However, if /i/ pairs with /y/ and /ø/, results would still be in line with previous findings (e.g., Obleser et al., 2004).

3.1 Method

3.1.1 Participants

Twenty-one native speakers of Standard German (13 females; age = 24.90, sd = 3.18) who grew up in German monolingual homes in various (mostly southern) regions of Germany participated in the experiment for a small pay. All participants reported that they had not learned any other language during their childhood. At the time of testing they were students at the University of Munich. None reported speech or hearing problems. One participant had to be replaced and therefore excluded from all analyses because she chose only one response option in all but two test trials irrespective of adaptor condition (99.44% of trials). Importantly, none of the participants had taken part in the previous experiment or pretests.

3.1.2 Materials, Design and Procedure

Part of the materials were the same as in Experiment 1; 68 adaptor words, those corresponding to the /i/, /y/, /u/ and /o/ conditions, were also used in this experiment. An additional set of 17 adaptors with /ø/ as stressed vowel was recorded by the same speaker as in Experiment 1 for the purposes of Experiment 2. Two renditions of each word were recorded and the ones that better matched the stimuli in the other adaptor conditions in perceived speech rate and F0 contours were selected.

Additionally, two nonwords (*Duhk-Dühk* (/duk/-/dyk/)) were also elicited four times in order to create the categorization continuum. Nonwords were used because most of the /u/-/y/ minimal pairs in German are morphologically related, which we tried to avoid for reasons of clarity (e.g., *Bruder* 'brother'-*Brüder* 'brothers'). Importantly, in previous literature the use of nonwords is common (Eimas & Corbit, 1973; Samuel, 2001; Tartter & Eimas, 1975, among others) and no difference has been reported on how selective adaptation operates on nonword and word stimuli (Samuel, 1997).

The 17-step continuum was built and pretested in a similar fashion to the *bog-Bug* continuum in Experiment 1. The experimental design and procedure were the same as in Experiment 1 except for the difference in adaptor conditions (/ø/ adaptors instead of /e/ adaptors) and the different categorization continuum. Importantly, in this case *Dühk* appeared written on the right side and *Duhk* appeared written on the left side. The two written words were accompanied by two pictures of non-objects to lead participants to assign a visual referent to the nonwords and to ensure that the procedure was consistent with that of Experiment 1. However, in Experiment 2 the written forms were also presented throughout the SA part to ensure that listeners did not forget about the nonword-picture associations while categorizing the continuum after exposure to a set of adaptors.

3.2 Results

As in Experiment 1, data were analyzed by means of two generalized linear mixed-effects models with a logistic linking function with response (*Duhk/Dühk*) as dependent variable. The fixed predictors were Step (-2, 0, 2), Adaptor condition (/i/, /y/, /u/, /ø/, /o/), and the interaction between the two. Step was centered on zero and, for Adaptor condition, the two vowels that formed the categorization continuum (/u/ and /y/) were mapped onto the intercept, each in one of the two models. Again, significance is interpreted based on a Bonferroni-corrected *alpha* criterion (.05/2 = .025). Both models were fit with a full random-effects structure. Figure 4 shows the fitted curves for categorization trials after each of the 5 adaptation conditions. It can be observed that adaptation in the back-vowel conditions (/o/, /u/) resulted in fewer *Duhk* (/u/) responses than exposure to adaptors with front vowels (/i/, /y/, /ø/). In addition, it appears that the curves for the two back-vowel adaptor conditions are very similar to each other, as are the curves for the front vowels.

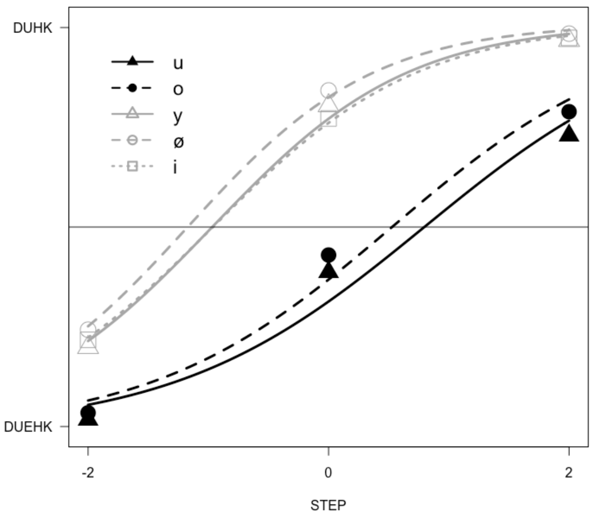

Figure 4. Predicted probability of responding "*Duhk*" as a function of Adaptor condition (/u/ and /o/ in black; /y/, /ø/ and /i/ in gray). Point symbols indicate the raw means by step by condition and the horizontal bar signals the 50% crossover point.

These observations were confirmed by statistical analyses. For the first model, the Adaptor condition /u/ was mapped onto the intercept. Responses to the test items in the /u/ Adaptor condition differed from those in the /i/, /ø/, and /y/ conditions (/i/: b = 2.82; z = 5.35; p < .001; /ø/: b = 3.41; z = 6.82; p < .001; /y/: b = 2.91; z = 6.77; p < .001). No difference was found between responses in the /u/ and /o/ conditions (*b* = 0.31; z = 0.73; p = .47). Additionally, there was an effect of Step (*b* = 1.85; z = 5.85; p < .001) but none of the interactions between Adaptor condition and Step were significant (all p > .19). That is, the effect of Adaptor condition was similar across the steps of the test continuum.

In the second model the /y/ Adaptor condition was mapped onto the intercept. Responses in the /y/ condition significantly differed from responses in the /o/ and /u/ conditions (/o/: *b* = -2.55; z = -5.57; p < .001; /u/: *b* = -2.88; z = -7.01; p < .001). However, no difference was found to the /i/ and /ø/ conditions (/i/: *b* = -0.10; z = -0.19; p = .85; /ø/: *b* = 0.55; z = 1.15; p = .25). Again, the effect of Step was significant (*b* = 2.08; z = 7.57; p < .001) but it did not interact with any of the Adaptor conditions (all p > .3). In sum, the front adaptor vowels grouped together in their SA effects of a vowel position continuum, as did the back adaptor vowels.

3.3 Discussion

Experiment 2 demonstrated that SA effects generalize by vowel position (back vs. front), just as they did in Experiment 1 by vowel height (high vs. mid). Adaptors with /i/, /y/ and /ø/ paired together in that they led to fewer front vowel responses (/y/) than adaptors with /u/ and /o/, which also patterned together in the direction and magnitude of their effects. Just as in Experiment 1, the overall acoustic similarity did not modulate the SA effects in Experiment 2. Non-endpoint adaptors with a similar F2 value to one of the endpoints but differing from it in F1 triggered shifts comparable to that of the endpoint itself. For instance, categorization after exposure to /ø/ adaptors was very similar to categorization after /y/ adaptors, regardless of the differences in F1 values between the two adaptor sets.

Crucially, /i/ also showed SA effects comparable to those of /y/ and /ø/, in spite of the fact that the F2 for /i/ is substantially higher than for /y/ and /ø/. This indicates that an exact match in the relevant acoustic dimension is not needed for reliable SA effects to be found. This is in line with previous findings (Boersma & Chladkova, 2011; Chladkova et al., 2015, 2017; Obleser et al., 2004). Our SA effects could therefore be predicted based on the phonological feature specification of the adaptor vowels. However, just as in Experiment 1, the observed results may also have an acoustic explanation: F2 values for /i/ are higher than /y/ but also much higher than /u/. That is, due the clear opposition between front and back vowels in terms of F2 the same outcome would be predicted if SA effects arose due to the encoding of acoustic-phonetic features linked to F2 (Mesgarani et al., 2014). Since position is related to one single acoustic cue, accounts of phonological or acoustic features cannot be disentangled. Importantly, despite the difference in absolute F2 values for /i/ and /y/ (approximately 1000 Hz, see Appendices B & C), acoustic contrast of the perceived F2 of adaptors and the (ambiguous) categorization stimuli (Diehl et al., 1985) could also explain the results.

In principle, according to a simple acoustic contrast account, adaptation to /i/ could be expected to trigger a bigger shift in categorization than adaptation to /y/ and /ø/, since it has been suggested that the magnitude of contrast effects increases the further the context stimuli (i.e., adaptors) are located from the category boundary probed in categorization (Ainsworth, 1977; Diehl, Elman & McCusker, 1978; Eimas, 1963; Healy & Repp, 1982; Miller, Connine, Schermer & Kluender, 1983). However, this is not always the case, since the increase in magnitude of contrast effects has been found to plateau (Miller et al., 1983). A possibility is thus that F2 values for the three vowels were already sufficiently far from the values of the ambiguous stimuli and thus the increase in magnitude of the effect had already reached ceiling. This would effectively explain why the categorization functions for the three front vowels in Experiment 2 were almost identical.

In sum, results of the first two experiments demonstrated that the two primary dimensions for vowels, height and position, exhibit very similar patterns of generalization of SA effects. These patterns fit with accounts of phonological features as the critical abstract units in vowel perception. However, since in both experiments the alternative explanation that listeners rely on acoustic information rather than abstract phonological features can explain the results as well, Experiment 3 set out to provide a stronger test for phonological features: whether SA effects would also pattern according to the complex phonological feature tenseness.

4.0 Experiment 3

Experiment 3 tested whether generalization of SA effects could be found as a function of the tenseness of the vowels in the adaptor words. Listeners were asked to categorize an /i/-/ɪ/ continuum and, in addition to adaptors with these vowels, they were exposed to sets of word adaptors with /e/, /o/ and /ɔ/ in stressed position (see Figure 5). The vowel inventories of a variety of languages, including German, have been characterized as containing oppositions between tense and lax vowels (Chomsky & Halle, 1968; Halle & Mohanan, 1985; Fischer-Jorgensen, 1990; Kroos, Hoole, Kühnert & Tillmann, 1997). From a featural perspective, contrasts such as /i/-/ɪ/ in German differ in just one feature, tenseness. In the vowel space, tense vowels are usually more peripheral than lax vowels because they are produced with enhanced muscular energy (Lindau, 1978). When the acoustic properties of tense vs. lax vowels are examined, they differ in three dimensions: F1, F2 and duration (see Methods). Differences in these three acoustic correlates of tenseness contrasts are implemented in a systematic manner, even though the absolute values may differ depending on the location of the contrast in the vowel space. Tense vowels like /i/ and /o/ are always longer than their lax counterparts /ɪ/ and /ɔ/. Likewise, all high and mid tense vowels have lower F1 values than their lax counterparts. Regarding F2, however, the relative position of tense and lax vowels varies depending on whether the contrast is front or back. Front tense vowels are more fronted (higher F2) than front lax vowels, and back tense vowels are backer (lower F2) than back lax vowels. That is why /i/ has a higher F2 than /ɪ/, but /u/ and /o/ have lower F2 values than /ʊ/ and /ɔ/, respectively.

Consequently, the /i/-/ɪ/ tenseness contrast provides us with the opportunity of examining whether an acoustically complex vowel distinction leads to similar effects of SA as the acoustically simple ones investigated in Experiments 1 and 2. Note that the results of the first two experiments could be explained by phonological features as well as acoustically-based accounts. For phonological features to withstand scrutiny, it is therefore necessary to show that results hold for all types of phonological features used to distinguish vowels and not only for those that are bound to their acoustic correlates in a one-to-one fashion.

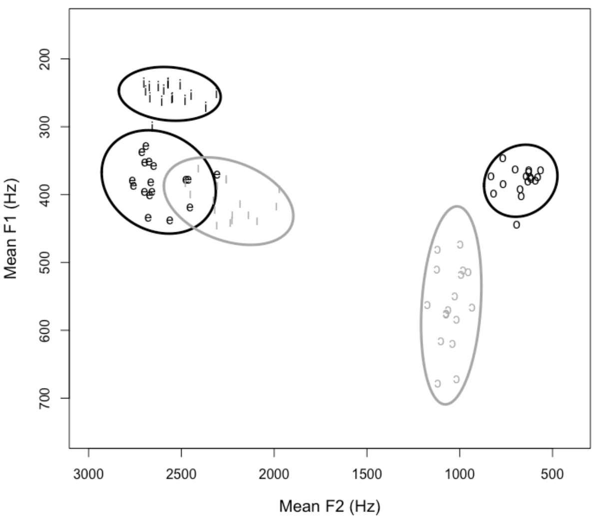

Figure 5. Mean F1 and F2 values for stressed vowels of adaptors used in Experiment 2. Tense vowels are in black, lax vowels are in grey.

A finding of generalization by tenseness, that is, comparable SA effects in one direction for /i/, /e/ and /o/ (more /ɪ/ responses) and the opposite effects for the lax vowels /ɪ/ and /ɔ/ (more /i/ responses), would be strong evidence that listeners abstract phonological features from the acoustic signal. This is because adaptor vowels that share the same feature specification for tenseness may differ greatly in their acoustics. Similarly, vowels that are acoustically similar in some dimensions may differ in tenseness. Compare, for instance, the continuum endpoints /i/ and /ɪ/ with the other tense-lax contrast used as adaptors, /o/-/ɔ/ (see Figure 5). The contrast /o/-/ɔ/ was purposefully chosen because it is the pair that maximally differs from /i/-/ɪ/ in terms of absolute values in F1 and F2. In addition, as mentioned above, differences in F2 between /o/ and /ɔ/ go in the opposite direction to the categorization contrast (i.e., *lower* F2 for tense /o/ compared to /ɔ/; *higher* F2 for tense /i/ compared to /ɪ/). Most importantly, both /o/ and /ɔ/ are spectrally closer to the lax endpoint vowel /ɪ/ than to the tense endpoint /i/. In fact, tense /o/ has very similar F1 values to lax /ɪ/ and /ɔ/ has even higher F1 values. The F2 values of both vowels are much lower than the values of either /i/ or /ɪ/ but still relatively closer to the latter. The only acoustic correlate that is exactly matched in the tenseness distinctions is duration (the two tense vowels are similarly longer than the lax vowels; see Appendices B and C). Therefore, the main question with regard to /o/ and /ɔ/ was whether they would pattern according to tenseness (i.e., in opposite directions) or spectral characteristics (i.e., same direction due to absolute F1 and possibly F2 values).

Finally, our last set of adaptors contained the tense vowel /e/. This vowel is tense and thus similarly long to /i/, and also shares F2 values with this endpoint vowel. In terms of F1, however, /e/ pairs with the lax endpoint /ɪ/ (see Figure 5). Consequently, its phonological specification for tenseness, its duration and the shared F2 values would predict that /e/ patterned with /i/, but the F1 values that are in fact almost identical to /ɪ/ may still drive the effects in the opposite direction.

In sum, if /o/ and /e/ show similar SA effects to /i/ and /ɔ/ to /ɪ/, irrespective of the fact that all three non-endpoint vowels (/e/, /o/, /ɔ/) are closer to /ɪ/ in F1, and /o/ and /ɔ/ also in F2 (see Figure 5), this would be strong evidence in line with phonological features. If no generalization was found as a function of tenseness, the present study would provide evidence that there are critical differences between the perception of acoustically simple (height, position) and acoustically complex (tenseness) vowel contrasts, just as suggested by studies on naïve discrimination of foreign contrasts (Ettlinger & Johnson, 2009; Kingston, 2003). Moreover, if /ɔ/, /o/, and /e/ were similar to /ɪ/ in their effects, this would suggest that the abstracted units need to be directly grounded into the specific acoustic characteristics of vowels and not on phonological grouping properties.

4.1 Method

4.1.1 Participants

Twenty-one native speakers of Standard German (17 females; age = 22.86, sd = 4.30), who grew up in German monolingual homes in various (mostly southern) regions of Germany participated in the experiment for a small compensation. All participants reported that they had not learned any other language during their childhood. At the time of testing they were students at the University of Munich. None reported speech or hearing problems One participant was replaced and therefore excluded from all analyses because she responded with only one option (either all "0" responses or all "1" responses) in all blocks. None of the participants had taken part in Experiment 1 or Experiment 2.

4.1.2 Materials, Design and Procedure

The adaptor sets with the tense vowels /i/, /o/, and /e/ were the same as in the previous experiments. Two additional sets of 17 adaptor words with /ɪ/ and /ɔ/ as the stressed vowel were recorded by the same native speaker of German as in Experiments 1 and 2. Two renditions of each word were recorded and the one that better matched the adaptors of the previous experiments in F0 contour and perceived speech rate were selected.

In addition, two words that form an /i/-/ɪ/ minimal pair (*Miete* 'rent (noun)' - *Mitte* 'center (noun)') were elicited multiple times in order to serve as endpoints for the categorization continuum. The continuum was created via morphing in the same manner as in the previous experiments. Note that now the time anchors around the vowel were also used to vary duration over the continuum. As before, a pretest was conducted to select 17 steps of the continuum such that the most ambiguous point would be close to the middle step for most participants.

The experimental design and procedure of the SA part were the same as in Experiments 1 and 2 except for the use of different adaptor conditions (/i/, /ɪ/, /e/, /o/, /ɔ/) and a different continuum in the pretest and categorization trials during test. For categorization, a picture for *Mitte* always appeared on the right side of the screen and its answer corresponded to the "0" key, while a picture for *Miete* appeared on the left side and could be chosen by clicking on the "1" key. The orthographic representations of the two words appeared below the picture during the pretest but not the SA part, just like in Experiment 1.

4.2 Results

Data were again first analyzed by means of two generalized linear mixed-effects models with a logistic linking function with response (*Miete/Mitte*) as dependent variable. The predictors were Step (-2, 0, 2), Adaptor condition (/i/, /ɪ/, /e/, /o/, /ɔ/), and the interaction between them. The two critical categorization vowels (/i/ and /ɪ/) were mapped onto the intercept in the two models. Step was centered on zero. A third model was fitted later to further explore the difference between the tense vowels /e/ and /o/. Therefore, significance is interpreted based on a Bonferroni-corrected *alpha* criterion that differs from those of the first two experiments because of the additional third analysis (i.e., p is considered significant if $p < .05/3 = .016$). P-values are again reported as rendered by the models. The random-effects structure of the models in Experiment 3 included a random intercept for Participant and random slopes for Adaptor condition and Step over participants but no interaction between the random slopes. This was due to convergence issues of the models. Model comparisons using log-likelihood ratio tests further showed that the simpler model without interaction fit the data better than the model with the full random-effects structure.

Figure 6 shows the fitted curves for categorization in each of the 5 adaptation conditions. It can be seen that exposure to /i/ adaptors triggered fewer *Miete* (/i/) responses than exposure to any other adaptor set, with lax vowel conditions (/ɪ/, /ɔ/) exhibiting very similar curves with comparatively many more *Miete* (/i/) responses and /o/ and /e/ lying somewhere in the middle. Note also that the curve for /e/ appears a little lower than that of /o/, that is, somewhat fewer *Miete* (/i/) responses were given following /e/ than /o/ adaptors.

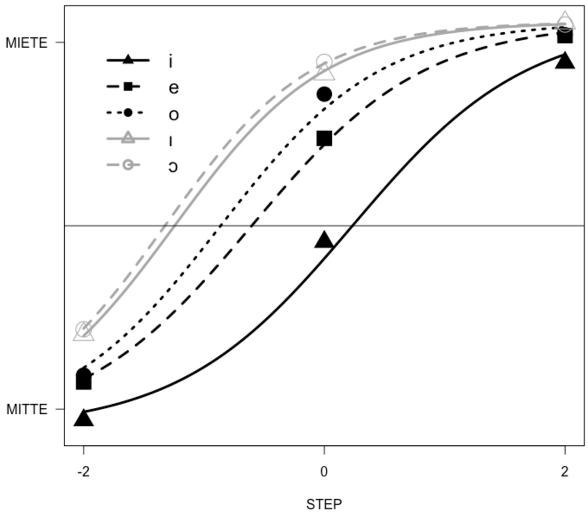

Figure 6. Predicted probability of responding "*Miete*" as a function of Adaptor condition (/i/, /e/ and /o/ in black; /ɪ/ and /ɔ/ in gray). Point symbols indicate the raw means by step by condition and the horizontal bar signals the 50% crossover point.

Results of the model with the /i/ condition mapped onto the intercept showed that categorization in the /i/ condition differed from all other conditions in that fewer *Miete* (/i/) responses were provided (/ɪ/: b = 3.26; z = 9.10; p < .001; /e/: b = 1.84; z = 6.60; p < .001; /o/: b = 2.36; z = 6.86; p < .001; /ɔ/: b = 3.28; z = 9.47; p < .001). Step was also significant (b = 2.11; z = 11.25; p < .001) but none of the interactions with Adaptor condition were (all p > .15).

In the second model the /ɪ/ Adaptor condition was mapped onto the intercept. Results indicated that categorization after exposure to the lax endpoint of the continuum (/ɪ/) did not differ from responses after exposure to /ɔ/ (b = 0.02; z = 0.07; p = .95) but differed from responses after the three remaining Adaptor conditions, that is, /i/ (b = -3.26; z = -9.09; p < .001), /e/ (b = -1.42; z = -5.49; p < .001) and /o/ (b = -0.90; z = -2.77; p < .01). Step was found to be significant (b = 2.32; z = 11.90; p < .001). No interaction between Adaptor condition and Step reached significance given our alpha criterion (all p > .04).

Given that Figure 6 suggests a difference between the categorization curves in the /e/ and /o/ conditions, we ran a third model with Adaptor condition /e/ mapped onto the intercept to test how it differed from the other conditions. Results showed that the numeric differences between the Adaptor conditions /e/ and /o/ were not statistically significant (b = 0.52; z = 1.64; p = .10) but that /e/ differed from all other conditions (/i/: b = -1.85; z = -6.59; p < .001; /ɪ/: b = 1.42; z = 5.47; p < .001; /ɔ/: b = 1.43; z = 4.12; p < .001). Again, the effect of Step was significant (b = 2.16; z = 11.41; p < .001) but none of the interactions were (all p > .25).

4.3 Discussion

Experiment 3 examined whether SA effects would generalize by vowel tenseness, that is, whether SA effects on the categorization of a tense-to-lax continuum occurred as a function of the specification of adaptors and continuum endpoints regarding tenseness (tense vs. lax). Note here that in Experiments 1 and 2 it was already shown that SA effects generalize by vowel height and vowel position (high vs. mid and back vs. front, respectively). However, height and position are determined by only one acoustic correlate. Tense versus lax contrasts such as /i/-/ɪ/, by contrast, exhibit differences in up to three acoustic dimensions, F1, F2 and duration. If effects depended on phonological features, a simple tense vs. lax opposition would be expected, mirroring the results of Experiments 1 and 2. If, on the contrary, effects were driven by differences in the acoustic cues differentiating tense and lax vowels, patterns should be determined by how close adaptor vowels and endpoints are in the acoustic space in these dimensions.

Results showed that repeated exposure to words with the lax vowel /ɪ/ resulted in fewer lax vowel (/ɪ/) responses than exposure to words with the tense vowel /i/. Adaptors with /ɔ/, the other lax adaptor vowel included in the experiment, triggered as few of these responses as /ɪ/, which could in principle be speaking in favor of phonological feature adaptation, since /ɪ/ and /ɔ/ are both lax but different in F1 and F2 and their effects differed from those of all three tense vowels. However, adaptors with the tense vowels /e/ and /o/ did not pair with tense /i/ in the direction of their effects. Instead, they resulted in patterns of responses that were intermediate between /i/ and the lax vowels. This last finding is not in line with predictions based on phonological features, which would predict similar effects for all tense vs. all lax vowels (Boersma & Chladkova, 2011; Chladkova, 2014; Eulitz & Lahiri, 2004; Lahiri & Reetz, 2010; Obleser et al., 2004; Scharinger et al., 2011).

Results for tenseness are nonetheless explainable considering the acoustic properties of the German vowels /e/, /o/ and /ɔ/ and the individual acoustic cues involved in a tense-lax contrast. As mentioned in the introduction to this experiment, our three non-endpoint adaptor vowels have F1 values that are quite similar to the /ɪ/ endpoint of our test continuum and clearly higher than /i/. However, only /ɔ/ is similarly short to /ɪ/, while /e/ and /o/ are more similar to /i/ in that they are much longer than /ɪ/. The statistical difference in the magnitude of the categorization shift between the lax vowel /ɔ/ and the tense vowels /o/ and /e/ can therefore be explained by the (mis)match in duration: while the former matches the /ɪ/ endpoint both spectrally (F1) and in duration, the latter only match in the spectral dimension. Interestingly, F2 seems not to be relevant for categorization, since /e/ and /o/ show similar effects despite the fact that /e/ shares F2 values with /i/, and besides, the categorization patterns following /ɪ/ and /ɔ/ adaptors are not statistically different regardless of their substantial differences in F2. In sum, the results of Experiment 3 indicate that the categorization of tenseness contrasts seems to be dependent on two individual acoustic cues relevant to categorization (F1 and duration), and not simply on the feature specification for tenseness (i.e., tense vs. lax).

5.0 General Discussion

In the present study, we tested the role of phonological features in perception. Phonological features have frequently been postulated as the basic units of perception, especially for vowels. Evidence in line with this has been mostly provided for vowel height and vowel position. These are two dimensions that are linked to one perceptual cue each (height: F1; position; F2). However, findings on acoustically complex features such as tenseness have been found to challenge accounts grounded on phonological feature perception (Ettlinger & Johnson, 2009; Kingston, 2003). We used the selective adaptation (SA) paradigm to investigate adaptation effects on vowel contrasts differing in vowel height (Experiment 1), position (Experiment 2) and tenseness (Experiment 3). In particular, we assessed how the categorization of each type of contrast was affected by adaptation to words containing vowels that resembled the endpoint vowels of the test continua to different extents, acoustically and in terms of their phonological feature specifications. Crucially, results showed that selective adaptation patterns could be predicted by the vowels' phonological features for the height and position contrasts, but not for the tenseness contrast. By contrast, adaptation patterns for all three types of contrast could be explained by the relationship between adaptor vowels and vowels of the test continua in terms of the acoustic cues used to distinguish the critical contrasts.

In the first two experiments, results were in line with full generalization by the phonological features for height and position, respectively. In Experiment 1, exposure to all adaptors with high vowels resulted in similar SA effects that could be clearly differentiated from the effects triggered by mid vowels, which were also similar to one another. In Experiment 2, the same binary pattern arose for vowel position: a clear opposition in SA effects was observed for back vowels vs. front vowels. However, note that vowel height and vowel position have only one acoustic correlate each. Therefore, since high vowels are tied to low F1 values and mid vowels to higher F1 values in a systematic manner (and similarly for position (back vs. front) and F2), there is no need to assume phonological features in vowel perception to explain these findings. The results can be accounted for by the perception of the single acoustic cues that are inextricably linked to said phonological features. The fact that an explanation in terms of phonological features fits with these results may hence be coincidental due to such one-to-one relationship between the proposed phonological features and their acoustic correlates.

Experiment 3 critically allowed us to tease apart phonological features and acoustically-based accounts for vowel perception. In this last experiment, the endpoint vowels differed in F1, F2 and duration, and non-endpoint adaptors could differ from the categorization endpoints in these three acoustic dimensions. A phonological feature account would characterize the /i/-/ɪ/ contrast as an opposition between high front vowels differing in their specification for tenseness. Consequently, were phonological features abstracted from the signal, the effects of adaptor vowels in Experiment 3 should be solely determined by their value for an abstract feature tenseness, just as SA effects could be argued to be determined by the features height and position in the first two experiments. The central finding of the present study was, however, that SA effects did not fully generalize as a function of tenseness. Instead, results of Experiment 3 rendered an adaptation pattern that was gradient as a function of how similar to the endpoints the non-endpoint vowels (i.e., /e/, /o/ and /ɔ/) were in two of the three acoustic dimensions distinguishing the two endpoints of the test continuum: F1 and duration (see Figures 5 and 6). Exposure to the adaptor sets with the two lax vowels resulted in similar categorization patterns because /ɔ/ and /ɪ/ are well matched in F1 and duration. Moreover, they both differed in their effects from the three adaptor sets with tense vowels (/i/, /e/, /o/). This overall tense vs. lax opposition could be in line with phonological feature abstraction, but the difference observed among the three tense vowels, in spite of their shared specification for tenseness, clearly challenges such an account. Adaptors with /e/ and /o/ differed from those with /i/ in their impact on categorization. An explanation for this difference based on the vowels' values for the acoustic cues to the categorization of the tense-lax distinction is, by contrast, easy to provide: /e/ and /o/ showed intermediate adaptation patterns between the two lax vowels and /i/ because their values for F1, on the one hand, and duration, on the other hand, matched one adaptor and mismatched the other in the opposite way: the F1 values of these two vowels are closer to the /ɪ/ endpoint but the duration values are closer to the /i/ endpoint (see Appendices B and C).

A possible account for the results in the present study therefore is that acoustically based phonetic units (i.e., "*acoustic* features") are abstracted from the signal at the pre-lexical level. This would be in line with studies arguing that the features that are encoded in the brain are not phonological but acoustic-phonetic in nature (Mesgarani et al., 2014). Our results agree with this proposal provided that one assumes that the impact on categorization of the acoustic-phonetic features extracted from the adaptors is in turn determined by the vowel contrast to be subsequently categorized. This accounts for the fact that, in Experiments 1 and 2, dimensions that are relevant to ascertain vowel identity in general but were not needed to distinguish between the two vowels spanning the continuum (F2 in Experiment 1; F1 in Experiment 2) were largely ignored, just as reported in previous studies (e.g., Chladkova et al., 2015). Under this account, in Experiments 1 and 2, the opposing patterns of high vs. mid vowels and front vs. back vowels, respectively, would be caused by the encoding of acoustic information from the adaptors and not by the abstraction of such acoustic information into a phonological feature with the specification "high" or "back". For Experiment 3, if acoustic-phonetic features were the units abstracted, more than one of the relevant cues could be encoded separately and therefore have an impact on the SA effects caused by the different adaptor vowels. This would allow for more complex patterns than a simple tense vs. lax opposition. Critically, listeners were indeed found to rely on more than one acoustic cue, with the use of both F1 and duration resulting in the three-level pattern (i.e., /i/ < /e/, /o/ < /ɪ/, /ɔ/) that is observable in Figure 6.

Note that, as mentioned in the Introduction, the abstraction of acoustic-phonetic features is not entirely incompatible with the notion of abstraction discussed in Boersma & Chladkova (2011), Chladkova et al., (2015) and Obleser et al., (2004). These authors assume that the abstracted phonological features need to be acoustically-based, which makes them not so different from the acoustic-phonetic features proposed by Mesgarani et al. (2014), even when they are consistently referred to as phonological. However, the problem with these studies is that they only focus on vowel height and vowel position, and are therefore unable to spell out the relationship between their proposed phonological features and the acoustics characteristics of vowels, since height and position are unidimensional and thus phonological feature specifications and values for the relevant acoustic cues go hand in hand. Consequently, their predictions for contrasts differing in acoustically complex features like tenseness are unknown.

A second acoustically-rooted explanation for our results could be that the effects observed are caused by acoustic contrast (Diehl et al., 1980, 1985; Lotto & Kluender, 1998; Holt, 2006; Mitterer, Csepe & Blomert, 2006). It has long been shown that the perception of individual sounds is affected by the context in which they are produced (e.g., Ladefoged & Broadbent, 1957; Mann, 1980; Miller & Liberman, 1979). Ladefoged and Broadbent (1957) showed that a [bVt] word where the vowel (V) was ambiguous between /ɪ/ and /ɛ/ was more likely to be identified as /bɛt/ than /bɪt/ when an immediate precursor sentence had a lowered F1 than when it had a raised F1. The lowered F1 of the precursor made the target vowel sound as having a relatively higher F1, which therefore led to it being perceived as /ɛ/. The precursor with raised F1, by contrast, made the F1 of the target vowel sound lower, and thus closer to /ɪ/. A similar type of contrast in the relevant dimensions for categorization for each contrast, F1 in Experiment 1, F2 in Experiment 2, and F1 and duration in Experiment 3, would indeed result in patterns approximating the ones we found (Diehl et al., 1978, 1980, 1985). A caveat to this account could be that in Experiment 3 but not in Experiment 2 we found graded effects depending on the match between adaptors and endpoints in absolute acoustic values. However, note that the acoustic differences leading to graded effects (or not) were not necessarily comparable in the two experiments: the gradience in Experiment 3 resulted from one vs. multiple matching cues (i.e., F1 vs. duration and F1), while in Experiment 2 there were absolute acoustic differences in one acoustic cue only (i.e., F2). Moreover, as already discussed, the lack of graded effects in Experiment 2 could be attributed to SA effects reaching a plateau state once adaptors and ambiguous categorization stimuli are sufficiently apart (Miller et al., 1983).

In sum, our results, in particular those of Experiment 3, indicate that listeners' categorization patterns after adaptation are not modulated by the adaptor vowels' specification for the phonological feature distinguishing the two categorization endpoints. An account relying on phonological features (Lahiri & Reetz, 2010; Obleser et al., 2004) would need to predict that all features used to contrast sounds should show a unitary behavior in perception, regardless of the acoustic composition thereof. Consequently, the difference between height and position on the one hand, and tenseness on the other hand, clearly argues against such an account. It could still be argued that phonological features could capture the present results if one assumed an inherent difference between acoustically simple and acoustically complex phonological features in that the latter are more strongly dependent on the acoustic composition of sounds than the former. Although such a division may be postulated, for the present study an explanation directly rooted in the individual acoustic cues to the different vowel contrasts can encompass results on height, position and tenseness. In the light of this, claiming that phonological features may still drive perception for height and position but tenseness is instead directly dependent on the acoustic makeup of vowels is less parsimonious than proposing a common acoustic origin for the results of the three experiments. Our study thus highlights the need to investigate dimensions other than vowel height and vowel position in order to reach a better understanding of native-language vowel perception and adds evidence to previous research questioning the role of phonological features in speech perception (Ettlinger & Johnson, 2009; Kingston, 2003; Mitterer, 2011; Mitterer, Cho & Kim, 2016; Mitterer & Reinisch, 2017; Reinisch & Mitterer, 2016; Reinisch et al., 2014). The SA effects reported in the present study seem to be instead modulated by the vowels' characteristics regarding each of the individual acoustic cues that are decisive for categorization. This strongly suggests that that the pre-lexical units abstracted from the speech signal and mapped onto representations stored in the mental lexicon are not phonologically but phonetically (i.e., acoustically) grounded.

Acknowledgements

We would like to thank Rosa Franzke for her help with testing participants and Holger Mitterer and Joseph V. Casillas for comments on a previous version of the paper. This project was funded by a grant from the German Research Foundation (DFG; grant nr. RE 3047/1-1) to the second author. Parts of the results were presented at the "International Workshop on Abstraction, Diversity and Speech Dynamics", May 2017, Herrsching am Ammersee, Germany.

Appendix A

List of adaptor words used in Experiments 1, 2 and 3. The experiments in which the adaptors were used appear in brackets

/i/-words (1, 2 & 3)	/o/-words (1, 2 & 3)	/y/-words (1 & 2)	/u/-words (1 & 2)	/e/-words (1 & 3)	/ø/-words (2)	/ɪ/-words (3)	/ɔ/-words (3)
aktiv	Brot	berühmt	Beruf	Allee	blöd	Beginn	Bombe
bieten	Dom	früh	Besuch	Fehler	böse	Begriff	Geschoss
Frieden	Hof	Frühling	Blume	geben	Brötchen	Bild	Gold
Gebiet	Hose	Gefühl	Blut	gehen	Flöte	Film	Gott
Klima	Krone	grün	Fußball	jeder	Föhn	Fisch	hoffen
Knie	Mode	kühl	gut	leben	fröhlich	Hilfe	Koch
Krieg	Monat	Kühlschrank	Jugend	Lehrer	Größe	Kind	Koffer
Kritik	Mond	Lüge	klug	lesen	Höhe	Kiste	Kopf
lieber	Montag	müde	Kuba	Meer	Höhle	Licht	Loch
Lied	Note	Mühe	Mut	Meter	König	Liste	oft
Liter	oben	prüfen	Ruf	nehmen	lösen	Pilz	Ost
Mieter	Probe	Süd	Ruhe	Regel	Löwe	Prinz	Post
sieben	Rose	südlich	Schuh	Regen	Möbel	Schiff	Sonne
Sieg	rot	süß	Schule	Schnee	Möwe	Sinn	Stoff
viel	Sohn	üben	Stufe	See	nötig	Stimme	Volk
wieder	Vogel	über	suchen	Thema	Öl	Tisch	voll
Ziel	Zoo	Wüste	Versuch	Zebra	schön	Wind	Woche

Appendix B

Mean duration (in milliseconds) and formant values (in Hertz; mean F1, F2 from 25% to 75% of the vowel) for all adaptor sets used in Experiments 1, 2 and 3. Standard deviations are in brackets. Exp. = experiment

	Acoustic Measurements Adaptors		
Adaptor set	Duration	F1	F2
/i/-words (Exp. 1, 2 & 3)	197 (34)	252 (16)	2563 (112)
/o/-words (Exp. 1, 2 & 3)	259 (39)	380 (22)	671 (81)
/y/-words (Exp. 1 & 2)	213 (43)	267 (14)	1628 (125)
/u/-words (Exp. 1 & 2)	200 (40)	277 (19)	647 (88)
/e/-words (Exp. 1 & 3)	246 (58)	381 (31)	2622 (126)
/ø/-words (Exp. 2)	283 (38)	378 (24)	1471 (91)
/ɪ/-words (Exp. 3)	83 (19)	410 (26)	2246 (143)
/ɔ/-words (Exp. 3)	97 (15)	563 (60)	1046 (66)

Appendix C

Mean duration (in milliseconds) and formant values (in Hertz; mean F1, F2 from 25% to 75% of the vowel) for the endpoint stimuli used to create the morphed continua for Experiments 1, 2 and 3. Exp. = experiment

	Acoustic Measurements Endpoints		
Endpoint	Duration	F1	F2
bog (Exp.1)	283	428	800
Bug (Exp.1)	277	301	628
Dühk (Exp. 2)	219	250	1644
Duhk (Exp. 2)	214	297	752
Miete (Exp. 3)	123	248	2581
Mitte (Exp. 3)	56	372	2201

Reference List

Ades, A. E. (1974). How phonetic is selective adaptation? Experiments on syllable position and vowel environment. *Perception & Psychophysics*, *16*(1), 61–66. doi:10.3758/bf03203251

Ainsworth, W. A. (1977). Mechanisms of selective feature adaptation. *Perception & Psychophysics, 21*(4), 365–370. doi:10.3758/bf03199488

Bakovic, E. (2000). *Harmony, dominance and control*. PhD Dissertation. Rutgers University, New Brunswick, NJ.

Boersma, P., & Chladkova, K. (2011). Asymmetries between speech perception and production reveal phonological structure. In *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 328–331).

Boersma, P., & Weenink, D. (2010). Praat: doing phonetics by computer. Version 5.4.22, http://www.praat.org/.

Bohn, O. S., & Flege, J. E. (1990). Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics, 11,* 303–328. http://dx.doi.org/10.1017/S0142716400008912

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. doi:10.1016/j.jml.2015.11.002

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, *106*(2), 707–729.

Browman, C. P., & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology, 6*(02), 201–251. doi:10.1017/s0952675700001019

Calabrese, A. (2000). The feature [Advanced Tongue Root] and vowel fronting in Romance. *Current Issues in Linguistic Theory, 59*. doi:10.1075/cilt.212.06cal

Chistovich, L., Fant, G., de Serpa-Leitao, A., & Tjernlund, P. (1966). Mimicking of synthetic vowels. *Quarterly Progress and Status Report, Speech Transmission Lab, Royal Institute of Technology, Stockholm*, *2*, 1–18.

Chladkova, K. (2014). *Finding phonological features in perception*. PhD dissertation, University of Amsterdam, the Netherlands.

Chladkova, K., Boersma, P., & Benders, T. (2015). The perceptual basis of the feature vowel height. In *Proceedings of the 18th International Congress of Phonetic Sciences* (article 711).

Chladkova, K., Podlipsky, V. J., & Chionidou, A. (2017). Perceptual adaptation of vowels generalizes across the phonology and does not require local context. *Journal of Experimental Psychology: Human Perception and Performance, 43*(2), 414–427. doi:10.1037/xhp0000333

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

Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *The Journal of the Acoustical Society of America, 116*(6), 3647–3658. doi:10.1121/1.1815131

Clopper, C. G., & Tamati, T. N. (2010). Lexical recognition memory across dialects. *The Journal of the Acoustical Society of America*, *127*(3), 1956-1956.

Clopper, C. G., & Tamati, T. N. (2014). Effects of local lexical competition and regional dialect on vowel production. *The Journal of the Acoustical Society of America*, *136*(1), 1–4.

Cooper, W. E. (1974). Selective adaptation for acoustic cues of voicing in initial stops. *Journal of Phonetics, 2*(4), 303–313.

Cooper, W. E., & Blumstein, S. E. (1974). A “labial” feature analyzer in speech perception. *Perception & Psychophysics, 15*(3), 591–600. doi:10.3758/bf03199307

Diehl, R. L. (1975). The effect of selective adaptation on the identification of speech sounds. *Perception & Psychophysics, 17*(1), 48–52. doi:10.3758/bf03203996

Diehl, R. L., Elman, J. L., & McCusker, S. B. (1978). Contrast effects on stop consonant identification. *Journal of Experimental Psychology: Human Perception and Performance, 4*(4), 599–609. doi:10.1037/0096-1523.4.4.599

Diehl, R. L., & Kluender, K. R. (1989). On the objects of speech perception. *Ecological Psychology, 1*(2), 121–144. doi:10.1207/s15326969eco0102_2

Diehl, R. L., Kluender, K. R., & Parker, E. M. (1985). Are selective adaptation and contrast effects really distinct? *Journal of Experimental Psychology: Human Perception and Performance, 11*(2), 209–220. doi:10.1037/0096-1523.11.2.209

Diehl, R. L., Lang, M., & Parker, E. M. (1980). A further parallel between selective adaptation and contrast. *Journal of Experimental Psychology: Human Perception and Performance, 6*(1), 24–44. doi:10.1037/0096-1523.6.1.24

Draxler, C., and Jänsch, K. (2004). SpeechRecorder—A universal platform independent multichannel audio recording software. *Proceedings of the Fourth International Conference on Language Resources and Evaluation*, (pp. 559–562).

Eimas, P. D. (1963). The relation between identification and discrimination along speech and non-speech continua. *Language and Speech*, *6*(4), 206–217.

Eimas, P. D., & Corbit, J. D. (1973). Selective adaptation of linguistic feature detectors. *Cognitive Psychology, 4*(1), 99–109. doi:10.1016/0010-0285(73)90006-6

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. doi:10.1159/000298584

Eulitz, C., & Lahiri, A. (2004). Neurobiological evidence for abstract phonological representations in the mental lexicon during speech recognition. *Journal of Cognitive Neuroscience, 16*(4), 577–583. doi:10.1162/089892904323057308

Fischer-Jørgensen, E. (1990). Intrinsic F0 in tense and lax vowels with special reference to German. *Phonetica, 47*(3-4), 99–140. doi:10.1159/000261858

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–277). Timonium, MD: York Press.

Foreit, K. G. (1977). Linguistic relativism and selective adaptation for speech: A comparative study of English and Thai. *Perception & Psychophysics, 21*(4), 347–351. doi:10.3758/bf03199485

Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review, 13*(3), 361–377. doi:10.3758/bf03193857

Godfrey, J. J. (1980). Comparison of consonantal and vocalic cues in selective adaptation. *Perception & Psychophysics*, *28*(2), 103–111.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*(2), 251–279.

Halle, M. (1977). Tenseness, vowel shift, and the phonology of the back vowels in Modern English. *Linguistic Inquiry*, *8*(4), 611–625.

Halle, M., & Mohanan, K. P. (1985). Segmental phonology of modern English. *Linguistic Inquiry*, *16*(1), 57–116.

Halle, M., & Stevens, K. (1969). On the feature [advanced tongue root]. *Quarterly Progress Report of the MIT Research Laboratory in Electronics*, *94*, 209–215.

Healy, A. F., & Repp, B. H. (1982). Context independence and phonetic mediation in categorical perception. *Journal of Experimental Psychology: Human Perception and Performance, 8*(1), 68–80. doi:10.1037/0096-1523.8.1.68

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

Holt, L. L., & Lotto, A. J. (2008). Speech perception within an auditory cognitive science framework. *Current Directions in Psychological Science, 17*(1), 42–46. doi:10.1111/j.1467-8721.2008.00545.x

Jakobson, R., Fant, C. G., & Halle, M. (1952). *Preliminaries to speech analysis: The distinctive features and their correlates*. Cambridge, MA: MIT Press.

Jongman, A., Fourakis, M., & Sereno, J. A. (1989). The acoustic vowel space of Modern Greek and German. *Language and Speech*, *32*(3), 221–248.

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

Kingston, J. (2003). Learning Foreign Vowels. *Language and Speech, 46*(2-3), 295–348. doi:10.1177/00238309030460020201

Kraljic, T., Brennan, S. E., & Samuel, A. G. (2008). Accommodating variation: Dialects, idiolects, and speech processing. *Cognition*, *107*(1), 54–81.

Kroos, C., Hoole, P., Kühnert, B., & Tillmann, H. (1996). Phonetic evidence for the phonological status of the tense–lax distinction in German. *The Journal of the Acoustical Society of America, 100*(4), 2691–2691. doi:10.1121/1.417038

Ladefoged, P., & Broadbent, D. E. (1957). Information conveyed by vowels. *The Journal of the Acoustical Society of America, 29*(1), 98–104. doi:10.1121/1.1908694

Lahiri, A., & Reetz, H. (2010). Distinctive features: Phonological underspecification in representation and processing. *Journal of Phonetics, 38*(1), 44–59. doi:10.1016/j.wocn.2010.01.002

Lindau, M. (1978). Vowel features. *Language, 54*(3), 541. doi:10.2307/412786

Lindau, M., Jacobson, L., & Ladefoged, P. (1972). The feature advanced tongue root. *UCLA Working Papers in Phonetics*, *22*, 76–94.

Llompart, M., & Simonet, M. (2017). Unstressed vowel reduction across Majorcan Catalan dialects: Production and spoken word recognition. *Language and Speech*. doi:10.1177/0023830917736019

Lotto, A. J., & Kluender, K. R. (1998). General contrast effects in speech perception: Effect of preceding liquid on stop consonant identification. *Perception & Psychophysics, 60*(4), 602–619. doi:10.3758/bf03206049

Maddieson, I., & Precoda, K. (1992). UCLA phonological segment inventory database. Los Angeles, CA: UCLA Phonetics Laboratory.

Mann, V. A. (1980). Influence of preceding liquid on stop-consonant perception. *Perception & Psychophysics, 28*(5), 407–412. doi:10.3758/bf03204884

Massaro, D. W., & Chen, T. H. (2008). The motor theory of speech perception revisited. *Psychonomic Bulletin & Review, 15*(2), 453–457. doi:10.3758/pbr.15.2.453

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

McQueen, J.M. 2005. Speech perception. In: Lamberts, K., Goldstone, R. (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*(6), 1113–1126.

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. doi: 10.1016/S0022-5371(81)90450-3

Mesgarani, N., Cheung, C., Johnson, K., & Chang, E. F. (2014). Phonetic feature encoding in human superior temporal gyrus. *Science, 343*(6174), 1006–1010. doi:10.1126/science.1245994

Miller, J. L., Connine, C. M., Schermer, T. M., & Kluender, K. R. (1983). A possible auditory basis for internal structure of phonetic categories. *The Journal of the Acoustical Society of America*, *73*(6), 2124–2133.

Miller, J. L., & Liberman, A. M. (1979). Some effects of later-occurring information on the perception of stop consonant and semivowel. *Perception & Psychophysics, 25*(6), 457–465. doi:10.3758/bf03213823

Mitterer, H. (2006). Is vowel normalization independent of lexical processing? *Phonetica, 63*(4), 209–229. doi:10.1159/000097306

Mitterer, H. (2011). The mental lexicon is fully specified: Evidence from eye-tracking. *Journal of Experimental Psychology: Human Perception and Performance, 37*(2), 496–513. doi:10.1037/a0020989

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. doi:10.1016/j.wocn.2016.03.001

Mitterer, H., Csepe, V., & Blomert, L. (2006). The role of perceptual integration in the recognition of assimilated word forms. *The Quarterly Journal of Experimental Psychology, 59*(8), 1395–1424. doi:10.1080/17470210500198726

Mitterer, H., & Reinisch, E. (2017). Surface forms trump underlying representations in functional generalisations in speech perception: The case of German devoiced stops. *Language, Cognition and Neuroscience, 32*(9), 1133–1147. doi:10.1080/23273798.2017.1286361

Mitterer, H., Reinisch, E., & McQueen, J. M. (2018). Allophones, not phonemes in spoken-word recognition. *Journal of Memory and Language, 98*, 77–92. doi:10.1016/j.jml.2017.09.005

Mitterer, H., Scharenborg, O., & McQueen, J. M. (2013). Phonological abstraction without phonemes in speech perception. *Cognition, 129*(2), 356–361. doi:10.1016/j.cognition.2013.07.011

Morse, P. A., Kass, J. E., & Turkienicz, R. (1976). Selective adaptation of vowels. *Perception & Psychophysics, 19*(2), 137–143. doi:10.3758/bf03204220

Obleser, J., & Eisner, F. (2009). Pre-lexical abstraction of speech in the auditory cortex. *Trends in Cognitive Sciences*, *13*(1), 14–19.

Obleser, J., Lahiri, A., & Eulitz, C. (2004). Magnetic brain response mirrors extraction of phonological features from spoken vowels. *Journal of Cognitive Neuroscience, 16*(1), 31–39. doi:10.1162/089892904322755539

Peirce, J. W. (2007). PsychoPy—psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1), 8-13. http://dx.doi.org/10.1016/j.jneumeth.2006.11.017

Pisoni, D. B., & Tash, J. (1975). Auditory property detectors and processing place features in stop consonants. *Perception & Psychophysics, 18*(6), 401–408. doi:10.3758/bf03204112

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

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

Remez, R. E. (1987). Neural models of speech perception: {A} case history. In S. Harnad (Ed.), *Categorical perception* (pp. 199–224). New York: Cambridge University Press.

Samuel, A. G. (1986). Red herring detectors and speech perception: In defense of selective adaptation. *Cognitive Psychology, 18*(4), 452–499. doi:10.1016/0010-0285(86)90007-1

Samuel, A. G. (1997). Lexical activation produces potent phonemic percepts. *Cognitive Psychology, 32*(2), 97–127. doi:10.1006/cogp.1997.0646

Samuel, A. G. (2001). Knowing a word affects the fundamental perception of the sounds within it. *Psychological Science, 12*(4), 348–351. doi:10.1111/1467-9280.00364

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

Savela, J. (2009). *Role of selected spectral attributes in the perception of synthetic vowels*. PhD Dissertation. University of Turku, Finland.

Sawusch, J. R., & Jusczyk, P. W. (1981). Adaptation and contrast in the perception of voicing. *Journal of Experimental Psychology: Human Perception and Performance, 7*(2), 408–421. doi:10.1037/0096-1523.7.2.408

Scharinger, M., Idsardi, W. J., & Poe, S. (2011). A comprehensive three-dimensional cortical map of vowel space. *Journal of Cognitive Neuroscience, 23*(12), 3972–3982. doi:10.1162/jocn_a_00056

Scharinger, M., Monahan, P. J., & Idsardi, W. J. (2012). Asymmetries in the processing of vowel height. *Journal of Speech Language and Hearing Research, 55*(3), 903–918. doi:10.1044/1092-4388(2011/11-0065)

Tartter, V. C., & Eimas, P. D. (1975). The role of auditory feature detectors in the perception of speech. *Perception & Psychophysics, 18*(4), 293–298. doi:10.3758/bf03199377

Wickelgren, W. A. (1969). Context-sensitive coding, associative memory, and serial order in (speech) behavior. *Psychological Review, 76*(1), 1–15. doi:10.1037/h0026823