This is a pre-print version of the article published in the Journal of Experimental Psychology: Human Perception and Performance, Vol 43(2), Feb 2017, 414-427. http://dx.doi.org/10.1037/xhp0000333 This version may not exactly replicate the authoritative document published in the APA journal. It is not the copy of record.

Perceptual adaptation of vowels generalizes across the phonology and doesn't require local context

Kateřina Chládková[1,2], Václav Jonáš Podlipský[3], Anastasia Chionidou[4]

[1]Cognitive and Biological Psychology, Institute of Psychology, University of Leipzig, Germany, Neumarkt 9-19; D-04109 Leipzig, katerina.chladkova@uni-leipzig.de
[2]Amsterdam Center for Language and Communication, University of Amsterdam, The Netherlands
[3]Department of English and American Studies, Palacký University Olomouc, Czech Republic
[4]Department of Theoretical and Applied Linguistics, School of English, Aristotle University of Thessaloniki, Greece

Abstract

Listeners usually understand without difficulty even speech that sounds atypical. When they encounter non-canonical realizations of speech sounds, listeners can make short-term adjustments of their long-term representations of those sounds. Previous research, focusing mostly on adaptation in consonants, has suggested that for perceptual adaptation to take place some local cues (lexical, phonotactic, or visual) have to guide listeners' interpretation of the atypical sounds. In the present experiment we investigated perceptual adaptation in vowels. Our first aim was to show whether perceptual adaptation generalizes to unexposed but phonologically related vowels. To this end, we exposed Greek listeners to words or non-words containing manipulated /i/ or /e/, and tested whether they adapted their perception of the /i/-/e/ contrast, as well as the unexposed /u/-/o/ contrast which represents the same phonological height distinction. Our second aim was to test whether perceptual adaptation in vowels requires local context. Thus, a half of our listeners heard the manipulated vowels in real Greek words, while the other half heard them in non-words providing no phonotactic cues on vowel identity. The results showed similar adjustment of /i/-/e/ categorization and of /u/-/o/ categorization, which indicates generalization of perceptual adaptation across phonologically related vowels. Furthermore, adaptation occurred irrespective of whether local context cues were present or not, suggesting that, at least in vowels, adaptation can be based on the distribution of auditory properties in the input. Our findings, confirming that fast perceptual adaptation in adult listeners occurs even for vowels, highlight the role of phonological abstraction in speech perception.

Keywords: perceptual adaptation; perceptual learning; vowel perception; perceptual generalization; vowel height; local context; lexical cues; phonotactic cues

Statement of public significance

When perceiving speech, listeners face enormous variability in speech-sound realization. We exposed Greek listeners to words pronounced by a speaker whose /i/'s or /e/'s sounded atypical, halfway between /i/ and /e/. Hearing these words caused changed perception not only of /i/ and /e/ but even of two related vowels, /u/ and /o/, which the listeners had never heard from the speaker. This reflects systematic organization of speech sounds in language users' minds. Moreover, listeners adjusted their perception even without feedback about whether an atypical sound had been intended as /i/ or /e/. This shows that adults can learn to process novel speech purely on the basis of its acoustic properties, possibly by a similar mechanism that infants employ in native language acquisition. Our findings demonstrate mutual interrelation between speech perception and long-term mental representations for sounds and confirm that ongoing perceptual learning is natural even for adults using their native language.

1. Introduction

Perception of speech exhibits remarkable flexibility. In everyday interaction, listeners promptly adapt to newly heard speech that often contains atypical realizations of speech sounds (Greenspan, Nusbaum, & Pisoni, 1988; Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995; Rosen, Faulkner, & Wilkinson, 1999; Clarke & Garrett, 2004; Bradlow & Bent, 2008). Listeners encounter outlier realizations of their native-language sound categories for various reasons. It may be due to long-term characteristics of the speaker (a regionally or socially different native accent, unfamiliar foreign accent, or idiosyncrasies of vocal-tract anatomy), the speaker's temporary characteristics (slurring, a cold, food in the mouth), or distortions in the channel (a poor recording). Despite the frequent occurrence of atypical speech tokens, two speakers of a language can usually easily understand one another even without prior experience with each other's productions. When a listener hears a non-canonical exemplar of a speech sound, contextual information can help them assign such a token to the appropriate speech sound category (for a review see Cutler, 2012).

Ganong (1980) demonstrated that when encountering a non-typical speech sound realization, listeners tend to perceive the sound as belonging to the category which is appropriate in the particular lexical context, i.e. in the particular word. Later research showed that the effect extends beyond the categorization task itself. It has been found that perceptual adaptation of speech sound categories happens even after very brief exposure to modified target sounds. Norris, McQueen, and Cutler (2003) exposed Dutch listeners to twenty Dutch /s/- or /f/-final words, such as *glass* or *cliff* to give an analogous example in English, where the final fricative was an ambiguous sound midway between [s] and [f]. Subsequently, listeners altered their perceptual categorization of a [s]-[f] continuum depending on the lexical context during exposure: listeners who had heard the ambiguous fricative in words where /s/ is expected learned to treat it as /s/, while those who had heard it in contexts for/f/ treated it as /f/. Crucially, in the categorization task, listeners identified *several* non-identical ambiguous tokens (i.e. also those to which they had not been exposed in the lexical decision task) as either /s/ or /f/. This indicates, the authors argue, that listeners adapted their native perceptual boundary between the two phonemes in that they shifted it in the appropriate direction along the [s]-[f] continuum. Kraljic and Samuel (2005) reported similar perceptual retuning for the [s]-[ʃ] continuum, the direction of which again depended on the lexical context in which the ambiguous [s]-[ʃ] fricatives had been presented.

These studies required listeners during the exposure phase to make explicit decisions about the lexical status of the words heard, asking whether they were existing words or not. Norris et al. (2003) found that perceptual adaptation along the [f]-[s] continuum occurred when listeners heard the ambiguous tokens in real words but not when they heard them in non-words. McQueen, Norris, and Cutler (2006b) tested whether explicit activation of the lexicon is necessary for perceptual adaptation to occur. The authors repeated Norris et al.'s (2003) experiment, but instead of requiring listeners to make lexical decisions, they asked them to count the number of items presented during the exposure phase. McQueen et al.'s (2006b) results were similar to those obtained earlier: perceptual adaptation occurred even without the explicit use of lexical knowledge during exposure. However, lexical information *was* in fact present in the design of their experiment: the exposure material consisted of real words carrying meaning in the listeners' language, and so they most likely did activate lexical representations (even without the explicit instruction to do so). In a later experiment, Cutler, McQueen, Butterfield, and Norris (2008) exposed listeners to non-words where the ambiguous [s/f] fricative was placed in a context where either /s/ or /f/ was phonotactically legal. The results showed that in the absence of word meaning, the phonotactics, i.e. pre-lexical information, was sufficient to guide the identification of the ambiguous fricative and cause perceptual retuning of /s/-/f/ categorization. In these experiments, the lexical or phonotactic knowledge served as top-down local-context cues for the listeners. Other research has used bottom-up cross-modality cues: a visual cue to the category identity of the ambiguous sounds (such as engagement of the lips) was also found to bring about perceptual recalibration (e.g. Bertelson, Vroomen, & De Gelder, 2003; Reinisch, Wozny, Mitterer, & Holt, 2014). In light of these findings, Samuel and Kraljic (2009) proposed that for perceptual learning to occur "some constraint on how the sound should be interpreted is necessary" (p. 1213).

McQueen, Cutler, and Norris (2006a), using lexical-context guiding, found that perceptual retuning was not specific to the lexical items presented in training but that it generalized across words, suggesting that recalibration is not item-specific. Kraljic and Samuel (2006) found that perceptual adaptation for sounds along the [d]-[t] voice-onset-time continuum generalized across different speakers. What is more, their results showed analogous adaptation for speech sounds ranging between [b] and [p] that were not included in the exposure phase. This means that listeners generalized the exposure-induced shift in their perceptual boundary to a non-exposed but phonologically related phoneme contrast. However, the generalization of perceptual recalibration is not without limit. Reinisch and Mitterer (2016), using lexical guiding, did not find generalization from the /b/-/d/ contrast to /m/-/n/ representing the same place-of-articulation distinction. This was in parallel to their earlier study using visual cueing (Reinisch et al., 2014) which failed to observe generalization not only across place contrasts but even to the same phoneme contrast in a different segmental context. To explain the apparent conflict between the generalization of perceptual retuning for voicing contrasts and the lack thereof for place contrasts, Reinisch and Mitterer (2016) argued that generalization occurs only when the acoustic cues for the identity of a speech sound are contained within the segment itself (as is the case with VOT that cues the voicing distinction in English), but not when they are distributed over the neighboring segments (as is the case with formant transitions in flanking vowels that cue the place distinctions in consonants), also highlighting the "need for acoustic similarity between exposure and generalization contrasts" (p. 106).

All of the studies reviewed so far and several others (see Samuel & Kraljic, 2009, for a comprehensive review) reported perceptual adaptation in consonants. Research into perceptual adjustments in vowels has mostly used different experimental paradigms than that introduced by Norris et al. (2003). First, several studies explored the effects of a precursor phrase on the perception of an immediately following vowel, inspired by Ladefoged and Broadbent's (1957) classic study. Using synthetic speech, Ladefoged and Broadbent (1957) showed that the same [bVt] word was more likely to be perceived as /bɛt/ than /bɪt/ when it followed the carrier phrase 'please say what this word is' with a lowered first formant (F1), which made the F1 of the target vowel seem relatively higher (i.e. more appropriate for /ɛ/ than for /ɪ/). This effect was replicated with natural speech (Ladefoged, 1989), but also reversed speech (Watkins, 1991; Watkins & Makin, 1994) and spectrally-rotated speech analogs (Sjerps, Mitterer, & McQueen, 2011) as precursors, indicating it is at least partially an auditory effect. Mitterer (2006) tested whether the effect would also occur when the precursor contained only a subpart of the vowel space (high front vowels), and when it contained non-words. The results did not show generalization of the perceptual adjustment to vowels that were not contained in the precursor but, as expected, perceptual adjustment occurred even when the precursor contained non-words. Accordingly, Mitterer (2006) concluded that this immediate perceptual adjustment was independent from the mechanism driving the medium-term adaptations of the perception of consonants (as documented by Norris et al., 2003, and the following literature).

The second line of research into perceptual adaptation concerning vowels used longer exposure phases and delayed tests, just like the studies on consonants, but instead of directly assessing changes of phoneme categorization, it measured a general 'tuning in' to the speaker's accent. For instance, Maye, Aslin, and Tanenhaus (2008) exposed listeners to a 20-minute familiar story synthesized in an accent of English with front vowels lowered in the vowel space (for instance, the word *witch* was realized as [wɛtʃ]). In a subsequent lexical decision task listeners were more likely (than in a baseline condition) to judge as existing words tokens whose front vowel was replaced by a lower front vowel. Maye et al.'s (2008) experiment thus showed that listeners could adjust the perception of their native vowel phonemes. An unresolved question is whether the perceptual adaptation in vowels generalizes to phonologically related contrasts. Recall that Mitterer (2006) found no evidence of such generalization using the carrier-phrase paradigm. Maye et al. (2008) argued that in their Experiment 1, listeners somewhat lowered also back vowels, which is why the authors proposed that a general perceptual lowering of the entire vowel system had occurred; no such effects for back vowels were found in their Experiment 2. On the basis of Maye et al.'s (2008) experiments, however, one cannot draw any firm conclusions about the generalization of vowel lowering from front to back vowels: their listeners always heard manipulated front vowels as well as *unmanipulated back* vowels in the exposure phase, which would have prevented listeners from making any front-to-back generalization.

Skoruppa and Peperkamp's (2011) findings indicated that phonological regularities of vocalic patterning indeed play a role in listeners' tuning in to atypical accents. Using a similar design as Maye et al. (2008), they exposed listeners to three artificial accents of French: front vowels either harmonized in the rounding feature with the preceding vowels, or they disharmonized, or some vowels harmonized while others disharmonized. At test, listeners trained on the harmonic accent or on the disharmonic accent more successfully recognized words with the harmony pattern they had been exposed to than listeners trained on the mixed-harmony accent. That suggests that the learning was feature-based: if the harmony patterns in the three accents are conceptualized in terms of individual vowel changes, they are equally complex, whereas if the rounding feature is considered, the harmonic and the disharmonic accent are less complex than the mixed-harmony accent.

There is at least one study that explored the recalibration of vowel perception using Norris et al.'s (2003) design. McQueen and Mitterer (2005) exposed Dutch listeners during a lexical decision task to vowels ambiguous between Dutch /i/ and /eː/ in lexical contexts for either /i/ or /eː/. They later tested /i/-/eː/ categorization, as well as categorization for two other Dutch vowel contrasts, /ɪ/-/ɛ/ and /ɑ/-/ɔ/. The results showed exposure-induced recalibration of the /i/-/eː/ boundary in some but not all test blocks. As regards generalization of recalibration, the results were similarly inconclusive: only generalization to the spectrally more dissimilar /ɑ/-/ɔ/ was observed, and only in one block. This lack of clear evidence of generalization parallels Mitterer's (2006) failure to find generalization using the carrier-phrase paradigm and is somewhat surprising, given the hypothesis based on research with consonants that generalization occurs for segments containing their own acoustic cues (Reinisch et al, 2014; Reinisch and Mitterer, 2016), which is true of vowels. It is possible that generalization of recalibration across vowel contrasts occurs for more symmetrical vowel contrasts than those included in McQueen and Mitterer's (2005) study, or in more symmetrical vowel systems in general.

The present study therefore tests perceptual adaptation for vowels and its generalization across the vowel system of Greek, a 5-vowel language with high front /i/, mid front /e/, low central /a/, mid back /o/, and high back /u/ (Fourakis, Botinis, & Katsaiti, 1999). We assess whether exposure to vowels with manipulated quality results in an adjustment of the category boundaries of these as well as non-exposed but phonologically related vowels: recall that such generalization has been attested for voicing contrasts but not for place-of-articulation contrasts in consonants (Kraljic & Samuel, 2006; Reinisch & Mitterer, 2016, respectively), and that the findings for vowels are inconclusive. We ask whether Greek listeners exposed to /i/ and/e/ with manipulated height adapt their perceptual boundary for these front vowels and whether they adapt their boundary also for the back vowels of corresponding height which are not included in the exposure. Participants are exposed either to words with lowered /i/s and words with unmanipulated /e/s, or to words with unmanipulated /i/s and words with raised /e/s. We then assess whether they lower or raise their /i/-/e/ boundary, as predicted by the direction of manipulation, and whether they do so for the /u/-/o/ boundary as well. The vowels /u/ and /o/ are phonologically related to /i/ and /e/: they share the height feature, so if listeners shift both the /i/-/e/ as well as the /u/-/o/ boundary, it will indicate that they adapt their phonological system as a whole and not only the exposed vowel contrasts. The present findings will thereby reveal whether adult listeners are able to adapt (at least on a short term basis) the abstract linguistic representations that they have acquired early in life for the vowels of their native language.

As noted above, Samuel and Kraljic (2009) proposed that some contextual guiding was necessary for perceptual learning to occur. However, their review was based almost exclusively on experiments with consonants. In contrast, research using the carrier-phrase paradigm (Mitterer, 2006) has shown that the short-term adjustment of vowel perception occurs even when the carrier phrase contains non-words. The present study, testing the medium-term perceptual adaptation in vowels, therefore again asks what type of information is the prerequisite of such adaptation: are local cues (such as lexical or phonotactic knowledge, or lip-reading) really necessary to guide listeners' identification of the non-canonical sounds, or is unassisted categorization of the auditory speech input sufficient? In this context, literature on speech sound acquisition by adults becomes relevant.

Studies on statistical or distributional learning of speech (mostly using vowel sounds) indicate that on the basis of the sounds' statistical distribution in the input, adult (and infant) listeners can learn to differentiate two speech sounds that in their native language fall within a single phoneme category (Maye & Gerken, 2000; Maye, Werker, & Gerken, 2002; Goudbeek, Cutler, & Smits, 2008; Escudero, Benders, & Wanrooij, 2011; Escudero & Williams, 2014; Wanrooij, Boersma, & van Zuijen, 2014; Ong, Burnham, & Escudero, 2015). For instance, in Goudbeek et al. (2008), native speakers of Spanish were exposed to vowel tokens drawn from two clusters in the auditory space (i.e. from a bimodal distribution) that represent a phonological contrast in Dutch but not in Spanish. Goudbeek et al. showed that on the basis of the sounds' distributions the Spanish listeners learned to differentiate the non-native contrast. Although learning seems to be more successful when feedback about category membership is given during the training phase (Goudbeek et al., 2008), or when attention to an auditory task is induced (Ong et al., 2015), participants in most distributional training experiments successfully learn to distinguish novel contrast on basis of the auditory information alone (but see also Wanrooij, Boersma, & Benders, 2015, who did not find a learning effect).

It is plausible that this mechanism plays a role in perceptual adaptation. Clayards, Tanenhaus, Aslin, and Jacobs (2008) showed that during a categorization experiment listeners adjust the crispness of their perceptual boundaries between /p/ and /b/ as they hear tokens from the /p/ and /b/ categories with a broadened variance. Liu and Holt (2015) showed that in the course of a categorization test listeners adjusted the relative weighting of two acoustic cues to a vowel contrast based on the distribution of these cues in the stimulus set. Therefore, even without local cues to the identity of non-canonical realizations of speech sounds, the statistical distribution of the acoustic properties of speech sounds in the input can assist the perceptual interpretation of the ambiguous sounds and hence trigger adaptation. That is, sounds that are ambiguous between two categories can be perceived as representing the one whose canonical realizations are absent in the input.

We thus hypothesize that when they encounter atypical realizations of speech sounds, adults are able to recalibrate their perceptual categories even without the involvement of local-context cues. In other words, we hypothesize that perceptual recalibration can happen on the basis of mere exposure to the phonetic environment, without any feedback or supervision from higher levels. In our experiment, besides testing the generalization of adaptation across phonologically related contrasts, we specifically ask whether perceptual adaptation in vowels occurs after auditory exposure either to words or to non-words providing no phonotactic cues to the identity of the atypical speech sounds. To this end, half of our Greek listeners are exposed to /i/- or /e/-manipulations in existent Greek words (i.e. when lexical cues to /i/ or /e/ identity are present), and the other half are exposed to /i/- or /e/-manipulations in non-words in which both /i/ and /e/ are phonotactically permissible (i.e. neither lexical nor phonotactic cues are available).

2. Method

2.1 Participants

Fifty-seven native speakers of Greek participated. One was excluded because no category boundary could be determined in her categorization responses to the /u/-/o/ contrast (see Section 3). The remaining 56 participants (45 females) formed 4 groups of 14 participants each, who received different types of exposures (see below). They came from various parts of Greece, although the majority (65%) was from Central Macedonia. The various regions of origins were distributed comparably across the four participant groups. Importantly for the present study, any differences between regional accents of Greek are not likely to be reflected in the acoustic characteristics of vowels (Sfakianaki, 2002). The participants' age ranged from 18 to 45 (mean age 23.5). Many of them had a good command of English but all were clearly Greek-dominant.

2.2 Stimuli

2.2.1 Word and non-word items for the exposure phase

We created a list of real Greek words and phonotactically well-formed non-words. They were di-, tri-, or quadrisyllabic. Approximately a half of the words and a half of the non-words were 'critical words', while the other half were 'fillers'. In the critical (non)words, there was always an /i/ or an /e/ in the stressed position while the remaining vowels were /a/s. Further, we selected each existent Greek critical word so that replacing the target vowel with its contrasting counterpart (i.e. either /i/ with /e/, or /e/ with /i/) would not produce another existent Greek word. This was to ensure there were lexical cues to the identity of the target vowel. In contrast, in each critical non-word both /i/ and /e/ were phonotactically permissible in the stressed syllable and at the same time each non-word with /e/ remained a non-word if the /e/ was replaced by /i/ and vice versa; this was to ensure there were no phonotactic or lexical cues to vowel identity in the non-words. Additionally, the list of items also included for each critical (non)word a non-word where the target vowel was replaced by its contrasting counterpart (e.g. for the real word καφές 'coffee', the list also contained the non-existent version καφίς). These items were not part of the exposure materials; they were included only to obtain measurements of the acoustic properties of both /i/ and /e/ in the same phonetic contexts. In the filler words, all the vowels were /a/s. Since our aim was testing whether any perceptual recalibration of the /i/-/e/ contrast induced by exposure to manipulated tokens of these vowels would generalize to the /u/-/o/ contrast, there were no instances of /u/ or /o/ in the critical words or the fillers whatsoever.

To ensure that exposure to manipulated vowel tokens was sufficient for recalibration to occur, the critical (non)words were presented more than once during the exposure phase; for the counts of each item type and the number of repetitions, see Table 1. For the entire lists of the words and non-words see Tables A1 and A2 in the Appendix.

Participants		Critical words				Fillers				Total
group	*n*	status	manipulation	number	repetitions	status	vowels	number	repetitions	trials
1	14	real words	original /i/; raised /e/	28; 28	2x; 2.5x	non-words	only /a/	49	1x	175
2	14	real words	lowered/i/; original /e/	28; 28	2.5x; 2x	non-words	only /a/	49	1x	175
3	14	non-words	original /i/; raised /e/	16; 16	3x; 4x	real words	only /a/	48	1x	160
4	14	non-words	lowered/i/; original /e/	16; 16	4x; 3x	real words	only /a/	48	1x	160

***Table 1**. The four groups of participants that were exposed to different manipulation directions (lowered /i/ vs raised /e/) and lexical status of the critical words (real words vs non-words). The table indicates the numbers and numbers of repetitions of each item type during exposure. The '2.5x' repetitions mean that the items were randomly assigned to two halves (differently for each participant) one of which was repeated twice and the other 3 times. In our between-subject design, critical real words were combined with non-word fillers, and vice versa.*

These real words and non-words were recorded by a native male speaker of Greek from Athens. The list of items was randomized and was recorded twice. The speaker was asked to read the items at a comfortable speech tempo and to repeat any item at which he hesitated or made a pronunciation error. One token of each recorded (non)word was selected to be included in the stimulus set for the lexical decision task (this was the token that was judged by the authors as best in terms of voice quality and auditory clarity, and it was mostly the speaker's first production of that word). The recoding was done in a sound-attenuated phonetics laboratory at the University of Amsterdam, using a Sennheiser MKH 105 T microphone and Tascam CDRW900 recorder, at the sampling rate of 44.1 kHz.

Vowel	Critical words		F1	F2	F3	F1	F2	F3	duration
			mean (and SD) in Hz			mean (and SD) in ERB			in ms
/i/	real words	original	332 (32)	2316 (88)	2833 (170)	7.81 (0.53)	22.14 (0.32)	23.81 (0.49)	111 (34)
		shifted	447 (32)	1898 (129)	2591 (177)	9.56 (0.44)	20.47 (0.57)	23.06 (0.56)	
	non-words	original	339 (49)	2317 (119)	2882 (243)	7.92 (0.82)	22.16 (0.41)	23.93 (0.69)	101 (36)
		shifted	457 (34)	1983 (230)	2605 (272)	9.71 (0.48)	20.81 (0.91)	23.08 (0.87)	
/e/	real words	original	560 (41)	1682 (84)	2492 (62)	11.03 (0.49)	19.48 (0.41)	22.75 (0.21)	103 (23)
		shifted	413 (47)	2028 (124)	2660 (120)	9.07 (0.70)	21.03 (0.50)	23.29 (0.38)	
	non-words	original	568 (50)	1728 (80)	2569 (103)	11.12 (0.62)	19.71 (0.38)	23.00 (0.33)	120 (41)
		shifted	414 (51)	2122 (136)	2762 (145)	9.07 (0.78)	21.41 (0.52)	23.60 (0.43)	

***Table 2**. The values of F1, F2, F3 and duration (means and standard deviations, SD) of /i/ and /e/ measured in the unaltered critical (non)words and in their manipulated versions. For formants, the table shows values in Hz as well as in ERB.*

The quality of vowels /i/ or /e/ in the recorded critical (non)words was manipulated using source-filter resynthesis in Praat (Boersma & Weenink, 2015; see Praat manual entry 'Source-filter synthesis 4. Using existing sounds'). We first measured the first three formants in the recorded /i/- and /e/-versions of every (non)word. The target vowels were then shifted so that their resulting spectral quality (in terms of the first three formant frequencies and their bandwidths) was psycho-acoustically (as measured in the equivalent rectangular bandwidth [ERB] scale) a little more than halfway between the original quality and the speaker's average quality for the contrasting front vowel. This means that the F1 of /i/ was raised, while its second and third formants (F2 and F3) were lowered. In contrast, the F1 of /e/ was lowered, while its F2 and F3 were raised. We interpolated the F1, F2, and F3 contours at the edges of the vowel (10% of the vowel's duration at each end), to preserve smooth transitions from and to the adjacent (unmanipulated) segments. In order to preserve the naturalness of the items, we modified only the lower three formants, keeping the original upper spectrum unchanged. The resynthesis was performed using a Praat script but it involved a visual inspection of the formant contour estimates obtained by linear predictive coding (Burg method). If the formant tracks did not correspond well with the formants seen in a spectrogram, either the prediction order and/or the Nyquist frequency were adjusted, or the tracks were corrected manually. Figure 1 plots the F1 and F2 values of the unaltered critical (non)words as well as of their resulting manipulated versions, and Table 2 lists the average values for F1–F3.

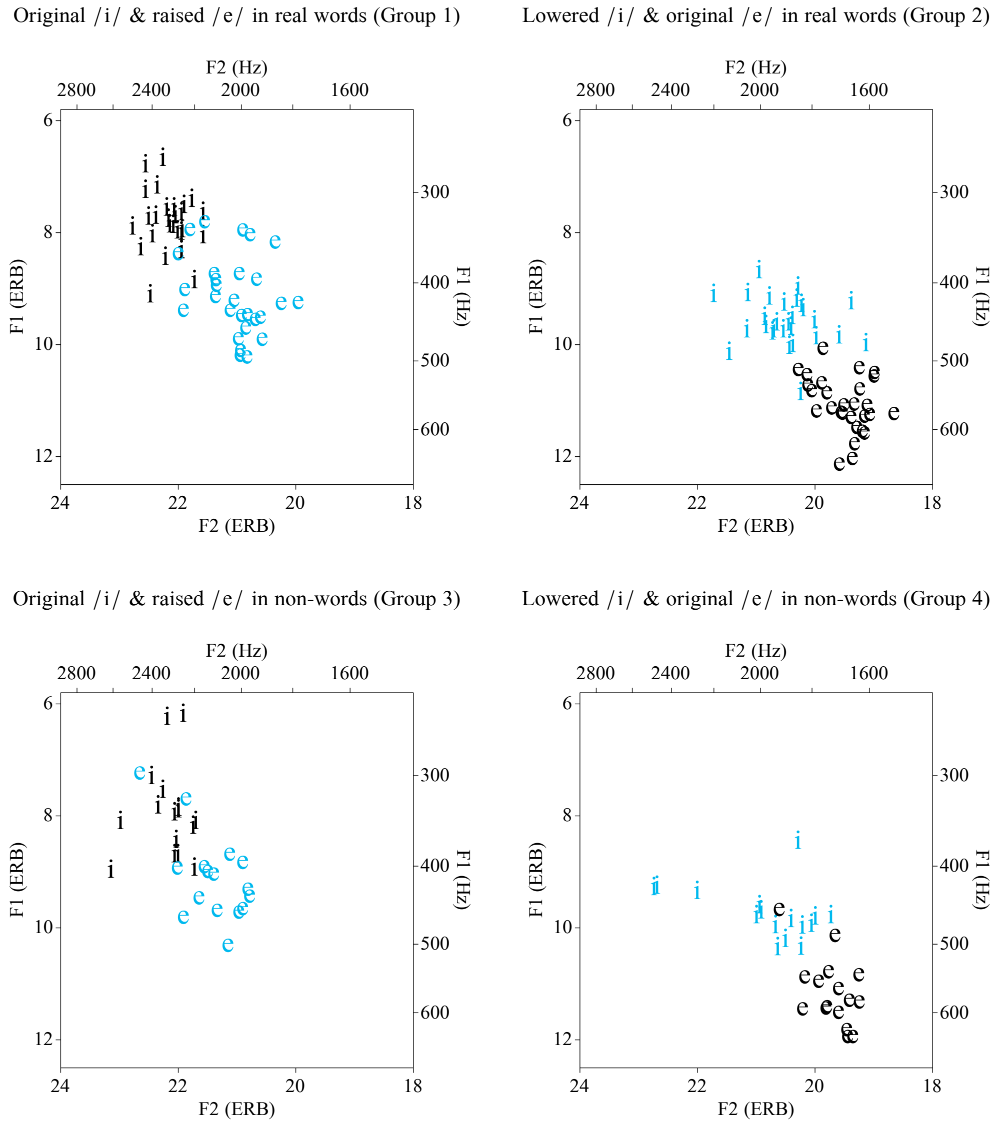

Figure 1. *Exposure stimuli: scatterplots showing F1 by F2 values of /i/ and /e/ measured in the unaltered critical (non)words and in their manipulated versions. The four plots represent the exposure materials presented to our four groups of participants (see Table 1). Black font represents original vowel qualities, and blue (grey) font manipulated*

vowels. Note that instead of substituting the target vowels with a single resynthesized ambiguous speech sound (as was done in many previous studies with consonants), we manipulated every single instance of /i/ or /e/ in the respective (non)words. This was to preserve the variability induced by the consonantal context, and thus to preserve naturalness of the exposure stimuli.

2.2.2 Vowel sets for testing post-exposure categorization

The same speaker who produced the words and non-words also recorded six isolated tokens of /i/, /e/, /u/, and /o/ each. The best 5 of each were selected. Five unique 12-item vowel continua were created for the front vowel contrast and 5 for the back contrast to be used in testing post-exposure vowel categorization. This was achieved by interpolating in 11 psycho-acoustically equal steps (ERB scale) within each of five /i/-/e/ token pairings and five /u/-/o/ pairings, again manipulating F1, F2, F3 and their bandwidths and preserving other acoustic characteristics (such as F0 and duration). Specifically, the five original tokens of /e/ and the five tokens of /o/ served as the staring points, and they were each resynthesized to produce the 11 shifted tokens until the quality of the respective counterpart in their pair was reached.[1] This yielded in total 60 physically different vowels for the /i/~/e/ set, and 60 vowels for the /u/~/o/ set. All are plotted in Figure 2. The minimum and the maximum F1 value of the resulting /i/~/e/ set were 221 Hz and 528 Hz, respectively. For the /u/~/o/ set, the minimum and maximum F1 values were 210 Hz and 483 Hz, respectively.[2]

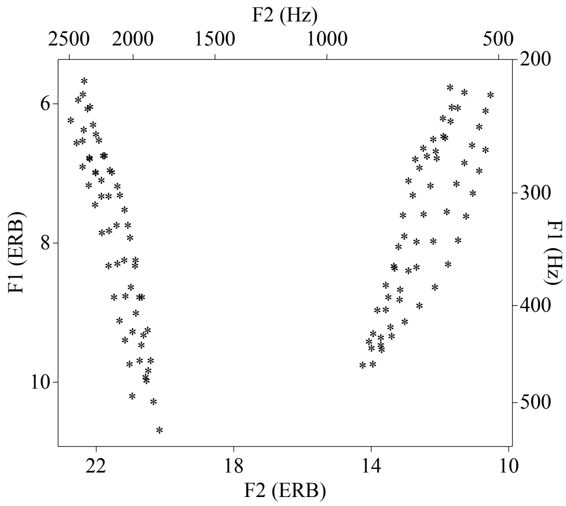

Figure 2. *An F1 by F2 plot of the isolated vowel tokens from the post-exposure categorization task. To the left: the 60 stimuli from the /i/~/e/ set; to the right: the 60 stimuli from the /u/~/o/ set.*

2.3 Procedure

The experiment was administered by a native speaker of Greek. It was conducted in a sound-treated booth at the Phonetics Lab of the School of English, Aristotle University of Thessaloniki, using Sennheiser HD 280 pro headphones and a Realtek HD Audio soundcard for the auditory presentation.

---

[1] The choice of /e/s and /o/s rather than /i/s and /u/s as input to the resynthesis was arbitrary. Importantly, the spectral qualities of the resulting vowel tokens covered the full mid vs. high range (see Figure 2).

[2] Note that the highest F1 values of the post-exposure /i/~/e/ set (i.e. F1 of the tokens at the /e/ end) were lower than some of the F1s of unaltered /e/s in the exposure (non)words. This was because we created the stimuli on the basis of naturally-produced vowel tokens and (non)words, preserving the natural acoustic variability of the vowels (even after shifting, as seen in Figure 1). Our speaker produced /e/s with somewhat higher F1s in (non)words than in isolation.

Eight of the 57 participants were tested in their homes, always in a small quiet room, using Sennheiser HD 202-II headphones and a laptop with a Conexant CX20671 soundcard.). The whole experiment was implemented as a Praat Demo window script (Boersma & Weenink, 2015). Prior to testing, the participants were unaware of the purpose of the experiment.

Listeners first performed a lexical decision task (the exposure phase) and were subsequently tested in a vowel identification task. In the lexical decision task, listeners were played a randomized set of word-items comprising the critical (non)words with the vowels /i/ and /e/ and fillers with the vowel /a/. Crucially, listeners were divided into 4 groups that differed in vowel manipulation direction and in the lexical status of the critical words (2 × 2 design). As shown in Table 1, one group of listeners was exposed to existent Greek words with lowered /i/s, existent Greek words with unaltered /e/s, and to non-words with /a/s; the second group was exposed to existent words with raised /e/s, existent Greek words with unaltered /i/s, and non-words with /a/; the third group was exposed to non-words with lowered /i/s, non-words with unaltered /e/s, and to existent words with /a/s; and the last, fourth, group was exposed to non-words with raised /e/s, non-words with unaltered /i/s, and to existent words with /a/s. Importantly, during the exposure phase, listeners did not hear any instances of the back vowels /o/ and /u/ at all. The number of repetitions of the stimuli during the exposure, and the total number of trials, has also been given in Table 1. In each trial, the participant heard a stimulus and then clicked on a button marked 'Είναι λέξη' "It is a word" or a button marked 'Δεν είναι λέξη' "It is not a word". There were 6 example trials at the beginning. The participants were offered to take a brief silent break 3 times during the lexical decision task.

The exposure phase was followed by a short break, during which the participants did not hear any speech (not even the experimenter's unless absolutely necessary). After the break, participants performed a vowel identification task. This task was divided in two blocks, whose order was counterbalanced across participants. In one block they were presented with the 60 isolated vowels from the /i/~/e/ set in random order without repetition, and in the other block with the 60 vowels form the /u/~/o/ set. After hearing a stimulus the participant clicked on one of two buttons marked 'ε' "e" versus 'η' "i" in one block, and 'ο' versus 'ου' "u" in the other and proceeded to the next trial. At the beginning of each block, there were 4 example trials with endpoint stimuli. Halfway through each block, and between blocks, participants could take a short silent break.

3. Results

We first computed the percentage of correct responses in the lexical decision task. The scores of individual listeners were 93% correct or higher. These high scores indicate that all the participants mostly perceived the lexical status of the stimuli as intended, even in the case of real words with manipulated vowels. Therefore, data from all listeners were included in the analysis of the subsequent vowel identification task.

The vowel identification data, that is the /i/ vs /e/ responses for the /i/~/e/ set, and the /u/ vs /o/ responses for the /u/~/o/ set, were submitted to a binomial logistic regression with F1 as the independent variable. Note that the vowels within each set differed to some extent also in their F2, F3 and duration.[3] Since the primary acoustic correlate of vowel height is F1, in our analysis we operationalized the high vs. mid vowel boundaries using the factor F1. For each participant, we thus computed the F1 value of the /i/-/e/ boundary and the F1 value of the /u/-/o/ boundary. An

[3] Recall that each set of 60 tokens consisted of 5 different resynthesized continua. The F2 and F3 thus varied across all 60 tokens, while vowel duration differed between the 5 continua (i.e. duration was the same for all 12 tokens per continuum).

ambiguous vowel with an F1 value exactly at the /i/-/e/ boundary has 0.5 probability of being labeled as /i/ and 1–0.5 probability of being labeled as /e/ (analogously for /u/-/o/). Therefore, the boundary location x along the F1 dimension was computed using the formula:

$$ln\frac{0.5}{1-0.5} = \beta_0 + \beta_1 x$$

where β_0 and β_1 are the logistic regression coefficients. Since $ln\frac{0.5}{1-0.5} = 0$, then

$$x = -\frac{\beta_0}{\beta_1}$$

Data of one participant were excluded because the analysis failed to find a boundary for the /u/-/o/ contrast; further analyses were done on data from the remaining 56 participants. Table 3 lists the β_0 and β_1 obtained for each manipulation direction and vowel set. Figures 3 and 4, respectively, plot the individual and averaged logistic regression curves for each manipulation direction and vowel set. The boundary locations were submitted to a repeated-measures analysis of variance with Manipulation direction (lowered /i/, raised /e/) and Lexical status (real words, non-words) as the between-subjects variables, and Vowel set (/i/~/e/, /u/~/o/) as the within-subjects variable.

Vowel set	Manipulation direction	β0 mean	95% c.i.	β1 mean	95% c.i.
/i/~/e/	raised /e/	34.076	28.803..39.349	−0.107	−0.123..−0.090
	lowered /i/	29.206	23.933..34.480	−0.086	−0.103..−0.069
/u/~/o/	raised /e/	31.214	25.541..36.886	−0.095	−0.112..−0.078
	lowered /i/	29.580	23.908..35.252	−0.087	−0.104..−0.070

Table 3*: Logistic regression coefficients (means and 95% confidence intervals) obtained for each Manipulation direction and Vowel set.*

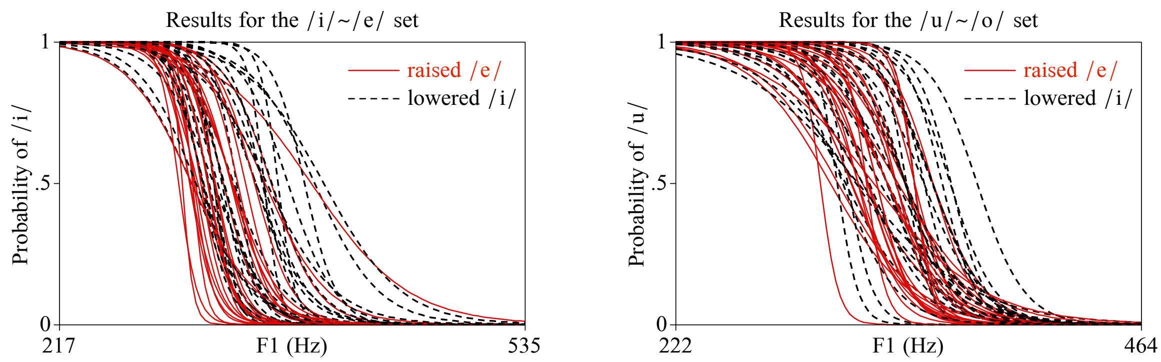

Figure 3. *Logistic regression curves modelling post-exposure categorization by individual listeners. Left panel: results for the /i/~/e/ set; right panel: /u/~/o/ set. Red (gray) solid lines: results for the groups exposed to raised /e/ and unmanipulated /i/; black dashed lines: groups exposed to lowered /i/ and unmanipulated /e/. Note the slightly different F1-range for each contrast.*

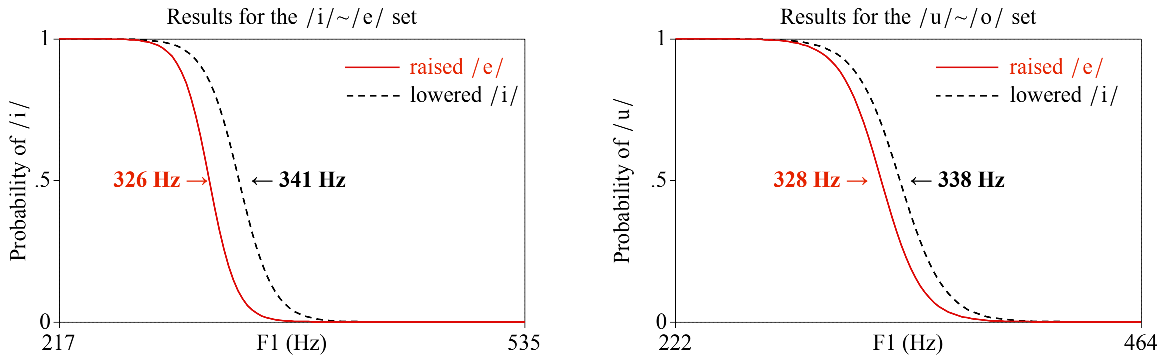

Figure 4. *Logistic regression curves and boundary locations (numbers next to arrows) averaged across participants in each group. Red (gray) solid lines: results for the groups exposed to raised /e/ and unmanipulated /i/; black dashed lines: groups exposed to lowered /i/ and unmanipulated /e/. Note the slightly different F1-range for each contrast.*

Factors	F (1,52)	*p*	partial η^2
Vowel set	0.017	.895	< .001
Lexical status	2.349	.131	.043
Manipulation direction	**7.158**	**.010**	**.121**
Vowel set * Lexical status	**4.827**	**.032**	**.085**
Vowel set * Manipulation direction	1.298	.260	.024
Lexical status * Manip. dir.	0.521	.474	.010
Vowel set * Manip. dir. * Lexical status	0.404	.528	.008

Table 4. *List of all significant (boldface) and non-significant effects.*

Factors yielding significant effects	Levels		Mean F1 boundary (Hz)	95% confidence interval
Manipulation Direction	/i/-lowered		339	333..346
	/e/-raised		327	320..334
Vowel set * Lexical status	non-words	/i/~/e/	340	331..349
		/u/~/o/	333	327..340
	real words	/i/~/e/	327	318..335
		/u/~/o/	333	326..339

Table 5. *Mean F1 boundary locations and their confidence intervals for cases in which significant main effects or interactions were detected.*

Table 4 lists all significant and non-significant main and interaction terms along with their F-statistics and effect size. Table 5 presents descriptive statistics for the obtained significant effects (means of boundary locations and their confidence intervals). As can be seen in Table 4, the analysis yielded a main effect of Manipulation Direction ($F[1,52] = 7.2$, $p = .010$, partial $\eta^2 = .121$); participants exposed to raised /e/ (and unmanipulated /i/) had the high-vowel versus mid-vowel boundaries at lower F1 values than participants exposed to lowered /i/ (and unmanipulated /e/). No significant interaction between Manipulation Direction and Vowel set was found. There was a significant interaction of Vowel set and Lexical status ($F[1,52] = 4.8$, $p = .032$, partial $\eta^2 = .085$); the /i/-/e/ boundary was at lower F1 values in participants who were exposed to vowel manipulations in real words than in those who were exposed to manipulated non-words.

No significant main effect of Lexical status or Vowel set was found, and no other interactions were significant. The absence of a significant interaction between Vowel set and Manipulation direction, and between Lexical status and Manipulation direction, indicates that perceptual adaptation happened in both vowel sets and in both word types. To examine the effects of manipulation direction within each vowel set and each word type, Figure 5 shows mean boundary locations and 95% confidence intervals for the /i/-/e/ and /u/-/o/ boundaries in the /i/-lowered and /e/-raised group, and for the high-vs-mid boundaries in real words and non-words in the/i/-lowered and /e/-raised group. Inspection of the confidence intervals shows that the /i/-lowered versus /e/-raised boundary difference for the /u/-/o/ set has the same direction and similar magnitude as the boundary difference for the /i/-/e/ set. Also, the /i/-lowered versus /e/-raised boundary difference in the group exposed to non-words has the same direction and similar magnitude as the boundary difference in the group exposed to real words[4].

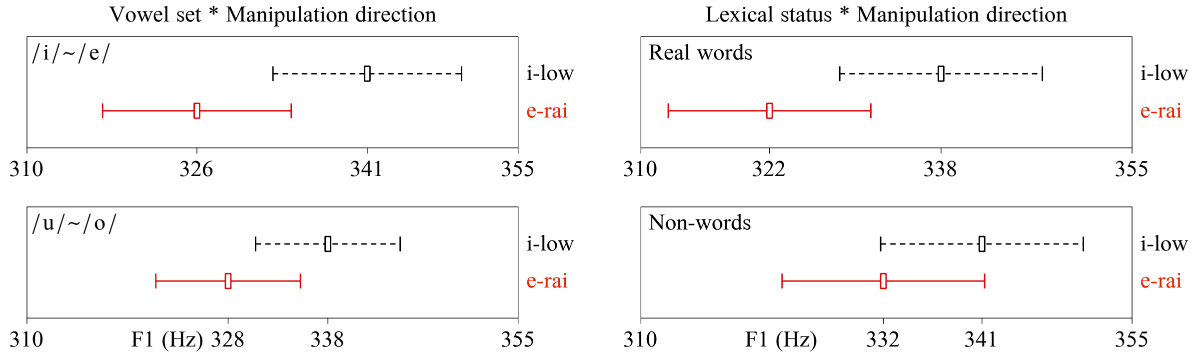

Figure 5*. Mean F1 locations of the mid-vs-high vowel boundary (small rectangles) and their 95% confidence intervals (whiskers) found for the group exposed to lowered /i/ (black dashed lines) and the group exposed to raised /e/ (red, or dark gray, solid lines). The left panel shows the mean boundary locations in the /i/~/e/ (top) and the /u/~/o/ vowel set (bottom). The right panel shows the mean boundary locations in real words (top) and non-words (bottom).*

4. Discussion

The present experiment investigated whether medium-term perceptual adaptation occurs for vowels, whether it generalizes to an unexposed vowel contrast, and whether it occurs in the absence of local context cues.

We found that short exposure to manipulated vowels during a lexical decision task led to a readjustment of the perceptual category boundaries in the expected direction in a subsequent identification task. The direction of manipulation had a significant effect on boundary location: Greek listeners exposed to lowered /i/s and typical /e/s had the mid-versus-high vowel boundaries at higher F1 values (i.e. lower in the vowel space) than listeners exposed to raised /e/s and typical /i/s. Furthermore, our data show a readjustment not only for the contrast between vowels /i/ and /e/, which had actually been manipulated and presented in the exposure, but also for the unexposed contrast between /u/ and /o/ (see Figures 4 and 5). This indicates generalization of perceptual

[4] As is seen in Figure 5 the confidence intervals of the /i/-/e/ and /u/-/o/ boundaries overlap to a somewhat larger extent in the group exposed to non-words than in the group exposed to real words. Table 4 shows that the effect size for the non-significant interaction of Manipulation direction * Lexical status is rather small, especially when one compares it to the relatively large effect of Manipulation direction, which we take as further support for our conclusion that recalibration occurs irrespective of whether listeners are exposed to real or nonexistent words.

adaptation from the manipulated front vowels to the unexposed back vowels of corresponding vowel height. In other words, listeners appear to have redefined their phonological mid-versus-high vowel boundary, and not just phoneme-specific boundaries.

Our results thus extend previous findings about the plasticity of vowel perception. First, we clearly show that perceptual recalibration of phoneme boundaries happens even for vowels. Note that vowels are speech sound categories that are acquired very early in life (already within the first six months; Kuhl, 2007). Yet, we showed that even for such long-established representations, adult listeners are able to make fine necessary perceptual adjustments only after very brief exposure. Second, research of immediate carrier-phrase-induced vowel perception adjustments found no generalization across the vowel space (Mitterer, 2006), and neither did the study testing medium-term lexically guided adaptation in Dutch vowels offer conclusive evidence of generalization (McQueen & Mitterer, 2005). Here we show that perceptual recalibration of vowel categories does generalize across phonologically related vowel contrasts, at least in a symmetrical vowel system, such as Greek.

The present findings that a modification of an auditory dimension triggered adaptation of phonological boundaries defined by categories such as the features high and mid bear relevance to theories of phonological representation. Our results suggest that there is a strong link between the auditory F1 dimension and a phonological property that distinguishes /i/ from /e/ and /u/ from /o/, which could well be called the feature of vowel height. The present result thus aligns well with theories in which the mental representations to which listeners link the phonetic information are phonological features (see e.g. Jakobson, Fant, & Halle, 1952; Stevens, 1989; Kingston, Diehl, Kirk, & Castleman, 2008; Chládková, Boersma, & Benders, 2015). At the same time, our finding that listeners adapted their boundary for a non-exposed contrast extends previous findings that cannot be interpreted in terms of strict exemplar views of phonology. Exemplar theories posit that listeners store and reuse the encountered realizations of words, which questions the existence of pre-lexical units (Johnson, 1997; Goldinger, 1998; Pierrehumbert, 2001). Such theories can account for the adaptation we found for the exposed /i/-/e/ contrast but not for its generalization to the non-exposed /u/-/o/ contrast unless some form of cross-category abstraction is assumed (McQueen et al., 2006a; Goldinger, 2007; Cutler, Eisner, McQueen, & Norris, 2010). In summary, our findings show that humans do not operate purely on the basis of amassed phonetic detail but relate ongoing phonetic inputs to a reduced set of abstract phonological properties that reflect systematic organization of speech sounds in language users' minds.

As noted in the Introduction, generalization to phonologically related contrasts in consonants has been attested for consonant voicing but not for place-of-articulation. Reinisch and colleagues (Reinisch et al., 2014; Reinisch and Mitterer, 2016) ascribed this difference to the fact that the voicing contrasts have acoustic cues found largely within the given segments themselves, which is not the case for place-of-articulation consonant contrasts, whose acoustic cues are located mainly in the neighboring segments. Our finding of generalization from a front to a back vowel-height contrast is not in disagreement with this account. The phonological features of a vowel, such as height, backness, or rounding, have straightforward correlates in the acoustic signal (such as F1 or F2 and F3) manifesting themselves within the vowel segment itself. Reinisch and Mitterer (2016) also emphasize the importance of acoustic similarity between the exposed segment and the segment to which recalibration generalizes. Our results suggest that such similarity may be of a considerably abstract form: our listeners could abstract away from the physical F1 shifts in the exposed manipulated /i/ or /e/ and adjust their perception of the unexposed /u/ and /o/ with analogous F1 differences. Therefore, even though there might be good reasons to question the role of phonological features in perceptual recalibration of consonantal place contrasts, some abstract (feature-like) property must underlie the generalization of perceptual adaptation for vowels that we observed. This is in line with Skoruppa and Peperkamp's (2011) conclusion that listeners draw feature-based inferences when learning accents with atypical vocalic patterning.

In the present data, we did not find an interaction between the direction of manipulation in the exposed vowels and the lexical status of the exposure (non)words. Thus, we have no evidence that perceptual adaptation only occurs with the involvement of contextual information (such as lexical knowledge) and not without it. The /i/-/e/ and /u/-/o/ boundaries seem to have been perceptually recalibrated irrespective of whether listeners were auditorily exposed to the manipulated vowels in real words or in non-words with no phonotactic cues to vowel identity (see Figure 5). Listeners who heard the ambiguous vowels in *real* words during the lexical decision task (i.e. during the exposure phase) could from the beginning of exposure associate the atypical sounds with the category cued lexically. In contrast, listeners who heard the atypical vowels in *non-words* had no lexical or phonotactic cues to help them assign the shifted sounds to the original vowel category. Still, a recalibration of that category seems to have taken place.

What could be the cause of the recalibration induced by non-words? Let us first consider the possibility that the effect was analogous to short-term speaker normalization, which is known to be independent of lexical access (Mitterer, 2006). However, our listeners were always exposed to (non)words containing the three Greek vowels /i/, /e/, or /a/ (representing all the height levels of the vowel system) only one of which was manipulated, /i/ or /e/, keeping the other two unchanged. Since the manipulation selectively affected the single vowel, rather than the height of all the exposed vowels, we do not ascribe the effect to a process of calibrating (or, normalizing) for the F1 range of the speaker at hand.

Rather, we propose that a statistical learning mechanism was at play. The vowel tokens that listeners heard during exposure were always sampled from two clouds in the perceptual space: one around a typical Greek /i/ and the other one midway between /i/ and /e/, or for another group of listeners, one around a typical Greek /e/ and the other one midway between /e/ and /i/. In the course of the exposure phase, listeners could therefore use the statistics of the sounds' distributions in the input as information about category identity: in the presence of the unmanipulated sounds near a typical /i/ or a typical /e/, the ambiguous [i/e] sounds could be associated with the competing category, whose canonical realizations never occurred. This explanation is in line with the attested effects of distributional learning demonstrating that speech sound comprehension can *continuously* adapt to updated distribution properties (Clayards et al., 2008; Kleinschmidt and Jaeger, 2015; Liu and Holt, 2015). We believe the present study is significant for theories that deal with category learning throughout the lifespan. It has been proposed that at least in early stages, young infants acquire the speech sounds of their language through an unsupervised learning mechanism (Maye & Gerken, 2000). As we argued above, such a learning mechanism in which no feedback or supervision from higher levels is involved seems to have been at play in our adult listeners, who adjusted their perception on the basis of the mere exposure to the atypical phonetic input. In that respect, our study supports the findings of a number of previous training experiments arguing that the cognitive learning mechanisms that are typically employed early in life seem to be available in adulthood (see e.g. Clayards et al., 2008; Escudero and Williams, 2015; but also Wanrooij et al., 2014 who did not find similar effects of training in adults and infants).

Our finding that medium-term perceptual adaptation happens irrespective of whether or not local cues are available is in disagreement with previous findings on consonants, which report the necessity of some forms of contextual cues (Samuel and Kraljic, 2009). Possibly, the present result could be specific to vowels. Vowels are acoustically louder, longer and their spectral properties are located in frequency regions were perceptual resolution is high, and in this sense they are auditorily more salient than consonants as a group. If artificial shifts of the acoustic properties of vowels are perceptually more salient than shifts in consonants, then perceptual recalibration for vowels may take place even in the absence of supporting information such as word meaning or phonotactics. Another potential explanation may lie in the fact that the mapping between auditory dimensions and phonological categories is more straightforward for vowels than it is for consonants. Unlike in vowels, most phonological contrasts in consonants are cued by multiple auditory dimensions: for instance, fricative contrasts are cued by several acoustic properties (Jongman, Wayland, & Wong, 2000) and the voicing contrast in English can be cued by as many as 17 different phonetic dimensions (Lisker & Abramson, 1964). Therefore, any change to a single auditory cue in vowels is likely to have more weight and can more robustly result in adaptation of the cue-mapping independent of the contexts.

Our results have implications for second-language (L2) speech sound acquisition. Recent research has demonstrated that lexically-guided perceptual recalibration happens in the L2 as well as in one's first language (Reinisch, Weber, & Mitterer, 2013; Drozdova, van Hout, & Scharenborg, 2014; Schertz, Cho, Lotto, & Warner, 2016), although it is not as robust (Drozdova, van Hout, & Scharenborg, 2015). One implication, which stems from our finding that perceptual adaptation can occur also in non-words, seems to be beneficial for L2 learners: they have the potential to learn to correctly perceive vowels and distinguish vowel contrasts in the second language even if they have not yet acquired the vocabulary, similarly to the way it happens in first-language acquisition (for a review see e.g. Kuhl, 2007). The other implication, arising from our finding that perceptual adaptation generalizes to non-exposed phonologically related vowels, can be beneficial as well as detrimental: learners may have the ability to generalize perceptual adjustments across phonologically related vowels, but that can be helpful only if they have correctly acquired the phonological properties of the L2 vowels, or if their first and second language have matching vowel systems.

Though unrelated to our research questions, the interaction between lexical status and vowel contrast (/i/-/e/ versus /u/-/o/) found in our experiment suggests that the perceived height of Greek /e/ depended on whether it was embedded in an existing word or a non-word. It has been reported previously that the height of the Greek mid vowels /e/ and /o/ is somewhat affected by stress (Lengeris, 2012; Nicolaidis & Sfakianaki, in press) and potentially also by speech tempo (see Fourakis et al., 1999). In addition, our results suggest that when meaningful context is available, listeners tend to identify /e/ with lower F1 values than when the vowel occurs in non-words. We can speculate that without the pressure of the lexicon listeners relax their height criterion and accept as /e/ also tokens with higher F1 values.

5. Conclusion

Our results demonstrate that the perceptual system of adult listeners shows a high degree of plasticity: unsupervised exposure to only a few dozen (non)words results in altered speech perception. Specifically, we corroborate previous findings that perception is malleable enough for recent exposure to non-canonical speech sounds to affect phoneme boundaries. We confirm that perceptual retuning of phoneme boundaries happens even for vowels. What is more, our experiment shows that such category boundary adjustment is not limited to the exposed vowels but it generalizes to vowels sharing a phonological feature (here, vowel height). Our results suggest that for such perceptual retuning to take place, local cues to the identity of the atypical sounds (the lexicon, phonotactic knowledge, lip-reading) are not a necessity, as long as listeners can use distributional information in the input.

6. Acknowledgments

During data analysis and writing KC was funded by the Netherlands Organization for Scientific Research (NWO) grant 446-14-012. The work of VJP was supported by the internal grant FPVC2014/23 from the Faculty of Arts, Palacký University Olomouc. The authors would like to thank Barbora Chládková and Eva Glykou for help with creating the lists of words and non-words, Diamantis Petropoulos–Petalas for recording the items, and Dirk Jan Vet for technical help with the recordings.

7. References

Bertelson, P., Vroomen, J., & De Gelder, B. (2003). Visual recalibration of auditory speech identification a mcgurk aftereffect. *Psychological Science*, 14(6), 592-597.

Boersma, P., & Weenink, D. (2015). Praat: doing phonetics by computer [Computer program]. Version 5.4.15, retrieved 1 August 2015 from http://www.praat.org/

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

Chládková, K., Boersma, P., & Benders, T. (2015). The perceptual basis of the feature vowel height. *Proceedings of ICPhS XVIII* 2015, Glasgow, article 711.

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

Clayards, M. A., Tanenhaus, M. K., Aslin, R. N., & Jacobs, R. A. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition*, 108 (3), 804–9. doi:10.1016/j.cognition.2008.04.004

Cutler, A. (2012). Native listening: The flexibility dimension. *Dutch Journal of Applied Linguistics,* 1(2), 169–187.

Cutler, A., Eisner, F., McQueen, J. M., & Norris, D. (2010). How abstract phonemic categories are necessary for coping with speaker-related variation. In C.Fougeron, B.Kühnert, M. D'Imperio, & N.Vallée (Eds.), *Laboratory phonology, 10*, Berlin: deGruyter (pp. 91-111).

Cutler, A., McQueen, J.M., Butterfield, S., & Norris, D. (2008). Prelexically-driven perceptual retuning of phoneme boundaries. *Proceedings of Interspeech 2008*, pp. 2056.

Drozdova, P., Van Hout, R., & Scharenborg, O. (2014). Phoneme category retuning in a non-native language. *Proceedings of Interspeech 2014,* pp. 553–557.

Drozdova, P., van Hout, R., & Scharenborg, O. (2015). The effect of non-nativeness and background noise on lexical retuning. *Proceedings of ICPhS XVIII* 2015, Glasgow, article 717.

Escudero, P., Benders, T., & Wanrooij, K. (2011). Enhanced vowel distributions facilitate the learning of second language vowels. *Journal of the Acoustical Society of America,* 130, EL206-EL212.

Escudero, P., & Williams, D. (2014). Distributional learning has immediate and long-lasting effects. *Cognition*, 133(2), 408-413.

Fourakis, M., Botinis, A., & Katsaiti, M. (1999). Acoustic characteristics of Greek vowels. *Phonetica*, 56, 28-43.

Ganong, W.F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance,* 6, 110–125.

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

Goldinger, S. D. (2007). A complementary-systems approach to abstract and episodic speech perception. In J. Trouvain W.J. Barry (Eds.), *Proceedings of the 16th international congress of phonetic sciences*. Saarbrucken: Pirrot Dudweiler (pp. 49-54).

Goudbeek M, Cutler A, & Smits R. (2008). Supervised and unsupervised learning of multidimensionally varying non-native speech categories. *Speech Communication*, 50, 109–25.

Greenspan, S.L., Nusbaum, H.C., & Pisoni, D.B. (1988). Perceptual learning of synthetic speech produced by rule. *Journal of Experimental Psychology: Learning, Memory and Cognition***,** 14, 421–433.

Jakobson, R., Fant, G., & Halle, M. (1952). *Preliminaries to speech analysis: the Distinctive Features and their Correlates*. Cambridge, MA: MIT Press.

Johnson, K. (1997). The auditory/perceptual basis for speech segmentation. *Ohio State University Working Papers in Linguistics,* 50, 101-113.

Jongman, A., Wayland, R., & Wong, S. (2000). Acoustic characteristics of English fricatives. *Journal of the Acoustical Society of America*, 108, 1252-1263.

Kingston, J., Diehl, R. L., Kirk, C. J., & Castleman, W. A. (2008). On the internal perceptual structure of distinctive features: The [voice] contrast. *Journal of Phonetics*, 36, 28–54.

Kleinschmidt, D.F., & Jaeger, T.F. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review,* 122, 148–203.

Kraljic, T., & Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology*, 51, 141-178.

Kraljic, T., & Samuel, A. G. (2006). Generalization in perceptual learning for speech. *Psychonomic bulletin & review*, *13*(2), 262-268.

Kuhl, P. K. (2007). Cracking the speech code: How infants learn language. *Acoustical Science and Technology*, 28(2), 71-83.

Ladefoged, P. (1989). A note on "Information conveyed by vowels". *The Journal of the Acoustical Society of America*, 85(5), 2223-2224.

Ladefoged, P., & Broadbent, D. E. (1957). Information conveyed by vowels. *The Journal of the Acoustical Society of America*, 29(1), 98-104.

Lengeris, A. (2012). Phonetic vowel reduction in Standard Modern Greek. In Z. Gavriilidou, A. Efthymiou, E. Thomadaki & P. Kambakis-Vougiouklis (Eds.), *Selected papers of the 10th International Conference on Greek Linguistics* (pp. 401-407). Komotini/Greece: Democritus University of Thrace.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: acoustical measurements. *Word*, 20(3), 384–422.

Liu, R., & Holt, L. L. (2015). Dimension-based statistical learning of vowels. *Journal of Experimental Psychology: Human Perception and Performance*, 41(6), 1783.

Maye, J., Aslin, R. N., & Tanenhaus, M. K. (2008). The weckud wetch of the wast: Lexical adaptation to a novel accent. *Cognitive Science*, *32*(3), 543-562.

Maye, J., & Gerken, L. (2000). Learning phoneme categories without minimal pairs. *Proceedings of the 24th Annual Boston University Conference on Language Development*, 2, 522–533.

Maye, J., Werker, J., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, 82, B101–B111.

McQueen, J. M., Cutler, A, & Norris, D. (2006a). Phonological Abstraction in the Mental Lexicon. *Cognitive Science*, 30, 1113–1126.

McQueen, J., & Mitterer, H. (2005). Lexically-driven perceptual adjustments of vowel categories. *Proceedings of PSP2005*, London, 233–236.

McQueen, J. M., Norris, D., & Cutler, A. (2006b). The dynamic nature of speech perception. *Language and Speech*, *49*(1), 101-112.

Mitterer, H. (2006). Is vowel normalization independent of lexical processing?. *Phonetica*, 63(4), 209-229.

Nicolaidis, K., & Sfakianaki, A. (in press). Acoustic characteristics of vowels produced by Greek intelligible speakers with profound hearing impairment I: Examination of vowel space. *International Journal of Speech-Language Pathology*. DOI: 10.3109/17549507.2015.1101155.

Norris, D., McQueen, J.M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology,* 47, 204-238.

Ong, J.H., Burnham, D., & Escudero, P. (2015). Distributional learning of lexical tones: a comparison of attended vs. unattended listening. *PLoS ONE* 10(7): e0133446. doi:10.1371/journal.pone.0133446

Pierrehumbert, J. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. L. Bybee & P. Hopper (Eds.), *Frequency and the Emergence of Linguistic Structure* (pp. 137-157). Amsterdam: John Benjamins.

Reinisch, E., & Mitterer, H. (2016). Exposure modality, input variability and the categories of perceptual recalibration. *Journal of Phonetics*, 55, 96–108.

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

Reinisch, E., Wozny, D., Mitterer, H., & Holt, L.L. (2014). Phonetic category recalibration: what are the categories? *Journal of Phonetics*, 45, 91–105.

Rosen S., Faulkner A., & Wilkinson L. (1999). Adaptation by normal listeners to upward spectral shifts of speech: implications for cochlear implants. *Journal of the Acoustical Society of America*, 106, 3629–3636.

Samuel, A. G., & Kraljic, T. (2009). Perceptual learning for speech. *Attention, Perception, & Psychophysics*, *71*(6), 1207-1218.

Schertz, J., Cho, T., Lotto, A., & Warner, N. (2016). Individual differences in perceptual adaptability of foreign sound categories. *Attention, Perception, & Psychophysics*, *78*(1), 355-367.

Sfakianaki, A. (2002). Acoustics Characteristics of Greek Vowels Produced by Adults and Children. In M. Makri-Tsilipakou (Ed.), *Selected Papers on Theoretical and Applied Linguistics from the 14th International Symposium* (pp. 383–394). Thessaloniki: Aristotle University.

Shannon R.V., Zeng F.G., Kamath V., Wygonski J., & Ekelid M. (1995). Speech recognition with primarily temporal cues. *Science*, 270, 303–304.

Sjerps, M. J., Mitterer, H., & McQueen, J. M. (2011). Constraints on the processes responsible for the extrinsic normalization of vowels. *Attention, Perception, & Psychophysics*, 73(4), 1195-1215.

Skoruppa, K., & Peperkamp, S. (2011). Adaptation to Novel Accents: Feature-Based Learning of Context-Sensitive Phonological Regularities. *Cognitive Science*, 35(2), 348-366.

Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics*, 17, 3–46.

Wanrooij, K., Boersma, P., & van Zuijen, T. (2014). Distributional vowel training is less effective for adults than for infants. A study using the mismatch response. *PLoS ONE* 9(10): e109806.

Wanrooij, K., Boersma, P., & Benders, T. (2015). Observed effects of "distributional learning" may not relate to the number of peaks. A test of "dispersion" as a confounding factor. *Frontiers in Psychology,* 6, article 1341.

Watkins, A. J. (1991). Central, auditory mechanisms of perceptual compensation for spectral-envelope distortiona). *The Journal of the Acoustical Society of America*, 90(6), 2942-2955.

Watkins, A. J., & Makin, S. J. (1994). Perceptual compensation for speaker differences and for spectral-envelope distortion. *The Journal of the Acoustical Society of America*, 96(3), 1263-1282.

Appendix

Critical words				Fillers	
Real /i/-words	English translation	Real /e/-words	English translation	Non-words with /a/	
Αθήνα	Athens	αίμα	blood	άκλασα	σκάλαρα
αντίκα	antique	δέρμα	skin	αμάλατα	σρανά
βύσμα	plug	ζακέτα	jacket	αρανά	τακαμάνα
καντίνα	canteen	καφές	coffee	δατάνα	τάλα
κατσαρίδα	cockroach	καναπές	sofa	γάντα	τάμακα
κλίμα	climate	κανέλα	cinnamon	θάκλα	τραλακά
κρασί	wine	καραμέλα	candy	θάμακα	τραμάλα
λαβίδα	tongs	καρέκλα	chair	κάρκα	τράμασα
λίβρα	pound	κέρατα	horns	κλάχαρα	τρανά
λίστα	list	κρέμα	cream	κλαμάρα	τραψά
μαγεία	magic	λέπρα	leprosy	κλαρασά	τσακανά
μαζί	together	μέντα	mint	κλαχαρά	φαθακά
μαυρίλα	blackness	νέκταρ	nectar	λάπα	φαμακά
μπύρα	beer	παρθένα	virgin	λαψάμα	φασάνα
νύχτα	night	πατέρας	father	νάτα	φαταλάσα
παγίδα	trap	πέρασμα	passage	ξάναλα	φραλά
παραλία	beach	πέτρα	stone	παθακά	χλαραπά
πρίσμα	prism	ρέγγα	herring	πλάματα	χράμασα
σίγμα	the letter “s”	ρεύμα	current	πλασακά	χρασακά
σήμα	signal	σαρδέλα	sprat	ρακάλα	χρατά
σταφίδα	raisin	σέλα	saddle	ρακάλασα	ψάτα
ταραχή	agitation	ταβέρνα	tavern	ρακάσα	
τρία	three	τέρμα	goal	ρακλασάμα	
φακίδα	freckle	τέχνασμα	trick	ραμάσα	
φύλακας	guard	φλέβα	vein	ραψά	
χαλί	carpet	φτέρνα	heel	σακαλά	
χαρτί	paper	χτένα	comb	σαλαμακά	
χρήματα	money	ψέμα	lie	σαπάλα	

Table A1. The list of real /i/- and /e/-words, and /a/-nonwords that were presented to the listeners who were exposed to manipulated vowels in real words.

Critical words		Fillers			
Non-words with /i/	Non-words with /e/	Real /a/-words	English translation	Real /a/-words	English translation
κλαχή	βραλέ	αδάμας	diamond	λάβα	lava
μίνατα	καλαπέ	άλας	salt	λάμπα	lamb
μλατίκανα	κλέρα	άναμμα	lighting up	μάζα	mass
ξαναλίνα	λέκανα	ανανάς	pineapple	μαμά	mum
ξίτα	μαναλέ	άνθρακας	carbon	μάρκα	brand
πατραμανί	μέσκα	αρακάς	peas	μπακλαβάς	baklava
πλίρα	μλατένακα	Άτλας	Atlas	μπαμπάς	daddy
ρακίλατα	πρατέλα	βλάκας	fool	μπανάνα	banana
ραχανίκα	ραδακέ	γάζα	gauze	ντάμα	queen
ρίλα	ρέκαταλα	γαλατάς	milkman	ξανά	again
τραλακατί	ταλαμακέ	γάτα	cat	Πάπας	pope
τραμαλί	ταραχέλα	γλάστρα	flowerpot	παπάς	priest
τρίμαλα	τέκα	γραβάτα	tie	πατάτα	potato
χαλακί	τέμαλα	δαγκάνα	grabber	πράγμα	thing
χαπαπλίσα	τραλέ	δράμα	drama	ράμπα	ramp
ψαλίρα	χακαπλέσα	ζάρα	wrinkle	ράτσα	breed
		θάλασσα	sea	σάλτσα	sauce
		θαμπάδα	blur	φάβα	split peas
		κάβα	liquor store	φαλάκρα	baldness
		κάλτσα	sock	φάντασμα	ghost
		καμάρα	arch	φράγμα	barrier
		κάπα	cape	χάντρα	bead
		κάρτα	card	χαρά	joy
		κατάρα	curse	ψαράς	fisherman

Table A2. The list of /i/- and /e/-nonwords, and real /a/-words that were presented to the listeners who were exposed to manipulated vowels in non-words.