Proceedings of ISCA Workshop on Plasticity in Speech Perception (PSP2005); London, UK; 15-17 June 2005

Short- and medium-term plasticity for speaker adaptation seem to be independent

Holger Mitterer

Max Planck Institute for Psycholinguistics
holger.mitterer@mpi.nl

Abstract

In a classic paper, Ladefoged and Broadbent [1] showed that listeners adapt to speakers based on short-term exposure of a single phrase. Recently, Norris, McQueen, and Cutler [2] presented evidence for a lexically conditioned medium-term adaptation to a particular speaker based on an exposure of 40 critical words among 200 items. In two experiments, I investigated whether there is a connection between the two findings. To this end, a vowel-normalization paradigm (similar to [1]) was used with a carrier phrase that consisted of either words or nonwords. The range of the second formant was manipulated and this affected the perception of a target vowel in a compensatory fashion: A low F2-range made it more likely that a target vowel was perceived as a front vowel, that is, with an inherently high F2. Manipulation of the lexical status of the carrier phrase, however, did not affect vowel normalization. In contrast, the range of vowels in the carrier phrase did influence vowel normalization. If the carrier phrase consisted of high-front vowels only, vowel categories shifted only for high-front vowels. This may indicate that the short-term and medium-term adaptations are brought about by different mechanisms.

1. Introduction

In studying how listeners compensate for the different formant-frequency ranges of different speakers, Ladefoged and Broadbent [1] presented listeners with target words with the structure bVt, and listeners had to decide whether the word was *bit* [bɪt], *bet* [bɛt], *bat* [bæt], or *but* [bʌt]. The target bVt was presented after the carrier phrase “please say what this word is”. The F1 and F2 range in the carrier phrase were manipulated. Listeners adapted to this change in range of formant frequencies. The same test word was more likely to be perceived as *bet* [bɛt] rather than *bit* [bɪt] if the F1 range in the carrier phrase was lowered. Lowering the F1 in the carrier phrase makes the F1 in the test word relatively higher, and a higher F1 is more appropriate for [ɛ] than for [ɪ]. This shows that listeners can adapt to a given speaker based on the short-term exposure of a carrier phrase.

Norris, McQueen, and Cutler [2] presented evidence for a medium-term adaptation to a particular speaker, which depends on lexical knowledge. Listeners were exposed to one of two lists of words and nonwords, and made lexical decisions to those items. One list contained twenty [f]-final words ending in an ambiguous fricative (midway between [f] & [s]) and twenty unambiguous [s]-final words, while the other list contained the same words but with the [f]-final words ending in unambiguous [f] and the [s]-final words ending with the same ambiguous fricative. A phonetic-categorisation task followed. Listeners exposed to the first list were more likely to perceive ambiguous fricatives on an [ɛf]-[ɛs] test continuum as [f] than listeners exposed to the second list. This perceptual-learning effect was found to depend on lexical knowledge, since it occurred if the ambiguous fricatives in the exposure phase were embedded in words but not if they were embedded in nonwords.

This lexically driven perceptual-learning effect is at least partly speaker-specific [3,4] and also occurs for vowels [5]. Therefore, the question rises whether vowel normalization in a Ladefoged-Broadbent paradigm [1] may also be influenced by lexical knowledge. Presently, the perceptual-learning paradigm—200-item exposure and test phase—has been conducted with the critical phones, to which the listeners should adapt, embedded in words and nonwords. An effect was only obtained if these critical phones occurred in words. It has not yet been explicitly investigated whether the lexical status of the carrier of the critical phones—in this case the vowels in the carrier phrase—in the short-term-exposure paradigm influences the degree of vowel normalization.

While some evidence suggests that vowel normalization may be determined by solely auditory processes [6], other data indicates that higher levels of processing also contribute to vowel normalization [7,8]. For instance, Johnson et al. [7] presented synthetic vowels and asked listeners to imagine that the vowels were produced by either a male or a female speaker. Listeners were more likely to label the same ambiguous vowel token as phonologically high (i.e. having a *low* F1) if they were asked imagine the vowel was produced by a female speaker rather than a male speaker. The listeners’ behavior is in line with a compensation for the higher F1 range of female speakers than male speakers, based on expectation alone. This suggests that vowel-normalization rests on both auditory and higher-level mechanisms. Accordingly, it not unlikely that lexical effects may also influence short-term adaptation in vowel normalization. It is, for instance, rather uninformative that a speaker produces an F2 at 2 kHz, because the F2 range of most speakers includes 2 kHz. If, however, the lexicon provides information that the F2 at 2 kHz occurred in an /i/, this indicates that this speaker has a below-average F2 range. Therefore, it is possible that the lexical status of the carrier phrase influences the degree of vowel normalization.

2. Experiment 1

In this experiment, vowel normalization was tested in target words that appeared in a carrier phrase that contained a wide variety of vowels (/u,i,ɑ,ɛ/). These vowels occurred in a carrier phrase that was either a meaningful sentence (*toen was hier* TARGET *gezegd* ’then was here TARGET said’) or a sequence of phonologically similar nonwords (n*oet fas tier* TARGET *ketegd* all Dutch nonwords). F2 values in the carrier phrase were either copied from natural utterances or in- or decreased by 20%, leading to low, medium, and high F2 carrier phrase. The subject had to perform a three-alternative forced-choice task and indicate whether the target was *keer* /ker/ 'time', *keur* /kør/ 'choice', or *koor* /kor/ 'choir'. If vowel-normalization is influenced by lexical status, one should expect a stronger effect of F2 range in the carrier phrase for the *word* carrier phrase than for the *non-word* carrier phrase.

2.1. Method

2.1.1. Participants

Eight native speakers of Dutch from the Max-Planck-Institute subject pool were paid for participation.

2.1.2. Materials

A male native speaker of Dutch was recorded saying multiple instances of the word and non-word carrier sentences containing one of the three different targets. For both target and carrier sentence used in the experiment, the consonants from the natural utterances were used and Klatt-synthesized vowels were spliced in. Synthesis parameters were estimated from the natural utterance. There were 11 different target vowels that formed the /ker/-/kør/-/kor/ continuum, all with an identical F1 starting with 380 Hz, going to 440 Hz at vowel midpoint and an endpoint of 550 Hz. F2 and F3 were manipulated as indicated in Table 1, with four intermediate steps between the displayed targets.

Table 1: Synthesis parameters for the extremes and the midpoint of target-vowel continuum in Hz.

Vowel (=Target)	Midpoint		Offset	
	F2	F3	F2	F3
/o/ (Target 0)	800	2350	1233	2350
/ø/ (Target 5)	1358	2350	1419	2350
/e/ (Target 11)	2200	2850	1700	2620

Parameters for the medium-F2 carrier phrase were based on the formant measurements in the natural utterance. For the low- and high-F2 version of the carrier phrase, F2 was in- or decreased by 20%. In order to prevent unnatural formant constellations, F3 was set at 1.2 times the F2 value, if the original F3 value was within a 20% range of the manipulated F2. Otherwise, F3 remained unchanged. With three versions of the vowels in the carrier phrase (low, medium, and high F2) and the two sets of consonantal portions (forming words and nonwords), this gives rise to six carrier phrases, in which the 11 targets occurred.

2.1.3. Procedure

Participants faced a computer screen with a four-button response box in front of them, with three buttons labeled *keer*, *keur*, and *koor*, respectively. After hearing a sentence, participants had to press one of the labeled buttons. It was stressed that the sentences could contain nonsense words.

Each of the 66 sentences—six carrier phrases crossed with eleven targets—was presented six times to each participant in a random order, with a short break after each 50 trials.

2.2. Results and Discussion

Figure 1 displays the aggregated proportion of perceived vowel frontness for each combination of Lexical Status (upper panel: words, lower panel: nonwords), F2 range (different symbols), and Target vowel (ordinate). The continuous lines represent the likelihood that the vowel was perceived as either /ø/ or /e/, that is more front than /o/, or 100% minus % /o/-responses. The dotted lines represent the likelihood of an /e/-response, that is more front than /ø/ and /o/. Accordingly, the areas under the dotted lines represent the proportion of /e/-responses, the areas between the dotted and the continuous lines the proportion of /ø/-responses, and the areas above the continuous lines the proportion of /o/-responses. The results show that the continuum endpoints were identified as intended as /o/ and /e/. Moreover, targets in the middle of the continuum are recognized almost exclusively as /ø/.

The data show a compensatory effect of F2 range on vowel identification: A lower F2 in the carrier phrase leads to more front-vowel responses, that is, an ambiguous F2 is more likely to be interpreted as “high” if it occurs in low-F2 carrier phrase. Ordinal-logistic-regression affirmed the statistical significance of these effects. More detailed analysis of the effect of F2 range for each individual target revealed a significant effect for the Targets 3 and 4 (/o/-/ø/ boundary) as well as the Targets 7 through 9 (/ø/-/e/ boundary).

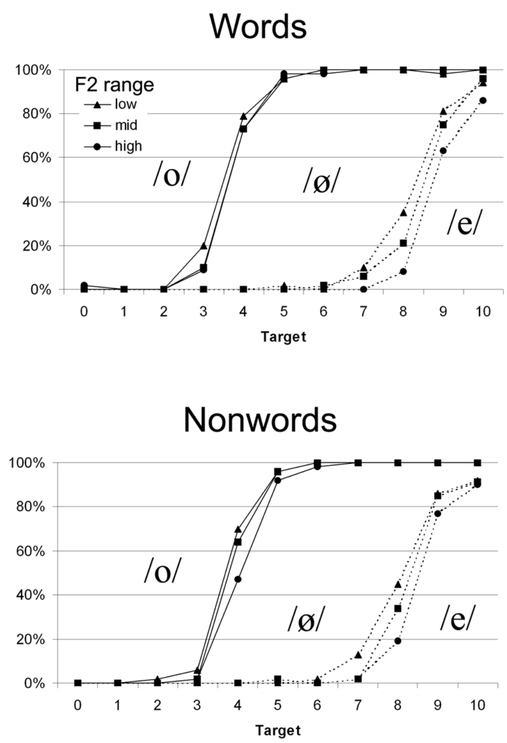

Figure 1: Proportion of vowel-responses in Experiment 1.

This formant-range effect replicates earlier results [1,6], which speaks for the validity of the data. Nevertheless, there is no effect of Lexical Status of the carrier phrase, the effect of F2-range was comparable in the word- and nonword-carrier conditions.

If vowel-normalization indeed rests on both auditory as well as higher-level mechanisms (see [6] and [7], respectively), the current experiment may, however, have been designed as to stack the deck in favor of signal-based normalization, and, as a consequence, leave little room for lexical effects to moderate the auditory effects. The carrier phrase contained the two point vowels [i] and [u], as well as the near-point vowel [ɑ]. This allows listeners to get a good estimate of both F1- and F2-range of the speaker, so that phonetic labels provided by the lexicon could not further help vowel normalization. A lexical effect may nevertheless be obtainable if the carrier phrase contains a narrower sample of vowels. I therefore ran another Experiment in which the carrier phrase contained high front vowels only. In this case, a signal-based strategy may not be completely successful if not informed about the phonetic labels of the high-front vowels it encounters. Embedding the vowel in words may provide the listener with those phonetic labels, which in turn could help the listener to interpret the small range of F1 and F2 encountered in the carrier phrase. Accordingly, a lexical effect on vowel normalization may occur if the carrier phrase contains only a small subset of the vowel space.

3. Experiment 2

The same targets as in Experiment 1 were used. The carrier phrases were changed to *weer is hier* TARGET *gezegd* (Word condition, 'again is here TARGET said') and *beeg it tier* TARGET *ketegd* (Nonword condition). These carriers only contained the front-high vowels[i],[ɪ],[e], and [ɛ]. (The schwa in *gezegd/ketegd* was deleted.)

3.1. Method

3.1.1. Participants

Ten native speakers of Dutch from the Max-Planck-Institute subject pool were paid for participation.

3.1.2. Materials and procedure

The same targets were used as in Experiment 1. The carrier sentence were generated from the natural consonantal parts with synthesized vowels spliced in. The formants estimated in the natural utterance were used as the medium F2 range carrier, and high and low F2-range carrier were generated by multiplying F2 by 0.8 and 1.2 respectively. F3 was only manipulated if necessary to prevent an overlap of F2 and F3 (cf. Experiment 1). Each of the 66 sentences—six carrier phrases crossed with eleven targets—was presented six times to each participant.

3.2. Results and Discussion

Figure 2 displays the aggregated proportion of perceived vowel frontness for each combination of Lexical Status (upper panel: words, lower panel: nonwords), F2 range (different symbols), and Target vowel (ordinate). The continuous lines represent the likelihood that the vowel was perceived as either /ø/ or /e/, that is more front than /o/, or 100% minus % /o/-responses. The dotted lines represent the likelihood that the vowel was perceived as /e/, that is more front than /ø/ and /o/. The data show a compensatory effect of F2 range on vowel identification: A lower F2 in the carrier phrase leads to more front-vowel responses, that is, an ambiguous F2 is more likely to be interpreted as "high" if it occurs in low-F2 carrier phrase. Ordinal-logistic-regression affirmed the statistical significance of these effects, and also showed that this effect was *not* modulated by the Lexical Status of the carrier phrase. More detailed analysis of the effect of F2 range for each individual target revealed a significant effect only for the Targets 7 through 10 (/ø/-/e/ boundary).

In comparison with Experiment 1, the current experiment replicates the vowel-normalization effect triggered by the F2 range in the carrier phrase. In contrast to Experiment 1, however, the F2 range does not have an influence on the /o/-/ø/ boundary, but only on the /ø/-/e/ boundary. This shows that vowel normalization is vowel-dependent. If the carrier phrase only consists of high-front vowels, phoneme boundaries are only adjusted for high-front vowels. Even if phonetic labeling of the altered vowels in the carrier phrase is supported by using words in the carrier phrase, listeners do not adjust all vowel boundaries, but only those in the vicinity of the vowels encountered in the carrier phrase.

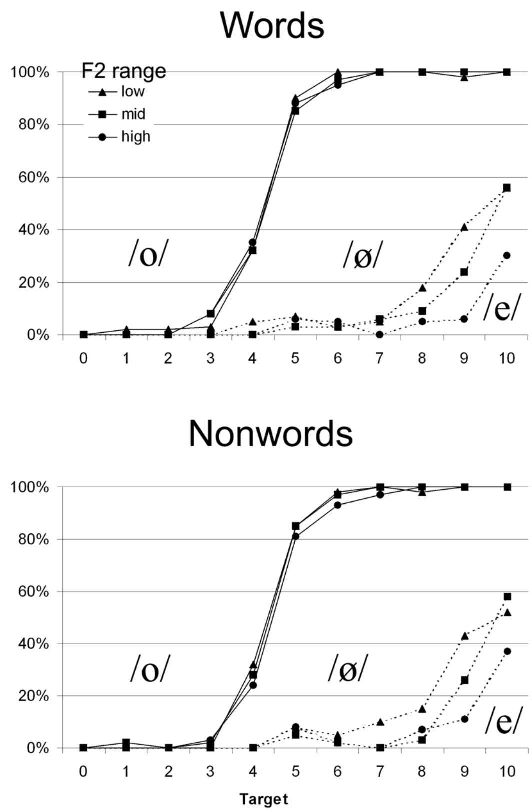

Figure 2: Proportion of vowel-responses in Experiment 2.

The clearest difference with the results of Experiment 1 is, nevertheless, the small percentage of /e/-responses in this experiment. While the percentages of /e/-responses approached ceiling level in Experiment 1, no more than 60% were observed in any of the cells in this experiment. This may be interpreted as a selective adaptation effect. It is a well-known finding that repeated exposure to a speech stimulus leads listeners to perceive a following stimulus as belonging to another category. This effect generalizes to similar speech sounds, so that adaptation to a voiced stops with one place of articulation leads listeners to perceive a following stop with a different place of articulation as unvoiced [9,10], independent of the place of articulation. In the current experiment, listeners heard three high-front vowels in the carrier phrase. Selective adaptation should accordingly lead listeners to perceive the target vowel as non high-front, which is exactly what is observed.

4. Discussion

In a series of two experiments, possible lexical effects on vowel normalization in the design introduced by Ladefoged and Broadbent [1] were tested. Previous investigations have shown that lexical effects play a role in medium term adaptation [2]. In Experiment 1, a carrier phrase with a diversity of vowels was used, and manipulating F2 in this carrier lead to a compensatory vowel-normalization effect on a F2-test continuum. Listeners were more likely to perceive a vowel which contains—within a given speaker's utterances—a lower F2, that is either /o/ or /ø/ rather than /e/ if the carrier phrase contained an elevated F2 range. In Experiment 2, the carrier phrase again consisted of either words or nonwords, but only contained high-front vowels rather than a diversity of vowels. This design change altered two aspects of the results. First of all, a vowel-normalization effect was only observed for high-front vowels. This indicates that vowel normalization is vowel-specific: If only high-front vowels are encountered in a carrier phrase, only phoneme boundaries of high-front vowels are adjusted. Secondly, presented a series of high-front vowels in the carrier phrase triggered selective adaptation, so that overall less high-front vowels were reported. This selective adaptation did, however, not obliterate the effect of F2-range in the carrier phrase on the /ø/-/e/ boundary.

The main purpose of the current experiments was to test the possible role of lexical status in the short-term adaptation paradigm [1], because , in the medium-term paradigm [2], adaptation only occurred if the critical phones were embedded in words. The results show that lexical knowledge is not necessary for and does not influence vowel identification in the short-term, vowel-normalization paradigm. It is important to note, however, that the manipulation of F2 range in the current experiment more or less resembles anatomically grounded speaker variation. It is possible that, if the manipulation of the carrier phrases differed in more idiosyncratic ways, resembling socio-phonetic variation, a lexical effect may nevertheless still be observed. Nevertheless, the current experiments show that the effects of formant range observed by Ladefoged and Broadbent [1] do not decrease if the carrier phrase consists of nonwords.

The current results also speak to the cue-weighting of intrinsic and external cues to vowel normalization. Nearey [11] conducted an experiment investigating the cue-weighting for intrinsic cues, such as f0 and higher formants, and extrinsic cues, such as second-formant range, to vowel-normalization. He concluded that both cues contribute to vowel normalization, but "it is clear the extrinsic ensemble effect dominates the changes" (p. 1201). The current results indicate that the cue-weighting for extrinsic factor is, however, dependent on the similarity of the vowels in the carrier phrase and the target vowel. Extrinsic factors may only play a role in vowel perception if the listener has been exposed to similar vowels previously, so that the cue-weighting strategies for extrinsic and intrinsic cues are in fact dynamic and not static.

5. Acknowledgements

The author wishes to thank James McQueen and Elizabeth Johnson for comments made on an earlier drafts of this paper.

6. References

[1] Ladefoged, P. and Broadbent, D. (1957). "Information conveyed by vowels" *J. Acoust. Soc. Am., Vol. 29*, 1957, 98-104.

[2] Norris D., McQueen J.M., and Cutler A, "Perceptual learning in speech", *Cogn. Psych., Vol. 47, 2003, 204-238.*

[3] Eisner, F. and McQueen, J.M., "The specificity of perceptual learning in speech processing", *Perc. & Psychophys, in press*.

[4] Kraljic, T. and Samuel, A.G., "How general is perceptual learning for speech?", manuscript submitted for publication.

[5] McQueen, J. & Mitterer, H., "Lexically-driven perceptual adjustments of vowel categories", *this volume*.

[6] Watkins, A. J., and Makin, S.J., "Perceptual compensation for speaker differences and for spectral-envelope distortion," *J. Acoust. Soc. Am., Vol. 96, 1994, 1263-1282*.

[7] Johnson, K., Strand, E.A., and D'Imperio, "Auditory-visual integration of talker gender in vowel perception," *J. Phonetics, 27*, 1999, 359-384.

[8] Evans, B.G. and Iverson, P., "Vowel normalization for accent: An investigation of best exemplar locations in northern and southern British English sentences", *J. Acoust. Soc. Am, Vol. 115,* 2004, 352-361.

[9] Eimas, P.D. and Corbit, J.D., "Selective adaptation and linguistic feature detectors.", *Cogn. Psych., Vol. 4,* 1973, 99-109.

[10] Remez, R.E., "Neural models of speech perception" in S. Harnad (Ed.), *"Categorical perception: The groundwork of cognition"* (pp.199-225). Cambridge, NY: Cambridge University Press.

[11] Nearey, T.D., "Static, dynamic, and relational properties in vowel perception," *J. Acoust. Soc. Am., Vol. 85,* 1989, 2088-2113.