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

Lexically-driven perceptual adjustments of vowel categories

James M. McQueen & Holger Mitterer

Max Planck Institute for Psycholinguistics
Nijmegen, The Netherlands
james.mcqueen@mpi.nl, holger.mitterer@mpi.nl

Abstract

We investigated the plasticity of vowel categories in a perceptual learning paradigm in which listeners are encouraged to use lexical knowledge to adjust their interpretation of ambiguous speech sounds. We tested whether this kind of learning occurs for vowels, and whether it generalises to the perception of other vowels. In Experiments 1 and 2, Dutch listeners were exposed during a lexical decision task to ambiguous vowels, midway between [i] and [e], in lexical contexts biasing interpretation of those vowels either towards /i/ or towards /e/. The effect of this exposure was tested in a subsequent phonetic-categorisation task. Lexically-driven perceptual adjustments were observed: Listeners exposed to the ambiguous vowels in /i/-biased contexts identified more sounds on an [i]-[e] test continuum as /i/ than those who heard the ambiguous vowels in /e/-biased contexts. Generalisation to other contrasts was weak and occurred more strongly for a distant vowel contrast (/ɑ/ vs. /ɔ/,) than for a near contrast (/ɪ/ vs. /ɛ/). In Experiment 3, spectral filters based on the difference between the exposure [i] and [e] sounds were applied to test stimuli from all three of the contrasts. Identification data of these filtered stimuli suggest that generalisation of learning across vowels does not depend on overall spectral similarity between exposure and test vowel contrasts.

1. Introduction

Lexical knowledge is able to influence adaptation of consonant categories in response to unusual pronunciations [1]. 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 in exactly 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 fricative in the exposure phase was embedded in words but not if it was embedded in nonwords.

Listeners can thus use lexical knowledge to adjust consonant categories while listening to a speaker who produces unusual tokens of those sounds. Such adjustments are useful for the listener since they make recognition of subsequent utterances by the same unusual speaker easier. Unusual speech occurs for several reasons, including socio-phonetic variation. Dialect differences are often carried by vowels, however. We therefore used the paradigm from [1] to test whether a lexically-driven learning effect could be obtained for vowels.

We also tested whether there is generalisation to previously untrained vowels. On the one hand, because socio-phonetic variation often is vowel-specific [2], one might predict that adaptations to a given vowel category pair should not lead to re-shaping of the complete vowel space, since this would not improve perception. One might therefore not expect generalisation to untrained vowel pairs. On the other hand, research with consonants using this exposure-test paradigm suggests that generalisation may depend on spectral similarity between the exposure and test stimuli [3,4,5]. Generalisation over the vowel space may therefore occur, as a function of spectral similarity.

There was one important modification of the exposure-test paradigm. Vowel acoustic shape varies substantially with phonetic context, unlike the reasonably stable characteristics of fricatives and stops used previously [1,3-5]. Figure 1 shows, for example, that spectra at vowel midpoints for the Dutch /i/-/e/ contrast in an alveolar [lVt] context differ from those in a velar/uvular [kVɹ] context (examples from the stimuli used in Experiments 1&2). To create natural-sounding stimuli we were thus forced to make different ambiguous vowels for each of the critical lexical contexts, in contrast to [1], for example, where exactly the same ambiguous fricative was used in all exposure contexts. This allowed us to test for a different kind of generalisation, namely, whether learning would emerge given a range of different acoustic tokens of the same phonological vowel contrast.

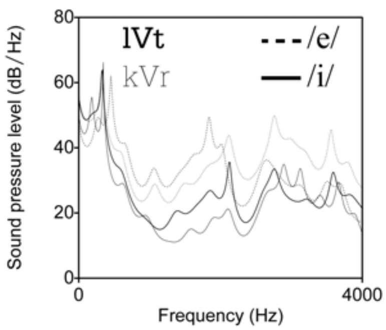

Figure 1: The LPC-smoothed spectra of the midpoints of the vowels /i/ (solid lines) and /e/ (dotted lines) in two contexts: alveolar (thick lines) and velar/uvular (thin lines).

2. Experiments 1 and 2

In the first two experiments, listeners were exposed to a list of words including twenty words containing /i/ in their final syllable and twenty matched words with /e/ in their final syllable. One group of listeners heard ambiguous vowels, midway between [i] and [e], in the /i/-words (e.g., *satelliet*, satellite, [satəlit]; [satəlet] is not a Dutch word), plus unambiguous versions of the /e/-words (e.g., *atleet*, athlete, [atlet]; again [atlit] is a Dutch nonword). A second group of listeners heard the reverse. After lexical-decision exposure, listeners performed a two-alternative forced-choice (2AFC) task on three continua: the trained contrast /ift/ vs. /eft/, a near-transfer contrast /ɪft/ vs. /ɛft/, and a far-transfer contrast /ɑft/ vs. /ɔft/. *Near* and *far* were defined by distances in F1/F2 space. Experiment 2 was identical to Experiment 1, except that the test continua were tested in a different order.

2.1. Method

2.1.1. Participants

92 native Dutch speakers from the MPI subject pool (aged 17 to 28) were paid to participate: 12 in the pre-test, 48 in Experiment 1 and 32 in Experiment 2.

2.1.2. Materials

We identified 20 pairs of polysyllabic words, in which the last syllable differed only in the vowel, which was either /i/ or /e/ (e.g., /bɑŋ.kir/, banker, & /vɛr.ker/, traffic): *parodie/procede; saffier/atmosfeer; alhier/beheer; hypocriet/concreet; bankier/ verkeer; kopie/coupé; satelliet/atleet; compromis/waarmee; stramien/fenomeen; apathie/paté; harmonie/abonnee; genie/ tournee; galerie/carré; seniel/rationeel; manier/meneer; graniet/magneet; ontzien/obsceen; steriel/tafereel; parasiet/ asceet; boetiek/apotheek*. In all 40 items a word could not be created by substituting one critical vowel for the other. These pairs were recorded by a male native speaker of Dutch, along with 160 fillers (60 words & 100 nonwords) that contained no high front vowels, and the endpoints of the test continua. The final syllables were excised from the 40 critical stimuli; mean duration and f_0 contour were equalized within each syllable pair. Eleven-step continua were created by digitally mixing the two waveforms in different proportions. This method thus captures at least some of the non-local cues to the vowel distinction. All 20 continua were presented in nonword contexts (e.g., [bɔr.kVr]) in a 2AFC identification pre-test. The most ambiguous syllable was chosen in each case.

The same procedure was used to create and select test continua for all three contrasts (*trained:* [ift]-[eft]; *near:* [ɪft]-[ɛft]; and *far:* [ɑft]-[ɔft]); all six of the endpoints of these continua are nonwords in Dutch. We chose test stimuli which, in the pre-test, were perceived as the higher member of the vowel opposition (i.e., /i/, /ɪ/, and /ɔ/) in 10%, 30%, 50%, 70%, and 90% of trials.

2.1.3. Procedure

In both experiments listeners first performed a lexical-decision task with the 40 critical items randomly mixed among the 160 fillers. Half of the participants in each experiment heard the 20 /i/-words with unambiguous final syllables and the 20 /e/-words with ambiguous final syllables (e.g., [bɑŋkir] & [vɛrkXr], where X denotes an ambiguous vowel), while the half heard the reverse (e.g., [bɑnkXr] and [vɛrker]).

After exposure, listeners performed 2AFC tasks on the 5-step [Vft] nonword-nonword continua. The two response alternatives in each block were the endpoint vowels; they were specified on a computer screen on each trial. Each block consisted of 6 repetitions of the 5 sounds in the following orders:

- Exp. 1: 1. [i]-[e], 2.[ɪ]-[ɛ], 3. [ɑ]-[ɔ], 4. [i]-[e]
- Exp. 2: 1. [ɪ]-[ɛ], 2. [ɑ]-[ɔ], 3. [i]-[e], 4. [i]-[e], 5. [ɪ]-[ɛ]

Furthermore, the order of presentation of the transfer contrasts (*near*: [ɪ]-[ɛ], *far*: [ɑ]-[ɔ]) in Blocks 2 and 3 in Experiment 1 and in Blocks 1 and 2 in Experiment 2 was reversed for half of the participants.

2.2. Results and Discussion

Analysis of the lexical decision data showed that the lexical manipulation was effective. Tokens with ambiguous vowels were mainly identified as words, but to a lesser degree than the unaltered words (Exp. 1: 86% vs. 94%; Exp. 2: 77% vs. 85%).

Logistic regression analysis of the identification data (see Figure 2) indicated that exposure conditions influenced the perception of the *trained* contrast to a statistically significant degree on the second test in Exp. 1 (Block 4) and the first test in Exp. 2 (Block 3). Only one statistically significant transfer effect was observed, on the *far* contrast in Exp. 2.

The results thus indicate that lexically-driven perceptual learning can be obtained with vowels, even if the exposure items differ in their acoustic make-up, and can generalise to other vowel contrasts. We find some puzzling order effects, however. In Exp. 1, the effect of exposure on the *trained* contrast is significant when tested the second time (in Block 4), but not when tested right after exposure. But in Exp. 2, there is no effect on the *trained* contrast in Block 4, while there is an effect in Block 3. These effects therefore cannot be attributed simply either to passage of time or to repetition of the *trained* contrast. It is possible that identification of

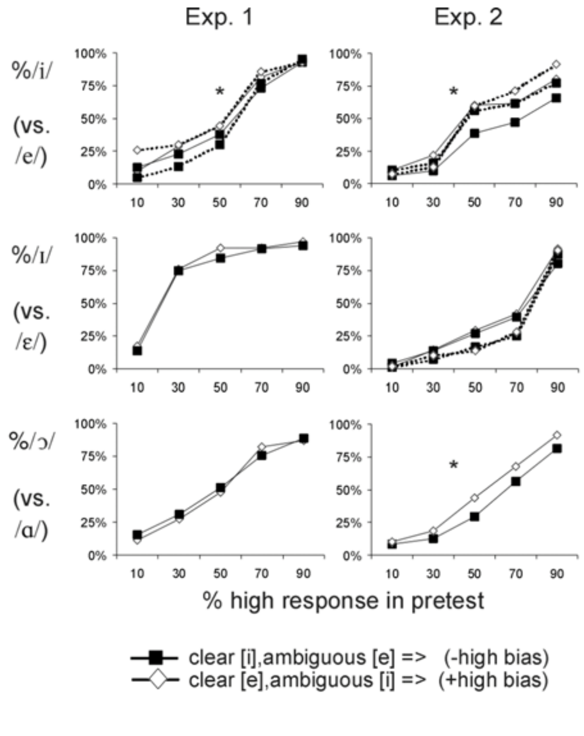

*Figure 2: Results from the identification task in Experiments 1 and 2. Dotted lines show results for repeated blocks; * indicates significant group effects.*

other vowel contrasts from the same speaker gave listeners more information about the speaker's vowel space and thus about the applicability of what was learned in the exposure phase. This may be why the learning effect on the trained contrast in both experiments is more pronounced after testing on the *near* and *far* contrasts.

More detailed analysis revealed some further effects of the order of presentation of the different test continua. As noted, the order of presentation of the two transfer continua was counterbalanced over participants. In Exp. 1, the participants tested on [ɑft]-[ɔft] in Block 2 showed the same training effect as was observed on these stimuli for all participants in Exp. 2. This effect, however, was counteracted by an opposite effect produced by the participants tested on [ɑft]-[ɔft] in Block 3 in Exp. 1. Furthermore, in Exp. 2, listeners who started with the [ɪft]-[ɛft] test continuum showed an effect in the expected direction (i.e., more /ɪ/ responses after clear [e], ambiguous [i] exposure), but the other participants (i.e., those who receieved these stimuli in Block 2) showed a reverse effect. Further research is required to ascertain which factor(s) cause the variability of the effects over test blocks on both the trained and transfer contrasts.

Nevertheless, we did find at least some evidence of generalisation. But this evidence is not easily explained in terms of distance in vowel space, because transfer was stronger for the *far* than for the *near* contrast. One possible explanation for this apparently paradoxical pattern is that the *far* contrast is in some way more similar to the *trained* contrast than the *near* contrast is. Spectra of the unambiguous vowels in the exposure phase (averaged over all 20 items) and of the test continuum endpoints are displayed in Figure 3.

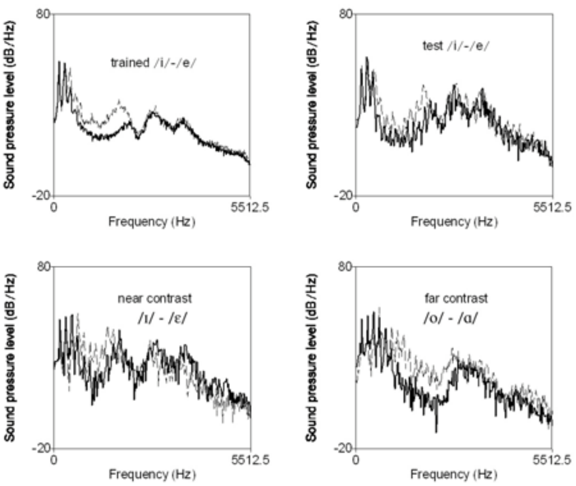

Figure 3: Spectra of the endpoint vowels in the training and test phases of Experiments 1 and 2. The thick lines represent the spectra of the higher vowels within each pair.

The *trained* and *far* contrasts show the same kind of difference in spectral shape: less energy in the 1-2 kHz region for the higher vowel. In contrast, the *near* contrast is defined by differences in the centre frequency of F2. Some theories of vowel perception emphasis the importance of spectral shape rather than formant frequencies as a major determinant of vowel perception ([6]). It is therefore possible that spectral similarity may drive the seemingly paradoxical transfer effects. This possibility was tested in Experiment 3.

3. Experiment 3

Filtering techniques were used — as it were — to test directly the impact of the difference in spectral shape of the training items from the earlier experiments on identification of the previous test stimuli. To this end, we generated a series of spectral filters based on the difference between the average exposure [i] and the average exposure [e]. These filters were applied to the 30% and 70% stimuli (i.e., steps 2 and 4) of all three test continua (see [7] for a similar method). If spectral similarity between exposure and test items determined whether perceptual learning occurred in Experiments 1 and 2, the difference filter should influence identification of stimuli from the *trained* contrast strongly, those from the *far* contrast less, and those from the *near* contrast least of all.

3.1. Method

3.1.1. Participants

12 native speakers of Dutch from the MPI subject pool were paid for their participation.

3.1.2. Materials

The six base stimuli (i.e., the 30% and 70% stimuli from the three continua) were filtered with [i] minus [e] filters (i.e., the spectral difference between the average exposure [i] and the average exposure [e]) in 9 different forms ranging from 100% to –100% in 25% steps. Figure 4 shows the ±100% and ±50% versions of the filter.

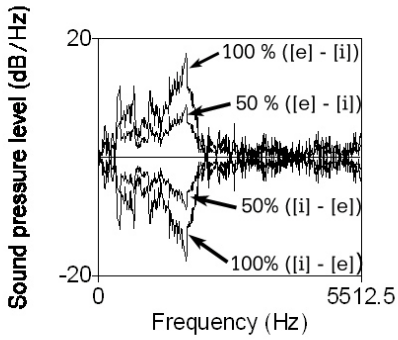

Figure 4: A subset of the filters used to generate the continua in Experiment 3.

3.1.3. Procedure

Six blocks were defined, one for each base stimulus. Within each block, each of the 9 filtered versions of the base stimulus was presented 10 times, in random order. Order of blocks was rotated over participants. The two response alternatives in each block were the endpoint vowels, which were specified on a computer screen on each trial.

3.2. Results and Discussion

Figure 5 displays the percentage of high-vowel choices for all 6 base stimuli, filtered in the nine different ways. The results show that applying the spectral difference of the exposure stimuli to the test continua leads to a perceptual switch for the trained contrast and the near contrast, but not the far contrast. It might therefore appear that spectral similarity cannot explain the pattern of generalisation found in Experiments 1 and 2. There is a potential difference between the earlier experiments and Experiment 3, however. It is possible that, in the perceptual-learning paradigm, the exposure conditions drew listeners attention to specific spectral differences within the 1-2 kHz region. In Experiment 3, however, the filter applied equally to all spectral differences, irrespective of frequency band. That is, the filtering in Experiment 3 may not have been specific enough to mimic the adjustments that were made in the learning experiments.

Experiment 3 certainly shows that spectral similarity can determine degree of change in vowel identification. For the *trained* contrast, because the filter was based on those vowels, filtering created the strongest effect on identification performance. For the *near* contrast, because of the overall similarity of those vowel's spectra to the trained contrast spectra, the filter still created an effect on identification, but a weaker one. But for the *far* contrast, because the overall spectral shape of the endpoint vowels is less like that of the trained vowels, filtering failed to create any effect on identification. It remains to be seen whether more specific filters could result in a stronger shift in vowel identification for the *far* than for the *near* contrast, and thus whether the generalisation effects in the perceptual-learning paradigm can indeed be explained as a function of fine-grained spectral similarity between the exposure and test vowel contrasts.

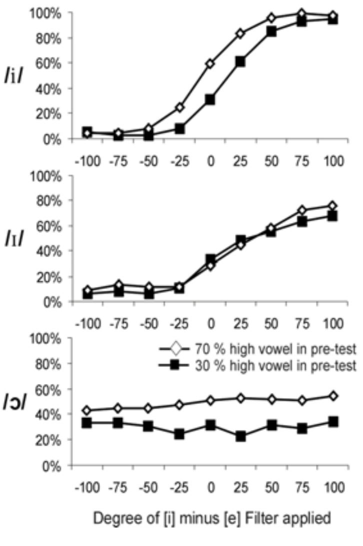

Figure 5: Identification results of Experiment 2.

4. General Discussion

Exposure to ambiguous vowels in lexically-biased contexts leads to adjustment of vowel categories. This result extends previous findings using a similar exposure-test paradigm in two ways. First, we have shown that lexically-driven perceptual learning is not limited to consonant contrasts. Previous studies have examined the place distinction in voiceless fricatives and the voicing distinction in stops [1,3-5]. Fricatives and stops have rather consistent acoustic properties over different phonetic contexts. In contrast, vowels are strongly coarticulated with their phonetic context [8]. We included this variability in the exposure stimuli. Nevertheless, similar perceptual-learning effects were found. Second, therefore, we have shown that perceptual learning is not limited to the interpretation of specific acoustic tokens. Instead, it appears to be learning about a more abstract phonological category distinction.

Some of our results, however, are truly puzzling. For example, the effect of exposure on the *trained* contrast did not appear in all blocks testing this contrast. It is possible that listeners were able to map the vowel space of the speaker better after identification of the *near* and *far* contrasts; this may be why the exposure effect was stronger on the second testing of the *trained* contrast in Experiment 1. But this leaves unresolved several other aspects of the variability of the effects over test blocks, in particular why we found an effect in the first test block in Experiment 2 on the *far* contrast. The patterns across blocks are not yet fully interpretable.

Finally, we found only limited evidence for generalisation of the learning to other vowels. Generalisation was not determined simply by distance in vowel space. Spectral analysis of the stimuli from Experiments 1 and 2 suggested that spectral similarity between exposure and test vowels might determine whether there is generalisation to untrained vowel contrasts. But Experiment 3 showed that overall spectral similarity between exposure and test vowel contrasts is not the reason why stronger generalisation was found for the *far* contrast than for the *near* contrast. It may be that spectral similarity in a more limited frequency range than was tested in Experiment 3 could nevertheless determine whether what is learned about the exposure vowel contrast is applied in the identification of other vowel contrasts.

5. References

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

[2] 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.*

[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] Kraljic, T. and Samuel, A. G., "Perceptual learning for speech: Is there a return to normal?", manuscript submitted for publication.

[6] Ito, M., Tsuchida, J., and Yano, M., "On the effectiveness of whole spectral shape for vowel perception", *J. Acoust. Soc. Am., Vol. 110, 2001, 1141-1149.*

[7] 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.*

[8] Lindblom, B. and Studdert-Kennedy, M.. On the role of formant transitions in vowel recognition. *J. Acoust. Soc. Am., Vol. 42, 1967, 830-843.*