NO LEXICALLY-DRIVEN PERCEPTUAL ADJUSTMENTS OF THE [x]-[h] BOUNDARY

Michaël A. Stevens[1], *James M. McQueen*[2], *& Robert J. Hartsuiker*[1]

[1]Ghent University, Belgium; [2]Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
michael.stevens@ugent.be, james.mcqueen@mpi.nl, robert.hartsuiker@ugent.be

ABSTRACT

Listeners can make perceptual adjustments to phoneme categories in response to a talker who consistently produces a specific phoneme ambiguously. We investigate here whether this type of perceptual learning is also used to adapt to regional accent differences. Listeners were exposed to words produced by a Flemish talker whose realization of [x] or [h] was ambiguous (producing [x] like [h] is a property of the West-Flanders regional accent). Before and after exposure they categorized a [x]-[h] continuum. For both Dutch and Flemish listeners there was no shift of the categorization boundary after exposure to ambiguous sounds in [x]- or [h]-biasing contexts. The absence of a lexically-driven learning effect for this contrast may be because [h] is strongly influenced by coarticulation. As [h] is not stable across contexts, it may be futile to adapt its representation when new realizations are heard.

Keywords: Perceptual learning, speech perception, regional accents

1. INTRODUCTION

When we listen to someone speaking, we have to deal with their idiosyncrasies in how they realize speech sounds. Part of the adaptation to specific talkers is achieved by a lexically-driven retuning of phoneme boundaries [8]. Norris et al. [8] used an exposure-test paradigm, in which listeners first heard a talker who produced an ambiguous fricative that was equally similar to [f] and [s]. One group of listeners was lexically-biased to interpret the ambiguous sound as [f] (they heard e.g. [wɪtlo?]; from *witlof*, 'chicory'; *witlos* is meaningless). The other group was biased to interpret it as [s] (they heard e.g. [na:ldbo?]; from *naaldbos* 'pine forest'; *naaldbof* is meaningless). In a subsequent test, the [f]-biased group categorized more sounds on a [f]-[s] continuum as [f] than the [s]-biased group. Listeners can thus compensate for the idiosyncratic production of a speech sound by shifting their phoneme boundaries after exposure to that sound in lexical contexts which determine that sound's identity.

The experiments reported here investigate whether this mechanism is also used to adjust to the systematic idiosyncrasies of speech sounds in different varieties of a language. Previous studies have shown that listeners adapt quite rapidly to foreign accents [1] and regional accents [4], but little is known about the processes by which this adaptation is achieved. One underlying process seems to be adjustment of vowel categories in response to talkers with different accents [2]. Londoners, who regularly interact with speakers of different accents, choose different realizations of a vowel as the best exemplar of that vowel when they hear the vowel in a sentence produced in a northern or southern English accent. People who are not familiar with the different accents do not make these adjustments. What drives these phoneme category adjustments is unclear. Are they driven by lexical information, like the adjustments made in [8]? Research has shown that adaptations made to one talker's speech are not readily applied to sounds uttered by another talker [3]. For example, adaptations can occur from a female talker to a male talker but not in the other direction [5], or across talkers for stops but not fricatives [6]. This lack of generalization makes sense because over-generalization of knowledge about how one talker speaks could impair recognition of another talker. In the case of accented speech, however, generalization to other talkers with the same accent would be appropriate.

To test whether such accent-specific perceptual adjustments are made, listeners from The Netherlands were exposed to words produced by a Flemish talker (Dutch spoken in The Netherlands and Flemish are two varieties of the same language). For one group of listeners, all instances of [x] during exposure were replaced by a sound between [x] and [h]. For a second group, all instances of [h] were replaced by that sound. Categorization responses to [x]-[h] sounds spoken by a second Flemish talker will show whether Dutch listeners handle the strange [x] or [h] as an idiosyncrasy of the exposure talker as an individual, or as a property of his Flemish accent. The use of the [x]-[h] contrast has the advan-

tage over the often-used [f]-[s] contrast in that the realization of the fricative sound at the beginning of a word such as *gek* (crazy) differs between Dutch and Flemish talkers. Flemish has a voiceless velar fricative (the "zachte g", [x]), while Dutch - at least that spoken in the north of the Netherlands - has instead a uvular fricative (the "harde g", [χ]). Furthermore, in the regional accent of West Flanders, there is virtually no [x]-[h] contrast, that is, words such as *gek* (crazy) and *hek* (fence) sound the same ([hɛk]). A talker who appears to produce an ambiguous fricative sound that is midway between [x] and [h] is therefore plausibly a Flemish talker whose accent is adjusting towards that of West Flanders.

2. EXPERIMENT 1

2.1. Method

2.1.1. Participants

Thirty-five native speakers of Dutch who were born and raised in The Netherlands were paid to take part; 18 were assigned to the within-talker condition, and 17 to the between-talker condition.

2.1.2. Stimulus construction

Twenty words containing just one [x] and no [h]'s and twenty words containing just one [h] and no [x]'s were selected from the Celex database. All were nouns or adjectives; average frequency was 12.5 per million. The words were selected so that the [x] or [h] occurred at the beginning of the second or third syllable of two- and three-syllable words. We could not select words that ended in the critical fricative because [h] does not occur at word endings in Dutch. Sixty filler words with the same lexical properties as the targets were selected and 100 filler nonwords were constructed to complete the stimulus set for the lexical decision task.

The stimuli were recorded by a male talker with a clear Flemish accent. Three versions of each target word were recorded: with the [x] or [h] sound pronounced as [x], as [h] and as [s]. The [x] and [h] versions of the fricative were cut out and equalized in length and F0 before they were mixed together to form the ambiguous fricative, which was spliced into the [s] context. The [s]-version was used as a carrier to avoid coarticulatory cues to either [x] or [h] [3]. Using this method an ambiguous fricative was constructed for each target separately. A natural version of each target was made in a similar way.

For the categorization test, tokens of the utterances [dɛtxu] and [dɛthu] were recorded by the talker who spoke the exposure stimuli and by another talker with a Flemish accent. Two [x]-[h] continua were made by digitally mixing each talker's natural fricatives in 101 different proportions. These fricatives were then put back into the [dɛt?u] context, and 7 steps from each speaker's continuum were selected using a pilot experiment.

2.1.3. Procedure & Design

There was a pretest, an exposure phase and a posttest. During the pre- and posttest, the 7 steps on the [dɛtxu]-[dɛthu] continuum from one of the talkers were presented once in a random ordering for 6 consecutive blocks. Listeners had to categorize the fricative they heard as either [x] or [h]. Each listener only heard one talker in the test phases, either the exposure talker (the *within-talker condition*) or the other talker (the *between-talker condition*).

In the exposure phase, 20 natural and 20 ambiguous targets were spread evenly across 200 trials of the lexical decision task. The listeners in the [x]-bias condition heard natural [h] sounds in the [h]-words and ambiguous sounds in the [x]-words. In the [h]-bias condition this was reversed: the [h]-words contained an ambiguous sound and the [x]-words contained a natural [x].

2.2. Results

2.2.1. Lexical decision

The results are shown in Table 1. Although responses to ambiguous targets were slower and less accurate than the reactions to natural targets, the lexical manipulation was effective: targets with ambiguous sounds were mainly identified as words.

Table 1: Lexical decision data, Experiments 1 & 2: Mean RT (ms, from word offset) and mean proportion correct responses.

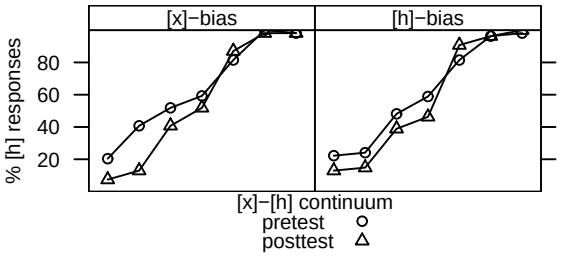
	filler words	natural targets	ambiguous targets
RT, Expt. 1	308	309	367
p(corr), Expt.1	.985	.977	.938
RT, Expt. 2	250	239	254
p(corr), Expt. 2	.989	.987	.981

2.2.2. Categorization, within-talker condition

The categorization responses were analyzed using a logistic regression model with phase (pre- or posttest), lexical condition ([x]- or [h]-bias), step on the [x]-[h] continuum (11, 26, 41, 46, 53, 69, 80) and replication block (1-12) as fixed factors and participant as a random factor.

Only the main effect of step and its interaction with replication block were significant. All main effects of and interactions with the factors phase and lexical conditions were not significant (all $p>.05$),

Figure 1: Categorization data, Dutch listeners, within-talker condition



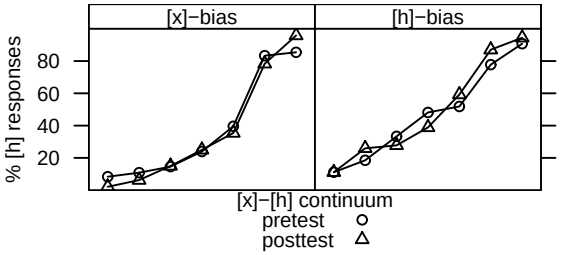
except from the third-order interaction of all four factors (z=2.30, p<.05). However, a likelihood ratio test revealed no advantage of the full model over a reduced model with only the factors step and replication block ($\chi^2_{(12)}$=14.96, p=.25). We therefore dropped the phase and lexical condition factors.

The reduced model showed that there were more [h]-responses to the more [h]-like stimuli (β=.073, z=5.21, p<.001), as shown in Fig. 1. Within the test blocks, the number of [x]-responses increased with replication block (β=-.066, z=-2.72, p<.01). This was mainly due to the most [x]-like stimuli (β=.0061, z=4.03, p<.001).

The expected interaction between phase and condition, which would indicate a lexically-driven learning effect, was not present. The main effect of phase was not significant either. The [x]-shift shown in Fig. 1 was thus caused by a shift towards [x] that happened within the test phases.

2.2.3. Categorization, between-talker condition

Figure 2: Categorization data, Dutch listeners, between-talker condition



These data were analysed in the same way, with similar results (see Fig. 2). Critically, no significant lexically-guided learning effect was found. The lack of a learning effect in this condition should be no surprise: one could not expect generalization across talkers of what is learned if nothing about the exposure talker is learned to begin with.

3. EXPERIMENT 2

In Experiment 1 we failed to observe the lexically-driven perceptual learning effect that has now been replicated several times [3, 5, 6, 8]. In order to interpret this null effect correctly, it is necessary to establish whether the expected learning effect occurs when both the talkers and the listeners have the same accent. Therefore, Experiment 2 is a replication of Experiment 1, but with Flemish listeners.

3.1. Method

Thirty-two native speakers of Dutch who were born and raised in Flanders received partial course credit for their participation; 16 were assigned to each condition. The experiment was otherwise identical to Experiment 1.

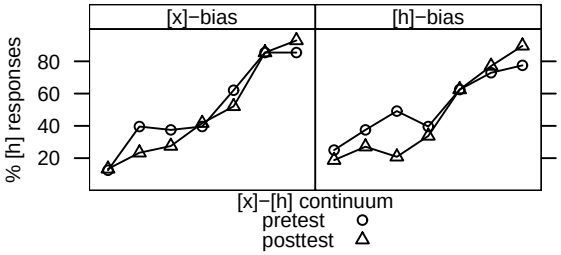
3.2. Results

3.2.1. Lexical decision

The results are shown in Table 1. There were virtually no differences between filler words, natural and ambiguous targets. The listeners accepted the ambiguous targets as words without hesitation.

3.2.2. Categorization, within-talker condition

Figure 3: Categorization data, Flemish listeners, within-talker condition

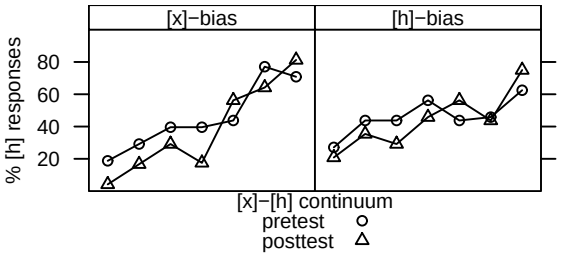


In the full model only the main effects of step and replication block were significant. The effects of all other factors and their interactions were not (all p>.05). A reduced model with only step and replication block provided as good a fit as the full model ($\chi^2_{(12)}$=14.79, p=.25). The reduced model showed that there were more [h]-responses to the more [h]-like stimuli (β=.039, z=6.70, p<.001). Within the test blocks, the number of [x]-responses increased with replication block (β=-.041, z=-2.20, p<.01). This was mainly due to the most [x]-like stimuli (β=.0025, z=2.57, p<.001).

The results were thus very similar to those of Experiment 1 (see Fig. 3). There was no lexically-driven learning and the number of [x]-responses increased over time.

3.2.3. Categorization, between-talker condition

Figure 4: Categorization data, Flemish listeners, between-talker condition



Again, no significant lexically-guided learning effect was found (see Fig. 4). As before, one cannot expect generalization of learning if nothing is learned.

4. GENERAL DISCUSSION

In two experiments we failed to observe a lexically driven perceptual adjustment of the [x]-[h] category boundary. We predicted more [x]-categorizations after exposure to the [x]-biasing words and more [h]-categorizations after exposure to the [h]-biasing words. Instead, in both groups the number of [x]-responses increased with replication block. The change of the categorization boundary was not caused by a sudden shift after the exposure phase, but by a continuous drift within the test phases. This drift was not present for the most [h]-like stimuli - participants accepted the most [x]-like stimuli as [x] more often with each repetition of the continuum.

Two methodological differences with [8] might have caused our failure to replicate. In [8], the ambiguous sound during exposure was always the same, and always word final. But studies that used a different sound for each target [5, 6], or where the critical phoneme occurred in the middle of the word [5] have successfully replicated the learning effect.

A more likely culprit is the phoneme contrast itself. Previously, more clearly defined contrasts have been used. The spectral properties of and the differences between the phonemes [f] and [s], for example, are well documented and rather stable across contexts. [h], however, is very variable due to coarticulation. This sound has no stable defining features and is not always recognized as a real phoneme [7]. It can be interpreted instead as the whispered onset of the following syllable-nuclear vocoid. For example the [h] in *he* [hi] is a whispered version of [i], whereas the [h] in *hoop* [hup] is a whispered version of [u].

To test whether there were any acoustical cues that could differentiate between the [x] and [h] sounds in our stimulus set, we computed the spectral moments, voicing degree, noise degree and the amount of high-frequency energy (above 1000 Hz) of the middle half of all the target fricatives. Voicing degree and noise degree were significantly different ($p<.001$) for [x] and [h] , but these two measures were highly correlated with each other. For [h] voicing degree was higher (67 vs. 35%) and noise degree was lower (46 vs. 80%). The high frequency energy in [h] was also lower (rms 190 vs. 231, $p<.01$). The spectral moments, that are often used to characterize different fricatives, did not differ between the two sounds. Only the spectral variance was slightly higher for [h] (2023 Hz vs. 1947 Hz, $p<.05$). Note that the analyses reported here only test for a difference between the two groups of sounds. A significant difference on a certain measure does not necessarily imply that this measure can predict how easy it is to differentiate between the two sounds. Even for noise degree the distributions of [x] (mean 80%, sd 18%) and [h] (mean 46%, sd 35%) had a reasonable overlap. Critically, most of the ambiguous sounds could be members of both categories (mean 58%, sd 33%).

It is therefore difficult to define the difference between the acoustic forms of [x] and [h] and the ambiguous sounds could have been instances of both categories. The variability of the sounds, and especially of [h], which may be defined almost entirely as a function of its context [7], means that it may be futile for listeners to adapt their [x]-[h] boundary.

5. REFERENCES

[1] Bradlow, A. R., Bent, T. 2003. Listener adaptation to foreign-accented speech. *Proc. 15th ICPhS* Barcelona, 2881-2884.

[2] Evans, B. G., Iverson, P. 2004. Vowel normalisation for accent: An investigation of best exemplar locations in northern and southern British English sentences. *J. Acoust. Soc. Am.* 115(1), 352–361.

[3] Eisner, F., McQueen, J. M. 2005. The specificity of perceptual learning in speech processing. *Percept. Psychophys.* 67(2), 224–238.

[4] Floccia, C., Goslin, J., Girard, F., Konopczynski, G. 2006. Does a regional accent perturb speech processing? *J. Exp. Psy. Hum. Perc. Perf.* 32(5), 1276–1293.

[5] Kraljic, T., Samuel, A. G. 2005. Perceptual learning for speech: Is there a return to normal? *Cog. Psy.* 51(2), 141–178.

[6] Kraljic, T., Samuel, A. G. 2007. Perceptual adjustments to multiple talkers. *J. Mem. Lang.* 56, 1–15.

[7] Laver, J. 1994. *Principles of phonetics*. Cambridge: CUP.

[8] Norris, D., McQueen, J. M., Cutler, A. 2003. Perceptual learning in speech. *Cog. Psy.* 47, 204–238.