Proceedings of the ICPhS XVII

17-21 August 2011 Hong Kong

香港城市大學
City University of Hong Kong

Proceedings of
the 17th International
Congress of Phonetic Sciences

17-21 August 2011
Hong Kong

www.icphs2011.hk

Edited by
Wai-Sum Lee & Eric Zee

Sponsored by: Department of Chinese, Translation and Linguistics, City University of Hong Kong

THE TIME COURSE OF PERCEPTUAL LEARNING

Katja Poellmann[a,b], *James M. McQueen*[a,c] *& Holger Mitterer*[a]

[a]Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands;
[b]International Max Planck Research School for Language Sciences, Radboud University, Nijmegen, the Netherlands;
[c]Radboud University Nijmegen, the Netherlands
katja.poellmann@mpi.nl; james.mcqueen@mpi.nl; holger.mitterer@mpi.nl

ABSTRACT

Two groups of participants were trained to perceive an ambiguous sound [s/f] as either /s/ or /f/ based on lexical bias: One group heard the ambiguous fricative in /s/-final words, the other in /f/-final words. This kind of exposure leads to a recalibration of the /s/-/f/ contrast, e.g. [4]. In order to investigate when and how this recalibration emerges, test trials were interspersed among training and filler trials. The learning effect needed at least 10 clear training items to arise. Its emergence seemed to occur in a rather step-wise fashion. Learning did not improve much after it first appeared. It is likely, however, that the early test trials attracted participants' attention and therefore may have interfered with the learning process.

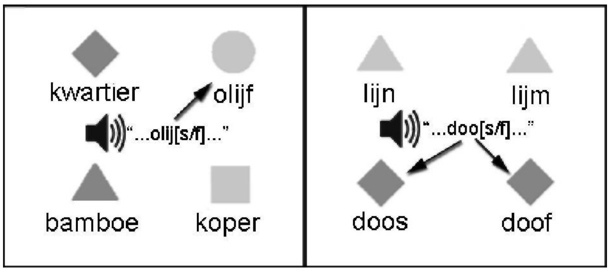
Keywords: perceptual learning, segmental idiosyncrasies, eye-tracking, time course

1. INTRODUCTION

Listeners can adapt to an odd pronunciation of a given phoneme. More precisely, listeners learn to interpret an ambiguous sound between /s/ and /f/ ([s/f]) as either /s/ or /f/ according to the training they received: Participants who heard [s/f] in words like "gira[s/f]" learned to interpret the sound as [f], while participants who heard [s/f] in words like "platypu[s/f]" learned to interpret it as [s] [1, 4, 5]. So far, all perceptual learning studies have used separate exposure and test phases. In the exposure phase, participants usually made lexical decisions on (non-)words they heard [1, 2, 4, 5]. In the test phase, they categorized a range of ambiguous sounds [1, 2, 4] or did a cross-modal priming task [5].

Exposure to just 10 instances of an ambiguous sound has been found to be enough to produce a stable perceptual learning effect [2]. However, the division of the experimental session into separate exposure and test phases has impeded a closer look at how learning emerges over time. The current study therefore used a variant of the visual-world eye-tracking paradigm [3] to examine the time course of perceptual learning. Participants saw displays as in Figure 1 and heard an instruction to click on one of the four words on the screen. Such trials can be learning trials when an ambiguous fricative appears in an unambiguous position (as in the left panel of Figure 1) or test trials when the ambiguous fricative appears at the end of a minimal pair (as in the right panel of Figure 1). This allows us to ask two questions: First, how many critical items are necessary to trigger perceptual learning? Second, is the learning process gradual or does it emerge suddenly in a step-wise fashion after a certain number of exposure trials?

Figure 1: Example of printed-word displays for a training trial (left panel) and a test trial (right panel). When hearing an [s/f]-bearing test word (e.g., *doo[s/f]*), participants should look to the /s/- or /f/-word according to their training. Participants saw only the four words and the four shapes on the screen.



2. METHOD

2.1. Stimuli

2.1.1. Stimulus selection

We selected three types of stimuli containing [s] or [f] which could be replaced by the ambiguous fricative [s/f]: 20 training items, in which the critical fricative could only be interpreted as [s] or [f] (e.g., *radijs* 'radish' and *olijf* 'olive'; *radijf* and *olijs* are nonwords in Dutch), 20 temporary minimal pairs, in which the ambiguous fricative could temporarily be interpreted as either [s] or [f], but was disambiguated later in the word (e.g., *gister* 'yesterday' - *giftig* 'toxic'), and 20 minimal pairs (e.g., *doos* 'box' - *doof* 'deaf'), which constituted the test items. The corresponding sets of stimuli were matched in terms of frequency (see Table 1). When pronounced with a natural fricative, the training items of one group served as contrast items for the other group. Thus, depending on the condition (/s/-bias or /f/-bias), it varied as to which words participants heard with natural or ambiguous fricatives (see Table 2).

Table 1: Mean frequencies per million of the different types of stimuli.

Items	/s/-words	/f/-words
training	9.9	9.5
test pairs	31.2	35.5
temporary minimal pairs	5.7	7.7

Table 2: Examples of the different stimulus types (natural and [s/f]-bearing) for both groups.

Items	/s/-bias group	/f/-bias group
training	radij[s/f]	olij[s/f]
contrast	olijf	radijs
test pairs	doo[s/f] + doof	doo[s/f] + doos
temporary minimal pairs	gi[s/f]ter	gi[s/f]tig

2.1.2. *Stimulus construction*

Digital recordings of the stimuli were made by a male native speaker of Dutch in a sound-proof booth, sampling at 44 kHz. All target items containing /s/ or /f/ were recorded in their natural version. The /f/-targets were additionally recorded with an [s] replacing the [f] (e.g., *olijf* 'olive' as well as *olijs*). When creating the ambiguous [s/f]-items, the [s]-versions were used: The [s] was removed and replaced with instances of [s/f], which were selected by means of two pretests. To reduce a coarticulatory [s]-bias, splicing points were chosen reasonably early in the vowel preceding the fricative. In addition, the amplitude of that vowel was decreased to match better an ambiguous sound that could still be interpreted as [f] by a native speaker of Dutch.

In order to find ambiguous fricatives midway between [s] and [f], two pretests were run. For the first pretest, 14 continua of VC-syllables were created. These contained all the vowels present in the 60 critical stimuli and a digital mixture of [s] and [f] varying in the proportions of the two fricatives in 11 steps (from 100% [s] 0% [f] via 90% [s] 10% [f] to 0% [s] 100% [f]). Participants were asked to categorize the fricatives of every second step (i.e., five steps per syllable) as either [s] or [f]. For each of the 14 vowels, the step whose value was closest to the ideal value of 50% of [s] responses was chosen as the ambiguous fricative for the second pretest.

As in the main experiment, the second pretest used [s/f]-bearing words embedded in complete sentences. Participants heard the carrier sentence "Klik op het woordje" ('Click on the word') followed by a member of an /s/-/f/-minimal pair (e.g., *doof*). These stimuli had either a natural [f], a natural [s], or an ambiguous fricative [s/f] (which varied according to the preceding vowel). Every participant heard and categorized each item four times. Ambiguous fricatives were selected per vowel context if they were categorized as [s] between 30% and 70%. For those outside this range, a new step was chosen in order to correct for the observed /f/- or /s/-bias.

2.2. Design and procedure

Each participant heard 240 instructions like "Klik op het woordje *doo[s/f]* onder het rooie ruitje" ('Click on the word *doo[s/f]* underneath the red diamond', see Figure 1) and was asked to follow these instructions using the computer mouse. Except for the test trials, there were always four different shapes (a circle, a diamond, a triangle, and a rectangle) on the screen. The two shapes associated with the target and its same-color competitor were presented in one of four colors (red, green, yellow and blue), the shapes paired with the two distractors were presented in one of the three remaining colors (see left panel in Figure 1 with gray-scale coding of different colors). In the test trials, the target and competitor shapes and those for the two distractors were the same. That is, there were only two different shapes on the screen (see right panel in Figure 1). This design was chosen in order to force participants to make a decision about every word they heard.

Eye movements were tracked using an Eyelink 1000 at a sampling rate of 1 kHz. The stimuli were presented in 20 miniblocks, each consisting of one training item, one contrast item, both members of a minimal pair (an [s/f]-bearing one serving as test item, the other one with a natural fricative serving as a second contrast item), one member of a temporary minimal pair (see Table 1), and seven filler trials. In order to be able to interpret the observed effects correctly, we used only one randomization of the 240 trials for all participants.

2.3. Participants

All participants were paid native speakers of Dutch. Twelve took part in each of two pretests and 44 in the main eye-tracking experiment.

3. RESULTS

3.1. Mouse-click responses

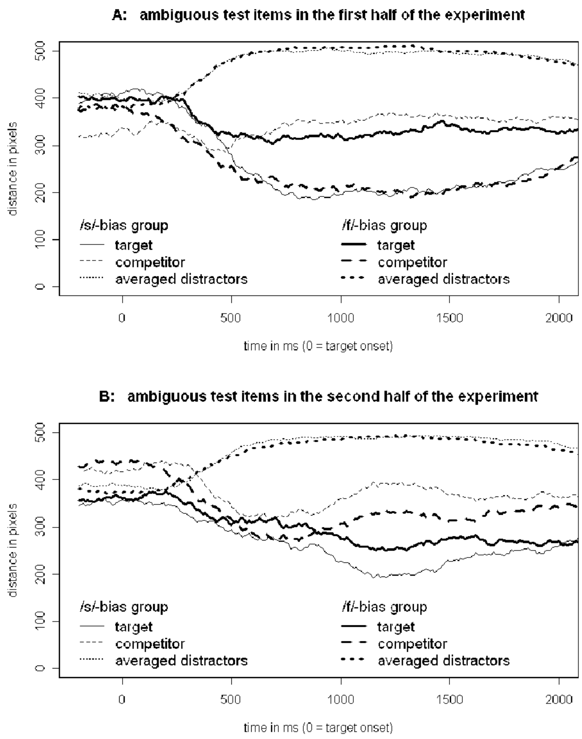
A linear mixed-effects model for the accuracy of the click responses showed an interaction of Group and Block. Separate analysis showed that the /s/-bias group performed above chance overall (68% correct responses) but did not improve over the course of the experiment. The /f/-bias group showed a significant improvement ($p < 0.05$) from 20% to 70% correct responses from the first to the last miniblock.

3.2. Eye-tracking data

Figures 2A and 2B show the eye-tracking results for the test trials of the first and second half of the experiment. Both groups heard members of /s/-/f/-minimal pairs pronounced with an ambiguous fricative (e.g., *doo[s/f]*). In the first half of the experiment (see Figure 2A), the /s/-bias group (represented by the thin lines) started to prefer their target words (e.g., *doos*) approximately 500 ms after target onset. At this point, the target and the competitor lines diverge. The /f/-bias group, however, behaved in exactly the same way. When listeners in the /f/-bias group heard *doo[s/f]*, they also tended to look to *doos* (their competitor) and not to the target *doof*. In the second half of the experiment (see Figure 2B), a major change in the behavior of the /f/-bias group (but not the /s/-bias group) can be observed. After first considering the competitor as a possible candidate, participants in the /f/-bias group prefer to look to their target (e.g., *doof*) from 900 ms onwards.

This is borne out by statistical analyses using linear mixed-effects models with Target Preference as dependent and Group and Block as independent variables. We used sliding 200 ms time windows from 400 ms to 1600 ms after target onset. There were clear group effects in all time windows (p_{max} = 0.0056), indicating the better performance of the /s/-bias group, and effects of Block ($p < 0.05$) from 1000 ms till 1500 ms, indicating an overall learning effect. The learning effect was numerically larger for the /f/-bias group, confirming a similar trend as was found in the click responses. However, an interaction of Group and Block (indicating significantly more learning in the /f/-bias group) was only marginal in the late time windows.

Figure 2: Distance of participants' looks to the target, competitor, and averaged distractors in pixels during the first half (Panel A) and second half (Panel B) of the experiment. Note that looks to an object lead to diminishing distances, so that smaller distances mean more looks.



3.3. Learning function

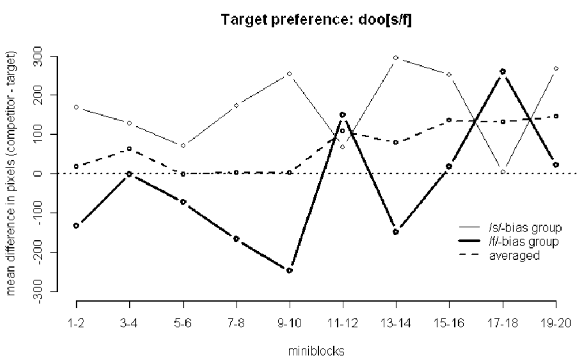
To calculate the learning function, we subtracted for each miniblock the target fixation distance from the competitor fixation distance in the time window from 1000 ms to 1500 ms after target onset and took the average value as representative for that miniblock. If participants have a preference for the target, this difference (competitor - target) will be positive.

Figure 3 shows the learning function averaged over two miniblocks for both groups separately as well as averaged over the two groups. Given that the target preference of the /s/-bias group was above chance throughout the experiment, participants in this group did well right from the beginning and improved only slightly. Apparently, there was no reason for them to adapt to the ambiguous sound as they already perceived it as /s/

in early miniblocks. The /f/-bias group started off quite badly, but improved during the experiment. That the two functions mirror each other so strongly is evidence for stimulus-specific biases in our materials. The auditory test stimuli used for each miniblock were identical for all participants, so a strong difference between the two groups indicates that the two [s/f]-bearing test items contributing to a given data point in Figure 3 either sound quite /s/-like or quite /f/-like. Overall, the results show that most items sounded more like /s/ with the exception of blocks 11/12 and 17/18.

The overall perceptual learning effect (driven mostly by the learning of the /f/-bias group) becomes evident after miniblock 10, as can be seen from the dashed line. The function rises quite suddenly at this point from the 0-line to values around 100 pixels and then does not rise much more. This suggests that perceptual learning occurs in a rather step-wise fashion.

Figure 3: Learning function averaged over two miniblocks for the /s/-bias group, the /f/-bias group, and averaged over both groups. Positive values on the y-axis (competitor - target) indicate more looks towards the intended target.



4. DISCUSSION

The aim of this study was to replicate perceptual learning effects with an eye-tracking paradigm and hence to investigate how learning emerges over time. As in [1, 4, 5], we were able to observe a learning effect after 20 critical items, as participants had heard 10 training items but also 10 [s/f]-bearing members of temporary minimal pairs (e.g., *gister - giftig*) after miniblock 10. However, in order to replicate the findings of [2], we should have found a learning effect from miniblock 5 onwards. A possible explanation for this difference may be that the temporary minimal pairs do not constitute training items, because there is insufficient lexical bias to guide perceptual learning at the moment the ambiguous fricative is heard. That is, the fragment *gi[s/f]* could contain an /s/ (continuing as *gister*) or an /f/ (continuing as *giftig*).

We can provisionally answer the two questions we posed. First, we found that learning apparently needs at least 10 clear training items to arise. Second, it seems that learning occurs in a step-wise fashion, as the learning effect that arose after 10 miniblocks did not get stronger with additional training.

A caveat, however, is that we observed quite a small learning effect. At least in cross-modal priming, learning appears to be "complete" after 20 training items [5]. In our data, however, the /f/-bias group still has the tendency to look at the /s/-words when hearing an ambiguous fricative, suggesting incomplete learning. A possible explanation for this discrepancy is that participants may have been aware of the purpose of the experiment. In a quarter of all trials, they saw a display like that in the right panel of Figure 1 and had to click on a member of a minimal pair. Thus, they may have paid more attention to the critical sounds and made conscious metalinguistic judgments about them. Put differently, the early test trials may have interfered with the learning process.

An interesting conclusion nevertheless arises from the fact that learning effects only occur late within a trial (> 900 ms after target onset). This suggests that perceptual learning may not influence first-pass perceptual processing, but only a reevaluation process when the input is ambiguous.

5. REFERENCES

[1] Eisner, F., McQueen, J.M. 2005. The specificity of perceptual learning in speech processing. *Perception & Psychophysics* 67, 224-238.

[2] Kraljic, T., Samuel, A.G. 2007. Perceptual adjustments to multiple speakers. *Journal of Memory and Language* 56, 1-15.

[3] Mitterer, H., McQueen, J.M. 2009. Processing reduced word-forms in speech perception using probabilistic knowledge about speech production. *Journal of Experimental Psychology: Human Perception and Performance* 35, 244-263.

[4] Norris, D., McQueen, J.M., Cutler, A. 2003. Perceptual learning in speech. *Cognitive Psychology* 47, 204-238.

[5] Sjerps, M.J., McQueen, J.M. 2010. The bounds on flexibility in speech perception. *Journal of Experimental Psychology: Human Perception and Performance* 36, 195-211.