

Perceptual Learning of /f/-/s/ by Older Listeners

Odette Scharenborg^{1,2}, Esther Janse^{3,1,2}, and Andrea Weber^{1,2}

¹ Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

² Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, The Netherlands

³ Centre for Language Studies, Radboud University Nijmegen, The Netherlands

{Odette.Scharenborg, Esther.Janse, Andrea.Weber}@mpi.nl

Abstract

Young listeners can quickly modify their interpretation of a speech sound when a talker produces the sound ambiguously. Young Dutch listeners rely mainly on the higher frequencies to distinguish between /f/ and /s/, but these higher frequencies are particularly vulnerable to age-related hearing loss. We therefore tested whether older Dutch listeners can show perceptual retuning given an ambiguous pronunciation in between /f/ and /s/. Results of a lexically-guided perceptual learning experiment showed that older Dutch listeners are still able to learn non-standard pronunciations of /f/ and /s/. Possibly, the older listeners have learned to rely on other acoustic cues, such as formant transitions, to distinguish between /f/ and /s/. However, the size and duration of the perceptual effect is influenced by hearing loss, with listeners with poorer hearing showing a smaller and a shorter-lived learning effect.

Index Terms: perceptual learning, older listeners, aging, hearing loss, human word recognition.

1. Introduction

Aging often affects sensitivity to the higher frequencies, which results in reduced sensitivity to phonetic detail in speech. The question arises whether this loss of sensitivity to speech detail affects the ability to learn non-standard pronunciations and to quickly tune into a speaker's pronunciations like young normal-hearing listeners can [1]. Using a perceptual learning paradigm, it has been shown that young listeners use both lexical and phonotactic knowledge to quickly retune their phonemic categories in response to ambiguous pronunciations [1],[2]. For instance, an ambiguous sound between /f/ and /s/ ([f/s]) will be learned as /s/ if heard in words such as *platypus*, but as /f/ in words such as *giraffe*. This learning generalises to words that have not been presented earlier [3], so that listeners interpret the previously unheard word [naɪf/s], as *nice* or *knife* depending on their previous exposure to *platypu[s/f]* or *gira[s/f]*.

Recent evidence suggests that this ability to adapt to unusual pronunciations seems to be stable throughout the life span [4],[5], at least for certain contrasts. In a large-scale lexically-guided perceptual learning experiment it was found that older Dutch listeners (60+ years old) are capable of perceptual learning that is comparable to young listeners [6]. This study used the contrast [l]-[ɹ], a contrast that mainly differs in the F3, which is much lower for [ɹ] than for [l], i.e., below 2000 Hz [7]. Since hearing loss usually affects the higher frequencies it can be assumed that older listeners are still able to hear the difference between the [l] and [ɹ] quite well. In this study, we investigate

whether lexically-guided perceptual learning occurs for a consonant contrast that has most information to differentiate between the two consonants in the higher frequencies and less in the lower frequencies, i.e., /f/ vs. /s/. /f/ and /s/ differ in their spectrum and the formant transitions.

[s] has more energy in the higher frequencies, concentrating around 5500 Hz, while [f] has a flatter spectrum, with the energy distributed more uniformly over the spectrum [7]. This information in the higher frequencies is (presumably) less strongly represented in many older listeners. Formant transitions are also quite informative; however, research has shown that Dutch young listeners do not rely on them [8]; instead the information in the fricative noise is more important. If older listeners still rely on information in the higher frequencies, their hearing loss may prevent perceptual learning because they cannot distinguish between /f/ and /s/ well enough. On the other hand, if older listeners have learned to rely on other cues instead (such as formant transitions which differ mostly around 2200 Hz and are thus less affected by age-related hearing loss), they may still show perceptual learning [7].

The questions addressed in this study are: 1) are older listeners capable of perceptual learning of a consonant contrast that differentiates between the two consonants in the higher frequencies? 2) (To what extent) Is individual amount of high-frequency hearing loss predictive of whether or not the category boundary is shifted to include the ambiguous sound? Moreover, the effect of hearing loss on the stability of the perceptual learning effect is investigated.

The experiment consisted of two parts (following [1],[2],[6]). First, listeners were exposed to an ambiguous [f/s] in Dutch words ending on either /f/ or /s/ during a self-paced lexical decision task (the exposure phase). Listeners were divided into two groups: one group was exposed to the ambiguous sound only in /f/-final words and the other group was exposed to the same ambiguous sound only in /s/-final words. In a subsequent self-paced phonetic categorisation task (the test phase), listeners were confronted with a range of ambiguous sounds from an [f]-[s]-continuum, appearing as the final phonemes of Dutch words, and were asked to decide whether the sound was /f/ or /s/. Both readings of the words yielded an existing Dutch word (e.g., *brief*[f/s] could be *brief* "letter" or *bries* "breeze").

2. Experimental Set-up

2.1. Participants

Fifty-three participants aged 60+ (13 M; mean age: 69.4, SD: 7.3), who were native speakers of Dutch were drawn from the MPI for Psycholinguistics subject pool and were paid for their

participation. Prior to the start of the experiment, hearing sensitivity was assessed with a Maico ST20 portable audiometer (air conduction thresholds only) for both ears at octave frequencies from 250 Hz through 8 kHz. No participants wore hearing aids. Mean pure-tone average (averaged over participants' thresholds at 1, 2 and 4 kHz in their better ear) was 24.8 dB HL (SD=16.7).

2.2. Materials

The critical items were identical to those used in [1]. For the exposure phase, 20 Dutch words with final /f/ (e.g., *olijf*, “olive”) and 20 Dutch words with final /s/ (e.g., *hagedis*, “lizard”) were used (no further /f/ and /s/ occurred in the words); there were no minimal pairs. Syllable structure, stress patterns, and word frequency were matched as far as possible. Sixty additional words without /f/ and /s/ and 100 nonwords were selected as fillers (following the same syllable length distribution as for the target words). The nonwords followed Dutch phonotactic rules and tended to become nonwords (i.e., were no longer consistent with any real Dutch words) before their final phonemes.

All words were produced in isolation by a female native speaker of Dutch and digitally recorded in a sound-attenuated booth at 44 kHz. She also recorded four minimal word pairs for the test phase: *brief-bries* (“letter”-“breeze”), *graf-gras* (“grave”-“grass”), *leef-lees* (“live”-“read”), *lof-los* (“praise”-“loose”) for use in the test phase. Subsequently, ambiguous versions of the 40 critical exposure words ending in /f/ and /s/ as well as of the test words were made. The ambiguous sound [f/s] for the exposure phase and the steps on the test continuum for the phonetic categorisation task were selected using a phonetic categorisation pretest on 10 older listeners who did not participate in the main experiment. The critical items contained eight different vowels preceding the final /f/ or /s/, in total. The selection of the ambiguous sounds was done separately for each vowel. Therefore, eight minimal pairs (including the four minimal pairs used in the test phase), one for each vowel were used to create the ambiguous sounds, and used in the pretest.

For the pretest, each of the eight minimal pairs, the final fricative was excised, zero-padded with 25 ms of silence, and subsequently morphed with its counterpart from the minimal pair to create an equally-spaced 11-step continuum using STRAIGHT [9] in Matlab. The ambiguous fricatives were then concatenated as final sounds to both items in the minimal pair.

During the pretest, six [f]-[s]-continuum steps for each of the eight minimal pairs were each presented six times binaurally over closed headphones. All participants were tested individually in a sound-treated booth. The task for the participants was to indicate by button press as quickly and as accurately as possible whether they heard the /f/-final reading or the /s/-final reading of the word. To help the listeners, the /f/-final word interpretation was always printed on the left side of the screen and the /s/-final word interpretation on the right side of the screen.

The total proportion of /s/-responses to each of the tested morphs (averaged over all eight minimal pairs) were calculated, and the most ambiguous morph was determined, which turned out to be step 4. Finally, the eight step 4 morphs were concatenated as final sounds to their vowel-consistent /f/- and /s/-final items, in the same manner as was done to create the stimuli for the pretest. This resulted in 40 stimulus pairs consisting of the same word ending in either a natural [f] or [s] or the selected ambiguous [f/s] sound. These stimuli were then used in the lexical decision task (exposure phase).

The test stimuli consisted of five versions of the four minimal pairs listed above, which were created by concatenating five versions of [f/s] (i.e., steps 1, 3, 4, 5, 7) as final sound to both items of the minimal pair.

2.3. Procedure

Two experimental-word lists were created in which the exposure items appeared in a pseudo-randomised running order, one for each experimental (between-subject) condition. The restrictions were that no critical item (i.e., no word ending in [f/s]) was allowed to appear in the first six words, and no two critical items could appear within a range of four words. Each list consisted of 200 words: the 100 nonwords, 60 filler words, 20 words ending in a natural [f] or [s], and 20 critical items, i.e., the /f/-final or /s/-final words ending in the ambiguous sound [f/s]. The difference between the two word lists was that one list contained only natural /f/-final words and /s/-final words ending in ambiguous [f/s], the other list contained the natural /s/-final words and the /f/-words ending in ambiguous [f/s]. Twenty-seven participants were presented with the list with ambiguous /f/-words, and 26 participants with the list with ambiguous /s/-words.

Participants were tested individually in a sound-treated booth. The stimuli were presented binaurally over closed headphones. Participants were asked to press a button as fast and accurately as possible when they heard a word (left button) or a non-word (right button). They were not informed about the presence of ambiguous sounds. Response times (RTs) were measured from item onset and adjusted by subtraction of item durations prior to analysis so as to measure from item offset.

Next, participants had to perform a phoneme categorisation test (test phase) on the four minimal pairs. They were asked to decide as fast and accurately as possible, by button press, whether they heard the /f/-final reading or the /s/-final reading of the word. The five ambiguous items of each word in each minimal pair (so each participant heard both source words) were each presented once per block (i.e., 40 items/block), and were newly randomised for each of a total of four blocks (120 items in total). To aid the participants, the /f/-interpretation of the stimulus was shown on the bottom left of the computer screen, and the /s/-interpretation of the stimulus on the bottom right.

The experiment is part of a larger experiment in which also background tasks were administered to the older listeners (after the categorisation task) in order to determine their performance on a battery of cognitive tests. The perceptual learning experiment lasted approximately 20 mins, and the background tasks (together) about 90 mins. The tasks were administered in one session with a break in the middle. We only present the results of the perceptual learning experiment here.

3. Results

Due to failure of the experimental software, the results of two participants who were exposed to the ambiguous sound in /f/-final words could not be used in the analyses. This left us with 51 participants: 26 participants who were exposed to the ambiguous sound in /s/-final words and 25 participants who were exposed to the ambiguous sound in /f/-final words.

3.1. Lexical decision

We first examined performance during the exposure phase. The percentage of ‘yes’ responses for the word filler items was

96.7% for the listener group exposed to the ambiguous sounds in the /f/-final words and 95.0% for the listeners exposed to the ambiguous sound in the /s/-final words. The percentage of ‘no’ responses to non-word filler items was 96.0% for the listener group exposed to the ambiguous sounds in the /f/-final words and 95.4% for the listeners exposed to the ambiguous sound in the /s/-final words.

Table 1 shows the mean percentage of ‘yes’ responses to the natural and the ambiguous versions of the /f/- and /s/-final words for the listeners who were exposed to the ambiguous sound in /f/- and /s/-final words. As Table 1 shows, listeners accepted most of the stimuli ending in the ambiguous [f/s] as words. Analyses carried out using generalised linear mixed-effects models showed that there were significantly more ‘yes’ responses to the natural stimuli than to the ambiguous stimuli ($\beta = 1.4446$, $SE = .4115$, $p < .001$). All results presented here were obtained with the best-fitting model (obtained after model comparisons). A subsequent analysis showed that there was no effect of hearing sensitivity on performance, nor were there any interactions between hearing sensitivity and word type.

Subsequently, mean response times (RTs), measured from stimulus offset, for ‘yes’ responses were analysed. These means are obtained after excluding all RTs that were more than 3 SDs slower than the mean RT. This led to the exclusion of 69 (=3.6%) trials. Mean RTs for the four item types are listed in Table 1. Analyses of the mean RTs using linear mixed-effects models showed that there were no significant differences in mean RTs for the four word types. An analysis into the effect of hearing sensitivity on mean RTs showed that listeners with worse hearing were slower to respond overall ($\beta = 4.064$, $SE = 1.402$, $p < .005$; p-value is based on Monte Carlo Markov Chain sampling). There were no interactions between hearing sensitivity and word type.

To summarise, the results showed that the listeners who were exposed to natural versions of the /f/-final words and to the ambiguous [f/s] in the normally /s/-final items often interpreted the ambiguous sound as /s/, whereas listeners who were exposed to [f/s] in the normally /f/-final words and natural versions of the /s/-final words interpreted [f/s] as /f/. Nevertheless, participants significantly more often judged natural stimuli as words than the stimuli ending in an ambiguous sound. Hearing sensitivity did not play a role in determining whether a stimulus is a word or a nonword, but did play a role in response times, with listeners with increased hearing loss responding more slowly.

3.2. Phonetic categorisation

The phonetic categorisation data were analysed using generalised linear mixed-effects models. The results presented here were obtained with the best-fitting model (after model comparisons). To show the stability of the perceptual learning effect over time, Figure 1 shows the proportion of /s/-responses for the five ambiguous stimuli in the phonetic categorisation task, for the four test blocks separately. The responses for the listeners who were exposed to the ambiguous [f/s] sound in the /f/-final words are indicated with ‘F’s. The responses for the listeners who were exposed to [f/s] only in the /s/-final words are indicated with ‘S’s.

We first analysed the data with respect to our first research question: are older listeners capable of perceptual learning of a consonant contrast that differentiates between the two consonants in the higher frequencies?

Table 1. Performance on the lexical decision task, mean percentage of ‘no’ responses and mean RTs for ‘yes’ responses in the two exposure conditions, for the natural and the ambiguous versions of the /f/- and /s/-final words.

	Natural fricatives		Ambiguous fricatives	
	/f/-final	/s/-final	/f/-final	/s/-final
Mean % yes	97.9	95.6	90.0	90.8
Mean RT yes	251	238	224	238

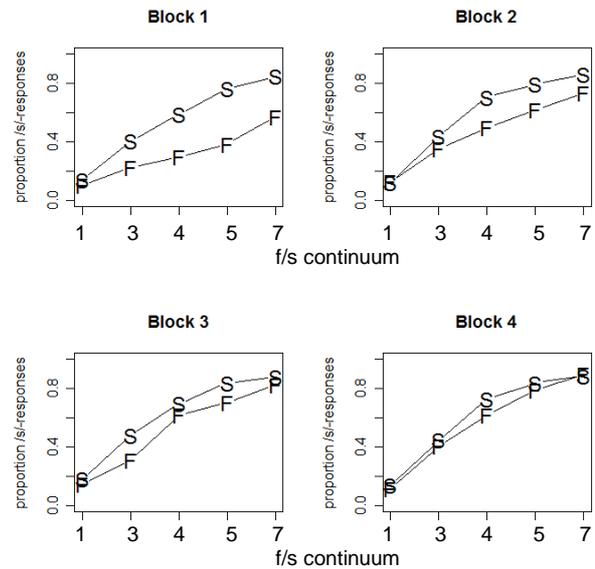


Figure 1. Total proportion of /s/ responses in the two exposure conditions for the five ambiguous test stimuli, over the four test blocks: S indicates the listeners who learned to map [s/f] onto [s]; F indicates the listeners who learned to map [f/s] onto [f].

As Figure 1 shows there is an effect of exposure condition on phonetic categorisation. This effect is called the lexically-guided perceptual learning effect. Listeners who were exposed to the ambiguous [f/s] in the /s/-final words were strongly biased to label the sounds on the continuum as /s/, while those listeners who were exposed to the ambiguous [f/s] in the /f/-final words were less likely to do so ($\beta = 1.6932$, $SE = .4076$, $p < .001$).

Significantly more /s/-responses were given to higher stimulus steps on the continuum, which is to be expected as ‘higher’ means more /s/-like ambiguous sounds ($\beta = .5354$, $SE = .0397$, $p < .001$). This effect was stronger for the higher stimulus steps in later test blocks (i.e., test block 2, 3 and 4; ($\beta = .1924$, $SE = .0242$, $p < .001$) and for the listeners who learned to map [f/s] onto /s/ than for the listeners who learned to map [f/s] onto /f/ ($\beta = .3233$, $SE = .0605$, $p < .001$). Significantly more /s/-responses were given in later test blocks ($\beta = .5873$, $SE = .0404$, $p < .001$), although this was less so for listeners who learned to map the ambiguous sound onto /s/ ($\beta = -.4389$, $SE = .0568$, $p < .001$). However, these effects reduced in later test blocks as shown by a 3-way interaction between exposure condition, stimulus step and test block ($\beta = -.1189$, $SE = .0350$, $p < .001$).

Finally, the interaction between exposure condition and test block was further investigated by analysing the effect of exposure condition in each separate block. This analysis showed that the perceptual learning effect was still present in test block 3 ($\beta = .8656$, $SE = .4358$, $p < .05$), but no longer in test block 4.

Subsequently, we investigated the second research question: to what extent is individual amount of high-frequency hearing loss predictive of whether or not the category boundary is shifted to include the ambiguous sound? To that end, the two exposure groups were taken together and a new category ‘learning-consistent’ was created in which the /f/-responses during the phonetic categorisation task of the group of listeners exposed to the ambiguous sound in /f/-final words and the /s/-responses of the group of listeners exposed to the ambiguous sound in the /s/-final words were combined. Moreover, we looked only at the stimulus steps of interest (i.e., the most ambiguous steps - steps 3, 4, and 5 in Figure 1). We investigated whether participants with more hearing loss gave fewer learning-consistent responses (i.e., fewer /f/-responses when they were exposed to the ambiguous sound in /f/-final words and fewer /s/-responses when they were exposed to the ambiguous sound in /s/-final words), particularly for the most ambiguous stimuli.

The results with the best-fitting generalised linear mixed-effects model showed that significantly fewer learning-consistent responses were given with more hearing loss ($\beta = -.0867$, $SE = .0311$, $p < .01$) and for later test blocks ($\beta = -.4376$, $SE = .04588$, $p < .001$). The latter is indicative of general ‘unlearning’. This unlearning is stronger for listeners with poorer hearing ($\beta = -.0213$, $SE = .0059$, $p < .001$). Finally, there was less unlearning for the listeners exposed to the ambiguous sound in the /s/-final group ($\beta = .5381$, $SE = .0642$, $p < .001$), which can also be clearly seen in Figure 1. The latter was less the case, however, for those with poorer hearing ($\beta = -.0225$, $SE = .0087$, $p < .01$). In other words, those listeners in the /s/-final group with more hearing loss had more unlearning than those with better hearing. This increased unlearning might be linked to the fact that listeners with hearing loss have more problems identifying /s/ than identifying /f/ [10].

4. General discussion and conclusions

This paper investigates two related questions regarding the lexically-guided perceptual learning effect: 1) are older listeners capable of perceptual learning given a consonant contrast that mainly differentiates between the two consonants in the higher frequencies? 2) (To what extent) Is individual amount of high-frequency hearing loss predictive of whether or not the category boundary is shifted to include the ambiguous sound? These questions are investigated in a lexically-guided perceptual learning study using a consonant contrast that distinguishes between the two consonants in the higher frequencies, thus exactly those frequencies that are known to suffer from age-related hearing loss: the /f/-/s/ contrast.

In the experiment, listeners were exposed to an ambiguous [f/s] in Dutch words ending in either /f/ or /s/. The ambiguous sound was created by morphing word-final [f] and [s]. A subsequent phonetic categorisation test revealed a significant difference in proportion of /s/-responses to an [f]-[s] continuum between the two listener groups that learned to interpret the ambiguous sound as either /f/ or /s/. There were more /s/-responses by the listeners who had been exposed to ambiguous [f/s] in the /s/-final words. So, to answer our first research question: yes, older listeners are capable of perceptual learning of a contrast that differentiates in the higher frequencies. This adaptation is also preserved over time. The perceptual learning effect was still present in the third test block, albeit no longer in the final test block.

The role of hearing sensitivity was investigated for both the lexical decision task and the phonetic categorisation task. Hearing loss did not interfere with lexical decision accuracy on the critical word items, but it did interfere with response times. Generally speaking, listeners with increased hearing loss were slower to respond than listeners with better hearing. The analysis of the phonetic categorisation results showed that hearing sensitivity interfered with perceptual learning. To answer our second research question: listeners with poorer hearing gave fewer learning-consistent responses and showed greater unlearning, i.e., they showed a smaller and less stable learning effect, especially those participants who were exposed to the ambiguous sound in /s/-final words. These results are in line with [6], who also showed that hearing loss resulted in a stronger unlearning effect. This effect of a smaller learning effect with poorer hearing may be due to the fact that the critical information present in the fricative noise is no longer available to them or not reliable enough. Further, hearing loss may also impede the use of other cues, such as formant transitions (see also [10]), even though this transition information is represented at somewhat lower frequencies.

To conclude, our results show that older listeners, with and without hearing loss, can still retune their phoneme categories to facilitate word recognition. However, hearing loss seems to interfere with the size and the stability of the perceptual learning effect, with a smaller learning effect and a stronger unlearning with poorer hearing.

5. Acknowledgements

The research by Odette Scharenborg is sponsored by the Max Planck International Research Network on Aging, Rostock, Germany. The research by Esther Janse is supported by a Vidi-grant from the Netherlands Organization for Scientific Research (NWO). Research by Andrea Weber is funded by the Max Planck Society, Munich, Germany.

6. References

- [1] Norris, D., McQueen, J.M., Cutler, A., “Perceptual learning in speech”, *Cognitive Psychology*, 47(2): 204-238, 2003.
- [2] Cutler, A., McQueen, J.M., Butterfield, S., Norris, D., “Prelexically-driven perceptual retuning of phoneme boundaries”, *Proceedings of Interspeech*, 2056-2056, 2008.
- [3] McQueen, J.M., Cutler, A., Norris, D., “Phonological abstraction in the mental lexicon”, *Cognitive Science*, 30(6): 1113-1126, 2006.
- [4] Peelle, J.E., Wingfield, A., “Dissociations in perceptual learning revealed by adult age differences in adaptation to time-compressed speech”, *JEP:HPP*, 31(6):1315-1330, 2005.
- [5] Adank, P., Janse, E., “Comprehension of a novel accent by young and older listeners”, *Psychology and Aging*, 25(3): 736-740, 2010.
- [6] Scharenborg, O., Janse, E. “Perceptual learning of liquids in older listeners”, *AMLaP*, Paris, France, 2011.
- [7] Rietveld, A.C.M., van Heuven, V.J., “Algemene Fonetiek”, Bussum: Dick Coutinho, 1997.
- [8] Wagner, A., Ernestus, M., Cutler, A. “Formant transitions in fricative identification: The role of native fricative inventory”, *JASA*:120, 2267-2277, 2006.
- [9] Kawahara, H., Masuda-Katsuse, I., Cheveigne, A., “Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction: possible role of a repetitive structure in sounds”, *Speech Communication*, 27:187-207, 1999.
- [10] Scharenborg, O., Janse, E. “Hearing loss and the use of acoustic cues in phonetic categorisation of fricatives”, To appear in *proceedings of Interspeech*, Portland, OR, 2012.