The neural correlates underlying lexically-guided perceptual learning

Scharenborg, Odette; Koemans, Jiska ; Smith, Cybelle ; Hasegawa-Johnson, Mark; Federmeier, Kara D.

DOI
10.21437/Interspeech.2019-2328

Publication date
2019

Document Version
Final published version

Published in
Proceedings of Interspeech 2019

Citation (APA)
Scharenborg, O., Koemans, J., Smith, C., Hasegawa-Johnson, M., & Federmeier, K. D. (2019). The neural correlates underlying lexically-guided perceptual learning. In G. Kubin, T. Hain, B. Schuller, D. E. Zarka, & P. Hodl (Eds.), *Proceedings of Interspeech 2019* (pp. 1223-1227). ISCA. https://doi.org/10.21437/Interspeech.2019-2328

INTERSPEECH 2019
September 15–19, 2019, Graz, Austria

The Neural Correlates Underlying Lexically-guided Perceptual Learning

Odette Scharenborg[1], *Jiska Koemans*[2], *Cybelle Smith*[3], *Mark Hasegawa-Johnson*[4], *and Kara D. Federmeier*[3,4]

[1]Multimedia Computing Group, Delft University of Technology, the Netherlands
[2]Radboud University Nijmegen, the Netherlands
[3]Department of Psychology, University of Illinois at Urbana-Champaign, USA
[4]Beckman Institute, University of Illinois at Urbana-Champaign, IL, USA
o.e.scharenborg@tudelft.nl

Abstract

There is ample evidence showing that listeners are able to quickly adapt their phoneme classes to ambiguous sounds using a process called lexically-guided perceptual learning. This paper presents the first attempt to examine the neural correlates underlying this process. Specifically, we compared the brain's responses to ambiguous [f/s] sounds in Dutch non-native listeners of English (N=36) before and after exposure to the ambiguous sound to induce learning, using Event-Related Potentials (ERPs). We identified a group of participants who showed lexically-guided perceptual learning in their phonetic categorization behavior as observed by a significant difference in /s/ responses between pretest and posttest and a group who did not. Moreover, we observed differences in mean ERP amplitude to ambiguous phonemes at pretest and posttest, shown by a reliable reduction in amplitude of a positivity over medial central channels from 250 to 550 ms. However, we observed no significant correlation between the size of behavioral and neural pre/posttest effects. Possibly, the observed behavioral and ERP differences between pretest and posttest link to different aspects of the sound classification task. In follow-up research, these differences will be further investigated by assessing their relationship to neural responses to the ambiguous sounds in the exposure phase.

Index Terms: adaptation, lexically-guided perceptual learning, neural correlates, ERP, human speech processing

1. Introduction

The speech signal is highly variable due to large intra-speaker variability. There is ample evidence that listeners deal with the variability of speech by adjusting their phoneme categories in a process called perceptual learning. Here, we focus on a special case of perceptual learning: listeners' ability to temporarily adjust their phoneme category boundaries [1] in response to ambiguous pronunciations of sounds using lexical and phonotactic information (e.g., [2][3], see for a review [4]). The retuning of phonemic categories helps listeners to understand new speakers and unfamiliar accents as it allows them to easily comprehend other words produced by those speakers [3].

Perceptual learning has been found for tones [5] and different types of sounds, e.g., for stops [6], fricatives [3][6][7][8][9], liquids [10][11][12], and vowels [13]. It has been found for younger listeners [14], older listeners [9][12], and for non-native listeners [10][15]. Lexically-guided perceptual learning has been found to be stronger the more often a listener considers words with an ambiguous sound as real words during exposure [12]; this suggests that there are differences in the amount of lexical guidance and subsequent category retuning between listeners. So, many factors that influence lexically-guided perceptual learning are clear. However, what remains unclear is what happens in the brain during lexically-guided perceptual learning; when and how is an ambiguous sound categorized as a (different) phoneme?

This paper is the first study that aims to shed light onto that question. Specifically, the current study aims to identify the neural correlates of lexically-guided perceptual learning in native and non-native listeners, as a first step towards constraining our understanding of the underlying neural mechanisms. Studies on the neural underpinnings of phoneme categorization have shown that event-related potentials (ERPs), aided by their millisecond temporal resolution, are well suited for the investigation of phoneme categorization during speech processing [16]. Hence, we collected ERPs during the experiment. Here, we report the non-native results.

In a typical lexically-guided perceptual learning paradigm [3], two listener groups are each exposed to the same ambiguous sound but in contrasting lexical environments, so that an ambiguous [f/s] in a context such as *loaf* is interpreted as /f/, while the same ambiguous [f/s] in a context such as *glass* is interpreted as /s/. If the ambiguous sound is presented consistently in contexts that always imply the same interpretation (always /s/ or always /f/), then listeners learn to accept the ambiguous token as a valid instantiation of the indicated phoneme. Lexically-guided perceptual learning is then observed by comparing the proportion of /s/ (or /f/) responses between the two listener groups in response to ambiguous stimuli (e.g., minimal pairs: *leaf/lease* or nonsense syllables: /ɛf-ɛs/), where the group exposed to [f/s] in the /f/-context gives significantly fewer /s/ responses than the group of listeners exposed to [f/s] in the /s/-context.

Here, we use a different set-up, wherein we investigate lexically-guided perceptual learning using a pretest-posttest design [7][17]. This allows us to use only one biasing context (the /f/ context in our case), which reduces the number of people that need to be tested by half. Lexically-guided perceptual learning is then observed by a significant decrease in /s/ responses from the pretest to the posttest.

2. Experimental Set-up

2.1. Participants

36 Dutch non-native listeners of English from the Radboud University, Nijmegen, the Netherlands subject pool were tested (15 males, mean age: 21.5, SD: 2.8). All participants were paid

http://dx.doi.org/10.21437/Interspeech.2019-2328

for participation in the experiment. No participants reported hearing or learning problems.

2.2. Materials

2.2.1. Exposure phase and short story

Twenty words containing /f/ and 20 words containing /s/ were chosen from SUBTLEX-US [18] (average frequency: 76.2 per million; range: 1-454 per million). All /f/ and /s/ appeared word-finally. Importantly, no other words contained /f/, so listeners never heard a normal /f/ but only ambiguous [f/s/] (as exposure to a normal instantiation of the sound reduces learning [19]). Table 1 lists the /f/- and /s/-final words. These 40 words were then used to compose a story, which contained no other words with an /f/. The final version of the story contained 438 words.

To ensure that the words were pronounced at a comparable speech rate, intonation, and style as the rest of the story, the target words were recorded as a part of the short story. In order to create the ambiguous stimuli, two versions of the story were recorded by a native female speaker of American English in a sound-damped booth using a Sennheiser ME 64 microphone. In one version of the text, the speaker produced all target words in their natural manner; in the second version, the speaker substituted all /f/ sounds with an /s/ (referred to as the reverse story). All versions of the text were recorded three times.

2.2.2. Creating the ambiguous stimuli

To create the ambiguous words in the short story, we followed the procedure in [10], in which two versions of each target word were morphed: one natural version of the target word (e.g., *loaf*) and another one where the target sound was substituted with its counterpart (e.g., *loas* (a nonword)).

To create the ambiguous items, the final /f/ sounds and the preceding vowel were excised at the positive-going zero crossings from the audio files of the natural short stories, using Praat [20]. Likewise, the final /s/ sounds and the preceding vowel from the corresponding reverse story were excised. Subsequently, the excised vowel+fricatives were zero-padded, so that there was 25 ms of silence at the beginning and at the end of the words. The pitch contours of the two items from each pair (e.g., *loaf-loas*) were equalized and the resulting words were morphed with the STRAIGHT algorithm [21] in MATLAB [22] to create an 11-step continuum from a version of the word where the interpretation of the ambiguous [f/s] sound was /f/ like (step 0) or /s/ like (step 10).

The most ambiguous sound was tested in a pilot study with 10 listeners (5M; mean age: 22.6; SD: 2.1) in which steps 1, 3, 5, 7, and 9 were presented auditorily in random order to the participants. Each stimulus was presented twice. The task for the participants was to indicate by button press as quickly and as accurately as possible whether the stimulus ended in /f/ or /s/. To help the listeners, the letter ‘f’ was always printed on the left side of the screen and ‘s’ on the right side of the screen. Subsequently, the total proportions of /f/ and /s/ responses to each of the steps of each of the continua were calculated. The most ambiguous step of each word individually was that step on the continuum that received approximately 50% of /f/ and 50% of /s/ responses. For eight words, a follow-up pilot was needed as the most ambiguous step was close to step 1 (3 participants; 3F; mean age: 22.3; SD: 2.5). Subsequently, the word with the most ambiguous step was spliced back into the story. None of the participants in the pilot studies participated in the main study or in another pilot.

Table 1. *The /f/ and /s/ words used in the short story.*

/f/ words	tough, bulletproof, loaf, sheriff, handkerchief, behalf, stiff, stuff, cliff, thief, safe, fluff, roof, staff, chef, mischief, belief, handcuff, proof, deaf
/s/ words	across, glorious, class, harmless, house, boss, less, dangerous, press, stress, precious, mouse, dress, pass, glass, kiss, nervous, chaos, worse, guess

2.2.3. Pretest and posttest

The stimuli for the pretest and posttest were identical and consisted of ambiguous versions of [ɛf]-[ɛs]. First, several versions of [ɛf] and [ɛs] were recorded by the same speaker who also produced the short story, and these items were morphed using the procedure described in the previous sub-section.

To determine the most ambiguous [f/s] items, all 11 [ɛf]-[ɛs]-continuum steps were each presented 10 times binaurally over closed headphones during the first pilot study. The task for the participants was to indicate by button press as quickly and as accurately as possible whether they heard [ɛf] or [ɛs]. The most ambiguous step was step 5. Consequently, steps 3 through 7 were used in the pretest and posttest.

2.3. Procedure

Participants were tested individually in a sound-proof booth. The intensity level of all stimuli (short story and words in the phoneme categorization task) was set at 60 dB SPL and was the same for all participants. The experiment was administered with Presentation software [23], and audio stimuli were presented binaurally through the same headphones as used for the pilot. Participants were comfortably seated in front of a computer screen in a sound-proof booth.

The pretest consisted of a phonetic categorization task in which the five steps of the [ɛf]-[ɛs]-continuum were auditorily presented 10 times each in random order (referred to as Block 1). As in the pilot study, listeners were asked to decide as quickly and accurately as possible whether they heard [ɛf] or [ɛs] by pressing the corresponding button.

Next, they saw an instruction on the computer screen informing them that they would be listening to a short story in English. To start the story, participants had to press a button. Once the story was finished, the participants were prompted to press another button to start the posttest. The posttest was identical to the pretest except it was presented twice with a different random order for both blocks (Block 2 is the first 50 stimuli; Block 3 is the second 50 stimuli). Thus, participants listened to 100 stimuli.

During the entire experiment, continuous EEG was collected from 32 active electrodes (Ag/AgCl, using the 10-10 system). Recordings were referenced to the left mastoid online and to the average of the left and right mastoids offline. Saccades and blinks were monitored by additional electrodes placed on the outer canthus of each eye and above and below the right eye. Impedances were generally kept below 5 KΩ, and never exceeded 10 KΩ. Data were sampled at 500 Hz and bandpass filtered online between .016 and 125 Hz.

3. Results

3.1. Behavioral study

Lexically-guided perceptual learning is defined as a significant reduction in /s/ responses from the pretest to the posttest. To determine learning, the percentage of /s/ responses in the pretest (Block 1) and the percentage of /s/ responses in the posttest

(Blocks 2 and 3) were compared using a z-test. The results showed that 18 participants showed a significant reduction in percentage of /s/ responses in the posttest ($z > 1.645$; 1-tailed, which corresponds to $p < 0.05$), while the other 18 participants did not show a significant reduction from pretest to posttest ($z < 1.645 = p > 0.05$). Note, two participants showed a significant increase from pretest to posttest, and these participants were grouped with the nonlearning group.

Figure 1 shows the proportion of /s/ responses for the group of participants who did not show learning (left panel) and the group of learners (right panel) for the three Blocks separately. The x-axis shows the step on the continuum of ambiguous sounds, with step 1 being a mixture of 30% /s/ and 70% /f/ and step 5 being a mixture of 70% /s/ and 30% /f/.

Ambiguous sounds were piloted to be ambiguous (pilot listeners labeled them as /s/ 50% of the time), but listeners in this study found them to be more /s/-like (88.4% /s/ responses in the pretest, averaged over all listeners). Despite the /s/ bias in the pretest, Figure 1 clearly shows the differences in response patterns between the participants in the learning and no learning group: where non-learning subjects exhibited highly similar proportions of /s/ responses for the pretest (black bullets) and the two blocks of the posttest, the proportion of /s/ responses for the two blocks for the posttest in the learning group is much lower than for the pretest. Table 2 shows the percentage of /s/ responses, averaged over all participants in the learning and nonlearning group, for the pretest and posttest separately. The difference between the learning and nonlearning groups (see bottom row) is 21 percentage points on average.

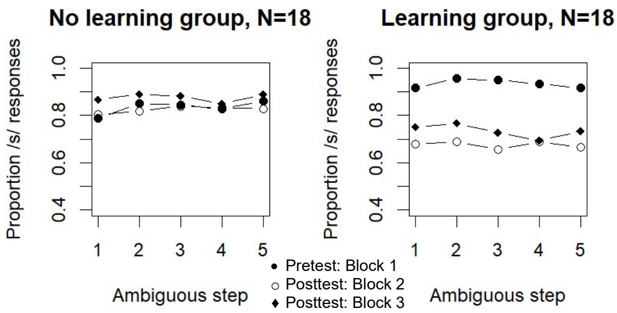

Figure 1. *Proportion of /s/ responses for the participants showing no lexically-guided perceptual learning and showing lexically-guided perceptual learning.*

Table 2. *Percentage of /s/ responses (and standard deviations) and ranges for the pretest and posttest over all participants and for the learning and nonlearning group separately, and the difference between the posttest and pretest.*

	Learning group		**Nonlearning group**	
	Mean (SD)	*Range*	*Mean (SD)*	*Range*
Pretest	93.4 (9.4)	64.0-100	83.4 (24.5)	22.0-85.0
Posttest	70.5 (17.0)	32.0-93.0	85.0 (20.3)	38.0-100.0
Difference	-22.9		1.6	

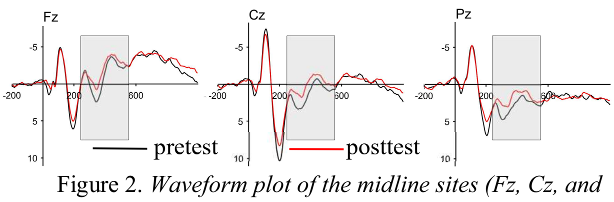

Figure 2. *Waveform plot of the midline sites (Fz, Cz, and Pz). Grey box indicates the ana lysis window.*

Table 3. *Mean amplitudes in microvolts (and standard deviations) for the pretest and posttest over all participants and for the learning and nonlearning group separately, and the difference between the posttest and pretest.*

	All	**Learning group**	**Nonlearning group**
Pretest	.9955 (3.7)	.8909 (3.1)	1.1002 (4.3)
Posttest	.0295 (2.9)	-.0052 (2.5)	.0642 (3.4)
Difference	-.9660	-0.8961	-1.0359

3.2. EEG study

In order to investigate the neural correlates underlying lexically-guided perceptual learning, we compared the EEG responses between the pretest and posttest. Epochs were extracted from 200 ms before the onset of frication (i.e., the first positive-going zero crossing of the frication) until 1000 ms after stimulus onset. Grand average mean amplitudes were calculated for the pretest stimuli and posttest stimuli (collapsed over Block 2 and Block 3) separately for each participant at each channel, after subtracting the prestimulus baseline, and averaged over all participants. To determine the region of interest (ROI), we visualized the grand average mean amplitudes for the pretest and posttest stimuli for each channel averaged over all participants. On the basis of this visual inspection, a broad window of 250-550 ms was chosen for medial central channels (FC 1/2, C 3/4/Z, CP 1/2, P 3/4/z), encompassing a P3b-like effect pattern in the data. Mean amplitudes in this ROI were computed for each participant and compared for the learning and nonlearning groups.

Figure 2 shows the waveform plot of the midline sites Fz, Cz, and Pz. The black and red lines show mean amplitudes at pretest and posttest, respectively, averaged over all participants. The difference between the two mean amplitudes in the window of interest increases towards the more parietal electrodes (i.e., from Fz to Pz).

Table 3 lists the mean amplitudes (and standard deviations) for the pretest and posttest over all participants and for the learning and nonlearning group separately. Moreover, differences between the mean amplitudes at pretest and posttest are presented. As Table 3 shows, the overall difference between the pretest and posttest is approximately 1 μV. A paired t-test showed that this difference was significant ($t(35)=2.436$, $p = 0.02008$). However, the difference between the pretest and posttest was similar for the learning and nonlearning groups ($t(17)=.182$, $p = .8579$).

3.3. Correlating the behavioral and EEG data

Previous studies have shown that listeners show different degrees of lexically-guided perceptual learning [12][17], which was also demonstrated in our behavioral results. To investigate whether the observed change in amplitude between pretest and posttest reflected lexically-guided perceptual learning, the size of the ERP effect was correlated by subjects with the degree of learning observed in the behavioral task. Specifically, the difference in P3b amplitude between the posttest and the pretest was correlated with the difference in percentage /s/ responses between the posttest and pretest. The correlational analysis showed a non-significant correlation of $r = 0.22$ ($t(34)=1.29$, $p = 0.205$, 95% confidence intervals [-0.1207399, 0.5087487]).

4. Discussion

In this study we used a pretest-posttest design to investigate lexically-guided perceptual learning. It has been argued that focusing listeners' attention on the ambiguity of a sound will inhibit this type of learning [9][19]. However, despite employing a phonetic categorization task at pretest and posttest which focused participants' attention on the ambiguous nature of the sound stimuli, we observed a clear lexically-guided perceptual learning effect, similar to [7][17], in about half of our participants. These results show that lexically-guided perceptual learning can take place under conditions where ambiguity is made explicit to the listener.

The number of participants showing no learning in our study is much higher than typically reported in the lexically-guided perceptual learning literature. This much larger group of nonlearners could possibly be attributed to the combination of two factors: focusing attention on the ambiguity [9][19] and phoneme representation bias. In lexically-guided perceptual learning experiments that do not use the pretest-posttest design, there is typically a small number of participants who do not show learning (e.g., [3][6][7][9]). This lack of learning is sometimes argued to be linked to the participants having a bias in their phoneme representations, such that they then observe ambiguous sounds as normal instantiations of that sound, which inhibits perceptual learning (e.g., [10]). Possibly, in addition to the small group of people who do not show lexically-guided perceptual learning due to a phoneme representation bias, there is a second group of people who do not show learning due to the focus of attention on the ambiguous sound. The variability in behavioral patterns is, however, useful for the purposes of looking for individual differences in neural responses, both from pretest to posttest and during the learning phase.

We observed a reliable change in ERP responses over the course of the experiment. The timing, distribution, and polarity of this effect was consistent with a P3b-like potential. The P3b component is a positive-going waveform sensitive to multiple aspects of attentionally-dependent stimulus evaluation and categorization (see [24] for review), reflects categorization of speech [25], and is influenced by distance to phonological category boundaries [16]. The current findings are in line with these previous findings. Given that lexically-guided perceptual learning entails a shift in the categorization boundary of the ambiguous stimuli, the amplitude change between pretest and posttest of the P3b component could index the amount of learning that occurred in response to the biasing context.

The P3b-like potential was larger in response to the pretest compared to the posttest sounds and was similar in size across the learning and nonlearning groups. Moreover, there was no significant correlation between the size of this effect and the amount of behavioral change. We thus cannot rule out the possibility that this ERP difference is not learning-related. For example, it could reflect a neural habituation response due to repeated exposure to the sound stimuli – although, in this case, it would be less likely that the observed effect is a P3b, as P3b responses typically do not show habituation [26]. Under this account, the decision to classify the ambiguous sound as /f/ or /s/ may be reflected in aspects of processing that are not time-locked or fall outside of our analysis window.

Nonetheless, it is still possible that the P3b amplitude reduction may partially reflect perceptual learning. In the pretest, both groups showed a strong bias to respond with /s/. Thus, although stimuli were piloted to be ambiguous, participants did not treat them as such. Possibly, during the pretest, the categorization of the sounds was relatively easy for the participants, since they heard them almost all as /s/. P3b responses are bigger for stimuli that are easier to categorize [27][28]. The exposure phase then seemed to make the stimuli more ambiguous, seen in the shift from strong /s/ bias toward mixed responding (see also the responses to Block 2 and Block 3 in Figure 1). Previous research has shown that if the sounds are harder to classify, the P3b will become smaller [29]. The fact that the pretest-posttest pattern shows up for both the learning group and the nonlearning group suggests that some people in the nonlearning group may also have started to hear the sounds as more ambiguous, but stuck with their original behavioral classification, unlike those in the learning group, who changed their explicit responding. If true, despite lexically-guided perceptual learning typically being indexed as a shift in responses to a phonetic categorization task on ambiguous items, phoneme category adaptation could then occur without an explicit change in categorization behavior.

Further examination of ERP response patterns during the exposure phase may assist us in teasing apart these competing explanations of the ERP change between pretest and posttest. Specifically, we plan to use both the level of behavioral change and the level of neurophysiological change as predictors of ERP patterns during the exposure phase. Thus, in the next step, we will investigate the neural responses to the 20 /f/ words in the story and examine whether changes in the neural responses to the ambiguous sounds can be observed over time and whether these are related to either the behavioral or ERP patterns observed for the pre-to-posttest comparisons. If responses to the ambiguous words are correlated with the amount of learning in the group that did demonstrate learning in their phonetic categorization behaviour, and similar responses are found in the nonlearning group, this could support an account in which the 'nonlearning' group did in fact adapt their phoneme category boundaries but failed to adjust their behavior at posttest.

5. Conclusions

We presented the first attempt to examine the neural correlates underlying lexically-guided perceptual learning, a process which allows listeners to quickly adapt their phoneme classes to ambiguous sounds. We compared the brain's responses to ambiguous sounds in Dutch non-native listeners of English before and after exposure to an ambiguous sound. At the level of categorization behavior, we found subsets of participants who show lexically-guided perceptual learning and those who do not. Moreover, we observed differences in the mean amplitude for the pretest and posttest in the ERP signal, in the form of a reduction in amplitude of a positivity over medial central channels from 250 to 550 ms. However, we did not observe a significant correlation between the behavioral and neural pretest/posttest effects. The observed differences between the pretest and posttest data will be further investigated by assessing their relationship to the neural responses to ambiguous words in the exposure phase.

6. Acknowledgements

This work was partially sponsored by a Vidi-grant from NWO (grant number: 276-89-003) and a Delft Technology Fellowship (TU Delft, the Netherlands) to O.S.. The authors are grateful to the Centre of Language Studies (CLS), Radboud University, Nijmegen for graciously providing the lab space for running the experiments, and to the student assistants at CLS who helped running the experiment.

7. References

[1] Clarke-Davidson, C., Luce, P. A., & Sawusch, J. R. (2008). Does perceptual learning in speech reflect changes in phonetic category representation or decision bias? *Perception & Psychophysics*, 70, 604–618.

[2] Cutler, A., McQueen, J.M., Butterfield, S., & Norris, D. (2008). Prelexically-driven perceptual retuning of phoneme boundaries. *Proceedings of Interspeech*, 2056-2056.

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

[4] Samuel, A. G., & Kraljic, T. (2009) Perceptual learning in speech perception. *Attention, Perception & Psychophysics*, 71, 1207-1218.

[5] Mitterer, H., Chen, Y., & Zhou, X. L. (2011). Phonological Abstraction in Processing Lexical-Tone Variation: Evidence From a Learning Paradigm. *Cognitive Science*, 35, 184-197. doi:10.1111/j.1551-6709.2010.01140

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

[7] Eisner, F., & McQueen, J. M. (2006). Perceptual learning in speech: Stability over time. *Journal of the Acoustical Society of America*, 119 (4). https://doi.org/10.1121/1.2178721

[8] Kraljic, T., & Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology*, 51, 141-178.

[9] Scharenborg, O., Weber, A., & Janse, E. (2015). The role of attentional abilities in lexically-guided perceptual learning by older listeners. *Attention, Perception, & Psychophysics*, 77 (2), 493–507. https://doi.org/10.3758/s13414-014-0792-2

[10] Drozdova, P., van Hout, R., & Scharenborg, O. (2016). Lexically-guided perceptual learning in non-native listening. *Bilingualism: Language and Cognition*, 19 (5), 914–920. doi:10.1017/S136672891600002X

[11] Scharenborg, O., Mitterer, H., & McQueen, J.M. (2011) Perceptual learning of liquids. *Proceedings of Interspeech,* Florence, Italy.

[12] Scharenborg, O., & Janse, E. (2013). Comparing lexically-guided perceptual learning in younger and older listeners. *Attention, Perception, and Psychophysics*, 75 (3), 525-536. doi: 10.3758/s13414-013-0422-4.

[13] McQueen, J. & Mitterer, H. (2005). Lexically-driven perceptual adjustments of vowel categories. *Proceedings of ISCA Workshop on Plasticity in Speech Perception*. Geneva, pp. 233-236.

[14] McQueen, J. M., Tyler, M., & Cutler, A. (2012). Lexical retuning of children's speech perception: Evidence for knowledge about words' component sounds. *Language Learning and Development*, 8, 317– 339.

[15] Reinisch, E., Weber, A., & Mitterer, H. (2013). Listeners retune phoneme categories across languages. *Journal of Experimental Psychology: Human Perception and Performance*, 39, pp. 75-86. http://dx.doi.org/10.1037/a0027979

[16] Toscano, J.C., McMurray, B., Dennhardt, J., & Luck, S.J. (2010). Continuous perception and graded categorization: Electrophysiological evidence for a linear relationship between the acoustic signal and perceptual encoding of speech. *Psychological Science,* 21, 1532-1540.

[17] Colby, S., Clayards, M., & Baum, S. (2018). The role of lexical status and individual differences for perceptual learning in younger and older adults. Journal of Speech, Language and Hearing Research. 61(8), 1855–1874 10.1044/2018_JSLHR-S-17-0392

[18] Brysbaert, M. & New, B. (2009) Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, 41 (4), 977-990.

[19] McAuliffe, M., & Babel, M. (2016). Stimulus-directed attention attenuates lexically-guided perceptual learning. *Journal of the Acoustical Society of America*, 140 (3), 1727–1738. https://doi.org/10.1121/1.4962529

[20] Boersma, P., & Weenink, D. (2005) Praat. Doing phonetics by computer (Version 5.1).

[21] Kawahara, H., Masuda-Katsuse, I., & Cheveigne, A (1999) 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.

[22] Matlab "R 2013a" (Software). *The MathWorks Inc.*

[23] Presentation "17.0" (Software). *Neurobehavioural Systems Inc.*

[24] Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical neurophysiology*, 118 (10), 2128-2148.

[25] Maiste, A.C., Wiens, A.S., Hunt, M.J., Scherg, M., & Picton, T.W.(1995). Event-related potentials and the categorical perception of speech sounds. *Ear and Hearing*, 16, 68–89.

[26] Debener, S., Kranczioch, C., Herrmann, C.S., & Engel, A.K. (2002). Auditory novelty oddball allows reliable distinction of top-down and bottom-up processes of attention. *Int. J. Psychophysiol.*, 46, 77-84.

[27] Johnson, R. (1986). A triarchic model of P300 amplitude. *Psychophysiology*, 23, 267–384. doi: 10.1111/j.1469-8986.1986.tb00649.x

[28] Kok, A. (2001). On the utility of P3 amplitude as a measure of processing capacity. *Psychophysiology,* 38, 557–577. doi: 10.1017/S0048577201990559

[29] Comerchero, M.D., & Polich, J. (1999). P3a and P3b from typical auditory and visual stimuli. *Clinical Neurophysiology,* 110, 24–30. doi: 10.1016/S0168-5597(98)00033-1.