Cognition 141 (2015) 121–126

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Recalibration of auditory phonemes by lipread speech is ear-specific

Mirjam Keetels, Mauro Pecoraro, Jean Vroomen *

Department of Cognitive Neuropsychology, Tilburg University, The Netherlands

ARTICLE INFO

Article history:
Received 4 February 2014
Revised 21 April 2015
Accepted 24 April 2015
Available online 14 May 2015

Keywords:
Speech perception
Phonetic recalibration
Lipreading
Generalization

ABSTRACT

Listeners quickly learn to label an ambiguous speech sound if there is lipread information that tells what the sound should be (i.e., phonetic recalibration Bertelson, Vroomen, & de Gelder (2003)). We report the counter-intuitive result that the same ambiguous sound can be simultaneously adapted to two opposing phonemic interpretations if presented in the left and right ear. This is strong evidence against the notion that phonetic recalibration involves an adjustment of abstract phoneme boundaries. It rather supports the idea that phonetic recalibration is closely tied to the sensory specifics of the learning context.

1. Introduction

Variability in the acoustic realization of speech segments is a well-known phenomenon that can arise from several sources, including individual differences between talkers and different acoustic environments. Despite this variability, listeners are remarkably successful at understanding and decoding speech. Recent evidence suggests that listeners may accommodate this variability by quickly adjusting the boundary between two phoneme categories if there is contextual information that disambiguates an otherwise ambiguous sound (phonetic recalibration). One source of contextual information stems from lipread speech, as listeners are more likely to label an ambiguous sound (/?/) halfway between /b–d/ as /b/ if the ambiguous sound was previously presented in the context of lipread /b/ rather than lipread /d/ (Bertelson, Vroomen, & de Gelder, 2003). Similar effects have been reported if the context consists of lexical knowledge about the possible words in the language (Kraljic & Samuel, 2005; Norris, McQueen, & Cutler, 2003; van Linden & Vroomen, 2007). Presumably, contextual information thus can 'teach' the auditory system how to categorize an ambiguous sound. Further research has demonstrated that phonetic recalibration by lipread speech is a rather automatic process (Baart & Vroomen, 2010) that actually leaves neural traces in auditory cortex (Kilian-Hutten, Valente, Vroomen, & Formisano, 2011a; Kilian-Hutten, Vroomen, & Formisano, 2011b).

A fundamental question about phonetic recalibration is the extent to which it generalizes to other tokens, speakers, and listening conditions. For lexically-induced recalibration, several authors argued that recalibration is used to fine-tune speech recognition to the idiosyncrasies of individual speakers. Generalization occurs if the ambiguous sound and its acoustic environment likely belong to the adapted speaker, but there is no generalization if the sound and its environment are too dissimilar (Cutler, Eisner, McQueen, & Norris, 2010; Eisner & McQueen, 2005; Kraljic & Samuel, 2006; Reinisch & Holt, 2014). *Lipread*-induced recalibration has also been found to be rather speaker- or even token-specific (van der Zande, Jesse, & Cutler, 2014). For example, Van der Zande et al. used an /aba–ada/ continuum to examine generalization

* Corresponding author at: P.O. Box 90153, 5000 LE, Tilburg, The Netherlands.

E-mail address: j.vroomen@uvt.nl (J. Vroomen).

http://dx.doi.org/10.1016/j.cognition.2015.04.019

of lipread-induced recalibration across two male speakers. The authors found large within-speaker effects, but lip-read induced recalibration was much smaller (though still significant) if the test stimuli came from the non-adapted speaker. Reinisch, Wozny, Mitterer, and Holt (2014) also reported robust lipread-induced recalibration if the same acoustic tokens were used during exposure and test, but there was no generalization across different vowel contexts (e.g., audiovisual exposure on /aba–ada/ had no effect on identification of /ibi–idi/), or across manner of articulation (e.g., exposure on /aba–ada/ did not affect /ama–ana/). This made the authors argue that allophones (context-specific phonemes) might be the targets of recalibration rather than (speaker-specific) phonemes.

To examine these views in more detail, we developed a new procedure that allowed us to examine whether the same ambiguous sound can be *simultaneously* adapted to two opposing phonemic interpretations. Thus, rather than testing for generalization of recalibration to new tokens, contexts, speakers or allophones, we exposed the left and right ear to *opposing* interpretations of the same sound. The critical question was whether phonetic recalibration is *ear-specific*. If recalibration targets abstract phonemes, this would be precluded because phonemes are central and abstract representations without separate tags for the left and right ear. A speaker- or token-specific account also predicts that acoustically identical tokens will be assigned to the same speaker or token, and the different training regimes for the left and right ear will then cancel each other. Possibly, though, lipread-induced recalibration is very specific for acoustic details. In the visual domain, psychophysical findings indeed indicate that perceptual learning is often restricted to the visual features that have been trained, such as a particular location in the visual field, the spatial configuration of the stimulus elements, or the specific eye (Crist, Kapadia, Westheimer, & Gilbert, 1997). Possibly, lipread-induced recalibration is equally specific for low-level details, and in this view it might then be possible that the same sound is categorized as /b/ if presented in the /b/-adapted ear, and as /d/ if presented in the /d/-adapted ear.

Here, we examined this hypothesis using the same stimuli as in Bertelson et al. (2003; Experiment 2). During a short exposure phase, an auditory ambiguous sound halfway between /b/ and /d/ (A?) was presented in alternating fashion in the left and right ear via headphones, while a left–right alternating video of a speaker was shown who articulated /b/ or /d/ (Vb or Vd). The A? was thus concurrently recalibrated toward /b/ in one ear by A?Vb, and toward /d/ in the other ear by A?Vd. After a short audiovisual exposure phase, auditory-only test sounds near the phoneme boundary were presented separately in each ear.

In order to control for a simple response bias or a 'prior' (Kleinschmidt & Jaeger, 2011) that reflects that one particular phoneme was heard at a specific side (e.g., participants respond /d/ simply because they heard /d/'s in the left ear in the foregoing exposure phase), we included, as in Bertelson et al. (2003; Experiment 2), audiovisual exposure stimuli that do *not* induce recalibration, namely audiovisual congruent stimuli with auditory *non*-ambiguous sounds: AbVb and AdVd. These sounds are perceptually very similar to their ambiguous counterpart (A?Vb and AbVb are both perceived as /b/; A?Vd–AdVd are both perceived as /d/) because the lipread information strongly captures the identity of the sound (Vroomen, van Linden, de Gelder, & Bertelson, 2007). Nevertheless, AbVb and AdVd do not induce recalibration because there is no conflict between the heard and lipread information that induces a shift in the phoneme boundary. In previous studies, these stimuli have sometimes induced contrastive aftereffects (i.e., *fewer* /b/ responses after exposure to AbVb than AdVd) indicative of selective speech adaptation (Eimas & Corbit, 1973), but this effect is usually very small as selective speech adaptation requires much larger amounts of exposure (Vroomen et al., 2007). Here, we nevertheless included these stimuli because they were ideal to serve as a baseline that controls for response bias.

2. Experiment 1

2.1. Method

Participants were sixteen students (4 male) from Tilburg University who received course credits in return (sample size was chosen based on previous research on phonetic recalibration, Bertelson et al., 2003). The study was approved by the ethics committee of Tilburg University, and was conducted in accordance with the ethical standards of the 1964 Declaration of Helsinki. All experiments were undertaken with the understanding and written consent of each subject

Stimuli have been described in Bertelson et al. (2003). In short, the video showed the face of a male native speaker of Dutch pronouncing /aba/ or /ada/. The videos were converted into strings of 71 bitmaps, each displayed for 30 ms on a 17 in CRT monitor (refresh rate 100 Hz, 800×600 pix resolution with 80-ms fade-in and 100-ms fade-out). The videos were presented against a dark background in a left–right alternating fashion (ISI = 2130 ms) at 9.2° from the center in one of two static gray frames (9.2° × 7.0° at 50 cm viewing distance; see Fig. 1). The original audio recordings were synthesized using the Praat program (Boersma & Weenink, 2005) to produce two 640-ms-long synthetic stimuli with 240-ms stop closure. Varying the frequency of the F2 formant by equal steps of 39 Mel provided a nine-step place-of-articulation continuum. The middle sound of the continuum was dubbed onto the two video recordings giving pairs A?Vb and A?Vd, and the two extremes of the continuum were dubbed in an audiovisual congruent fashion giving pairs AbVb and AdVd.

The experiment consisted of 64 'mini-blocks'. Each mini-block comprised 12 audiovisual exposures followed by 6 auditory-only test trials. During exposure, the face articulating /aba/ and /ada/ was presented in alternating fashion on the left and right side (6 exposures at each side). The sounds (ambiguous or non-ambiguous) were presented in the same left–right alternating fashion via closed headphones (Sennheiser HD201). Dedicated software ensured accurate audiovisual synchronization within a single refresh rate. Half of the mini-blocks contained the

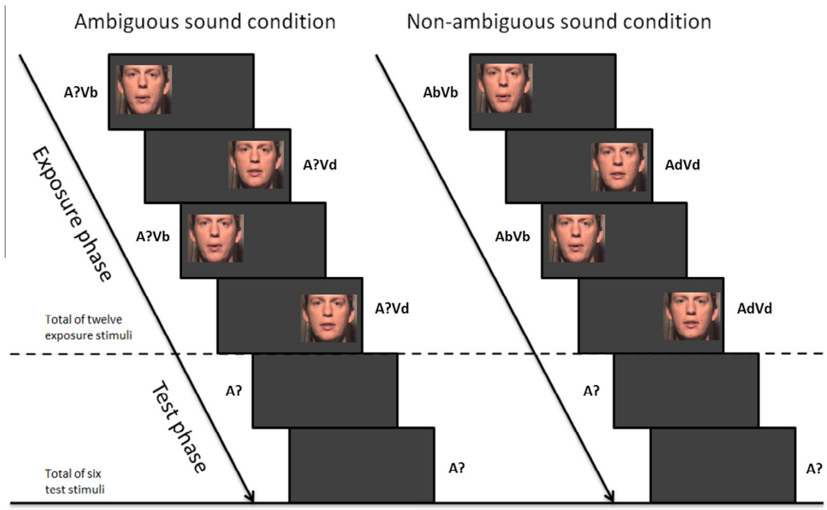

Fig. 1. Example of a mini-block. Each mini-block consisted of an exposure phase of 12 left–right alternating audiovisual stimuli followed by 6 auditory-only test sounds. Participants reported by key-press whether they heard /b/ or /d/ at test.

auditory ambiguous sounds yielding A?Vb and A?Vd, the other half contained the auditory non-ambiguous sounds yielding AbVb and AdVd. The ambiguity of the exposure sound (ambiguous or non-ambiguous) switched after 8 consecutive mini-blocks while the identity of the exposure stimulus on the left (/b/ or /d/) varied randomly after each mini-block. Participants were asked to watch the video during exposure, and this was monitored by the experimenter via a closed camera circuit.

Following exposure, there was an auditory-only test during which the 3 sounds closest to the phoneme boundary (A?-1, A? and A?+1 for sounds closer to the /b/ and /d/ boundary, respectively) were presented once to each ear in a random fashion, thus yielding 6 test trials per mini-block (3 on the left, 3 on the right). The screen remained dark during test. Participants decided whether the test sound was /aba/ or /ada/ by pressing the ‘b’- or ‘d’-key on a standard keyboard. The next trial was delivered 1s after a response was detected

2.2. Results

One participant was removed because of incomplete data. Of the remaining participants, the proportion of /d/-responses was calculated as a function of whether the ear was exposed to A?Vb, A?Vd, AbVb or AdVd. Fig. 2 shows the group-averaged proportion of /d/-responses for each condition. Most importantly, there were substantially more /d/-responses if the test sound was delivered in the ear exposed to auditory ambiguous A?Vd than A?Vb (a 12% difference). In the auditory non-ambiguous baseline condition, this difference (AdVd vs. AbVb) was much smaller (4%). This was statistically confirmed in a 2 (Exposure sound ambiguous/non-ambiguous) × 2 (Ear exposed to /b/ or /d/) × 3 (Test sound) overall analysis of variance (ANOVA) on the log-odds transformed proportion of /d/-responses. There were significant effects of Ear, $F(1, 14) = 30.55$, $p < .001$, Test sound, $F(2, 13) = 44.95$, $p < .001$, and no interaction between Ear and Test sound, $F(2, 13) = .71$, $p = .51$. The theoretically important interaction between the Ambiguity of the exposure sound and whether the Ear was exposed to /b/ or /d/ was significant, $F(1, 14) = 6.89$, $p < .05$. Separate post hoc t-test (Bonferroni corrected) revealed that the 12% difference between the A?Vb- and A?Vd-exposed ear was greater than zero, $t(14) = 7.49$, $p < .001$, while the 4% difference between AbVb- and AdVd-exposed ear was not different from zero, $t(14) = 2.28$, $p < .078$).

2.3. Discussion

Experiment 1 demonstrates that the left and right ear can be simultaneously recalibrated to a different phonemic interpretation of the same ambiguous sound. This is strong evidence that recalibration carries information about the ear at which it was adapted. In Experiment 2, we further examined this idea by changing the exposure procedure so that we did not need a baseline for response biases.

3. Experiment 2

Previous research has demonstrated that an auditory non-ambiguous stimulus from the *opposite* category to which an ambiguous phoneme is calibrated (a contrast stimulus) can enhance recalibration (Vroomen et al., 2007; Experiment 3). As an example, recalibration of A? toward /d/ by exposure to A?Vd is enhanced if during exposure the contrast stimulus AbVb is present. In research on lexical recalibration, this use of contrast stimuli is in fact a standard procedure (e.g., Norris et al., 2003). Possibly, this enhancement occurs because the contrast stimulus serves as an anchor for recalibration, though other explanations remain possible. Here, we mixed the adapting and contrast stimuli (A?Vd + AbVb) for adaptation toward /d/ , and (A?Vb + AdVd) for adaptation toward

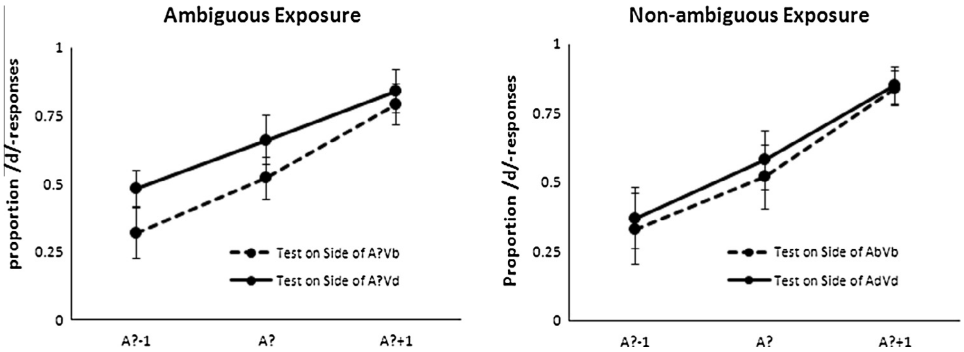

Fig. 2. Results of Experiment 1: *Left panel.* The proportion of /d/-responses after simultaneous exposure to auditory ambiguous sounds A?Vd and A?Vb as a function of the test sounds. There were more /d/-responses if test sounds were delivered in the ear exposed to A?Vd than A?Vb. Bars represent 95% confidence intervals. *Right panel.* The same data after exposure to auditory non-ambiguous sounds AbVb and AdVd. With non-ambiguous sounds, there was no difference between the AdVd- and AbVb-exposed ears.

/b/. Participant thus received the /d/-mix in one ear, and the /b/-mix in the other ear. The advantage of this is that *equal* amounts of /b/ and /d/ are perceived in the left and right ear. Post hoc interviews confirmed that participants were completely unaware of the different mixes, and explicit response strategies that might contaminate the picture seem extremely unlikely.

3.1. Method

Fifteen new students (6 male) from Tilburg University participated. Stimuli and procedures were as in Experiment 1, except that stimuli with ambiguous and non-ambiguous sounds were mixed. For recalibration toward /d/, we presented 3 VdA? + 3 VbAb (the '/d/-mix') in alternating order; for recalibration toward /b/ we used 3VbA? + 3 VdAd (the '/b/-mix'). Each mini-block contained a /b/- and a /d/-mix (12 exposure stimuli in total), each presented to one ear in left–right alternating order. There were 32 mini-blocks in total. Half of the participants received the /b/-mix in their left ear and the /d/-mix in their right ear, for the other half the order was reversed

3.2. Results

The proportion of /d/-responses was calculated for each participant and for each ear as to whether it was exposed to the /b/- or /d/-mix (see Fig. 3). Most importantly, there were substantially more /d/-responses if the test sound was delivered in the ear previously exposed to the /d/-mix rather than /b/-mix (a 17% difference). In the 2 (Ear exposed to /b/- or /d/-mix) $\times$ 3 (Test sound) ANOVA on the log-odds transformed proportion of /d/-responses there was a significant effect of auditory test stimulus, $F(2, 13) = 57.16$, $p < .001$, and a significant effect of whether the Ear was exposed to the /b/- or /d/-mix, $F(1, 14) = 7.56$, $p = .02$, reflecting the 17% difference. The interaction between Ear and Test sound was not significant, $F(2, 13) = .83$, $p = .46$. These results thus again confirm that the two ears can be simultaneously recalibrated toward /b/ and /d/.

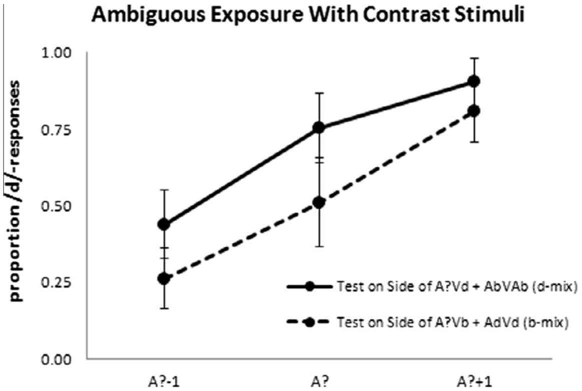

Fig. 3. Results of Experiment 2. The proportion of /d/-responses after simultaneous exposure to A?Vd + AbVb (the /d/-mix) in one ear and A?Vb + AdVd (the /b/-mix) in the other ear. Bars represent 95% confidence intervals. There were more /d/-responses if test sounds were delivered in the ear that had received the /d/-mix than the /b/-mix.

4. General discussion

Listeners quickly learn to label an ambiguous speech sound in accordance with lipread information that tells what the sound should be. Here, we demonstrate that the left and right ear can be *simultaneously* adapted to two opposing phoneme categories: The same ambiguous sound was thus perceived as /b/ if presented in the /b/-adapted ear, and as /d/ if presented in the /d/-adapted ear. This counter-intuitive result of ear-specific recalibration is compelling evidence that lipread-induced phonetic recalibration is closely tied to the sensory specifics of the learning context.

Previous studies already found that lipread- (Reinisch et al., 2014) and lexically-induced phonetic recalibration (Mitterer, Scharenborg, & McQueen, 2013) is more context- or token-specific than previously thought. For example, Reinisch et al. reported that listeners readily adjust their phonetic boundary if the same acoustic tokens /aba–ada/ tokens were used during exposure and test, but there was a lack of generalization across different vowel contexts or different articulatory features. These null-findings for generalization might in principle be attributed to a lack of power, but this argument seems untenable in the light of the present results where the same sound was simultaneously adapted in two different directions. Apparently, phonetic recalibration is not only token- and context-specific, but also depends on ear-of-exposure. The latter finding is particularly surprising if one realizes that auditory learning (of e.g., frequency discrimination) usually transfers almost completely across ears (Delhommeau, Micheyl, & Jouvent, 2005).

A theoretically important question, then, is why phonetic recalibration is so sensory-specific. In visual research, it has often been found that perceptual learning is highly specific for low-level sensory-specific features like frequency-specific orientation and location of visual stimuli that can be linked to the tuning curves of cells in (primary) visual cortex V1 (Fahle, 2005). For auditory speech, though, such detailed ideas about neural implementation are currently lacking, and theorizing is therefore more open for speculation. One argument for specificity of recalibration is that it is a mechanism that allows listeners to accommodate unusual speech input. For any learning mechanism, there must be a balance between, on the one hand, being flexible and being able to generalize to new tokens, speakers, and acoustic environments, and, on the other hand, being robust and to restrict adjustments to the local context so as to avoid catastrophic interference (Samuel, 2011). Our results highlight that speech recognition rather prefers to maximize stability over its lifelong experience, and that learning is thus very specific to the situation encountered.

Our results also shed new light on the findings by Eisner and McQueen (2005) who argued that recalibration by lexical information is speaker-specific. They exposed participants to an ambiguous sound halfway between /f/ and /s/ of a female speaker. The ambiguous sound occurred at the end of Dutch words that normally end in /f/ (e.g., *witlof*, meaning 'chicory') or /s/ (e.g., *naaldbos*, 'pine forest'). Listeners who heard the ambiguous sound in the context of /f/-final words later categorized more items on an /ef/–/es/ continuum as /f/ than listeners who heard this sound in /s/-final words. Critically, the vowel of the test stimulus (/e/) was either produced by the same female speaker that participants had been exposed to, or it was replaced by a different but similar female voice, or by a male speaker's voice. The fricative portion was always produced by the original female speaker. Although participants in the latter two groups believed that the test items were produced by a new speaker, recalibration was obtained for all three groups. In contrast, when the entire test token, including the fricative portion, was produced by a male speaker, no recalibration was obtained. In the light of the present result, these findings are understandable because they suggest that the *sensory* match of the /s/–/f/ portion was critical, rather than the higher-level information about the speaker's identity.

This is not to say, though, that higher-order information plays no role. Kraljic, Samuel, and Brennan (2008) examined lexical recalibration of a fricative place (/s/–/sh/) contrast, and reported that lexical recalibration occurred if the ambiguous /s/–/sh/ sound was embedded in disambiguating words like *episode* or *vacation*. Interestingly, listeners did not recalibrate if the odd pronunciation could be attributed to an alternative factor, such as a pen in the speaker's mouth. It remains to be examined whether such higher-order factors are at stake in the present situation. One relevant factor might be the identity of the speaker's face. We used a video of the same speaker that alternated between the left and right. Listeners likely attributed this to a single speaker. Alternatively, though, one might argue that the sounds in the left and right ear were attributed to the two individuals of an identical twin. Listeners then may maintain a different interpretation of the same sound as long as the individual speakers can be separated in space. To check this, future studies might examine whether different interpretations of the same sound can be maintained if the speaker does not change position, and whether recalibration becomes more manifest if the same ambiguous sound is dubbed onto two different faces.

It also remains to be examined to which extent lipread-induced recalibration is different from lexical recalibration. Many studies have demonstrated that lipread speech and lexical information bias the phonetic percept of auditory speech via the "McGurk" effect and the "Ganong" effect, respectively. Surprisingly, though, a McGurk percept does not induce selective speech adaptation like the Ganong-effect does (Pufahl & Samuel, 2014). Also, lipread speech seems to be able to influence compensation for coarticulation (Mitterer, 2006; but see Holt, Stephens, & Lotto, 2005; Vroomen & de Gelder, 2001), whereas lexical influences seem problematic to replicate. Lipread speech and lexical context thus operate differently, and it remains for future studies to examine whether lexical recalibration is as closely tied to the sensory details as lipread-induced recalibration appears to be.

Author contribution

JV and MK developed the study concept. All authors contributed to the study design. Testing and data collection were performed by MK and MP. JV drafted the manuscript. All authors approved the final version of the manuscript for submission.

References

Baart, M., & Vroomen, J. (2010). Phonetic recalibration does not depend on working memory. *Experimental Brain Research, 203*(3), 575–582.

Bertelson, P., Vroomen, J., & de Gelder, B. (2003). Visual recalibration of auditory speech identification: A McGurk aftereffect. *Psychological Science, 14*(6), 592–597. http://dx.doi.org/10.1046/j.0956-7976.2003.psci_1470.x.

Boersma, P., & Weenink, K. (2005). Praat: Doing phonetics by computer (Version Version 4.3.14). Retrieved from www.praat.org.

Crist, R. E., Kapadia, M. K., Westheimer, G., & Gilbert, C. D. (1997). Perceptual learning of spatial localization: specificity for orientation, position, and context. *Journal of Neurophysiology, 78*(6), 2889–2894.

Cutler, A., Eisner, F., McQueen, J., & Norris, D. (2010). How abstract phonemic categories are necessary for coping with speaker-related variation. In C. Fougeron, B. Kűhnert, & D. I. M. N. Valée (Eds.), *Laboratory phonology 10*. Berlin: de Gruyter.

Delhommeau, K., Micheyl, C., & Jouvent, R. (2005). Generalization of frequency discrimination learning across frequencies and ears: Implications for underlying neural mechanisms in humans. *Journal of the Association for Research in Otolaryngology, 6*(2), 171–179. http://dx.doi.org/10.1007/s10162-005-5055-4.

Eimas, P. D., & Corbit, J. D. (1973). Selective adaptation of linguistic feature detectors. *Cognitive Psychology, 4*, 99–109.
Eisner, F., & McQueen, J. M. (2005). The specificity of perceptual learning in speech processing. *Perception & Psychophysics, 67*(2), 224–238.
Fahle, M. (2005). Perceptual learning: Specificity versus generalization. *Current Opinion in Neurobiology, 15*(2), 154–160. http://dx.doi.org/10.1016/j.conb.2005.03.010.
Holt, L. L., Stephens, J. D. W., & Lotto, A. J. (2005). A critical evaluation of visually moderated phonetic context effects. *Perception & Psychophysics, 67*(6), 1102–1112. http://dx.doi.org/10.3758/bf03193635.
Kilian-Hutten, N., Valente, G., Vroomen, J., & Formisano, E. (2011a). Auditory cortex encodes the perceptual interpretation of ambiguous sound. *Journal of Neuroscience, 31*(5), 1715–1720. http://dx.doi.org/10.1523/Jneurosci.4572-10.2011.
Kilian-Hutten, N., Vroomen, J., & Formisano, E. (2011b). Brain activation during audiovisual exposure anticipates future perception of ambiguous speech. *Neuroimage, 57*(4), 1601–1607. http://dx.doi.org/10.1016/j.neuroimage.2011.05.043.
Kleinschmidt, D. F., & Jaeger, T. F. (2011). A Bayesian belief updating model of phonetic recalibration and selective adaptation. In *Paper presented at the 2nd ACL workshop on cognitive modeling and computational linguistics*.
Kraljic, T., & Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology, 51*(2), 141–178.
Kraljic, T., & Samuel, A. G. (2006). Generalization in perceptual learning for speech. *Psychonomic Bulletin & Review, 13*(2), 262–268.
Kraljic, T., Samuel, A. G., & Brennan, S. E. (2008). First impressions and last resorts – how listeners adjust to speaker variability. *Psychological Science, 19*(4), 332–338.
Mitterer, H. (2006). On the causes of compensation for coarticulation: Evidence for phonological mediation. *Perception & Psychophysics, 68*(7), 1227–1240. http://dx.doi.org/10.3758/bf03193723.
Mitterer, H., Scharenborg, O., & McQueen, J. M. (2013). Phonological abstraction without phonemes in speech perception. *Cognition, 129*(2), 356–361. http://dx.doi.org/10.1016/j.cognition.2013.07.011.
Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology, 47*(2), 204–238.
Pufahl, A., & Samuel, A. G. (2014). How lexical is the lexicon? Evidence for integrated auditory memory representations. *Cognitive Psychology, 70*, 1–30. http://dx.doi.org/10.1016/j.cogpsych.2014.01.001.
Reinisch, E., & Holt, L. L. (2014). Lexically guided phonetic retuning of foreign-accented speech and its generalization. *Journal of Experimental Psychology: Human Perception and Performance, 40*(2), 539–555. http://dx.doi.org/10.1037/a0034409.
Reinisch, E., Wozny, D. R., Mitterer, H., & Holt, L. L. (2014). Phonetic category recalibration: What are the categories? *Journal of Phonetics, 45*, 91–105. http://dx.doi.org/10.1016/j.wocn.2014.04.002.
Samuel, A. G. (2011). Speech perception. *Annual Review of Psychology, 62*(62), 49–72.
van der Zande, P., Jesse, A., & Cutler, A. (2014). Cross-speaker generalisation in two phoneme-level perceptual adaptation processes. *Journal of Phonetics, 43*, 38–46. http://dx.doi.org/10.1016/j.wocn.2014.01.003.
van Linden, S., & Vroomen, J. (2007). Recalibration of phonetic categories by lipread speech versus lexical information. *Journal of Experimental Psychology: Human Perception and Performance, 33*, 1483–1494.
Vroomen, J., & de Gelder, B. (2001). Lipreading and the compensation for coarticulation mechanism. *Language and Cognitive Processes, 16*(5–6), 661–672.
Vroomen, J., van Linden, S., de Gelder, B., & Bertelson, P. (2007). Visual recalibration and selective adaptation in auditory-visual speech perception: Contrasting build-up courses. *Neuropsychologia, 45*(3), 572–577.