Language and Speech
http://las.sagepub.com

Recalibration of Phonetic Categories by Lipread Speech: Measuring Aftereffects After a 24-hour Delay

Jean Vroomen and Martijn Baart

Language and Speech 2009; 52; 341
DOI: 10.1177/0023830909103178

The online version of this article can be found at:
http://las.sagepub.com/cgi/content/abstract/52/2-3/341

Published by:
SAGE
http://www.sagepublications.com

Recalibration of Phonetic Categories by Lipread Speech: Measuring Aftereffects After a 24-hour Delay

Jean Vroomen, Martijn Baart

Tilburg University, The Netherlands

Key words

aftereffects

audiovisual speech

lipreading

recalibration

Abstract

Listeners hearing an ambiguous speech sound flexibly adjust their phonetic categories in accordance with lipread information telling what the phoneme should be (recalibration). Here, we tested the stability of lipread-induced recalibration over time. Listeners were exposed to an ambiguous sound halfway between /t/ and /p/ that was dubbed onto a face articulating either /t/ or /p/. When tested immediately, listeners exposed to lipread /t/ were more likely to categorize the ambiguous sound as /t/ than listeners exposed to /p/. This aftereffect dissipated quickly with prolonged testing and did not reappear after a 24-hour delay. Recalibration of phonetic categories is thus a fragile phenomenon.

1 Introduction

Lipread information can help listeners by telling how to interpret a speech sound that initially might be ambiguous. Imagine, for example, a speaker who pronounces an ambiguous sound intermediate between /b/ and /d/ in the context of the sentence “Could you please pass me the b/dutter.” By looking at the speaker’s face, listeners may notice that the lips were closed during pronunciation of the ambiguous sound, which is typical for /b/ but not for /d/. Moreover, there is also lexical knowledge informing the listener that the ambiguous sound should be /b/ rather than /d/, because “butter” but not “dutter” is a word in English. Numerous studies have shown that when listeners are asked to categorize the ambiguous sound, they do indeed use lipread and lexical information (Ganong, 1980; Sumby & Pollack, 1954). In addition, there is not only an immediate or online effect of the context on sound categorization, but there is also an aftereffect because the next time listeners hear the same sound, they have learned from the past and now perceive the initially ambiguous “b/d” sound as /b/ right away (Bertelson, Vroomen, & de Gelder, 2003; Eisner & McQueen, 2005, 2006; Kraljic & Samuel, 2005, 2006, 2007; Norris, McQueen, & Cutler, 2003; van Linden & Vroomen, 2007; Vroomen, van Linden; de Gelder, & Bertelson, 2007; Vroomen, van Linden, Keetels,

Address for correspondence. Jean Vroomen, Tilburg University, Dept. of Psychology, Warandelaan 25000 LE Tilburg, The Netherlands; e-mail: <j.vroomen@uvt.nl>

Language and Speech 0023-8309; Vol 52(2/3): 341–350; 103178; DOI:10.1177/0023830909103178
http://las.sagepub.com

de Gelder, & Bertelson, 2004). The occurrence of this aftereffect demonstrates that listeners have adjusted the phonetic categories of their language so as to adapt to the new sound. What is at present unknown, though, is how long this adaptive shift lasts over time because some have reported that phonetic recalibration is fragile and dissipates within minutes (Stevens, 2007; van Linden & Vroomen, 2007; Vroomen et al., 2004), while others have found that recalibration is robust (Kraljic & Samuel, 2005) and can last for hours (Eisner & McQueen, 2006). Here, we further examined the robustness of phonetic recalibration by testing listeners immediately after exposure and then re-testing them after a 24-hour delay.

Lipread-induced recalibration has first been demonstrated by Bertelson, Vroomen, and de Gelder (2003). They exposed listeners to an ambiguous sound intermediate between /aba/ and /ada/ (A?) dubbed onto a face articulating either /aba/ or /ada/ (A?Vb or A?Vd). Participants shortly exposed to A?Vb tokens reported in a subsequent auditory-only speech identification test more /aba/ responses than when exposed to A?Vd. This was taken as a demonstration that the visual information had shifted the interpretation of the ambiguous auditory phoneme in its direction. The same study also showed that when a non-ambiguous sound was dubbed onto a congruent face (AbVb or AdVd), the proportion of responses consistent with the visual stimulus decreased. Participants exposed to AbVb thus reported fewer /aba/ responses than when exposed to AdVd. This was interpreted as a sign of selective speech adaptation (Eimas & Corbit, 1973) in which it is the repeated presentation of a non-ambiguous speech sound by itself (and thus in the absence of any conflict between auditory and visual information) that causes a reduction in the frequency with which that token is reported in subsequent categorization trials. Selective speech adaptation probably reveals fatigue of some of the relevant processes, most likely acoustic or phonetic in nature, although criterion setting may also play some role (Samuel, 1986).

In a follow-up study, it was explored how long recalibration and selective speech adaptation would last over time (Vroomen et al., 2004). Participants were again exposed to audiovisual exposure stimuli that contained either the non-ambiguous or ambiguous auditory speech tokens (AbVb, AdVd, A?Vd, or A?Vb). Immediately after exposure, participants then categorized 60 auditory-only ambiguous speech tokens as /aba/ or /ada/. This allowed us to trace aftereffects as a function of time of testing. Results showed that aftereffects induced by ambiguous versus non-ambiguous sounds dissipated at different rates: Whereas recalibration effects were transient and lasted only about six-to-twelve test trials, selective speech adaptation effects were robust and were present even after 60 test trials. This difference in dissipation rates provided further evidence that the two phenomena resulted from distinct underlying mechanisms. It also showed that the transient nature of recalibration was not due to some particularity of the test itself (like participants trying to equally distribute the two response alternatives) because aftereffects of ambiguous and non-ambiguous sounds were tested in the same way.

In contrast with the transient nature of lipread-induced recalibration, studies on lexical recalibration have reported much more stable effects over time. Norris, McQueen, and Cutler (2003) were the first to demonstrate lexically-induced recalibration. They spliced an ambiguous fricative intermediate between /f/ and /s/ onto Dutch words normally ending in /s/ (e.g., radijs; *radish*) or /f/ (e.g., witlof*; chicory*). Exposure to the ambiguous sound embedded in words normally ending in /s/ (a /s/-biasing context) resulted in more /s/ responses on subsequent categorization trials if compared to the /f/-biasing context, thus revealing recalibration (or, in the authors' words, "perceptual learning"). When the ambiguous speech sound was spliced onto pseudo-words, no boundary shift was observed indicating that the shift was caused by lexical information proper. Others have since demonstrated the same phenomenon. For example, Kraljic and Samuel (2005) exposed listeners to a speaker whose pronunciation of the sound /s/ or /ʃ/ was ambiguous (halfway between /s/ and /ʃ/). Following an exposure phase, participants were tested for recalibration either immediately after exposure, or after a 25-min silent intervening task. Aftereffects were actually numerically bigger after the delay, indicating that simply allowing time to pass did not cause learning to fade. Even longer-lasting aftereffects were reported by Eisner and McQueen (2006). They exposed listeners to a story in which they learned to interpret an ambiguous sound as /f/ or /s/. Results showed that perceptual adjustment measured after 12 hours was as robust as when measured immediately, and equivalent aftereffects were found when listeners heard speech from other talkers in the 12-hour interval or when they could sleep.

An obvious difference between studies that report robust versus fragile aftereffects is that the former used lexical information to induce recalibration, whereas fragile effects have been obtained with lipread speech. This difference in the way recalibration is induced, though, is unlikely to be relevant for the robustness of the effect because lipread effects tend, in general, to be bigger than lexical effects (e.g., Brancazio, 2004). This was confirmed in a study where lexical- and lipread-induced recalibration were compared directly with each other (van Linden & Vroomen, 2007). Listeners were exposed to an ambiguous sound halfway between /t/ and /p/ that was either dubbed onto a face articulating /t/ or /p/, or the sound was embedded in Dutch words normally ending in /t/ (e.g., 'groot', *big*) or /p/ (knoop, *button*). Following exposure to a lipread or lexical t- or p-bias, participants categorized auditory ambiguous tokens as /t/ or /p/. Results showed that the lipread-induced aftereffects tended to be somewhat bigger in size than the lexically-induced aftereffects, but both effects dissipated equally fast. It remains therefore unclear why some studies observed aftereffects to last for hours, while others reported fast dissipation.

One clue, though, may come from a procedural difference that has been demonstrated to play a role. Studies reporting robust aftereffects not only use the ambiguous sound that presumably drives recalibration (e.g., the ambiguous s/f-sound as embedded in the f-biasing context 'witlo/?/'; *witlo**f** = chicory*), but listeners are also exposed to the non-ambiguous sound from the opposite phoneme category (in this example the non-ambiguous /s/ as embedded in radij**s**; *radish*). It has been demonstrated that the presence of this contrast phoneme during the exposure phase increases the size of the aftereffect (van Linden & Vroomen, 2007). There are various reasons, besides the already mentioned selective speech adaptation, why this may occur: the non-ambiguous contrast stimulus might, for example, serve as an anchor, or it might provide a comparison model for another stimulus. Aftereffects thought to reflect recalibration could in this way be boosted because listeners set the criterion for the phoneme boundary in between the ambiguous token and the extreme one. Alternatively, participants may also adopt a tendency to judge anything that is not a clear /s/ as an /f/. Either way, the obtained aftereffect will then reflect the contribution of two distinct processes. One is related to recalibration proper (i.e., a shift in the phoneme boundary that is meant to reduce the conflict between the auditory and lexical information), while the other might be a strategic and longer lasting criterion setting operation that depends on the presence of two phonemes from opposing categories.

To explore this possibility, we addressed whether lipread-induced aftereffects become robust when contrast stimuli are presented during the exposure phase. Recalibration was induced by exposing participants to an ambiguous speech sound /?/ halfway between /t/ and /p/ that was combined with the non-ambiguous visual articulation of /t/ or /p/, (A?Vt for the t-biased group, and A?Vp for the p-biased group). In addition, non-ambiguous contrast stimuli from the opposing category were presented: ApVp for the t-biased group, and AtVt for the p-biased group. Following exposure to these stimuli, participants categorized ambiguous speech sounds from a /t/–/p/ continuum immediately after exposure and—to examine whether the effects were robust—were re-tested after a 24-hour delay.

2 Method

2.1 Participants

Twenty native speakers of Dutch (18–25 years old) with normal hearing and normal seeing participated. Half of them was biased towards /t/, the other towards /p/.

2.2 Materials

The same stimuli were used as in van Linden and Vroomen (2007). Stimulus creation started with a video and audio recording of a male native speaker of Dutch. An auditory ambiguous sound intermediate between /t/ and /p/, henceforth /?/, was created using the Praat speech editor (http://www.praat.org). The /?/ was created from a recording of /ot/ of which the second (F2) and third (F3) formant were varied so as to create a 10-step /ot/–/op/ continuum. The steady state-value of the F2 in the vowel was 950 Hz and 72 ms in duration. The transition of the F2 was 45 ms, and its offset frequency varied from 1123 Hz for the /t/-endpoint to 600 Hz for the /p/-endpoint in ten equal Mel steps. The F3 had a steady state value of 2400 Hz in the vowel, and the offset frequency of the transition varied from 2350 Hz for the /t/-endpoint to 2100 Hz for the /p/-end point in ten equal Mel steps. The silence before the final release of the stop consonant was increased in 6 ms steps from 22 ms for the /t/-endpoint to 82 ms for the /p/-endpoint. The waveforms of the aspiration part of the final release of /p/ and /t/ (134 ms) were mixed from natural /p/ and /t/ bursts in relative proportions to each other. The resulting continuum sounded natural with no audible clicks.

For the exposure stimuli, the ambiguous sound /?/ was spliced into recordings of four different pseudowords such as woo/?/ (‘woot’ and ‘woop’ are both non-words). The audio was then dubbed on the video of the face that articulated either ‘woop’ or ‘woot’. For the auditory test tokens, /?/ was spliced into the pseudoword soo/?/.

2.3 Procedure

Participants were tested individually in a soundproof booth. Half of the participants was biased towards /t/ and exposed to A?Vt and ApVp; the other half of the participants was biased towards /p/ and exposed to A?Vp and AtVt. Participants were seated at a distance of 60 cm in front of a 17-inch CRT-monitor on which the video fragments were presented. The audio samples were presented via two speakers (JBL Media 100WH/230) placed on the left and right of the monitor. The video fragments were 10 × 9.5 cm in size, and were shown against a black background. Loudness peaked at 70 dBa. A regular keyboard was used for data-acquisition. During the test, participants were instructed to press the p-key upon hearing 'soop' and t-key upon hearing 'soot'.

The whole experiment consisted of four phases: a calibration phase, a training phase, an exposure-test phase, and a second test phase after 24 hours.

2.3.1 Calibration

For each individual participant, it was determined which token was the most ambiguous one of the continuum. All test tokens were presented ten times in pseudo random order and participants were asked to indicate whether they heard 'soot' or 'soop'. The 50% crossover point was then determined via a logistic procedure. The token nearest this point served as the most ambiguous stimulus /?/ for subsequent testing.

2.3.2 Training

To acquaint participants with the test procedure, they categorized the most ambiguous /?/ token, and the two tokens nearest to this stimulus; the more 'p'-like token /?–1/ and the more 't'-like token /?+1/. Each of the three tokens was presented 20 times in pseudo-random order.

2.3.3 Exposure—test

Participants were presented five blocks of 16 exposure stimuli followed by 60 test trials. In the t-biased condition, each exposure block contained eight A?Vt stimuli and eight contrast stimuli ApVp in random order. In the p-biased condition, each exposure block contained eight A?Vp and eight AtVt stimuli. To ensure that participants were watching the monitor during the exposure phase, catch trials were included consisting of the short appearance (100 ms) of a small white dot on the upper lip of the speaker. Upon detecting a catch trial, participants pressed a special key.

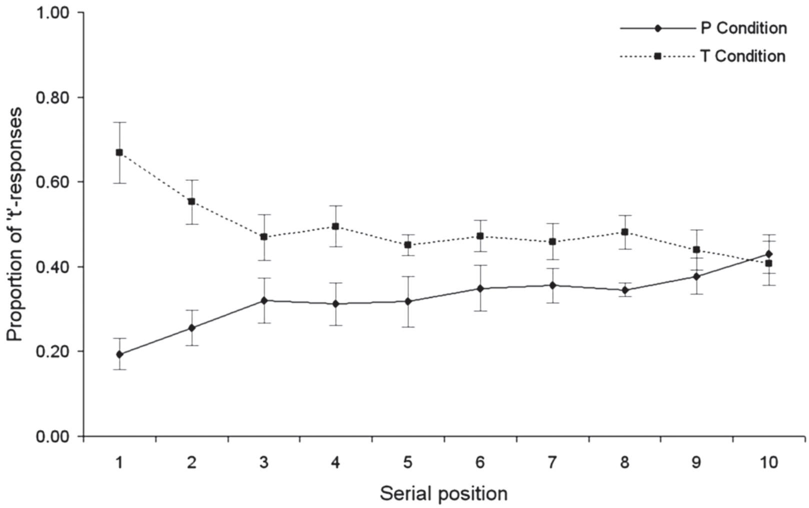
Each of the five exposure blocks was immediately followed by 60 auditory-only test trials. In the test phase, the three test tokens (soo/?–1/, soo/?/, and soo/?+1/) were presented 20 times in counterbalanced order. Participants were asked to indicate whether they heard 'soop' or 'soot' by pressing the 'p' or 't' key.

2.3.4 Re-test after a 24-hour delay

The second test was delivered 24 hours after exposure. Participants were presented five blocks of 60 auditory-only test stimuli. Stimuli and procedure were the same as

Figure 1

Proportion of /t/-responses as a function of the serial position in the immediate test. Error bars represent 1 standard error of the mean



in the immediate test, except that participants were not exposed anew to the exposure trials. Instead they tried to solve a Rubik's cube (a visual puzzle) during a one-minute interval between successive blocks.

3 Results

The most ambiguous stimulus ranged between tokens 3 and 7 of the ten synthesized test tokens. In the training phase, 50 % of the stimuli were judged as /t/ in the t-biased group and 51% in the p-biased group, indicating that the proportion of /t/- and /p/-responses before exposure was alike in both groups. During exposure, 99% of the catch trials were detected, indicating that participants kept their eyes fixed on the monitor.

To measure aftereffects and their dissipation, the 60 test trials were binned into 10 serial positions with each position representing the mean average number of /t/-responses of six consecutive test trials. The group-averaged proportion of /t/-responses is shown in Figure 1 (immediate test) and Figure 2 (delayed test). High values indicate more /t/- and thus less /p/-responses.

Aftereffects were calculated as in previous studies by subtracting the proportion of /t/-responses following /p/-bias from /t/-bias. Figure 3 shows the group-averaged difference in /t/- versus /p/-biased groups.

For the immediate test, a 2 (/t/- vs. /p/-bias) × 10 (test token position) ANOVA on the proportion of /t/-responses showed a significant main effect of exposure condition, $F(1, 18) = 12.14$, $p < .003$, because there were, as expected, more /t/-responses following /t/-bias than /p/-bias. This is the basic recalibration effect. The interaction with test token position was also significant, $F(9, 162) = 7.59$, $p < .001$, as the difference between the two groups dissipated with prolonged testing. Separate t-tests (Bonferroni

Figure 2

Proportion of /t/-responses as a function of the serial position in the delayed test. Error bars represent 1 standard error of the mean

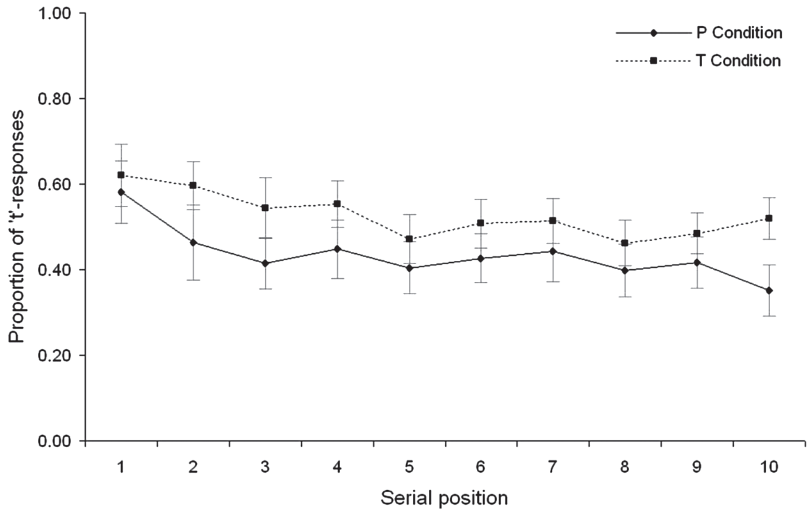
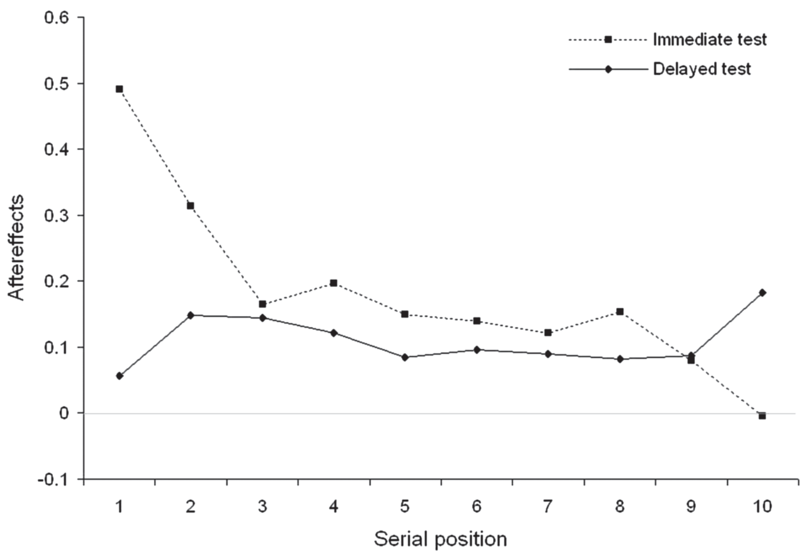


Figure 3

Aftereffects as a function of the serial position in the test



corrected for multiple comparisons) showed that aftereffects were bigger than zero up to test token position 2, thus representing the first 12 test trials. On test trials 1–6, the /t/-biased group gave a substantial 48% more /t/-responses than the /p/-biased group, and on test trials 7–12 this difference was still 30%. The same ANOVA on the data of the delayed test showed that after 24 hours, none of those effects was significant anymore. There was no overall difference between the /t/- and /p/-biased groups, $F(1, 18) = 1.72$, $p < .206$, and the interaction with test token position was also not significant, $F < 1$. Thus, despite substantial aftereffects in the immediate test, they dissipated fast and did not survive the delay.

Finally, we examined aftereffects per test block, because in delayed testing there might have been an aftereffect in the first test block only. In the 5 (block) × 2 (/t/- vs. /p/-bias) × 10 (test token position) ANOVA, there was no main effect of block in the immediate and delayed test, $F(4, 72) = 1.08$, $p < .372$ and $F(4, 72) < 1$, respectively, and block did not interact with any of the other factors, all p's > .15. Visual inspection confirmed that aftereffects were essentially the same across all five test blocks.

4 Discussion

Participants were biased to categorize an ambiguous speech sound as /t/ or /p/ by using two different kinds of exposure stimuli. On the one hand, we exposed participants to an auditory ambiguous sound combined with non-ambiguous lipread speech (A?Vt for the t-biased group and A?Vp for the p-biased group). Presumably, for these stimuli it is the lipread information that informs listeners on how to interpret the ambiguous sound. The phonetic conflict between the heard and lipread information thus induces a shift in the phoneme boundary that reduces the conflict (i.e., recalibration proper). This shift can in subsequent testing be observed as an aftereffect. Secondly, participants were also exposed to contrast stimuli containing a non-ambiguous sound from the opposing phoneme category (ApVp for the t-biased group and AtVt for the p-biased group). The presence of this contrast phoneme was expected to boost and possibly stabilize aftereffects as it might help in settling where the phoneme boundary should be. Exposure to both kinds of stimuli indeed resulted in a substantial aftereffect (i.e., a 46% difference), but only on the first test token positions. With prolonged testing, the aftereffect dissipated quickly and it did not reappear following a 24-hour delay. Apparently, the presence of contrast stimuli did thus not stabilize aftereffects.

This result raises the question what it is that drives phonetic representations to be re-adjusted back to normal that quickly. This is a particularly intriguing question if it is realized that others have observed (lexical) recalibration to be extremely robust against various unlearning conditions whereby listeners even heard "good" examples of previously adjusted tokens (Kraljic & Samuel, 2005). The simplest potential mechanism might be time itself whereby perceptual adjustments just "fade away." In the absence of any speech input, phonetic representations would then revert to their prior settings. Previously, though, we showed, with the same stimuli as used here, that this does not obtain for lipread-induced recalibration because aftereffects did not become smaller when a three-minute silent interval intervened between the exposure phase and the test (van Linden & Vroomen, 2007, Experiment 4). Lipread-induced recalibration is thus not fragile as such.

Another possibility is that the test procedure itself induces dissipation. The test involves a large number of trials in which ambiguous sounds are presented. It might be that listeners adjust their response criterion in the course of testing such that the two response alternatives are chosen about equally often. One would then expect the test to be most sensitive on the first few trials, while differences between conditions will become washed out with prolonged testing. We and others have indeed observed that aftereffects become smaller with prolonged testing (see, e.g., Kraljic & Samuel, 2006). However, against this interpretation of response equilibration is the finding that in one of our a previous studies, there was no dissipation of aftereffects caused by selective speech adaptation, despite the same test as here being used (Vroomen et al., 2004). Thus, after being exposed to, say, the non-ambiguous tokens AbVb, participants were less likely to report /b/ *during the whole test*. The test itself does thus not induce response equilibration. One notable difference, though, between selective speech adaptation and recalibration is that in the exposure phase of selective speech adaptation, no ambiguous speech sounds are presented whose phoneme boundary is shifted towards one or the other side of the continuum. For recalibration, it might thus be that the shift in the phoneme boundary by itself causes participants to be flexible during the test as well. Recalibration effects would, in this view, thus be fragile because participants are in a "shifting-mood." Further tests, though, are needed to examine this speculation more thoroughly and explore the conditions under which recalibration remains stable or returns to normal again.

Another potentially important factor affecting the stability of recalibration may be the acoustic or phonetic nature of the stimulus that is adapted. Previous studies either used fricatives (/f/–/s/, /s/–/ʃ/) or stop consonants (/p/–/t/, /b/–/d/, and /d/–/t/). Fricatives tend to produce large shifts that are long-lasting, and are primarily speaker- or token-specific (Eisner & McQueen, 2005; Kraljic & Samuel, 2005, 2007). Stop consonants, though, tend to produce smaller shifts that do not seem to last as long (Kraljic & Samuel, 2006; van Linden & Vroomen, 2007), but that generalize across speakers (Kraljic & Samuel, 2006, 2007). Further tests are required to explore whether the acoustic nature of the stimuli explains the difference in stability over time.

Another aspect that may play a role is the use of filler items. One of the classic learning principles is that massed-trials produce a weaker learning effect than spaced-trials. In our previous studies, we always presented massed-trials of adapters with either no filler items separating the critical items, or—as in the present study—only a few contrast stimuli (Bertelson et al., 2003; van Linden, Stekelenburg, Tuomainen, & Vroomen, 2007; van Linden & Vroomen, 2007, 2008; Vroomen et al., 2007; Vroomen et al., 2004). Others, though, reporting long-lasting effects used lots of filler items separating each of the critical items (Eisner & McQueen, 2006; Kraljic & Samuel, 2005, 2006; Norris et al., 2003). Typically, about 20 critical items containing the ambiguous phoneme were interspersed among 180 fillers items. At present, it remains to be explored whether recalibration of phonetic categories is sensitive to this variable and whether it follows the same classic learning principle.

To conclude, we found that aftereffects induced by lipread recalibration were fragile despite the presence of contrast phonemes during the exposure phase. The size of the aftereffect was most boosted by the presence of contrast stimuli, but these stimuli did not make the effect more robust. The robustness of lexical aftereffects reported by others (Eisner & McQueen, 2006; Kraljic & Samuel, 2005) are, most likely, therefore not caused by the presence of contrast stimuli as such.

References

BERTELSON, P., VROOMEN, J., & de GELDER, B. (2003). Visual recalibration of auditory speech identification: A McGurk aftereffect. *Psychological Science*, **14**, 592–597.

BRANCAZIO, L. (2004). Lexical influences in audiovisual speech perception. *Journal of Experimental Psychology-human Perception and Performance*, **30**, 445–463.

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**, 224–238.

EISNER, F., & McQUEEN, J. M. (2006). Perceptual learning in speech: Stability over time. *Journal of the Acoustical Society of America*, **119**, 1950–1953.

GANONG, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception & Performance*, **6**, 110–125.

KRALJIC, T., & SAMUEL, A. G. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology*, **51**, 141–178.

KRALJIC, T., & SAMUEL, A. G. (2006). Generalization in perceptual learning for speech. *Psychonomic Bulletin & Review*, **13**, 262–268.

KRALJIC, T., & SAMUEL, A. G. (2007). Perceptual adjustments to multiple speakers. *Journal of Memory and Language*, **56**, 1–15.

NORRIS, D., McQUEEN, J. M., & CUTLER, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, **47**, 204–238.

SAMUEL, A. G. (1986). Red herring detectors and speech perception: In defence of selective adaptation. *Cognitive Psychology*, **18**, 452–499.

STEVENS, M. (2007). *Perceptual adaptation to phonological differences between language varieties*. Unpublished Ph.D. thesis, University of Gent, Gent.

SUMBY, W. H., & POLLACK, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*, **26**, 212–215.

Van LINDEN, S., STEKELENBURG, J. J., TUOMAINEN, J., & VROOMEN, J. (2007). Lexical effects on auditory speech perception: An electrophysiological study. *Neuroscience Letters*, **420**, 49–52.

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.

Van LINDEN, S., & VROOMEN, J. (2008). Audiovisual speech recalibration in children. *Journal of Child Language*, **35**, 809–822

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**, 572–577.

VROOMEN, J., van LINDEN, S., KEETELS, M., de GELDER, W., & BERTELSON, P. (2004). Selective adaptation and recalibration of auditory speech by lipread information: dissipation. *Speech Communication*, **44**, 55–61.