Recalibration of phonetic categories by lipread speech versus lexical information

van Linden, S.; Vroomen, J.

Published in:
Journal of Experimental Psychology. Human perception and performance

Document version:
Publisher's PDF, also known as Version of record

Publication date:
2007

Link to publication

Citation for published version (APA):
Recalibration of Phonetic Categories by Lipread Speech Versus Lexical Information

Sabine van Linden and Jean Vroomen
Tilburg University

Listeners hearing an ambiguous phoneme flexibly adjust their phonetic categories in accordance with information telling what the phoneme should be (i.e., recalibration). Here the authors compared recalibration induced by lipread versus lexical information. Listeners were exposed to an ambiguous phoneme halfway between /t/ and /d/ dubbed onto a face articulating /t/ or /d/ or embedded in a Dutch word ending in /t/ (e.g., groot [big]) or /d/ (knoop [button]). In a posttest, participants categorized auditory tokens as /t/ or /d/. Lipread and lexical aftereffects were comparable in size (Experiment 1), dissipated about equally fast (Experiment 2), were enhanced by exposure to a contrast phoneme (Experiment 3), and were not affected by a 3-min silence interval (Experiment 4). Exposing participants to 1 instead of both phoneme categories did not make the phenomenon more robust (Experiment 5). Despite the difference in nature (bottom-up vs. top-down information), lipread and lexical information thus appear to serve a similar role in phonetic adjustments.

Keywords: recalibration, speech perception, perceptual learning, lipreading, lexical information

Frequently, people encounter speakers with unfamiliar accents who are difficult to understand. But, in natural speech, there are other information sources that can help listeners by telling them how to interpret speech sounds that initially might be ambiguous. Two potentially important information sources are visual information from the articulators of the face—here referred to as lipread speech—and lexical knowledge. As an example, imagine an unknown 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 might have noticed that the lips were closed during pronunciation of the ambiguous sound, which is typical for /b/ but not /d/. Lexical knowledge also informs 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). What is less known, though, is that next time listeners hear the same sound, they may 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, 2006; Kraljic & Samuel, 2005, 2006; Norris, McQueen, & Cutler, 2003). The occurrence of such an aftereffect is taken as an indication that listeners have adjusted the phonetic categories of their language so as to adapt to the new situation. Here we address whether the phonetic adjustment—or recalibration—differs when it is evoked by lipread information versus lexical knowledge. The following sections provide brief reviews of the evidence for lipread and lexical recalibration and a discussion of why the two might differ.

Recalibration of phonetic boundaries by lipread speech has first been demonstrated by Bertelson et al. (2003). They presented participants an ambiguous sound intermediate between /aba/ and /ada/ dubbed onto a face articulating either /aba/ or /ada/. After exposure to the auditory ambiguous speech sound combined with a face articulating /aba/, participants reported more /aba/ responses on subsequent auditory categorization trials if compared with exposure to a face articulating /ada/. The occurrence of such an aftereffect demonstrates the basic recalibration phenomenon. That is, listeners learned to categorize the ambiguous sound in accord with the lipread information it was previously combined with. A control experiment also demonstrated that when the sound was a nonambiguous /aba/ sound dubbed onto the congruent articulatory gestures, there were fewer /aba/ responses on subsequent categorization trials of the auditory ambiguous stimulus if compared with exposure to a nonambiguous /ada/, thus revealing selective speech adaptation (Eimas & Corbit, 1973). Selective speech adaptation reveals fatigue of some of the relevant processes (Eimas & Corbit, 1973; Samuel, 1986) and strongly depends on repeated exposure to nonambiguous speech sounds. Of note, subsequent tests showed that participants were unable to distinguish the ambiguous from nonambiguous exposure stimuli when asked to do so in a discrimination task, thus excluding the possibility that the results were caused by deliberate response strategies (Vroomen, van Linden, Keetels, de Gelder, & Bertelson, 2004).

Recalibration driven by lexical knowledge was first demonstrated by Norris et al. (2003). They spliced an ambiguous fricative intermediate between /l/ and /s/ onto Dutch words normally ending in /l/ (e.g., radijs [radish]) or /s/ (e.g., witlof [chicory]). Exposure to the ambiguous sound embedded in words normally ending in an /l/ (an /l/-biasing context) resulted in more /l/ responses on sub-
sequent categorization trials if compared with the /fl/-biasing context, thus revealing recalibration (or, in the authors’ words, “perceptual learning”). When the ambiguous speech sound was spliced onto pseudowords, the authors did not observe a boundary shift, indicating that the shift was caused by lexical information proper. Others have since demonstrated the same phenomenon. For example, Kraljac and Samuel (2005) exposed listeners to a speaker whose pronunciation of the sound /s/ was ambiguous (halfway between /s/ as the initial sound in cigarette and /s/ as in shop).

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 listeners learned to interpret an ambiguous sound as /fl/ or /fl/. Results showed that perceptual adjustment measured after 12 hr was as robust as measured immediately after learning. Equivalent effects were found when listeners heard speech from other talkers in the 12-hr interval or when they could sleep.

At first sight, it may seem that phonetic recalibration driven by lipread and lexical information is much alike. Both potentially rely on the same mechanism in the sense that the phonetic boundary between two speech categories is adjusted in accordance with disambiguating information that tells what the sound should be. Whether the information stems from lipread speech or lexical knowledge might be immaterial from this point of view. There are, however, potentially important differences between the two information sources that justify further exploration. One is concerned with the fact that lipreading, by its nature, is very different from lexical knowledge; the other is that studies on lipread recalibration have found more transient effects than those on lexical recalibration. Both issues are dealt with in the following sections.

In the literature on speech perception, there are two sets of theories that explain the roles of lipreading (Green, 1998; Massaro, 1987; Summerfield, 1987) and lexical knowledge (McClelland & Elman, 1986; Norris, McQueen, & Cutler, 2000) in speech perception. There is, however, very little cross-talk between the two, and both have largely been developed independently of one another. As concerns lipreading, numerous studies have shown that seeing a person speak has a profound effect on speech perception. One of the most striking demonstrations is the McGurk effect in which listeners report to hear /da/ when in fact they are presented with the auditory clear /ha/ combined with a face articulating /ga/ (McGurk & MacDonald, 1976). Although there is some debate about whether integration of the auditory and visual signal occurred early or late (Schwartz, Robert-Ribes, & Escudier, 1998; Vroomen, 1992), it seems clear that listeners integrate the two information sources at or before phonetic classification.

Whereas the use of so-called bottom-up lipread information in speech perception is undisputed, the status of top-down lexical effects remains much more debated. There are numerous studies that have shown that the lexical status of an utterance (is it a word or a pseudoword?) matters to speech perception. This has been demonstrated in paradigms like phoneme categorization (Ganong, 1980), phoneme monitoring (Culler, Mehler, Norris, & Segui, 1987), and phoneme restoration (Samuel, 1981). As an example, Ganong (1980) observed that an ambiguous phoneme between /d/ and /fl/ preceding ask tended to be categorized as /fl/ presumably because task but not dask is a word (if compared with ash; see also Pitt, 1995). It is not clear, though, whether this shift reflects a genuine perceptual phenomenon or a response bias, and some have taken the stance that lexical knowledge is actually not used in online speech processing proper (Norris et al., 2000).

Besides this potential difference between lipread speech and lexical knowledge in the online processing of speech sounds, there are other distinctions that might be crucial. Developmental studies have shown that there is a close link between lipreading and speech perception from very early on in life (Kuhl & Meltzoff, 1996), whereas for lexical information, it seems logical that it can only start to emerge when the lexicon starts to develop. Lipread and lexical information also differ in the time at which the information becomes available in online speech processing. Because of anticipatory articulation, lipread information can be available even before the speech signal is heard (Munhall & Tohkura, 1998), whereas lexical effects are typically slower and are usually only obtained after the word is recognized (Fox, 1984; Pitt, 1995). Lipreading can also result in stronger effects on speech perception than lexical information. For example, whereas lexical effects are typically found with ambiguous or degraded speech (Ganong, 1980; Warren, 1970), lipread information can alter the perception of auditory clear speech sounds as demonstrated in the McGurk effect (McGurk & MacDonald, 1976).

The previous examples might lead one to expect that lipreading will have more profound effects on speech perception than lexical information, as only lipreading provides perceptual and anticipatory information about the ongoing speech signal with a strong impact on the perceived sound. Such a pattern was indeed observed in a study by Brancacio (2004), who compared lipread and lexical effects on phoneme categorization. He observed that the effects of lipreading were strong and present throughout in slow, medium, and fast responses if compared with lexical effects that were smaller and only reliable in the medium and particularly slow responses. To the extent that these categorization effects relate to recalibration, one might also expect that lipread recalibration is stronger and more robust than is lexical recalibration.

Surprisingly, though, this does not seem to be the case. When one compares studies on lipread and lexical recalibration, it appears that lipread aftereffects are more fragile. For example, Vroomen et al. (2004) reported rapid dissipation of lipread aftereffects with prolonged testing. They presented 50 audiovisual exposure stimuli followed by 60 posttest trials and observed that aftereffects had dissipated after only 6 posttest trials. Contrary to the fast dissipation of lipread aftereffects, studies on lexical recalibration report stable effects over time. For example, after participants were exposed to only 20 ambiguous speech sounds, Kraljac and Samuel (2005) observed no decline in the magnitude of lexical aftereffects when posttests were presented either immediately or after 25 min. Lexical aftereffects were also resistant to various types of unlearning conditions in the 25-min interval, and they only diminished when participants heard the critical phoneme spoken unambiguously by the speaker of the exposure phase. Even larger intervals were used by Eisner and McQueen (2006), who found lexical aftereffects to remain stable over a period of even 12 hr subsequent to exposure to only a short story.

For the time being, it is unclear why lipread aftereffects seem to dissipate faster than lexical aftereffects because the various studies are difficult to compare. They not only differ in the use of lipread
versus lexical information but also in the nature of the phonemes (syllable-initial stops vs. word-final fricatives or stops) and various other experimental procedures. For example, studies on lexical recalibration typically presented during the exposure phase not only the ambiguous sound that presumably drives recalibration (e.g., witlof) but also the unambiguous sound from the opposite phoneme category (e.g., radijs). At present, it is unknown whether the presence of this sound boosts aftereffects (e.g., by enhancing the contrast). Furthermore, the short-lived lipread aftereffects reported by Vroomen et al. (2004) are, in principle, not mutually exclusive with the long-lasting lexical aftereffects reported by others (Eisner & McQueen, 2006; Kraljic & Samuel, 2005, 2006; Norris et al., 2003) because lipread aftereffects have hitherto not been analyzed as to how they survive a long silent period.

In the present study, we tried to resolve these issues by directly comparing aftereffects evoked by lipread and lexical information using the same paradigm and the same test stimuli. Experiment 1 served as a check that our newly created lipread and lexical exposure stimuli did indeed induce a bias effect in phoneme categorization and corresponding aftereffects. Experiment 2 explored dissipation of lipread and lexical aftereffects by measuring them as a function of the serial position in the posttest. Experiment 3 explored whether the presence of a contrast phoneme in the exposure phase enhanced aftereffects. In Experiment 4, a 3-min silent interval was introduced between the exposure phase and the posttest to check whether aftereffects dissipate if no testing intervenes, and Experiment 5 tested whether it mattered whether participants were exposed to one versus both phoneme categories.

### Experiment 1

Lipread and lexical exposure stimuli were created that biased an ambiguous sound halfway between /t/ and /p/ toward either /t/ or /p/. In the critical part of the experiment, participants were exposed to these stimuli (i.e., lipread or lexical /t/ or /p/ exposure) for a short time, immediately followed by a short posttest phase in which auditory tokens near the phoneme boundary were categorized as /t/ or /p/. These exposure posttest phases were presented several times during the experiment, with exposure toward /t/ or /p/ in random order. Experiment 1 served as a check that the exposure stimuli did indeed induce a bias effect on categorization and a corresponding aftereffect. The experiment also allowed us to compare the size of lipread and lexical aftereffects. It was expected that lipread speech would evoke bigger bias effects on categorization than would lexical information as it is known that lipreading can affect even auditory clear speech tokens, whereas lexical effects are typically only obtained when the auditory stimuli are ambiguous. No predictions could be made concerning the magnitude of lipread and lexical aftereffects because relevant data were not available.

### Method

**Participants.** Twenty-nine first-year students (mean age = 19 years 6 months), all native speakers of Dutch, participated in return for course credits.

**Materials.** An auditory ambiguous sound intermediate between /t/ and /p/, henceforth /?/, was created using the Praat speech editor (http://www.praat.org). For the effects of lipreading, the ambiguous sound /?/ was embedded in pseudowords like /wo?/ and dubbed on the video of a face that articulated either /wop/ or /wot/. For inducing lexical effects, we embedded /?/ in a context that made up a /t/ or /p/ word. As an example, when /?/ was embedded in /groo?/, it made up a /t/ word because /groot/ (big) but not /groop/ is a word in Dutch. Similarly, when /?/ was embedded in /knoo?/, it made up a /p/ word because /knoop/ (button) but not /knoop/ is a word in Dutch. Eight words and eight pseudowords were recorded: four ending in /ot/ and four ending in /op/. Words and pseudowords were matched on number of syllables (mono-, bi-, or trisyllabic), and they contained no other instances of /p/ or /t/. The average (logarithmic) frequency of occurrence (per million words) of the /t/ words was 1.066, and for /p/ words it was 0.764 (Baayen, Piepenbrock, & van Rijn, 1993). An overview of the exposure stimuli is provided in Table 1.

A male native speaker of Dutch was recorded on digital audio- and videotape (Philips DAT recorder and Sony PCR-PC2E mini DV). The /?/ was created from another recording of /ot/ of which the second and third formant were varied so as to create a 10-step /ot/-/op/ continuum. The steady state value of the second formant in the vowel was 950 Hz and 72 ms in duration. The transition of the second formant was 45 ms, and its offset frequency varied from 1123 Hz for the /t/-endpoint to 600 Hz for the /p/-endpoint in 10 equal Mel steps (a perceptual scale of pitches judged by listeners to be equal in distance from one another). The third formant 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/-endpoint in 10 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.

The lexical exposure stimuli were created by exciting naturally produced /op/ and /ot/ portions from words and replacing them with /?/. For the effects of lipreading, the lipread ambiguous sound /?/ was embedded in pseudowords like /wo?/. For inducing lexical effects, we embedded /?/ in a context that made up a /t/ or /p/ word. As an example, when /?/ was embedded in /groo?/, it made up a /t/ word because /groot/ (big) but not /groop/ is a word in Dutch. Similarly, when /?/ was embedded in /knoo?/, it made up a /p/ word because /knoop/ (button) but not /knoop/ is a word in Dutch. Eight words and eight pseudowords were recorded: four ending in /ot/ and four ending in /op/. Words and pseudowords were matched on number of syllables (mono-, bi-, or trisyllabic), and they contained no other instances of /p/ or /t/. The average (logarithmic) frequency of occurrence (per million words) of the /t/ words was 1.066, and for /p/ words it was 0.764 (Baayen, Piepenbrock, & van Rijn, 1993). An overview of the exposure stimuli is provided in Table 1.

### Table 1

<table>
<thead>
<tr>
<th>Auditory</th>
<th>Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lipread exposure stimuli</td>
<td></td>
</tr>
<tr>
<td>Foo?</td>
<td>Foop</td>
</tr>
<tr>
<td>Woo?</td>
<td>Woop</td>
</tr>
<tr>
<td>Kafoo?</td>
<td>Kaffoot</td>
</tr>
<tr>
<td>Dikasoo?</td>
<td>Dikasoot</td>
</tr>
<tr>
<td>Lexical exposure stimuli</td>
<td></td>
</tr>
<tr>
<td>Knoo?</td>
<td>Knoop (knob)</td>
</tr>
<tr>
<td>Hoo?</td>
<td>Hoop (hope)</td>
</tr>
<tr>
<td>Siroo?</td>
<td>Siroop (syrup)</td>
</tr>
<tr>
<td>Microscoop?</td>
<td>Microscope</td>
</tr>
<tr>
<td>Groot?</td>
<td>Groot (big)</td>
</tr>
<tr>
<td>Vloot?</td>
<td>Vloot (fleet)</td>
</tr>
<tr>
<td>Devoo?</td>
<td>Devoot (devout)</td>
</tr>
<tr>
<td>Idioo?</td>
<td>Idioot (idiot)</td>
</tr>
</tbody>
</table>
with the synthesized token /o?/. This resulted in t words like groo? and p words like knoo?. For the lipread exposure stimuli, pseudowords were used like woo? (i.e., neither woop nor woot is a word in Dutch) dubbed onto the video of the speaker pronouncing woop or woot. The video showed the speaker’s face up to his eyes so that the speaker’s mouth, mandible, and cheeks were visible. Videos were digitized at 352 × 288 pixels at 25 frames per second. All video fragments had a 10-frame (250-ms) fade in and fade out, with the natural synchronization between audio and video left intact. The posttest trials were made by replacing the naturally produced /ot/ from the pseudoword /sot/ with tokens from the /ot/-/op/ continuum so that listeners heard a continuum that varied from /sot/ to /sop/.

Procedure. Participants were tested individually in a soundproof booth. The videos were presented on a 17-in. (43.2-cm) monitor connected to a computer. The video filled about one third of the screen (10 × 9.5 cm) and was surrounded by a black background. The sound was presented via a Fostex 6301B (www.fostexinternational.com) speaker placed underneath the monitor. Loudness was 70 dBA when measured at ear level. Participants were seated at a distance of 60 cm in front of the screen.

Testing involved four phases: a calibration phase, a training phase, an exposure phase intermitted by posttest trials for testing visually or lexically induced recategorization, and a categorization phase to examine the goodness ratings of the exposure stimuli.

Calibration. The most ambiguous token of the /sot/-/sop/ continuum was, for each individual listener, determined in the calibration phase. All tokens of the /sot/-/sop/ continuum were presented 10 times in random order at a 1.5-s intertrial interval (ITI). Participants pressed a P or T on a keyboard upon hearing /sop/ or /sot/, respectively. The obtained s-shaped identification curve was then fitted with a logistic procedure, and the item nearest to the 50% crossover point served as the participant’s most ambiguous stimulus /?/ during subsequent exposure and testing.

Training. In the training phase, participants were acquainted with the posttest procedure. The three most ambiguous tokens—the boundary token /?/ and the two tokens nearest to the boundary token, /?/-1/ for the more p-like token and /?/+1/ for the more t-like token—were presented for identification. Each of the three tokens was presented 20 times with a 1.5-s ITI in pseudorandomized order. Responses were given as before.

Exposure posttest. The exposure posttest phase consisted of 2 sessions: 1 for testing lexical recategorization and 1 for lipread recategorization. Session order was counterbalanced over participants. Within each session, 10 exposure posttest blocks were presented: 5 blocks biasing toward /p/ and 5 blocks biasing toward /t/. The p or t blocks were presented in quasi-randomized order, with no more than 2 successive p or t blocks in a row. One exposure posttest block consisted of 8 exposure stimuli (500-ms ITI), immediately followed by 6 posttests trials (1.5-s ITI). The 8 exposure stimuli consisted of 2 presentations of each of the 4 different exposure stimuli that biased toward either /p/ or /t/ (presentation order of the exposure stimuli counterbalanced). The 6 identification posttests trials consisted of 2 triplets of the individually determined three most ambiguous tokens of the /sot/-/sop/ continuum (/?/-1/, /?/, and /?/+1/). Presentation order of the test tokens was counterbalanced so that each test token occurred equally often on each of the 6 serial positions of the posttest.

During exposure, we did not give participants a phonetic task, but to ensure that they were looking at the video during lipread exposure, we had them monitor the face for the occasional appearance of a small white dot (100 ms) on the upper lip of the speaker (catch trial). During lexical exposure, participants were viewing a white fixation cross against a black screen. On catch trials, the fixation cross changed into the small dot (100 ms). Participants pressed a special key upon detecting a catch trial. Each session contained eight catch trials.

Rating of the exposure stimuli. In the final part of the experiment, participants, rated on a 7-point Likert-scale, the /p/-/t/ quality of the /?/ as embedded in the lexical and lipread exposure stimuli. They were asked to circle 1 upon hearing a clear /t/, 7 upon hearing a clear /p/, and 4 when indecisive about the identity of the consonant. The stimuli were presented in two blocks (lipread and lexical), each block containing five repetitions of each of the eight exposure stimuli.

Results

The participants’ most ambiguous token ranged from 3 to 7 on the 10-point continuum. Participants detected 82% of the catch trials, indicating that they were looking at the screen during exposure.

Bias. We first examined whether the ambiguous phoneme /?/ was perceived as intended when embedded in the lipread and lexical exposure stimuli. The bias effect was calculated by taking the difference in the ratings of /?/ when embedded in t versus p stimuli. The average ratings were 2.15 and 6.15 for lipread t and p stimuli and 4.11 and 5.32 for lexical t and p stimuli, respectively. In a 2 (lipread vs. lexical information) × 2 (/p/ or /t/ stimulus) analysis of variance (ANOVA) on the ratings, there was an overall difference between lipread versus lexical stimuli, F(1, 28) = 8.17, p < .01, as the lipread /?/ was rated as more t-like than was the lexical /?/. More important, there was a main effect of /p/ versus /t/ stimuli, F(1, 28) = 175.37, p < .001, indicating that the lipread stimuli induced bigger bias effects (i.e., the 4.2 difference in the ratings of /?/ when embedded in t versus p stimuli) than did the lexical stimuli (a 1.21 difference). Separate t tests confirmed that lipread bias, t(28) = 13.61, p < .001, and lexical bias, t(28) = 5.09, p < .001, were both significantly bigger than zero.

Aftereffects. Posttest trials were likely to be labeled in accordance with the previously presented exposure stimuli (/p/ or /t/ exposure), thus showing that there was indeed lipread and lexical recategorization. To compute aftereffects, we calculated the mean percentage of T responses on the posttest trials for each exposure stimulus, pooling over the three different test stimuli (/?/-1/, /?/, and /?/+1/). Aftereffects were then calculated by subtracting the percentage of T responses following /p/ exposure stimuli from /t/ exposure stimuli. For lipread stimuli, the thus computed aftereffect was 20%; for lexical exposure stimuli, it was 9%. An ANOVA showed that aftereffects were, in general, bigger than zero, F(1, 28) = 37.53, p < .001, and that even though aftereffects induced by lipread exposure stimuli were numerically bigger, they were not significantly different from lexical aftereffects, F(1, 28) = 2.56, p = .121.
We also explored whether there was a relation between the size of the bias and the aftereffect. There was a general tendency that participants with large bias effects also had large aftereffects. For lipread stimuli, the correlation just failed to reach significance, \( r(n = 29) = .317, p = .094 \), whereas it was significant for lexical stimuli, \( r(n = 29) = .437, p < .02 \). When the size of the bias was entered as a covariate in the comparison of lipread versus lexical aftereffects, there was no sign that lipread and lexical aftereffects were different from each other \( (F < 1) \). The size of lipread and lexical aftereffects was thus comparable, in particular if the difference in bias was taken into account.

**Discussion**

Exposure to lipread and lexical stimuli resulted in aftereffects that were comparable in size. The aftereffects could be interpreted as the manifestation of recalibration because the ambiguous sound was identified in accord with previously seen (lipread) or heard (lexical) information. The information that induces the shift can thus be bottom-up lipread information or top-down lexical knowledge. As predicted, lipread information resulted in a stronger bias effect on phoneme categorization than did lexical information, but at this stage, there is no reason to maintain that there is a difference in the size of the aftereffects induced by the two information sources. In the following experiments, we explored whether lipread and lexical aftereffects would last equally long by measuring dissipation.

**Experiment 2**

Experiment 2 explored other potential differences between aftereffects induced by lipread and lexical information. One such difference is the rate at which the two effects dissipate. Previous studies have shown that lipread aftereffects dissipate quickly, whereas lexical aftereffects seem to last much longer (Eisner & McQueen, 2006; Kraljic & Samuel, 2005, 2006; Vroomen et al., 2004). Here, we compared the two directly by measuring dissipation of lipread and lexical aftereffects over the course of prolonged testing.

**Method**

Thirty new first-year psychology students (mean age = 18 years 7 months) participated in the experiment. All were native speakers of Dutch. Stimuli and procedures were as in Experiment 1, except that the number of posttest trials was increased from 6 (in Experiment 1) to 60, thus allowing the measurement of dissipation of lipread and lexical aftereffects. The 60 posttest trials consisted of 20 triplets of the participants’ three most ambiguous stimuli (/t/-1/0, /p/; and /r/-1/0) presented in counterbalanced order. Testing lasted approximately 2.5 hr, with regular pauses interspersed.

**Results**

The participants’ most ambiguous token varied from 3 to 7. On average, 96% of the catch trials were detected.

**Bias.** The average ratings of the /p/ phoneme when embedded in /t/ and /p/ exposure stimuli were 2.03 and 5.92 for the lipread stimuli and 4.49 and 5.36 for lexical exposure stimuli, respectively. In a 2 (lipread vs. lexical information) × 2 (/t/ or /p/ exposure stimulus) ANOVA, the effect of information type, \( F(1, 29) = 18.07, p < .001 \), the effect of exposure stimulus, \( F(1, 29) = 225.51, p < .001 \), and the interaction between the two, \( F(1, 29) = 81.02, p < .001 \), were significant. As in Experiment 1, both lipread and lexical stimuli induced bias effects, with lipread stimuli being more potent than lexical stimuli (3.89 vs. 0.88). Separate \( t \) tests confirmed that both effects were bigger than zero: \( t(29) = 17, 45, p < .001 \) for lipread stimuli, and \( t(29) = 3.67, p < .005 \) for lexical stimuli.

**Aftereffects.** In order to measure dissipation of lipread and lexical aftereffects, we first binned responses on the 60 posttest trials into 10 serial positions, with each position representing the average number of \( T \) responses on 6 consecutive posttest trials. Aftereffects were then calculated by subtracting the proportion of \( T \) responses following /p/ exposure from /t/ exposure for each of the 10 serial positions. Figure 1 shows the thus computed aftereffects, with positive numbers reflecting more responses consistent with the previous exposure stimuli (i.e., recalibration). As is clear from this figure, both lipread and lexical exposure stimuli induced positive aftereffects but only on the first serial positions of the test. Lipread aftereffects started out stronger and lasted somewhat longer (up to the third serial position or until Posttest Trials 13–18) than did lexical aftereffects (lasting up to the first serial position only or Posttest Trials 1–6), but both effects dissipated quickly. There was thus no sign that lexical recalibration would last longer than lipread recalibration.

A 2 (lipread vs. lexical information) × 10 (test token position) ANOVA on the aftereffects showed that, on average, aftereffects were bigger than zero, \( F(1, 29) = 7.23, p < .015 \), thus indicating that there were more \( T \) responses following /t/ exposure than /p/ exposure. There was no overall difference in the size of lipread and lexical aftereffects, \( F(1, 29) = 1.24, p < .274 \). The main effect of test token position was significant, indicating that aftereffects became smaller when more test trials were presented, \( F(9, 261) = 5.60, p < .001 \). The interaction between information type and test token position was significant, \( F(9, 261) = 4.61, p < .001 \). Separate \( t \) tests showed that lipread aftereffects were significantly bigger than zero up to Serial Position 3 (i.e., until Test Trials 13–18; all \( p \)s at least \( < .05 \)), whereas lexical aftereffects were significant only at Serial Position 1 (i.e., until Test Trials 1–6). Paired \( t \) tests also showed that lipread aftereffects were bigger than were lexical aftereffects on Serial Position 1, \( t(29) = 3.88, p = .001 \), and Serial Position 2, \( t(29) = 3.57, p = .001 \).

The correlation between the amount of bias in the categorization responses and the aftereffects (on the first serial position only) was not significant for lipread stimuli, \( r(n = 30) = .078, p = .68 \), or for lexical stimuli, \( r(n = 30) = .339 p = .067 \). The difference between lipread and lexical aftereffects on the first and second serial position was not significant anymore, \( F(1, 28) = 2.10, p = .158 \), when the difference in bias (3.88 vs. 0.87) was entered as a covariate in a 2 (lipread vs. lexical information) × 2 (test token position) analysis of covariance, indicating that the magnitude of the lipread and lexical aftereffects was comparable in size if the difference in bias was taken into account.

**Discussion**

Lipread and lexical exposure stimuli evoked aftereffects that dissipated quickly with prolonged testing. Contrary to expectations...
Based on previous studies (Eisner & McQueen, 2006; Kraljic & Samuel, 2005; Vroomen et al., 2004), there was no sign that lexically induced aftereffects lasted longer than did lipread aftereffects. If anything, lipread aftereffects tended to last somewhat longer, a result that was well accounted for by the fact that lipread stimuli also exerted a stronger bias effect than did lexical stimuli. These results suggest that there is no fundamental difference in the duration of lipread and lexical aftereffects. However, they also leave unexplained why others reported lexical aftereffects to last much longer than the ones observed here (Eisner & McQueen, 2006; Kraljic & Samuel, 2005). In the following experiment, we therefore explored whether the presence of contrast stimuli in the exposure phase, as frequently used by others, enhances aftereffects.

Experiment 3

A potentially relevant difference between studies exploring lipread and lexical aftereffects concerns the use of contrast stimuli. Studies reporting long-lasting lexical aftereffects presented during the exposure not only words with ambiguous sounds but also filler words with nonambiguous sounds taken from the opposite side of the phoneme continuum (Eisner & McQueen, 2006; Kraljic & Samuel, 2005). For example, in the exposure phase of Norris et al. (2003) in which an ambiguous s/f sound was biased toward /s/, there were not only exposure stimuli like radij? which supposedly drives recalibration (i.e., the lexical information teaches that the s/f sound is /s/ rather than /f/) but also contrast stimuli containing the nonambiguous sound /f/ (e.g., witlof). The presence of these contrast phonemes might, possibly, cause selective speech adaptation. For example, the /f/ in witlof may cause a fatigue of /f/ detectors such that the s/f sound is perceived as /s/: radij? via recalibration and witlof via selective speech adaptation. For these reasons, Norris et al. (2003) included a control condition in which it was actually checked—and discarded—that selective speech adaptation was at stake. However, because there is no inherent reason why contrast stimuli should be present in the exposure phase, it might be more appropriate to exclude them from the exposure phase right away, rather than controlling for them. Moreover, the recalibration stimuli might interact with the contrast stimuli because the two phonemes together create a contrast that can trigger criterion-setting operations resulting in long-range aftereffects. Possibly, then, previous studies using contrast stimuli have overestimated the contribution of lexical recalibration. In Experiment 3, we tested this possibility by repeating Experiment 2 but included contrast stimuli in the exposure phase. If contrast stimuli do indeed enhance aftereffects, then lipread and lexical aftereffects might become bigger or more stable in time.

Method

Twenty-four new first-year psychology students (mean age = 20 years 1 month) participated. Stimuli and procedures were as in Experiment 2, except that nonambiguous contrast stimuli were included in the exposure phase. A single exposure block contained 16 stimuli: 8 stimuli with the ambiguous phoneme that biased /?/ toward /?/ toward either /?/ or /p/ and 8 contrast stimuli that contained clear tokens of the nonambiguous contrast phoneme. Participants might thus hear the lexical exposure stimulus know? (biasing /?/ toward /p/) and the contrast stimulus groot (which contains the nonambiguous sound /l/) in a single exposure block. The order of the exposure stimuli was quasi-randomized, with no more than three exposure stimuli or contrast stimuli in a row.
Results and Discussion

The participants’ most ambiguous token varied from 3 to 8. Participants detected on average 94% of the catch trials, indicating they were attending the screen.

Bias. Both lipread and lexical exposure stimuli induced a bias effect. The average ratings of lipread /t/ and /p/ stimuli were 1.96 and 5.86 points, respectively, and for lexical stimuli, the ratings were 3.69 and 4.93 points, respectively. In a 2 (lipread vs. lexical stimulus) × 2 (p or t stimulus) ANOVA, there was no overall difference between lipread and lexical stimuli, F(1, 23) = 3.00, p = .097. The effect of t versus p stimulus was significant, F(1, 23) = 182.46, p < .001, as was the interaction with information type, F(1, 23) = 32.41, p < .001. As before, /t/ was rated as more t-like when embedded in t stimuli than p stimuli, and lipread information exerted a stronger bias than did lexical information. Separate t tests showed that the 3.89 bias effect of lipread stimuli was significant, t(23) = 12.93, p < .001, as was the 1.24 effect of lexical stimuli, t(23) = 4.13, p < .001.

Aftereffects. Figure 2 displays aftereffects induced by lipread and lexical stimuli in Experiment 3, together with those of Experiment 2. As is clear from the figure, contrast stimuli indeed enhanced the magnitude of the aftereffects, but aftereffects still dissipated quickly. A 2 (lipread vs. lexical information) × 10 (test token position) ANOVA on the aftereffects of Experiment 3 showed that aftereffects were significantly bigger than zero, F(1, 23) = 27.99, p < .001. There was no overall difference between lipread and lexical aftereffects (F < 1), and the effect of serial position of the test token was significant, F(9, 207) = 34.15, p < .001, as aftereffects became smaller when more test tokens were presented. The interaction between information type and test token position was also significant, F(9, 207) = 3.52, p < .001. Separate t tests showed that lipread aftereffects were bigger than zero up to Test Token Position 3 (Posttest Trials 1–18), whereas lexical aftereffects were significant up to Test Token Position 2 (Posttest Trials 1–12; all ps < .001). There was no relation between the magnitude of the bias effect and the aftereffect on the first test token position for lipread, r(n = 24) = .016, p = .941, or lexical, r(n = 24) = .247, p = .245, stimuli.

To analyze whether the contrast stimuli boosted aftereffects, we compared aftereffects of Experiments 2 and 3 in a 2 (Experiment 2 vs. 3) × 2 (lipread vs. lexical information) × 10 (test token position) ANOVA. The aftereffects of Experiment 3 were, in general, bigger than were those of Experiment 2, F(1, 52) = 5.25, p < .03, demonstrating that contrast stimuli indeed enhanced the magnitude of the aftereffects. The effect of test token position was significant, F(9, 468) = 35.18, p < .001, as aftereffects in both experiments dissipated with prolonged posttesting. The interaction between test token position and experiment, F(9, 468) = 8.54, p < .001, was also significant indicating that aftereffects of Experiment 3 were bigger than were those of Experiment 2 on the first test token positions only. Separate t tests confirmed that lipread aftereffects were bigger in Experiment 3 than in Experiment 2 on Test Token Positions 1–3, whereas lexical aftereffects were bigger on Test Token Positions 1 and 2 (all ps < .01). Contrast stimuli with auditory nonambiguous sounds thus enhanced the magnitude of lipread and lexical aftereffects but only at the beginning of the test. Of note, there was no sign that aftereffects would also become more stable when contrast stimuli were included. This issue about the stability of the aftereffects was further explored in Experiment 4.

Figure 2. Experiments 2 and 3: Aftereffects as a function of the serial position in the posttest. Following exposure to an auditory ambiguous speech token combined with lipread or lexical information, the proportion of responses consistent with the lipread or lexical information increased (i.e., recalibration) on the first test token positions of subsequently delivered posttest trials. The presence of a contrast phoneme during the exposure phase (Experiment 3) increased this effect, but aftereffects did not become more stable in time (if compared with Experiment 2). Error bars represent plus or minus one standard error of the mean.
Experiment 4

It may seem that the short-lived aftereffects reported in Experiments 2 and 3 are in contradiction with long-lasting lexical aftereffects reported by others (Eisner & McQueen, 2006; Kraljic & Samuel, 2005). In studies reporting lexical aftereffects, there is typically an interval between the exposure phase and the beginning of the posttest that may vary from approximately 1 min (Norris et al., 2003) to 25 min (Kraljic & Samuel, 2005) up to 12 hr (Eisner & McQueen, 2006). This raises the question why aftereffects in the present study dissipate within the first minute of testing. Presumably, aftereffects dissipate because the phoneme boundary is re-adjusted back to normal. This readjustment, though, may occur either because time simply passes or because listeners change their criterion while being tested. For example, one possibility is that the response criterion is adjusted in due course of testing such that the two response alternatives are chosen about equally often. There might then be no real difference between the short-lived aftereffects reported here and the stable aftereffects reported by others, because both can remain stable until the posttest phase starts.

To explore whether aftereffects indeed remain stable as long as no new test tokens are encountered, we introduced a 3-min silent interval between the end of the exposure phase and the beginning of the posttest phase. If the mere passing of time causes dissipation, one would expect aftereffects as observed in Experiments 2 and 3 to have dissipated completely following a 3-min interval. Alternatively, a silent interval might not harm aftereffects if dissipation occurs as a consequence of criterion shifts during the posttest.

Method

Twenty-nine new students (mean age = 18 years 7 months) participated in Experiment 4. Stimuli and procedures were as in Experiment 2, except that a 3-min silent interval was introduced between the end of the exposure phase and the beginning of the posttest trials. Participants tried to solve a Rubik’s cube during this interval. A short tone, 10 s before the first posttest trial, warned participants for the upcoming test phase.

Results and Discussion

The participants’ most ambiguous token varied from 3 to 6. Participants detected on average 89% of the catch trials.

Bias. Both lipread and lexical stimuli induced bias effects. The average ratings of /l/ were 2.77 and 6.09 when embedded in lipread t and p stimuli and were 4.82 and 5.56 for lexical t and p stimuli, respectively. A 2 (lipread vs. lexical information) × 2 (p or t stimulus) ANOVA showed that /l/ was rated as more t-like with lipread information than with lexical information, F(1, 28) = 9.63, p < .005. The effect of t versus p stimulus was significant, F(1, 28) = 98.26, p < .001, as was the interaction F(1, 28) = 48.90, p < .001, indicating that lipread stimuli induced bigger bias effects than did lexical stimuli. Separate t tests confirmed that the 3.32 bias effect of lipread stimuli was significant, t(28) = 12.11, p < .001, as was the 0.74 bias effects of lexical stimuli, t(28) = 2.67, p < .015.

Aftereffects. Aftereffects were calculated as in Experiment 2 by subtracting the proportion of T responses following /p/ exposure for each of the 10 serial positions. Aftereffects are shown in Figure 3, together with those of Experiment 2. Of note, aftereffects survived the 3-min silent interval, indicating that the mere passing of time did not make aftereffects disappear.

A 2 (lipread vs. lexical information) × 10 (test token position) ANOVA on the aftereffects of Experiment 4 confirmed that aftereffects were significantly above zero, F(1, 28) = 4.98, p < .05. There was no overall difference between lipread and lexical exposure stimuli (F < 1), and the effect of test token position was again significant, F(9, 252) = 2.76, p < .005, as aftereffects decreased with prolonged testing. The interaction between information type and test token position was not significant, F(9, 252) = 1.18, p = .306. Despite the fact that the interaction was not significant, we conducted separate t tests so that a comparison could be made with the analysis of Experiment 2. For the lipread stimuli, aftereffects were bigger than zero on Serial Positions 1–4 (Posttest Trials 1–24), whereas for the lexical stimuli, aftereffects were bigger than zero only on Serial Position 1 (Posttest Trials 1–6; all ps < .05).

To check whether the 3-min silent interval affected aftereffects, we also compared the results of Experiment 4 with those of Experiment 2 in a 2 (Experiment 2 vs. 4) × 2 (lipread vs. lexical information) × 10 (test token position) ANOVA, with experiment as a between-subjects variable. Of note, the main effect of experiment and its interaction with information type and serial position were all nonsignificant (all Fs < 1). The effect of test token position was again significant, F(9, 513) = 7.85, p < .001, as aftereffects decreased when more test tokens were presented. Inspection of Figure 3 shows that there was also a tendency that lipread aftereffects following the 3-min interval were smaller on the first serial position if compared with the no-interval condition, but the second-order interaction among experiment, information type, and serial position was not significant, F(9, 513) = 1.72, p = .081. There was no relation between the size of the bias effect and the aftereffect on the first serial position (Posttest Trials 1–6) for lipread, r(n = 29) = –.001, p = .997, and lexical, r(n = 29) = .21, p = .275, stimuli. However, when the same correlations were computed across the four experiments, it appeared that for the lexical stimuli, participants with big bias effects also displayed big aftereffects, r(n = 112) = .262, p < .01, but there was no such relation for lipread stimuli, r(n = 112) = .068, p = .48.

The results of Experiment 4 thus essentially showed that a 3-min silent interval between the exposure phase and the posttest did not make aftereffects disappear. Lipread and lexical aftereffects remained stable until the posttest phase started and only then did they disappear quickly. In the final experiment, we further explored reasons for this quick dissipation. Here we addressed whether aftereffects disappear because participants in previous experiments were exposed to both p and t biasing contexts, rather than just a single context.

Experiment 5

Aftereffects as measured in Experiments 1–4 relied on the effect of the most recently encountered exposure stimuli. These exposure stimuli varied from block to block, so that participants were biased toward both phoneme categories (i.e., /p/ and /t/). This procedure allowed us to measure aftereffects in a within-subjects design. Studies on lexical aftereffects, though, have used a between-
subjects design and have exposed participants to a single context only (Eisner & McQueen, 2006; Kraljic & Samuel, 2005). Possibly, this difference affects the robustness of the phenomenon because in the latter case, phoneme boundaries are not continuously adjusted to one side of the continuum or the other (but see Vroomen et al., 2004). To check whether this difference in procedure indeed matters, we exposed participants of Experiment 5 to only a single context (i.e., lipread or lexical /p/ or /t/ exposure).

**Method**

Sixty new students (mean age = 19 years 3 months) were randomly assigned to one of four groups (15 participants per group). Each group was exposed to only one out of four possible combinations of lipread or lexical /p/ or /t/ exposure stimuli. Participants were presented five exposure posttest blocks.

**Results**

The participants’ most ambiguous token varied from 3 to 7. On average, 98% of the catch trials were detected.

Unlike in the previous experiments, aftereffects could not be computed on an individual basis because the lipread and lexical /t/ versus /p/ exposure conditions were between-subjects variables. Figure 4 therefore shows the average proportion of T responses on the posttests as a function of the serial position for each of the four groups separately. As is clear from this figure, in the initial phase of the posttest (up to Serial Position 2 or Posttest Trials 1–12), there were more T responses for the /t/-exposed groups (lipread and lexical) than for the /p/-exposed groups, but this difference disappeared with prolonged testing.

This generalization was supported by a 2 (lipread vs. lexical information) × 2 (/p/ vs. /t/ exposure) × 10 (test token position) ANOVA on the proportion of T responses, with information type and exposure phoneme as between-subjects variables. In the ANOVA, there was a significant interaction between exposure phoneme and test token position, $F(9, 504) = 8.26, p < .001$, indicating that the /t/-exposed groups had more T responses than did the /p/-exposed groups in the beginning of the test. There was no overall difference in number of T responses between the lipread and lexical exposure groups, nor were any of the other interactions significant (all $F$s < 1). Separate $t$ tests confirmed that the lipread /t/-exposed group had more T responses on Test Token Positions 1 and 2 (Trials 1–12) than did the /p/-exposed group, whereas for the lexical /t/-exposed groups, this was the case for Serial Position 1 (all $ps < .05$) and marginally so for Test Token Position 2, $t(28) = 1.66, p = .10$.

To examine whether aftereffects were different for participants exposed to one versus both phoneme categories (i.e., Experiments 2 vs. 5), we also computed a group averaged aftereffect for Experiment 5 by subtracting, per serial position and information type, the proportion of T responses of the /p/-exposed groups from the /t/-exposed groups. This group-averaged aftereffect can then be compared with the individually determined aftereffects of Experiment 2 (see Figure 5). As is clear from Figure 5, lipread aftereffects of Experiments 2 and 5 were very much alike: Both were significant in the beginning of the test and dissipated with pro-

![Figure 3. Experiments 2 and 4: Aftereffects as a function of the serial position in the posttest. Following exposure to an auditory ambiguous speech token combined with lipread or lexical information, the proportion of responses consistent with the lipread or lexical information increased (i.e., recalibration) on the first test token positions of subsequently delivered posttest trials. A 3-min pause between the exposure phase and the posttests (Experiment 4) did not affect the magnitude or the rate of dissipation of the aftereffects (if compared with Experiment 2). Error bars represent plus or minus one standard error of the mean.](image)
The lexical aftereffects of Experiment 5 were bigger than those of Experiment 2 on the first test token positions but were smaller on later positions. There was thus no sign that aftereffects were, in general, bigger or would last longer if participants were exposed to one instead of both phoneme categories.

**General Discussion**

The present study shows that when listeners hear an ambiguous phoneme, they flexibly adjust the phonetic categories of their language in accordance with disambiguating information that tells what the phoneme should be (i.e., recalibration). The disambiguating information can be lipread or lexical, and both information sources induce similar-sized aftereffects (Experiment 1). Results also show that lipread and lexical aftereffects dissipate about equally fast in due course of testing (Experiment 2). The presence of a contrast phoneme during the exposure phase enhances the size of the aftereffects (Experiment 3), but aftereffects do not become more stable in time. Of note, though, aftereffects do not become smaller when a 3-min silent interval intervenes between the exposure phase and the posttest (Experiment 4). This indicates that recalibration as such is not fragile but that other factors possibly related to test procedure itself may explain why aftereffects dissipate quickly during testing. One such factor we further explored was whether participants were biased in consecutive exposure phases toward one or both phoneme categories. The results showed that this factor was not critical because aftereffects of participants exposed to one or both phoneme categories were comparable in size and duration (Experiment 5). Taken together, the results show that aftereffects induced by lipread and lexical information are very much alike. From a functional perspective, there was no difference between bottom-up perceptual information or top-down stored lexical knowledge: Both information sources were used in the same way to adjust the boundary between two phonetic categories.

It remains to be explained why others have observed long-lasting lexical aftereffects (Eisner & McQueen, 2006; Kraljic & Samuel, 2005), whereas here we found that aftereffects dissipate fast, even if no other information is encountered that tells what the ambiguous phoneme should be. One possible explanation for the fast dissipation is that listeners change their criterion while being tested. For example, it may be that listeners adjust their response criterion in due course of testing such that the two response alternatives are chosen about equally often. If so, there might be no real difference between the short-lived aftereffects reported here and the stable aftereffects reported by others, because both remain stable until the posttest phase starts. Another prediction is that if others would measure aftereffects as a function of the serial position in the posttest, they might observe that aftereffects become smaller with prolonged testing (see, e.g., Kraljic & Samuel, 2006). Alternatively, though, it may also be that the rate of dissipation depends on the acoustic nature of the stimuli. In the present case, syllable-final stop consonants were used that varied in place of articulation (/p/-/t/), whereas others used fricatives (/f-s/ and /s-S/; Eisner & McQueen, 2006; Kraljic & Samuel, 2005) or syllable-initial voiced–voiceless stop consonants (/d-t/; Kraljic & Samuel, 2006). If the stability of the phenomenon depends on the acoustic nature of the cues (e.g., a more ambiguous cue is more likely to shift than is a less ambiguous cue), one may observe that aftereffects differ in this respect as well.
Another robust finding was that lipread bias on phoneme categorization was bigger than lexical bias. This result is in line with previous findings showing that lipreading exerts a stronger effect on the percept of a heard sound than does lexical information (Brancazio, 2004). What is surprising is that from the point of view where information is defined as a reduction in uncertainty, there is actually little reason why lipread information should have a stronger impact than should lexical information, because lipread information is in fact not less constraining than lexical information. As an example, for lexical exposure stimuli like *groo?*, the word context specifies a single interpretation of */t?/*, because only *groot* and not *groop* is a word in Dutch. On the other hand, for lipread information, one could argue that the final consonant in the pseudoword *woot* is not very distinctive because */t/* is difficult to lipread (Montgomery, Walden, & Prosek, 1987). This then raises the question why lipread information, if not more constraining than lexical information, has nevertheless a stronger impact on phoneme categorization? The answer to this question probably relates to the distinction made earlier between bottom-up versus top-down information. Lexical influences will always be relatively weak and will only operate when the input signal is weak or ambiguous. In contrast, lipread information has, from the very beginning, more weight because it is an inherent input property of the speech signal. From this point of view, then, there is a quantitative distinction between bottom-up and top-down information because lipreading outweighs lexical information in relative contribution to the immediate phonetic percept.

From a theoretical point of view, it is of interest that there was a correlation between the amount of lexical bias in phoneme categorization and the size of the lexically induced aftereffect. Listeners who displayed the biggest lexical bias also showed the biggest lexical aftereffects. This finding speaks to the hotly debated issue about whether lexical information can actually influence how people hear speech sounds. Despite the fact that numerous studies have shown that lexical information can bias phoneme categorization, an important limitation is that these studies rely on subjective reports of what listeners hear. For this reason, some have argued (Norris et al., 2000) that bias effects are not informative about speech processing proper, because they reflect decision-level influences rather than true perceptual effects.

One way to address this concern is to look for the consequences of contextual effects in a situation in which listeners do not make a decision about the speech signal itself. The paradigm used here in which aftereffects were measured is an example in case. To explain the existence of (lexically induced) aftereffects, the same theorists have argued that there is a distinction between offline lexical feedback used for learning—which is supposedly real and of benefit to speech perception—and online lexical feedback as embodied in an interactive model of spoken word recognition—which is supposedly not real and even harmful (Norris et al., 2000). However, if this strict distinction between online and offline feedback is valid, there is actually little reason why the two measures of these phenomena (i.e., the bias in phoneme categorization and aftereffects) should be related, given that they reflect different domains. Because our results, though, show that there is a correlation between lexical bias and aftereffects, it may be more useful to integrate the two measures into a more coherent framework.

Figure 5. Experiments 2 and 5: Aftereffects as a function of the serial position in the posttest. Following exposure to an auditory ambiguous speech token combined with lipread or lexical information, the proportion of responses consistent with the lipread or lexical information increased (i.e., recalibration) on the first test token positions of subsequently delivered posttest trials. Aftereffects were not different for participants exposed to one versus both phoneme categories. Error bars represent plus or minus one standard error of the mean.
One plausible mechanism might be that listeners flexibly adjust their phoneme boundary whenever an ambiguous sound is heard in a disambiguating context. This adjustment occurs fast and instantly (Vroomen, van Linden, de Gelder, & Bertelson, 2007), and it may for this reason show up as an immediate bias when one is asked to rate or identify the phoneme, whereas upon later testing, the adjustment is still manifest to be observable as an aftereffect. Bias in phoneme categorization and aftereffects are, on this view, thus caused by the same mechanism and they reflect a (retrospective) change in the criterion of the phoneme boundary. A testable prediction that follows from this notion is that there will be no recalibration if the exposure stimuli do not evoke a bias as well.

References


