LANGUAGE AND COGNITIVE PROCESSES, 2001, *16* (5/6), 661–672

Lipreading and the compensation for coarticulation mechanism

Jean Vroomen
Tilburg University, Tilburg, The Netherlands

Beatrice de Gelder
Tilburg University, Tilburg, The Netherlands, and Université Libre de Bruxelles, Brussels, Belgium

Listeners compensate for coarticulatory influences of one speech sound on another. We examined whether lipread information penetrates this perceptual compensation mechanism. Experiment 1 replicated the finding that when an /as/ or /aʃ/ sound preceded a /ta/-/ka/ continuum, more velar stops were perceived in the context of /as/. Experiments 2 and 3 investigated whether the same phoneme boundary shift would be obtained when the context was lipread instead of heard. An ambiguous sound between /as/ and /aʃ/ was dubbed on the video of a speaker articulating /as/ or /aʃ/. Subjects relied on the lipread information when identifying the ambiguous fricative sound as /s/ or /ʃ/, but there was no corresponding boundary shift in the following /ta/-/ka/ continuum. These results indicate that biasing of the fricative by lipread information and compensation for coarticulation can be dissociated.

A tacit assumption in psycholinguistic research has been that tasks involving the explicit identification and/or discrimination of phonemes (like phoneme monitoring or phoneme categorisation) tap much of the same processes as those involved in the implicit phonetic processing of natural speech. However, the idea that explicit phoneme categorisation directly reflects implicit phonetic processes may be too simplistic. In the present study, we indeed present results showing that explicit phoneme

---

Requests for reprints should be addressed to Jean Vroomen, Tilburg University, Dept. of Psychology, P.O. Box 90153, 5000 Le Tilburg, The Netherlands. E-mail: j.vroomen@kub.nl
The experiments were carried out as part of the PhD thesis of Vroomen (1992).

http://www.tandf.co.uk/journals/pp/01690965.html DOI: 10.1080/01690960143000092

identification on the one hand, and so-called "compensation for coarticulation"—a process presumably at stake in the phonetic processing of natural speech—on the other, can dissociate.

Previously, it has been shown that ambiguous stops between /t/ and /k/ tend to be heard as /k/ after /s/ and as /t/ after /ʃ/ (Mann & Repp, 1981). The explanation given for this finding is that the perceptual system compensates for the coarticulation of the fricatives and stops which occurred in speech production. During production of a fricative-stop pair, the place of articulation of the stop is supposed to shift towards that of the fricative. A /k/ in the context of /s/ is thus more anterior than in a neutral context. The perceptual system compensates for this by adjusting the category boundary such that an "anterior" /k/ will nevetheless be heard as a velar /k/ when preceded by /s/. This effect has been replicated with an ambiguous fricative (henceforth /?/) midway between /s/ and /ʃ/ that was embedded in a lexical context biasing it towards /s/ or /ʃ/ (Elman & McClelland, 1988). More /k/ responses were obtained after christma? than after fooli?, as if listeners had heard /s/ in christma? and /ʃ/ in fooli?. Ever since, this finding has been taken as one of the prime examples of how high-order lexical information can affect phoneme perception, that is, of a proper "top-down" effect.

More recently, though, this picture has become more complicated. Pitt and McQueen (1998) argued that Elman and McClelland (1988) had confounded lexical context with transitional phoneme probabilities (TPs). Pitt and McQueen reasoned that an /s/ is more likely to follow /ə/ (as in christmas), and /ʃ/ is more likely to follow /I/ (as in foolish), and they argued that the results of Elman and McClelland should be attributed to this difference in TPs rather than to lexical biases. When Pitt and McQueen controlled words for TPs (e.g., as in juice and bush), they only found biases in the labelling of the fricative (listeners were lexically biased in that they tended to report bush when hearing bu? and juice when hearing jui?), but no subsequent shift in a /t/-/k/ continuum that immediately followed these words. In contrast, nonwords with TPs that biased the fricative either toward /s/ (as in /mi?/) or /ʃ/ (as in /neɪ?/) had an effect on labelling of both fricative and stop. Pitt and McQueen concluded that TPs rather than lexical effects were responsible for triggering the compensation for coarticulation mechanism observed by Elman and McClelland. They also argued that the dissociation between lexical bias on the fricative (bu? and jui?) with no accompanying shift in the /t/-/k/ continuum is a problem for TRACE (McClelland & Elman, 1986). This model has a single phoneme level that serves two functions, namely: (1) as output for explicit decisions about phonemes, and (2) as an intermediate processing stage between features and words. In such an architecture, biases in fricative identification and compensation for coarticulation should affect the same processing level, and so the effects should come and go hand-in-hand, without dissociation. In contrast, models like Merge (Norris, McQueen, & Cutler, 2000) are able to handle this dissociation because in Merge there is a distinction between phoneme decision nodes, which are lexically biased, and phonetic input nodes which are immune to lexical biases.

However, the TP-interpretation of the compensation for coarticulation data of Merge has also not been undisputed. Samuel (this issue) argues that the short monosyllabic words used by Pitt and McQueen (1998) may not have been potent enough to trigger a lexical top-down effect on compensation for coarticulation. He reports a lexical effect with long trisyllabic words while controlling TPs (i.e., bigram frequency). This finding is also in line with Elman and McClelland (1988; Experiment 3) who obtained a compensation for coarticulation effect with words in which differences in vowels (and hence TPs as defined by Pitt and McQueen) were explicitly controlled. A lexical effect on compensation for coarticulation may thus emerge with long, but not with short words.

One aim of the present study was a further attempt to dissociate biases on fricative identification from biases on the identification of the following stops. This dissociation has become important in the literature because it serves as evidence favouring Merge over TRACE (Norris et al., 2000). It seemed therefore important to show that the dissociation could also be found with materials other than short words controlled for TPs. We avoided the difficulty of using different TPs by combining the same auditory stimulus with lipread information. Fowler, Brown, and Mann (2000) conducted a study very similar to the present one. They used tokens of an auditory /da/-/ga/ continuum preceded by /al/ or /ar/. As originally reported by Mann (1980), the authors found that listeners identified ambiguous /da/-/ga/ syllables more often as /ga/ following /al/ than following /ar/. The same effect was found when instead of clear auditory /l/ or /r/, an ambiguous sound between /l/ and /r/ was dubbed on the video of a speaker saying /l/ or /r/ (a variant of the well-known 'McGurk effect'; McGurk & MacDonald, 1976). They observed that lipread information affected identification of the auditory ambiguous liquid, and more important, had an effect on identification of the following /da-ga/ continuum. Lipread information thus triggered compensation for coarticulation.

In the present study, we used the same methodology as in Fowler et al. (2000), but with fricative-stop tokens (an /a?/ as context token that preceded a /ta/-/ka/ continuum while dubbed on the video of a speaker saying /as/ or /aʃ/). The idea was that identification of the auditory ambiguous /?/ would be biased towards /s/ or /ʃ/ depending on the lipread information with which it was combined. The question was whether this bias in the fricative would trigger compensation for coarticulation so that an ambiguous stop between /t/ and /k/ would be heard as /k/ after lipread /s/ and as /t/ after lipread /ʃ/. If such an effect were obtained, it would show, as in Fowler et al., that lipread information penetrates early processes in speech perception. This would be in line with other evidence showing that lipread information can affect pre-phonetic processing stages (e.g., Green & Miller, 1985), and that, in general, there are cross-modal interactions at early stages of auditory and visual information processing (e.g., Vroomen & de Gelder, 2000a). Alternatively, if lipread information biases identification of the fricative, but has no effect on compensation for coarticulation, it would be evidence in favour of the distinction made in Merge (Norris et al., 2000) between implicit phonetic and explicit phonemic processes.

EXPERIMENT 1

To ensure that there were no stimulus confounds, we first wanted to replicate the finding that when an /as/ or /aʃ/ sound precedes a /ta/-/ka/ continuum, more velar stops are perceived in the context of /as/ if compared to /aʃ/.

Method

Subjects. Ten native speakers of Dutch took part in the experiment. None of them reported any problems with their hearing or sight.

Stimuli. A nine-step /ta/-/ka/ continuum was created that was preceded by /as/or /aʃ/. Natural tokens of /ta/, /ka/, /as/, /aʃ/, /asta/, /aska/, /aʃta / and / aʃka / were produced by a male native speaker of Dutch and recorded in a sound-damped booth. The speaker was also recorded on Sony U-Matic video for the audio-visual experiments. The auditory stimuli were low-pass filtered at 9.8 kHz, and then digitised at 20 kHz with a 12 bit analog-to-digital converter. The nine-step /ta/-/ka/ continuum was created by adding in proportion the amplitudes of the waveforms of the first 55 ms of a /t/ and /k/. The proportion of the /k/ increased from .2 to 1.0 in nine steps of .1 such that the /t/ endpoint contained .2 /k/ and .8 /t/, while the /k/ endpoint contained 1.0 /k/ and 0. /t/. The vocalic portion of the /ka/ sound was appended at the zero crossing of the so created stimuli.

From naturally produced /as/ and /aʃ/ tokens, the /s/ and /ʃ/ were cut out with the cut being made at a zero crossing. Then, the /ʃ/ was made to be of equal length as the /s/ by cutting out the final 64 ms, making each stimulus last 175 ms. The /s/ and /ʃ/ were appended at a zero crossing to the vocalic part of /as/ to make /as/ and /aʃ/ tokens. Finally, the /as/ and /aʃ/ tokens were prepended to the /ta/-/ka/ continuum with a 150 ms silence interval between them. The stimuli sounded natural and there were no discontinuities in the waveform audible as clicks. The stimuli were presented over two loudspeakers (Philips 420 car speakers with the peak amplitude set at 61 dB-A) in front of the subjects.

Procedure. The stimuli were presented in two blocks of 90 stimuli each. Within each block, context sound (/as/ and /aʃ/) and token of the /ta/-/ka/ continuum was randomised. Each block was made up of five sequences of the eighteen possible combinations of context sound and token of the /ta/-/ka/ continuum.

Each subject participated in two sessions of 7 min each. Subjects pressed one out of four buttons to indicate whether they had heard /asta/, aska/, /aʃta/, or /aʃka/. The next stimulus was presented 2 s after a button press. As a warm-up, each block was preceded by nine different tokens.

Results and discussion

Figure 1 shows the identification functions of the /asta/-/aska/ and /aʃta/-/aʃka/ continua. The effect was exactly as predicted: an /as/ context increased the percentage of /ka/ responses compared with the /aʃ/ context.

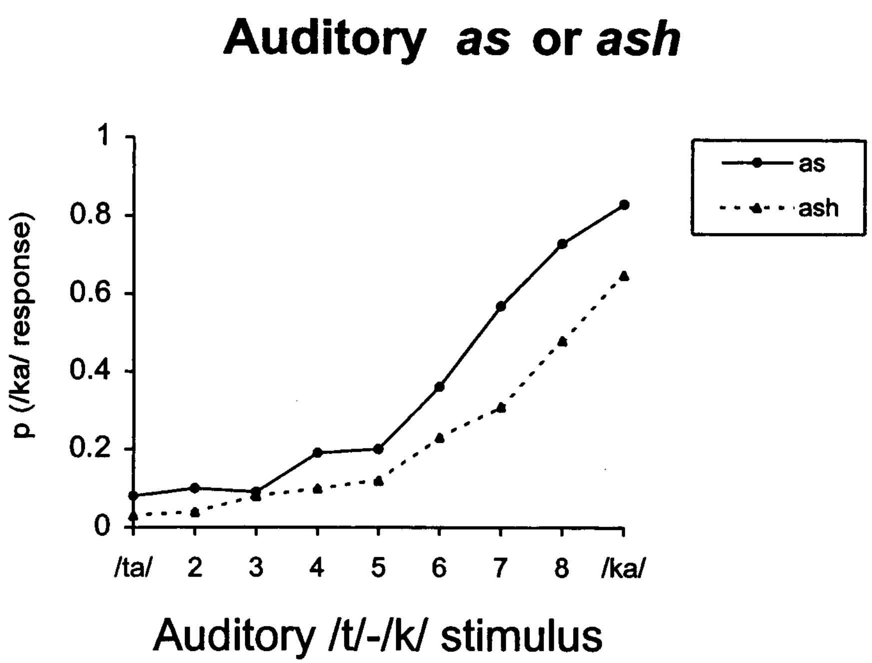

Figure 1. The effect of an auditory /s/ or /ʃ/ sound on the perceived place of articulation of a following stop consonant.

The proportion of /ka/ responses was submitted to a 2 (context) × 9 (auditory levels) analysis of variance (ANOVA). As expected, identification functions rose from left to right as the target stimuli varied along the /ta/-/ka/ continuum, $F(8, 72) = 31.55$, $p < .001$. The effect of context was significant because more /ka/ responses were given when /as/ was the context than when /aʃ/ was the context, $F(1, 9) = 15.07$, $p < .001$. The interaction between context and the auditory information was significant indicating that the context effect was largest for the most ambiguous /ta/-/ka/ tokens, $F(8, 72) = 2.25$, $p < .05$. As can be seen in Figure 1, the most ambiguous part was near the /k/ endpoint of the continuum. This was mainly caused by three subjects who perceived less than 67% /k/ (i.e., 66%, 50%, and 50%) on even the most /k/-like stimulus (the /as/ context combined with the /ka/ endpoint). Individual analyses, however, confirmed that all subjects had the shift in the predicted direction. When a /ʃ/ was heard, the proportion of /ʃ/ responses (i.e., /aʃta/ plus /aʃka/) was .939 while it was .227 when an /s/ was heard, $F(1, 9) = 10.95$, $p < .009$. Thus, as expected, the heard context phoneme contributed to the identification of the fricative (a shift of 71.2%) and it shifted identification of the following stop. Experiment 1 thus replicated the basic compensation for articulation effect and shows that there were no confounds with the stimuli created for this series of experiments.

EXPERIMENT 2

Experiment 2 tested whether lipread information for the fricative can influence labelling of the fricative and whether, in addition, it affects labelling of the stop continuum.

Method

Subjects. Ten new subjects drawn from the same population were paid for their participation.

Stimuli. The nine stimuli from the /ta/-/ka/ continuum were now preceded by an /a?/. These sound tracks were then dubbed on the video of the speaker articulating either /aska/ or /aʃka/. The ambiguous /?/ was made in a similar way to the /ta/-/ka/ continuum, namely by summing the waveforms of a /s/ and /ʃ/. The tokens were spliced out from a naturally produced /as/ and /aʃ/ and they were made of equal length. By taking a proportion of .8 /s/ and .2 /ʃ/ a /?/ sound was made which was ambiguous between /s/ and /ʃ/. The rest of the /a?ta/-/a?ka/ tokens were exactly the same as in Experiment 1. The tokens were dubbed onto the video recording of the speaker saying /aska/ or /aʃka/. During the pronunciation of /s/ in /aska/, the speaker's lips were retracted, and during the pronunciation of /aʃ/ in /aʃka/, the lips were rounded. The pronunciation of the /k/ phoneme in the video recording was visually indistinguishable from /t/ since both phonemes belong to the same viseme cluster (i.e., a group of phonemes that are visually indistinguishable from each other; Walden, Prosek, Montgomery, Scherr, and Jones, 1977). Lipreading thus distinguished between /s/ and /ʃ/, but not between /t/ and /k/. The audio track was synchronised with the video at the onset of the initial vocalic portion of the stimuli. This resulted in natural looking stimulus events in which no desynchronisation of the audio track and video could be detected.

Procedure. The stimuli were presented in 10 lists of random sequences of the 18 possible stimuli (nine auditory levels and two visual articulations). There was a 4 s pause between the successive stimulus events and a 10 s pause between successive sequences of 18 trials. Prior to testing, subjects were given 10 practice trials. Subjects were tested individually in a sound-damped booth. They viewed a 63-cm television monitor that presented both the auditory and visual dimensions of the speech stimuli. Subjects were seated at a distance of about 2 m from the monitor. The audio channel was set at a comfortable listening level with the peak amplitude at approximately 61 dB-A. Testing lasted about 20 min. Subjects were asked to respond orally whether they heard /asta/, /aska/, /aʃta/, or /aʃka/. Responses were written down by the experimenter.

Results and discussion

Figure 2 displays the proportion of /ka/ responses (i.e., the proportion of /aska/ plus /aʃka/ responses) as a function of the auditory levels, separately for the visual /as/ or /aʃ/ context. An ANOVA performed on the proportion of /ka/ responses indicated that the identification functions rose from left to right as the target stimuli varied from /ta/ to /ka/, $F(8, 72) = 169.05$, $p < .001$. The visual pronunciation of /as/ or /aʃ/ had a marginally significant effect on the identification of the /ta/-/ka/ stimuli, $F(1, 9) = 4.45$, $p = .064$. The interaction between the visual and auditory information of the stimuli was not significant, $F(8, 72) < 1$. However, although there was a small effect of the lipread context, it was in the opposite direction of what was expected. Instead of /as/, it was /aʃ/ that increased the proportion of /ka/ responses from .463 for the /as/ context to .483 for /aʃ/.

Performing a logit transformation on the data showed that the mean phoneme boundary in the /as/ context was 5.31 stimulus units while it was 5.15 in the /aʃ/ context, $F(1, 9) = 4.09$, $p = .074$. There was thus no phoneme-boundary shift towards the /ta/-end of the continuum when a visual /as/ pronunciation was seen. Individual analysis confirmed that only two out of ten subjects had a shift in the /ta/ direction.

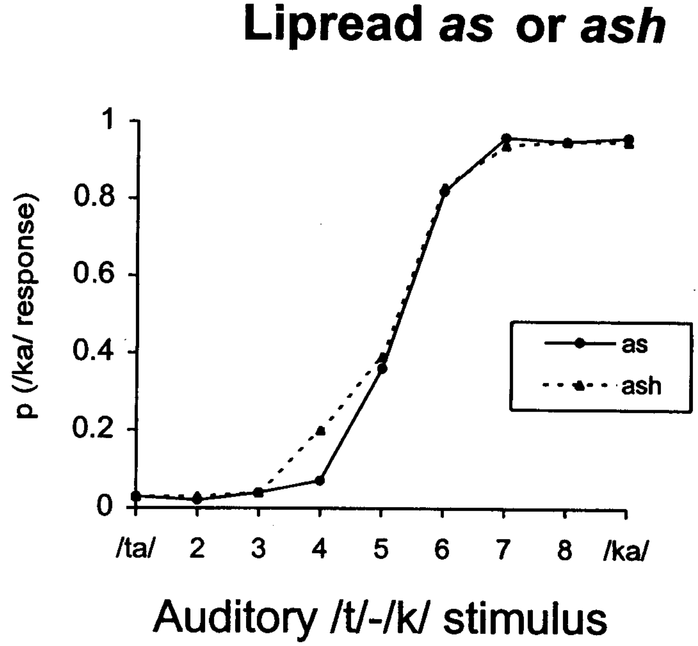

Figure 2. The effect of a lipread /s/ or /ʃ/ on the perceived place of articulation of a following stop consonant.

An explanation for the failure of a visual /as/ to increase the proportion of /ka/ responses could be that the lipread information did not affect the identification of the fricative /s/ or /ʃ/. Since we asked our subjects to report both context and target stimulus, we were able to examine this. When an /s/ was seen, the proportion of /s/ responses (i.e., /asta/ plus /aska/) was .510 while it was .292 when a /ʃ/ was seen, $F(1, 9) = 8.14$, $p < .02$. Thus lipreading did indeed contribute to the identification of the fricative (a shift of 22%) , but not the stop. One might, however, argue that the effect of lipreading was too small for an effect to be observed on the following /ta/-/ka/ continuum, since only 61% of the context was correctly identified. We therefore tried to amplify the impact of lipreading in a second analysis by discarding all responses that were not in agreement with the lipread information. In the case a visual /s/ was seen, all responses with /ʃ/ were rejected (i.e., /aʃta/ and /aʃka /) and similarly, when a /ʃ/ was seen, all responses with /s/ (i.e., /asta/ and /aska/) were rejected. Thus, the lipread context was in this analysis always correctly identified. An ANOVA performed on these corrected measures indicated that the auditory variable was significant, $F(8, 56) = 99.28$, $p < .001$, but there was again neither an effect of lipreading, $F(1, 7) < 1$, NS, nor was the interaction significant, $F(8, 56) < 1$, NS. (Two subjects had to be discarded from this analysis because, on some steps on the continuum, the appropriate proportions could not be computed unless a division was made by zero.) The mean phoneme boundary was 5.31 stimulus units in the /as/ context, while it was 4.95 in the /aʃ/ context, $F(1, 7) = 1.53$, $p = .256$. Thus, the perceived fricative did not exert any influence on the phoneme boundary of the following /ta/-/ka/ continuum, even though in this analysis all subjects reported consistently /as/ or /aʃ/ depending on the context.

We were, however, still not satisfied with these results because in the latter analysis about 39% of the data had to be discarded. Moreover, it might be the case that coarticulatory effects would have been found if the /ta/-/ka/ tokens were more ambiguous since compensation effects are usually largest for the most ambiguous stimuli. In the final experiment, we therefore decreased the volume of the audio. It was hoped that this would boost the contribution of vision in the identification of the context and that the auditory information of the /ta/-/ka/ tokens would become more ambiguous.

EXPERIMENT 3

The same stimuli were used as in the previous experiment, except that the volume of the audio was lowered so as to increase the relative contribution of the lipread information.

Method

Subjects. Seven new subjects, drawn from the same subject pool as before, were paid for their participation.

Stimuli and procedure. The same as in Experiment 2, except that the volume of the audio was set at a level in which it became hard to distinguish /t/ from /k/ (the peak of the amplitude was at 51 dB-A).

Results and discussion

Figure 3 displays the proportion of /ka/ responses (i.e., the proportion of /aska/ plus /aʃka/ responses) as a function of the auditory levels, separately for the visual /as/ or /aʃ/ context. We first investigated whether subjects relied on vision in the identification of the context sound. The proportion of /s/ responses when /s/ was seen was .84, whereas it was .21 when a /ʃ/ was seen, $F(1, 6) = 39.23$, $p < .001$. Thus, lipreading contributed to the identification of the fricative (a shift of 63%) and the visual influence was boosted when compared with Experiment 2 (a 22% shift). Moreover, as

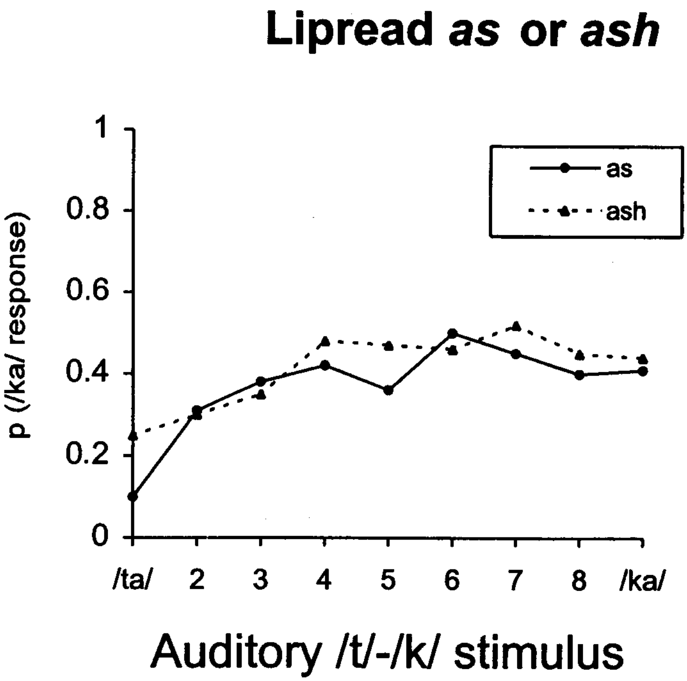

Figure 3. The effect of a lipread /s/ or /ʃ/ on the perceived place of articulation of a following stop consonant with the audio level at a low amplitude.

can be seen in Figure 3, identification of the /ta/-/ka/ tokens was much more difficult since the usual steep categorical identification functions were not obtained. However, there was again no effect of the visual information on the /ta/-/ka/ continuum. Lipreading an /s/ thus did not elevate the proportion of /ka/ responses.

An ANOVA performed on the proportion of /ka/ responses confirmed that the identification functions rose from left to right as the target stimuli varied from /ta/ to /ka/, $F(8, 48) = 3.16, p < .006$. The visual pronunciation of /as/ or /aʃ/ had a small and non-significant effect on the identification of the /ta/-/ka/ stimuli, $F(1, 6) = 4.12, p = .09$, but the effect was again in the opposite direction of what was found when /as/ or /aʃ/ were heard. When an /s/ was lipread, the proportion of /k/ responses was .39 whereas it was .44 when a /ʃ/ was lipread. (We report here only the responses in which the context was correctly identified. Including responses in which the context was incorrectly lipread did not change the results.) Thus, although we succeeded in increasing the contribution of vision and decreasing the quality of the /t/-/k/ tokens, there was still no effect of compensation for coarticulation.

GENERAL DISCUSSION

In the present study we showed that in a fricative-stop pair, lipread information biased identification of the fricative, but had no effect on compensation for coarticulation of the following stop. We thus observed a dissociation similar to the one reported by Pitt and McQueen (1998). They found that short words (bu? and jui?) biased identification of the fricative without consequences on labelling of the following stop. This pattern of results is accounted for by the Merge model (Norris et al., 2000) in which there is distinction between phonetic processes (biased by TPs) and phonemic decisions (biased by lexical information). On this account, lipread information would, like lexical information, affect phonemic decisions, but not phonetic processes.

This interpretation, though, is difficult to reconcile with Fowler et al. (2000) who found an effect of lipread information on compensation for coarticulation in a liquid-stop pair. It remains to be investigated whether the difference between our study and their study stems from the different phonemes that were used (fricative-stop versus liquid-stop). One possibility is that subjects may find it more difficult to distinguish lipread /s/ from /ʃ/ than distinguishing lipread /r/ from /l/. Fowler et al. obtained a 77% shift in identification of the liquid, whereas we obtained a 22% (Expt. 2) and 63% (Expt. 3) shift of the fricative. Although our lipread information was possibly weaker than in Fowler et al., our shifts are comparable with Pitt and McQueen (1998) who obtained TP-induced shifts of 17% and 37%, but who nevertheless obtained an effect on compensation for coarticulation. So at first sight, it seems that our fricative bias was within the range in which effects on compensation for coarticulation can be observed.

Despite these conflicting data about the stage at which lipreading affects speech processing, it is nevertheless clear that compensation for coarticulation can dissociate from labelling of the context phoneme. It shows that variables that affect explicit categorisation of phonemes may not affect processes involved in the phonetic processing of natural speech. There are other studies that have pointed out a similar contrast between processes involved in explicit phoneme identification and those involved in the processing of natural speech. From the neuropsychological literature it is known that the ability to identify CV syllables is a poor predictor of auditory comprehension deficits. On this basis, Hickok and Poeppel (2000) speculate that there is, akin to vision, a ventral cortical pathway for word recognition and a dorsal pathway for sublexical discrimination and identification. Norris et al. (2000) also make a strict distinction between phoneme units and decision units (for comments, see Vroomen & de Gelder, 2000b). Moreover, it is well-known that illiterates, Chinese, and dyslexics have problems with tasks requiring manipulation of speech segments at a subsyllabic level, while at the same time they have no apparent deficit in spoken word recognition (Bertelson, de Gelder, Tfouni, & Morais, 1989; de Gelder, Vroomen, & Bertelson, 1993). This all suggests that speech perception is less transparent than psycholinguistic tasks assume.

REFERENCES

Bertelson, P., de Gelder, B., Tfouni, L., & Morais, J. (1989). The metaphonological abilities of adult illiterates: new evidence of heterogeneity. *European Journal of Cognitive Psychology*, *1*, 239–250.

de Gelder, B., Vroomen, J., & Bertelson, P. (1993). Effects of alphabetic reading competence on language representation in bilingual Chinese subjects. *Psychological Research*, *55*, 315–321.

Elman, J.L., & McClelland, J.L. (1988). Cognitive penetration of the mechanisms of perception: Compensation for coarticulation of lexically restored phonemes. *Journal of Memory and Language*, *27*, 143–165.

Fowler, C.A., Brown, J.M., & Mann, V.A. (2000). Contrast effects do not underlie effects of preceding liquids on stop-consonant identification by humans. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 877–888.

Green, K.P., & Miller, J.L. (1985). On the role of visual rate information in phonetic perception. *Perception and Psychophysics*, *38*, 269–276.

Hickok, G., & Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Sciences*, *4*, 131–138.

Mann, V.A. (1980). Influence of preceding liquid on stop-consonant perception. *Perception & Psychophysics*, *28*, 407–412.

Mann, V.A., & Repp, B.H. (1981). Influence of preceding fricative on stop consonant perception. *Journal of the Acoustical Society of America*, *69*, 548–558.

McClelland, J.L., & Elman, J.L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*, 1–86.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*, 746–748.

Norris, D., McQueen, J.M., & Cutler, A. (2000). Merging information in speech recognition: feedback is never necessary. *Behavioral and Brain Sciences*, *23*, 299–370.

Pitt, M.A., & McQueen, J.M. (1998). Is compensation for coarticulation mediated by the lexicon? *Journal of Memory and Language*, *39*, 347–370.

Samuel, A. (this issue). Some empirical tests of Merge's architecture. *Language and Cognitive Processes*, *16*, 709–714.

Vroomen, J., & de Gelder, B. (2000a). Sound enhances visual perception: Cross-modal effects of auditory organization on vision. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 1583–1590 .

Vroomen, J., & de Gelder, B. (2000b). Why not model spoken word recognition instead of phoneme monitoring? *Brain and Behavioral Sciences*, *23*, 349–350.

Vroomen, J. (1992). *Hearing voices and seeing lips: Investigations in the psychology of lipreading.* Unpublished doctoral dissertation, Tilburg University, Tilburg.

Walden, B.E., Prosek, R.A., Montgomery, A.A., Scherr, C.K., & Jones, C.J. (1977). Effects of training on the visual recognition of consonants. *Journal of Speech and Hearing Research*, *20*, 130–145.