Perception & Psychophysics
2006, 68 (7), 1227-1240

On the causes of compensation for coarticulation: Evidence for phonological mediation

HOLGER MITTERER
Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

This study examined whether compensation for coarticulation in fricative–vowel syllables is phonologically mediated or a consequence of auditory processes. Smits (2001a) had shown that compensation occurs for anticipatory lip rounding in a fricative caused by a following rounded vowel in Dutch. In a first experiment, the possibility that compensation is due to general auditory processing was investigated using nonspeech sounds. These did not cause context effects akin to compensation for coarticulation, although nonspeech sounds influenced speech sound identification in an integrative fashion. In a second experiment, a possible phonological basis for compensation for coarticulation was assessed by using audiovisual speech. Visual displays, which induced the perception of a rounded vowel, also influenced compensation for anticipatory lip rounding in the fricative. These results indicate that compensation for anticipatory lip rounding in fricative–vowel syllables is phonologically mediated. This result is discussed in the light of other compensation-for-coarticulation findings and general theories of speech perception.

Hockett (1955) once described the problem of speech recognition with the metaphor of recognizing colored eggs on a conveyor belt after they have been crushed by a wringer. After the wringer, the different colors run into each other, and it is difficult to say where a given color starts and ends. This metaphor correctly indicates that speech sounds are neither separable—that is, there is no point in time at which the speech signal is influenced by only one phoneme—nor invariant, because the acoustic form for a given phoneme will be influenced by the surrounding phonemes. One example is the case of fricative–vowel syllables in American English. Although /s/ is supposed to be pronounced with unrounded lips, a following rounded vowel leads to some anticipatory lip rounding, making the fricative more /ʃ/-like. How are such coarticulated phonemes then recognized? A number of studies have shown that coarticulation is compensated for in perception (e.g., Beddor & Krakow, 1999; Fowler & Brown, 2000; Liberman, Delattre, & Cooper, 1952; Mann, 1980; Mann & Repp, 1981). Mann and Repp (1980), for instance, showed that listeners will still accept a fricative with some cues for lip rounding as /s/ if it is followed by a rounded vowel, whereas the same fricative is interpreted as /ʃ/ if followed by an unrounded vowel. Listeners thus take the context into account in a compensatory way when making phonetic decisions. The underlying mechanisms that cause compensation for coarticulation are hotly debated. In the present article, I will contrast an auditory account with phonological accounts (see Figure 1).

Auditory accounts hold that context sensitivity is pervasive in perception in general (Kluender, Coady, & Kiefte, 2001; Warren, 1999) and that perceptual systems have evolved to cope with the lack of invariance in the environment. Lack of invariance in the environment is caused by, among other forces, inertia. Inertia is also one cause of coarticulation (Farnetani, 1997; Whalen, 1990). Given that the auditory system has evolved to deal with inertia, it can compensate for coarticulation through general auditory perception. One particular mechanism that has been put forward is *spectral contrast*, which arises from adaptation at different levels along the auditory-processing chain (see, e.g., Holt, 2005; Holt, Lotto, & Kluender, 2000). It is assumed that although coarticulation with a preceding /l/ leads to a higher *F*3 in a velar stop /g/—making it more /d/-like—listeners adapt to the high *F*3 in the preceding /l/. This adaptation decreases the sensitivity for frequencies in the higher part of the *F*3 skirt for the [g], thereby decreasing the perceived *F*3 center frequency and canceling out the /l/'s coarticulatory influence. Evidence for spectral contrast stems from, for instance, an experiment by Lotto and Kluender (1998), who used stimuli with the structure "sine wave sound + /{d, g}V/." Lotto and Kluender found that listeners gave more /g/ responses after a high sine wave sound with a frequency similar to the *F*3 center frequency of [l] than after a lower sine wave sound with a frequency similar to the *F*3 center frequency of [r]. This indicates that the frequency content of the speech context is sufficient to induce compensation for coarticulation. Substitution of speech sounds with nonspeech sounds has been shown to be effective—or better, not effective in

I thank Roel Smits for discussions and his encouragement in developing this project and Alexandra Jesse, Michael Tyler, Carol Fowler, Anne Cutler, and two anonymous reviewers for their comments on previous versions of the manuscript. Marloes van der Goot and Marieke van Heugtem are to be thanked for their help in running the experiments, as well as Jan Peter de Ruiter for his convincing lip rounding. Correspondence concerning this article should be sent to H. Mitterer, Max-Planck Institut für Psycholinguistik, 6500 AH Nijmegen, The Netherlands (e-mail: holger.mitterer@mpi.nl).

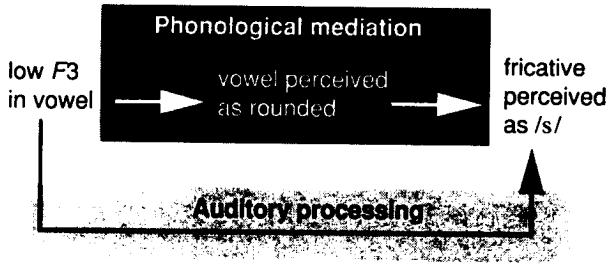

Figure 1. Two possible accounts for a context effect that occurs in fricative–vowel syllables in Dutch, as well as in English.

the sense that the nonspeech sounds create effects similar to those of speech sounds—in later investigations of the liquid–stop case (Blomert, Mitterer, & Paffen, 2004; Holt & Lotto, 2002; Lotto, Sullivan, & Holt, 2003), in CVC syllables (see Holt et al., 2000, for the speech effect reported by Lindblom & Studdert-Kennedy, 1967) and in VCV syllables (Coady, Kluender, & Rhode, 2003). Mitterer and colleagues (Mitterer, Csépe, & Blomert, 2006; Mitterer, Csépe, Honbolygo, & Blomert, 2006) showed that contrast effects are restricted neither to spectral contrasts nor to right-to-left effects. They found that general auditory processes may be involved in compensation for a regressive (i.e., left-to-right) manner assimilation—essentially, an anticipatory coarticulation—in Hungarian, in which amplitude modulation was the crucial acoustic cue.

Rather than being a consequence of auditory processing, context effects may be phonologically mediated by the perceived identity of the context sounds (see Figure 1). In this vein, two quite distinct theories argue that speech perception takes the speech production mechanisms into account. *Motor theory* (e.g., Liberman, 1996; Liberman & Whalen, 2000) argues that listeners recover the intended speech gestures by testing which intended gestures could account for the acoustic input—that is, by analysis-by-synthesis. According to the *direct perception* account (Fowler, 1996), such an inference is not necessary, because listeners directly perceive the articulatory gestures. Compensation then occurs because listeners parse the speech signal along gestural lines (cf. Fowler & Smith, 1986) and ascribe the lip rounding during the frication part of a syllable such as [su] to the rounded vowel and, thereby, perceive a context-invariant fricative gesture without intrinsic lip rounding. Despite the conceptual differences between these accounts, both assume that coarticulation is compensated for because the listener directly perceives or reconstructs the invariants in the production mechanisms.

A third possibility within the class of phonological accounts is a statistical-learning account: Listeners may learn how phonemes interact in connected speech (see, e.g., Gaskell, 2003, for a model of a possible learning mechanism). There is some evidence that learning is involved in compensation for assimilation. Beddor and colleagues (Beddor, Harnsberger, & Lindemann, 2002; Beddor & Krakow, 1999; Darcy, Peperkamp, & Dupoux, in press) showed that languages differ in their fine-grained detail of coarticulatory pattern and that listeners' compensatory patterns are adjusted to that. If listeners from different language backgrounds are presented with identical items, they show more compensation for native-like patterns of coarticulation.

Learning accounts are not fundamentally at odds with auditory (Diehl, Lotto, & Holt, 2004; Holt, Lotto, & Kluender, 2001) or gestural (see, e.g., Best, 1995; Gibson, 1979, p. 141) theories. However, for any given pattern of coarticulation, the absence of an effect of learning has been taken as evidence for either an auditory account (Lotto, Kluender, & Holt, 1997) or a gestural account (Fowler & Dekle, 1991; Mann, 1986).

Pursuing a phonological-learning account, Smits (2001a, 2001b) investigated compensation for coarticulation in Dutch fricative–vowel sequences, in which a palatal /ʃ/ or an alveolar fricative /s/ is followed by a rounded or unrounded vowel. The phonetic implementation of the fricative contrast differs from American English (the language in Mann & Repp, 1980) on at least two counts. The Dutch /s/ is less "sharp" in quality than its American counterpart, and the place contrast is not enhanced by lip rounding, since both fricatives are unrounded (cf. Booij, 1995). The distinction between the fricatives is carried mainly by the frequency of the fricative pole (FP), which is higher for the alveolar than for the palatal fricative. Smits (2001a, 2001b) showed theoretically—on the basis of a simple tube model—and empirically that anticipatory lip rounding leads to a lower FP if fricatives are followed by a rounded vowel [y] than if these fricatives are followed by an unrounded vowel [i]. This lowering is larger for the alveolar fricative than for the palatal fricative, so that the difference in FP frequency between [s] and [ʃ] is smaller with a following [y] than with a following [i]. Smits argued that listeners *learn* the dependency of FP frequency on the roundness of surrounding vowels and compensate for this by accepting fricatives with lower poles as instances of [s] if they *perceive* the following vowel as the vowel [y], as opposed to cases in which they are in front of the unrounded [i].

In order to differentiate a phonological from an auditory effect, Smits (2001a) applied logistic regression models to a large data set of fricative–vowel identifications. He first accounted for a possible auditory effect by using the third formant of the vowel ($F3$)[1] as a predictor for fricative identification. Unsurprisingly, $F3$, a correlate of lip rounding, influenced fricative identification in a compensatory manner. A low $F3$, indicating lip rounding, inclined listeners to make a *high*[2] decision (i.e., [s]) for the fricative. However, even after accounting for a possible auditorily driven context effect, a significant amount of variance could still be accounted for by assuming a learned phonological mediation: That is, the (fuzzy) vowel categorization influenced fricative categorization, even after accounting for an auditory effect. Smits (2001a, 2001b) called this the *hierarchical categorization* of phonemes, given that the decision for the vowel influences the decision made for the fricative. The hierarchical categorization is supposed to be a consequence of a statistical-learning algorithm, which has picked up that FPs are lower if fricatives are followed by rounded vowels.

One problem with this analysis is the colinearity of the predictors $F3$, standing for an auditory effect, and *vowel identification*, standing for a learned phonological effect. Because $F3$ influences *vowel identification*, the predictors are correlated. Hence, alternative interpretations are possible. An advocate of auditory effects might point out that vowel identification is a nonlinear function of $F3$. The fact that vowel identification explains *fricative identification* better and above any effect of $F3$ may still be explained as an auditory effect: Using vowel identification and $F3$ as predictors for fricative identification allows $F3$ to influence fricative identification not only linearly, but also nonlinearly.

Conversely, an advocate of a phonological dependency of fricative and vowel identification might argue that the contribution of $F3$ to fricative identification is simply a consequence of $F3$'s driving vowel identification, which, in turn, influences fricative identification. Indeed, in some of the analyses of Smits (2001a), removing $F3$ as a predictor for fricative identification did not impair prediction accuracy if the predictor vowel identification was used in the regression model.

The aim of the experiments reported in this article is to separate these conflicting accounts. After replicating the compensation effect in Experiment 1, I manipulated the acoustic properties of the context without influencing vowel identification in Experiment 2 and manipulated vowel identification without modifying $F3$ in Experiment 3. If compensation for coarticulation is a consequence of auditory processing, there should be a context effect in Experiment 2, but not in Experiment 3. Conversely, phonological mediation is implicated if there is an effect in Experiment 3, but not in Experiment 2.

EXPERIMENT 1
Replicating the Compensation-for-Coarticulation Effect

Smits (2001a) showed that Dutch listeners compensate for anticipatory lip rounding in fricative–vowel syllables. Listeners are more likely to perceive a fricative as /s/ if it is followed by an [y] than if it is followed by an [i]. In that experiment, Smits asked participants to identify the syllables in four sessions, with the first session purely for familiarization. This led to systematic identifications of the endpoints. Informal pretesting for the present study indicated that untrained participants had great difficulties in identifying Smits's stimuli systematically in only one session. In order to achieve systematic identification performance without extensive training, I slightly increased the range of the FP and $F3$ frequencies. Experiment 1 evaluated whether similar compensation for anticipatory lip rounding in fricative–vowel syllable results can be found with these slightly changed stimuli.

Method

Participants. Ten members of the Max Planck Institute for Psycholinguistics (MPI) participant pool participated in the study. All were native listeners of Dutch, with no known hearing impairment, and were between 17 and 25 years of age. As with all Dutch university students, these participants had a good command of English (cf. Akker & Cutler, 2003; Weber & Cutler, 2004), and all of them had at least some knowledge of a third language, either German or French.

Materials. The stimuli were synthesized using Praat 4.0 (Boersma & Weenink, 2004) with synthesis parameters similar to those in Smits (2001a). Fricatives were created by filtering a white noise source (sampling rate = 16 kHz) by a two-formant filter. The source had a duration of 0.18 sec, and the noise was faded in linearly over the first 100 msec, stayed level for 50 msec, and was faded out to one half of the maximal amplitude over the last 30 msec. In order to prevent a click at the end, an additional linear fade-out to zero was applied for the last 5 msec. The filter had a fixed second formant at 6.5 kHz (bandwidth [BW] = 2.6 kHz), whereas the first formant (later referred to as the FP) had a center frequency range from 2890 to 3410 Hz (BW = 0.1 * center frequency) in seven equal steps in barks. Vowels were created using a 0.2-sec pulse train falling from 135 to 115 Hz as the source. The second and third harmonics were amplified in order to create a more natural sounding source. The amplitude of the source started at one sixth of the maximum and reached maximum at 40 msec, stayed level for 110 msec, and was reduced to half over the last 50 msec, with linear interpolation between indicated time points. In addition, a 5-msec linear fade-out was applied to prevent clicks. All formants except $F3$ were fixed: $F1$ = 250, BW1 = 60; $F2$ = 1,900, BW2 = 80; $F4$ = 3,200, BW4 = 130; $F5$ = 4,000, BW5 = 150; $F6$ = 5,500, BW6 = 200; $F7$ = 6,500, BW7 = 200. The $F3$ ranged from 2400 to 2725 Hz (BW = 110), using seven equal steps in bark. The seven fricative and the seven vowel stimuli (= 49 stimuli) were concatenated in that order without an intervening silence.

Procedure. The experiment was run with the participants facing a computer screen and with a four-button response box in front of them. All the participants first completed a short familiarization session. In this familiarization, they heard four examples of the four endpoint stimuli that arise at the minimal and maximal FP and $F3$. They were instructed to indicate which of the four syllables they had heard by pressing one of four buttons labeled "sie," "suu," "sjie," and "sjuu," the Dutch orthographic transcriptions of the phonological forms /si/, /sy/, /ʃi/, and /ʃy/, respectively. The instructions stressed accuracy. After pressing a button, the participants received feedback as to whether they had heard the stimulus as intended, in the form of a happy or sad cartoon face. If the listeners failed to react within a 2-sec deadline, a stopwatch appeared on the screen. After the training trials, the experiment started, in which no feedback was provided. During the experiment, each listener heard each of the 49 stimuli 10 times and categorized them.

Design. Separate logistic regression analyses were carried out with the fricative and vowel responses as dependent variables and FP and $F3$ frequency as independent variables. Reactions were coded so that *high* responses ([s] for the fricative and [i] for the vowel) were coded as 1 and *low* responses as 0 ([ʃ] for the fricative and [y] for the vowel). In this and all the other regression models presented here, predictors were normalized to a range from 0 to 1, respecting the height dimension (in terms of acoustic frequency) in the stimulus. Hence, positive regression weights indicate that a high value in a predictor variable makes a high response more likely.

Results and Discussion

A logistic regression analysis for the complete data set showed that the perception of the vowels (see Figure 2A) was influenced by $F3$ ($\beta = 2.89, p < .001$) and FP ($\beta = 0.24, p < .05$). As in the data of Smits (2001a), vowels with a lower $F3$ were more likely to be perceived as a rounded /y/. Moreover, a high FP in the fricative triggered /i/ responses. This is akin to the result that Fowler (2006) called

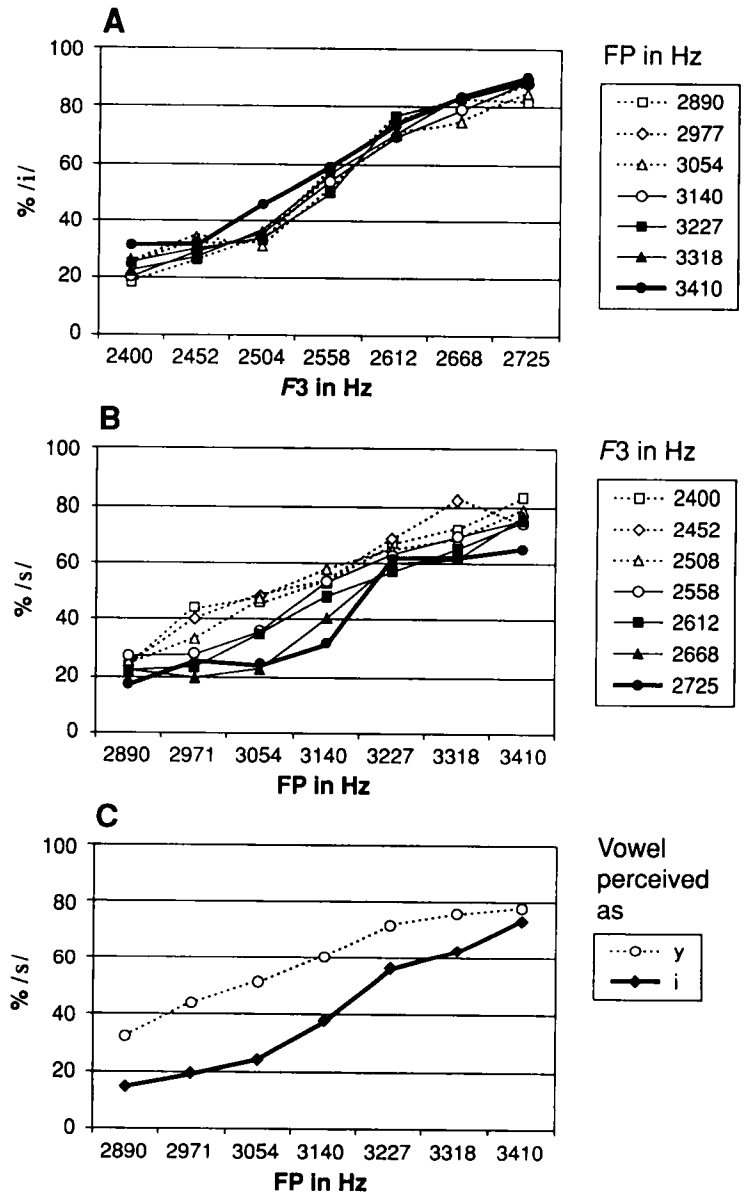

Figure 2. Results from Experiment 1. (A) Percentages of vowels perceived as [i] depending on the *F*3 (abscissa) and fricative pole (FP) of the preceding fricative (different functions). (B) Percentages of fricatives perceived as [s] depending on the FP (abscissa) and *F*3 of the following vowel (different functions). (C) Percentages of fricatives perceived as [s] depending on the FP (abscissa) and perceived vowel identity (different functions).

the *companion finding*. A high FP is used as an indication of the presence of a following unrounded vowel, maybe because high FPs do not occur with rounded vowels. The perception of the fricatives (see Figure 2B) as /s/ was more likely if the FP was high ($\beta = 3.91, p < .001$) and if the *F*3 was low ($\beta = -0.88, p < .001$). The latter effect reflects compensation for coarticulation; a low *F*3 is likely to occur in a rounded vowel, and a rounded vowel lowers FPs and, therefore, leads the listener to accept lower FPs for [s].

Another noteworthy aspect of the data falls out of the logistic regression analysis if applied to single subjects. In line with the assumption of a phonological mediation (i.e., *F*3 influences vowel identification, which in turn influences fricative identification), there was a negative correlation between the β-weights of *F*3 on vowel identification and fricative identification ($r = -.68, p < .05$). This indicates that the compensatory—thus, negative—influence of *F*3 on fricative identification was stronger

for participants who showed a larger than average positive influence of $F3$ on vowel identification.

The data were also subjected to the more traditional method of an ANOVA, pooling the data for each listener and predicting the percentages of fricative and vowel responses, with FP and $F3$ as predictors. In this analysis, there was also, for the fricative responses, a significant influence of FP [$F(6,54) = 29.42$, $p < .001$] and $F3$ [$F(6,54) = 13.35$, $p < .001$]. The interaction was also significant [$F(36,324) = 1.76$, $p < .05$], indicating that the context effect was smaller for the less ambiguous fricatives. For the vowel identifications, the effect of $F3$ was significant [$F(6,54) = 18.27$, $p < .001$] and the effect of FP was almost significant [$F(6,54) = 2.09$, $p = .07$], whereas the interaction was not ($F < 1$).

It might be argued that the present results are difficult to interpret because the identification functions do not reach 0% and 100% at the continuum endpoints. However, the stimulus continua did not, on the one hand, invoke a bias toward one kind of response and did, on the other hand, give rise to a strong context effect. For the fricative identifications, the figure shows a minimum of 18% and a maximum of 82% of /s/ identifications over all the experimental cells. Incidentally, this endpoint consistency is nearly identical to that of Smits (2001a), who found a minimum of 17% and a maximum of 82%. Moreover, over all stimuli, the mean percentage of /s/ responses was 49.4%. This shows not only that the endpoints have a similar distance to the 50% point, but also that the listeners did not have a strong bias toward /s/ or /ʃ/ responses overall. Although the endpoint consistency with the vowels is less symmetrical with a minimum of 18% and a maximum of 90% /i/ identifications, the overall mean of 55.2% indicates that there was no strong bias toward /i/ responses. Most important, there was a clear context effect, which can be visualized more clearly if the fricative responses are plotted contingent on the vowel responses, instead of the vowel stimuli (see Figure 2C vs. 2B). The difference in fricative identification approaches 30%. This figure also nicely illustrates the question at hand. Listeners give more /s/ responses if they perceive the vowel as /y/. However, the perception of /y/ is correlated with a low $F3$ in the vowel. So what does cause the context effect, the low $F3$ or the perception of the vowel as rounded (see Figure 1)? Experiment 2 explored the first possibility (the direct path), and Experiment 3 the second (the mediated path).

EXPERIMENT 2A

Given the presence of a compensation and a companion effect, this experiment investigated whether the acoustic properties of the context sounds are sufficient to trigger these effects. To this end, $F3$ should be modified independently of vowel identification. This was achieved by replacing $F3$ with a single sine wave sound at the critical formant center frequency (following, e.g., the example of Lotto & Kluender, 1998). In this case, the context sound varies in frequency, substituting for $F3$ without carrying phonological information. Although a combination of frequency and amplitude modulated sine waves may be perceived as speech (Remez, Rubin, Berns, Pardo, & Lang, 1994), it is difficult to see how a single steady-state sine wave might evoke a phonetic or phonological percept. If such sounds produce context effects similar to those of speech sounds, a strong case can be made for an auditorily based context effect. Both context effects, the compensation and the companion effects, were investigated, by, first, replacing the vowel with a single sine wave at the center frequency of $F3$ and, second, replacing the fricative by a single sine wave at the center frequency of FP. If there is an auditory effect, lower sine wave sounds, replacing the vowel, should trigger more /s/ responses (the compensation effect), and a high sine wave, replacing the fricative, should trigger /i/ responses (the companion effect).

Method

Participants. Ten members of the MPI participant pool participated in the study. All were native listeners of Dutch, with no known hearing impairment, and were between 17 and 25 years of age.

Materials. The same speech stimuli as those in Experiment 1 were used, and the fricatives and vowels were presented with nonspeech sounds to replace the vowel or the fricative, respectively. The nonspeech sounds were steady-state sine wave sounds with the center frequencies of the speech stimuli (FP in fricatives and $F3$ in vowels). These sounds were then multiplied by the amplitude envelope of the fricatives and the vowels and were equated in overall loudness with the speech stimuli in the following way. The original speech stimuli were filtered with a band-pass filter (one-third octave) centered at the FP or $F3$ frequency, in order to estimate the loudness in the critical band around FP or $F3$. The sine wave substitutes were then equated in loudness (in sones) with the filtered speech sounds. As a consequence, the sine wave substitutes were lower in acoustic energy overall than the speech sounds but had as much energy in a critical bandwidth around $F3$ and the FP as the original vowels and fricatives.[3]

In order to reduce the number of conditions, only the first, third, fifth, and last steps of each continuum were used as a template for a sine wave substitute. Probing whether the compensation effect can be triggered by a nonspeech sound was done with the 28 fricative–tone "syllables" that arose by concatenating one of the seven fricatives with one of the four sine wave sounds having the frequency of the first, third, fifth, or last step of the $F3$ continuum. Similarly, there were 28 tone–vowel "syllables" (one of the four sine wave sounds replacing the fricative plus one of the seven vowels). These stimuli were presented binaurally over headphones (Sennheiser HD250) in a sound-attenuated booth, using a computer controlled by the NESU software.

Procedure. The experiment was run with the participants facing a computer screen and a four-button response box. All the participants first completed a short familiarization session. In this familiarization, they heard each of the four endpoint stimuli with the minimal or maximal FP and $F3$ four times. They were instructed to indicate which of the four syllables they heard by pressing one of four buttons labeled "sie," "suu," "sjie," and "sjuu," the Dutch orthographic transcriptions of the phonological forms /sɪ/, /sy/, /ʃɪ/, and /ʃy/. The instructions stressed accuracy. After pressing a button, the participants received feedback as to whether they had heard the stimulus as intended, in the form of a sad or happy cartoon face. If the listeners failed to react within 2 sec, a stopwatch appeared on the screen. After the 16 training trials, the labels on the response buttons were removed, and the experiment started.

The experimental session consisted of two blocks. In one block, the participants identified fricatives followed by a sine wave substitute for the vowel, and in the other block, the participants identified vowels preceded by a sine wave substitute for the fricative. Half of

the participants started with fricative identification; the other half with vowel identification. In each session, each of the 28 stimuli was presented 15 times. The participants were told to use only the two inner buttons of the four-button response box to give responses. The assignment of target sounds to the left and right buttons was indicated by two figures on the left and right halves of the computer screen (e.g., "s" or "sj"). No explicit feedback was given, but the chosen alternative was highlighted on the screen. On every 60th trial, the participants had the opportunity to take a short break.

Design. The results were analyzed using logistic regression analyses for fricative and vowel identification sessions separately. For fricative identification, the independent variables were FP frequency (seven levels) and sine wave vowel substitute frequency (four levels; $F3_s$, "s" for sine wave). For vowel identification, the independent variables were sine wave fricative substitute frequency (four levels; FP_s), and $F3$ frequency (seven levels). The dependent variable was identification of the fricatives and vowels (coded as /ʃ/, /y/ = 0; /s/, /i/ = 1, analogous to the higher FP and $F3$ in [s] and [i], respectively).

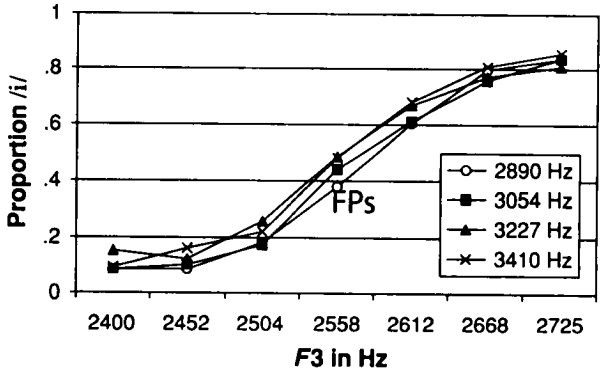

Figure 3. Mean proportions of [i] identifications depending on $F3$ (abscissa) and the preceding sine wave fricative substitute (different functions) in Experiment 2A.

Results and Disctussion

Vowel identification. Figure 3 shows the mean percentages of /i/ identifications for all combinations of FP_s and $F3$ frequency. Logistic regression analyses for the group, as well as for individual subjects, indicated that both independent variables influenced vowel identification. Besides the expected positive effect of $F3$ ($\beta = 4.39$, $p < .001$), there was also a small positive effect of the preceding sine wave ($\beta = 4.39, p < .05$), which was caused by the monotonic increase of the relative frequencies of /i/ responses with increases in the pitch of the preceding sine wave (2890 Hz, 41.5%; 3054 Hz, 42.4%; 3227 Hz, 46.1%; 3410 Hz, 46.5%). This effect is similar to the companion finding observed in Experiment 1, where a high FP made it more likely that the vowel would be perceived as unrounded. The effects of $F3$ and FP_s on vowel identification also proved significant when evaluated with an ANOVA for the pooled percentages for each participant [FP_s, $F(3,27) = 5.67, p < .01$; $F3$, $F(6,54) = 63.34, p < .001$], whereas their interaction did not ($F < 1$). This result indicates that the "companion finding" of compensation for coarticulation can be elicited by a nonspeech sound. The fact that a fricative with a high FP induces the perception of the vowel as unrounded is classically interpreted as evidence for gestural parsing of the rounding gesture. The present result indicates that such an articulatory inference—or direct perception of gestures—may not be necessary to account for this effect. Instead, the acoustic differences between the fricatives, and not their phonological interpretation, seem to be sufficient to trigger such an integrative effect.

Fricative identification. Figure 4 shows the mean percentages of /s/ identification for all combinations of FP frequency and $F3_s$. The overall logistic regression analysis seems to indicate that the results were similar to those of Experiment 1. There is a significant positive β-weight for FP ($\beta = 0.53, p < .001$) and a negative β-weight for $F3_s$ ($\beta = -0.62, p < .001$), showing a compensatory influence of the nonspeech replacement for the vowel that is similar to the compensatory influence of the vowel itself. A similar picture arises in an ANOVA analysis [FP, $F(6,54) = 3.87, p < .01$; $F3_s$, $F(6,54) = 3.87, p < .05$; FP × $F3_s$, $F(18,162) = 1.43, p > .1$]. The effect of the sine wave sound is caused by a monotonic *decrease* of /s/ responses with higher sine wave sounds (2400 Hz, 48.2%; 2504 Hz, 45.4%; 2612 Hz, 36.7%; 2725 Hz, 35.2%). This

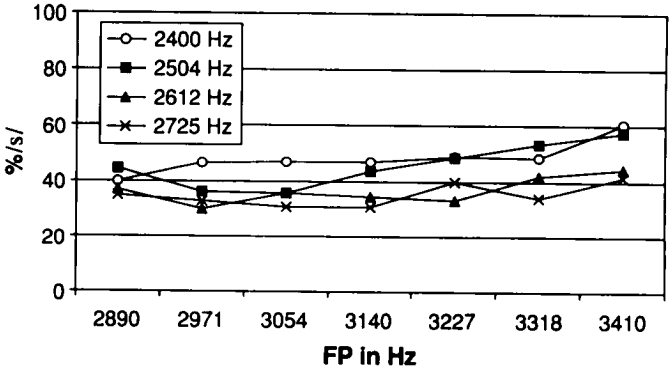

Figure 4. Mean percentages of [s] identifications depending on fricative pole (FP) frequency (abscissa) and the following sine wave vowel substitute (different functions) in Experiment 2A.

replicates the results obtained with speech sounds in Experiment 1.

However, an examination of the individual data reveals a puzzling aspect (see Table 1): No listener used both FP and $F3_s$ to distinguish fricatives. Four listeners used only FP, 3 listeners used only $F3_s$, and 3 listeners failed to show any significant effects, with the first participant answering /ʃ/ on all the trials. That means that only the listeners who failed to use FP for fricative identification showed a compensatory effect of the sine wave sound. It is important to note that the mean β-weight for FP was significantly smaller in this experiment than in Experiment 1 [$t(18) = 3.70$, $p < .01$]. Similarly, the endpoints of the FP continuum were identified more consistently in Experiment 1 than in this experiment (Experiment 1, 2890 Hz → 77% /ʃ/, 3410 Hz → 76% /s/; Experiment 2A, 2890 Hz → 62% /ʃ/, 3410 Hz → 52% /s/), with much shallower identification functions than in Experiment 1. This indicates that the fricative noises were less distinguishable if presented as the only speech sound. This makes it difficult to interpret the effect of the sine wave substitute. Therefore, Experiment 2B was run to test the effect of a sine wave context sound with more consistent fricative identification.

EXPERIMENT 2B

Why are the fricative noises less distinguishable when presented with a single sine wave sound in the vicinity of $F3$ than when presented with a vowel? Unlike stops (see Liberman, 1996), steady-state fricatives clearly can be produced and perceived as isolated speech sounds without an accompanying vowel. The effect of the vowel may be ascribed to a form of speaker normalization (cf. Ladefoged & Broadbent, 1957; Nearey, 1989), because the vowel's $f0$ and the higher formants may provide a frame of reference for the fricative noises. Therefore, the fricative identification task was repeated with a slight modification: The fricative noise was not only followed by a sine wave vowel substitute, but also preceded by the vowel [ɛ], which is neutral in terms of lip rounding.

Method

Participants. Eleven listeners participated in the study. Six were members of the MPI participant pool, and 5 students, recruited from a senior high school, came to the Institute for a school project. All were native listeners of Dutch. The responses of 1 listener were discarded because he reported having a hearing problem.

Material. The fricatives and sine wave sounds substituting for a following vowel were the same as those in Experiment 2A. The only change with regard to Experiment 2A was that the fricative plus sine wave vowel substitute quasisyllables were now preceded by an additional constant vowel. This preceding vowel was synthesized with a length of 100 msec and a steady $f0$ (125 Hz) and with an $F1$, $F2$, and $F3$ at 600, 1700, and 2500 Hz, respectively (BW at 10% of the formant frequency). The other formants were identical to the other vowels.

Procedure and Design. The procedure and the design were identical to those for the fricative identification part of Experiment 2A, except for two procedural changes: The listeners were now trained on the endpoints of the fricative continuum of [ɛ]–fricative syllables, using a two-alternative forced choice task (/s/ or /ʃ/). Explicit feedback was given during training. This training consisted of 16 trials and was repeated if the participant made more than five errors, which was the case for 2 out of the 10 participants, who needed two sessions to pass the criterion. After training, there were two experimental sessions separated by a compulsory break of at least 2 min. In each session, each of the 28 stimuli (seven fricatives crossed with four different sine wave sounds) was presented 12 times.

Results and Discussion

Figure 5 shows the mean percentages of [s] identifications depending on FP frequency on the abscissa and different functions for each vowel sine wave frequency. Although the fricatives were the same as those in Experiment 2A, the listeners distinguished the endpoints in a more systematic fashion, with an endpoint consistency of 82% and 85% for the low and high ends of the FP continuum (Experiment 2A: 62% and 52%). This suggests that introducing a vowel may provide listeners with a kind of anchor for perceiving the fricative sounds as different from one another.[4] The logistic regression analyses also revealed, besides the obvious effect of FP frequency on fricative identification, an integrative effect of the $F3_s$ (see Table 2). In contrast to the previous experiment, there was a tendency for the listeners to respond /s/ (i.e., *high*) more often in the case of a high sine wave vowel substitute, which is opposite to the compensation effect observed for the speech sounds. It is interesting to note, however, that the overall integrative effect varied over listeners. Due to this interindividual variability, an ANOVA on the mean percentages of /s/ responses revealed only a significant effect of FP [$F(6,54) = 94.2$, $p < .001$], whereas the effect of $F3_s$ and the interaction failed to reach significance [$F(3,27) = 1.4$, $p > .1$; $F(18,162) = 1.3$, $p > .1$].

Most important, the present experiment nevertheless shows that it is unlikely that compensation for coarticulation in fricative–vowel sequences could be due to an auditory effect. Although an effect akin to compensation triggered by a nonspeech sound was found in Experiment 2A,

Table 1
β-Weights for Logistic Regressions With Fricative Pole (FP) and $F3$ Sine Wave Substitute as Independent Variables and Fricative Reaction as a Dependent Variable in Experiment 2A for All and Individual (P1–P10) Participants

	Participant										
	All	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
FP	0.53***	–	0.02	0.80***	0.00	−0.27	−0.05	1.51***	1.86***	2.32***	0.28
$F3_s$	−0.62***	–	−0.67***	−0.11	−0.72***	0.16	−4.16***	−0.1	0.61	−1.33	−0.57

*** $p < .001$.

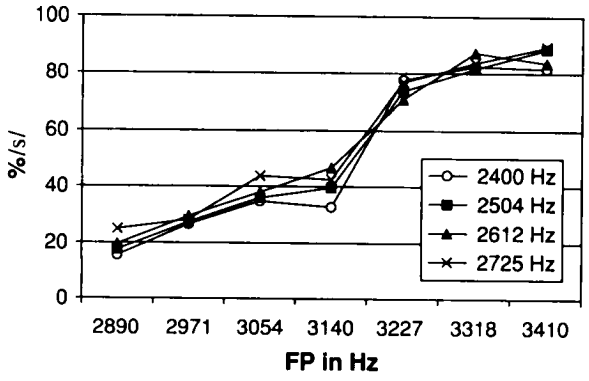

Figure 5. Mean percentages of [s] identifications depending on fricative pole (FP) frequency (abscissa) and the following sine wave vowel substitute (different functions) in Experiment 2B.

this effect occurred only for the participants who did not use FP to distinguish the fricative. When fricative identification was improved experimentally in Experiment 2B, the compensatory effect triggered by the nonspeech sound vanished. Consequently, Experiments 2A and 2B are compatible with the assumption that the perception of fricative–vowel sequences involves a phonological mediation, in the sense that vowel identification influences fricative identification. This possibility was further investigated in the next experiment.

EXPERIMENT 3

Changing vowel identification influenced fricative identification in Smits (2001a) and in Experiment 1 of the present study. In these experiments, vowel identification was driven by acoustic *F*3. In the present experiment, vowel identification was modified independently of *F*3 by using visual cues for speech recognition (Johnson, Strand, & D'Imperio, 1999; McGurk & MacDonald, 1976). In a classic article, McGurk and McDonald presented mismatching natural auditory and visual stop–vowel syllables to listeners. They found that the identity of the visual syllables influenced the identification of the audiovisual stimulus. When, for instance, a visual [ga] and an acoustic [ba] were presented, the listeners were most likely to report hearing /da/. The labial gesture (in this case, closing or not closing) especially influenced the perception of the acoustic syllable. The listeners seldom reported hearing a labial stop if the visual display did not show a labial closure, whereas the presence of a visual labial closure often led the listeners to perceive a labial stop. This shows that visual context effects are, unsurprisingly, most effective for cases of speech gestures with clearly visible signatures. Fortunately for the present case, lip rounding is a highly salient visual cue, which should influence the identification of an acoustically ambiguous vowel between a rounded [y] and an unrounded [i] (cf. Johnson et al., 1999). The question, then, is whether listeners also accept more fricatives as [s] if the visual display leads them to perceive the following vowel as [y]. Such an effect would be strong evidence for a phonological mediation of the context effects observed in Experiment 1.

An audiovisual design has already been applied to study compensation for coarticulation in liquid–stop (Fowler, Brown, & Mann, 2000; Holt, Stephens, & Lotto, 2005) and the fricative–stop (Vroomen, 1992; Vroomen & de Gelder, 2001) cases. These experiments showed that perceivers can use visual cues to compensate for coarticulation in these sequences (Fowler et al., 2000). However, visual effects are restricted to visual information that is presented simultaneously with the stop (Holt et al., 2005). Visually influencing the perception of the context sounds—the liquid or the fricative—is not sufficient to induce listeners to adjust their stop boundaries (Holt et al., 2005; Vroomen, 1992).

Method

Participants. Ten members of the MPI participant pool participated in the study. All were native listeners of Dutch with no known hearing impairment.

Materials. A subset of the acoustic stimuli used in Experiment 1 was selected for this experiment. The whole range of fricative noises was used, and for the vowel the fourth to sixth steps of the [y]-to-[i] continuum were used. This gave rise to 21 acoustic fricative–vowel syllables. In order to create audiovisual stimuli, a phonetically naive male native speaker of Dutch was digitally video recorded (25 frames per second) saying [si], [ʃi], [sy], and [ʃy] several times. Visual recordings for the experiment were selected so that a tight coupling between visual and auditory stimuli was achieved. Therefore, one recording for each syllable type with a fricative-to-syllable duration ratio (46%–47%) similar to the fricative-to-syllable duration ratio of the synthetic stimuli (47%) was chosen. The length of the natural utterance was measured, and the video was accelerated by 7%–29%, so that the natural recording was similar in length to that of the synthetic stimuli. For the preparation of the final stimuli, the videos were then cut to a length of 1 sec, with the natural sound beginning at the fifth frame. For each of the four videos, a fade-in and fade-out were created by interpolating between a black frame and the first or last frame of the video over five frames. The original sound was then replaced by 1 of the 21 acoustic fricative–vowel

Table 2
β-Weights for Logistic Regressions With Fricative Pole (FP) and *F*3 Sine Wave Substitute as Independent Variables and Fricative Reaction as a Dependent Variable in Experiment 2B for All and Individual (P1–P10) Participants

	Participant										
	All	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
FP	3.62***	4.33***	3.50***	1.89***	3.48***	4.19***	5.88***	6.83***	2.80***	3.80***	9.57***
$F3_s$	0.23**	0.23	−0.63**	1.26***	0.08	0.53*	0.11	−0.00	−0.01	−0.15	0.38

*$p < .05$. **$p < .01$. ***$p < .001$.

syllables. With the four videos mouthing [si], [ʃi], [sy], or [ʃy], this gave rise to 84 stimuli.

Procedure. The procedure was similar to that for the pretest in Experiment 2. Each participant was first familiarized with the acoustic stimuli with the same training regime before the experimental session. In the experimental session, the participants responded to each of the 84 stimuli eight times. Every 60th trial, the participants had the opportunity to take a short break.

Each trial started with a fixation cross appearing on the center of the screen. After 500 msec, the video was displayed at the center of the screen. After the onset of the video, the participants had 3.7 sec to respond.

In order to encourage attention to the visual stimulus, 3% of all the trials were catch trials. In these trials, the word *stop* appeared over the face of the speaker. The participants were instructed not to respond to these stimuli. If, nevertheless, a response was given, a warning tone (1 kHz, 1 sec) was played to the participants.

Design. The results were analyzed, separately for fricative and vowel choices as dependent variables, with the four independent variables *visual fricative*, *visual vowel*, *F3 frequency*, and *FP frequency*. The effects of the independent variables were assessed with logistic regression analyses with predictors normalized to a range from 0 to 1. Reactions were coded so that a positive β-weight indicates that *high* values in the predictor (such as visual [i] or a high $F3$) lead to *high* responses (/ʃ/, /y/ = 0; /s/, /i/ = 1).

Results

General performance indicated that the participants paid attention to the visual display. The response rate to catch trials was much lower (9 participants, < 7%; 1 participant, 39%) than the response rate to experimental trials (>95% for all the participants). Figure 6 shows the mean percentages of /si/, /ʃi/, /sy/, and /ʃy/ responses for each of the 84 stimuli. Panels A–D show the responses to the visual syllables [si], [ʃi], [sy], and [ʃy], respectively. The abscissa indicates the values for $F3$ and FP. Using the abscissa for two independent variables creates the "zig-zag" pattern, because the $F3$ frequencies repeat after three steps, being paired with a new FP frequency. The two darker areas (black and dark gray) indicate the responses containing the vowel [i], whereas the two lighter areas (white and light gray) are associated with /y/ responses. The border between the dark gray area and the white area is the border between /ʃ/ and /s/ responses. This border can thus be read as the more familiar identification functions from a two-alternative forced choice task. There is a clear effect of the visual vowel on vowel responses: In panels A and B for visual [si] and [ʃi] there is more *dark* area (indicating /si/ and /ʃi/ responses) than in panels C and D for visual [sy] and [ʃy]. In order to evaluate statistically which independent variable influenced the response choice, logistic regression analyses were applied separately for the vowel and the fricative responses. As in the previous analyses, positive β-weights indicate that higher values in the predictor variable lead to higher probabilities for /i/ and /s/ responses. In contrast, negative β-weights indicate compensatory influences.

Vowel identification. Table 3 shows the β-weights for vowel responses for individual participants and for the whole group. An ANOVA using the percentages of /i/ responses confirmed the logistic regression results for the whole group: Vowel identification is influenced by the visual vowel [$F(1,9) = 45.2, p < .001$; visual [i], 90% /i/ responses; visual [y], 21% /i/ responses], $F3$ [$F(2,18) = 5.1, p < .05$; a monotonic increase in /i/ responses with $F3$ from 50% to 60%], and FP [$F(6,54) = 4.4\ p < .01$; higher FPs tended to give rise to fewer /i/ responses, with a minimal number of /i/ responses of 53% for the second highest FP and a maximum of 58% for the second lowest FP], whereas the visual fricative did not have an effect [$F(1,9) = 1.1, p > .1$]. None of the interactions turned out to be significant. Because the visual vowel influenced the vowel responses, it is possible to test whether the visually induced vowel identification shift also influenced fricative identification.

Fricative identification. Table 3 shows the β-weights for fricative responses for the whole group, as well as for individual participants. An ANOVA using the aggregated percentages of /s/ responses confirmed the logistic regression results for the whole group. Fricative identification was influenced by the visual vowel [$F(1,9) = 37.32, p < .001$; visual [i], 46% /s/; visual [y], 63% /s/], $F3$ [$F(2,18) = 17.8\ p < .001$; a monotonic decrease in /s/ responses with a rising $F3$ from 59% to 50%], and FP [$F(6,54) = 60.8\ p < .001$; a monotonic increase in /s/ responses with increases in FP from 17% to 89%], whereas the visual fricative did not have a significant effect [$F(1,9) = 1.5, p > .1$]. The indirect influence of the visual vowel on fricative identification is most evident in Figure 7, which plots the mean percentages of /s/ identifications for each FP frequency and visual syllable, averaged over $F3$ frequency. In addition, there were two significant interactions. The two-way interaction between visual vowel and FP was significant [$F(6,54) = 7.3, p < .001$], due to the smaller effect of the visual vowel for the endpoints of the FP continuum. Moreover, the three-way interaction between visual vowel, visual fricative, and FP was significant [$F(6,54) = 2.9, p < .05$], which is due to the fact that the effect of the visual fricative was more pronounced on the endpoints of the FP continuum in the case of a visual [i], but in the middle of the FP continuum in the case of a visual [y] (see Figure 7).

According to the hypothesis of a phonologically mediated effect, acoustic and visual context variables should influence fricative identification only if they also influence vowel identification. The logistic regression results for the individual participants reveal such a pattern. All the participants who showed a significant integrative influence of the visual vowel on vowel identification (i.e., all but P4) also showed a significant compensatory influence of the visual vowel information on fricative identification. The 7 participants who showed a compensatory influence of the acoustic vowel information included the 5 participants who had a significant positive β-weight for the influence of acoustic vowel information on the vowel response, plus 1 participant whose β-weight was marginally significant in the analysis with vowel response as the dependent variable (P3). In contrast, the 3 participants whose fricative identification was not influenced by the acoustic $F3$ of the vowel also failed to use the acoustic $F3$ for vowel identification. That indicates that, overall, acoustic and visual information about vowel identity influenced fricative identification—and accordingly, compensation for coarticulation—only if the

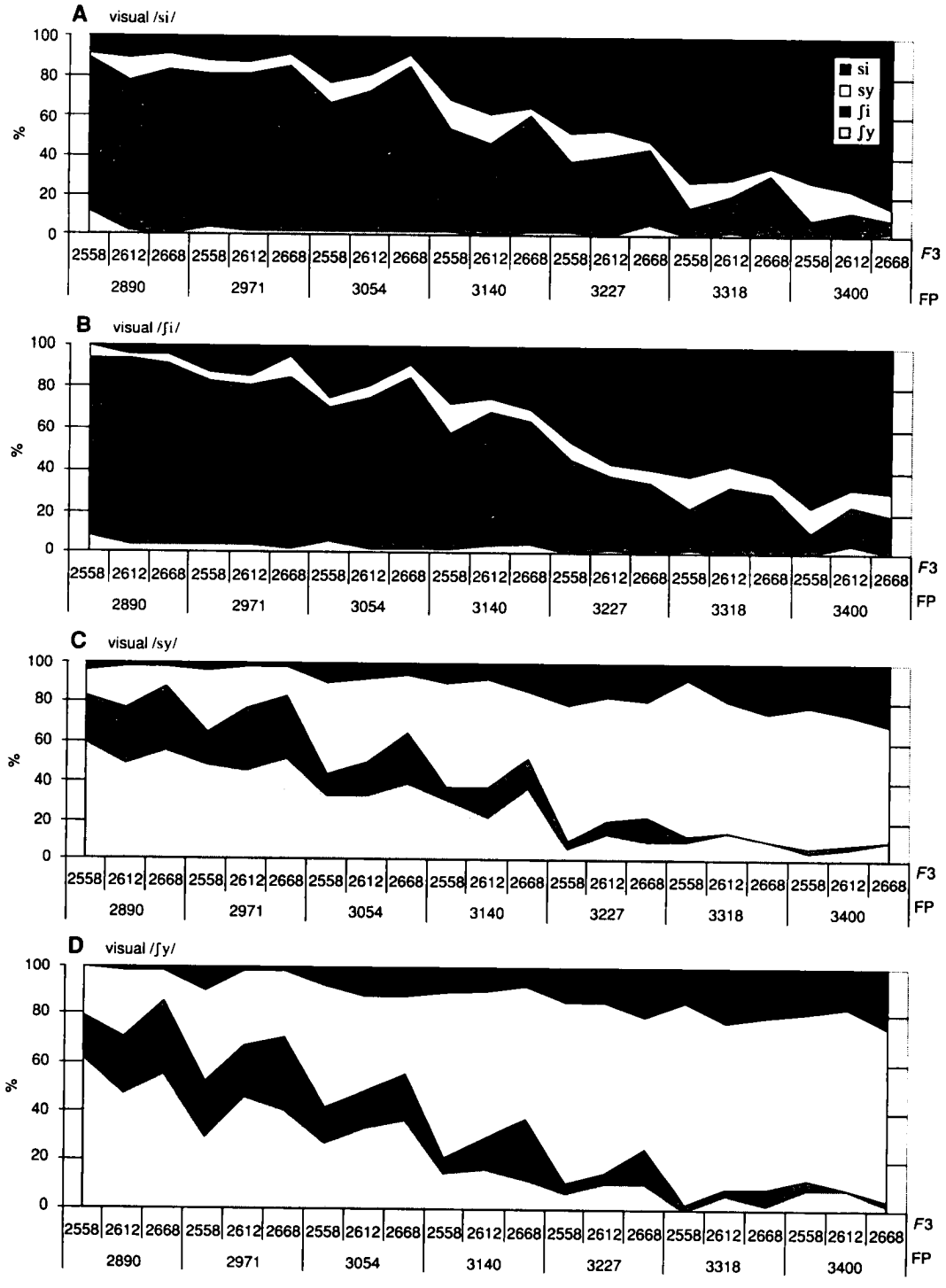

Figure 6. Mean percentages of [si] *sie*, [sy] *suu*, [ʃi] *sjie*, and [ʃy] *sjuu* identifications for each combination of *F*3, fricative pole (FP) frequency (abscissa), and visual syllables (different panels) in Experiment 3. See the text for further explanations.

information influenced vowel identification as well. There was one striking anomaly with regard to this pattern that a significant compensatory influence of vowel information presupposes an influence of that information on vowel identification. P10 showed a large but nonsignificant negative influence of *F*3 on vowel identification but still used *F*3 in a compensatory way for fricative identification. This was a consequence of this participant's strong reliance on the visual vowel, with only 4 out of 667 valid responses being inconsistent with the visual vowel. The negative value

Table 3
β-Weights for Logistic Regressions With Fricative Pole (FP), *F*3, Visual Vowel [V(V)], and Visual Fricative [V(F)] as Independent Variables and Vowel Reaction and Fricative Reaction as Dependent Variables for All and Individual (P1–P10) Participants in Experiment 3

	Participant										
	All	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
						Vowel Reaction					
*F*3	0.82***	1.21***	0.74***	0.57	0.15	0.41*	0.05	2.45***	0.32	1.97***	−1.69
V(V)	3.57***	6.65***	3.23***	7.87***	0.30	5.71***	5.69***	1.56***	5.55***	6.59***	11.22***
FP	−0.35**	0.00	−0.19**	−0.21	−0.04	0.03	−0.03	−0.93**	0.09	−0.96	−1.58
V(F)	0.05	0.37	−0.28	−0.90	−0.03	0.10	−0.16	−0.18	0.05	0.16	1.11
						Fricative Reaction					
*F*3	−0.49***	−0.36*	−0.56***	−0.34*	−0.04	−0.55***	0.04	−0.96**	−0.13	−0.69**	−1.25***
V(V)	−0.98***	−0.89***	−0.78**	−2.96***	−0.32	−1.23***	−0.98***	−2.01***	−1.09***	−0.44*	−1.48***
FP	3.98***	0.49***	1.39***	1.19***	0.53***	1.24***	0.42***	7.55***	1.58***	3.04***	6.62***
V(F)	0.20**	0.15	0.06	0.07	−0.05	0.06	1.42***	0.01	−0.09	−0.03	0.42

*$p < .05$. **$p < .01$. ***$p < .001$.

of the β-weight for *F*3 rests solely on the relation of *F*3 and vowel response on these four trials, because for all the other trials, behavior is sufficiently predicted by the visual vowel. Accordingly, this departure from the general trend of a coincidence of compensatory effects of vowel information on fricative identification and direct effects of vowel information on vowel identification can be disregarded, because it stems from a rather small empirical basis.

Discussion

The results show that the visual vowel influenced fricative responses, whereas the visual fricative failed to do so for all but 1 participant. The latter result is most likely due to the fact that the Dutch fricative contrast between [s] and [ʃ] is a place contrast only (alveolar vs. palatal) and is not enhanced by lip rounding, so that there is little visual evidence that distinguishes the two fricatives. However, the visually induced vowel identification shift also influences perception of the preceding fricative. This speaks against an auditory account of this context effect but is in line with a phonologically mediated effect. The individual results provide additional evidence for such a phonologically mediated effect. On an individual basis, the visual and acoustic vowel information failed to influence fricative identification if they also failed to influence vowel identification: Only those participants who used an acoustic or a visual cue for vowel identification also showed a compensatory influence of this cue on fricative identification. This also fits with the results from Experiment 1, in which the participant with the smaller influence of *F*3 on vowel identification showed less of a compensatory influence of *F*3 on fricative identification. It appears that any compensatory influence of the visual or acoustic cues to vowel identity on fricative identification arises only if these cues influence vowel identification.

An alternative explanation for the present result might be that the visual cues for [s] were more salient in the visual syllable [sy] than in [si], whereas the visual cues for [ʃ] were more salient in [ʃi] than in [ʃy] (see Holt et al., 2005). This would explain the higher likelihood of [s] responses with the visual vowel [y] than with the visual [i], without assuming a visual influence on compensation for coarticulation. In order to test the possibility of such a "hidden" McGurk effect, I restricted the analysis to the visual syllables [si] and [ʃy] and tested the influence of the visual vowel and the two acoustic variables FP and *F*3 frequency on the fricative reaction. In this analysis,

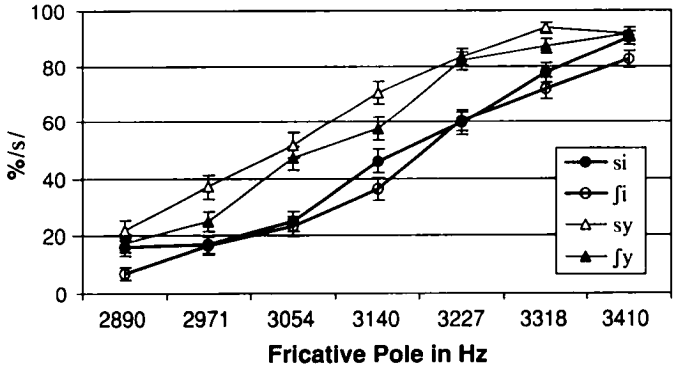

Figure 7. Mean percentages of [s] identifications in Experiment 3 depending on visual syllable (different functions) and fricative pole (FP) frequency (abscissa). Values are collapsed over all *F*3 frequencies.

the visual vowel is confounded with the visual fricative in the direction opposite to that for the compensatory effect of the vowel that induces /s/ reactions in [Cy] syllables. Despite this restriction and the fact that any McGurk effect should lead to more /ʃ/ reactions after [y], there is still a significant compensatory influence ($\beta = -0.64$, $p < .001$). The overall bias for more /s/ responses to the visual [ʃy] syllable than to the visual [si] syllable remains (58.2% vs. 47.4%), although it is understandably weaker, due to the influence of the visual fricative.

GENERAL DISCUSSION

In this article, mechanisms driving compensation for coarticulation in the perception of fricative–stop sequences were investigated. Previous results by Smits (2001a) had indicated that compensation was phonologically mediated, in that the perception of the fricative was dependent on the phonological categorization of a following vowel. Smits (2001a) based his conclusion on logistic regression analyses using both *F*3 and vowel identification as predictors. However, the colinearity of these predictors allowed alternative interpretations; so, in the experiments reported here, the acoustic parameters and vowel identification were varied independently.

Experiment 1 established two effects: a compensation effect (more /s/ responses for vowels with a lower *F*3) and a companion effect (more /i/ responses for fricatives with a higher FP). In Experiments 2A and 2B, possible auditory effects were investigated by using sine wave substitutes for both the fricative and the vowel and testing their effects on vowel and fricative identification, respectively. In Experiment 2A, the results indicated that there was an integrative auditory effect of the fricative on the vowel identification. This integrative effect is similar to the companion effect, that coarticulation is used to identify the segment causing the coarticulation. The present results indicated that general auditory processes may contribute to the companion effect, if one accepts the logic of the speech–nonspeech comparison (but see Fowler, 1990, 2006).

In contrast, a sine wave substitute for the vowel influenced fricative identification in a compensatory way in Experiment 2A. Interpretation of this effect was complicated by the fact that the listeners did not seem able to make valid judgments about fricative identity. In Experiment 2B, therefore, the fricative identification was made easier by introducing a preceding neutral vowel. With a preceding constant vowel, fricative identification improved, and the following sine wave ceased to have a compensatory effect on fricative identifications. Instead, a small integrative effect was again observed.

In Experiment 3, vowel identification was manipulated independently of the acoustic parameters of the vowel, using audiovisual displays. This manipulation clearly influenced the fricative identification as well. Despite the fact that there was only a very limited direct influence of the visual information on the fricative identification—only 1 out of 10 listeners showed evidence of a direct McGurk effect on fricative identification—there was a large indirect effect of the visual vowel on fricative identification. This indicates that these compensatory effects were not driven by the acoustic parameters of the speech sounds involved but were mediated by the perceived identity of the speech sound.

There are two different possible accounts for this phonological mediation. Smits (2001a) assumed that listeners would act as pattern classifiers and, hence, learn the codependency using a method of *hierarchical* categorization, where the categorization of segment x depends on the perception of segments $x \pm 1$. Theories of speech perception involving gestures as objects of perception provide an alternative interpretation for the present results. A speech perception module with reference to production mechanisms (Liberman, 1996; Liberman & Whalen, 2000) would recognize that an alveolar fricative gesture plus lip rounding for the vowel leads to FP frequencies that are otherwise observed for palatal fricatives. Hence, lower FPs would still be accepted as instances of an alveolar fricative in front of a rounded vowel. Similarly, a direct perception account (e.g., Fowler, 1996) would argue that the perceived lip rounding in the fricative is parsed from the fricative and attributed to the vowel, also leading to compensation for coarticulation. The present results do not allow one to decide between these different accounts. Smits (2001a), however, showed that the vowel context changes not only the location of the fricative boundary, but also its steepness. This result was predicted from his statistical-learning algorithm, because the difference in FP between the alveolar and the palatal fricatives was smaller in front of a rounded vowel. It is as yet unclear how this difference in boundary steepness could be accounted for in the framework of gestural theories.

One aspect of the present results requires further attention. Although the nonspeech sounds failed to trigger the contrastive, compensatory effect, they triggered an integrative effect that was similar to the companion finding: Fricatives with a high FP led to more /i/ responses than did fricatives with a low FP. Similarly, high sine wave substitutes for the fricative led to more /i/ responses. Similar integrative effects are also evident in the data of Fowler (1992), who investigated the possibility of durational contrast mechanisms, using nonspeech sounds only. Although previous efforts with nonspeech sounds have consistently shown contrastive effects (Blomert et al., 2004; Holt & Lotto, 2002; Lotto et al., 2003), it is not completely unexpected to find such effects of integration. Not only contrast effects are pervasive in perception (e.g., Warren, 1999); also, integrative effects are well documented, such as loudness integration (e.g., Florentine, Buus, & Poulsen, 1996) and the concept of a window of integration for generating "auditory objects" (Bregman, 1990; Yabe et al., 1998). Accordingly, both integration and contrast occur in auditory processing. Moreover, individual results for Experiment 2B revealed that integrative and contrastive effects can co-occur. This means that previous results showing the absence of contrastive effects of nonspeech sounds on other nonspeech sounds (Fowler, 1992; Fowler et al., 2000) do not necessarily discredit the assumption

that auditorily driven contrastive effects help to bring about compensation for coarticulation. Instead, the absence of contrastive effects may be due to a competition of contrastive and integrative mechanisms. Such an assumption would, however, inoculate auditory contrast effects against any attempt at falsification. Therefore, more research is necessary to indicate when and how integrative or contrastive effects occur.

The present data nevertheless indicate that compensation for coarticulatory lip rounding in fricative–vowel sequences is not based on general auditory processes. Instead, it seems that phonological mediation is pivotal. Does this conclusion apply to other compensation-for-coarticulation phenomena? As has already been noted, there has been an extended discussion about the case of liquid–stop sequences; and experiments similar to the present ones have previously been conducted both by Lotto and colleagues (Holt et al., 2005; Lotto & Kluender, 1998) and by Fowler and colleagues (Fowler, 2006; Fowler et al., 2000). Those results seem to contrast with the present results: Sine wave analogues are able to trigger compensation for coarticulation (Blomert et al., 2004; Lotto & Kluender, 1998), whereas visually influencing the perception of the liquid does not affect the perception of a following stop (Holt et al., 2005; Vroomen & de Gelder, 2001). Moreover, compensation for coarticulation in liquid–stop sequences is independent of language experience (Lotto et al., 1997; Mann, 1986). Hence, there is a large amount of evidence to suggest that compensation for coarticulation in liquid–stop sequences is a consequence of auditory processing.

The opposite results for liquid–stop sequences and for fricative–vowel sequences seem to indicate that one cannot argue for a single cause for compensation for coarticulation. Instead, the compensation mechanisms seem to depend on the type of coarticulation. This conclusion is buttressed by the findings that indicate different mechanisms of compensation for phonological assimilation and /t/ deletion (Mitterer & Blomert, 2003; Mitterer & Ernestus, 2006). In some cases, compensation arises in audition, but in other cases, compensation is phonologically mediated, possibly by covariate learning. It may, nevertheless, be possible to make a more specific statement about the causes of compensation. Two other context effects for which learning seems crucial are the cases of the perception of stop voicing influenced by the *f*0 in the vowel (Holt et al., 2001) and the trading relation of silence and *F*1 onset in the perception of [s] versus [st] onsets (Sinnott & Saporita, 2000). Effects seem to be based on audition for liquid–voiced-stop sequences, fricative–unvoiced-stop sequences, and even liquid assimilation in Hungarian. In the latter cases, there is a strong overlap in the acoustic parameters of the two speech sounds. However, for the case of stop voicing, VOT and *f*0 are acoustically quite dissimilar. For the present case, FP frequencies and *F*3 frequencies in the vowel are separated by up to one-half octave. In such cases, there is little room for auditory interactions. However, for liquid–stop sequences, identification of place in the liquid and in the stop depend heavily on *F*3s, which are in similar frequency areas. Hence, it may be cautiously assumed that general auditory processes may play a more important role when acoustic cues are similar and, hence, engage similar neuronal populations in the auditory cortex (Scott & Wise, 2004). Other dependencies may need to be learned or may be compensated for by reference to speech production.

REFERENCES

Akker, E., & Cutler, A. (2003). Prosodic cues to semantic structure in native and nonnative listening. *Bilingualism: Language & Cognition*, **6**, 81-96.

Beddor, P. S., Harnsberger, J. D., & Lindemann, S. (2002). Language-specific patterns of vowel-to-vowel coarticulation: Acoustic structures and their perceptual correlates. *Journal of Phonetics*, **30**, 591-627.

Beddor, P. S., & Krakow, R. A. (1999). Perception of coarticulatory nasalization by speakers of English and Thai: Evidence for partial compensation. *Journal of the Acoustical Society of America*, **106**, 2868-2887.

Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 167-200). Baltimore, MD: York Press.

Blomert, L., Mitterer, H., & Paffen, C. (2004). In search of the auditory, phonetic and/or phonological problems in dyslexia: Context effects in speech perception. *Journal of Speech, Language, & Hearing Research*, **47**, 1030-1047.

Boersma, P., & Weenink, D. (2004). *Praat 4.0*. [Computer software]. Amsterdam: University of Amsterdam, Institute of Phonetic Sciences.

Booij, G. (1995). *The phonology of Dutch*. Oxford: Oxford University Press, Clarendon Press.

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.

Coady, J. A., Kluender, K. R., & Rhode, W. S. (2003). Effects of contrast between onsets of speech and other complex spectra. *Journal of the Acoustical Society of America*, **114**, 2225-2235.

Darcy, I., Peperkamp, S., & Dupoux, E. (in press). Bilinguals play by the rules: Perceptual compensation for assimilation in late L2-learners. In J. Cole & J. Hualde (Eds.), *Papers in laboratory phonology 9: Change in phonology*. Berlin: Mouton de Gruyter.

Diehl, R., Lotto, A. J., & Holt, L. L. (2004). Speech perception. *Annual Review of Psychology*, **55**, 149-179.

Farnetani, E. (1997). Coarticulation and connected speech processes. In W. J. Hardcastle & J. Laver (Eds.), *The handbook of phonetic sciences* (pp. 371-404). Oxford: Blackwell.

Florentine, M., Buus, S., & Poulsen, T. (1996). Temporal integration of loudness as a function of level. *Journal of the Acoustical Society of America*, **99**, 1633-1644.

Fowler, C. A. (1990). Sound-producing sources as objects of perception: Rate normalization and nonspeech perception. *Journal of the Acoustical Society of America*, *89*, 2905-2909.

Fowler, C. A. (1992). Vowel duration and closure duration in voiced and unvoiced stops: There are no contrast effects here. *Journal of Phonetics*, **20**, 143-165.

Fowler, C. A. (1996). Listeners do hear sounds, not tongues. *Journal of the Acoustical Society of America*, **99**, 1730-1741.

Fowler, C. A. (2006). Compensation for coarticulation reflects gesture perception, not spectral contrast. *Perception & Psychophysics*, **68**, 161-177.

Fowler, C. A., & Brown, J. M. (2000). Perceptual parsing of acoustic consequences of velum lowering from information for vowels. *Perception & Psychophysics*, **62**, 21-32.

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 & Performance*, **26**, 877-888.

Fowler, C. A., & Dekle, D. J. (1991). Listening with eye and hand: Cross-modal contributions to speech perception. *Journal of Experimental Psychology: Human Perception & Performance*, **17**, 816-828.

FOWLER, C. A., & SMITH, M. R. (1986). Speech perception as "vector analysis": An approach to the problems of invariance and segmentation. In J. S. Perkell & D. H. Klatt (Eds.), *Invariance and variability in speech processes* (pp. 123-139). Hillsdale, NJ: Erlbaum.

GASKELL, M. G. (2003). Modelling regressive and progressive effects of assimilation in speech perception. *Journal of Phonetics*, **31**, 447-463.

GIBSON, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.

HOCKETT, C. F. (1955). *A manual of phonology*. Baltimore: Waverly.

HOLT, L. L. (2005). Temporally nonadjacent nonlinguistic sounds affect speech categorization. *Psychological Science*, **16**, 305-312.

HOLT, L. L., & LOTTO, A. J. (2002). Behavioral examinations of the level of auditory processing of speech context effects. *Hearing Research*, **167**, 156-169.

HOLT, L. L., LOTTO, A. J., & KLUENDER, K. R. (2000). Neighboring spectral content influences vowel identification. *Journal of the Acoustical Society of America*, **108**, 710-722.

HOLT, L. L., LOTTO, A. J., & KLUENDER, K. R. (2001). Influence of fundamental frequency on stop-consonant voicing perception: A case of learned covariation or auditory enhancement? *Journal of the Acoustical Society of America*, **109**, 764-774.

HOLT, L. L., STEPHENS, J. D. W., & LOTTO, A. J. (2005). A critical evaluation of visually moderated phonetic context effects. *Perception & Psychophysics*, **67**, 1102-1112.

JOHNSON, K., STRAND, E. A., & D'IMPERIO, M. (1999). Auditory–visual integration of talker gender in vowel perception. *Journal of Phonetics*, **27**, 359-384.

KLUENDER, K. R., COADY, J. A., & KIEFTE, M. (2001). Sensitivity to change in perception of speech. *Speech Communication*, **41**, 59-69.

LADEFOGED, P., & BROADBENT, D. E. (1957). Information conveyed by vowels. *Journal of the Acoustical Society of America*, **27**, 98-104.

LIBERMAN, A. M. (1996). *Speech: A special code*. Cambridge, MA: MIT Press.

LIBERMAN, A. M., DELATTRE, P. C., & COOPER, F. S. (1952). The role of selected stimulus variables in the perception of unvoiced stop consonants. *American Journal of Psychology*, **65**, 497-516.

LIBERMAN, A. M., & WHALEN, D. W. (2000). On the relation of speech to language. *Trends in Cognitive Sciences*, **4**, 187-196.

LINDBLOM, B. E., & STUDDERT-KENNEDY, M. (1967). On the role of formant transitions in vowel recognition. *Journal of the Acoustical Society of America*, **42**, 830-843.

LOTTO, A. J., & KLUENDER, K. R. (1998). General contrast effects in speech perception: Effect of preceding liquid on stop consonant identification. *Perception & Psychophysics*, **60**, 602-619.

LOTTO, A. J., KLUENDER, K. R., & HOLT, L. L. (1997). Perceptual compensation for coarticulation by Japanese quail (*Coturnix coturnix japonica*). *Journal of the Acoustical Society of America*, **102**, 1134-1140.

LOTTO, A. J., SULLIVAN, S. C., & HOLT, L. L. (2003). Central locus for nonspeech context effects on phonetic identification (L). *Journal of the Acoustical Society of America*, **113**, 53-56.

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

MANN, V. A. (1986). Distinguishing universal and language-dependent levels of speech perception: Evidence from Japanese listeners' perception of English "l" and "r." *Cognition*, **24**, 169-196.

MANN, V. A., & REPP, B. H. (1980). Influence of vocalic context on perception of the [sh]–[s] distinction. *Perception & Psychophysics*, **28**, 213-228.

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.

MCGURK, H., & MACDONALD, J. (1976). Hearing lips and seeing voices. *Nature*, **264**, 746-748.

MITTERER, H., & BLOMERT, L. (2003). Coping with phonological assimilation in speech perception: Evidence for early compensation. *Perception & Psychophysics*, **65**, 956-969.

MITTERER, H., CSÉPE, V., & BLOMERT, L. (2006). The role of perceptual integration in the recognition of assimilated words forms. *Quarterly Journal of Experimental Psychology*, **59**, 1305-1334.

MITTERER, H., CSÉPE, V., HONBOLYGO, F., & BLOMERT, L. (2006). The recognition of phonologically assimilated words does not depend on specific language experience. *Cognitive Science*, **30**, 451-479.

MITTERER, H., & ERNESTUS, M. (2006). Listeners recover /t/s that speakers reduce: Evidence from /t/-lenition in Dutch. *Journal of Phonetics*, **34**, 73-103.

NEAREY, T. D. (1989). Static, dynamic, and relational properties in vowel perception. *Journal of the Acoustical Society of America*, **85**, 2088-2113.

REMEZ, R. E., RUBIN, P. E., BERNS, S. M., PARDO, J. S., & LANG, J. M. (1994). On the perceptual organization of speech. *Psychological Review*, **101**, 129-156.

SCOTT, S. K., & WISE, R. J. S. (2004). The functional neuroanatomy of prelexical processing in speech perception. *Cognition*, **92**, 13-45.

SINNOTT, J. M., & SAPORITA, T. A. (2000). Differences in American English, Spanish, and monkey perception of the *say–stay* trading relation. *Perception & Psychophysics*, **62**, 1312-1319.

SMITS, R. (2001a). Evidence for hierarchical categorization of coarticulated phonemes. *Journal of Experimental Psychology: Human Perception & Performance*, **27**, 1145-1162.

SMITS, R. (2001b). Hierarchical categorization of coarticulated phonemes: A theoretical analysis. *Perception & Psychophysics*, **63**, 1109-1139.

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

VROOMEN, J., & DE GELDER, B. (2001). Lipreading and the compensation for coarticulation mechanism. *Language & Cognitive Processes*, **16**, 661-672.

WARREN, R. M. (1999). *Auditory perception: A new analysis and synthesis*. Cambridge: Cambridge University Press.

WEBER, A., & CUTLER, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory & Language*, **50**, 1-25.

WHALEN, D. H. (1990). Coarticulation is largely planned. *Journal of Phonetics*, **18**, 3-35.

YABE, H., TERVANIEMI, M., SINKKONEN, J., HUOTILAINEN, M., ILMONIEMI, R. J., & NÄÄTÄNEN, R. (1998). Temporal window of integration of auditory information in the human brain. *Psychophysiology*, **35**, 615-619.

NOTES

1. Dutch [i] and [y] differ in other formants (mostly *F*2) as well. Smits (2001a) manipulated *F*3 only in order to simplify the data analysis. For the sake of consistency, I followed his example in my experiments.

2. Throughout this article, the term *high* will have an acoustic, and not a phonological, meaning. Although both [i] and [y] vowels are phonologically high, they differ in the height of the third formant, which is higher for [i].

3. This measure was taken to prevent any masking artifact. The proponents of gestural accounts (Fowler et al., 2000) have shown that the effects of nonspeech sounds on speech perception may be caused by masking, which would not occur in speech perception. Given the forward direction of the masking, the masking should mainly occur peripherally (see Warren, 1999, p. 63). As a reaction, the advocates of auditory effects have shown that such effects persist even if the stimulus arrangement prevents peripheral masking and, hence, has to be attributed to more central auditory mechanisms (Holt, 2005; Holt & Lotto, 2002). An emphasis on central mechanisms may, however, unduly minimize a possible function that peripheral mechanisms may have for speech perception. If nonspeech stimuli mimic the frequency-specific amplitude relations of the speech stimuli—as they also did in the second experiment of Lotto and Kluender (1998)—any peripheral masking could not be labeled an artifact, because similar masking effects would occur in speech. Therefore, I purposefully allowed peripheral masking to occur by presenting both fricatives and vowels and their sine wave substitutes binaurally.

4. It cannot be ruled out that the short training session, which focused on the fricatives in Experiment 2B but on the whole syllables in Experiment 2A, contributed to the difference in results as well. However, Experiment 2B was designed to improve fricative identification, and in that it succeeded, be it through the different short preexperimental training or the added vowel.

(Manuscript received October 1, 2004;
revision accepted for publication December 8, 2005.)