

Coping with phonological assimilation in speech perception: Evidence for early compensation

HOLGER MITTERER and LEO BLOMERT

Universiteit Maastricht, Maastricht, The Netherlands

The pronunciation of the same word may vary considerably as a consequence of its context. The Dutch word *tuin* (English, *garden*) may be pronounced *tuim* if followed by *bank* (English, *bench*), but not if followed by *stoel* (English, *chair*). In a series of four experiments, we examined how Dutch listeners cope with this context sensitivity in their native language. A first word identification experiment showed that the perception of a word-final nasal depends on the subsequent context. Viable assimilations, but not unviable assimilations, were often confused perceptually with canonical word forms in a word identification task. Two control experiments ruled out the possibility that this effect was caused by perceptual masking or was influenced by lexical top-down effects. A passive-listening study in which electrophysiological measurements were used showed that only unviable, but not viable, phonological changes elicited a significant mismatch negativity. The results indicate that phonological assimilations are dealt with by an early prelexical mechanism.

One of the biggest problems for human and machine speech recognizers is to deal with the enormous amount of variation between different instances of the same spoken word. Even a single speaker may pronounce the same word quite differently depending on the phonetic context. In Dutch, for instance, the word *tuin* (*garden*) may be pronounced [tœym] if the segment following the intended /n/ has a labial place of articulation, such as /b/ in *tui(n/m)bank* (*garden bench*; see Booij, 1995). This process is called *nasal place assimilation*. The matter is further complicated by the fact that this assimilation process is not obligatory and occurs in approximately 60% of the cases (Van Heuven & Van Berg, 1982). This leads to the problem that, on the one hand, listeners must be able to map the different forms [tœyn] and [tœym] onto the same meaning. On the other hand, listeners have to make a distinction between other forms that also vary only in the place of articulation of a final nasal but differ in meaning. Take, for instance, the Dutch minimal pair *duin* (*dune*) and *duim* (*thumb*), which differ only in the place of articulation of a final nasal—precisely the feature that is changed by assimilation.

Early psycholinguistic research paid little attention to this aspect of the invariance problem (see Huckvale, 1999). It was assumed that word recognition processes might treat such variation as noise (Norris, 1982). Such an assumption was fostered by the finding that listeners

often fail to notice small mispronunciations (Cole, 1973). Furthermore, variation may be corrected by top-down influences (McClelland & Elman, 1986). In a recent review of these matters, however, Norris, McQueen, and Cutler (2000) argued that the spoken word recognition system is quite sensitive to acoustic variation (see, e.g., Marslen-Wilson, Moss, & van Halen, 1996). In addition, Norris et al. argued that top-down influences are neither necessary for explaining existing word recognition data nor beneficial for more efficient recognition. Although the existence of top-down influences is still a matter of debate, results indicate that possible top-down influences are rather small (Samuel, 1996). The sensitivity of the word recognition system is nicely illustrated by the fact that even subphonetic variation can hinder lexical access. Utman, Blumstein, and Burton (2000) altered words with initial unvoiced plosives by reducing the voice onset time (VOT). These alterations were subphonetic, because these words were still identified as starting with an unvoiced plosive. However, in an identity-priming paradigm, these altered words produced less priming than did words with a canonical VOT.

This shows that the word recognition system does not generally tolerate variation. Therefore, phonological assimilations pose a problem for the spoken word recognition system. Due to the sensitivity of the word recognition system, the acoustic consequences of assimilation will be picked up and cannot be ignored, because this would blur the distinction between other words (*duin* [*dune*] and *duim* [*thumb*]).

A Regressive-Inference Account

Place assimilations do not occur at random but are constrained by the following phonological context. If a word-final /m/ is followed by a nonlabial segment, the

The authors thank Paul Boersma for making his software package Praat available and providing support so swiftly, as well as two anonymous reviewers for their insightful comments on an earlier version of this article. Correspondence concerning this article should be addressed to H. Mitterer, Max-Planck-Institut für Psycholinguistik, Postbus 310, 6500 AH Nijmegen, The Netherlands.

/m/ has to be ascribed to the underlying word form. If a word-final /m/ is, however, followed by a labial segment (/b/, /p/, or /m/), the pronounced /m/ might correspond to an underlying /n/ that has been assimilated. Given this rule-like behavior, assimilations may be compensated for in perception by a mechanism that uses the phonological context to regressively infer the underlying place of articulation of a word-final nasal. Henceforth, we will use the expression *perceptual compensation for phonological assimilation* to refer to a mechanism that helps phonologically assimilated utterances activate a lexical entry. That is, such a mechanism makes it possible for [tœym] to activate *tuin*. Gaskell, Hare, and Marslen-Wilson (1995) have introduced a computational model for such a mechanism. The input to this model comprises traditional phonological features. At the input level, these phonological features are perceived as articulated. This implies that the consequences of assimilation pass unaltered through all acoustic and phonetic processing stages. That is, the utterance *lean bacon* is perceived as containing an /m/ up to a phonological level. Only then does a hidden layer connected to a single recurrent network influence the activation of phonological feature nodes and correct for assimilations. The activations of coronal segments are not corrected by this structure (*coronal* refers to the phonological feature corresponding to an alveolar place of articulation), reflecting the fact that coronal segments do not occur as the product of regressive English place assimilations (see Collins & Mees, 1996). This is due to the asymmetric nature of place assimilation. Although an intended /n/ can be pronounced as /m/, an intended /m/ can never be pronounced as /n/. Therefore, a pronounced /n/ has to correspond to an intended /n/. If, however, a pronounced /m/ is followed by another labial segment, this segment activates the unit corresponding to a coronal place of articulation. This form of phonological inference reflects the fact that the utterance *lean_bacon* is a possible assimilated pronunciation of *lean bacon*. Thus, the mechanism proposed by Gaskell et al. compensates for place assimilation by a process of regressive inference. That is, the interpretation of segment *n* is influenced by segment *n + 1*.

Gaskell and Marslen-Wilson (1996) have provided experimental evidence for this regressive-inference account. They used assimilated tokens such as *lean* pronounced with a final /m/ in a cross-modal lexical decision task. While listening to sentences, participants performed a lexical decision to visually presented words. If the visual target was preceded by the same word in the auditory stream, reactions were faster, a so-called identity-priming effect. In the second experiment, changed tokens (*lean* → *leam*) were presented in two conditions. In one condition, the change of the final nasal was viable (*lean bacon*), whereas in the other condition, the change was unviable (*leam gammon*). If the change was viable, both changed and unchanged tokens led to a comparable priming effect. In the unviable-context condition, however, the changed tokens (*leam gammon*) showed signifi-

cantly reduced priming effects, as compared with the unchanged tokens. This speaks for the existence of a phonological inference mechanism that infers the intended *deep form* from the perceived *surface* utterance. However, these data do not allow one to pinpoint the locus of these effects, because priming in a lexical decision task is a measure that leaves open many possibilities for interpretation, including strategic effects (see Neely, 1991).

The strongest evidence for a regressive-inference account stems from a phoneme-monitoring study by Gaskell and Marslen-Wilson (1998). They used phrases containing phonologically viable and unviable changes. While listening to these phrases, participants performed a phoneme-monitoring task, monitoring for a segment with an alveolar place of articulation. The participants were more likely to (mis)detect a target with an alveolar place of articulation when the context allowed for an assimilation of the segment in question. This effect was present in words and in nonwords. That is, the participants were more likely to report hearing a /t/ in the phrase *prayp bearer* than in the phrase *prayp carrier*. In the phrase *prayp bearer*, the final /p/ in the nonword *prayp* could have a deep form with a final /t/. (A /t/ can be assimilated and become a /p/ if the following segment is a /b/.) This effect was moderated by the lexical status of the carrier words. Misdetections of coronal segments, indicative of regressive phonological inference, were more likely in word targets than in nonword targets. This led Gaskell and Marslen-Wilson (1998) to extend the model of Gaskell et al. (1995) by including lexical influences on phonological inference. In the earlier model, phonological inference was assumed to be purely prelexical. The presence of an effect in nonword targets, however, points in the direction of a prelexical nature of the phonological inference mechanism.

This interpretation needs further consideration. Warren (1971, 1999, 2000; Warren, Bashford, & Gardner, 1990) has argued that “listeners do not, and indeed cannot, perceive phonemes in running speech directly, but that their presence is inferred following prior identification of larger units” (Warren, 1999, p. 172). That is, in this view, phoneme detection in a phoneme-monitoring task occurs after, and not prior to, phonotactic and lexical processing of the stimulus. Indeed, event-related potentials (ERPs) show that lexical information is available before phoneme monitoring can be performed (Thierry, Doyon, & Démonet, 1998). Therefore, further clarification is needed as to how far the lexical influences on the perceptual compensation for phonological assimilation were genuine or task-related in the study of Gaskell and Marslen-Wilson (1998).

The Present Experiments

The evidence in favor of a regressive-inference account is based mainly on studies employing cross-modal priming and phoneme-monitoring tasks. Gow (2001) argued that the pattern of results in favor of a regressive-

inference mechanism is due to a disruption of lexical processing: On hearing *lean*, participants may activate the lexical candidate *lean* and expect a subsequent word to start with a labial. When this expectation is not fulfilled, lexical processing is disrupted, with increasing reaction times as a consequence.

To avoid the possibility of such alternative interpretations, we used a two-alternative forced-choice (2AFC) task with feedback. In this task, participants had to decide whether a Dutch word (*tuin* [*garden*]) was pronounced canonically or with a word-final /m/. It may be argued that this 2AFC task is less open to alternative interpretations, such as *disruption of lexical processing* (Gow, 2001). Although the 2AFC task does not exclude the possibility of lexical influences (see, e.g., Ganong, 1980), it may be argued that neither phoneme awareness nor lexical processing is necessary for one to perform the task. This can be derived from the fact that the 2AFC is easily adaptable for nonliterate preschool children (e.g., Nittrouer, 1996) and even nonhuman species (e.g., Lotto, Kluender, & Holt, 1997; Schulze & Scheich, 1999; Sinnott & Saporita, 2000).

In order to investigate the perception of assimilation tokens with a 2AFC task, words were presented in isolation or in a phrase that did or did not allow assimilation. According to a regressive-inference account, the pronunciation change (*tuin* → *tuim*) should be easy to detect in isolation but difficult to detect in a viable context. This hypothesis was investigated in Experiment 1. In Experiments 2 and 3, a 2AFC task was also used in order to test alternative interpretations of the results of Experiment 1. Experiment 2 tested whether masking contributes to the perceptual compensation for place assimilation. Experiment 3 tested whether lexical processes influence the perception of place assimilation.

Finally, we tested whether evidence for perceptual compensation for phonological assimilation can be found when participants are only listening passively to viable and unviably changed phrases. In Experiment 4, we therefore used a paradigm that did not involve an active decision by the participant: the passive-oddball paradigm. In this paradigm, participants hear a train of stimuli that consists of a standard stimulus (e.g., 85%) and a deviant stimulus (e.g., 15%). The participants do not have to react to the stimuli and are either reading a book or watching a silent videotape. When the electroencephalogram (EEG) is measured and averaged—time-locked to the stimulus onset—to obtain ERPs, the ERP to the deviant stimulus is more negative than the ERP to the standard stimulus, starting around 100 msec after stimulus onset. This negativity is called *mismatch negativity* (MMN; see, e.g., Näätänen, 1992, 1995; Schröger, 1998). MMN is interesting for the present purposes, because MMN seems to reflect the perceptual, and not the physical, difference between the standard–deviant stimulus pair (Näätänen & Winkler, 1999). Näätänen, Schröger, Karakas, Tervaniemi, and Paavilainen (1993) have shown that a stimulus pair consisting of two complex tone stim-

uli induces an MMN only when participants are able to discriminate between the two stimuli with a high level of accuracy. In this study, a group of participants learned to discriminate between two stimuli over the course of the experiment. These participants did not show an MMN in the first block of passive listening but did so in the last block, after their discrimination performance had increased. This shows that MMN does not reflect the physical but, rather, the perceptual difference between the standard and the deviant. Along similar lines, Näätänen et al. (1997) obtained a comparable result with vowels as stimuli. They measured the MMN to two vowel stimuli in Finnish and Estonian participants. The main result of this study showed that the MMN was enlarged if both the standard and the deviant were vowel prototypes in the native language of the listener. Again, the same physical difference between a standard and a deviant gave rise to different MMNs, depending on the native language of the listener (see also Winkler et al., 1999).

These findings show that the MMN is an interesting tool for our purposes, because it is sensitive to the perceptual distance between the members of a stimulus pair. If there is a regressive compensation mechanism that compensates for phonological variation, this should reduce the perceptual distance between a canonical and a changed utterance. That is, the perceptual distance between *tuinbank* (*garden bench*) and *tuimbank* should be smaller than the distance between *tuinstoel* (*garden chair*) and *tuimstoel*. In the first example, the change of the /n/ to /m/ is viable according to the assimilation rule. In the second example, however, the change is not viable. Accordingly, the MMN should be smaller to a *tuinbank*–*tuimbank* stimulus pair (viable condition) than to a *tuinstoel*–*tuimstoel* stimulus pair (unviable condition). This was the hypothesis that we set out to test in Experiment 4.

EXPERIMENT 1

In the first experiment, we used edited natural speech to investigate how participants would handle phonological assimilation in the 2AFC task. The participants had to indicate whether the Dutch word *tuin* (*garden*) was pronounced canonically or with a changed place of articulation for the final nasal—that is, *tuim*. This was tested by presenting the target words in isolation or in a context that did (. . . *bank*) or did not (. . . *stoel*) allow for the assimilation of a word-final nasal.

The fact that the stimulus materials for the present experiment were also to be used in a passive-oddball paradigm later on put certain constraints on the stimulus material. An MMN may be observed not only if two stimuli differ in their phonological content, but also if they differ in pitch, intensity, or length. This implies that one cannot use arbitrary utterances with and without a change of the final phoneme and use them in a passive-oddball paradigm. As a result, one might find the MMN to be smaller in the viable condition, not because of phonological content, but because the two viable utterances

(*tuinbank*–*tuimbank*) differed less in pitch contour, length, or intensity than did the unviable stimulus pair (*tuinstoel*–*tuimstoel*). To avoid this potential confound, we used cross-splicing to generate the stimulus materials. Thus, this first experiment also served to test whether cross-spliced materials would lead to a stable perception of the place of articulation of a final nasal.

Method

Participants

Eight students of Maastricht University participated in the experiment. Six were female, 2 were male. None reported any history of hearing loss.

Design

A two-factorial design was constructed, with the first factor, nasal murmur, measured at two levels (/n/ and /m/) and the second factor, context, measured at three levels: no context; a viable context (*bank*), in which assimilation (*tuimbank*) could occur in natural speech; and an unviable context (*stoel*), in which assimilation could not occur. The three context conditions were blocked. The no-context condition ([tœyn] *tuin* and [tœym] *tuim*) was always presented first. The order of presentation of the two subsequent context blocks (viable-context block and unviable-context block) was counter-balanced between subjects. Each block consisted of 40 trials—that is, 20 repetitions of the two stimuli presented in a random order. In order to focus the participants' attention on the phonetic details of the stimuli, the participants received explicit feedback after every trial.

Stimulus Materials and Apparatus

A male native speaker of Dutch was selected, because of his low f_0 (75–80 Hz), which did not vary greatly between utterances. The sonority of the speaker's voice contributed to the naturalness of the cross-spliced stimuli. Multiple pronunciations of *tuinbank* [tœynbɑŋk], *tuimbank* [tœymbɑŋk], *tuinstoel* [tœynstuːl], and *tuimstoel* [tœymstuːl] were digitally recorded with a sampling rate of 44.1 kHz. Utterances were digitally bandpass filtered (40–5200 Hz) and resampled at 11025 Hz. Experimental stimuli were created by cross-splicing. First, [tœyn] was spliced from an utterance of [tœynstuːl]. From two other utterances of [tœynstuːl] and [tœynbɑŋk], the nasal murmurs were spliced. The nasal murmurs were equated in length and energy (using a root mean square [RMS] measure). Due to the constraint of splicing at zero-crossings, the nasal murmurs differed slightly in length. However, this was less than 1 msec. The edited nasal murmurs were concatenated with [tœy], resulting in the experimental stimuli [tœyn] and [tœym] used in the isolation condition.

The context words *stoel* [stuːl] and *bank* [bɑŋk] were each spliced from two other utterances. The [bɑŋk] utterance was produced without prevoicing. To make the utterance sound natural, a

25-msec period of silence was used to emulate a closure. By cutting 5 msec of the friction noise of the /s/ in [stuːl], the length of this sound was made equivalent to the length of the [bɑŋk] sound, including the silent period (see Table 1 for a detailed description of the nasal murmurs and the context sounds).

The stimuli were played to the participants by a DOS-operated computer using the software package ERTS (Behringer, 1996). The output of the SoundBlaster soundcard was redirected to four loudspeakers (JBL Control 25) within a sound-attenuated booth with a Spirit Folio mixer and was amplified by a Yamaha P4050 power amplifier. The maximal sound pressure level was 70 dB(A).

Procedure

The participants entered the sound-attenuated booth and were seated facing a monitor and an array of buttons. Two of the buttons were labeled *tuin* and *tuim*. The participants were instructed to press one of two buttons upon hearing a target. Instructions stressed accuracy, without special emphasis on speed of response. After each trial, the participants received visual feedback in the form of the words *right* and *wrong* on the monitor, indicating whether their identification had been correct.

Results

For each participant, the percentage of correct identifications was calculated for each condition. The mean percentages of correct identifications may be found in Table 2. The descriptive data show that performance was near ceiling in the no-context and the unviable-context conditions. In the viable-context condition, however, performance levels dropped.

A two-factor repeated measures analysis of variance (ANOVA) on the percentages of correct identifications, with the factors of nasal murmur (/n/ or /m/) and context (no context, unviable context [*stoel*], viable context [*bank*]), revealed a significant influence of context [$F(2,14) = 38.49, p < .001$]. The nasal murmur factor was close to significance [$F(1,7) = 4.79, p = .065$], and the interaction between the two factors was significant [$F(2,14) = 8.19, p < .01$]. Given the significant interaction, we examined the effect of context for each of the nasal murmurs separately. For the alveolar nasal murmur, context produced only a trend [$F(2,14) = 3.1, p < .1$]. For the labial nasal murmur, context influenced percentages of correct responses significantly [$F(1,14) = 56.8, p < .001$]. Post hoc Newman–Keuls tests ($p < .05$) revealed that performance was worse in the viable-context condition (69.3% correct) than in the two other conditions (98.8% correct for the no-context condition and 96.3%

Table 1
Acoustic Properties of the Stimuli Used in the Experiments

Stimulus	Formant 1			Formant 2			Formant 3			Formant 4		
	C	B	A	C	B	A	C	B	A	C	B	A
/n/	360	180	34.7	1500	470	25.3	2170	40	17.1	3460	180	18.7
/m/	357	160	33.9	1308	420	21.4	1960	100	20.5	3010	355	9.3
/s/	1870	520	4.9	2550	285	5.7	3095	270	15.0	3892	290	11.1
/b/	593	77	41.1	1130	175	30.4	1995	60	20.2	3379	30	18.4

Note—Values for the nasal murmurs and the context sounds were derived from formant estimations provided by Praat (Boersma & Weenink, 1999). Number of assumed formants were adjusted manually to achieve a stable solution for each speech sound. C, center frequency in hertz; B, bandwidth in hertz; A, relative amplitude in decibels.

Table 2
Percentages of Correct Identifications
(With Standard Deviations) in Experiment 1

Stimulus	Context					
	No Context		Unviable (<i>stoel</i>)		Viable (<i>bank</i>)	
	%	SD	%	SD	%	SD
<i>tuin</i>	96.3	5.8	96.9	2.6	87.5	15.8
<i>tuim</i>	98.8	2.3	96.3	5.1	69.3	11.5

correct for the unviable-context condition). The latter conditions did not differ significantly. In sum, the interaction indicates that the participants had an overall tendency to (mis)perceive the changed stimulus in the viable-change condition (*tuimbank*) as a canonical pronunciation (*tuinbank*).

Discussion

The first point to be stressed is that our stimulus generation method was justified: Cross-splicing the nasal murmur led to a stable perception of the place of articulation. In the no-context condition, performance was near ceiling. This result is worth noting, because it is not self-evident whether cross-splicing nasal murmurs is sufficient for creating a stable percept of the place of articulation. Usually, formant transitions within the vowel part of the signal contribute to the perception of place of articulation in nasals. However, Repp and Svastikula (1988) found that nasal murmur is a more important cue for postvocalic than for prevocalic nasals. In addition, we purposefully chose the word for this study in order to minimize the formant transition cue. Formant transition cues are expected to be weaker in cases of high vowels and in cases of diphthongs. Formant transitions into nasal murmurs are more salient for low vowels, because this affords more movement of the articulators to achieve the closure necessary for the nasal. On the basis of, for instance, the window model of speech-sound production (Keating, 1990), formant transitions in diphthongs are constrained in order to convey the diphthongal nature of the sound and leave less room for coarticulatory anticipation of the following segment's place of articulation. Given that the to-be-identified nasal in this experiment was postvocalic and was preceded by a diphthong with a high target, it is therefore not completely unexpected that the nasal murmur turned out to be a sufficient cue for a stable perception of place of articulation.

The stable perception of place of articulation also did not suffer from the introduction of an arbitrary context. Performance was equivalent and close to ceiling level in both the no-context and the unviable-context conditions. However, in the viable-context condition, the participants not only made more errors, but also showed a significant tendency to (mis)perceive the *tuimbank* stimulus as *tuinbank*. This result supports the assumption that phonological variation is compensated for regressively: An /m/ followed by a stimulus with a labial place of ar-

ticulation is interpreted as an /n/. As a consequence, the *tuimbank* stimulus is perceived as *tuinbank*. These results are consistent with the assumption of a regressive compensation mechanism dealing with place assimilation. It is difficult to see how strategic higher level processes could influence the present results obtained with a comparably simple task (as Gow, 2001, assumed for the data of Gaskell & Marslen-Wilson, 1998).

One might argue that the present result is compromised by the fact that only two stimuli were used. However, the present result extends earlier similar findings (Gaskell & Marslen-Wilson, 1996, 1998) for which a large number of stimuli were employed. Therefore, it seems unlikely that the context sensitivity evidenced in the present experiment was due to reasons different from those for the context sensitivity in these earlier studies. However, it seems worthwhile to consider two alternative interpretations of the present results. First, amplitude differences between the context words might have caused the results. The /b/ in the viable context (*bank*) has a larger amplitude than the /s/ in the unviable context (*stoel*). The results might, thus, not have been due to the viability of the assimilation in the different contexts but might have been caused by backward masking or unspecific distraction caused by the stimulus with the larger amplitude. Second, the results might have been caused by lexical, rather than phonological, inference. Both the *tuinbank* and the *tuimbank* utterances are legal pronunciations of an existing word. If lexical knowledge about the possible pronunciations of *tuinbank* (*garden bench*) influences identification performance, *tuimbank* stimuli might trigger *tuinbank* responses. To investigate these alternative interpretations (masking and lexical influence), we conducted another two identification experiments in which the same stimuli were used.

EXPERIMENT 2

The purpose of this experiment was to test whether the results obtained in Experiment 1 could be attributed to a perceptual compensation mechanism for place assimilations or were due to masking. Masking could be responsible for the results of Experiment 1, because the context words differed in the amplitude of the first segment. The stop in the viable-context condition was higher in amplitude than the fricative in the unviable-context condition (see Table 1). The poor performance in the viable-context condition in Experiment 1 might have been caused by the large amplitude of the context word's onset. If masking had been responsible for the results in the first experiment, the participants should be less accurate when the *tuin* and *tuim* stimuli were followed by a context consisting of large-amplitude band noises.

To test this possibility, we created three band noises that had the same RMS as did the 25 msec of maximal amplitude of the *bank* (*bench*) stimulus. A drop in performance might occur as a consequence of a frequency-specific

masking or a general distraction effect. In the case of a frequency-specific effect, it is conceivable that the low second formant of /b/ might be responsible for a masking effect. If energy in the nasal of the target is reduced in perception because of masking effects, participants might hear an /n/ instead of an /m/ only if the /m/ were followed by a low-frequency masker. On the other hand, it might also be possible that the result of Experiment 1 was due to an unspecific distraction by any large-amplitude context sound. In this case, any band noise should lead to a decrease in accuracy in identifying the nasal murmur in *tuin* and *tuim*.

Method

Participants

Eight students of Maastricht University participated in the experiment. None reported any history of hearing loss: none had participated in Experiment 1. They were paid for their participation.

Design

A two-factorial design was constructed, with the first factor, nasal murmur, measured at two levels (/n/ and /m/) and the second factor, context, measured at five levels: no context, the viable context from the previous experiment (*bank*), and three band noises with energy in labial, alveolar, and s-frication frequency areas. All the conditions were blocked.

Stimulus Materials and Apparatus

The speech stimuli and the apparatus were the same as those in the previous experiment. Three band noises were generated with the help of the software package Praat 3.9 (Boersma & Weenink, 1999). A white noise 100 msec in duration was generated. This noise was used as a template and filtered with a Hanning-band window (smoothing 100 Hz) to generate three band noises. A labial noise was generated with energy in the area of 0.5–1.5 kHz (corresponding roughly to the distinctive second formant of /b/). An alveolar noise was generated with energy in the area of 1.5–2.5 kHz. An s-frication noise was generated with energy in the area of 2.5–3.5 kHz. (The spectrum of the friction in the *stoel* context revealed a maximum at 3 kHz.) After filtering, the band noises were equated in RMS with the RMS of the maximum-amplitude area of the /b/ in the *bank* context. Band noises were concatenated with the target words with a 25-msec silent interval, mimicking the presentation of the *bank* context.

Procedure

The procedure was similar to that in Experiment 1. All the participants first completed the no-context condition. Then half of the participants heard the stimuli in the *bank* context and continued with the s-frication noise, the alveolar noise, and the labial noise contexts. The other half of the participants heard the stimuli in the order of labial noise context, s-frication noise context, alveolar noise

context, and, finally, the *bank* context. These two orders cross-balanced the position of the two probably most difficult conditions (the *bank* context and the labial noise context).

Results

The percentages of correct identifications are summarized in Table 3. As in Experiment 1, performance was worst in the *bank*-context condition. In addition, there was no visible influence of the band noises on the percentages of correct identifications. These observations were borne out by a repeated measures ANOVA with nasal murmur (*tuin* or *tuim*) and context (none, labial noise, alveolar noise, friction noise, and *bank*) as factors. The context factor did reach significance [$F(4,28) = 17.9, p < .001$]. The nasal murmur factor was not significant [$F(1,7) = 3.6, p < .1$].

However, the interaction of the two factors was significant [$F(4,28) = 36.1, p < .001$]. To investigate the nature of the interaction, we examined the effect of context on each level of the nasal murmur factor separately. For the alveolar nasal murmur, context influenced performance significantly [$F(4,28) = 3.8, p < .05$]. Post hoc Newman–Keuls tests ($p < .05$) revealed that performance in the viable-context condition (85.6% correct) was worse than that in all the other conditions (range, 93.1%–95.5% correct). All other comparisons failed to reach significance. For the labial nasal murmur also, context influenced performance [$F(4,28) = 30.6, p < .001$]. Post hoc Newman–Keuls tests ($p < .05$) revealed that performance was worse in the viable-context condition (63.6% correct) than in all the other conditions (range, 94.1%–95.6% correct). The latter conditions did not differ significantly. For both nasal murmurs, performance was worse in the viable-context condition than in the other conditions. We then tested whether the context effect of the viable-context condition was larger for the labial nasal murmur than for the alveolar nasal murmur. To this end, we computed a *viable-context effect* by subtracting the percentage of correct responses in the viable-context condition from the mean percentage of correct responses in all the other conditions. The decrease in the percentage of correct responses in the viable-context condition, in comparison with all the other conditions, was larger [$t(7) = 10.2, p < .001$] for the labial nasal murmur (31.9%) than for the alveolar nasal murmur (8.6%). This reflects the fact that the participants made most of the errors in response to the changed stimulus in the vi-

Table 3
Percentages of Correct Identifications (With Standard Deviations) in Experiment 2

Stimulus	Context									
	None		Labial Noise		Alveolar Noise		S-frication Noise		<i>bank</i>	
	%	SD	%	SD	%	SD	%	SD	%	SD
<i>tuin</i>	93.8	9.0	95.5	7.9	93.1	14.9	94.2	6.4	85.6	12.0
<i>tuim</i>	95.6	6.8	95.6	5.5	94.0	8.9	96.4	5.5	63.6	16.5

able condition—that is, they were inclined to hear *tuimbank* as *tuinbank*.

Discussion

These results revealed that the results obtained in Experiment 1 were not caused by amplitude differences between the viable- and the unviable-context sounds. None of the three band noises led to a decrease in accuracy. Neither unspecific distraction by large-amplitude sounds nor frequency-specific masking was responsible for the results obtained in Experiment 1. It may therefore be concluded that the discrimination between final nasals with different places of articulation is not impaired by large-amplitude context sounds presented immediately after the target. The interpretation of the results of both experiments in terms of a perceptual compensation mechanism seems valid. Furthermore, the drop in accuracy in the viable-context condition, as well as the tendency to (mis)perceive the *tuimbank* stimulus as *tuinbank*, also replicated the results of Experiment 1. However, in the present experiment, there also was a significant drop in the percentages of correct responses for the alveolar nasal murmur. A similar trend was observed in Experiment 1. We refrain from giving an interpretation for this effect until the discussion of Experiment 4, when additional evidence will be available. Nevertheless, Experiment 2 showed that the results of Experiment 1 were not due to masking. In the next experiment, we investigated possible lexical contributions to the regressive context effect observed in the two previous experiments.

EXPERIMENT 3

In Experiment 1, as well as in Experiment 2, the participants tended to (mis)perceive the *tuimbank* stimulus as *tuinbank*. The purpose of this third experiment was to investigate the possibility that lexical inference caused those misperceptions. If the lexical information that *tuimbank* is a legal variation of *tuinbank* has become available to the listener, this might trigger a *tuinbank* reaction in response to a *tuimbank* stimulus. One possible strategy for ruling out this possibility would be to use nonwords (see Gaskell & Marslen-Wilson, 1998). We took, however, a different road and replicated Experiment 1 with German participants who had knowledge of Dutch. The reason for doing so was twofold. First, this allowed a direct comparison with Experiment 1 by using identical stimuli. Second, it has been argued that nonwords might access the lexicon by means of analogy with similar words (Marslen-Wilson & Warren, 1994). It was highly unlikely that German listeners would be able to perceive the Dutch stimulus words by analogy with similar words. The German word for *garden* [gartən] is completely dissimilar to the Dutch word [tœyn]. Moreover, the diphthong in *tuin* (/œy/) does not exist in the German language. An analogy with any German word would, therefore, be difficult. Even if one assumes that German listeners would first assimilate [tœyn] to

their native phonological categories (following Best's, 1994, perceptual assimilation model), they would arrive at [dɔɪn] or [tɔɪn], both being nonwords in German. Consulting a lexical database (CELEX; Baayen, Piepenbrock, & Gulikers 1995) revealed that the nearest neighbor was [nɔɪn] (English, *nine*), which is acoustically quite different. Thus, the use of the same stimuli as those in Experiment 1, but now with German listeners, would allow one to estimate possible lexical involvement by using the same stimuli as words with Dutch listeners and as nonwords with German listeners. If the results of Experiment 1 were to be replicated, it would be unlikely that this was due to lexical inference. It has to be noted, however, that German has a rule for nasal place assimilation. Therefore, it could not be ruled out that knowledge of a rule of place assimilation as such might also play a role.

Method

Participants

Ten native speakers of German participated. One had just received his first lesson in Dutch; all the others had no knowledge of Dutch. It was verified that this 1 participant did not know that *tuin* [tœyn] was the Dutch word for *garden*. The sample was quite heterogeneous, with an age range from 20 to 58 years and with quite different professional occupations.

Design, Materials, and Procedure

The design and materials were identical to those in Experiment 1. Given that one might expect nonnative speakers to show some initial problems in judging the, for them, foreign language materials, we doubled the length of each block so that, in effect, every stimulus would be heard 40 times.

Apparatus

The experiments were done with a mobile lab. Stimulus presentation was controlled from a laptop by a Delphi 5.0 (Borland, 1999) program written by the first author. The sound output of the computer was amplified (Radio Design Labs, ST-PH 1) and was played to the participants via headphones (Sony MDR-V900). The experiment was done in a quiet room, with only the experimenter and the participant present.

Results

The percentages of correct identifications are displayed in Table 4 (see also Figure 1). A repeated measures ANOVA showed that the percentages of correct identifications were significantly influenced by the context [$F(2,18) = 11.3, p < .005$].

In addition, the nasal murmur factor was significant [$F(1,9) = 9.6, p < .05$], as was the interaction between the factors [$F(2,18) = 7.6, p < .005$]. To investigate the

Table 4
Percentages of Correct Identifications (With Standard Deviations) in Experiment 3

Stimulus	Context					
	No Context		Unviable (<i>stoel</i>)		Viable (<i>bank</i>)	
	%	SD	%	SD	%	SD
<i>tuin</i>	97.0	2.8	98.3	2.6	92.8	9.3
<i>tuim</i>	97.5	2.4	97.0	4.0	72.0	23.9

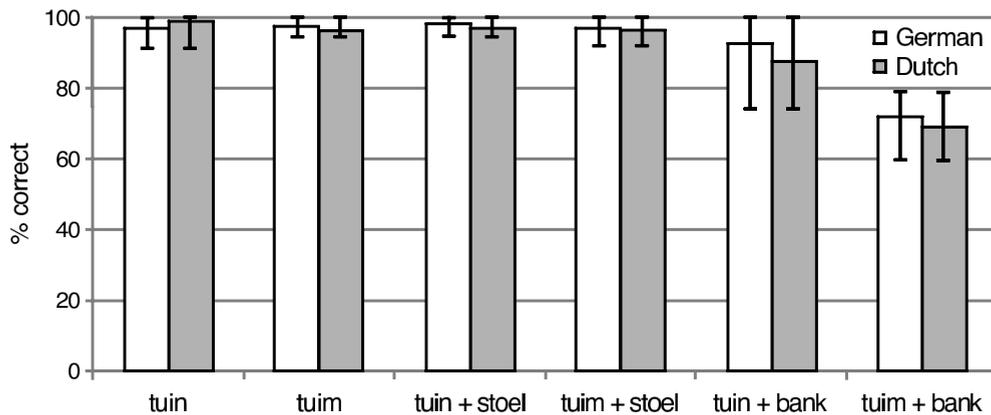


Figure 1. Mean percentages of correct identifications and confidence intervals in Experiment 1 (Dutch sample) and Experiment 3 (German sample).

nature of the interaction, we examined the effect of context on each level of the nasal murmur factor separately. For the alveolar nasal murmur, context produced only a weak trend [$F(2,14) = 2.1, p < .15$]. For the labial nasal murmur, context influenced percentages of correct responses significantly [$F(1,14) = 11.4, p < .005$]. Post hoc Newman–Keuls tests ($p < .05$) revealed that the level of performance was lower in the viable-context condition (72.0% correct) than in the two other conditions (97.0% correct for the no-context condition and 97.5% correct for the unviable-context condition). The latter conditions did not differ significantly. This interaction revealed that the German listeners replicated the performance of the Dutch participants in Experiment 1; they also made the most errors in (mis)perceiving *tuimbank* as *tuinbank*.

Post Hoc Between-Experiment Comparisons

Combining the Dutch and the German results. To allow a more detailed comparison, the results of Experiments 1 and 3 were combined in a repeated measures ANOVA with native language as a between-subjects variable. The descriptive data are presented in Figure 1, showing similar trends in both samples. In the statistical analysis, the context factor was significant [$F(2,32) = 35.7, p < .001$]. The nasal murmur factor was also significant [$F(1,16) = 13.6, p < .005$], as was the interaction between nasal murmur and context [$F(2,32) = 15.0, p < .001$]. However, neither the between-subjects native language variable nor its interaction with any of the within-subjects variables was significant (all $F_s < 1$). Familiarity with the stimulus words did not have any measurable influence on identification performance.

Testing for transitory effects. The results of the previous experiments indicated that the participants tended to (mis)perceive the *tuimbank* stimulus as *tuinbank*. During the experiments, we often observed that the participants were initially puzzled when negative feedback was provided in the viable-context block. This might have led to a change in performance over the course of the critical

block. Therefore, it was tested whether accuracy improved over the course of the critical viable-context block. The accuracy for the first and the second halves of this block were compared: Accuracy did not differ between the first and the second halves of the viable-context block in any of the experiments [Experiment 1, 77.5% vs. 79.4%, $t < 1$; Experiment 2, 78.9% vs. 70.0%, $t(7) = 2.0, p = .09$; Experiment 3, 81.3% vs. 83.6%, $t < 1$]. On average, accuracy was almost identical for the first (79.2%) and the second (77.7%) halves of the viable-context block over all three experiments ($t < 1$).

Discussion

The results render it unlikely that the relatively poorer identification in the viable-context condition and the bias toward perceiving a final /n/ in response to the *tuimbank* stimulus was a consequence of lexical inference. The results of the German participants did not differ from those of the Dutch participants. The results were not only qualitatively similar, but also quantitatively comparable. A combined statistical comparison of the present experiment with Experiment 1 showed that neither the main effect of native language nor any interactions involving this factor showed a trend toward significance. This is in slight contrast with the results of Gaskell and Marslen-Wilson (1998), who found a stronger context effect in the word than in the nonword condition. In our study, however, it did not matter whether the stimuli were perceived as words or nonwords.

In addition to testing for effects of lexical inference, we also evaluated the possibility that the participants needed to adjust their strategy only in the viable-context condition. If this were the case, the comparatively low level of accuracy in the viable-context condition in the viable-context condition block should be confined to the first trials. However, in all three experiments, an equivalent number of errors were made in the first and the second halves of the critical block. These data indicate that the effect was not transitory but was robust over the course

of the experiment. Although the participants clearly showed an awareness response to negative feedback in the viable-context block, they were not able to devise a strategy that led to a level of performance near ceiling level as in the other blocks of the experiment. An interpretation of the results of Experiment 1 favoring a regressive-compensation mechanism, free of perceptual masking and lexical inference effects, might be compromised by the fact that the experimental task also recruited attentive and decision-making processes.

EXPERIMENT 4

In this experiment, a rather different method was used to provide evidence for regressive compensation for phonological assimilation. The evidence previously brought forward to substantiate this claim was based on data collected by means of experimental tasks that always required an active decision on the part of the participant (Gaskell & Marslen-Wilson 1996, 1998; Gow, 2002). This also holds for our own Experiments 1–3. It may, therefore, be argued that compensation for phonological assimilation occurs mainly as a consequence of decision processes and not as a result of an automatic perceptual mechanism. Outside the laboratory, listeners do not actively decide on the place of articulation of word-final nasals. Therefore, the aim of this fourth experiment was to find evidence for the perceptual compensation for phonological variation by means of a task that did not afford a decision by the participant. This could best be achieved by using a passive-oddball experiment while measuring MMN (see the introduction for the rationale).

Two different trains of stimuli were used. One train consisted of a canonical form and a viable variant (*tuinbank* and *tuimbank*); the other train consisted of a canonical form and an unviable variant (*tuinstoel* and *tuimstoel*). The acoustic differences between the standard and the deviant stimuli within each pair were identical. If we were to find a difference in the size of the MMN elicited in a passive-oddball paradigm, such that the MMN to the viable-context pair was smaller than that to the unviable-context pair, this would provide evidence for the assumption that there is a compensation mechanism for phonological assimilations that is perceptual in nature.

Method

Participants

Eight students from Maastricht University took part in the experiment, 5 female and 3 male. The participants were paid for participation. None of the participants reported a history of hearing problems. None of the participants had participated in any of the previous experiments.

Design

As in the previous experiments, a two-factorial design was used. The first factor was the identity of the nasal murmur (/n/ or /m/). The canonical pronunciation with an alveolar nasal murmur was used as the standard stimulus, and the labial nasal murmur was used

as the deviant stimulus of the oddball series. The second factor was the context (viable [*bank*] vs. unviable [*stoel*]). The stimuli were presented in four blocks. Each block consisted of 800 stimuli (83% standards and 17% deviants) and lasted about a quarter of an hour. Blocks alternated between the viable context (standard [*tuinbank*] and deviant [*tuimbank*]) and the unviable context (standard [*tuinstoel*] and deviant [*tuimstoel*]). Presentation order was counter-balanced across participants.

While ERPs were measured, the participants did not react overtly to the stimuli and, therefore, did not receive any feedback. Therefore, we administered an identification posttest after the passive-oddball task. In this task, the participants listened to the target stimuli in the no-context condition, as well as in the viable and the unviable contexts. In this posttest, no feedback was given, in order not to deviate from the ERP experimental design.

Materials, Apparatus, and Procedure

The stimulus materials and apparatus were the same as those in Experiment 1. After preparation for EEG recording was finished, the participants were seated in a sound-attenuated room and were instructed to relax and watch a silent movie. Meanwhile, they heard a train of stimuli with 1.15 sec between the onsets of consecutive stimuli. Each stimulus had a duration of 617 msec. After completing the four blocks, the participants judged the stimuli in a 2AFC task (see Experiment 1) without feedback.

Electrophysiological Recording and Data Reduction

The nose-referenced EEG (0.1–125 Hz; sampling rate, 256 Hz) was recorded with a 32-channel electrode cap covering frontal, central, temporal, and parietal scalp areas. Blinks and vertical eye movements were monitored with electrodes placed at the sub- and supraorbital ridges of the right eye. Lateral eye movements were monitored by a bipolar montage, using two electrodes placed on the right and the left external canthus. All electrode impedances (EEG and EOG) were kept below 5 k Ω . ERPs were obtained by averaging the EEG time-locked to the onset of the sound. The ERPs were bandpass filtered digitally from 1 to 30 Hz. The signal was normalized using a baseline that was calculated with the mean amplitude from 30 msec before stimulus onset to the onset of the mismatching nasal murmur, which was 170 msec poststimulus onset (see Näätänen et al., 1993). Epochs exceeding $75 \mu\text{V}$ on any of the channels (including the eye channels) were rejected, leading to a rejection of about 30% of the trials. The individual ERP averages were obtained from 272 epochs in each condition by using all the deviant stimuli and the same number of standard stimuli randomly drawn from all the standard stimuli, excluding the standards directly following deviants. MMN was quantified in the following way (see Schröger, 1998). An area measure was obtained from the electrode location Fz, which is the most common electrode for MMN measurement. Mean amplitude was calculated for a 100-msec window starting 100 msec after the onset of the mismatching nasal murmur. These values were used for statistical analysis.

Results

Electrophysiological Measurement

The grand-average waveforms from a selection of electrodes—giving an impression of the scalp distribution—are displayed in Figure 2. There is a clearly defined negative deflection for the deviant stimulus in the unviable-context condition, but not in the viable-context condition. In the viable-context condition, the mean amplitude measurements were $-0.23 \mu\text{V}$ for the standard stimulus and $-0.38 \mu\text{V}$ for the deviant stimulus. This leads to an MMN of $-0.15 \mu\text{V}$, which was not signifi-

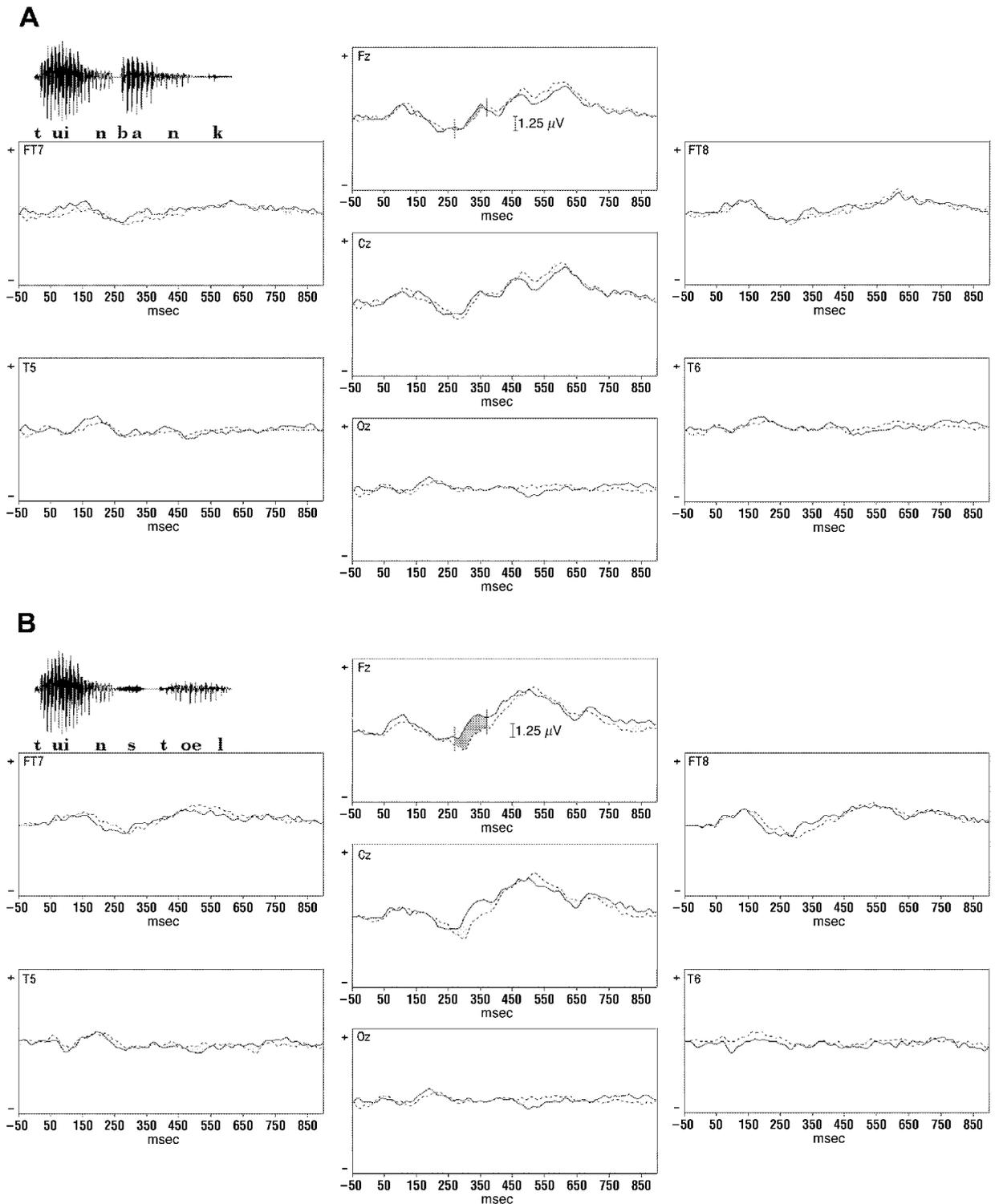


Figure 2. Grand average waveforms from Fz, Cz, Oz, FT7, FT8, T5, and T6. The electrodes are arranged in a manner analogous to their positions on a head that is “looking” toward the top of the page. The upper panels (A) show the event-related potentials (ERPs) in the viable-context (*bank*) condition; the lower panels (B) show the ERPs in the unviable-context (*stoel*). The solid lines represent the standards (*tuinbank* and *tuinstoel*); the dotted lines represent the deviants (*tuimbank* and *tuimstoel*). In the upper left corners, an oscillogram of a sound used is displayed as it aligns in real time with the ERPs. The gray shadings indicate the area in the Fz waveforms used for the estimation of the mismatch negativity.

cant ($t < 1$). In the unviable-context condition, the mean voltage values were $0.34 \mu\text{V}$ for the standard stimulus and $-0.75 \mu\text{V}$ for the deviant stimulus. This equals an MMN of $1.09 \mu\text{V}$, which was significant [$t(7) = 4.55, p < .005$]. To test whether the two MMN waves differed from each other, a two-factor repeated measures ANOVA was performed with the factors of nasal murmur (standard /n/ vs. deviant /m/) and context (viable vs. unviable). The nasal murmur factor was significant [$F(1,7) = 10.0, p < .025$], indicating an overall MMN. The context factor did not reach significance ($F < 1$). Most important, the interaction between factors was significant [$F(1,7) = 9.9, p < .025$], indicating that the change in nasal murmur affected the electrophysiological response differently depending on the viability of the change.

Behavioral Posttest

The mean percentages of correct identifications are displayed in Table 5. A repeated measures ANOVA confirmed that context did influence performance [$F(2,14) = 92.7, p < .001$].

The nasal murmur factor was also significant [$F(1,7) = 68.3, p < .001$], as was the interaction between the two factors [$F(2,14) = 57.9, p < .001$]. In order to investigate the nature of the interaction, the effect of context was evaluated separately for the alveolar and the labial nasal murmurs. Context did not influence performance for the alveolar nasal murmur ($F < 1$) but did so for the labial nasal murmur [$F(2,14) = 87.4, p < .001$]. Post hoc Newman-Keuls tests ($p < .05$) revealed that performance in the viable-context condition (35%) was worse than that in the unviable-context condition (96.9%) and the no-context condition (96.3%). The latter two conditions did not differ significantly.

Discussion

In this experiment, we found evidence for a regressive compensation for phonological variation by using a passive-oddball paradigm in which participants were not required to make active decisions. MMN was significant only for the unviable deviant stimulus, but not for the viable deviant, despite the same acoustical difference between the viable and unviable deviants and their respective standards. It should be noted here that this is the first report to show that MMN reflects auditory context in speech perception.

The main result of this fourth experiment was that the MMN in the unviable-context condition was significantly larger than the MMN in the viable-context condi-

tion. If one interprets MMN as a measure of perceptual distance (see the introduction), the main result translates into the finding that a canonical phonological form and a viable change are separated by less perceptual distance than are a canonical form and an unviable change. This result provides strong evidence for the existence of regressive context effects in the perception of place assimilation.

On a general level, MMN results are ambiguous with regard to the level at which the observed context sensitivity is located. Previous studies indicated that MMN is sensitive to phonological (Phillips et al., 2000) and lexical (Pulvermüller et al., 2001; but see Wunderlich & Cone-Wesson, 2001) levels. However, influences from lexical processes seem, in our case, rather unlikely, given the time frame of the MMN. The MMN was already measurable immediately at the onset of the context word. At this point, the participants heard [tœyns ...] or [tœynbɑ ...]. It was thus not yet clear that the complete stimulus was an existing compound noun. Therefore, the present experiment provides converging evidence for a regressive compensation mechanism for place assimilation that is perceptual in nature and very likely prelexical.

The behavioral posttest yielded two additional results. Unlike in the previous experiments, no feedback was provided here. First, the context effect in the viable condition for the labial nasal murmur was much larger than those in the previous experiments. Second, there was no trend for a drop in accuracy for the alveolar nasal murmur. Such a drop was significant in Experiment 2 and was observed as a trend in Experiments 1 and 3. In the discussion of Experiment 2, we promised to discuss this here. Previously, we noted that the feedback initially elicited a puzzled reaction by the participants when they performed the viable-context block. It seems likely that feedback led the participants to adopt a strategy in which they tried to "balance" their answers over the possible answer categories, given that negative feedback is mostly a consequence of choosing the canonical-form response. Although this strategy neither obliterated the effect that most errors were made in reaction to the viable assimilated form nor led to an increase of the level of performance over the course of the block, it induced some errors for the stimulus with the alveolar nasal murmur, too. The assumption that the errors arising from responding *tuim* in case of the [tœynbɑŋk] stimulus might be due to the attempt to prevent the "opposite" error (responding *tuin* in case of the [tœymbɑŋk] stimulus) is buttressed by the fact that most of the participants from Experiments 1–3 (23 out of 26, 88.5%) made their first error in reaction to the [tœymbɑŋk] stimulus. This is significantly different ($\chi^2 = 5.1, p < .025$) from the 72.7% that would be expected if all the errors observed were to be distributed equally over the blocks.

GENERAL DISCUSSION

Four experiments were conducted to investigate how listeners cope with the variation caused by place assimilation in continuous spoken word recognition. In the

Table 5
Percentages of Correct Identifications (With Standard Deviations) in the Posttest of Experiment 4

Stimulus	Context					
	No Context		Unviable (<i>stoel</i>)		Viable (<i>bank</i>)	
	%	SD	%	SD	%	SD
<i>tuin</i>	96.3	2.1	99.4	0.6	96.9	2.1
<i>tuim</i>	96.3	1.6	96.9	1.3	35.0	6.1

first experiments, the participants had to indicate whether the Dutch word *tuin* (English, *garden*) was pronounced canonically or with a changed place of articulation for the final nasal—that is, [tœym]. This was tested by presenting the target words in isolation or in a context that did (... *bank*; English, *bench*) or did not (... *stoel*; English, *chair*) allow for the /n/ to /m/ change of a word-final nasal. Identification performance was near ceiling in both the no-context and the unviable-context conditions. However, in the viable-context condition, the participants not only made more errors, but also showed a significant tendency to (mis)perceive the *tuimbank* stimulus as *tuinbank*. Two more behavioral experiments were conducted, to control whether these results were (partly) caused by masking or lexical top-down effects. Experiment 2 showed that large-amplitude noise bands presented directly after the target stimulus did not lead to a decrease in performance. This rules out the possibility that the differences in amplitude between the viable context (*bank*) and the unviable context (*stoel*) caused the difference in identification accuracy. In Experiment 3, German participants listened to the Dutch utterances, which sounded unlike any German word, and the original pattern of results was replicated. This makes it unlikely that the results can be attributed to some kind of lexical inference.

Finally, Experiment 4 ruled out the possibility that the effects were attributable to attentive and/or decision processes. The MMN for the viable standard–deviant pair *tuinbank*–*tuimbank* was smaller than that for the unviable pair *tuinstoel*–*tuimstoel*. In addition, the fact that the MMN was measured in the same time frame in which the context words were presented indicates that the lexical status of the stimulus hardly influences the context-sensitive compensation mechanism.

Although both the electrophysiological and the behavioral results indicate a prelexical mechanism for perceptual compensation for phonological assimilation, the precise level at which the perception of place assimilations is vulnerable to context effects needs further clarification. Gaskell and Marslen-Wilson (1996, 1998) proposed a regressive phonological inference mechanism working on abstract phonological representations. This implies that such an assimilated utterance as [tœymbaŋk] is perceived “correctly”—that is, indicating a labial place of articulation for the nasal—at auditory and/or phonetic levels of processing. Currently, however, there is no evidence indicating the supposed *context insensitivity* of earlier auditory and phonetic processing stages. It is noteworthy that even in the present experiments, in which explicit feedback directed the participants’ attention to acoustic and phonetic details, the participants were still influenced by context, and this effect did not dissipate over the course of the present experiments. Although this does not preclude the possibility that other measures may show early context-insensitive processing of place assimilations, it seems worthwhile to consider the possibility of context sensitivity at an auditory (e.g., Lotto & Kluender, 1998; Lotto et al., 1997) or a phonetic (Fowler, Brown, & Mann, 2000) level more closely.

At least two lines of reasoning can be put forth in favor of such an assumption. First, a compensation mechanism at a subsymbolic level may be better suited to deal with some ecological facets of assimilations. In Dutch, nasal place assimilation happens in only 60% of the cases in which phonology would allow it (Van Heuven & Van Berg, 1982). In addition, Gow and Hussami (1999) reported subcategorical differences between intended and assimilated labials (see also Nolan, 1992). Hence, place assimilations are optional and often only partial in speech-sound production. In some respect, this is mirrored by our finding that the participants were significantly beyond chance level in recognizing the *m* in [tœymbaŋk] but were far from perfect. This suggests that compensation for place assimilation is probably only partial, as has been observed in other cases of context-dependent speech perception (Beddor & Krakow, 1999; Fowler & Brown, 2000). Therefore, the equation might be that partial compensation by the listener complements partial assimilation by the speaker to achieve successful communication. In its current form, the model of phonological inference assumes that compensation for assimilation is an all-or-none phenomenon. As such, it is difficult to account for graded effects. Although graded effects may be accommodated by allowing phonological inference to be a graded instead of an all-or-none phenomenon, a subsymbolic mechanism may be still more efficient. At a subsymbolic level, auditory/phonetic detail—information about speech rate and the amount of hypoarticulation—is still available. Such variables may help to differentiate labial segments from coronal segments that have been assimilated by labials.

A second argument in favor of a subsymbolic locus of compensation for assimilation stems from proposals in the field of phonology (Hura, Lindblom, & Diehl, 1992; Seo, 2001; Steriade, 2001). These linguistic approaches suggest that assimilation rules are *perceptually licensed*; that is, the attested assimilation rules tend to lead to perceptually inconspicuous changes from the canonical form. The changes in speech production (assimilation rules) are not independent of the perceptual abilities of the listener. However, the invariance problem caused by place assimilation has almost without exception been treated as a purely perceptual problem, comparable to the problem of color constancy despite constantly changing illumination conditions during the day. An alternative account suggests that languages and, therefore, assimilation rules have an evolutionary trajectory of their own, during which they adapt to the properties of the human (in the present case) perceptual system (see, e.g., Deacon, 1997; de Boer, 2000; Steels, 1999). Since context sensitivity is a general perceptual principle (Warren, 1999), it might be expected that in the shaping of place assimilation rules, use was made of the already present context-sensitive mechanisms for dealing with invariance problems (Lotto et al., 1997; Mitterer, 2000).

Very recently, a related account of compensation for assimilation was proposed, which supports our suggestion of a subsymbolic locus for the processing of assimilated speech (Gow, 2002). This feature-parsing account builds

on the evidence that the assimilated segment bears information for both a labial and the original alveolar place of articulation (see above). Thus, the /m/ in [tɛymbɑŋk] actually bears feature cues for an alveolar and a labial nasal. Compensation for articulation is then achieved by parsing the cues for the labial place of articulation from the assimilated segment (the /m/ in [tɛym]), ascribing this information to the assimilating segment (the /b/ in [bɑŋk]). After this feature parsing, the assimilated segment bears cues only for the intended alveolar place of articulation. Feature parsing is assumed to rely on principles of auditory grouping (Bregman, 1990), which deconvolutes the overlapping feature cues.

In principle, this feature-parsing account seems compatible with the kind of subsymbolic inference account that we propose. Gow (2002) assumed that the perception of an assimilated segment is not an all-or-none but a graded phenomenon based on Gestalt principles. As Gow admits, his *feature-parsing* account still leaves open the precise nature of the mechanisms that help us overcome the invariance problem caused by assimilatory processes. Therefore, in terms of Marr (1982), feature parsing may be considered a description at a computational level, whereas the mechanisms proposed for context effects in speech perception can be considered descriptions at an algorithm level. Indeed, conceptions of general auditory contrast effects (Lotto & Kluender, 1998), direct perception of articulatory gestures (Fowler et al., 2000), or covariant learning on an acoustic-phonetic level (Holt, Lotto, & Kluender, 2001) may all constitute algorithms that contribute to the computational goal of feature parsing. Future research should indicate to what extent these conceptions of context sensitivities in speech perception may be able to explain context effects in the perception of place assimilations.

REFERENCES

- BAAYEN, R. H., PIEPENBROCK, R., & GULIKERS, L. (1995). *The CELEX lexical database* (Release 2) [CD-ROM]. Philadelphia: University of Pennsylvania, Linguistic Data Consortium.
- 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.
- BEHRINGER, J. (1996). *Experimental run time system* (Version 3.18) [Computer software]. Darmstadt: Technische Hochschule Darmstadt.
- BEST, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In J. C. Goodman & H. C. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 167-224). Cambridge, MA: MIT Press.
- BOERSMA, P., & WEENINK, D. (1999). *Praat 3.9: A system for doing phonetics with the computer* [Computer software]. Amsterdam: University of Amsterdam.
- BOOI, G. (1995). *The phonology of Dutch*. Oxford: Oxford University Press, Clarendon Press.
- BORLAND (1999). *Delphi 5.0* [Computer programming language]. Scotts Valley, CA: Author.
- BREGMAN, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.
- COLE, R. A. (1973). Listening for mispronunciations: A measure of what we hear during speech. *Perception & Psychophysics*, **13**, 153-156.
- COLLINS, B., & MEES, I. M. (1996). *The phonetics of English and Dutch*. Leiden: Brill.
- DEACON, T. W. (1997). *The symbolic species: The co-evolution of language and the brain*. New York: Norton.
- DE BOER, B. (2000). Self-organization in vowel systems. *Journal of Phonetics*, **28**, 441-465.
- 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.
- GANONG, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception & Performance*, **6**, 110-125.
- GASKELL, M. G., HARE, M., & MARSLÉN-WILSON, W. D. (1995). A connectionist model of phonological representation in speech perception. *Cognitive Science*, **19**, 407-439.
- GASKELL, M. G., & MARSLÉN-WILSON, W. D. (1996). Phonological variation and inference in lexical access. *Journal of Experimental Psychology: Human Perception & Performance*, **22**, 144-158.
- GASKELL, M. G., & MARSLÉN-WILSON, W. D. (1998). Mechanisms of phonological inference in speech perception. *Journal of Experimental Psychology: Human Perception & Performance*, **24**, 380-396.
- GOW, D. W. (2001). Assimilation and anticipation in continuous spoken word recognition. *Journal of Memory & Language*, **45**, 133-159.
- GOW, D. W. (2002). Does English coronal place assimilation create lexical ambiguity? *Journal of Experimental Psychology: Human Perception & Performance*, **28**, 163-179.
- GOW, D. W., & HUSSAMI, P. (1999, November). *Acoustic modification in English place assimilation*. Paper presented at the meeting of the Acoustical Society of America, Columbus, OH.
- 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.
- HUCKVALE, M. (1999). Opportunities for re-convergence of engineering and cognitive science accounts of spoken word recognition. In *Speech, language, and hearing: Work in progress* (Vol. 11, pp. 62-75). London: University College London, Department of Phonetics. Available at <http://www.phon.ucl.ac.uk/home/shl/>.
- HURA, S. L., LINDBLOM, B., & DIEHL, R. (1992). On the role of perception in shaping phonological assimilation rules. *Language & Speech*, **35**, 59-72.
- KEATING, P. A. (1990). The window model of coarticulation: Articulatory evidence. In J. Kingston & M. Beckman (Eds.), *Papers in laboratory phonology I* (pp. 451-470). Cambridge: Cambridge University Press.
- 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.
- MARR, D. (1982). *Vision*. San Francisco: Freeman.
- MARSLÉN-WILSON, W. D., MOSS, H. E., & VAN HALEN, S. (1996). Perceptual distance and competition. *Journal of Experimental Psychology: Human Perception & Performance*, **22**, 1376-1392.
- MARSLÉN-WILSON, W. D., & WARREN, P. (1994). Levels of perceptual representations and processes in lexical access: Words, phonemes, and features. *Psychological Review*, **101**, 653-675.
- MCCLELLAND, J. L., & ELMAN, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, **18**, 1-86.
- MITTERER, H. (2000). Are connected-speech processes selected for communicative consequences? In B. de Boer & P. Vogt (Eds.), *Proceedings of the Benevolang Workshop 2000* (pp. 31-34). Brussels: Vrije Universiteit Brussel.
- NÄÄTÄNEN, R. (1992). *Attention and brain function*. Hillsdale, NJ: Erlbaum.

- NÄÄTÄNEN, R. (1995). The mismatch negativity: A powerful tool for cognitive neuroscience. *Ear & Hearing*, **16**, 6-18.
- NÄÄTÄNEN, R., LEHTOKOSKI, A., LENNES, M., CHEOUR, M., HUOTILAINEN, M., IIVONEN, A., VAINO, M., ALKU, P., ILMONIEMI, R. J., LUUK, A., ALLIK, J., SINKKONEN, J., & ALHO, K. (1997). Language specific phoneme representation revealed by electric and magnetic brain responses. *Nature*, **385**, 432-434.
- NÄÄTÄNEN, R., SCHRÖGER, E., KARAKAS, S., TERVANIEMI, M., & PAAVILAINEN, P. (1993). Development of a memory trace for a complex sound in the human brain. *NeuroReport*, **4**, 503-506.
- NÄÄTÄNEN, R., & WINKLER, I. (1999). The concept of auditory stimulus representation in cognitive neuroscience. *Psychological Bulletin*, **125**, 826-859.
- NEELY, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 264-336). Hillsdale, NJ: Erlbaum.
- NITTROUER, S. (1996). Discriminability and perceptual weighting of some acoustic cues to speech perception by 3-year-olds. *Journal of Speech & Hearing Research*, **39**, 278-297.
- NOLAN, F. (1992). The descriptive role of segments: Evidence from assimilations. In G. J. Docherty & R. D. Ladd (Eds.), *Gestures, segment, prosody* (Papers in Laboratory Phonology, Vol. 2, pp. 261-280). Cambridge: Cambridge University Press.
- NORRIS, D. (1982). Autonomous processes in comprehension: A reply to Marslen-Wilson and Tyler. *Cognition*, **62**, 714-719.
- NORRIS, D., MCQUEEN, J. M., & CUTLER, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral & Brain Sciences*, **23**, 299-324.
- PHILLIPS, C., PELLATHY, T., MARANTZ, A., YELLIN, E., WEXLER, K., POEPEL, D., MCGINNIS, M., & ROBERTS, T. (2000). Auditory cortex accesses phonological categories: An MEG mismatch study. *Journal of Cognitive Neuroscience*, **12**, 1038-1055.
- PULVERMÜLLER, F., KUJALA, T., SHRYROV, Y., SIMOLA, J., TIITINEN, H. J., ALKU, P., ALHO, K., MARTINKAUPPI, S., ILMONIEMI, R. J., & NÄÄTÄNEN, R. (2001). Memory traces for words as revealed by the mismatch negativity. *NeuroImage*, **14**, 607-616.
- REPP, B. H., & SVASTIKULA, K. (1988). The perception of the [m]-[n] distinction in VC syllables. *Journal of the Acoustical Society of America*, **83**, 237-247.
- SAMUEL, A. G. (1996). Does lexical information influence the perceptual restoration of phonemes? *Journal of Experimental Psychology: General*, **125**, 28-51.
- SCHRÖGER, E. (1998). Measurement and interpretation of the mismatch negativity. *Behavior Research Methods, Instruments, & Computers*, **30**, 131-145.
- SCHULZE, H., & SCHEICH, H. (1999). Discrimination learning of amplitude-modulated tones in Mongolian gerbils. *Neuroscience Letters*, **261**, 13-16.
- SEO, M. (2001). A perception-based study of sonorant assimilation in Korean. *OSU Working Papers in Linguistics*, **55**, 43-69.
- SINNOTT, J. M., & SAVORITA, T. A. (2000). Differences in American English, Spanish, and monkey perception of the *say-stay* trading relation. *Perception & Psychophysics*, **62**, 1312-1319.
- STEELES, L. (1999). *The talking heads experiment: Vol. I. Words and meaning* (Special preedition). Brussels: Vrije Universiteit Brussel.
- STERIADE, D. (2001). Directional asymmetries in place assimilation: A perceptual account. In E. Hume & K. Johnson (Eds.), *The role of speech perception in phonology* (pp. 219-250). San Diego: Academic Press.
- THIERRY, G., DOYON, B., & DÉMONET, J. F. (1998). ERP mapping in phonological and lexical semantic monitoring tasks: A study complementing previous PET results. *NeuroImage*, **8**, 391-408.
- UTMAN, J. A., BLUMSTEIN, S. E., & BURTON, M. W. (2000). Effects of subphonetic and syllable structure variation on word recognition. *Perception & Psychophysics*, **62**, 1297-1311.
- VAN HEUVEN, V. J., & VAN BERG, I. (1982). Sprekerstrategie en de perceptieve dominantie van het woordbegin [Speakers' strategies and the perceptual dominance of word beginnings]. *Glott*, **5**, 31-48.
- WARREN, R. M. (1971). Identification times for phonemic components of graded complexity and for spelling of speech. *Perception & Psychophysics*, **9**, 345-349.
- WARREN, R. M. (1999). *Auditory perception: A new analysis and synthesis*. Cambridge: Cambridge University Press.
- WARREN, R. M. (2000). Phonemic organization does not occur: Hence no feedback. *Behavioral & Brain Sciences*, **23**, 350-351.
- WARREN, R. M., BASHFORD, J. A., JR., & GARDNER, D. A. (1990). Tweaking the lexicon: Organization of vowel sequences into words. *Perception & Psychophysics*, **47**, 423-432.
- WINKLER, I., LEHTOKOSKI, A., ALKU, P., VAINO, M., CZIGLER, I., CSÉPE, V., AALTONEN, O., RAIMO, I., ALHO, K., LANG, H., IIVONEN, A., & NÄÄTÄNEN, R. (1999). Pre-attentive detection of vowel contrasts utilizes both phonetic and auditory memory representations. *Cognitive Brain Research*, **7**, 357-369.
- WUNDERLICH, J. L., & CONE-WESSON, B. K. (2001). Effects of stimulus frequency and complexity on the mismatch negativity and other components of the cortical auditory-evoked potential. *Journal of the Acoustical Society of America*, **109**, 1526-1537.

(Manuscript received March 5, 2002;
revision accepted for publication March 5, 2003.)