THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY
2006, 59 (8), 1395–1424

The role of perceptual integration in the recognition of assimilated word forms

Holger Mitterer
Universiteit Maastricht, Maastricht, The Netherlands

Valéria Csépe
Institute for Psychology of the Hungarian Academy for Sciences, Budapest, Hungary

Leo Blomert
Universiteit Maastricht, Maastricht, The Netherlands

We investigated how spoken words are recognized when they have been altered by phonological assimilation. Previous research has shown that there is a process of perceptual compensation for phonological assimilations. Three recently formulated proposals regarding the mechanisms for compensation for assimilation make different predictions with regard to the level at which compensation is supposed to occur as well as regarding the role of specific language experience. In the present study, Hungarian words and nonwords, in which a viable and an unviable liquid assimilation was applied, were presented to Hungarian and Dutch listeners in an identification task and a discrimination task. Results indicate that viably changed forms are difficult to distinguish from canonical forms independent of experience with the assimilation rule applied in the utterances. This reveals that auditory processing contributes to perceptual compensation for assimilation, while language experience has only a minor role to play when identification is required.

The invariance problem in speech perception presents itself to the observer as a "variance problem". Different utterances of the same word vary dramatically depending on the speaker, the emphasis in the sentence, and the immediate phonetic/phonological context. The immediate phonological context of a word can in some circumstances lead to phonological assimilation—that is, word-final or word-initial segments may be modified by onsets or codas of the surrounding words. Assimilation is a rather strong form of context-dependent variation that, superficially, neutralizes a phonemic contrast. This raises the question of how words can still be recognized despite the deviation from the citation form when they are assimilated.

Studies concerned with the recognition of assimilated word forms mostly investigated cases of place assimilation in Germanic languages (Coenen, Zwitserlood, & Bölte, 2001; Gaskell &

Correspondence should be addressed to Holger Mitterer, Max-Planck Institute for Psycholinguistics, Postbus 310, 6500 AH Nijmegen, The Netherlands. Email: holger.mitterer@mpi.nl

This work was supported by a grant from the Dutch Scientific Organization (NWO, Project 048.011.046) in cooperation with the Hungarian Scientific Research Fund (OTKA, Project Number N 37282) awarded to Leo Blomert. We thank Dénes Szűcs, Annett Ragó, and Nina Davids for their help in running the experiments.

Marslen-Wilson, 1996, 1998; Gow, 2001, 2002, 2003; Mitterer & Blomert, 2003). This type of assimilation typically applies in C#C sequences in which the word-final consonant takes over the place of articulation of the second, word-initial consonant (e.g., *lean bacon* → *leam bacon*). It has generally been found that a changed form (*leam*) is accepted as an instance of the intended word (*lean*) only if the context allows the change to occur as the result of a phonological assimilation (. . . *bacon*). Assimilated forms are perceived as a violation, if the context does not allow assimilation (. . . *salami*): Gaskell and Marslen-Wilson (1996, see also Gow, 2002) showed that assimilated forms such as *leam* prime a lexical decision to a visual target—that is, the written word denoting the original word (i.e., *lean*)—only if the phonological context allows the assimilation, but not if the context does not allow the assimilation. Similarly, they (Gaskell & Marslen-Wilson, 1998) showed in a phoneme-monitoring task that the /m/ in *leam* is more likely to be—falsely—recognized as an /n/ if the phonological context allows an assimilation. Using event-related potentials, Mitterer and Blomert (2003) showed that compensation is a fast and automatic process. In a passive oddball paradigm, participants were exposed to an oddball series while watching a silent movie. The oddball series consisted of a Dutch word pronounced canonically ([tœyn . . .], English "garden") as standard, and deviants were either viable alternative pronunciations ([tœymbaŋk], English "garden bench") or unviable alternatives (*[tœymstuˈl], English "garden bench"). A deviation-elicited mismatch-negativity component could only be observed in the case of the unviable alternative. This shows that preattentive perceptual mechanisms distinguish a canonical pronunciation and an unviable alternative, whereas viable alternatives are "perceived" as not significantly different from the canonical pronunciations. In summary, a number of divergent techniques ranging from cross-modal identity priming to automatically evoked brain potentials indicate that assimilations are perceived in a context-sensitive way. This implies some form of compensation for assimilation, contrary to claims of underspecified recognition (Lahiri & Marslen-Wilson, 1991; Lahiri & Reetz, 2002).

Two different accounts for perceptual compensation for phonological assimilation have been proposed so far. First, Gaskell and Marslen-Wilson (1998) put forward a model of phonological inference. According to this view, listeners learn the assimilation rules of their native language implicitly during acquisition. Being exposed to the rule of coronal place assimilation, an English listener learns that a labial nasal followed by a labial obstruent, as in *leam bacon* may actually correspond to an intended alveolar nasal that has been assimilated by the labial obstruent (see Gaskell, Hare, & Marslen-Wilson, 1995). This proposed learning mechanism implies that compensation is located at a phonological processing level. First, the input is categorized phonologically in a context-insensitive way—that is, an assimilated utterance as "leam bacon" is first categorized as having a labial nasal. When the following segment is phonologically categorized as a labial plosive, the listeners "knows" that the previous labial nasal may be a consequence of nasal place assimilation and regressively infers that the previous nasal is possibly alveolar. This regressive-inference mechanism is also assumed to be influenced by the lexicon in a top-down manner. That is, the regressive inference is stronger if it changes a nonword into a word (as is the case in *freighp* → *freight*, but not for *preighp* → *preight*).

A second account of compensation for phonological assimilation has been proposed by Gow (2001, 2002, 2003). According to Gow's feature-parsing account, the assimilated segment—for instance, the /m/ in *leam bacon*—should be considered as bearing cues for both a labial and the original alveolar place of articulation. That is, assimilated segments differ from intended nonalveolar segments (as in "arm chair") and are hybrids in terms of place of articulation. This fits well with the description of phonological assimilation given from the perspective of gestural phonology (e.g., Browman & Goldstein, 1992) that assimilation arises because of gestural overlap. This assumption has been corroborated by acoustic measurements of assimilated segments

(Gow & Hussami, 1999; Nolan, 1992) and is now widely accepted (see Gaskell, 2003). According to the feature-parsing account, the feature cues for both the underlying and assimilated place of articulation are extracted from the signal. Compensation for assimilation is then achieved by parsing the cues for the labial place of articulation from the assimilated segment (the /m/ in "leam"), and ascribing this information to the assimilating segment (the /b/ in "bacon"). After this feature parsing, the assimilated segment is only associated with the cues for the intended alveolar place of articulation. Gow assumes that feature parsing is governed by grouping principles as proposed by Bregman (1990). Therefore, it is not necessary to acquire any knowledge about assimilation rules in order to achieve compensation for assimilation.

These two accounts assume that compensation for assimilation occurs at a level at which the acoustic input has already been transformed into phonological features or feature cues. However, it is conceivable that compensation for assimilation arises at earlier auditory processing levels. The present evidence with regard to compensation for phonological assimilation does not rule out such a possibility. In this paper, we investigate the possibility of such an account. Our proposal is based on the framework developed by Kingston and Macmillan (Kingston & Macmillan, 1995; Macmillan, Kingston, Thorburn, Dickey, & Bartels, 1999). In this framework it is assumed that speech perception involves at least two processing stages. First, the acoustic input is converted into a multidimensional perceptual space. Within this perceptual space, decision rules associate regions of the perceptual space with certain phonological categories (for similar views, see Nearey, 1990; Smits, 2001a, 2001b).

Context sensitivity can arise at both processing stages. If context sensitivity arises as a consequence of *perceptual integration* of target and context, then the position of the target in the perceptual space is already influenced by the context. This may in turn lead to a different phonological categorization by a (context-independent) decision rule. However, a context effect can also arise at the level of phonological categorization, without an influence of context on the representation of the input in the perceptual space. In the latter case, the context modifies the decision rule. Then, the decision rules associate different regions of the perceptual space with different phonological categories depending on the context.

To give an example, consider the assimilated utterance "freighp bearer". Let us assume that in this utterance, the actual F2 offset is lowered so that it is more compatible with the interpretation as a labial [p] than an alveolar [t]. Due to compensation for assimilation, the [p] in "freighp" is nevertheless perceived as a /t/ if it occurs in the context of a [b]. If this context effect is a consequence of perceptual integration of target and context, it would mean that phonological context leads to an increase of the perceived F2 offset frequency relative to the actual F2 offset. Alternatively, the F2 offset is perceived faithfully; however, the phonological context changes the decision rules so that lower F2 offsets are still accepted as instances of [t].

Here, we want to argue that perceptual compensation for phonological assimilation may arise as consequence of perceptual integration of target and context at early auditory levels of processing. The perceptual-integration account proposes that the acoustic properties linked to the production of assimilated and assimilating segments interact in auditory processing. Hence, in hearing the assimilated version of *garden bench*, [gardəm bɛntʃ], neither the feature nor the feature cues for the labiality of the final nasal are extracted. In contrast, the acoustic information in the following context "overwrites" the acoustic effects of assimilation before features or feature cues are extracted.

Such a "perceptual integration" might be produced by early auditory processes (cf. similar proposals by Delgutte, 1997; Holt & Lotto, 2002; Lotto, Kluender, & Holt, 1997; and Summerfield & Assmann, 1989). One candidate process to bring about compensation for assimilation at such an early level is "perceptual contrast". Repp (1983) concluded that a psycho-acoustically driven perceptual contrast arises in VCCV sequences if the two consonants are very similar to each other. That is, the first consonant in a C_1C_2 sequence was likely to be perceived as different from C_2 when the two consonants were similar to each other. Such a contrast effect would be especially suitable to deal with assimilations, which render C_1 and C_2 rather similar. Contrast effects are not restricted to cases in which the CC sequence is made up of two identical consonants (as in ab_1b_2a), but also arise in coarticulated fricative-stop and liquid-stop sequences (Mann, 1980; Mann & Repp, 1981). Therefore, it is likely that similar effects may occur in phone strings, in which one phone has been assimilated. Hence, we propose that basic auditory mechanisms producing contrast effects not only may explain "compensation for coarticulation", but may also be involved in "compensation for assimilation".

How do these accounts differ? First, the phonological-inference account predicts that a listener needs extensive experience with an assimilation rule in order to be able to compensate for a given assimilation. In contrast, both the feature-parsing and the perceptual-integration account attribute compensation for assimilation to factors that do not require learning of assimilation rules, such as perceptual grouping and context effects in auditory processing. There is yet little evidence that directly investigates compensation for assimilation cross-linguistically. Darcy (2002) investigated whether assimilated French words could still be recognized by English listeners. She found that only native speakers of English with a strong command of the French language were able to compensate for assimilation in French.

Secondly, the accounts differ with regard to the question at which level compensation for assimilation arises. The phonological-inference account assumes that the assimilated and assimilating segment are processed independently until, for both segments, phonological features are extracted. Then, the listener infers that a labial nasal may in fact be an underlying alveolar nasal if it is followed by a labial obstruent. The feature-parsing account assumes that assimilated and assimilating segments are processed independently up to a level at which feature cues are extracted. Then, due to grouping by similarity, the cues for labiality in an assimilated *leam* are grouped with the cues for labiality in the context *bacon*. Finally, the perceptual-integration account assumes that the "conflicting feature cues" in the assimilated segment are not extracted. Instead, the acoustic information linked to the production of the assimilated and the assimilating segment interact during auditory processing, thus decreasing the saliency of the conflicting information. Therefore, the feature cues, and consequently the phonological features, extracted from an assimilated utterance "leam bacon" will be similar to the feature cues extracted from a canonical pronunciation "lean bacon". The differences between the accounts with regard to the level at which compensation occurs can be captured by the statement that compensation for assimilation is assumed to occur before phonological-feature extraction (perceptual integration), during phonological-feature extraction (feature parsing), or after phonological-feature extraction (phonological inference).

Somewhat related to the question of level of processing is the question of lexical influences on compensation for assimilation. Gaskell and Marslen-Wilson (1998) found that phonological inference seems to be influenced by the lexicon: Compensation for assimilation was more likely to occur if the perceptual deassimilation of [m] to [n] rendered a word (e.g., "leam" → "lean") than in cases where it rendered a nonword. This result was, however, not replicated by Mitterer and Blomert (2003), who used a simpler task than that of Gaskell and Marslen-Wilson (two-alternative forced choice vs. phoneme monitoring). Based on their results, Gaskell and Marslen-Wilson suggested that phonological inference was influenced by the lexicon. The feature-parsing account may accommodate a lexical influence on compensation by invoking the notion that learned schemata may influence grouping (cf. Bregman, 1990). Words may be viewed as such schemata, which then make the grouping of the conflicting labial feature cues with the assimilated segment less likely. That is, a word such as "lean" is more likely to "repel" evidence for labiality than is a word such as "rum". What is problematic, however, is that not only the presence but also the absence of a "word superiority effect" in compensation for assimilation can be accommodated by the same feature-parsing account. The grouping process can proceed without using schemata, and, in this case, no lexical effect is expected. The perceptual-integration account makes a more specific prediction. Because auditory processing is probably independent of the lexical status of the input, compensation for assimilation should be independent of any higher level processing.

In the current paper, we present five experiments with relevant evidence to distinguish the three accounts for compensation for assimilation. In order to evaluate the effect of language experience on the perception of assimilated utterances, we compare the results from listeners with experience with an assimilation rule to results from listeners without such experience. In order to evaluate the level at which level compensation occurs, we used, first, words and nonwords as assimilated targets, and, second, we contrasted performance on an identification and discrimination task. Finally, we tested the speech specifity of the effects by using nonspeech analogues.

In order to evaluate the role of language experience, we needed an assimilation rule that is unknown to some of the participants. This speaks against the use of the well-investigated case of coronal place assimilation. This rule occurs in English, and it is difficult to find participants who do not have any experience with English. Therefore, we investigated a manner-assimilation rule that occurs in the relatively isolated Fin-Ugric language Hungarian. According to the phonology of Hungarian (see Olsson, 1992, p. 57; and Siptár & Törkenczy, 2000, p. 182), an apical lateral that is followed by an apical trill may also be pronounced as a trill. Thus, the Hungarian word for "from the left" /bɔlroːl/ may be pronounced [bɔrroːl], but the Hungarian word for "at the left" /bɔlnaːl/ may not be pronounced *[bɔrnaːl]. (/bɔl/ is the Hungarian word for "left", while /roːl/ and /naːl/ are directional case suffixes.)

These stimuli were first presented to Hungarian listeners. As "naïve" control participants, we used Dutch participants. A good feature of this comparison is that Dutch has a phonological distinction between /l/ and /r/, but no assimilation rule involving /l/ and /r/. The /l/ is mostly realized as an apical lateral. The phonetic implementation of the /r/ phoneme varies considerably within the Netherlands, with an uvular trill as the most common exemplar. However, the Hungarian standard of an apical trill is used by a subgroup of Dutch speakers, especially in the northern and central parts (Verstraeten & Van de Velde, 2001). Thus, Dutch listeners should clearly hear an opposition between a liquid and a trill, as they are familiar with apical trills and laterals. However, the phonology of Dutch has no assimilation rule similar to the liquid assimilation rule in Hungarian. Neither Booij (1995), from a generative point of view, nor Ernestus (2000), based on a survey of phoneme realizations in casual Dutch, report that a word-final [l] can be assimilated by a following [r]. Thus, this series of experiments allows us to estimate the influence of specific language experience on the perception of assimilated utterances and investigates the level at which compensation occurs.

EXPERIMENT 1: IDENTIFICATION

In this first experiment, we make use of a two-alternative forced choice task (2AFC). Listeners have to decide whether they hear [bɔl] "left", the canonical form, or the changed form [bɔr]. If compensation for Hungarian liquid assimilation is similar to compensation for major place assimilation in Germanic languages, listeners should make a clear distinction between a canonical form [bɔlnaːl] "at the left" and an unviable variant *[bɔrnaːl], but should not distinguish a canonical form [bɔlroːl] "from the left" from a viable alternative [bɔrroːl]. In order to test this, we created a [Cɔl]-[Cɔr] continuum and presented it in three conditions: in isolation, in a context that allows a change from /l/ to /r/, and in a context that does not allow this change. Identification was tested in a 2AFC task in which no feedback was supplied.

The main impetus of this experiment was to investigate whether this basic pattern of compensation for assimilation depends on experience with the assimilation and the lexical status of the assimilated word. Therefore, the identification task was performed by three groups. First, the stimuli were presented in the original form to Hungarian listeners. Second, the same stimuli were presented to Hungarian listeners with a changed initial consonant to form a nonword ([zɔl]). Finally, the Hungarian word was also presented to Dutch listeners. The comparison of the Hungarian word stimuli and nonword stimuli, presented to Hungarian listeners, allows us to evaluate whether there is an influence of the lexical status of the stimulus word on compensation for assimilation. Second, probing the perception of Hungarian utterances by Dutch listeners allows us to test whether experience with liquid assimilation is critical for compensation for liquid assimilation.

Method

Participants

All participants were psychology students from the University of Maastricht or the Pázmány Péter Catholic University of Piliscsaba and participated for course credit. All participants were free of a history of hearing problems. The Hungarian participant group hearing the original word utterances consisted of 10 listeners (6 female, 4 male) with a mean age of 22.5 years. All participants had some foreign language experience. English (9 participants) and German (4 participants) accounted for the majority of the foreign-language knowledge. In addition, 2 participants were acquainted with Italian and Latin, respectively. However, nobody in the sample had any knowledge of Dutch.

The Hungarian group listening to the Hungarian nonwords consisted of 12 participants (6 female, 6 male) with a mean age of 21.7 years. Of these, 2 participants had no foreign-language skills. The other 10 participants all spoke some English, and 7 participants also spoke some German. In addition, 1 participant spoke some Esperanto. Nobody in the sample had any knowledge of Dutch.

In the Dutch group, there were 3 male and 9 female participants, with a mean age of 21.1 years. All participants spoke English and either German or French as foreign languages. However, no participant had any knowledge of Hungarian.

Materials

A female native speaker of Hungarian was recorded uttering multiple tokens of the canonical form [bɔlnaːl] (English "at the left") and an unviable variant *[bɔrnaːl] and the canonical [bɔlroːl] form (English "from the left") and a viable alternative [bɔrroːl]. The Hungarian context suffixes of the cases called delative (*ról* "from where?") and adsessive (*nál* "where at?") were chosen, because they are phonetically quite similar. Both start and end with a sonorant and contain a long vowel. Any differences that these stimuli cause in the perception of the preceding segments can thus not be attributed to gross acoustic differences (overall amplitude, presence of voicing) between the context sounds.

The sample frequency for the recording of the natural utterances was 22050 Hz. Recordings were band-pass filtered from 130 to 8000 Hz. From one utterance of [bɔlnaːl], the first syllable was spliced out and edited with the software package PRAAT 4.0 (Boersma & Weenink, 2002). This syllable was then edited in order to create a continuum of speech sounds from the original apical lateral to an apical trill, using the purposefully assimilated utterances as a template.

A linear-predictive-coding analysis with 16 predictors yielded a stable solution and was used in order to estimate source and filter for this utterance. Given the identical place of articulation, the primary cue for the lateral/trill distinction is the presence of amplitude modulation (AM) in the trill (cf. Ladefoged & Maddieson, 1996). For the edited sound stimuli, one cycle of AM (20 Hz) was added to the estimated source using five steps from 0 dB (no AM) to 12 dB (strongest AM). This was done by editing the intensity function of the source as estimated by PRAAT. The middle panel of Figure 1 shows the intensity as

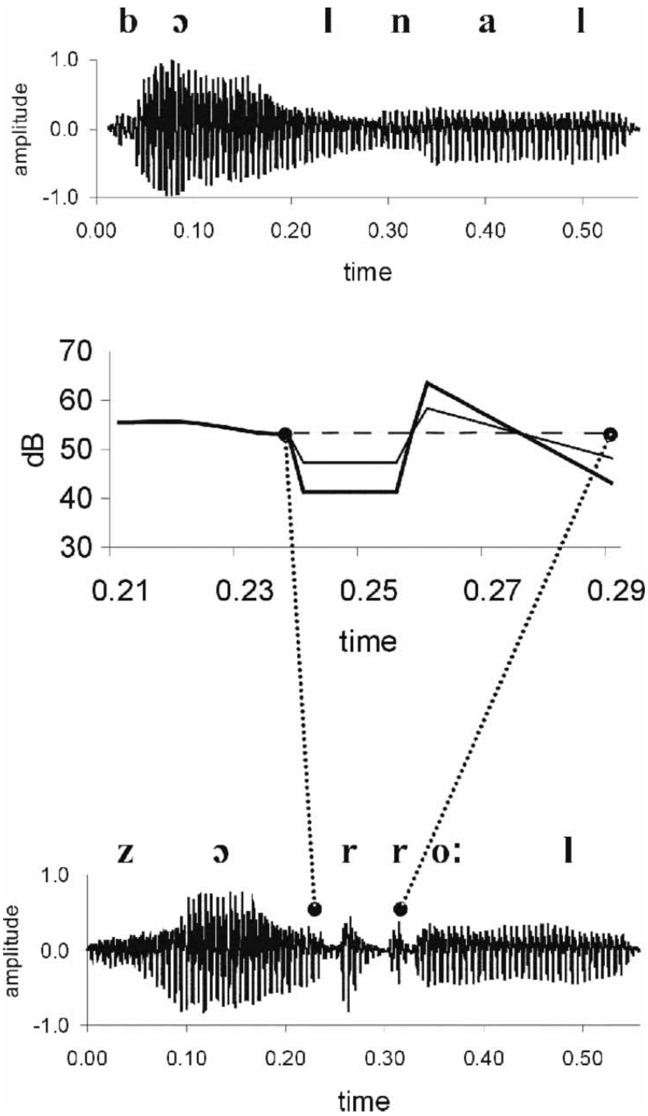

Figure 1. *The top panel shows the stimulus "balnal" without AM. The short silence at the beginning is due to the shortening of the prevoice bar. The middle panel shows the parameters used to create AM during the /l/ in "bal" and "zal". The bottom panel shows the result of this editing, but here with the nonword stimulus "zarrol". The first AM at 0.265 s is due to the manipulation of an original lateral, while the second peak at 0.315 s stems from the natural suffix "rol".*

manipulated for the stimulus without AM, the stimulus with AM of 6 dB, and the stimulus with AM of 12 dB. The source was multiplied by these edited intensity functions.

Formant estimations showed similar formant transitions into the consonant for laterals and trills, but a slightly lower third formant (200 Hz) for the resonant period of the trill. Therefore, the third formant was also lowered in five steps of 0.1 bark. Filtering the edited sources with the edited filters yielded five stimuli, in which the first stimulus ended in an final apical lateral, and the last stimulus emulated a final apical trill. In order to convey a casual speaking style, prevoicing was shortened from 40 ms to 27 ms by cutting two complete cycles in all stimuli starting with a voiced labial stop.

In order to create the Hungarian nonwords, the filter was manipulated for the initial 105 ms of the utterance in order to emulate the estimated filter function in natural utterances of an alveolar voiced fricative of the speaker. For the first 75 ms, the filter settings were stable, and from 75 to 105 ms, the filter settings were interpolated from the /z/ settings to the original /ɔ/ settings using the cosine function in [0, π]. This yields a slow initiation of the formant movement, a maximal acceleration at the midpoint, and slowing down again toward the endpoint. In addition, the original voice source was attenuated to one half of the original amplitude, and noise was added for the first 75 ms. From 75 ms to 105 ms the noise was faded out, and the original voiced source was restored to the original level using linear interpolation. The acoustic changes, with which to achieve a continuum ranging from an apical lateral to an apical trill in the postvocalic position, were identical to the Hungarian word continuum. These [zɔl]–[zɔr] stimuli were 13 ms longer than the [bɔl]–[bɔr] stimuli, reflecting that the acoustic signal in onset position is longer for fricatives than stops, because the closure associated with a stop is not marked within a speech stream in onset position.

The context case suffixes [naːl] (English "to" in answer to the question "where?" as in "to your left") and [roːl] (English "from") were spliced from other utterances, equalized in overall energy, and concatenated with the Hungarian word and nonword stimuli. The amplitude relation of words and nonwords to the case suffices was edited to emulate the amplitude relation of the first (always stressed) syllable to the second syllable in the natural utterances. Figure 1 shows two stimuli that resulted from the editing procedure. The upper panel shows the stimulus [bɔlnaːl] in which the first syllable has no amplitude modulation, and the second syllable is the case suffix [naːl] that does not allow the assimilation of /l/ to /r/. The lowest panel shows the stimulus [zɔrroːl] in which the original /b/ onset was manipulated as described above. The first AM peak at 0.265 s is the result of the editing of the intensity function as described above. The second peak at 0.315 s stems from the natural trill of the case suffix [roːl]. These stimuli were presented to the participants via headphones (SONY MDR-V 900) using E-prime (Psychology Software Tools, Version Beta 5.0, for the Hungarian listeners with Hungarian words), the Presentation software (Neurobehavioral Systems, for the Hungarian listeners with Hungarian nonwords), or ERTS (Behringer, 1996, for the Dutch listeners).

Procedure

Experiments were run with participants facing a computer screen, and instruction was given in written form via the computer screen. All participants were instructed that they were going to hear the Hungarian word for left, *bal*—or the nonword, *zal*—spoken by a Hungarian speaker who sometimes make an error, and pronounces *(b/z)al* as *(b/z)ar*. Participants were asked to indicate, after hearing a disyllable, whether the speech sound was *(b/z)al* or *(b/z)ar*. In order to prevent any stimulus–response incompatibility for the Hungarian participants, participants were instructed to hit the left key when hearing *bal* (Hungarian for "left") and the right key when hearing *bar* (of either the computer keyboard for Hungarian listeners, or the left button of a ERTS response box for the Dutch listeners).

A period of 500 ms before hearing a stimulus, the computer screen displayed the two answer alternatives (*bal* vs. *bar* or *zal* vs. *zar*) on the left and right of the screen, corresponding to the response key allocation. After hearing a stimulus participants had 2.5 s to respond. If no response was given in this time, a feedback screen asked participants to respond faster. If a response was given, the word on the screen corresponding to the response alternative was moved up and to the margin of the computer screen by 6 pixels while the other alternative disappeared. This indicated to the participants that their answer had been registered by the computer.

Conditions of context suffixes (no suffix, unviable suffix [naːl], and viable suffix [roːl]) were blocked. The order of presentation was randomized within blocks. All participants started by judging the stimuli without context suffixes. Then, the presentation order of the blocks was counterbalanced over participants. In every participant group, half of the participants first heard the stimuli with the viable suffix, while the other half first heard the stimuli with the unviable suffix.

Design

The design entails three independent variables: one between-subject variable with three levels and two within-subjects variables with three and five levels, respectively. The between-subject variable was listener group with the levels native listener–word, native listener–nonword, and non-native listener. The two within-subject variables were context (none, viable, unviable) and AM depth (five levels). There were 15 measurements for every cell in this design. The dependent variable is the percentage of "bal" responses calculated from the 15 trials per cell.

Results

The mean percentages of [Cɔl] responses are shown in Figure 2 for all participant groups. A repeated measures analysis of variance (ANOVA) with participant group as between-subjects variable and AM depth and context (none, viable, unviable) as within-subject variables revealed a significant effect of AM depth, $F(4, 124) = 290.85, p < .001$, a significant effect of context, $F(2, 62) = 24.93, p < .001$, and a marginally significant effect for the variable listener group, $F(2, 31) = 2.94, p < .1$. These simple effects were, however, qualified by significant two-way interactions between context and AM depth, $F(8, 248) = 39.10, p < .001$, AM depth and listener group, $F(8, 124) = 3.07, p < .005$, and a significant three-way interaction, $F(16, 248) = 2.65, p < .005$.

In order to investigate the nature of these interactions, we broke down the factorial design to evaluate simple effects that were not qualified by an interaction (cf. Keppel, 1991, p. 236). First, we examined the effect of the factors context and listener group for each level of AM depth separately. If the interaction of this analysis was significant, we then examined the effect of group for each level of context. In cases where there was a significant effect of listener group, the source of this effect was investigated by linearly independent planned comparisons to evaluate the effect of wordness (Hungarian listeners with words vs. Hungarian listeners with nonwords) and the effect of native language (Hungarian listeners vs. Dutch listeners).

This procedure yielded the following results (see Table 1 for an overview): For the stimuli with no AM modulation, there was a significant effect of context, $F(2, 62) = 12.3, p < .001$, and group, $F(2, 31) = 8.9, p < .01$. These effects were, however, qualified by a significant interaction of group and context, $F(4, 62) = 2.7, p < .05$. Separate ANOVAs for all levels of context showed that there was an effect of group for the stimulus in the viable context, $F(2, 31) = 6.3, p < .01$, but only a trend in the unviable context, $F(2, 31) = 3.1, p < .1$, and no significant effect for the stimulus without context, $F(2, 31) = 1.8, p > .1$. Planned comparisons for the viable-context stimulus yielded a significant effect for native language, $t(31) = 3.2, p < .05$, but no effect of wordness, $t(31) = 1.6, p > .1$. This was caused by the fact that Dutch listeners were less inclined to hear a lateral (71.5%) in the viable context than were Hungarian listeners (87.9%), even when there was no AM.

For the stimuli with 3-dB AM, there was a significant effect of context, $F(2, 62) = 7.1, p < .01$, and group, $F(2, 31) = 8.3, p < .01$, but no interaction, $F(4, 62) = 1.2, p > .1$. Planned comparisons for the listener group showed that wordness did not have a significant influence, $t(31) = 1.1, p > .1$, but native language did, $t(31) = 4.0, p < .001$. Dutch listeners were overall less inclined to hear a lateral in this condition (73.9%) than were Hungarian listeners (88.9%). Post hoc tests (HSD, $p < .05$) for the different levels of the context factor showed that the stimuli in the no-context condition were more often labelled as a lateral (91.3%) than those in the viable-context

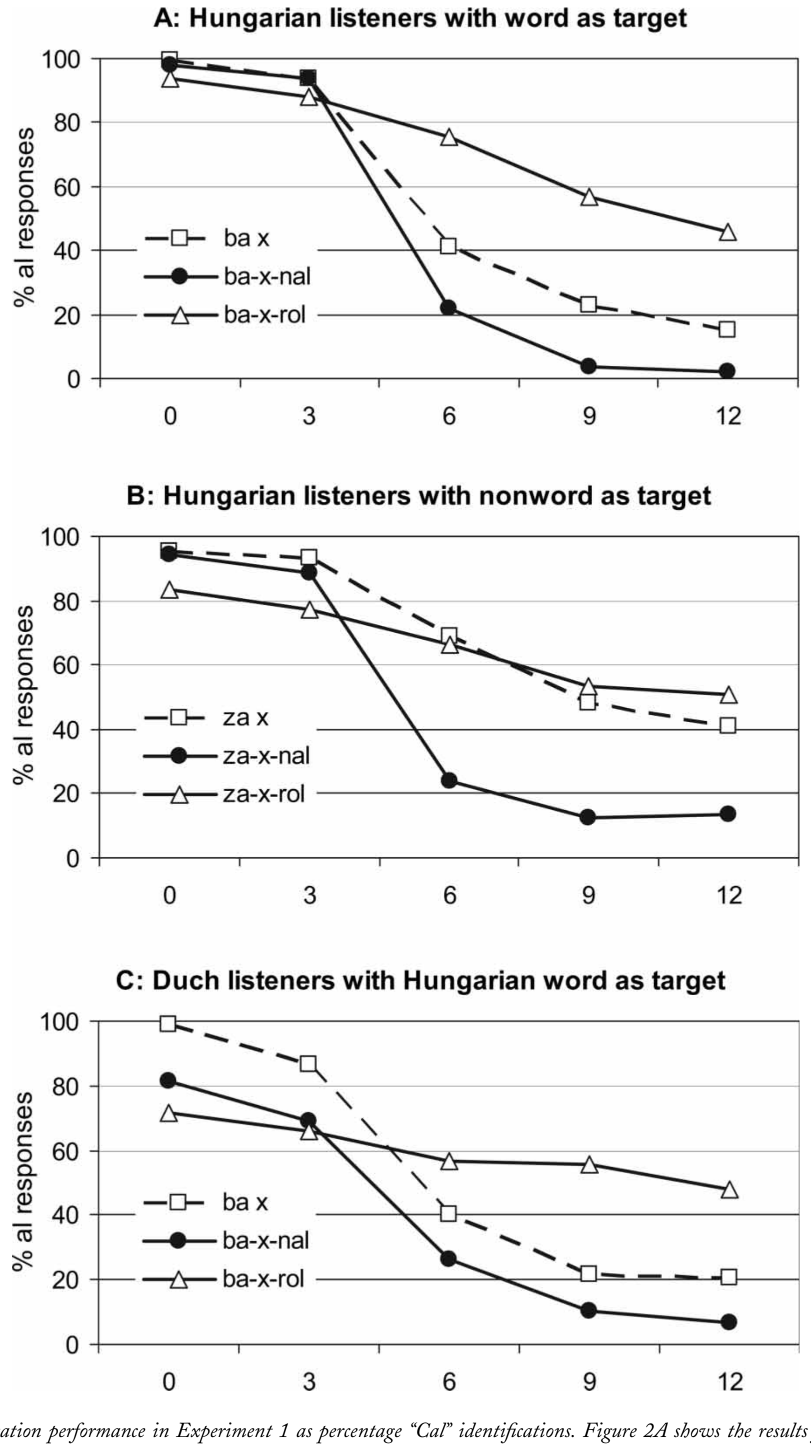

Figure 2. *Identification performance in Experiment 1 as percentage "Cal" identifications. Figure 2A shows the results for the Hungarian listeners with Hungarian words, Figure 2B shows the results for the Hungarian listeners with Hungarian nonwords, and Figure 2C shows the results for the Dutch listeners with Hungarian words.*

Table 1. *Results of the ANOVAs with group and context as predictors on the percentage of /l/-responses for each level of AM in Experiment 1*

				Group at context		
AM	*Group*	*Context*	*Interaction*	*None*	*Viable*	*Unviable*
0	—	—	*	*ns*	Hu > Du	*ns*
3	Hu > Du	n > {u,v}	*ns*	—	—	—
6	—	—	**	N > W	Hu > Du	*ns*
9	*ns*	v > n > u	*ns*	—	—	—
12	*ns*	v > n > u	*ns*	—	—	—

Note: AM = amplitude modulation (in dB). W = word. N = nonword. Hu = Hungarian listeners. Du = Dutch listeners. n = no context. v = viable context. u = unviable context. Irrelevant tests—depending on whether the interaction between group and context is significant—are indicated by (—). Effects of group were analysed by planned comparisons (word vs. nonword and Hungarian vs. Dutch listeners) and effects of context by HSD post hoc tests between all levels (no context, viable context, unviable context). Signs (<, >) are based on percentage of /l/-responses. Hence N > W indicates that significantly more /l/-responses were given in case of the nonword stimulus.

*$p < .5$; **$p < .01$; *ns*: not significant.

(77.1%) and unviable-context (83.7%) conditions. The two context conditions did not differ significantly.

For the stimulus with 6-dB AM, there was a significant effect of context, $F(2, 62) = 40.8$, $p < .01$, and no effect of group, $F(2, 31) = 1.9$, $p > .1$. In addition, there was a significant interaction, $F(4, 62) = 3.7$, $p < .01$. Separate ANOVAs for each level of context yielded an effect of group for the stimulus without context, $F(2, 31) = 3.8$, $p < .05$, and the viable-context stimulus, $F(2, 31) = 3.7$, $p < .05$, but not for the unviable-context stimulus ($F < 1$). For the stimulus without context, planned comparisons revealed a significant effect of wordness, $t(31) = 2.2$, $p < .05$, but no effect of native language, $t(31) = 1.5$, $p > .1$. The stimulus in the word condition was more likely to be perceived as a trill (58.8%) than was the stimulus in the nonword condition (31.1%). For the stimulus in the viable condition, there was no effect of wordness, $t(31) = 1.4$, $p > .1$, but a significant effect of native language, $t(31) = 2.5$, $p < .025$. The stimulus with medium AM in the viable context condition was more likely to be perceived with the "canonical" lateral by the Hungarian listeners (70.4%) than by the Dutch listeners (56.9%).

For the stimuli with 9- and 12-dB AM, there was a significant effect of context—9 dB, $F(2, 62) = 39.0$, $p < .001$; 12 dB, $F(2, 62) = 39.0$, $p < .001$—while both the group variable—9 dB, $F(1, 31) = 1.5$, $p > .1$; 12 dB, $F(1, 31) = 3.0$, $p > .1$—and the interaction—9 dB, $F(4, 62) = 1.9$, $p > .1$; 12 dB, $F < 1$—failed to reach significance. For both levels, post hoc tests (HSD) revealed that all context conditions differed significantly from each other in how often the stimuli were labelled as a lateral: no context, 31.4% at 9-dB AM and 26.2% at 12-dB AM; unviable context, 8.9% at 9-dB AM and 7.7% at 12-dB AM; viable context, 55.4% at 9-dB AM and 48.2% at 12-dB AM.

In summary, the results show that the AM was successful in leading to a perceptual change from /l/ to /r/ in the unviable condition and, to a lesser degree, in the no-context condition. In the viable-context condition, however, even strong AM modulation does not lead to a clear percept of an apical trill. This general pattern is moderated by the fact that Dutch listeners were less likely to perceive the "canonical" lateral in the viable context condition than were the Hungarian listeners. In addition, there was a difference between the word and nonword conditions in how the medium AM was perceived without context. Moderate AM was more likely to be perceived as a trill in a word than in a nonword. Most importantly, however, all listener groups perceived the maximal AM modulation as a trill in the unviable context, but as ambiguous in the viable context condition.

Discussion

First of all, the current experiment showed that compensation for liquid assimilation in Hungarian seems to be similar to compensation for major place assimilation of word-final nasals and stops in Germanic languages (Coenen et al., 2001; Gaskell & Marslen-Wilson, 1996, 1998; Gow, 2002; Mitterer & Blomert, 2003). A changed form, such as [bɔr], is more likely to be perceived as changed in a context that does not allow assimilation. If [bɔr] occurs in a context that allows assimilation, it is perceived as similar to the canonical form of the Hungarian word for "left", /bɔl/.

More importantly, the current results indicate that compensation for liquid assimilation is not completely dependent on experience with this assimilation. The Dutch listeners showed a similar overall pattern to that of the Hungarian listeners. They clearly perceived an opposition between a lateral and a trill in the unviable context but there was no clear category shift in the viable context. Nevertheless, there were differences between Hungarian and Dutch listeners. While both groups had a steeper identification function in the unviable context than in the viable context, only the native listeners showed a bias towards the "canonical" form. A possible interpretation of this pattern is that perceptual integration of target and context makes the distinction between [bɔl] and [bɔr] more difficult in the [roːl] context than in the [naːl] context. This affects all listener groups. Dutch listeners are therefore uncertain as to what they are actually hearing and basically identify all forms ambiguously, near 50% [l] identifications. However, Hungarian listeners may have developed a bias to resolve such a conflict toward [l], because they have learned that [l], but not [r] is likely to undergo assimilation. There are different possibilities of how to account for this language-specific effect. One possibility is that Hungarian listeners developed a form of partial underspecification: Hungarian listeners may need little evidence for an underlying /l/ to perceive it, while unambiguous acoustic evidence for /r/ is needed for Hungarian listeners to perceive an underlying /r/. (In the categorical version of underspecification theory, e.g., Lahiri & Reetz, 2002, listeners are assumed to always assume, in this case of assimilation, an underlying /l/ when presented with strong or weak evidence for either /l/ or /r/.) A second possibility is that the bias is independent of assimilatory patterns: Listeners may use language-specific knowledge about the likelihood of a given phoneme, in a given position, and this biases them toward the perception of this phoneme (cf. Pitt & McQueen, 1998).

A second question investigated in this experiment was whether compensation for assimilation is influenced by the lexical status of the assimilated word. This does not seem to be the case, although the task used previously reflected lexical influences (Ganong, 1980; McQueen, 1996). In the viable and unviable context, the speech sound continuum from [Cɔl] to [Cɔr] was perceived similarly in both the word and the nonword condition. This replicated earlier results by Mitterer and Blomert (2003). They showed that German and Dutch listeners compensated for assimilation equivalently when presented with a Dutch word that was a German nonword to which an assimilation has been applied that was viable in the German and the Dutch language.

An unexpected result was the finding that perception of the speech sound continuum was more categorical in the unviable context than in the no-context condition. A likely interpretation of this result is that, in utterance-final position, a trill with one AM period is not an acceptable phonetic implementation of a trill, but, within an utterance, this is a valid implementation of a trill. Therefore, we find a strong perceptual switch in the unviable context condition (one AM period within a phrase), but less so in the no-context condition (one AM period in phrase-final position).

In the Introduction, we argued that the three theoretical accounts—phonological inference, feature parsing, and perceptual integration—differ with regard to, first, the role of language experience, second, whether the lexicon influences compensation for assimilation, and third, the level at which compensation occurs. The current findings provide evidence that, first, language experience influences the perception of assimilated forms, although language-independent effects also drive the perception of assimilated forms. Secondly, the lexical status of the assimilated form does not seem to influence compensation for assimilation. This leaves one issue unresolved. It is still unclear whether compensation for phonological assimilation is based on speech-specific processes or on a perceptual integration of target and context on auditory levels of processing. In order to investigate this, we used a discrimination task, in which participants were not asked to label the stimuli, but to discriminate them. Using a discrimination task should incline listeners to probe an auditory level of analysis (see Beddor & Krakow, 1999, for a similar view). If the context sensitivity in the perception of assimilated utterances is based on a late, speech-specific processing stage, listeners should be influenced less by context in a discrimination task. If, however, perceptual compensation for phonological assimilation is based on early processing levels, listeners should be influenced by context in the discrimination task just as in the identification task.

EXPERIMENT 2: DISCRIMINATION

In this experiment, we used the same stimuli as those in Experiment 1 but employed a discrimination task. The phonological-inference account and the feature-parsing account predict that performance in a discrimination task is not influenced by the phonological categorization of the stimuli in question. Both assume that speech-specific information about the consequences of assimilation are extracted from the signal, either in the form of "feature cues" or in the form of phonological features. This information should allow listeners to make a distinction between assimilated and canonical form, no matter what the context is. However, this only holds if higher level representations, which are affected by compensation for assimilation, do not influence performance in the discrimination task. How can this be achieved? Gerrits (2001, Gerrits & Schouten, 2004; see also Schouten, Gerrits, & van Hessen, 2003) provided a comprehensive analysis of identification and discrimination tasks. Listeners in a discrimination task can base their responses on either phonological labelling or auditory representations. For the current purpose, we are interested in a discrimination task that leads participants to adopt the latter strategy, relying on auditory representations. Gerrits showed that a four-interval-oddity (4I-oddity) task leads to this mode of responding: Categorical recoding, even if available, did not influence task performance. In the experiments of Gerrits, both participants, who were most or least efficient in discriminating stimuli along a voiced-unvoiced-stop continuum, were equally proficient in within-category discrimination as in between-category discrimination.

In the 4I-oddity task, four stimuli presented at a constant interstimulus interval (ISI), of which three are identical (the standard), and one is different (the "odd"), are presented to the listener. The listener is (correctly) informed that either the second or the third is the "odd", and the task of the listeners is to indicate whether the second or third stimulus is the odd one.

In this task, the likelihood that participants are able to discriminate between two stimuli is the same for within-category and between-category pairs, given the same acoustic difference. Therefore, we may expect a dissociation between identification and discrimination for assimilated utterances, if compensation for assimilation is based on phonological processing, be it in the form of phonological inference or feature parsing. Both accounts predict that auditory levels of processing are not subject to context effects. In contrast, a perceptual-integration account predicts a context effect even in a discrimination task.

Method

Participants

The same three participant groups as those in Experiment 1 participated in Experiment 2. That is, 10 Hungarian participants performed the discrimination task with the Hungarian word stimuli, 12 Hungarian participants performed the task with the Hungarian nonword stimuli, and 12 Dutch participants performed the task with the Hungarian word stimuli. The discrimination task followed the identification task in all groups.

Materials

The speech sound stimuli were the same as those in Experiment 1. Sound files for the discrimination task were created before the experiment to prevent online timing problems. For each trial, four sounds were concatenated with an ISI of 425 ms. The standard stimulus was always the sound without AM, and the deviant was one of the stimuli with AM. This odd stimulus was either in the second or the third position in the train of four stimuli.

Procedure

Experiments were run with participants facing a computer screen. Instructions were also presented via the computer screen. All participants were instructed that they would hear a series of four stimuli, in which either the second or the third stimulus differed from the other stimuli. They were explicitly instructed that two sounds, which might be written in the same way, might still differ in how they sounded. Participants were asked to indicate which one was the odd after hearing a series of four speech sounds.

A period of 500 ms before hearing a stimulus consisting of four speech sounds, the computer screen displayed the two answer alternatives (the Hungarian or the Dutch words for "two" and "three") on the left and right of the screen, corresponding to the response key allocation. After hearing a train of four speech sounds, participants had 3 s to respond. If no response was given in this time, a feedback screen asked participants to respond faster. If a response was given, feedback indicated whether the choice was correct or not.

Conditions of context suffixes (unviable suffix [naːl] and viable suffix [roːl]) were blocked in four blocks of 40 trials each. Within blocks, order of presentation was randomized. The presentation order of the blocks was counterbalanced over participants. In every participant group, half of the participants heard the stimuli in the viable context in the first and third blocks and the stimuli in the unviable context in the second and fourth blocks. For the other half, the unviable-context stimuli were presented in the first and third blocks and the viable-context stimuli in the second and third blocks.

Design

In order to prevent a combinatorial explosion of conditions, we did not probe all possible contrasts. As we are interested in how far listeners are able to detect a mismatch with a canonical pronunciation, only the original stimulus with a lateral /l/ was used as standard in the 4I-oddity task, while the stimuli with differing degrees of AM served as odds. The design then entails three independent variables: one between-subject variable with three levels and two within-subject variables with two and four levels, respectively. The between-subject variable is listener group with the levels native listener–word, native listener–nonword, and non-native listener. The two within-subject variables are context (viable or unviable) and difference in AM (ΔAM) between the standard and the odd stimulus with four levels from 3-dB ΔAM to 12-dB ΔAM.

A total of 10 trials each were presented with the odd stimulus in the second and third positions for all eight conditions arising from the crossing of the ΔAM and the context factor. The dependent variable d' was calculated from the 20 trials per cell. We arbitrarily defined the second position as the target. Hence, a correct response in a trial with the odd stimulus at the second position is counted as a hit, while an error in a trial with the odd stimulus at the third position is counted as a false alarm. In order to calculate d', we reduced the range of the percentages of correct responses by 0.1% to 99.9%.

Results

Figure 3 shows the mean d' data for all conditions. An ANOVA with ΔAM, context (viable vs. unviable), and listener group revealed a significant

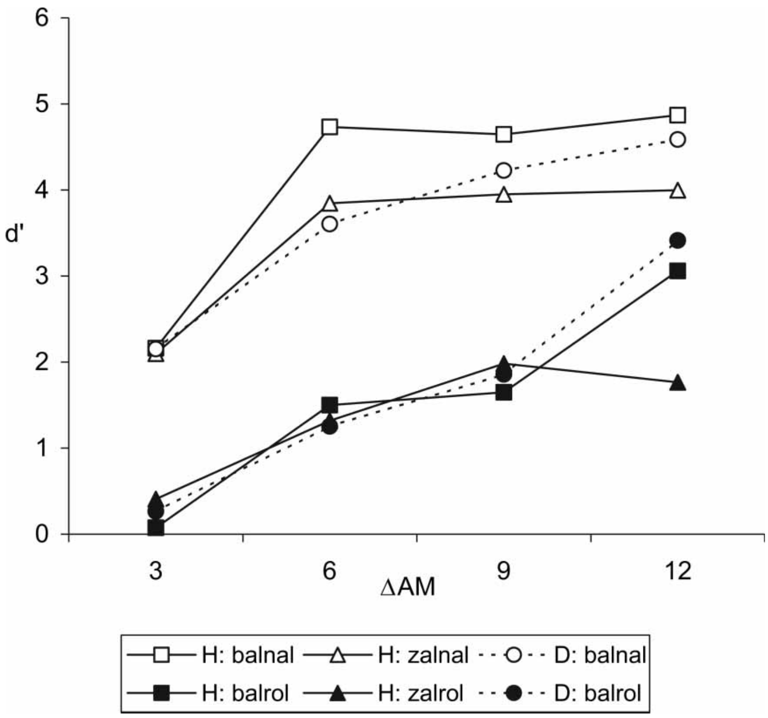

Figure 3. *Discrimination performance in Experiment 2. The open symbols represent the data for the unviable-context condition, and the closed symbols represent the unviable conditions. The squares represent the data of the Hungarian listeners with the Hungarian word, the triangles represent the data of the Hungarian listeners with the Hungarian nonword, and the circles represent the data of the Dutch listeners with the Hungarian word.*

effect of context, $F(1, 31) = 115.8$, $p < .001$, and a significant effect of ΔAM, $F(3, 93) = 49.0$, $p < .001$. These two within-subject variables did not interact significantly, $F(3, 93) = 1.7$, $p > .1$. Neither the main effect of listener group ($F < 1$) nor any of its interactions—by context, $F < 1$; by ΔAM, $F(5, 93) = 1.9$, $p = .08$; by context-by-ΔAM, $F < 1$—was significant.

The main effect of context shows that discrimination was significantly better in the unviable context ($d' = 3.76$) than in the viable context ($d' = 1.54$). In order to evaluate the main effect of ΔAM, post hoc tests (HSD, $p < .05$) were performed. This showed that the level 3-dB ΔAM ($d' = 1.19$) was significantly different from all other levels. The levels 6-dB ΔAM ($d' = 2.71$) and 9-dB ΔAM ($d' = 3.05$) did not differ significantly. In addition, discrimination performance was better at the maximal ΔAM ($d' = 3.62$) than at the level 6-dB ΔAM.

Discussion

We examined whether the perception of phonological changes is also influenced by context in a discrimination task. This was clearly the case: Deviations from a canonical form [Cɔl] were less salient in the viable context than in the unviable context. This was independent of the lexical status of the assimilated form and independent of experience with the assimilation rule in question.

The results bear out the predictions of the perceptual-integration account. This account predicted that discrimination of assimilated and canonical utterances should be worse in a context that allows assimilation than in a context that does not allow assimilation. Moreover, this effect should be independent of specific language experience, which is in agreement with our results.

These results are difficult to explain in terms of the phonological-inference account. First of all, this account predicts strong effects of language experience, and these were absent in the discrimination task. Moreover, the account argues that compensation is "undone" perceptually only after phonological features are extracted from the input. This implies that the phonological context of an assimilated utterance should not impede discrimination of an assimilated and canonical utterance.

However, to make a clear distinction between a perceptual-integration account and a feature-parsing account is more difficult. First of all, both accounts predict similar behaviour for listeners with and without experience with the assimilation rule. Moreover, both accounts predict a reasonably "early" locus of compensation for assimilation. If it is assumed that auditory stimulus representations not yet affected by feature parsing cannot be probed by a discrimination task, the feature-parsing account can accommodate the current results. To avoid such a "loophole", we made use of a discrimination task, which seems to induce discrimination based on auditory features only (Gerrits & Schouten, 2004). However, Gerrits and Schouten investigated discrimination performance with single vowels and CV utterances. There is recent evidence that discrimination performance in the 4I-oddity task is influenced by categorization performance if CVCCVC utterances are used (Kingston, in press). This result makes it more likely that discrimination performance is influenced by the feature-parsing process.

In this context, it is interesting to note that the feature-parsing account bears a resemblance to the theory of direct perception (see, e.g., Fowler, 1996). Within the framework of the theory of direct perception of speech, it is also assumed that the speech perception system "parses" context-dependent variation from a target and assigns it to its source (Fowler & Brown, 2000). In comparison, Gow (2003) argued that the evidence for the place of articulation of the assimilating segment is parsed from the assimilated segment and assigned to the assimilating segment. The theory of direct perception assumes that listeners do not perceive proximal acoustic signals, but the distal speech gestures that produce them. Accordingly, listeners have difficulty in detecting acoustic differences between stimuli, which arise from the same or similar gestures (Fowler & Smith, 1986). Failures to detect context-dependent variation in an appropriate context in a discrimination task (Fitch, Halwes, Erickson, & Liberman, 1980) are assumed to lend support to the assumption that speech gestures are the objects of perception (see Fowler, 1996, p. 1740). Hence, the current finding that context effects in the perception of assimilations are also evident in discrimination is in fact supportive for an account in terms of a theory of direct perception of speech gestures as well as motor theory (Libermann, 1996). A similar argument can be made for the feature-parsing account.

How can the perceptual-integration account then be distinguished from the feature-parsing account and the more general theories such as motor theory or direct perception? The perceptual-integration account makes different predictions from those of the feature-parsing account with regard to the effects of nonspeech sounds (see Fowler, Brown, & Mann, 2000, and Lotto & Kluender, 1998, respectively). According to the perceptual-integration account, the auditory processing of target and context integrates on early nonlinguistic levels of processing. Therefore, the "currency" in which integration occurs is an auditory one. That is, there must be an identifiable *auditory feature* of the context phoneme that leads to a perceptual integration. Whether this auditory feature occurs in a speech sound or in a nonspeech sound should therefore not matter. In contrast, the feature-parsing or a direct-realist account assumes that evidence for the *phonological* or *gestural feature* of the assimilating segment is parsed from the assimilated segment in order to be assigned to the assimilating segment. Therefore, the context must carry the appropriate phonological or gestural feature in order to be able to "catch" the evidence for the assimilating feature that is parsed from the assimilated segment (see Gow, 2003, Exp. 3). Thus, the feature-parsing account and gestural accounts on the one hand and the perceptual-integration account on the other differ in their assumption of the speech specificity of the processes involved. This implies that nonspeech context sounds should lead to similar context effects as those for speech sounds according to the perceptual-integration account, but not according to a feature-parsing or a gestural account.

The value of such demonstrations is, however, not undisputed. Fowler (1990, in press) argued that effects with nonspeech sounds may or may not mirror the effects of speech sounds, depending on what source is ascribed to the nonspeech sounds. In the framework of direct perception, listeners perceive any auditory input, including nonspeech stimuli, as a signal for a distal event. Depending on the distal event that the nonspeech sounds induce the listenere to perceive, nonspeech sounds may trigger similar or dissimilar effects as speech sounds do (see especially Fowler, 1990). Moreover, similar results obtained with speech and nonspeech sounds may not be taken as evidence that the same mechanisms underlie both effects. The similarity of results may just be coincidental. Nevertheless, recent neuroimaging results provide some arguments for the validity of the speech–nonspeech comparison. Scott and Wise (2003) showed that speech and nonspeech sounds are processed by similar structures in early auditory areas (i.e., core, belt, and parabelt in the nomenclature of Rauschecker, 1998) given similar acoustic structure. Only areas in the more frontal parts of the superior temporal lobes, which are less specifically involved in auditory perception, distinguish speech from nonspeech. Given the overlapping cortical areas of processing of speech and nonspeech, it is unlikely that similar results with, in some respect, acoustically similar speech and nonspeech sounds are caused by different perceptual mechanisms.

EXPERIMENT 3

In the first two experiments, we showed that Dutch and Hungarian listeners react to phonological changes in Hungarian utterances in a context-sensitive way: The difference between /l/ and /r/ is clearly perceived in a context that does not allow assimilation, but is difficult to make if the context allows assimilation. In Experiments 1 and 2, we found two effects that were independent of language experience. Listeners discriminated assimilated and nonassimilated word forms worse in a viable context that allowed assimilation, and identification functions were more shallow in the viable context. A language-specific bias in identification performance nonwithstanding, we now test whether the effects, which seem to be independent of language experience, are better explained in terms of a speech-specific account, such as the feature-parsing or a gestural account, or in terms of an auditory account, such as the perceptual-integration account. To this end, we test whether nonspeech context sounds are able to generate similar context effects to the speech contexts. In adherence to our earlier methodology, perception was probed by means of an identification and a discrimination task.

Method

Participants

A total of 10 students (9 female, 1 male) of the University of Maastricht participated in the experiment and were paid for participation. All participants were native speakers of Dutch, free of any known hearing problems, and right-handed. Participants were aged from 19 to 24 years (mean age 20.5 years). All participants were fluent speakers of the English language. In addition, some of the participants also spoke some German and French. One participant also spoke Indonesian. However, no participant had any knowledge of Hungarian.

Materials

The same speech targets were used as those in the previous experiments. The nonspeech context sounds were created in the following way. A 2-s stretch of white noise was generated and then convolved with the speech context sounds [naI] and [roːl] used in the Experiments 1 and 2. This generates a sound that approximates the long-term spectrum of the speaker, but does not contain any speech-like information. A 300-ms piece was extracted form the middle part of the edited 2 s of noise, and a linear fade-out was applied to the last 30 ms of the stimulus. The noise was then equated in mean amplitude with the speech context sounds. This stimulus was used as the nonspeech analogue of the context speech sound [naI], which does not allow the phonological change from [bɔl] to [bɔr]. This sound will be called −AM in order to indicate that this sound does not contain AM. In order to create a nonspeech analogue of the context [roːl], which allows the change to occur, the same AM technique as that for the target stimulus was applied to the noise to create a +AM stimulus that was identical to the −AM stimulus in every respect but the presence of AM.[1] These context sounds were concatenated with the speech stimuli in order to generate the experimental stimuli. The nonspeech context sound followed the target stimulus directly, just as did the natural-speech context sounds. These stimuli were presented to the participants sitting in a sound-attenuated booth over headphones (Sennheiser HMD 25−1) using the ERTS program (Behringer, 1996).

Procedure

The procedure was the same as that in Experiment 1 for the identification task and Experiment 2 for the discrimination task. Participants performed identification and discrimination in this order within one session.

Design

The design of the identification task entails the two independent variables AM in the target syllable and context with the levels +AM and −AM (cf. Experiment 1). Each of the 10 cells of this design was presented to each participant 16 times, and the dependent variable is the percentage of *bal* responses for every cell of the design.

The design of the discrimination task entails two independent variables ΔAM between standard and odd stimulus, with four levels from 3 to 12 dB, and context with the levels +AM and −AM. For each of the eight cells of this design, 10 trials each were presented with the odd stimulus in the second and third positions. From these 20 trials per cell, the dependent variable d' was calculated.

Results

Identification task

Table 2 shows the mean percentage of "bal" responses for every cell of the design. The descriptive data show that the identification function is steeper in the −AM condition than in the +AM condition. A repeated measures ANOVA performed on the these data revealed a significant effect of AM depth, $F(4, 36) = 40.77$, $p < .001$, but no main effect of context ($F < 1$). However, there was a significant interaction of AM depth and context, $F(4, 36) = 4.76$, $p < .01$, which shows that the +AM context led to a significantly shallower identification function.

Discrimination task

Table 3 shows the mean d' scores for every cell of the design. A repeated measures ANOVA performed on the these data revealed a significant effect of ΔAM, $F(3, 27) = 12.71$, $p < .001$, a significant main effect of context, $F(1, 9) = 16.38$, $p < .005$, and the interaction between these factors, $F(3, 27) = 3.78$, $p < .05$. The interaction is due to the fact that the effect of context was significant for all levels of the factor ΔAM—6 dB, $t(9) = 2.97$, $p < .05$; 9 dB, $t(9) = 3.04$, $p < .05$; 12 dB, $t(9) = 3.62$, $p < .02$—except the 3-dB level ($t^2 < 1$).

[1]The stimuli as well as the PRAAT scripts to generate them are available at http://www.mpi.nl/world/persons/private/holmit/proflinks.html.

Table 2. *Mean percentages of "bal" identifications in Experiment 3*

	AM depth in target				
Context	*0*	*3*	*6*	*9*	*12*
+AM	73.6	59.6	40.8	29.1	22.2
−AM	94.4	81.7	41.3	23.2	21.5

Note: AM = amplitude modulation (in dB).

Table 3. *Mean d′ scores in Experiment 3*

	ΔAM				
Context	*3*	*6*	*9*	*12*	*Mean*
+AM	0.32	1.13	1.52	1.33	1.08
−AM	1.26	3.56	4.06	3.90	3.20

Note: AM = amplitude modulation (in dB).

Discussion

The nonspeech sounds clearly influenced the perception of the speech sound continuum. Identification was more uncertain and discrimination more problematic when the context sound carried an AM. This mirrors the findings obtained in Epxeriments 1 and 2. Therefore, the present results support the assumption that compensation for assimilation rests to a rather substantial degree on general auditory perceptual principles.

The present results are difficult to reconcile with the assumptions of a feature-parsing account. This account assumes that the feature cues for in the assimilated [r] in the [bɔrroːl] are parsed from this segment and are attributed to the [r] in the case suffix [roːl]. The nonspeech targets used in this experiment probably did not carry evidence for phonological features; there should be no context to which the feature cues of the assimilated target can be assigned. Accordingly, there should be no effect of the nonspeech sounds on the perception of the phonologically changed forms. Similar arguments can be made for gestural accounts as the direct-realist theory or the motor theory. In contrast with these predictions, the present results show that there is a clear context effect with nonspeech sounds that mirrors the context effects caused by speech sounds.

One possibility to counter this argument is to argue that the context sound was too much like speech and therefore led to a preliminary detection of the features for an apical trill in the context. By design, the noise stimuli used had a similar long-term spectrum as that for speech. Therefore, it may be argued that such a noise, with added AM, sounds similar to a whispered /r/. Such a sound might lead to an activation of a phonological feature, which, in turn, would allow phonological feature parsing. To rule out such an alternative explanation, and in order to test the generality of the results of Experiment 3, we used a less speech-like sound in Experiment 4.

EXPERIMENT 4

In this experiment, we tested whether the results of Experiment 3 depend on the spectral composition of the carrier sound to which the AM is applied. Thus, we used a pure tone with a frequency of 400 Hz. In contrast to the sounds used in Experiment 3, the spectral composition of this sound is completely unlike speech. If the pattern of context effects as observed in Experiment 3 is replicated, this indicates that AM occurring in nonspeech sounds is perceptually integrated with the AM in the speech sounds, independent of the spectral composition of the context. Again, we tested the impact of these nonspeech sounds on the identification and discrimination of the target speech sounds.

Method

Participants

A total of 9 students of the University of Maastricht and 1 nonscientific staff member participated in the study. They were paid for participation. All participants were female, and all participants but one were right-handed. Of the participants, 9 were aged 20 to 24, and 1 participant was aged 50. All participants were native speakers of Dutch and were free of any known hearing impairment. All participants spoke English fluently. In addition, some of the participants spoke some German of French or both of these languages. One participant also spoke Spanish fluently. However, none of the participants was familiar with Hungarian.

Materials

The targeted speech stimuli were the same as those in Experiment 1. The −AM context sound was a 0.3-s pure tone with a frequency of 400 Hz. In order to prevent clicks, a linear fade-in and fade-out (10 ms) was applied to this nonspeech sound. The +AM sound was created by multiplying the −AM sound with the amplitude envelope of the +AM sound from Experiment 3. These nonspeech sounds were concatenated with the target speech sounds.

Procedure and design

Procedure and design were the same as those in Experiment 3.

Results

Identification

The results of the identification experiment, expressed in percentage "bal" responses are shown in Table 4. The results replicate the pattern observed in Experiment 1, showing that the identification function is shallower in the case of the +AM context. A repeated measure ANOVA revealed a significant effect of AM depth, $F(4, 36) = 33.9$, $p < .001$, but no main effect of context ($F < 1$). However, the interaction between the two factors is significant, $F(4, 36) = 3.08$, $p < .05$, replicating a similar effect in the previous experiment.

Discrimination

The results of the discrimination experiment, expressed in mean d' scores are shown in Table 5. A similar pattern as that in Experiment 1 is observed; discrimination performance was worse in the +AM condition. A repeated measure ANOVA revealed a significant effect of context, $F(1, 9) = 7.56$, $p < .05$, and a significant effect of ΔAM, $F(3, 27) = 17.52$, $p < .001$. The interaction between these two factors was not significant ($F < 1$).

Discussion

The results of the present experiment show that the context effect found in Experiment 3 does not depend on an overall spectral overlap between the AM in the speech target and the AM in the nonspeech context. In both the identification task and the discrimination task remarkably similar results have been obtained. The present results therefore indicate a perceptual integration of the AM in the speech sound target and the AM in the pure tone. It seems that amplitude modulations are perceptually integrated irrespective of the frequency range in which they occur.

However, there is one difference between the results of Experiment 3 and those of Experiment 4. In the previous discrimination experiment, there was an interaction between context and ΔAM: The effect of context was significant at all but the lowest level of ΔAM. In the present experiment, the overall significant effect of context did not interact with ΔAM. It should be noted that the overall level of discrimination performance was better in Experiment 2. Therefore, the

Table 4. *Mean percentages of "bal" identifications in Experiment 4*

	AM depth in target				
Context	*0*	*3*	*6*	*9*	*12*
+AM	67.4	48.6	23.3	13.2	11.3
−AM	78.9	55.1	15.0	4.3	3.1

Note: AM = amplitude modulation (in dB).

Table 5. *Mean d' scores in Experiment 4*

	ΔAM				
Context	*3*	*6*	*9*	*12*	*Mean*
+AM	0.53	2.37	2.80	3.58	2.32
−AM	1.65	4.34	4.48	4.51	3.82

Note: AM = amplitude modulation (in dB).

interaction between the independent variables context and ΔAM in Experiment 1 may be ascribed to a floor effect. At the lowest level of ΔAM, discrimination performance was near chance, which may have made it more difficult to detect a significant effect of context.

Nevertheless, Experiments 3 and 4 show that nonspeech stimuli are able to influence the perception of phonological changes. These results support a perceptual-integration account for compensation for assimilation. The results are difficult to explain in terms of the feature-parsing account (Gow, 2002, 2003) or within the frameworks of direct realism (Fowler, 1996) or motor theory (Liberman, 1996). These accounts assume that context effects only arise if the context carries evidence for phonological features or speech gestures. This might have been possible for the noise stimuli used in Experiment 3. However, it is difficult to see how the speech-perception system could mistake the AM in the pure-tone context used in Experiment 4 to indicate the presence of phonological features or a speech gesture. This buttresses the assumption that the perception of phonological changes is strongly influenced by basic auditory processes. In the next experiment, we aim to find converging evidence for this assumption.

EXPERIMENT 5

In this experiment, we modulated the frequency of a nonspeech context sound instead of its amplitude. Single-cell recordings in early cortical auditory areas have revealed that cells reacting selectively to a particular frequency of AM show a similar response to frequency-modulated tones (Wang, Liu, & Liang, 2003). This result seems counterintuitive at first glance. If one considers the tonotopic organization of auditory cortex, however, it is evident that frequency modulation (FM) and AM lead to similar effects on a given single cell. Consider a cell with a characteristic frequency of f Hz responding to a FM tone with f Hz at the centre of the FM. Given a sinewave form of the FM, this cell gets more or less the same input as for an AM tone with a modulation frequency of $2f$. This is due to the fact that the centre of the FM is passed through twice within one cycle of FM. For cells with a characteristic frequency at the lower or higher limits of the FM, the input will be similar to that for an AM tone and a FM tone with a modulation frequency of f. This follows from the fact that the upper and lower frequency limits are only passed through once within one cycle of FM. Therefore, we expect that FM may in principle also cause the same context effects as those for AM on the speech sound continuum.

Method

Participants

A total of 12 participants (9 female and 3 male) from the same subject pool as that in the previous experiments participated in this experiment. They were paid for their participation. A total of 2 of the participants were left-handed, and the rest were right-handed. The participants were aged 18 to 25 years (mean age 21.8 years). All participants were native speakers of Dutch and free of any known hearing impairment. All participants spoke English fluently. All but 1 participant had additional foreign-language proficiencies: A total of 8 participants spoke French, 4 of them fluently; 7 participants spoke German, of which 3 spoke fluently; 3 participants spoke Spanish, and 1 of these also spoke Surinamese. In addition, 1 participant spoke Frisian, a Germanic language still spoken in the northern part of The Netherlands. However, none of the participants was familiar with Hungarian.

Materials

The same speech targets and the stationary −AM stimulus as those in the previous experiments were used. This sound was a steady pure tone with a frequency of 400 Hz and thus a tone without FM. Hence, we call this sound the −FM sound. In order to create a +FM sound, FM at the rate of 50 Hz and a depth 100 Hz in a sinusoidal fashion was added. That leads to a minimum of 300 Hz and a maximum of 500 Hz, which leads a 0.95 bark range between centre (400 Hz) and minimum and a 0.89 bark deviation from centre to maximum. This is a rather strong FM depth, which is not too strongly skewed due to the nonlinear relation between the Hertz scale and perceived height.

Procedure and design

The procedure and design were the same as those in Experiment 3, with the only difference that the independent variable context now had the levels ±FM instead of the ±AM.

Results

Identification task

The results of the identification experiment as percentage "bal" responses are shown in Table 6. The results show that the identification function is shallower in the case of the +FM context. A repeated measure ANOVA revealed a significant effect of AM depth, $F(4, 44) = 79.74$, $p < .001$, but no main effect of context ($F < 1$). However, the interaction between the two factors is significant, $F(4, 44) = 4.20$, $p < .01$, replicating a similar effect in the previous experiments.

Discrimination task

The results of the discrimination experiment, expressed in mean d' scores, are shown in Table 7. As in the previous experiment, discrimination performance is worse in the context with a modulation, which is a FM here. A repeated measure ANOVA revealed a significant effect of context, $F(1, 11) = 10.63$, $p < .01$, and a significant effect of ΔAM, $F(3, 27) = 8.99$, $p < .001$. The interaction between these two factors was not significant ($F < 1$).

Table 6. *Mean percentages of "bal" identifications in Experiment 5*

	AM depth in target				
Context	*0*	*3*	*6*	*9*	*12*
+FM	83.9	65.2	33.3	16.7	9.9
−FM	93.8	71.4	24.5	6.8	6.3

Note: AM = amplitude modulation (in dB). FM = frequency modulation.

Table 7. *Mean d′ scores in Experiment 5*

	ΔAM				
Context	*3*	*6*	*9*	*12*	*Mean*
+FM	0.71	2.09	2.00	2.37	1.74
−FM	1.51	2.63	2.58	3.49	2.61

Note: AM = amplitude modulation (in dB). FM = frequency modulation.

Discussion

The present experiment succeeded in generating a similar context effect to that in the earlier AM experiments. The identification function was shallower in the case of the FM context than in the case of a steady-state context. As in the AM experiments, FM impaired discrimination, just as the original speech sound context did.

An interpretation favouring an auditory basis of the effect is fostered by an electrophysiological study (Mitterer, Csépe, & Blomert, 2003). This study showed that context effects in assimilated utterances could be detected with event-related potentials (ERPs) during passive listening. An ERP component called mismatch negativity (MMN), which indicates a perceptual distance between two stimuli (see, e.g., Näätänen & Winkler, 1999), displayed context sensitivity when presented with the phonological change [bɔl] → [bɔr]. The MMN was smaller if the change occurred in a phonological context that allowed the change (i.e., [bɔrroːl]) than if the change occurred in a context that did not allow the change (i.e., *[bɔrnal]). Especially interesting for the current purposes is the fact that the MMN inverted polarity at the mastoid electrodes, a known characteristic of a classic MMN. Although exact localization is difficult with ERP, the polarity reversal of the MMN may indicate a generation of the effect in early cortical auditory areas (Colin, Radeau, Soquet, Demolin, Colin, & Deltenre, 2002), which do not show a specialization for speech perception (cf. Scott & Wise, 2003).

An additional argument may be taken from the fact that FM and AM have been shown to lead to similar effects in tonotopically organized areas. Both AM and FM lead to an increase and decrease in energy at a particular frequency. Although both interpretations are admittedly speculative, they are buttressed by the report of Scott and Wise (2003) that results from functional-imaging studies indicate a similar early cortical localization for the processing of AM and FM.

GENERAL DISCUSSION

Gaskell and Marslen-Wilson (1996) showed that the perception of phonological changes is context sensitive: A phonological change does not impede the recognition of the changed word, if, and only if, the context allows the phonological change to happen as a consequence of a lawful phonological assimilation. In this paper, we conducted five experiments to examine the processes that drive this *compensation for assimilation.* We first replicated the compensation for assimilation result with a Hungarian assimilation rule: The Hungarian word /bɔl/ (English "left") is recognized in [bɔrroːl] but not in *[bɔrnal], because /l/ may be changed to /r/ in Hungarian before /r/ but not before /n/. In order to examine a possible lexical contribution of this effect, we use a nonword target. With the Hungarian nonword /zɔl/, the same results as with the word /bɔl/ were obtained. Moreover, we tested whether experience with the Hungarian language contributed to compensation for assimilation. Although Dutch listeners had no experience with the assimilation rule in question, they behaved similarly to Hungarian listeners. They had more difficulties in differentiating the forms [bɔl] and [bɔr] if these were followed by [. . . roːl], which allows liquid assimilation, than if they were followed by [. . . naːl], which does not allow liquid assimilation,. However, unlike Hungarian listeners, Dutch listeners were less likely to perceive an [l] in forms that were presented in the viable context [. . . roːl] than were Hungarian listeners.

A second experiment investigated whether similar context effects could be found by using a discrimination task. If context effects are not evident in a discrimination task—that is, canonical [bɔl] and [bɔr] are equally well discriminable in a viable and an unviable phonological context—then an auditory account for compensation for assimilation is difficult to maintain. However, the results indicated that [bɔl] and [bɔr] were less likely to be distinguished in the viable context [. . . roːl] than in the unviable context [. . . naːl]. This effect was again independent of the lexical status of the assimilated word and the experience of the listeners with the assimilation rule. This result indicated that compensation for assimilation to a large degree does not depend on experience with an assimilation rule.

The context effect observed in the discrimination task can be accounted for by a general-auditory account or a speech-specific account such as the feature-parsing account or the more general theories of speech perception, such as the motor theory, or the application of the theory of direct perception to speech perception. In order to distinguish these theoretical interpretations, three experiments using nonspeech sounds were conducted. Both the feature-parsing account and a direct-perception account argue that the phonological or gestural features supplied by the context trigger the context effect. Because nonspeech sounds do not carry any phonological or gestural features, they should therefore not trigger any context effect. In contrast with this prediction, all three experiments showed that nonspeech sounds cause similar context effects to those caused by speech sounds. Experiment 3 showed that the critical feature is the presence of AM in the context. For a context effect to occur, it does not matter whether this AM occurs in a speech sound, as in Experiments 1 and 2, or whether the AM occurs in nonspeech noise as in Experiment 3. Experiment 4 validated this finding by showing that this holds even if the speech sound and the context sound are spectrally dissimilar. Finally, Experiment 5 showed that a similar effect was obtained if a frequency-modulated context sound was used instead of an amplitude-modulated sound.

How do the three different accounts fare in the light of these data? The perceptual-integration account we proposed is clearly in line with the data. As predicted, compensation for assimilation does to a large degree not depend on experience with an assimilation rule (Experiments 1 and 2). Secondly, the context effects do not depend on the lexical status of the target (Experiments 1 and 2), nor do context effects vanish in a discrimination task (Experiment 2). The crucial prediction differentiating an auditory account from the more language-specific accounts was that nonspeech context sound should lead to similar effects to those for speech context sounds. This prediction was borne out. One obvious objection to this conclusion is that the context effects seem to diminish as the auditory properties of the nonspeech context deviate more from the auditory features of the speech sounds. The context effects were largest in Experiments 1 and 2 (speech context sounds), and got successively smaller as the context became less speech-like (Experiments 3 to 5). However, if auditory processing is responsible for the context effects, as we assume, the exact auditory make-up of the stimuli should influence the size of the context effects. Perceptual integration is conceivably less likely to occur as two sounds become less similar. Spectral similarity and periodicity information should therefore moderate the size of the context effect, just as they do in the different experiments. The main finding in Experiments 3 to 5 is that context effects occur with nonspeech sounds, which are qualitatively similar to the context effects caused by speech sounds.

An obvious question for the perceptual-integration account is why the human auditory system should be equipped with a processing mechanism that contributes to compensation for assimilation. It has been argued that such specific innate dispositions to solve rather specific problems are evolutionary quite unlikely (Elman et al., 1996). It is, however, possible that it is actually speech production that respects the constraints of the auditory system and thus adapts to these constraints. Consider the following analogy (from Clark, 1997): A space traveller with an ultra-adaptationistic view observing mankind might report that humans have evolved in order to sit comfortably on chairs. What this imaginary space traveller failed to notice is that chairs are artifacts that humans have adapted to fit them. Similarly, language, and thus speech production, is in some respect also an artifact, which is amenable to change. Therefore, it is conceivable that the system of phonological changes in a given language adapts to the auditory abilities of the listener (see also de Boer, 2000; Hume & Johnson, 2001; Hura, Lindblom, & Diehl, 1992; Kohler, 1990; Ohala, 1990; Seo, 2001; Steriade, 2001). In this view, we need not assume that aspects of the auditory system are extremely well adapted to the environment, but rather that languages use existing properties of the perceptual system. An affirmation of this conclusion would require evidence that such context sensitivities as reported here can also be found in nonhuman species. Context effects—such as compensation for speaking rate (Stevens, Kuhl, & Padden, 1988) and for coarticulation (Lotto et al., 1997)—have already been reported for nonhuman species. It remains to be seen whether similar effects can be obtained for compensation for phonological changes.

Just as the data are in line with the predictions from the perceptual-integration account, they disconfirm the predictions of the phonological-inference account (Gaskell, 2003; Gaskell & Marslen-Wilson, 1996, 1998). This account predicted that specific language experience is crucial for compensation for assimilation to occur, that compensation occurs at a phonological level of processing, and hence lower levels should be unaffected by context. Finally, nonspeech sounds should not be able to trigger effects akin to compensation for assimilation. Our data did not bear out these predictions. However, in Experiment 1, we observed some effects of specific language experience. Hungarian listeners showed a greater bias towards perceiving [l], especially in the context that allowed assimilation. We noted in the discussion of Experiment 1 that there are two ways to account for this result. Native listeners may have developed an asymmetric use of acoustic evidence, with little evidence needed to perceive an /l/, while clear evidence is needed to perceive /r/. Alternatively, perception may be biased by the phonotactic properties of the native language. This may also resolve the apparent conflict between the current finding that language-independent processes drive compensation for assimilation and Darcy's (2002) finding that specific language experience matters for compensation for assimilation. Darcy used a word-monitoring task, which obviously would reflect such language-dependent perceptual biases as observed here. This does not imply that the results by Darcy are artifacts of such higher level processes. Instead, these results reflect the use of higher level knowledge in resolving the partial ambiguity that arises as a consequence of phonological assimilation. Such higher level influences are more clearly visible with tasks that probe higher level processing. Accordingly, there is no effect of either lexical status or native-language experience in a discrimination task (Experiment 2), while effects of language experience arise in an identification task (Experiment 1) and word monitoring (Darcy, 2002). Lexical effects and discourse-level context effects were observed in phoneme-monitoring and cross-modal priming tasks (Gaskell & Marslen-Wilson, 1998, 2001). However, such experience-dependent effects build on language-independent context effects arising on an auditory level of processing. In line with this interpretation, Gaskell and Marslen-Wilson (2001) found an additive effect of phonetic and discourse context on compensation for assimilation.

Finally, how does the feature-parsing account—and related accounts derived from gestural theories of speech perception—fare in the light of the present data? These accounts can handle the finding that compensation for assimilation does to a great extent not depend on specific language experience (e.g., Fowler, Best, & McRoberts, 1990; Mann, 1986). With regard to the fact that context effects persist in a discrimination task, an interpretation is less straightforward. According to the theory of direct perception, speech is directly perceived in terms of gestures, and this is the only, to use a Gibsonian term, resonance of speech there is. Similarly, motor theory (Liberman, 1996) assumes that the "capture" of a sound by the speech-perception module prevents this sound from being perceived in auditory terms. Both theories leave no room for a possible auditory representation that could support discrimination of canonical and viably assimilated utterances. In the framework of feature parsing (Gow, 2002, p. 174), it is explicitly assumed that a surface representation arises in the course of speech perception, which reflects the difference between assimilated and canonical utterances. This surface representation should be able to allow context-independent discrimination of canonical and assimilated targets. Accordingly, the fact that compensation for assimilation is reflected in a discrimination task is not easily accounted for in the framework of the feature-parsing account.

The three accounts, which assume speech-specific but learning-independent compensation for assimilation (feature-parsing, direct-perception, and motor-theoretic accounts), however, cannot easily explain the fact that nonspeech sounds create similar context effects to those for speech sounds, because these accounts assume that the linguistic and not the auditory properties of the context trigger compensation for assimilation. Feature parsing, or parsing of articulatory gestures from the assimilated segment, is supposed to occur because the context carries articulatory or phonological features that attract this mismatching information. Accordingly, nonspeech sounds carry no phonological or articulatory features and should not trigger the context effects that they do trigger. However, as we noted, it may be argued that the effect of speech and nonspeech sounds are only coincidentally similar (Fowler, 1990, in press). Although this argument is difficult to dismiss, recent neuroimaging results (see, e.g., Scott & Wise, 2003) show that the same cortical areas process speech and acoustically similar nonspeech sounds.

It is therefore noteworthy that Gow (2003; Gow & Im, 2004) rejected an interpretation in auditory terms. He favours a feature-parsing account, because the perception of assimilation also induces progressive context effects, which seem difficult to capture in an auditory framework. Gow (2001) showed that a segment, which can trigger assimilation, is recognized faster—as measured by reaction times in a phoneme-monitoring task—if it is preceded by an assimilated segment than if it is preceded by an unchanged segment. This effect is also independent of specific language experience with assimilation rules (Gow & Im, 2004). The monitoring advantage for assimilating segments is explained by the assumption that, due to feature parsing, evidence for the upcoming segment is already available during the perception of the assimilated segment, but not if the preceding segment is not assimilated. This interpretation may be questioned on two grounds. Because phoneme monitoring is a task that presupposes segmentation skills that are not part of the speech-perception ability (Liberman, 1996, chap. 24), a task-specific explanation is also possible for the progressive context effect. Any task imposed on a subject gives rise to task-specfic stimulus–response mappings (Neumann, 1990). Hence, in asking a subject to monitor for a phoneme such as [d], listeners may set up stimulus–response mappings for auditory features associated with voicing. If such features then occur in the assimilated segment they may prime a detection response. This even occurs if this information is masked by later occurring stimuli (Klotz & Neumann, 1999; Vorberg, Mattler, Heinecke, Schmidt, & Schwarzbach, 2003). Accordingly, the existence of progressive context effects does not challenge a perceptual-integration account for compensation for assimilation. Furthermore, it should be noted that in order to achieve compensation for assimilation, the regressive context effect is crucial, because the regressive context effect prevents a mismatch between the assimilated form ([bɔr]) and the canonical form ([bɔl]), which is probably a part of the lexical representation (Gaskell & Marslen-Wilson, 1996; Gow, 2001).

In summary, the current results show that language-independent auditory processes contribute to compensation for liquid assimilation in Hungarian. Compensation for assimilation may proceed differently for different assimilatory processes (see especially Gow & Im, 2004). Indeed, strong language-dependent perception of phonological assimilations has been reported for palatal-to-velar fricative assimilation in German, triggered by the frontness of the preceding vowel (Weber, 2001), and place assimilation of moraic nasals in Japanese (Otake, Yoneyama, Cutler, & van der Lugt, 1996). While such results show that our result does not generalize to all other phonological assimilations, it may be possible to estimate for which phonological assimilations our results hold. The two cases above share at least two properties, which differentiate them from Hungarian liquid assimilation. First, the assimilations in question do not blur a phonemic contrast in the language in which they are applied. German does not distinguish a palatal and a velar fricative, and Japanese does not use place to distinguish nasals. Secondly, these assimilations are obligatory so that the sequences [ix] and [nb] are illegal in German and Japanese, respectively. Finally, both assimilation rules lead to rather gross acoustic changes. In particular, the distinction between palatal and velar fricatives is very salient (cf. Kohler, 1990). In contrast, assimilation rules for which compensation for assimilation has been observed—for instance, Hungarian liquid assimilation and place assimilation in nasal and stops—are optional and tend to be gradient (Browman & Goldstein, 1992), and the acoustic consequences of assimiltion seem less salient. Accordingly, auditory processes may only play a role in the perception of phonological assimilations that fit this latter description.

Our results also may not generalize to connected-speech processes other than phonological assimilations. In connected speech, stronger reductions than phonological assimilations occur, such as extreme reduction of forms (e.g., the Dutch phrase /ɪn idər xəfɑl/ *in ieder geval* "in any case" may be pronounced as [ifɑl], see Ernestus, 2000) or deletion of word-final /t/ (e.g., the Dutch word /kɑst/ *kast* "cupboard" may be pronounced as [kɑs]). Ernestus, Baayen, and Schreuder (2002) showed that word-specific knowledge and semantic context are crucial in recognizing extreme reductions. Mitterer and Ernestus (in press) showed that higher level processes play a greater role in compensation for /t/-deletion than in compensation for assimilation: Exactly the same manipulation of lexical status that failed to influence compensation for assimilation in the current paper and in the paper of Mitterer and Blomert (2003) influenced compensation for /t/-deletion.

Nevertheless, the current results show that language-independent auditory processes contribute to compensation for optional and gradient phonologcial assimilations. We propose that this compensation occurs because the auditory properties of assimilated and assimilating segments interact so that the perceptual difference between an assimilated and a canonical form is reduced in auditory processing. This, in turn, is in line with the assumption that speech production does to some degree respect basic auditory principles (see, e.g., Boersma, 1997; Hume & Johnson, 2001). However, auditory processing is not sufficient to explain compensation for assimilation. Auditory processing alone seems to lead to a decreased salience of a phonological contrast (here, contrast between [r] and [l]) in a context that blurs this phonological contrast by assimilation (here [r]). In addition, language experience seems to induce a bias towards the phonological category that undergoes assimilation, which is important for achieving complete compensation. By allowing such learned biases, a perceptual-integration account seems at present to be the best option to account for compensation for assimilation.

Original manuscript received 18 November 2004
Accepted revision received 27 April 2005
First published online 20 September 2005

REFERENCES

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, Germany: Technische Hochschule Darmstadt.

Boersma, P. (1997). *Functional phonology* [Dissertation]. Utrecht, The Netherlands: LOT.

Boersma, P., & Weenink, D. (2002). PRAAT 4.0: A system for doing phonetics with the computer [Computer software]. Amsterdam, The Netherlands: University of Amsterdam.

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

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

Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica, 49*, 155–180.

Clark, A. (1997). *Being there: Putting brain, body, and world together again.* Cambridge, MA: MIT Press.

Coenen, E., Zwitserlood, P., & Bölte, J. (2001). Consequences of assimilation for word recognition and lexical representation. *Language and Cognitive Processes, 15*, 535–564.

Colin, C., Radeau, M., Soquet, A., Demolin, D., Colin, F., & Deltenre, P. (2002). Mismatch negativity evoked by the McGurk–MacDonald effect: A phonetic representation within short-term memory. *Clinical Neurophysiology, 113*, 495–506.

Darcy, I. (2002). Online processing of phonological variation in speech comprehension: The case of assimilation. In S. Hawkins & N. Nguygen (Eds.), *Temporal integration is the perception of speech* (p. 32). Grenoble, France: ISCA.

de Boer, B. (2000). Self-organisation in vowel systems. *Journal of Phonetics, 28*, 441–465.

Delgutte, B. (1997). Auditory neural processing of speech. In W. J. Hardcastle & J. Laver (Eds.), *The handbook of phonetic sciences* (pp. 507–538). Oxford, UK: Blackwell.

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development.* Cambridge, MA: MIT Press.

Ernestus, M. (2000). *Voice assimilation and segment reduction in Dutch* [Dissertation]. Utrecht, The Netherlands: LOT.

Ernestus, M., Baayen, R. H., & Schreuder, R. (2002). The recognition of reduced word forms. *Brain and Language, 81*, 162–173.

Fitch, H. L., Halwes, T. E., Erickson, D. M., & Liberman, A. M. (1980). Perceptual equivalent of two acoustic cues for stop-consonant manner. *Perception & Psychophysics, 27*, 343–350.

Fowler, C. A. (1990). Sound-producing sources as objects of perception: Rate normalization and non-speech perception. *Journal of the Acoustical Society of America, 88*, 1236–1249.

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

Fowler, C. A. (in press). Compensation for coarticulation reflects gesture perception, not spectral contrast. *Perception & Psychophysics*.

Fowler, C. A., Best, C. T., & McRoberts, G. W. (1990). Young infants' perception of liquid coarticulatory influences on following stop consonants. *Perception & Psychophysics, 48*, 559–570.

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 and Performance, 26*, 877–888.

Fowler, C. A., & Smith, M. (1986). Speech perception as "vector analysis": An approach to the problems of segmentation and invariance. In J. Perkell & D. Klatt (Eds.), *Invariance and variability of speech processes* (pp. 123–136). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

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

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

Gaskell, M. G., Hare, M., & Marslen-Wilson, W. D. (1995). A connectionist model of phonological representation in speech perception. *Cognitive Science, 19*, 407–439.

Gaskell, M. G., & Marslen-Wilson, W. D. (1996). Phonological variation and inference in lexical access. *Journal of Experimental Psychology: Human Perception and Performance, 22*, 144–158.

Gaskell, M. G., & Marslen-Wilson, W. D. (1998). Mechanisms of phonological inference in speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 380–396.

Gaskell, M. G., & Marslen-Wilson, W. D. (2001). Ambiguity resolution and spoken word recognition: Bridging the gap. *Journal of Memory and Language, 44*, 325–349.

Gerrits, E. (2001). *The categorization of speech sounds by adults and children* [Dissertation]. Utrecht, The Netherlands: LOT.

Gerrits, E., & Schouten, M. E. H. (2004). Categorical perception depends on the discrimination task. *Perception & Psychophysics, 66*, 363–376.

Gow, D. W. (2001). Assimilation and anticipation in continuous spoken word recognition. *Journal of Memory and Language, 45*, 133–159.

Gow, D. W. (2002). Does English coronal place assimilation create lexical ambiguity. *Journal of Experimental Psychology: Human Perception and Performance, 28*, 163–179.

Gow, D. W. (2003). Feature parsing: Feature cue mapping in spoken word recognition. *Perception & Psychophysics*, *65*, 575–590.

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, USA.

Gow, D. W., & Im, A. M. (2004). A cross-linguistic examination of assimilation context effects. *Journal of Memory and Language, 51*, 279–296.

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

Hume, E., & Johnson, K. (2001). The interplay of perception and phonology. In E. Hume & K. Johnson (Eds.), *The role of speech perception in phonology* (pp. 3–26). New York, NJ: Academic Press.

Hura, S. L., Lindblom, B., & Diehl, R. (1992). On the role of perception in shaping phonological assimilation rules. *Language & Speech, 35*, 59–72.

Keppel, G. (1991). *Design and analysis: A researcher's handbook* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Kingston, J. (in press). Ears to categories: New arguments for autonomy. In S. Frota, M. Vigario, & M. Freitas (Eds.), *Proceedings of the First Conference on Phonetics and Phonology in Iberia.* Berlin, Germany: Mouton de Gruyter.

Kingston, J., & Macmillan, N. A. (1995). Integrality of nasalization and f1 in vowels in isolation and before oral and nasal consonants—a detection-theoretic application of the Garner paradigm. *Journal of the Acoustical Society of America, 97*, 1261–1285.

Klotz, W., & Neumann, O. (1999). Motor activation without conscious discrimination in metacontrast masking. *Journal of Experimental Psychology: Human Perception and Performance, 25*, 976–992.

Kohler, K. J. (1990). Segmental reduction in connected speech in German: Phonological facts and phonetic explanations. In W. J. Hardcastle & A. Marchal (Eds.), *Speech production and speech modelling* (pp. 69–92). Dordrecht, The Netherlands: Kluwer.

Ladefoged, P., & Maddieson, I. (1996). *Sounds of the world's languages.* Oxford, UK: Blackwell.

Lahiri, A., & Marslen-Wilson, W. (1991). The mental representation of lexical form: A phonological approach to the lexicon. *Cognition, 38*, 245–294.

Lahiri, A., & Reetz, H. (2002). Underspecified recognition. In C. Gussenhoven & N. Warner (Eds.), *Laboratory phonology 7* (pp. 637–676). Berlin, Germany: Mouton de Gruyter.

Liberman, A. M. (1996). *Speech: A special code.* Cambridge, MA: MIT 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 Acoustic Society of America, 102*, 1134–1140.

Macmillan, N. A., Kingston, J., Thorburn, R., Dickey, L. W., & Bartels, C. (1999). Integrality of nasalization and F-1. II. Basic sensitivity and phonetic labeling measure distinct sensory and decision-rule interactions. *Journal of the Acoustical Society of America, 106*, 2913–2932.

Mann, V. A. (1980). Influence of preceding liquid on stop-consonant perception. *Perception and 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. (1981). Influence of preceding fricative on stop-consonant perception. *Journal of the Acoustic Society of America, 69*, 548–558.

McQueen, J. M. (1996). Phonetic categorization. *Language and Cognitive Processes, 11*, 655–664.

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. (2003). Compensation for phonological assimilation in perception: Evidence from Hungarian liquid assimilation. *Proceedings of the 15th International Congress of Phonetic Sciences*, 2321–2324.

Mitterer, H., & Ernestus, M. (in press). Listeners recover /t/s that speakers reduce: Evidence from /t/-lenition in Dutch. *Journal of Phonetics.*

Näätänen, R., & Winkler, I. (1999). The concept of auditory stimulus representation in cognitive neuroscience. *Psychological Bulletin, 125*, 826–859.

Nearey, T. M. (1990). The segment as a unit of speech perception. *Journal of Phonetics, 18*, 347–373.

Neumann, O. (1990). Direct parameter specification and the concept of perception. *Psychological Research, 52*, 207–215.

Nolan, F. (1992). The descriptive role of segments: Evidence from assimilations. In G. J. Docherty & R. D. Ladd (Eds.), *Papers in laboratory phonology: II. Gestures segment, prosody* (pp. 261–280). Cambridge, UK: Cambridge University Press.

Ohala, J. J. (1990). The phonetics and phonology of aspects of assimilation. In J. Kingston & M. Beckman (Eds.), *Papers in laboratory phonology I: Between the grammar and the physics of speech* (pp. 258–275). Cambridge, UK: Cambridge University Press.

Olsson, M. (1992). *Hungarian phonology and morphology.* Lund, Sweden: Lund University Press.

Otake, T., Yoneyama, K., Cutler, A., & Van der Lugt, A. H. (1996). The representation of Japanese moraic nasals. *Journal of the Acoustical Society of America, 100*(6), 3831–3842.

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

Rauschecker, J. P. (1998). Cortical processing of complex sounds. *Current Opinion in Neurobiology, 8*, 516–521.

Repp, B. (1983). Bidirectional contrast effects in the perception of VC–CV sequences. *Perception & Psychophysics, 33*, 147–155.

Schouten, B., Gerrits, E., & van Hessen, A. (2003). The end of categorical perception as we know it. *Speech Communication, 41*, 71–80.

Scott, S. K., & Wise, R. J. S. (2003). PET and fMRI studies of the neural basis of speech perception. *Speech Communication, 41*, 23–24.

Seo, M. (2001). A perception-based study of sonorant assimilation in Korean. *OSU Working Papers in Linguistics, 55*, 43–69.

Siptár, P., & Törkenczy, M. (2000). *The phonology of Hungarian.* Oxford, UK: Oxford University Press.

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

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

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). New York, NJ: Academic Press.

Stevens, E. B., Kuhl, P. K., & Padden, D. M. (1988, November). *Macaques show context effects in speech perception.* Paper presented at 116th Meeting of the Acoustical Society of America, Honolulu, HI, USA.

Summerfield, Q., & Assmann, P. F. (1989). Auditory enhancement and the perception of concurrent vowels. *Perception & Psychophysics, 45*, 529–536.

Verstraeten, B., & Van de Velde, H. (2001). Sociogeographical variation of/r/in standard Dutch. In H. Van de Velde & R. van Hout (Eds.), *'r-atics: Sociolinguistic, phonetic and phonological characteristics of* /r/ (pp. 45–62). Brussels, Belgium: Etude & Travaux.

Vorberg, D., Mattler, U., Heinecke, A., Schmidt, T., & Schwarzbach, J. (2003). Different time courses for visual perception and action priming. *Proceedings of the National Academy of Sciences, 100*, 6275–6280.

Wang, X., Lu, T., & Liang, L. (2003). Cortical processing of temporal modulations. *Speech Communication, 41*, 107–121.

Weber, A. (2001). Help or hindrance: How violation of different assimilation rules affects spoken-language processing. *Language and Speech, 44*, 95–118.