

Inequality Across Consonantal Contrasts in Speech Perception: Evidence From Mismatch Negativity

Sonia A. Cornell

University of Konstanz and University of Munich

Aditi Lahiri

University of Oxford

Carsten Eulitz

University of Konstanz

The precise structure of speech sound representations is still a matter of debate. In the present neurobiological study, we compared predictions about differential sensitivity to speech contrasts between models that assume full specification of all phonological information in the mental lexicon with those assuming sparse representations (only contrastive or otherwise not predictable information is stored). In a passive oddball paradigm, we studied the contrast sensitivity as reflected in the mismatch negativity (MMN) response to changes in the manner of articulation, as well as place of articulation of consonants in intervocalic positions of nonwords (manner of articulation: [edi ~ eni], [ezi ~ eni]; place of articulation: [edi ~ egi]). Models that assume full specification of all phonological information in the mental lexicon posit equal MMNs within each contrast (symmetric MMNs), that is, changes from standard [edi] to deviant [eni] elicit a similar MMN response as changes from standard [eni] to deviant [edi]. In contrast, models that assume sparse representations predict that only the [ezi] ~ [eni] reversals will evoke symmetric MMNs because of their conflicting fully specified manner features. Asymmetric MMNs are predicted, however, for the reversals of [edi] ~ [eni] and [edi] ~ [egi] because either a manner or place property in each pair is not fully specified in the mental lexicon. Our results show a pattern of symmetric and asymmetric MMNs that is in line with predictions of the featurally underspecified lexicon model that assumes sparse phonological representations. We conclude that the brain refers to underspecified phonological representations during speech perception.

Keywords: speech perception, mental lexicon, phonology, place and manner of articulation features, mismatch negativity

In spoken language comprehension, the human brain must be able to resolve variation in the speech signal. A question of vital importance is how the perceptual system of a listener copes with the enormous variability in the speech input, be they rule-governed or genuine mispronunciations. Speech perception becomes even more challenging as the number of categories of speech sounds increases. For example, to decipher oral stop consonants in German and English (e.g., [b], [p], [d], [t], [g], and [k]), listeners have to distinguish between six phonemes, whereas Bengali listeners have to differentiate 16 phonemes. What could be a plausible strategy to help listeners minimize perception errors and assist in detecting phonemes even under worst case conditions such as a densely packed perceptual space and noisy environment?

One possible strategy could be to keep contrast sensitivity asymmetric, that is, making it easier for the perceptual system to

correctly detect, for example, a phoneme such as [d] after an [n], rather than vice versa, an [n] after a [d]. This strategy makes sense whenever it is possible to compensate for an inferior detection rate by employing another featural dimension to detect a contrast. Here, the overall detection rate across all possible contrasts is higher as compared with a situation in which the contrast sensitivity is kept symmetric for all relevant contrasts. A prerequisite for an asymmetric contrast sensitivity is that not all features that can be extracted from the speech signal are also represented in the mental lexicon. A model assuming that not all phonological information is fully specified is the *featurally underspecified lexicon* (FUL) model (Eulitz & Lahiri, 2004; Lahiri & Marslen-Wilson, 1991, 1992; Lahiri & Reetz, 2002, 2010). Contrary to the view that all variants the listener is exposed to are mentally stored (cf. Bybee, 2001; Johnson, 1997; Ranbom & Connine, 2007), propo-

Sonia A. Cornell, Department of Linguistics, University of Konstanz, Konstanz, Germany, and University Hospital Munich, Department of Child and Adolescent Psychiatry and Psychotherapy, University of Munich, Munich, Germany; Aditi Lahiri, Department of Linguistics, Philology and Phonetics, University of Oxford, Oxford, England; Carsten Eulitz, Department of Linguistics, University of Konstanz.

This research was partly supported by a grant from the German Research Foundation (DFG) awarded to Carsten Eulitz and Aditi Lahiri (EU 39/6-1

and 6-2) using equipment provided by the University of Konstanz and the DFG to support the SFB 471. We are very thankful to Henning Reetz for his help with the description of stimuli in providing input and scripts for the pitch and intensity contours, as well as the sound frequency values. We would also like to thank A. Bobrov, G. Salagan, and N. Bekemeier for the technical support and A. Wetterlin for further help on the article.

Correspondence concerning this article should be addressed to Carsten Eulitz, Department of Linguistics, University of Konstanz, D-78457 Konstanz, Germany. E-mail: carsten.eulitz@uni-konstanz.de

nents of more abstract lexical representations argue that not all variants are stored in the mental lexicon. Experimental findings showing an asymmetric sensitivity for certain but not all contrasts support models assuming underspecified representations. The FUL model assumes the underspecification of the place of articulation feature [coronal] (sounds are articulated with the tip and/or blade of the tongue) as well as of the manner of articulation feature [plosive] (release of air flow after a complete closure being made in the vocal tract). The present experiments compared a number of contrasts involving mentally underspecified and fully specified speech sounds. We show that whenever mentally underspecified speech sounds are contrasted and used as frequent stimuli in an oddball paradigm, the contrast sensitivity is asymmetric. However, symmetric contrast sensitivity is found whenever fully specified speech sounds are contrasted or when underspecified speech sounds are used as rare stimuli in an oddball paradigm.

Central to the FUL model is that irrespective of whether the variants are allophonic (so entirely predictable as in the case of American English flaps; Pitt, 2009) or neutralizing (i.e., the contrast is neutralized as in the case of assimilations of phonemes), the representation will assist in resolving the variation. Consequently, in an assimilatory context, the nasal consonant in *rain* is underspecified for its place of articulation feature [coronal], that is, it is not stored in the mental representation. As a result, the acoustic manifestations of [rain], [raim], or [rait] (*raindrop*, *rainbow*, *raincoat*) would all be possible variants of the underlying partially specified /n/ in the mental representation of the word *rain*.

A benefit of assuming that certain features in phonological systems are underspecified is a considerable reduction in the mental workload when dealing with variability in speech, which then also has consequences for the search and identification mechanisms. The more one stores, the more the recognition process has to deal with in order to ultimately identify the correct word.

A crucial difference between FUL and other models dealing with speech variants is that all other approaches depend on both experience and context. These models are able to explain how the human comprehension system compensates for variations resulting from speech assimilation of adjacent segments at a prelexical level. This ultimately means that a listener must have been exposed to the assimilated or partially assimilated versions of, for example, *hand* in appropriate contexts (as in work by Gaskell and his colleagues), or else the word could not be recognized in the variant **hambag* for *handbag*. This prerequisite is not necessary when using a model that allows underspecified features. The feature representation according to the FUL model is not just based on local contexts, but on a universal system of contrasts that play a role in diachronic and synchronic alternations of languages. Consequently, in this model, underspecification is context-independent and the reason behind postulating underspecified features like [coronal] is not solely based on assimilations. This feature is not only underspecified in possible assimilation contexts, for example, word finally, but also intervocalic word medially (Friedrich, Eulitz, & Lahiri, 2006) or utterance initially (Friedrich, Lahiri, & Eulitz, 2008). In these contexts, assimilation is not possible and thus could not account for the acceptance of variations as in **doy* for *boy* or **homey* for *honey*. It seems that certain word variations do not interfere less with lexical access during speech recognition.

The claim here is not that experience has no influence, but rather that there are certain inherent asymmetries in the phonological

system of languages that cannot be attributed to context effects alone (cf. Gaskell & Marslen-Wilson, 1996, 1998, 2001; Gaskell & Snoeren, 2008) or cues in the signal (cf. Gow, 2002, 2003). *Underspecification* simply infers that the lack of certain information in the mental lexicon can lead to asymmetric perception, which in turn helps the listener to resolve variation in any context—not only in environments that are prone to assimilation.

Further neurophysiological evidence for underspecification involving context-free variation in vowels can be found in Eulitz and Lahiri (2004). This study investigated the perception of three vowels, the coronal vowels [e] and [ø] and the dorsal (articulation with the dorsum of the tongue) vowel [o] in German, using a mismatch negativity (MMN) paradigm. When contrasting coronal and dorsal vowels, the same acoustic standard–deviant contrasts triggered asymmetric MMNs when they were reversed in their role as standards or deviants. The MMN asymmetry for similar acoustic/phonetic differences between pairs of vowels in isolation was claimed to reflect underspecified phonological representations in the brain. These differential MMN asymmetries recently have been replicated for the same vowels embedded in words and pseudo-words (Cornell, Lahiri, & Eulitz, 2011).

Although most experiments involving changes in segmental features have compared real words against mispronounced words by changing segmental features (e.g., initial *pig* ~ **tig*, final *pun* ~ **pum*), fewer studies have examined the consequences in activation by comparing minimally different word pairs such as *pin* ~ *tin*. Could one expect similar asymmetries in such instances? In earlier gating experiments by Nix, Gaskell, and Marslen-Wilson (1993) and Marslen-Wilson, Nix, and Gaskell (1995), word pairs were used such as *lake* ~ *late* in the contexts of *cruise* where *late* could become assimilated to *lake* preceding the word *cruise*. These studies found a coronal–noncoronal asymmetry. In a forced-choice task, the listeners responded to the stimuli equally as coronal or noncoronal. However, in a gating study in which the listeners were free to respond with whatever word they thought they were hearing, the results were different. When the vowel transition information was present but the release of the consonant was not yet heard, listeners surprisingly gave 52% coronal words (e.g., *late*) as responses although what they were hearing was *lake*. In contrast, the coronal words such as *late* were responded to correctly. The authors concluded that “the listeners’ perceptual interpretation is not linked in an obligatorily one-to-one manner to the phonetic input” and that “a phonetically unambiguous labial or velar gesture can be reinterpreted as a lexically coronal” (p. 296).

Results supposed to contradict predictions of the FUL model have been reported by Mitterer (2011). In a series of four eye-tracking experiments, he compared predictions about the fixation behavior to competitor words based on his optimal-perception account with predictions of the FUL model. For Experiments 1–3 in which the FUL model predicted differences, he found no statistically significant effects. However, for Experiment 4 in which the optimal-perception account predicted a difference and according to his interpretation the FUL model predicted no difference, a statistically significant effect was found. A closer look at his analysis reveals two crucial points that we do not understand: First, in Experiment 1, he observed a nonsignificant trend in the direction predicted by the FUL model. However, a much longer time interval was used only in Experiment 1 to parameterize the dependent variable. In Experiments 2–4, he used a 400-ms interval

starting at 200 ms after target onset, but in Experiment 1 the interval was twice as long. Moreover, as seen in Figure 2 of his article, there seems to be a visible difference between the critical conditions (as predicted by the FUL model) in the first half of that interval and almost no difference in the second half. Therefore, we are not sure that the treatment of the data for Experiment 1 was impartial because one can get the impression that a uniform parameterization for all the experiments might have led to a significant difference in Experiment 1, which would have been in favor of the predictions by the FUL model.

Second, in Experiment 4, he based his predictions of the FUL model on the scoring formula reported in Lahiri and Reetz (2002). The score was actually intended primarily for their automatic speech recognition system rather than an actual score for humans. This is because it would be extremely difficult to ensure how many features are correctly extracted by a human listener, whereas it is possible to do so for the automatic speech recognition system. Consequently, we have never used the algorithm to make precise hypotheses in our own experiments. Nevertheless, even if we accept the algorithm as Mitterer did, the formula does predict the asymmetry that in fact was obtained by Mitterer's Experiment 4. The algorithm takes into account features extracted from the signal as well as those from the lexicon (clearly stated in Lahiri & Reetz, 2002, 2010; see footnote 1). For Mitterer's experiment, the crucial feature is [voice], which is both extracted and represented for [b], but nothing parallel is extracted from [p]. This was misinterpreted by Mitterer and led to wrong predictions in his Experiment 4.¹ Taken together, in Experiment 4, the optimal-perception account predicted the same pattern of results as the FUL model, and reanalyzing the data of Experiment 1 using the same approach as in the remaining experiments may even change the outcome of Experiment 1. Consequently, the experiments reported in Mitterer may not contradict predictions of the FUL model in the way he implied.

Summing up, there is a considerable amount of literature reporting behavioral and neurophysiological evidence for the asymmetric behavior of [coronal] place of articulation in speech perception and production (cf. Cornell et al., 2011; Eulitz & Lahiri, 2004; Friedrich et al., 2006, 2008; Gaskell & Marslen-Wilson, 1996, 1998; Gumnior, Zwitserlood, & Bólte, 2005; Lahiri & Reetz, 2002, 2010; Snoeren, Gaskell, & Di Betta, 2009; Wheeldon & Waksler, 2004; Zimmerer, Reetz, & Lahiri, 2009). However, this issue is still a topic of controversy in the literature (e.g., Gaskell & Marslen-Wilson, 1996, 1998, 2001; Gow, 2002, 2003; Mitterer, 2003, 2011; Mitterer & Bloomert, 2003; Tavabi, Elling, Döbel, Pantev, & Zwitserlood, 2009). All studies report an asymmetry in word-nonword coronal-noncoronal pairs. Yet, the cause for the asymmetry is subject to different interpretations: Gaskell and colleagues (1996, 1998, 2001, 2008) claim that the cause is contextual inference, whereas Lahiri and colleagues (cf. Cornell et al., 2011; Eulitz & Lahiri, 2004; Friedrich et al., 2006, 2008; Lahiri & Reetz, 2002, 2010) maintain that representational issues are at play. In what follows, we examine nonword pairs, systematically varying different phonetic and phonological features, in situations in which mispronunciation cannot be caused by assimilation alone.

In the neurobiological experiments reported here, we examined consonantal manner of articulation as well as place of articulation contrasts in intervocalic positions, using the MMN technique as in Eulitz and Lahiri (2004). The article is organized as follows. We

first lay out why the MMN is appropriate for this type of research and then discuss the phonological contrasts we examined, justifying our assumptions regarding which of the features are underspecified.

Mismatch Negativity

The MMN is a component of event-related brain activity that is considered to be an automatic or preattentive change detection response of the brain because it reflects the brain's reaction to changes in the acoustic sensory input. The MMN can be elicited in the absence of the subject's attention (Näätänen, 1992; Schröger, 1996), suggesting that a preattentive echoic memory trace of the preceding stimuli is used as a template against which incoming sounds are compared. It also has been assumed to reflect long-term memory traces for native language phonemes (Dehaene-Lambertz, 1997; Näätänen et al., 1997), syllables (Shtyrov et al., 2000), and lexical representations of words (for review, see Pulvermüller & Shtyrov, 2006), as well as the size of native vowel inventories (Hacquard, Walter, & Marantz, 2007). Earlier studies also have demonstrated that the MMN is a robust measure to study aspects of feature specification of segments (e.g., Cornell et al., 2011; Eulitz & Lahiri, 2004; Phillips et al., 1995, 2000).

¹ The correct calculations of the scores according to Lahiri and Reetz, (2002, 2010) are the following:

$$\frac{(\text{No. of matching features})^2}{(\text{No. of features from the signal}) \times (\text{No. of features from the lexicon})}$$

The higher the score, the better the match. Below, we list all of the features we assume are part of the lexical representation and extracted from the signal of the initial consonant of the three words [baart] [paart] [taart].

Features in the lexicon

Word	Features in the lexicon	Number of features
/taart/	obstruent, consonantal	2
/paart/	obstruent, consonantal, labial	3
/baart/	obstruent, consonantal, labial, voice	4

Features from the signal

Word	Features from the signal	Number of features
[taart]	obstruent, consonantal coronal, plosive	4
[paart]	obstruent, consonantal labial, plosive	4
[baart]	obstruent, consonantal labial, voice, plosive	5

Now, if we compare the words as heard by the listener [baart] and [paart] with the lexical representation of /taart/, the following scores emerge:

/taart/ representation – [paart] signal

$$\frac{2^2}{4 \times 2} = 4/8 = 0.5.$$

/taart/ representation – [baart] signal

$$\frac{2^2}{5 \times 2} = 4/10 = 0.4.$$

As one can see, the score for [b] to /t/ is 0.4, and [p] to /t/ is 0.5. That is, [b] > /t/ is a worse match than [p] > /t/.

Reliable MMNs are obtained in the oddball paradigm, in which the subject is presented with occasional infrequent, deviant stimuli within a series of standard stimuli (Schröger, 1998). The auditory oddball can also be described in terms of a regular relationship between the sounds in which the standard is a repetition of a frequent sound pattern, a regularity representation, and the deviant reflects the violation of this regularity (Winkler, 2007).

In speech perception, it can be assumed that the sound percept created by the deviant corresponds more to the surface representation of speech, whereas the repeated processing of standard stimuli before the deviant activates representations with an information structure closer to that in the mental lexicon, that is, that of the underlying representation. The change detection due to the deviant reflects—besides the acoustic change—the comparison of the surface representation with the underlying mental representation preactivated by the standard (Eulitz & Lahiri, 2004).

Crucially, certain featural differences such as strident (sounds made by directing a stream of air with the tongue toward the sharp edge of the teeth) for fricatives can only be studied using consonantal speech sounds. Although consonants have been used in a variety of MMN experiments (for review, see Näätänen, 2001; Pulvermüller & Shtyrov, 2006), to answer phonologically motivated research questions, fine-grained featural differences in consonants have been used in only a handful of studies. One group of studies examined categorical perception differences between place of articulations by manipulating F2 or F2 and F3 transitions for /ba/, /da/, /ga/ in several steps (from five to nine; e.g., Maiste, Wiens, Hunt, Scherg, & Picton, 1995; Shafer, Schwartz, & Kurtzberg, 2004). The basic finding for all of these experiments was that the MMNs for across-category differences were invariably higher than for within-category differences. A further study by Bonte, Mitterer, Zellagui, Poelmans, and Blomert (2005) examined the relevance of phonotactic probabilities between intervocalic [tk/ft/ts] in Dutch. The results supported the view that phonotactic probabilities were reflected in the MMN amplitude. Our focus in these experiments was on intervocalic consonants, varying both place and manner of articulation, examining the relevance, if any, for the theory of underspecification.

We are concerned with the consonantal contrasts *place* ([coronal] ~ [dorsal]) and *manner* ([strident] ~ [nasal]). The notion of underspecified representations is not only relevant for [coronal], but is applicable to other phonological features as well (cf. Scharinger, Reetz, & Lahiri, 2009, for vowel height features). A basic assumption in phonological representation is that neither consonants nor vowels are indivisible wholes; rather, they are made up of features. For consonants, these are distinguished in terms of place and manner of articulation features. Consider the initial consonants of the English words *do*, *go*, *zoo*, and *no*. The sounds [d] and [g] are voiced stops denoted by the feature [plosive], the latter has a [dorsal] place of articulation and the former [coronal]. The fricative [z] is a [strident] and [n] is [nasal] (allowing air to escape through the nose). The consonants [d], [n], [z] share the same [coronal] place of articulation but differ with respect to features pertaining to manner of articulation. Table 1 lists the surface and underlying features of the relevant consonants according to the FUL model.

Not all of these properties have the same status, that is, there are asymmetric relationships among them. Besides observing asymmetries between the features [coronal] and [dorsal], the manner

Table 1
Place and Manner of Articulation Features of the Four German Consonantal Segments

Consonant	Surface representation extracted from the signal		Underlying lexical representation activated in the lexicon	
	Place	Manner	Place	Manner
g	dorsal	plosive	dorsal	[]
d	coronal	plosive	[]	[]
n	coronal	nasal	[]	nasal
z	coronal	strident	[]	strident

Note. The second column gives the features that are extracted from the speech signal, and the right column lists the features in the underlying representation and how they are stored according to the FUL model. [] = features assumed to be underspecified in the mental lexicon.

features [plosive] and [nasal] are also asymmetric in their pattern of alternations, and we assume that this is also reflected in their mental representations. First, [plosive + nasal] sequences hardly exist word medially in English and German. There are no [bm] sequences in either language, no [dn] sequences in German, and very few [gn] sequences in either: *kidney*, *magnet* (English); *Magnet* (German). Second, [nasal + plosive] sequences such as [nd], [mb], or [ng] can often become all nasal, but not the reverse. One often hears this in American English dialects in words such as *winter* and *tendency* when the two consonants are followed by an unstressed vowel (Gussenhoven, 1986). Historically, we also find evidence for [nasal + plosive] sequences becoming [nasal], but never [plosive]. For instance, *number* in English corresponds to *Nummer* in German in which the [mb] sequence is totally assimilated; *lamb* in both languages is pronounced without a final [b]. We assume that [plosive] is underspecified, whereas [nasal] is specified.

Neurobiological studies up to now have shed light mainly on the underspecification of place of articulation showing asymmetries between [coronal] and [dorsal] or [labial] (e.g., Cornell et al., 2011; Eulitz & Lahiri, 2004; Friedrich et al., 2006, 2008), but have hardly considered any other featural dimension. The present study was designed to fill this gap and investigate the status of the representation for other features relevant for consonants and closely examine whether the underspecification assumption is indeed a basic principle in the functional organization of the mental lexicon. The features we focus on here are the place features [coronal] and [dorsal] as well as the manner features [nasal], [strident], and [plosive]. Based on asymmetries in diachronic and synchronic alternations, we again propose asymmetries in representation and processing. Thus, we assume that [coronal] and [plosive] are underspecified, and [dorsal], [nasal], and [strident] are not. We now turn to the experiments.

Experiment 1

We manipulated consonants in two conditions contrasting place of articulation in stops (coronal/dorsal), and within coronal sounds, the manner of articulation (nasal/strident) was also manipulated. The experimental stimuli were embedded in identical nonword VCV structures using the sequences [edi] ([coronal]) versus [egi]

([dorsal]) and [eni] ([nasal]) versus [ezi] ([strident]). An important extension to the place contrast was the manipulation of the manner of articulation, which was experimentally tested here for the first time. To ensure that the consonantal variation effects could not be attributed to assimilation, we used nonword medial consonants that thus require disyllabic stimuli.

According to our hypothesis, [coronal] place of articulation and [plosive] manner of articulation are underspecified, whereas the other features are specified in the mental representations. The feature details of the phonological representations of the consonants under study are given in column 3 of Table 2. The fourth column lists the features that are extracted from the signal, and the final column lists the predicted levels of conflict between phonological features in surface and underlying representation assuming underspecification. [nasal] and [strident], for instance, conflict with each other. Because [coronal] is unspecified, when [dorsal] is extracted from the signal, it does not conflict with coronal sounds. This is not the case in the opposite directions, however: [coronal] conflicts with [dorsal]. Here, we predicted an asymmetry of MMNs within the reversal of consonantal contrasts, presented as standard and deviant. In line with other studies (Cornell et al., 2011; Eulitz & Lahiri, 2004), we expected larger MMN amplitudes in conditions with a feature conflict compared with conditions with no featural conflict.

In the first two conditions, place features were contrasted: In the [d]_{/g/} condition (for labeling, see Table 2), the feature [coronal] is extracted from the deviant [d] and conflicts with the mentally specified [dorsal] sound, which has been preactivated by the standard /g/. In the reversed condition, that is, [g]_{/d/}, a nonconflicting situation occurs in that the feature [dorsal] is extracted from the deviant [g] and maps onto the preactivated representation of the standard /d/, which contains no information about the place of articulation. Consequently, the MMN amplitude for these acoustically equalized conditions [d]_{/g/} (conflict) and [g]_{/d/} (nonconflict) should be larger for the conflicting condition compared with the nonconflicting condition; that is, an asymmetric MMN pattern is expected.

More symmetry in MMN responses is expected for [n] and [z] of our manner contrast because both features, [nasal] and [strident], are assumed to be fully specified for their manner of articulation. Consequently, they should conflict equally in both directions. Thus, the [n]_{/z/} as well as the [z]_{/n/} condition should result in similar MMN amplitudes.

Mental lexicon models assuming full specification would predict equal MMNs between all contrasting consonants with varying place and manner features ([coronal], [dorsal], [plosive], [nasal], or [strident]) because it is assumed that all phonological information is fully stored (Bybee, 2001). Because our stimuli were designed in a way that context remained constant—they all had the same vowel contexts—these models would presumably predict equal MMN activation patterns. Differences in MMN would then probably be attributed to acoustic characteristics.

Method

Subjects. Twenty-nine students from the University of Konstanz participated in our study. All were native German speakers with normal hearing who passed a standard hearing screening evaluating individual hearing thresholds for 200–8000 Hz (Audio Console Version 2.4.8; Oscillo Hearing Instruments, Inmedico, Denmark). They all had normal or corrected vision and no history of neurological or psychiatric disorders. All subjects were paid for their participation or received course credit. The data of three subjects were excluded from all subsequent analysis because of excessive eye movement artifacts. Of the remaining 26 participants (age range 20–32 years, mean age = 24.92 years, *SEM* = 0.65; 13 women), all were right-handed as assessed by the Edinburgh Handedness Inventory (Oldfield, 1971).

Stimuli. The experimental stimuli used as standard and deviant for the critical comparisons of Experiment 1 were the VCV syllables [egi], [edi], [eni], and [ezi] (see Figure 1), all of which are nonwords in German. Both vowels surrounding the critical consonants are acoustically long. Multiple repetitions of these stimuli along with the disyllable [epi] were recorded by a male speaker. The stimuli were recorded on a digital recorder at a sampling rate of 44 kHz. The recordings were digitized and edited into individual segments using the speech analysis program PRAAT (Boersma & Weenink, 2007).

Three variants of each stimulus type with matched F0 frequencies were used in the experiment. All were cross-spliced and prepared in the following way. Three [e] vowels from the syllable [epi] were cross-spliced on to three variants of the syllables [-di], [-gi], [-ni], and [-zi]. The [e] vowels were taken from the syllable [epi] to prevent further coarticulation effects because [p] has a different place of articulation from the critical consonants. This resulted in nine VCV syllables, all with the same three [e] vowels

Table 2
Predictions Made About the Amount of Feature Conflict in the Four Experimental Conditions of Experiment 1

Contrast	Experimental conditions [deviant] _{/Standard/}	Features in the lexical representation (activated by the standard)		Features in the surface representation (extracted by the deviant)		Hypotheses based on full specification	Hypotheses based on underspecification
		MoA	PoA	MoA	PoA		
a	[g]_{/d/}	[]	[]	[plosive]	[dorsal]	conflict	nonconflict with []
	[d] _{/g/}	[]	[dorsal]	[plosive]	[coronal]	conflict	conflict with [dorsal]
b	[n]_{/z/}	[strident]	[]	[nasal]	[coronal]	conflict	conflict with [strident]
	[z] _{/n/}	[nasal]	[]	[strident]	[coronal]	conflict	conflict with [nasal]

Note. The standard–deviant contrasts are identical in each pair of inversions. The first pair of inversions, (a) [g]_{/d/} ~ [d]_{/g/}, reflects a change in place of articulation; the second pair, (b) [z]_{/n/} ~ [n]_{/z/}, reflects a change in manner of articulation. The row in bold type highlights the only completely nonconflicting condition [g]_{/d/} assuming underspecification. MoA = manner of articulation; PoA = place of articulation; [] = underspecified feature.

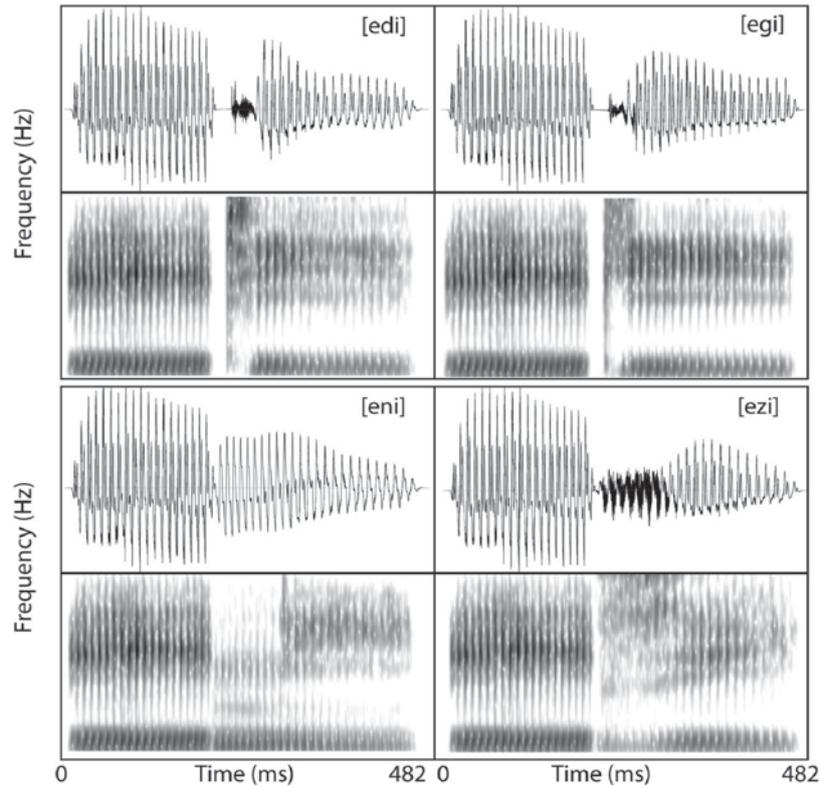


Figure 1. Oscillograms (above) and spectrograms (below) of the acoustic stimuli used in Experiment 1: [edi], [egi], [eni], and [ezi]. The standard and deviant stimuli were maximally matched for their acoustic properties in all conditions.

attached to three variants of the second syllables. This way, acoustic differences between the stimuli were minimized to reduce the likelihood that any observed MMN differences were caused by a particular acoustic feature. Furthermore, pitch and intensity measures were controlled as far as possible. Each set of VCV syllables showed comparable variation in pitch and intensity (see Figure 2). Pitch and intensity were equal between [edi] and [egi]; they varied to a large extent between [eni] and [ezi]. This was closely watched in the outcome of the MMN, however.

All items were controlled for length—200 ms for [e], 281–283 ms for the second syllable, giving a total duration of 481–483 ms. It is important to note that the consonantal onset varied slightly between the [plosive] consonants ([d], [g]) and between the [nasal] [n] and [strident] [z]. The place and manner of articulation information of the [n] and [z] began right after the offset of the [e] after 200 ms as is common in natural speech. However, the plosive, [d] started with a period of silence after the offset of [e] because there is usually no prevoicing in German. The closure duration lasted 20 ms before the burst onset; consequently, the onset of the place and manner information began slightly later compared with the [n] and [z]. However, because the two plosive sounds, as well as the [nasal] and [strident] sounds, were being compared with each other, but not among each other, that is, the comparison remained within the same category, these latency differences did not create a problem.

All stimuli were equalized for intensity (root-mean-square power [RMS]) to ensure minimal acoustic differences between the

stimuli. Because the RMS varied naturally between the experimental stimuli, an additional loudness rating was conducted to ensure identical perception between the consonants. For this purpose, 27 subjects listened to all contrasts via headphones with a gradually differing loudness (e.g., [eni] with 70 dB was contrasted with [edi] with 66 dB, 68 dB, and 70 dB, and vice versa). Each subject was asked to judge whether each pair sounded equally loud. Based on these perception results, the following amplitudes were chosen to induce equal perceptual loudness between the sound contrasts: 74 dB for the initial vowel [e], 68 dB for [di] and [gi], 70 dB for [ni] and 66 dB for [zi]. In addition, the intensities of the initial and final 20 ms of each stimulus were reduced linearly (linear fade in, linear fade out; Adobe Audition 2.0) to ensure a smooth onset and offset of the stimuli. By using three variants of each stimulus type, we introduced some acoustic variability to simulate more natural speech perception, thus forcing the processing system to map the acoustic signals onto more abstract representations (Eulitz & Lahiri, 2004). All four sounds were evaluated for their word-medial frequencies (see Table 3) by calculating their log-values from cumulative frequency counts weighted for word frequency based on the CELEX corpus (Baayen, Piepenbrock, & Gulikers, 1995). The consonantal sounds differ in their individual frequency counts in medial intervocalic position (V[C]V), with [n] and [g] being fairly similar ([n] log-value 2.97, [g] log-value 2.48), both being slightly higher than [d] (log-value 0.91), and [z] with the lowest frequency (log-value 0.37; see second column in Table 3). In addition, distributional frequencies of the speech sounds (phonot-

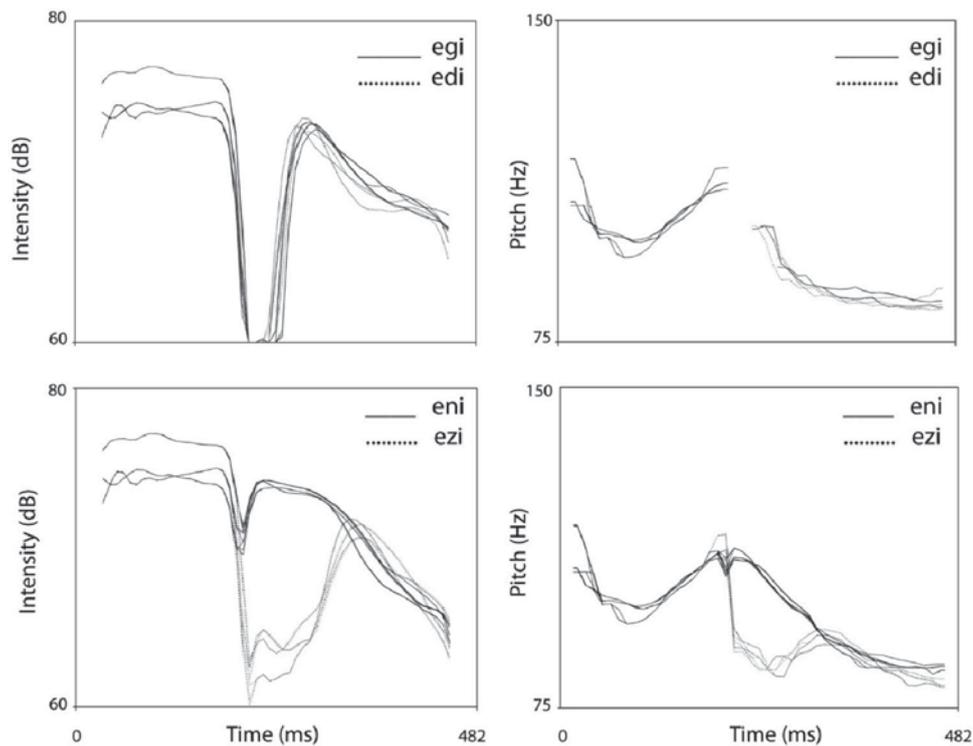


Figure 2. Waveforms show the pitch and intensity contours of the VCV syllables used in Experiment 1 plotted in their contrasts [edi] ~ [egi] and [eni] ~ [ezi]. Three variants for each VCV syllable were selected. The plosive place contrast [edi] ~ [egi] is very similar in pitch and intensity; the manner contrast [eni] ~ [ezi] differs quite substantially in pitch and intensity. Note that the three [e:] sounds are the same across all four VCV syllables.

actic probabilities) were checked for the first critical vowel–consonant ([eC]) syllables (see third column in Table 3). Bonte et al. (2005) showed that the distributional probabilities of phoneme clusters have an influence on the processing of speech as well as the amplitude characteristics of the MMN. The phonotactic probabilities of our stimuli were determined by log-values of the cumulative frequency counts of phoneme sequences weighed for word frequency of the German word form database (CELEX corpus; Baayen et al., 1995). The [eC] sequences differed as well,

Table 3
Log-Values of the Frequency Counts of Phonemes and Phoneme Sequences in the Studied VCV Syllables

Nonword stimuli	V[C]V	[eC]V	CV[C]VC
[eni]	2.97	3.12	4.48
[ezi]	0.37	1.12	4.03
[edi]	0.91	1.48	4.02
[egi]	2.48	0.26	4.26

Note. Frequency counts were weighed for word frequency and are based on token counts of the CELEX corpus. [...] = phonemes for which counts are given in the table; C = consonant; V = vowel; e.g., V[C]V of the second column gives the *individual* frequency counts for the four consonants in medial position; [eC]V of the third column shows the phonotactic probability counts for [en], [ez], [ed] und [eg] with a vowel following. CV[C]VC in the fourth column gives the frequency counts for /n/, /z/, /d/, and /g/ in an intervocalic position with further sounds before and after the vowels (e.g., words such as *Ge[n]jus* “genius” or *Ge[n]us* “genus”).

with [en] having the highest probability and [eg] the lowest probability (log-values: [en] 3.12 > [ed] 1.48 > [ez] 1.12 > [eg] 0.26). For words that had consonants initially and finally, that is, in the environment of CV[C]VC, the frequencies of the critical consonants hardly differed from each other (see fourth column in Table 3).

Procedure. Each participant was presented with a passive oddball paradigm while their electroencephalogram (EEG) was recorded. The experimental session consisted of four blocks. In each block, nonwords were combined pairwise, with each nonword serving as a standard as well as a deviant. This made it possible to compare MMN responses with equalized contrasts and subtract the same stimuli (identity MMN) to guarantee the same physical differences between the sounds. During the study, 680 standards (85%) and 120 deviants (15%) were presented for each of the four sequences. The stimuli were presented using a fixed interstimulus interval (ISI) of 500 ms binaurally via headphones (stimulus onset asynchrony [SOA] 981–983 ms). A pseudorandom stimulus sequence was created so that there were at least three and at most eight standard stimuli between two deviants ($M = 5.5$). The order of blocks was counterbalanced across subjects.

Data acquisition and analysis. Participants were seated in an electrically and acoustically shielded chamber and instructed to watch a silent movie of their choice and ignore the nonwords presented to them via headphones. EEGs were continuously recorded (Brain Vision, Brain Amp DC; 2 × 32 Ch.) from 64

electrode positions (plus one ground electrode; Easy Cap, Montage M10% to 10% System) against Cz as a reference and using a sampling rate of 500 Hz. Interelectrode impedances were kept below 5 k Ω . Vertical and horizontal eye movements (EOGs) were co-registered bipolarly with an additional electrode located on the forehead to correct the EEG raw data for EOG artifacts using the algorithm implemented in Brain Electric Source Analysis (BESA; MEGIS Software GmbH, Gräfelfing, Germany). During the study, subjects were asked to sit quietly and avoid excessive eye and other movements.

Further offline data processing included a digital bandpass filter set to 1–30 Hz and a standardization from 64 channels to 81 channels. Recording epochs with EOGs were corrected, and whenever the maximum voltage criterion exceeded 85 μ V at any EEG channel epoch, other recording artifacts were rejected before averaging. Channels were interpolated that were very noisy (e.g., due to muscle artifacts). Maximal seven electrodes were interpolated with maximal two neighboring electrodes. For each participant, the averaged MMN responses contained at least 75% accepted deviant trials in each condition. Standards immediately following deviants were not included in the analysis. All responses were re-referenced offline against right and left mastoids for further analysis (Schröger, 1998).

Event-related potential (ERP) averages were computed using epochs of 800 ms in duration including a 100-ms prestimulus baseline. The MMN was obtained by subtracting the ERP response of the standard from that of the equivalent deviant stimulus across blocks, for example, [edi] as deviant minus [edi] as standard of the reversed block, resulting in a same-difference waveform, the so-called *identity MMN*. This method provides a reliable measure not confounded by variation in ERP morphology that may result from physical stimulus differences per se.

Based on visual inspection of the grand-average waveform, we used the peak amplitude of the MMN waveform of every individual condition for parameterization. The MMN was expected around 150–250 ms after the offset of the first syllable (change onset; the set of initial syllables was identical for all conditions), that is, in a time window around 350–450 ms.

For statistical analysis, we used the MMN peak latency and the mean amplitude of the MMN as dependent variables. The peak latency was determined as the latency of the most negative point in the difference waveform of the Fz electrode of each subject for each condition. The mean amplitude of the MMN was measured over a 40-ms time window of the Fz electrode (rereferenced against linked mastoids) centered at the peak latency of the averaged difference waveform, separately for each subject for all four conditions.

The statistical analyses (using SPSS Statistics 15.0) were restricted to paired *t* tests because more complex factorial designs were difficult to formulate. Full balancing across factors was impossible for the present study in which different featural dimensions were varied and other possible confounds (physical differences between stimuli, number of conditions to signal to noise relationship, etc.) had to be avoided. Furthermore, the statistical model was restricted to two independent variables, the two pairs of inversion with an equal acoustic change subjected to two paired *t* tests ($\alpha = .05$) with the pair of inversion PLACE CHANGE ([g]_{/d/} ~ [d]_{/g/}) and MANNER CHANGE ([z]_{/n/} ~ [n]_{/z/}). The *t* values together with the

uncorrected degrees of freedom and the corrected *p* values are reported.

Results

MMN responses in the expected time window of the grand-average difference waveforms were significant in all four conditions, largest at the Fz electrode, but with enormous variations in size: [d]_{/g/}: $t(25) = -5.37$; [g]_{/d/}: $t(25) = -2.13$; [n]_{/z/}: $t(25) = -10.38$; [z]_{/n/}: $t(25) = -10.18$; $p < .05$ (see Figures 3, 4, and 5).

The MMNs for the asymmetrical PLACE CHANGE contrast ([g]_{/d/} ~ [d]_{/g/}) were significantly different, $t(25) = 2.55$, $p < .017$. As predicted by the underspecification hypothesis, the asymmetrical pair of inversion for place of articulation showed a larger MMN in the conflict condition, when the place feature [coronal] of the deviant [d] mapped onto the preactivated specified feature [dorsal] of the standard /g/ in the underlying form. A smaller MMN amplitude in the reversed condition [g]_{/d/} was found, when the [dorsal] deviant mapped onto the standard /d/, which is assumed to be underspecified for [coronal] (see Figures 4a and 5a).

The MMN amplitudes of the bidirectional conflicting manner conditions MANNER CHANGE ([n]_{/z/} ~ [z]_{/n/}) did not differ from each other, $t(25) = -0.43$, $p > .1$, *ns*, as expected, having equal manner conflicts in both directions of presentation, one conflict for the feature [nasal] and one for [strident] (see Figures 4b and 5b).

The interesting point is whether the above-reported differential MMN effect between the PLACE CHANGE and the MANNER CHANGE is systematic in nature. To examine this question, we calculated and statistically compared the differences between the contrasts. The analyses revealed a significant difference between the PLACE CHANGE and MANNER CHANGE contrasts, $t(25) = 2.35$, $p < .026$, showing largely enhanced MMN amplitudes for the MANNER CHANGE contrast (see Figures 4 and 5).

Latencies were further analyzed for the PLACE CHANGE and MANNER CHANGE contrasts. The MMN latencies of the two plosive sound conditions did not differ from each other as expected, $t(25) = 1.58$, $p > .1$, *ns*. Similarly, there was no significant difference for the MANNER CHANGE contrast, $t(25) = -1.93$, $p < .065$, *ns*.

Discussion

This first MMN experiment examined the fine structure of phonological specifications with respect to the underspecification of the feature [coronal] with stop consonants and the sensitivity of the MMN using manner features. This was done as a further step in generalizing the results for place of articulation in vowels to other sounds such as stops (in medial word position) and feature dimensions (manner of articulation). The underspecification of the [coronal] place of articulation has been demonstrated in other MMN studies for isolated vowels (Eulitz & Lahiri, 2004) and vowels embedded in linguistically more complex structures (Cornell et al., 2011), as well as for sounds in different word positions (e.g., medial word position: Friedrich et al., 2006; initial word position: Friedrich et al., 2008). When embedded into vowel contexts, the consonantal contrasts under study showed supporting

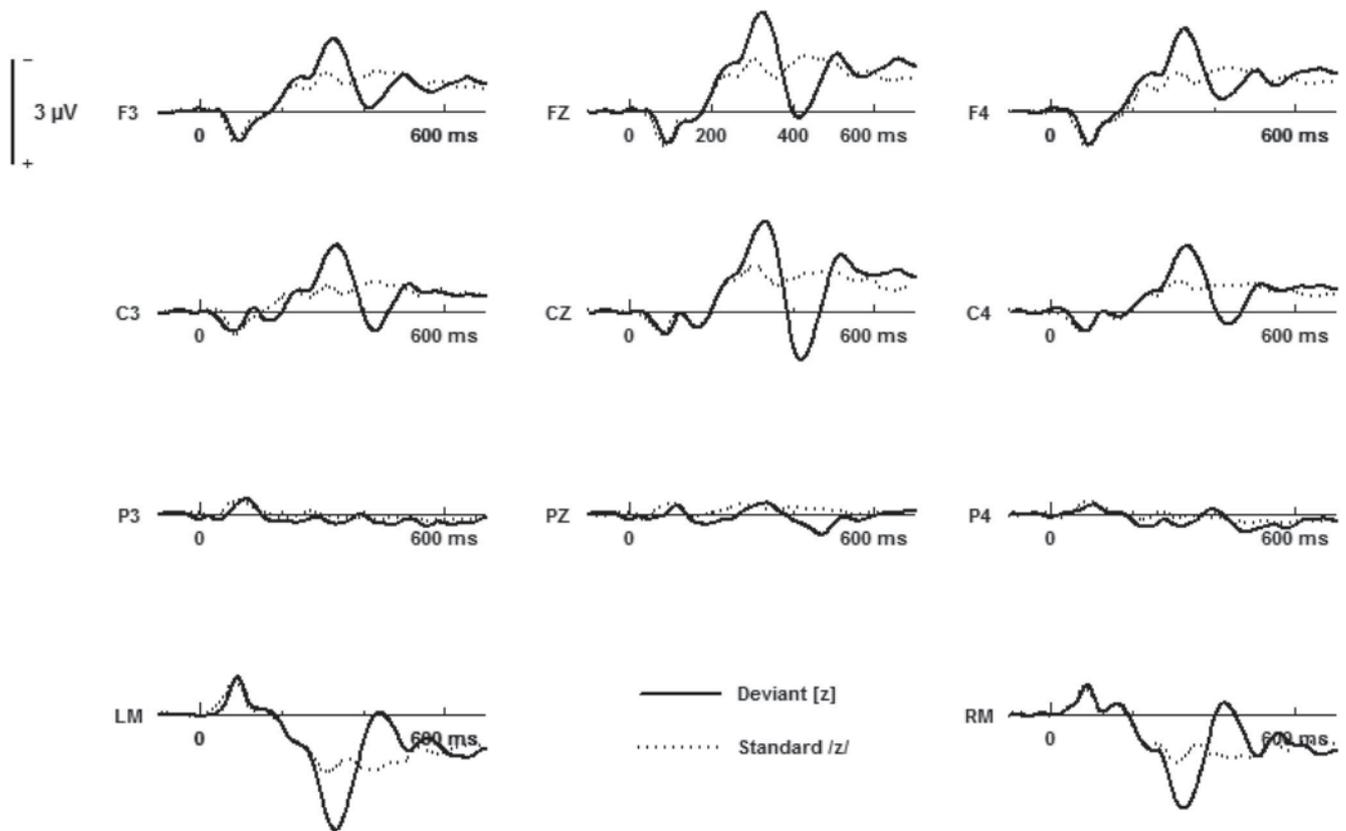


Figure 3. Sample grand-average waveforms of 11 electrode positions for the manner conflicting condition $[z]_{/n/}$. Figure plots the event-related potentials of the standard $/z/$ (dotted line) and deviant $[z]$ (solid line) waveforms. Note that the MMN $[z]_{/n/}$ is calculated by the difference waveform of the same stimuli $[z]$ across blocks in the context of $[n]$. The effect is largest at Fz and Cz, with the typical MMN reversal at the mastoids (Lm, Rm; for presentation purposes, these waveforms are average referenced).

evidence for the underspecification of the [coronal] place of articulation in stop consonants. Larger MMN amplitudes were found when a place of articulation conflict occurred compared with a nonconflicting condition. According to the FUL model (Lahiri & Reetz, 2002, 2010), a conflict in this study occurred when [coronal] place of articulation information was extracted from the acoustic signal when the deviant $[d]$ was heard between repetitions of standard stimuli, that is, $/g/$, which would have preactivated the [dorsal] place of articulation. In the reversed case, when $[g]$ was the deviant and $/d/$ was the standard, no conflict was expected. Models that assume fully specified phonological information would predict similar conflicts for all directions of change, resulting in similar MMN response patterns (Bybee, 2001). What inference-based models such as those of Gaskell and colleagues (Gaskell, 2003; Gaskell et al., 1995; Gaskell & Marslen-Wilson, 1998) would predict is difficult to know because the sound changes here were not context-dependent.

Our present findings concur with the predictions of the FUL model. We found the expected asymmetrical MMN pattern for the place of articulation conflict \sim nonconflict contrast ($[d]_{/g/} \sim [g]_{/d/}$), that is, higher amplitude when the place feature [coronal] from the $/d/$ is mapped onto the [dorsal] from the $/g/$ in the mental lexicon, and symmetrical MMN amplitudes for

our manner condition, the conflict \sim conflict contrast ($[n]_{/z/} \sim [z]_{/n/}$).

It is interesting that the phonetic contrasts evoked MMNs, which differed in the two feature dimensions under investigation. The manner contrast produced much larger MMNs than the place contrast, which also shows that the MMN indeed is a sensitive measure for studying manner features of articulation. Another explanation for the significantly smaller MMNs for place contrast might be the fact that they are stop consonants that are acoustically less salient than nasal and strident sounds. These findings are backed up by previous studies examining phonetic contrasts between stop consonants that manipulate mainly place of articulation (Diesch & Luce, 1997; Maiste et al., 1995; Pettigrew, Murdoch, Kei, et al., 2004; Pettigrew, Murdoch, Ponton, et al., 2004; Shafer et al., 2004). For example, Pettigrew, Murdoch, Ponton, et al. (2004) found poor MMN responses to their fine acoustic speech contrasts $[de/ge]$ and $[day/gay]$ and proposed that when carefully controlled methodological designs and strict methods of analysis are applied, robust MMN responses to fine-grained phonetic contrasts may be difficult to obtain. Studies by Dalebout and Stack (1999) and Dalebout and Fox (2000) also found that MMN

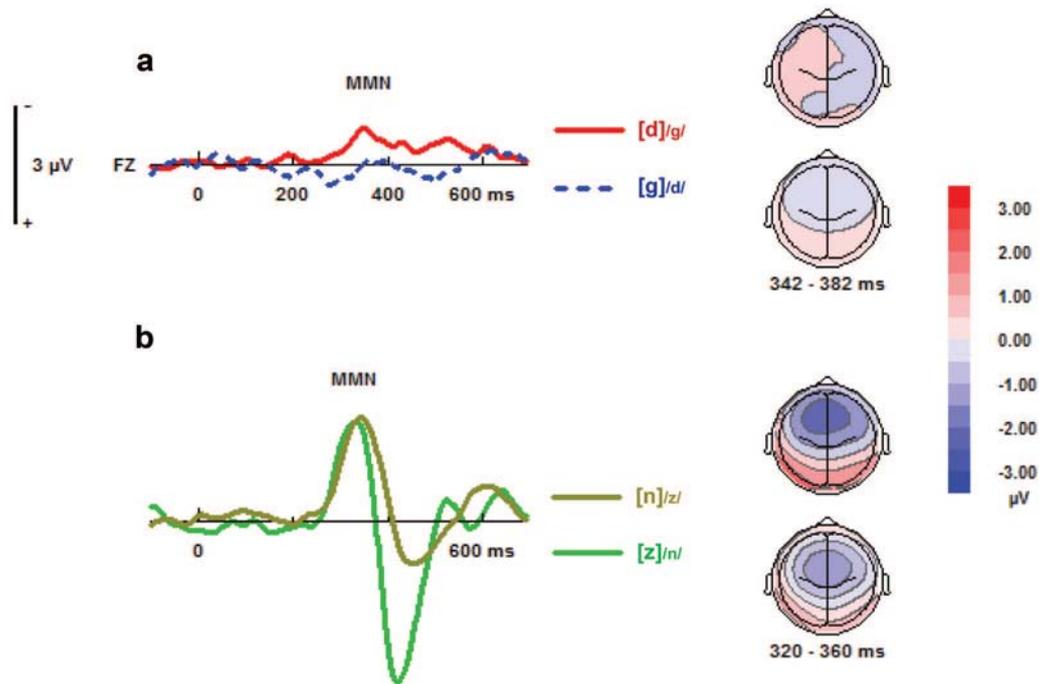


Figure 4. Grand-average difference waveforms of all four conditions at Fz (rereferenced against the right mastoid [Rm], the place change pair of inversion, (a) [d]_{/g} (red) and [g]_{/d} (blue, dotted), and the manner change pair, (b) [z]_{/n} (olive) and [n]_{/z} (green), with their topographical maps (right). Note that the only nonconflicting condition is shown in dotted lines.

responses to fine acoustic contrasts such as [d/g] may not always be obtained. Nevertheless, we did find small MMNs for our fine-grained place feature contrasts in the medial position with the predicted asymmetrical response pattern, which suggests that even in minimal acoustic contrasts, the assumed difference in mental representation can be detected.

In sum, MMN asymmetries were found for pairs of consonants with varying place features such as [dorsal] ~ [coronal], taking a further step in generalizing the underspecification of

the feature [coronal] to other speech sounds. Furthermore, MMN symmetries were found for pairs of consonants with varying manner features ([nasal] ~ [strident]).

We now turn to another contrast using the same feature dimension, this time contrasting [plosive] and [nasal], while keeping place constant. As we noted in Table 1, [plosive] is underspecified in the mental lexicon but [nasal] is not. Consequently, we again expected an asymmetry in MMN patterns, but this time for the manner of articulation.

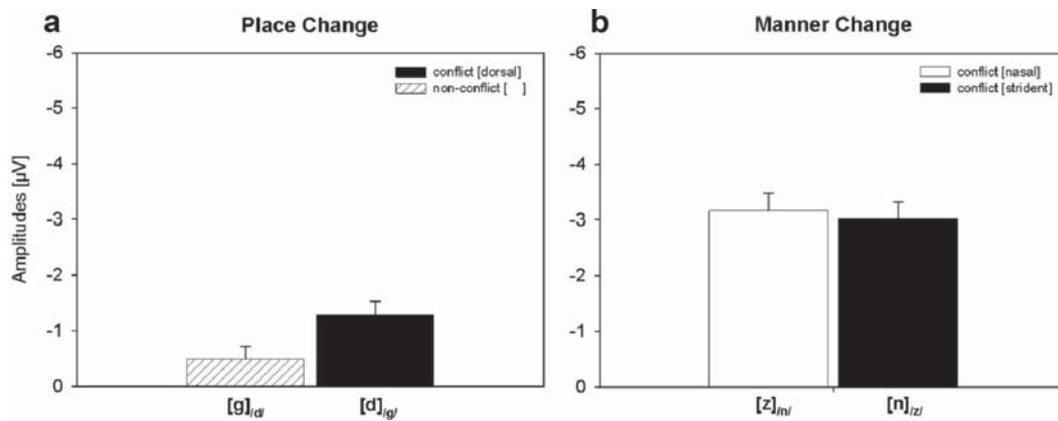


Figure 5. Mismatch negativity (MMN) mean peak amplitudes of all four conditions at Fz, the place change (a) and the manner change contrast (b). The striped bar indicates the only nonconflicting condition. Error markings on top of the bars represent the standard error of the mean.

Experiment 2

In this experiment, we contrasted [edi] versus [eni] and repeated the other contrast [eni] versus [ezi] to ensure that we could replicate our findings (see Table 4). We applied the same logic as in Experiment 1. Because [plosive] is unspecified, when [nasal] or [strident] are extracted from the speech signal, they do not conflict with it in the underlying representation. Within the comparisons of our second experiment, manipulating only manner of articulation features—in the [n]_{/d/} condition when the feature [nasal] extracted from the deviant [n] maps onto the preactivated representation of the standard /d/, which contains no information about the manner of articulation—a nonconflicting situation occurs. In the reversed condition, [d]_{/n/}, the feature [plosive], extracted from the deviant [d], would conflict with the mentally specified [nasal] created by the standard /n/. Consequently, the MMN amplitude for these acoustically equalized conditions [d]_{/n/} (conflict) and [n]_{/d/} (non-conflict) should be larger for the conflicting condition compared with the nonconflicting condition; that is, an asymmetric MMN pattern would be expected. As for the second manner contrast [nasal] and [strident], we again expected similar and symmetric MMN patterns between [n]_{/z/} and [z]_{/n/} because both manner features conflict in both directions of stimulus presentation. The predictions of the conditions in Experiment 2 are laid out in Table 4.

Method

Subjects. A total of 26 subjects took part in the second EEG study. They were selected with the same handedness, psychological, neurological, and perceptual criteria as in Experiment 1. Data from three subjects had to be excluded because of excessive EOGs, cardiac, or muscle artifacts. Thus, the following analysis is based on the remaining 25 participants (age range 19–31 years; mean age = 25.44 years, *SEM* = 0.65; 13 women). None of the subjects in Experiment 1 participated in Experiment 2.

Stimuli. The experimental stimuli used as standard and deviant for the critical comparisons in Experiment 2 were the same VCV syllables as in Experiment 1, [eni], [ezi], and [edi] (see Figure 1) plus a new disyllable [egi].

Again, pitch and intensity measures were controlled for as far as possible; however, they varied considerably between the contrasts (see Figure 6).

It is important to remember that the consonantal onset varies slightly between the [plosive] consonant ([d]) and between the [nasal] [n] and [strident] [z] because of the silent period at the beginning of the [d]. We were aware that this may add to latency differences in the MMN response for the condition that contrast [nasal] and [plosive] consonants. This was, however, considered in the parameterization of the data.

Procedure. Each participant was presented with a passive oddball paradigm as in Experiment 1 while their EEG was recorded. The experimental session consisted of six blocks, with four blocks containing pairs of the nonwords under study (see Table 4) and two blocks serving as a control condition for another study.² In each experimental block, nonwords were again combined pairwise, with each serving as a standard as well as a deviant. During the study, 680 standards (85%) and 120 deviants (15%) were presented for each of the four sequences, with an SOA of

981–983 ms. The experimental sequences, therefore, were equal to Experiment 1.

Data acquisition and analysis. The data acquisition of Experiment 2 was the same as for Experiment 1, in which participants were seated in an electrically and acoustically shielded chamber and instructed to watch a silent movie of their choice and ignore the nonwords while their EEGs were recorded.

The offline processing of the data also corresponded to that of Experiment 1. For each participant, the averaged MMN responses contained at least 77% accepted deviant trials in each condition.

For the statistical analysis, the MMN peak latency and the mean amplitude of the MMN were again used as dependent variables in a time window around 350–450 ms.

The statistical analyses were again restricted to paired *t* tests because more complex factorial designs were difficult to formulate. The comparisons were restricted to two pairs of inversion with an equal acoustic change. Paired *t* tests ($\alpha = .05$) for the pairs of inversion MANNER CHANGE ASYMMETRY, (c) [n]_{/d/} vs. [d]_{/n/}, and MANNER CHANGE SYMMETRY, (b) [z]_{/n/} vs. [n]_{/z/}, as control condition, were calculated.

Results

A clear MMN in the expected time window of the grand-average difference waveforms was observed in all four experimental conditions: [d]_{/n/}: $t(22) = -10.90$; [n]_{/d/}: $t(22) = -10.23$; [n]_{/z/}: $t(22) = -8.92$; [z]_{/n/}: $t(22) = -8.39$; $p < .000$ (see Figures 7 and 8). The grand-average difference waveforms of the two contrasts can be seen in Figure 7. Figure 8 shows the mean MMN peak amplitudes for both contrasts as bar graphs. The MMN amplitudes of the bidirectionally conflicting manner features MANNER CHANGE SYMMETRY, (b) [n]_{/z/} and [z]_{/n/}, did not differ from each other, $t(22) = -0.535$, $p > .1$, *ns*. This was expected when manner conflicts in both directions, that is, for the feature [nasal] after [strident], as well as for the feature [strident] after [nasal] in their functions as standard and deviant. The MMNs for the asymmetric manner contrast MANNER CHANGE ASYMMETRY revealed significant effects, (c) [d]_{/n/} versus [n]_{/d/}: $t(22) = 3.122$, $p < .005$. As predicted by the FUL model, the asymmetric manner pair shows a larger MMN in the conflict condition, when the manner feature [plosive] of the deviant [d] maps onto the preactivated specified feature [nasal] of the standard /n/. A reduced MMN amplitude was found in the reversed condition [n]_{/d/}, when the [nasal] deviant maps onto the underspecified [plosive] standard.

The question again asks whether the above-reported differential MMN effect between the MANNER CHANGE SYMMETRY and the MANNER CHANGE ASYMMETRY is systematic in nature. The differences between the contrasts were calculated and statistically compared to address this point. The analyses revealed a significant difference between the MANNER CHANGE SYMMETRY, (b) [nasal] \sim [strident], and the MANNER CHANGE ASYMMETRY, (c) [nasal] \sim [plosive], contrast, $t(22) = 2.276$, $p < .033$.

² A further VCV syllable [egi] was used as standard and deviant in two blocks for control purposes. These conditions are not directly related to the present research question and the results will be reported elsewhere.

Table 4

Predictions Made About the Amount of Feature Conflict in the Four Experimental Conditions of Experiment 2

Contrast	Experimental conditions [deviant] _{/Standard/}	Features in the lexical representation (activated by the standard)		Features in the surface representation (extracted by the deviant)		Hypotheses based on full specification	Hypotheses based on underspecification
		MoA	PoA	MoA	PoA		
c	[n]_{/d/}	[]	[]	[nasal]	[coronal]	conflict	nonconflict with []
	[d] _{/n/}	[nasal]	[]	[plosive]	[coronal]	conflict	conflict with [nasal]
b	[n] _{/z/}	[strident]	[]	[nasal]	[coronal]	conflict	conflict with [strident]
	[z]_{/n/}	[nasal]	[]	[strident]	[coronal]	conflict	conflict with [nasal]

Note. The standard–deviant contrasts are identical in each pair of inversions. The first pair of inversions, (c) [n]_{/d/} ~ [d]_{/n/}, reflects an asymmetric change in manner of articulation; the second pair, (b) [z]_{/n/} ~ [n]_{/z/}, reflects the symmetric change in manner of articulation under the assumption of underspecified mental representations. The row in bold type highlights the only completely nonconflicting condition [n]_{/d/} assuming underspecification. MoA = manner of articulation; PoA = place of articulation; [] = underspecified feature.

Because the stimuli had natural timing differences in the onsets between the [plosive] sounds and between the [nasal] and [strident] sounds (see Method, Stimuli subsection), which might have influenced the MMN latency responses for the [nasal] ~ [plosive] contrast [d]_{/n/} ~ [n]_{/d/}, we conducted further latency analysis; however, there were no significant latency differences found between asymmetric manner contrast MANNER CHANGE

ASYMMETRY, [d]_{/n/} versus [n]_{/d/}: $t(22) = -2.021, p > .1, ns$. The minimal onset change between the plosive and the nasal sound did not seem to affect the MMN latency. Furthermore, the [nasal] and [strident] contrast also showed no difference in latency, [n]_{/z/} versus [z]_{/n/}: $t(22) = 0.031, p > .1, ns$, as expected, which corresponds to the fact that they did not differ in their onset of change.

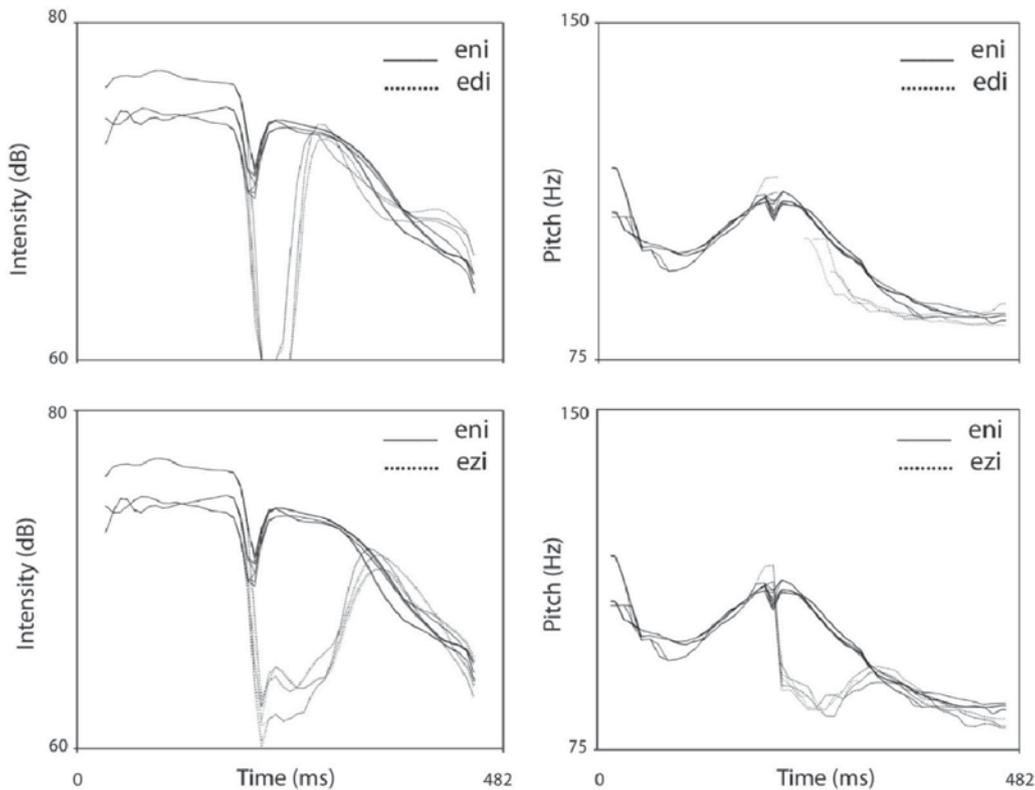


Figure 6. Waveforms show the pitch and intensity contours of the VCV syllable contrasts used in Experiment 2: [edi] ~ [eni] (above) and [eni] ~ [ezi] (below). Three variants for each VCV syllable were selected. The manner contrasts [eni] ~ [edi] and [eni] ~ [ezi] differ quite substantially in pitch and intensity. Note that the three [e:] sounds are the same across all four VCV syllables.

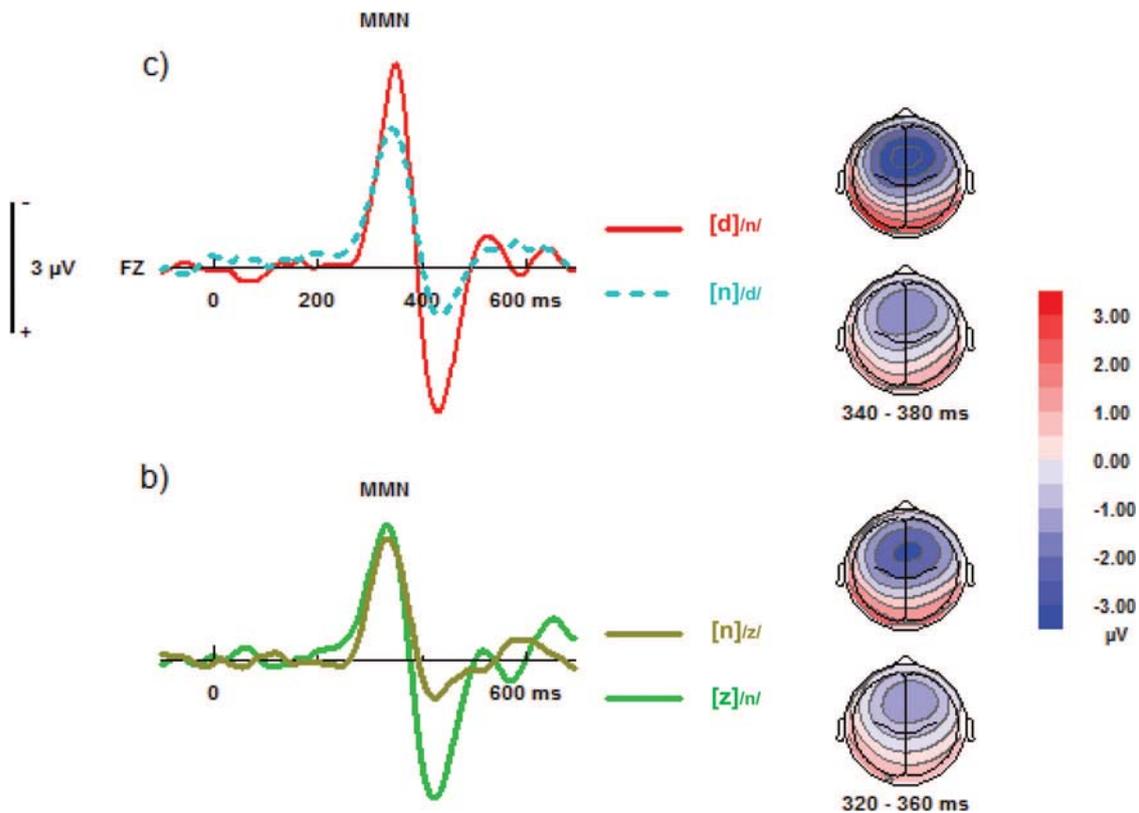


Figure 7. Grand-average waveforms of the frontal electrode (Fz) for all four conditions, the Manner Change Asymmetric pair of inversion, (c) [d]_{/n/} (red) and [n]_{/d/} (azure), and the Manner Change Symmetric pair of inversion, (b) [z]_{/n/} (green) and [n]_{/z/} (olive), with their equivalent topographical maps. These presented mismatch negativities (MMNs) are rereferenced against their right mastoids.

Discussion

In this second study, we compared the brain responses to sounds varying only in manner features. In both contrasts, the comparison focused on change of manner features [nasal], [plosive], [strident], keeping place of articulation constant. The specific contrasts were (c) [nasal] versus [plosive] and (b) [nasal] versus [strident].

According to our hypothesis, in contrast (c) an asymmetric MMN pattern was expected in the reversal of the [plosive] and [nasal] manner of articulations, each serving as standard and deviant. In contrast (b), we assumed symmetric MMN patterns having conflicting conditions for both [strident] and [nasal] manners of articulation. The [nasal] ~ [strident] contrast evoked exactly the same MMN responses as in Experiment 1. The new conflict condition, [nasal] ~ [plosive], showed an asymmetric MMN amplitude difference. The obtained asymmetry was in accordance with predictions assuming the underspecification of [plosive] in the mental lexicon.

In models in which all information of sound structure is stored (Bybee, 2001; Johnson, 1997), the prediction would be that differences between standard and deviant stimuli would be reflected in MMN brain responses on the basis of acoustic discrepancies or statistical probabilities. Again, because the changes in the sounds for the present study were not context-dependent, predictions from models in which contextual expe-

rience plays a major role in dealing with phonological deviance (Gaskell & Marslen-Wilson, 1996, 1998; Gow, 2003) could not be tested here.

General Discussion

The current study examined the fine structure of speech sounds during speech perception by means of event-related brain potentials using the MMN paradigm. We compared the brain responses to sounds varying in place and manner features. In an oddball paradigm, one expects the brain responses to reflect the degree of difference between the standard and deviant sounds. Thus, the similarity or dissimilarity between the two provided us with a means for measuring the degree of contrast between various sounds.

The MMN amplitudes varied considerably across conditions. The MMN response is affected, among others, by the deviant's characteristics that are at stake, as well as the mismatch at various levels between the standard and deviant. Being aware of this, the strategy in our experiments was the following: (i) Examine minimal differences; (ii) compare the reversals, that is, having both stimuli as standards and deviants; and (iii) add control conditions that are close to the critical conditions but in which predictions of different types of models are the same. With this ternary approach,

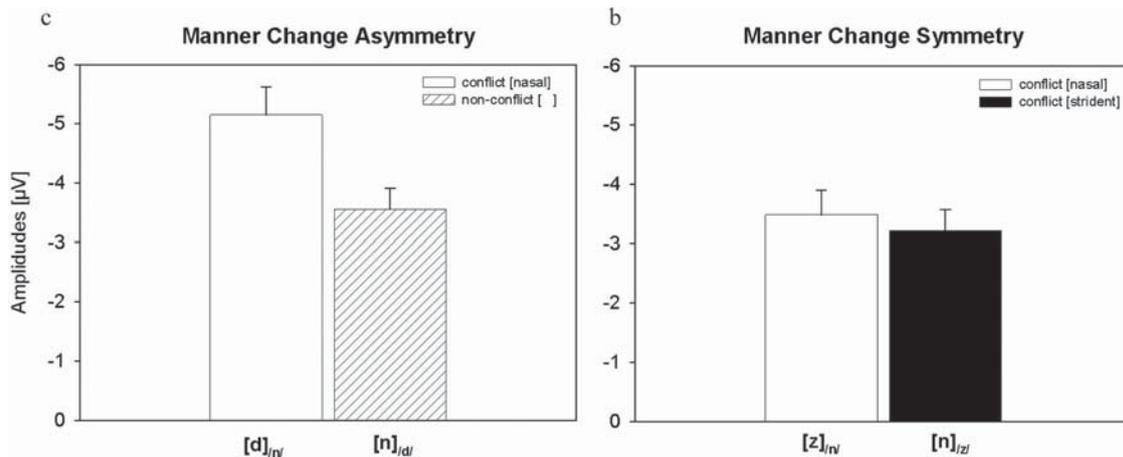


Figure 8. Mismatch negativity (MMN) mean peak amplitudes of all four conditions, the pair of inversion Manner Change Asymmetry (c), and the pair of inversion Manner Change Symmetry (b). Error markings on top of the bars represent the standard error of the mean.

we can interpret asymmetries in the way we have done, even when the MMN amplitudes varied considerably across conditions.

We discuss the results in light of our predictions as well as in terms of other hypotheses. The results for all contrasts are fully in line with the predictions of the FUL model. We found an asymmetric MMN activation pattern for our conflict versus nonconflict place ($[d]_{g/} \sim [g]_{d/}$) and manner contrasts ($[d]_{n/} \sim [n]_{d/}$) and symmetric MMN amplitudes for the two-way manner conflict ($[n]_{z/} \sim [z]_{n/}$) across both experiments. Furthermore, a significant difference was found between the symmetrical and asymmetrical conditions, suggesting different processes between the asymmetric and the symmetric pair. The predictions laid out in Tables 2 and 4 for the three (four) contrasts appear to hold; nevertheless, we consider alternative explanations.

The MMN has been shown to be sensitive to many different factors such as pitch changes, as well as intensity, duration, and ISI differences (for review, see Näätänen, 2001). Our two studies controlled for all these factors not only within the stimuli but also by using the identity MMN approach. Is there still perhaps the possibility that can be accounted for purely based on the acoustic characteristics of these stimuli? The consonantal sounds used in the experiment—[g], [d], [n], [z]—differ quite substantially in their broad acoustic properties, particularly pitch and intensity (see Figures 2 and 6). For example, [n] is a sonorant, whereas [d] is an obstruent, and consequently the differences in MMN pattern in contrast (c) could indeed be a pure acoustic difference. One could argue that a strong acoustic cue would cause a higher MMN. However, in our results, the MMN amplitude of the deviant [d] compared with the standard /n/ is higher than the deviant [n] compared with standard /d/. For an acoustic explanation to hold to explain our MMN amplitude asymmetries in this contrast, one needs to claim that the release of [d] is acoustically stronger than that of [n], which is not what we see in our stimuli. There is more reason to doubt a purely acoustic explanation if we look at contrast (b) in which the nasal [n] and the strident [z] are acoustically (pitch and intensity) rather different, but we found no differences in their MMN activation pattern. Again, if this were due to the acoustics alone, it would be difficult to argue that the strident and nasal

stimuli are equal in their acoustic strength. Similarly, the place contrast in (a) is among stops that are acoustically quite alike. However, the similar pair of stops elicited different MMN amplitudes, with the deviant [d] showing a higher amplitude than the deviant [g].

We believe that the equal MMN in contrast (b), along with the asymmetric MMN in contrasts (a) and (c), support our claim that [nasal] and [strident] are both specified in the mental representation and that these features conflict with each other in the same way, whereas [coronal] and [plosive] are different, that is, underspecified in the mental lexicon.

Among other likely parameters that could affect our results, a possible influence of phonotactic probabilities might be of relevance. Bonte et al. (2005) reported higher MMN amplitudes for nonwords with high phonotactic probability (*notsel*) as compared with the low probability condition (*notkel* and *notfel*). The distributional probabilities of [eC] sequences in our experiment were lowest for the cluster with [eg] followed by [ez], and highest for – (see Method sections and Table 3). According to the results of Bonte et al., the prediction for the present study would be a higher MMN amplitude for [eni], and [edi] compared with [ezi] and [egi]. However, the pattern of MMN differences in the present study was totally different. The conditions with the least difference in phonotactic probabilities showed large MMN differences ($[edi] < [eni]$), whereas the conditions with a large difference in phonotactic probabilities showed no MMN difference ($[ezi] = [eni]$). Thus, contrary to Bonte et al., no MMN amplitude difference between our high versus low phonotactic probability condition was found.

Furthermore, our findings cannot be explained by individual sound frequency effects of our stimuli. The intervocalic frequency (V[C]V) turns out to be highest for [n], slightly lower for [g], and lowest for [z] (see Table 3). Again, one could argue that a high sound frequency deviant would elicit a higher MMN response; however, in our results, the MMN amplitude of the deviant [d] compared with the standard /n/ is increased compared with the reversed condition. In addition, the largest sound frequency difference is between [n] and [z], but here we found equal MMN

amplitudes. These results show a pattern that cannot be explained by frequency effects or phonotactic probability influences.

With respect to more probabilistic connectionist models such as contextual integration (Gaskell & Snoeren, 2008; Gaskell et al., 1996, 1998) or feature parsing (Gow, 2002, 2003) or others examining word-final variation (e.g., Mitterer & Blomert, 2003), these studies examined consonant alternations in contexts in which assimilation could take place (e.g., *grain basket* > *graim basket*). Thus, these studies have not investigated phoneme contrasts word medially, in which no assimilation context is available. Our goal was to examine the differences between phoneme types in which the alternations are not governed by other phoneme contexts but only differ in terms of their hypothesized complexity in representation. We examined intervocalic medial consonants in which the vocalic contexts were deliberately kept constant. The stimuli were also cross-spliced to maintain as much acoustic similarity as possible. Thus, context could not have played a role in explaining these results.

In conclusion, our results provide, perhaps for the first time, electrophysiological evidence during speech perception for differential contrast sensitivity of the human brain for place and in particular manner of articulation differences that were equalized on the acoustic scale. Whereas certain manner differences such as [nasal] versus [strident] (/n/ vs. /z/) were equally contrastive for both orders of presentation, others such as [nasal] and [plosive] (/n/ vs. /d/) or [coronal] and [dorsal] (/d/ vs. /g/) turned out to be dependent on the direction of featural change. Among all possible reasons for our MMN differences discussed above, the fine structure of phonological representations in the mental lexicon explains the pattern of results best. If we assume underspecification of the manner feature [plosive] and the place features [coronal] in the mental lexicon, the asymmetric MMN amplitudes in the [nasal] ~ [plosive] and [coronal] ~ [dorsal] conditions compared with the symmetric MMN amplitudes in the [nasal] ~ [strident] conditions can be explained best. All in all, the results appear to support the assumption that the brain does refer to abstract phonological representations during speech perception that do not necessarily have to be fully specified. These experiments have revealed that MMN is a reasonably reliable and sensitive measure to study details of a variety of phoneme contrasts that differ in their phonological representations.

References

- Baayen, R., Piepenbrock, R., & Gulikers, L., (1995). *The CELEX lexical database* (Release 2) [CD-ROM]. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.
- Boersma, P., & Weenink, D., (2007). *PRAAT: Doing phonetics by computer* (Version 4.6.38). Amsterdam, the Netherlands: Institute for Phonetic Sciences.
- Bonte, M., Mitterer, H., Zellagui, N., Poelmans, H., & Blomert, L. (2005). Auditory cortical tuning to statistical regularities in phonology. *Clinical Neurophysiology*, *116*, 2765–2774. doi:10.1016/j.clinph.2005.08.012
- Bybee, J., (2001). *Phonology and language use*. Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511612886
- Cornell, S. A., Lahiri, A., & Eulitz, C. (2011). What you encode is not necessarily what you store: Evidence for sparse feature representations from mismatch negativity. *Brain Research*, *1394*, 79–89. doi:10.1016/j.brainres.2011.04.001
- Dalebout, S., & Fox, L. (2000). Identification of the mismatch negativity in the responses of individual listeners. *Journal of the American Academy of Audiology*, *11*, 12–22.
- Dalebout, S., & Stack, J. (1999). Mismatch negativity to acoustic differences not differentiated behaviorally. *Journal of the American Academy of Audiology*, *1*, 388–399.
- Dehaene-Lambertz, G. (1997). Electrophysiological correlates of categorical phoneme perception in adults. *NeuroReport*, *8*, 919–924. doi:10.1097/00001756-199703030-00021
- Diesch, E., & Luce, T. (1997). Magnetic mismatch fields elicited by vowels and consonants. *Experimental Brain Research*, *116*, 139–152. doi:10.1007/PL00005734
- Eulitz, C., & Lahiri, A. (2004). Neurobiological evidence for abstract phonological representations in the mental lexicon during speech recognition. *Journal of Cognitive Neuroscience*, *16*, 577–583. doi:10.1162/089892904323057308
- Friedrich, C. K., Eulitz, C., & Lahiri, A. (2006). Not every pseudoword disrupts word recognition: An ERP study. *Behavioral and Brain Functions*, *2*, 36. doi:10.1186/1744-9081-2-36
- Friedrich, C. K., Lahiri, A., & Eulitz, C. (2008). Neurophysiological evidence for underspecified lexical representations: Asymmetries with word initial variations. *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 1545–1559. doi:10.1037/a0012481
- Gaskell, M. (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: A Multidisciplinary Journal of Artificial Intelligence, Linguistics, Neuroscience, Philosophy, Psychology*, *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. doi:10.1037/0096-1523.22.1.144
- 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. doi:10.1037/0096-1523.24.2.380
- Gaskell, M. G., & Marslen-Wilson, W. D. (2001). Simulating parallel activation in spoken word recognition. In M. H. Christiansen & N. Chater (Eds.), *Connectionist psycholinguistics* (pp. 76–105). Westport, CT: Ablex.
- Gaskell, M. G., & Snoeren, N. (2008). The impact of strong assimilation on the perception of connected speech. *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 1632–1647. doi:10.1037/a0011977
- Gow, D. W., Jr. (2002). Does English coronal place assimilation create lexical ambiguity? *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 163–179. doi:10.1037/0096-1523.28.1.163
- Gow, D. W., Jr. (2003). Feature parsing: Feature cue mapping in spoken word recognition. *Perception & Psychophysics*, *65*, 575–590.
- Gunnior, H., Zwitserlood, P., & Bölte, J. (2005). Assimilation in existing and novel German compounds. *Language and Cognitive Processes*, *20*, 465–488. doi:10.1080/01690960444002174
- Gussenhoven, C. (1986). English plosive allophones and ambisyllabicity. *Gramma*, *10*, 119–141.
- Hacquard, V., Walter, M., & Marantz, A. (2007). The effects of inventory on vowel perception in French and Spanish: An MEG study. *Brain and Language*, *100*, 295–300. doi:10.1016/j.bandl.2006.04.009
- Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In K. Johnson & J. W. Mullenix (Eds.), *Talker variability in speech processing* (pp. 144–166). San Diego, CA: Academic Press.

- Lahiri, A., & Marslen-Wilson, W. (1991). The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition*, 38, 245–294. doi:10.1016/0010-0277(91)90008-R
- Lahiri, A., & Marslen-Wilson, W. D. (1992). Lexical processing and phonological representation. In G. Docherty & R. Ladd (Eds.), *Papers in laboratory phonology: II, Gesture, segment, prosody* (pp. 229–260). Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511519918.010
- Lahiri, A., & Reetz, H. (2002). Underspecified recognition. In C. Gussenhoven & N. Warner (Eds.), *Laboratory phonology VII* (pp. 637–677). Berlin, Germany: Mouton de Gruyter.
- Lahiri, A., & Reetz, H. (2010). Distinctive features: Phonological underspecification in processing. *Journal of Phonetics*, 38, 44–59. doi:10.1016/j.wocn.2010.01.002
- Maiste, A. C., Wiens, A. S., Hunt, M. J., Scherg, M., & Picton, T. W. (1995). Event-related potentials and the categorical perception of speech sounds. *Ear and Hearing*, 16, 68–89. doi:10.1097/00003446-199502000-00006
- Marslen-Wilson, W. D., Nix, A., & Gaskell, M. G. (1995). Phonological variation in lexical access: Abstractness, inference, and English place assimilation. *Language and Cognitive Processes*, 10, 285–308. doi:10.1080/01690969508407097
- Mitterer, H. (2003). *Understanding Gardem bench: Studies on the perception of assimilated word forms* (Unpublished doctoral dissertation). Maastricht, the Netherlands: Universiteit Maastricht.
- Mitterer, H. (2011). The mental lexicon is fully specified: Evidence from eye-tracking. *Journal of Experimental Psychology: Human Perception and Performance*, 37, 496–513. doi:10.1037/a0020989
- Mitterer, H., & Blomert, L. (2003). Coping with phonological assimilation in speech perception: Evidence for early compensation. *Perception & Psychophysics*, 65, 956–969. doi:10.3758/BF03194826
- Näätänen, R. (1992). *Attention and brain function*. Hillsdale, NJ: Erlbaum.
- Näätänen, R. (2001). The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology*, 38, 1–21. doi:10.1111/1469-8986.3810001
- Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., . . . Alho, K. (1997, January 30). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*, 385, 432–434. doi:10.1038/385432a0
- Nix, A., Gaskell, M. G., & Marslen-Wilson, W. D. (1993). Phonological variation and mismatch in lexical access. In *Proceedings of the third European conference on speech communication and technology* (pp. 685–688). Berlin, Germany: ESCA.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh Inventory. *Neuropsychologia*, 9, 97–113. doi:10.1016/0028-3932(71)90067-4
- Pettigrew, C. M., Murdoch, B. M., Kei, J., Chenery, H. J., Sockalingam, R., Ponton, C. W., . . . Alku, P. (2004). Processing of English words with fine acoustic contrasts and simple tones: A mismatch negativity study. *Journal of the American Academy of Audiology*, 15, 47–66. doi:10.3766/jaaa.15.1.6
- Pettigrew, C. M., Murdoch, B. E., Ponton, C. W., Finnigan, S., Alku, P., Kei, J., . . . Chenery, H. J. (2004). Automatic auditory processing of English words as indexed by the mismatch negativity, using a multiple deviant paradigm. *Ear and Hearing*, 25, 284–301. doi:10.1097/01.AUD.0000130800.88987.03
- Phillips, C., Marantz, A., McGinnis, M., Pesetsky, D., Wexler, K., Yellin, A., . . . Rowley, H. (1995). Brain mechanisms of speech perception: A preliminary report. *MIT Working Papers in Linguistics*, 26, 125–163.
- Phillips, C., Pellathy, T., Marantz, A., Yellin, E., Wexler, K., Poeppel, D., . . . Roberts, T. (2000). Auditory cortex accesses phonological categories: An MEG mismatch study. *Journal of Cognitive Neuroscience*, 12, 1038–1055. doi:10.1162/08989290051137567
- Pitt, M. A. (2009). How are pronunciation variants of spoken words recognized? A test of generalization to newly learned words. *Journal of Memory and Language*, 61, 19–36. doi:10.1016/j.jml.2009.02.005
- Pulvermüller, F., & Shtyrov, Y. (2006). Language outside the focus of attention: The mismatch negativity as a tool for studying higher cognitive processes. *Progress in Neurobiology*, 79, 49–71. doi:10.1016/j.neurobio.2006.04.004
- Ranbom, L. R., & Connine, C. M. (2007). Lexical representation of phonological variation in spoken word recognition. *Journal of Memory and Language*, 57, 273–298.
- Scharinger, M., Reetz, H., & Lahiri, A. (2009). Levels of regularity in inflected word form processing. *The Mental Lexicon*, 4, 77–114. doi:10.1075/ml.4.1.04sch
- Schröger, E. (1996). A neural mechanism for involuntary attention shifts to changes in auditory stimulation. *Journal of Cognitive Neuroscience*, 8, 527–539. doi:10.1162/jocn.1996.8.6.527
- Schröger, E. (1998). Measurement and interpretation of the mismatch negativity. *Behavior Research Methods, Instruments, & Computers*, 30, 131–145. doi:10.3758/BF03209423
- Shafer, V. L., Schwartz, R. G., & Kurtzberg, D. (2004). Language-specific memory traces of consonants in the brain. *Cognitive Brain Research*, 18, 242–254. doi:10.1016/j.cogbrainres.2003.10.007
- Shtyrov, Y., Kujala, T., Palva, S., Ilmoniemi, R. J., & Näätänen, R. (2000). Discrimination of speech and of complex nonspeech sounds of different temporal structure in the left and right cerebral hemispheres. *NeuroImage*, 12, 657–663. doi:10.1006/nimg.2000.0646
- Snoeren, N. D., Gaskell, M. G., & Di Betta, A. M. (2009). The perception of assimilation in newly learned novel words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35, 542–549. doi:10.1037/a0014509
- Tavabi, K., Elling, L., Dobel, C., Pantev, C., & Zwitserlood, P. (2009). Effects of place of articulation changes on auditory neural activity: A magnetoencephalography study. *PLoS ONE*, 4, e4452. doi:10.1371/journal.pone.0004452
- Wheeldon, L., & Waksler, R. (2004). Phonological underspecification and mapping mechanisms in the speech recognition lexicon. *Brain and Language*, 90, 401–412. doi:10.1016/S0093-934X(03)00451-6
- Winkler, I. (2007). Interpreting the mismatch negativity. *Psychophysiology*, 21, 147–163.
- Zimmerer, F., Reetz, H., & Lahiri, A. (2009). Place assimilation across words in running speech: Corpus analysis and perception. *Journal of the Acoustical Society of America*, 125, 2307–2322. doi:10.1121/1.3021438