Applied Psycholinguistics **33** (2012), 55–82
doi:10.1017/S0142716411000105

Perception of long-distance coarticulation: An event-related potential and behavioral study

MICHAEL GROSVALD and DAVID CORINA
University of California at Davis

Received: August 21, 2009 Accepted for publication: June 20, 2010

ADDRESS FOR CORRESPONDENCE
Michael Grosvald, Center for Mind and Brain, University of California at Davis, 267 Cousteau Place, Davis, CA 95618. E-mail: mgrosvald@ucdavis.edu

ABSTRACT
In this study we explore listeners' sensitivity to vowel to vowel (VV) coarticulation, using both event-related potential (ERP) and behavioral methodologies. The stimuli used were vowels "colored" by the coarticulatory influence of other vowels across one, three or five intervening segments. The paradigm used in the ERP portion of the study was intended to elicit the mismatch-negativity (MMN) component, a negative deflection typically seen at central midline scalp sites about 200 ms after the presentation of a "deviant" acoustic stimulus occurring among a train of "standard" acoustic stimuli. VV coarticulation at near and medium distances was associated with significant MMN-like effects, which however were not observed in response to the longest distance coarticulatory contrasts. Subjects' ERP results did not predict their performance on the behavioral task, which found evidence of listener sensitivity to even the furthest distance coarticulatory effects. Although the MMN has previously been shown to be sensitive to phonemic contrasts, this is the first study using ERP methodology to investigate the subphonemic processing associated with the perception of coarticulation.

In this paper we report the results of a study in which we explore listeners' sensitivity to the effects of transconsonantal vowel to vowel (VV) coarticulation. Although previous work has examined the perceptibility of such contrasts behaviorally, to the best of our knowledge, the present work represents the first investigation of perceptual sensitivity to coarticulatory effects in which event-related potential (ERP) methodology has been used.

VV COARTICULATION

Coarticulation may be generally described as the articulatory influence of one linguistic element (e.g., a consonant or vowel) on another (for an overview, see Hardcastle & Hewlett, 1999). Research exploring the factors influencing and constraining the production and perception of coarticulation informs theories of related phenomena such as lexical processing, speech production, and language change.

© Cambridge University Press 2011 0142-7164/11 $15.00

https://doi.org/10.1017/S0142716411000105 Published online by Cambridge University Press

Although it has long been known that adjacent segments in the flow of speech exert coarticulatory influence on each other, the systematic study of coarticulation across intervening segments began relatively recently. In particular, the systematic investigation of VV coarticulation began with Öhman's (1966) study of English, Swedish, and Russian. The strength of VV effects is now known to be influenced by diverse factors such as syllabic stress (Cho, 2004; Fletcher, 2004), prosodic domain boundaries (Cho, 1999, 2004), and a language's vowel inventory (Manuel, 1990; Manuel & Krakow, 1984), in addition to the identities of the vowels themselves and the intervening segments.

Some researchers (e.g., Gay, 1977) suspected initially that such effects would not occur over substantial distances (i.e., across many intervening segments); but later studies such as those of Recasens (1989), Magen (1997), and Grosvald (2009) have shown otherwise. For instance, Magen (1997) recorded [bVbəbVb] sequences spoken by four English speakers and found significant coarticulatory effects between the first and final vowels (i.e., across three intervening segments), meaning that such effects can cross foot boundaries and multiple syllable boundaries. However, this was not so for all four speakers in the study, showing that VV effects vary not just from language to language or context to context, but also from speaker to speaker. Magen's finding was replicated and extended by Grosvald (2009), who analyzed sentences spoken by 20 English speakers and found that the vowels [i] and [a] exerted significantly different anticipatory effects on preceding vowels across as many as five intervening segments. As was the case with Magen's study, however, there was a great deal of interspeaker variation in this regard.

The existence of long-distance coarticulatory effects has important implications for theories of speech production and perception. The aforementioned research on anticipatory effects in particular, which shows that articulatory planning can extend over a range that extends across multiple syllable boundaries, illustrates the inadequacy of earlier speech production theories positing syllable-sized planning units (e.g., Kozhevnikov & Chistovich, 1965; Stetson, 1951). More recent approaches attempting to deal with coarticulatory effects between nonadjacent segments include coproduction models (e.g., Fowler, 1983) and Keating's window model (1988, 1990a, 1990b). (See also Abry & Lallouache, 1995; Bell-Berti & Harris, 1981; Boyce, 1990; Fowler, 1980, 1983; Fowler & Saltzman, 1993; Öhman, 1967; Perkell & Chiang, 1986; Saltzman & Munhall, 1989; Smith, 1992, 1995; for an extensive overview of coarticulation models, see Farnetani & Recasens, 1999.)

To the extent that they are perceptible, even if only to a minority of listeners, long-distance anticipatory effects are also relevant to theories of speech perception. For example, regarding the role of the listener in sound change, Ohala (1981, 1994) has suggested that acoustic byproducts of articulatory processes may sometimes be misinterpreted as linguistically significant by listeners, who may then reproduce these effects rather more deliberately, ultimately leading to language change. For example, vowel harmony may be the outcome of perceptible VV coarticulation that has been grammaticalized (see Ohala, 1994; Przezdziecki, 2000). The strongest support for this hypothesis in the present context would be a correlation between tendency to coarticulate and perceptual sensitivity to coarticulatory effects; however, Grosvald (2009) investigated this possibility and found no such correlation. It should be noted however that such a correlation is not strictly required by the

https://doi.org/10.1017/S0142716411000105 Published online by Cambridge University Press

language-change hypothesis, because in a scenario like that suggested by Ohala, only a minority of speakers and listeners might be involved in the earlier stages of the change, so that a production–perception correlation in the speech community as a whole would not necessarily be expected. In a synchronic sense, anticipatory coarticulation might also be seen as aiding listeners as they process the incoming speech stream, because such coarticulation offers early information about the nature of upcoming segments.

Although the general question of how perceptible coarticulatory effects are has been examined in previous work (e.g., Beddor, Harnsberger & Lindemann, 2002; Fowler, 1981; Lehiste & Shockey, 1972; Martin & Bunnell, 1982), the first research focusing on listeners' sensitivity to long-distance effects specifically was the perception study presented in Grosvald (2009). In that study, listeners performed a behavioral oddball task: they heard vowel sounds that had been differently "colored" by coarticulatory effects of the vowels [i] and [a] at various distances (across one, three, or five intervening segments) and were asked to respond to the differently colored stimuli. Even the longest distance effects were perceptible to some, although not all, of the listeners. Here we are interested in the same basic issue investigated in that experiment (i.e., how sensitive listeners might be to VV coarticulatory effects at various distances), but we explore the issue from a different perspective through the use of ERP methodology.

This technique involves the recording of brain wave (EEG) data, and it can provide insight into mental processes that occur whether or not subjects are consciously aware of them (for a thorough overview, see Luck, 2005). Groups of neurons firing in response to particular types of stimuli produce positive or negative electrical potentials at characteristic locations on the scalp during particular time frames. When these signals are recorded, time locked to the stimuli of interest, and averaged over many trials, the background noise tends to zero and the response patterns consistently associated with the relevant stimuli remain. Such patterns are called ERP components.

One of the advantages of this approach is that it will allow an exploration of the topography and time course of the processing associated with the perception of these coarticulatory effects. In addition, previous studies have sometimes reported sensitivity differences between behavioral and ERP methodologies, often in favor of the latter. For example, McLaughlin, Osterhout, and Kim (2004) reported evidence that, even in the earliest stages of acquisition of a new language, learners' neural response to linguistically inappropriate forms, as indexed by a component known as the N400, changes significantly, even in the absence of sensitivity as measured behaviorally (see also Osterhout et al., 2008). In the present study we focus on another ERP component: mismatch negativity (MMN).

THE MMN COMPONENT

The paradigm used in our ERP study targeted the MMN component, which is generally seen at frontocentral and central scalp sites approximately 150 to 250 ms after the presentation of a "deviant" acoustic stimulus occurring among a train of otherwise similar ("standard") acoustic stimuli. Specifically, the MMN is the negative deflection seen in the difference wave obtained by subtracting the

https://doi.org/10.1017/S0142716411000105 Published online by Cambridge University Press

response to the standard stimuli from that of the deviant stimuli. The MMN can be elicited, for example, by an occasional high-frequency tone occurring among a series of low-frequency tones (Näätänen, Gaillard, & Mäntysalo, 1978). Generally, smaller differences between the deviant and standard stimuli are associated with the MMN responses of smaller amplitude and greater peak latency (Näätänen, Paavilainen, Rinne, & Alho, 2007). The MMN has counterparts in other sensory modalities such as the visual (Alho, Woods, Algazi, & Näätänen, 1992; Tales, Newton, Troscianko, & Butler, 1999), somatosensory (Kekoni et al., 1997), and olfactory (Krauel, Schott, Sojka, Pause, & Ferstl, 1999) senses and in other technological and methodological approaches such as magnetoencephalography (Hari et al., 1984), positron emission tomography (Tervaniemi et al., 2000), and functional magnetic resonance imaging (Celsis et al., 1999).

The MMN is generally thought to reflect the outcome of an automatic process comparing the just occurred auditory event with a memory trace formed by the preceding auditory events; this may be referred to as the "model adjustment" hypothesis (Näätänen, 1992; Tiitinen, May, Reinikainen, & Näätänen, 1994). The MMN is believed to have generators located bilaterally in the primary auditory cortex and in the prefrontal cortex, whose respective roles are thought to involve sensory memory and cognitive (comparative) functions (Giard, Perrin, Pernier, & Bouchet, 1990; Gomot, Giard, Roux, Barthelemy, & Bruneau, 2000). Rinne, Alho, Ilmoniemi, Virtanen, and Näätänen (2000) have found that during this process the temporal (auditory cortex) generators act earlier than those in the prefrontal cortex, supporting the hypothesis that the outcome of sensory processing in the auditory cortex is passed to the prefrontal cortex, where an attention-switching operation takes place. An alternative account of the MMN, the "adaptation hypothesis" (Jääskeläinen et al., 2004), will be discussed later.

A distinctive characteristic of the MMN is that it can be elicited even if subjects are not actively attending the stimuli, such as when they read a book or watch a silent video (Näätänen, 1979, 1985; Näätänen et al., 1978) or while they sleep (Sallinen, Kaartinen, & Lyytinen, 1994). Therefore, researchers carrying out the MMN studies do not depend on the ability or willingness of subjects to focus on behavioral tasks (regarding the possible influence of attention on the occurrence of the MMN in some contexts, see also Alain & Izenberg, 2003; Näätänen, 1991; Näätänen et al., 2007; Woldorff, Hackley, & Hillyard, 1991). Because of its automatic nature, the MMN has proven to be a valuable tool in clinical investigations where issues of auditory perception and memory encoding are relevant (for a review, see Kujala, Tervaniemi, & Schröger, 2007). This property of the MMN is also of value in the present study because it permits us to compare subjects' sensitivity, using essentially the same paradigm, in two different contexts: one incorporating a behavioral task in which subject attention is devoted to distinguishing the relevant contrasts, and another consisting of a passive-listening situation in which subjects are instructed not to pay conscious attention to the stimuli at all. In the behavioral and the MMN experiments, the same stimuli were used and were presented in exactly the same way, the only difference being the presence or absence of a response task.

Of the most importance for the present study, the MMN is known to be sensitive to linguistic phenomena such as phonemic distinctions (for an overview, see

https://doi.org/10.1017/S0142716411000105 Published online by Cambridge University Press

Näätänen et al., 2007). Moreover, studies in which multiple exemplars of each phonemic category are used have provided evidence that two subcomponents of the MMN may be distinguished: an "acoustic" one, bilaterally generated and reflecting general acoustic differences, and a "phonetic" one, generated in most subjects in the left hemisphere only (see Näätänen et al., 1997; Shestakova et al., 2002). The present study uses multiple exemplars as well, but here the contrasts are subphonemic in nature, because they are due to coarticulatory effects that are quite subtle in some cases. Thus, it it is not clear at the outset whether these contrasts can be expected to be treated during processing as belonging to different categories in this sense. The present experiment may therefore be able to establish bounds on the kinds of conditions in which one or both of these subcomponents may be elicited.

That the distinction between cross-phonemic and subphonemic contrasts might be important in the present context is supported by previous work showing categorically different responses associated with these two kinds of linguistic variation. For example, recent work by Tilsen (2009) shows an interesting interplay between coarticulatory and dissimilatory effects whose occurrence was dependent on exactly this kind of cross-category versus subcategory distinction. In Tilsen's study, subjects performed a primed-shadowing task in which they had to produce either an /a/ or /i/ vowel on each trial. At the beginning of a trial, subjects heard a "prime" stimulus (usually /a/ or /i/, sometimes a beep), which was then followed by a "target" stimulus (/a/, /i/, or a beep, the latter never following a beep prime). Subjects were required to shadow the target stimuli, except in the case of beep targets, in which case they were required to shadow the prime. The use of nonvowel (beep) primes and targets allowed Tilsen to manipulate subjects' expectations, pushing them toward planning one vowel response or another. In addition, the /a/ and /i/ prime stimuli were sometimes subtly altered, so that they were slightly more central (in terms of their first and second formants) than otherwise. The results overall were that subjects' productions of /a/ and /i/ vowels showed *assimilatory* effects in the context of subphonemic priming, and *dissimilatory* effects in the context of cross-phonemic priming. For example, subjects' responses when shadowing an /a/ target tended to be *more* central following a centralized /a/ prime than following an uncentralized /a/ prime, but /a/ responses tended to be *less* /i/-like after an /i/ prime.

Although the present study does not make use of a production task, results such as Tilsen's illustrate the fundamental importance of the subcategory versus cross-category distinction, showing that subtle perceptual data can be rapidly processed and incorporated into other aspects of listeners' behavior. The nature of such production experiments does not allow the precise determination of when these kinds of distinction are made in the course of speech processing, but the ERP technique, because of its temporal sensitivity, can provide such information. To the extent that a consistent response like the MMN is found, its time course may help to delimit boundaries on the time required for phonemic versus subphonemic distinctions to be made. If we see an MMN-type effect that is similar in kind to that observed in earlier studies in which the deviants and standards belonged to different phonemic categories (e.g., with a similar scalp distribution but smaller amplitude and/or greater latency, in line with earlier studies showing a smaller

https://doi.org/10.1017/S0142716411000105 Published online by Cambridge University Press

MMN for subtler contrasts), this would suggest that the subphonemic versus cross-phonemic distinction is noticed and acted on later. In contrast, if the MMN is insensitive to subphonemic contrasts, or if a qualitatively different kind of effect is seen, this would suggest that the MMN component reflects a stage of processing at which a determination has already been made as to whether the relevant difference is cross-categorical.[1]

PROJECT OVERVIEW

We now introduce a behavioral study that examines individual listeners' sensitivity to VV coarticulation at various distances. This is followed by a presentation of the closely related ERP study. The behavioral study, which confirms the results of Grosvald (2009), finds that (a) VV effects are readily perceived by listeners at closer distances (i.e., across fewer intervening segments) and (b) with more intervening segments, coarticulatory effects weaken and are correspondingly more difficult for listeners to detect. However, even VV effects across as many as five intervening segments are perceptible to some listeners (about 10% of subjects in the present study). The ERP study uses the same vowel stimuli and method of presentation and finds significant MMN-like responses to VV effects across up to three intervening segments. Finally, an analysis of the relationship between the behavioral and ERP results is presented.

METHOD

Listeners

The 17 participants (6 female, 11 male; age range = 18–24 years, age mean = 19.4 and $SD = 1.8$) were undergraduates at the University of California at Davis who received course credit for participating. All were native speakers of English, but as students at a university with a foreign-language requirement, tended to have had some exposure to at least one other language. Two indicated that they had substantial knowledge of another language, but the group results did not differ depending on whether they were excluded, so their data were included in the analysis. All subjects had normal hearing, were uninformed as to the purpose of the study, and gave informed consent in accordance with established Institutional Review Board procedures at UC Davis.

Creation of stimuli

For the purposes of this perception study, we felt it was desirable to use natural-language stimuli instead of creating synthesized sounds. In part, this was because stimuli reflecting long-distance coarticulatory effects were to be used, so that the creation of synthesized stimuli would require specific decisions about what sorts of effects might be considered realistic and which would not. Because the domain of long-distance VV coarticulation is still relatively unexplored, this would raise questions about the validity of whatever results might be found. Similarly, because multiple (nonidentical) tokens were to be used in each condition, as discussed

https://doi.org/10.1017/S0142716411000105 Published online by Cambridge University Press

below, it seemed best to rely on the variation present in speakers' actual utterances rather than attempt to recreate such variation artificially.

Following the perception study of Grosvald (2009), the stimuli used here were schwas ([ə]) or schwalike vowels taken from recordings of speakers saying the following sentences, which each speaker spoke six times:

"It's fun to look up at a key." (context = [i])
"It's fun to look up at a car." (context = [a])

The final (context) vowels, [i] and [a], exert different coarticulatory influence on the preceding (target) vowels; Grosvald (2009) found that for some speakers, such effects can extend over at least three vowels' distance. Further, using behavioral methods, that study found that even such long-distance effects are perceptible to some listeners. The vowels undergoing these coarticulatory effects will be referred to here as Distance 1, Distance 2, and Distance 3 target vowels, according to their distance from the context vowel. Therefore the Distance 1, 2, and 3 target vowels are the vowels in the words "a," "at," and "up," respectively (i.e., the reverse of their order in the original sentences). Schwa was chosen to serve as the target vowel because stimuli reflecting long-distance effects were desired, and schwa is known as being particularly susceptible to the coarticulatory influence of other vowels (Alfonso & Baer, 1982; Fowler, 1981).[2]

The tokens needed were [i]- and [a]-colored schwas in Distance 1, 2, and 3 conditions. Because individual recordings might have distinguishing characteristics that could aid listeners in determining their distinctiveness independent of vowel quality, four recordings of schwa in each distance condition and vowel context were selected to be presented interchangeably during each presentation of that vowel type. Four recordings from one female speaker were used for each context ([i] and [a]) for each Distance 1 and 2, whereas four recordings from another female speaker were taken for each context for the Distance 3 condition.[3] Therefore, the total number of tokens used was

$$2\,([i]\text{ vs. [a] context}) \times 3\,(\text{Distance 1, 2, or 3 condition}) \times 4\,\text{copies of each} = 24.$$

These tokens were normalized (resynthesized) in Praat (Boersma & Weenink, 2005) for duration, amplitude, and f0, according to the values shown in Table 1, which also shows the average first and second formant frequencies for the four stimuli used for each distance condition and each context. The context-related F1 and F2 differences were all significant at all three distances, except for the F1 difference at Distance 3.

The stimuli were organized into cycles consisting of eight stimuli each, with one [i]-colored schwa randomly positioned within seven [a]-colored schwas. These cycles were grouped into three blocks, with 40 or 80 cycles per block. The number of cycles in the Distance 1 block was set at only 40 because previous research (see Grosvald, 2009) and a pilot ERP study had indicated that Distance 1 schwas were easily distinguishable by subjects and that the MMN-like responses were readily elicited in this condition. Blocks for the Distances 2 and 3 stimuli consisted of 80 cycles each.

https://doi.org/10.1017/S0142716411000105 Published online by Cambridge University Press

Table 1. *Stimulus attributes*

Distance Condition	Duration (ms)	Amplitude (dB)	f0 (Hz)	F1 Mean, Contexts [i] and [a] (Hz)	F2 Mean, Contexts [i] and [a] (Hz)
1	65–70	70	120	455, 599	2398, 1700
2	55–60	70	150	574, 610	1907, 1821
3	75–80	70	200	711, 717	1645, 1549

Note: The duration, amplitude, and f0 values used to normalize the tokens used in distance conditions 1, 2, and 3 and the mean first and second formant frequencies of the set of stimuli associated with each context vowel at each distance.

Stimuli were presented 1.2 to 1.4 s apart within each cycle of eight stimuli, with short blink breaks of 2.8 to 3.1 s between each cycle and longer breaks of 10.8 to 11.1 s every 10 cycles. Relatively long interstimulus intervals were used in order to permit behavioral responses during the second (behavioral) half of the experiment, which used the same presentation of the same stimuli. Each cycle of eight vowels lasted about 15 s, and blocks of 40 or 80 cycles lasted about 10 or 20 min, respectively. The sequencing of the ERP portion of the experiment is shown in Figure 1. Although the figure shows the three blocks in distance condition order 1, 2, and 3, ordering of the three blocks for a given subject was made randomly. Because this experiment was intended to evoke the MMN, subjects did not have a response task, so they simply sat in a chair while the sounds played and were asked stay alert by watching a silent film playing on a portable DVD player positioned in front of the subject. After participating in the MMN experiment, subjects performed a behavioral perception task with the same stimuli (to be discussed presently) so that behavioral and MMN responses could be compared.

During the experiment, subjects were seated in a small (approx. 10 feet by 12 feet) sound-attenuated room in a comfortable chair facing a high-quality loudspeaker (Epos, Model ELS-3C) placed 36 in. away on a table 26 in. high. The stimuli (stored as .wav files) were delivered using a program created with Presentation software (Neurobehavioral Systems, Albany, CA), which also recorded subject responses. The tokens had been standardized in Praat at 70 Db and were delivered at this amplitude, as verified by measurement on a sound level meter (Radio Shack Model 33–2055).

EEG recording

EEG data were recorded continuously from 32 scalp locations at frontal, parietal, occipital, temporal, and central sites using silver chloride electrodes attached to an elastic cap (BioSemi). Vertical and horizontal eye movements were monitored by means of two electrodes placed above and below the left eye and two others located adjacent to the left and right eye. All electrodes were referenced to the average of the left and right mastoids. The EEG was digitized online at 256 Hz and filtered offline below 30 Hz. Scalp electrode impedance threshold values were

https://doi.org/10.1017/S0142716411000105 Published online by Cambridge University Press

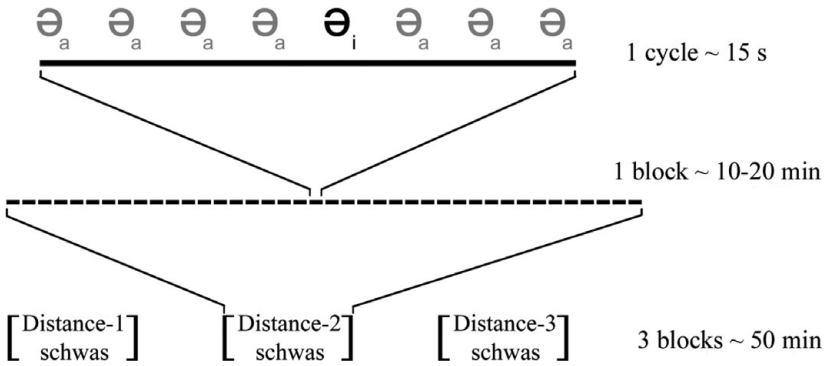


Figure 1. Sequencing of the event-related potential experiment. Each block consisted of 40 consecutive cycles of eight vowels each for the Distance 1 condition and 80 cycles for the Distance 2 and 3 conditions, with one randomly placed [i]-colored schwa per cycle. The subscript after a phonetic symbol indicates its coarticulatory "coloring," that is, [$ə_a$] for an [a]-colored schwa and [$ə_i$] for an [i]-colored schwa.

set at 20 kΩ. Epochs began 200 ms before stimulus onset and ended 600 ms after. After inspection of subjects' data by eye, artifact rejection thresholds were set at ±100 μV and rejection was performed automatically. ERP averages over epochs were calculated for each subject at each electrode for each context (standard [a] and deviant [i]) and distance condition (1, 2, or 3). Analysis was performed using EEGLAB (Delorme & Makeig, 2004). Two subjects were excluded from the group results because of a persistently high proportion of rejected trials due to artifacts (over 50%), leaving 15 participants whose data were used in the analysis.

Behavioral task

All subjects completed the ERP experiment before beginning the behavioral task, because the latter involved debriefing subjects about the purpose of the study. Afterward, in preparation for the behavioral task, subjects were given some information about the purpose of the research. They were told that language sounds can affect each other, that this can occur over some distance, that people can sometimes detect this, and that they were about to begin a task in which their own perceptual abilities along these lines were to be tested. They were told that they would be hearing vowel sounds that had been coarticulatorily "colored" by nearby [i] or [a], with some of these vowels therefore sounding more like [i] than the others. So that they would not be discouraged by the more difficult contrasts, they were told (a) that the task would progress from relatively easy to more difficult, and (b) that subjects who have taken part in such experiments sometimes say later that they felt they had been answering mostly at random in response to the more difficult contrasts, but often turn out to have performed at significantly better than chance levels. (This turned out to be the case in the present study as well.)

https://doi.org/10.1017/S0142716411000105 Published online by Cambridge University Press

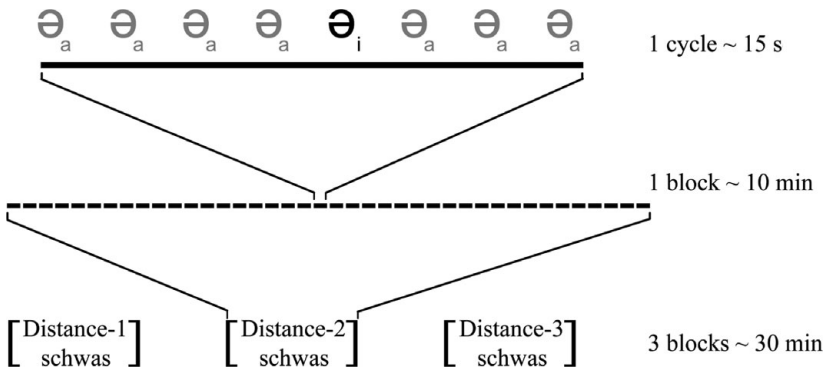


Figure 2. Sequencing of the behavioral task. Each block consisted of 40 consecutive cycles of eight vowels each, with one randomly placed [i]-colored schwa per cycle.

In addition, all subjects were given an easy warm-up task that was about 1 min long, during which easily distinguishable [i] and [a] sounds (not schwas) were played at the same rate and ratio as their counterparts in the actual task. To respond, subjects used a keyboard that was placed on their lap or in front of them on the table, whichever they felt was more comfortable. Subjects were told to hit a response button when they heard a sound that seemed more like "[i]" than the others. After completing this warm-up, they were told that the actual task would be the same in terms of pacing and goal (responding to the occasional [i]-like sounds) but more challenging. Impressionistically, the [i]-colored schwas do have a noticeably [i]-like quality to them, most obviously for Distance 1 and to some extent for Distance 2; participants' feedback as well as the results presented here indicate that subjects understood the task once they completed the warm-up.

Figure 2 illustrates the structure of the perception experiment, which consisted of three blocks, with one block for each distance condition 1, 2, and 3, in that order. The task was arranged so that each subject began with relatively easy discriminations, which we hoped would keep them from being discouraged as they progressed to the more difficult contrasts. Each block consisted of 40 cycles, each of which consisted in turn of eight consecutive schwa tokens, one of which was [i] colored and the other seven of which were [a] colored. Therefore, to perform with 100% accuracy, a subject would have to respond 40 times per block by correctly responding to the one [i]-colored schwa in each of the 40 cycles in that block. The [i]-colored tokens were randomly positioned between the second and eighth slot in each cycle, so that such tokens never occurred consecutively. As discussed earlier, the interstimulus interval varied randomly between 1200 and 1400 ms, which provided a reasonable amount of time for subjects to respond when they thought they had just heard an [i]-colored vowel.[4] Subjects were not told about the structure of cycles within blocks. However, because they had performed the warmup task, they had a sense of how often the [i]-colored schwas would tend to occur. Participants could choose to take short breaks between blocks if they wished or could proceed straight through the whole experiment.

https://doi.org/10.1017/S0142716411000105 Published online by Cambridge University Press

RESULTS

This section presents the behavioral results before the ERP study results, even though subjects took part in the two experiments in the opposite order. The reason for this is that this sequence of presentation provides an immediate quantitative measure of sensitivity for each subject in each condition, as well as for the subject pool overall. The individual subject measures will also be used later to separate subjects into subgroups for a follow-up to the main ERP analysis.

Behavioral study: Group results

For an analysis of the results of the behavioral experiment, the raw scores are not an appropriate measure; thus, the d' statistic from signal detection theory will be used instead (see Gourevitch & Galanter, 1967; Macmillan & Creelman, 1991). For the Distance 1, 2, and 3 conditions the results were significant for the set of study participants taken as a whole, with group d' scores of 3.86, 0.90, and 0.29, respectively. The details of these outcomes and the individual subject results are presented in the following section.

Behavioral study: Individual results

Table 2 displays the values of d' obtained for each listener in each distance condition, together with their associated significance levels. Consistent with related work in Grosvald (2009), a large amount of variation among listeners is evident in the perception of VV coarticulation. The results show that all respondents were readily able to distinguish [i]- and [a]-colored schwas in the Distance 1 condition; many respondents had near-perfect results here. In contrast to nearer distance effects, subjects' sensitivity to longer distance VV coarticulation was quite variable. The Distance 2 contrasts clearly presented more of a challenge to listeners than those at Distance 1 and seemed to represent something of a threshold, inasmuch as seven respondents' scores did not reach significance while 10 others' scores did. The Distance 3 condition was by far the most challenging, and respondents appear to have answered mostly at random, although two had scores that were significantly better than chance.

Table 2 also provides the results for the entire subject group. As might be expected, results were well above chance for both the Distance 1 and Distance 2 tasks. Further, although only a few subjects scored significantly better than chance in the Distance 3 task, the collective group results were also above chance levels.

Although the perception results obtained here offer some insight into differences among listeners in their perception of subtle coarticulatory effects, the methodology used might underestimate listeners' sensitivity, partly because vowels were excised from their natural contexts and played in isolation and because of the direct-questioning nature of the task ("Which vowel sounds different?"). In one example of a different approach, Scarborough (2003) cross-spliced recordings of words in such a way as to create stimuli that varied in the consistency of their coarticulatory patterns and then had listeners perform a lexical decision task. Stimuli that were consistent with naturally occurring coarticulation patterns were

https://doi.org/10.1017/S0142716411000105 Published online by Cambridge University Press

Table 2. *Behavioral results for individual subjects*

Subject	Distance 3	Distance 2	Distance 1
1	0.41	0.70	4.15***
2	0.94*	1.16*	4.65***
3	0.10	−0.07	2.56***
4	0.00	0.79	3.92***
5	0.44	0.69*	3.83***
6	−0.05	1.03*	3.38***
7	−0.20	−0.32	4.29***
8	0.49	0.57	2.89***
9	−0.18	1.21*	3.97***
10	0.44	0.53	4.65***
11	1.26*	2.04**	3.97***
12	0.17	0.53	4.29***
13	−0.07	1.93***	4.63***
14	0.27	1.72**	3.45***
15	0.40	0.94*	3.97***
16	0.15	1.46**	4.29***
17	0.59	0.80*	4.65***
Whole group	**0.29***	**0.90*****	**3.86*****

Note: The table shows each subject's d' measure for each of the three distance conditions.
*$p < .05$. **$p < .01$. ***$p < .001$.

generally associated with faster reaction times. The present study use of ERP methodology represents yet another approach seeking to better understand the perception of these subtle contrasts.

ERP study results

The ERP group results for Distances 1 and 2 show effects that have a distribution similar to that expected for the MMN component, and that, as expected, are more negative in the context of the deviant stimuli, as can be seen in Figure 3 and Figure 4. For the Distance 3 context, in contrast, such effects seem to be largely absent. Figure 3 presents topological maps of the effects at all three distances, and the associated waveforms are presented in Figure 4. Although the latency of the MMN component is typically expected to fall in the neighborhood of 200 ms, the effects seen here appear to be strongest closer to 300 ms. Previous research has shown that the amplitude and peak latency of the MMN is modulated by task difficulty, with larger and earlier MMN responses associated with greater deviations from the standard (e.g., Näätänen, 2001; Tiitinen et al., 1994), so it is possible that the late effects seen here reflect an MMN-like response whose greater latency is due to the subtlety of these subphonemic differences. For these reasons, for the purposes of the group testing whose results are about to be presented, mean amplitude over the time interval from 275 to 325 ms was used.

https://doi.org/10.1017/S0142716411000105 Published online by Cambridge University Press

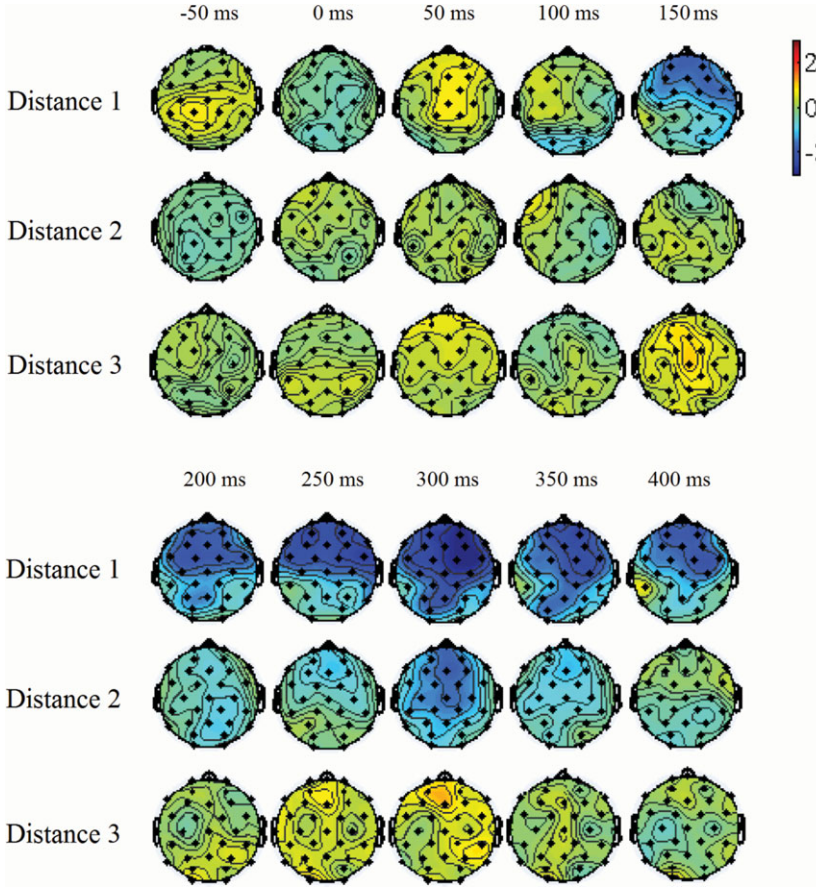


Figure 3. Topographical maps showing grand averages at each 50-ms interval in the time range of −50 to 400 ms for Distance 1, 2, and 3 conditions in the top, middle, and bottom rows, respectively. Units on the scale are in microvolts. [A color version of this figure can be viewed online at journals.cambirdge.org/aps]

To assess the significance of these effects, analysis of variance (ANOVA) testing was first performed on the group data for each distance condition with within-subject factors of hemisphere (left, mid, or right), anteriority (anterior, central, or posterior),[5] and vowel context ([i] or [a]) in the time range of 275–325 ms. In all cases, Greenhouse–Geisser and Sidak adjustments were performed where appropriate and are reflected in the results reported here.

Highly significant effects and interactions related to electrode site and context vowel were found at Distance 1 and Distance 2. At Distance 1, there were significant main effects of hemisphere, $F\,(1.36, 19.0) = 11.0, p < .01$, anteriority, $F\,(2, 28) = 25.5, p < .001$, and vowel context, $F\,(1, 14) = 14.9, p < .01$, and interactions of hemisphere–anteriority, $F\,(2.04, 28.5) = 6.42, p < .01$, and hemisphere–vowel,

https://doi.org/10.1017/S0142716411000105 Published online by Cambridge University Press

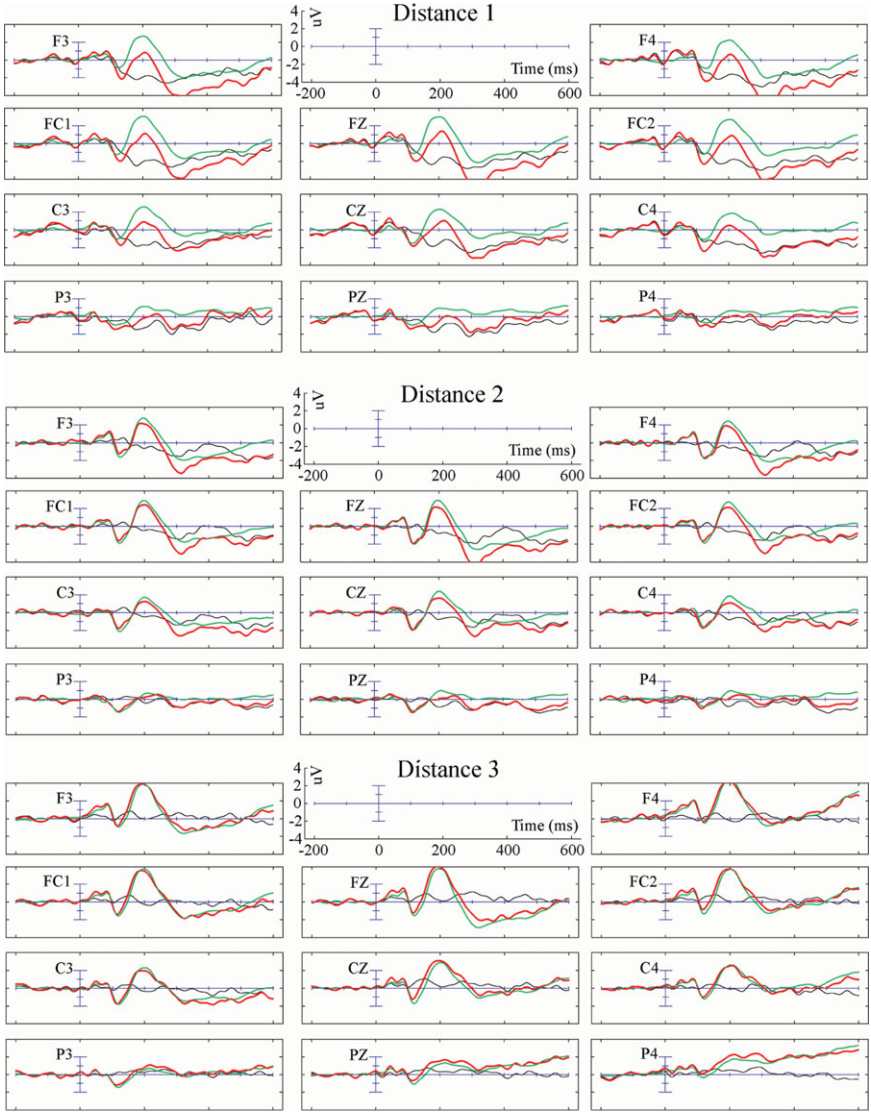


Figure 4. Grand-average waveforms at selected electrode sites in the time range of −200 to 600 ms for each of Distance 1, 2, and 3 conditions at top, middle, and bottom, respectively. In these graphs in the online color version, thick green represents standard ([a]), very thick red represents deviant ([i]), and thin black represents the difference (deviant minus standard). Negative is plotted downward. [A color version of this figure can be viewed online at journals.cambirdge.org/aps]

https://doi.org/10.1017/S0142716411000105 Published online by Cambridge University Press

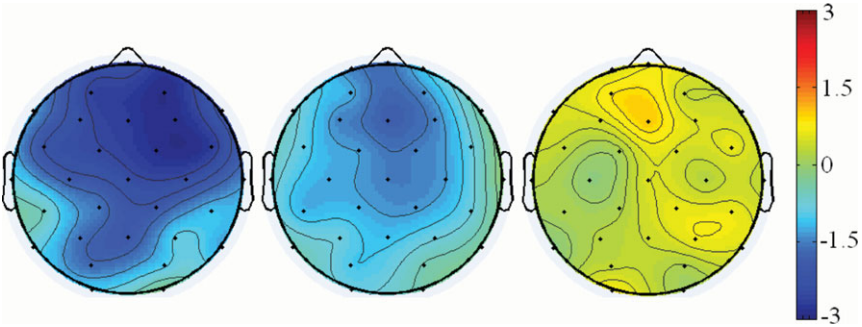


Figure 5. The topographic distribution of the mismatch-negativity-like effects at 300 ms from stimulus onset at Distance 1 and 2, and the marginally significant positivity found at that time point for the Distance 3 condition from left to right, respectively. Units on the scale are in microvolts. [A color version of this figure can be viewed online at journals.cambirdge.org/aps]

F (2, 28) = 3.49, $p < .05$. At Distance 2, there were main effects of hemisphere, F (2, 28) = 15.2, $p < .001$, anteriority, F (1.22, 17.1) = 14.9, $p < .01$, and vowel context, F (1, 14) = 11.1, $p < .01$, and interactions of hemisphere–anteriority, F (1.78, 24.9) = 4.82, $p < .05$, and hemisphere–vowel, F (2, 28) = 4.48, $p < 0.05$. All effects were in the expected direction, with greater negativity in the deviant context relative to the standard context, showing up most strongly at frontal midline sites in both of the Distance 1 and 2 conditions. In the Distance 1 condition, this negativity was somewhat stronger in the right hemisphere than in the left hemisphere, as can be seen in Figure 3, but this difference was not significant in a pairwise comparison. At Distance 3, only the effects of hemisphere and anteriority were significant, F (1.33, 18.6) = 7.94, $p < .01$ and F (1.46, 20.5) = 10.2, $p < .01$, respectively; the effect of context vowel only approached marginal significance, F (1, 14) = 2.25, $p = .156$. The effects of hemisphere and anteriority again reflect that the greatest voltage differences were seen in the front midline region. However, for Distance 3 this difference was in the contrary to expected direction, with greater *positivity* in the deviant context.

Next, ANOVAs with vowel context and electrode as factors were performed on a restricted set of electrode sites (FZ, CZ, PZ, AF3, AF4, FC1, FC2, CP1, and CP2) located in the central midline region, where the MMN is typically expected and where the strongest effects were seen. These sites will be referred to here as "MMN sites." Results were similar to those found in the earlier ANOVAs, with highly significant outcomes for vowel context only seen at Distances 1 and 2 and at best a marginal outcome at Distance 3: for the respective electrode and context at Distance 1, F (3.04, 42.6) = 12.0, $p < .001$, and F (1, 14) = 15.3, $p < .01$; at Distance 2, F (1.80, 25.2) = 11.3, $p < .001$, and F (1, 14) = 13.1, $p < .01$; and at Distance 3, F (1.91, 26.7) = 3.97, $p < .05$, and F (1, 14) = 2.32, $p = .15$. Figure 5 shows the topographic distribution of these effects in each distance condition.

The contribution to these effects from the left hemisphere is no more prominent than that of the right hemisphere, indicating that the present findings are characteristic of an "acoustic" MMN with no evidence of the specifically "phonetic" MMN

https://doi.org/10.1017/S0142716411000105 Published online by Cambridge University Press

subcomponent discussed earlier (Näätänen et al., 1997). Because the contrasts in the present study were subphonemic in nature, this is perhaps not surprising. However, given the results of the behavioral study, in which subjects were instructed (successfully) to respond to the more "[i]-like" sounds, these outcomes suggest that subjects were able to route the outcome of acoustic-level processing through linguistic channels when needed in order to perform a language-related task in which the relevant contrasts did not cross category boundaries.

Latency

Because the timing of these effects is later than would generally be expected for an MMN effect, a further breakdown of their temporal properties was made. To this end, ANOVAs like those carried out earlier for the 275- to 325-ms interval were performed over a series of 100-ms intervals in 50-ms steps from stimulus onset up to 500 ms later. Table 3 provides the outcomes of these tests. For comparison, the two rightmost columns show (a) the results for the interval from 275 to 325 ms after stimulus onset, where the effects were strongest in this study, and (b) the results for the interval from 100 to 300 ms, where the MMN is typically expected. Significant results for Distance 1 begin to be seen relatively early at approximately 100 ms. This may indicate that some but not all subjects began exhibiting a differential response by this time to the differently colored vowels. For Distance 2, the results indicate that the effect is concentrated in the later time frame. As noted earlier, context-related effects in the Distance 3 condition tended to be in the contrary to expected direction, but these differences were not significant.

For Distance 1 and perhaps to some degree at Distance 2, the effects appear rather prolonged. Together with the waveforms shown earlier in Figure 4, this raises the question of whether more than one component may be involved or, perhaps just as likely, whether these effects reflect contributions from different subgroups of subjects. Some subjects may have a more delayed response to these subphonemic contrasts than others do, leading to smearing of the effects when examined at the whole-group level. To examine this possibility further, subjects were next separated into two groups based on the behavioral results discussed earlier. The results of that investigation are presented in the following section.

Relationship to behavioral results

For the purposes of this analysis, the group of 15 ERP subjects was broken down into two subgroups, based on their behavioral (d') scores. Recall that the Distance 1 task was quite easy, with all subjects performing near ceiling levels, whereas the Distance 3 task was so difficult that few subjects performed at significantly better than chance levels. In contrast, the Distance 2 task, on which about half of the subjects performed at better than chance levels while the rest did not, provides a convenient means for breaking the subject group into two nearly equal size subgroups, with seven in what will be called the "insensitive" group and the other eight falling into the "sensitive" group.

Table 4 presents the results of a latency analysis like that whose outcome was previously shown for the entire subject group in Table 3, but now with separate

https://doi.org/10.1017/S0142716411000105 Published online by Cambridge University Press

https://doi.org/10.1017/S0142716411000105 Published online by Cambridge University Press

Table 3. *Latency results for the entire subject group (n = 15)*

	Interval (ms)										
	0–100	50–150	100–200	150–250	200–300	250–350	300–400	350–450	400–500	275–325	100–300
All sites											
Dist 1	*ns*	*ns*	*	***	**	**	**	*	*ns*	**	**
Dist 2	*ns*	*ns*	*ns*	*ns*	*	*	†	*ns*	*ns*	**	*ns*
Dist 3	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*
MMN sites											
Dist 1	*ns*	*ns*	*	**	**	**	**	*	†	**	**
Dist 2	*ns*	*ns*	*ns*	*ns*	*	*	*	*ns*	*ns*	**	†
Dist 3	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*

Note: Dist, distance. The outcome of significance testing of the mean amplitude difference between vowel contexts in the indicated time windows.
†$p < .10$. *$p < .05$. **$p < .01$. ***$p < .001$.

https://doi.org/10.1017/S0142716411000105 Published online by Cambridge University Press

Table 4. *Latency results by subject subgroup*

	Interval (ms)										
	0–100	50–150	100–200	150–250	200–300	250–350	300–400	350–450	400–500	275–325	100–300
Sens. ($n = 8$)											
Dist 1	*ns*	†	**	**	**	**	**	**	*	**	**
Dist 2	*ns*	*ns*	*ns*	*ns*	†	†	*ns*	*ns*	*ns*	*	*ns*
Dist 3	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*
Insens. ($n = 7$)											
Dist 1	(*)	*ns*	*ns*	†	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*
Dist 2	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*
Dist 3	(*)	(*)	(*)	(*)	(†)	*ns*	*ns*	*ns*	*ns*	(†)	(*)

Note: Dist, distance. Sensitivity (Sens) was determined by each subjects' performance on the behavioral task at Dist 2. Outcomes are given for significance testing of the mean amplitude difference between vowel contexts in the indicated time windows. Results in parentheses indicate an outcome in the unexpected direction.

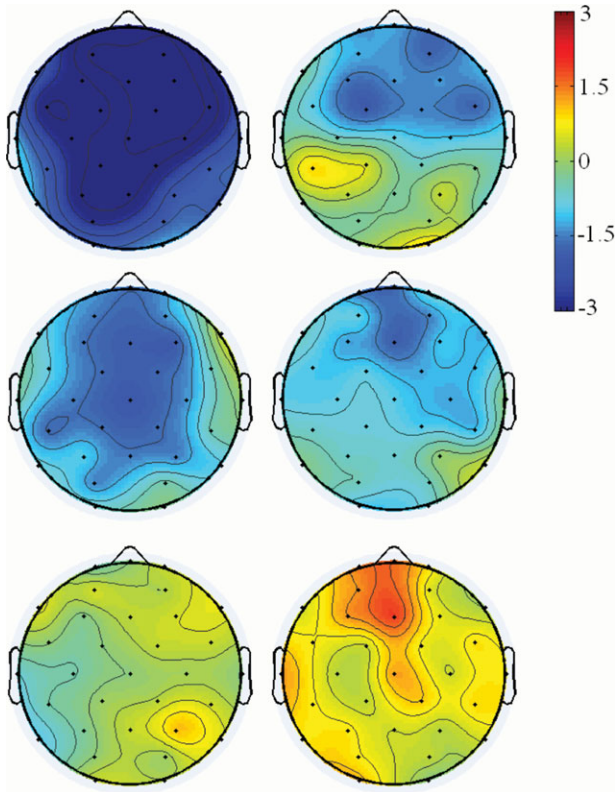


Figure 6. The topographic distribution of the effects seen at 300 ms after stimulus onset, broken down by subject group. The column at the left shows the results for the subjects deemed "sensitive" according to their d' scores, whereas the right column gives corresponding results for "insensitive" subjects. The top, middle, and bottom rows correspond to Distance 1, 2, and 3 conditions, respectively. Units on the scale are in microvolts. [A color version of this figure can be viewed online at journals.cambirdge.org/aps]

analyses performed for the insensitive and sensitive subject subgroups. Topological maps of the responses of the two groups as of 300 ms after stimulus onset are presented in Figure 6. The results are quite different for the two groups, with no significant negative effects seen for the insensitive group, even at Distance 1. This is true even though all subjects performed at well above chance levels in the Distance 1 behavioral task, and the insensitive and sensitive groups did not differ significantly in their performance of the Distance 1 task (respective mean d' scores $= 3.82$ and 3.94); $t\,(8.81) = 0.36$, $p = .73$. For Distances 1 and 2, the insensitive subjects do show a negative trend in the deviant condition, but it is clearly much weaker than that of the sensitive group.

One other noteworthy difference between the two subject subgroups is the weak but significant positivity shown by insensitive subjects prior to 300 ms

https://doi.org/10.1017/S0142716411000105 Published online by Cambridge University Press

after stimulus onset in the Distance 3 condition and much earlier in the Distance 1 condition. The very early onset of these positive effects may be cause for suspicion that it is spurious, although Jääskeläinen et al.'s (2004) "adaptation hypothesis," alluded to earlier, may hint at an alternative explanation. Jääskeläinen et al.'s assertion is that the MMN is in fact illusory, being caused by decreased neuronal response, "neuronal adaptation," during the train of similar (standard) auditory stimuli, resulting in attenuation and increased latency of another component, the N1. In this scenario, the subtraction of the averaged standard stimuli response from the averaged deviant stimuli response (from which the MMN is derived) results in an apparently distinct but in fact illusory component, which is merely an artifact of these differences in N1 behavior between the standard and deviant conditions. Although there is limited support for the adaptation hypothesis (for a discussion, see Garrido, Kilner, Stephan, & Friston, 2009), the model adjustment hypothesis remains more widely accepted, and has been assumed here. However, following a similar line of thought, it might be supposed that subjects' MMN response in the present experiment is a combination of a number of subcomponents, some positive and some negative, occurring during the same general time frame. If so, perhaps the insensitive subjects have a greater tendency toward a reduced negative subcomponent in some circumstances, resulting in a net positivity. Of course, this must remain a very tentative hypothesis at this time. In any case, it is interesting that the insensitive subjects' ERP response differs from the others' even in cases where the two subgroups' behavioral results were very similar, as in the Distance 1 condition.

Previous researchers have sometimes found the magnitude of subjects' MMN response to deviant stimuli was correlated with their sensitivity as measured behaviorally in either linguistic (Aaltonen, Eerola, Lang, Uusipaikka, & Tuomainen, 1994) or nonlinguistic (Lang et al., 1990) contexts. Unlike those studies, an analysis of the data obtained in the present experiment did not find evidence of such a relationship. The correlation of ERP response (measured as the difference in mean amplitude at the designated MMN electrode sites between the deviant and standard contexts) and d' scores obtained in the behavioral study was not found to be significant in *any* of the three distance conditions. ERP response and behavioral scores were also not correlated when each was averaged over the three distance conditions. This was the case for ERP response as measured in the 275- to 325-ms time interval used in this study, as well as in the 100- to 300-ms interval more typically associated with the MMN. Moreover, a few individuals did not appear to generate an MMN-type response even in the easier listening conditions, in which all subjects scored well above chance levels in the behavioral task.[6]

A related issue is whether magnitude of the MMN response in an easier condition might predict behavioral outcomes in more difficult conditions. To address this possibility, subjects' mean amplitude of the MMN response for the Distance 1 condition (taken as the context-dependent mean amplitude difference over the interval of 275–325 ms) was examined, together with their behavioral scores in the Distance 2 condition. A negative correlation was found, as might be expected (the MMN is by definition negative), but this was not statistically significant ($r = -.25$, $p = .37$).

https://doi.org/10.1017/S0142716411000105 Published online by Cambridge University Press

DISCUSSION

This project investigated the sensitivity of listeners to subphonemic vowel contrasts associated with VV coarticulation across various numbers of intervening segments. The behavioral results replicate findings of Grosvald (2009), confirming that such effects are well perceptible to listeners at closer distances and that a minority of listeners show sensitivity even to coarticulatory effects occurring across several intervening segments. The parallel ERP study, which targeted the MMN component, has allowed us to explore listener sensitivity to these contrasts in a passive listening situation in which subjects' attention was engaged elsewhere. Although earlier studies have shown that the MMN can be elicited in similar listening conditions involving phonemic contrasts, here we have found that an MMN (or MMN-like) effect can be seen even in response to subphonemic contrasts. However, this effect is attenuated and occurs later for progressively more subtle contrasts. For the furthest distance, most subtle contrasts, no evidence of an MMN-like effect was found, which together with the behavioral results seems to establish a boundary of sorts for listener sensitivity to such effects. A number of other interesting points emerge from these findings.

That the ERP results closely mirrored the behavioral outcomes is of interest in itself, inasmuch as one goal of this study was to explore whether there might be listener sensitivity to subphonemic variation (here, attributable to coarticulation) not captured by the behavioral data. As noted in the introductory section, researchers have often reported significant ERP responses in subjects who did not evidence sensitivity as measured in parallel behavioral tasks. When this issue has been investigated in regard to the MMN component in particular, previous studies have reported mixed results. Näätänen et al. (2007) review a number of studies illustrating their view that "in general, the MMN sensitivity to small stimulus changes seems quite well to correspond to the behavioural discrimination thresholds, which holds both with normal subjects and clinical populations" (p. 2558). However, the authors also note that some researchers have reported an MMN response in subjects who showed no evidence of sensitivity as measured behaviorally. For example, Tremblay, Kraus, and McGee (1998) reported that over the course of 10 days of training on a listening task in which subtle differences in voice onset time had to be discerned, many subjects produced a significant MMN response days before performing behaviorally at significantly better than chance levels. Although one can easily imagine that people might be neurally sensitive to a range of coarticulatory patterns that they cannot explicitly detect, no evidence for this was found in the present study, other than the tentative positivity findings.

It is also of interest that the effects seen here in response to *sub*phonemic contrasts were similar to the MMN effects found in earlier studies examining listener sensitivity to *cross*-phonemic contrasts. Although these findings are consistent with the general observation noted by previous researchers (see Näätänen et al., 2007), which is more similar standard and target stimuli are associated with a later and smaller MMN response, our study has provided specific information about the time course and topography of these effects as they apply to VV coarticulation, at least in the first two of the three distance conditions examined here. At Distance 1,

https://doi.org/10.1017/S0142716411000105 Published online by Cambridge University Press

the onset and maximum of the MMN response occurred during approximately the 100- to 200-ms and 150- to 250-ms time intervals, respectively, whereas the corresponding intervals for the Distance 2 condition were later at 200 to 300 ms and 275 to 325 ms, respectively. In terms of topography, these results are consistent overall with that typically associated with the MMN. However, in the Results Section we noted the lack of evidence for a stronger left-hemisphere contribution to the observed effects that would be characteristic of a "phonetic" as opposed to "acoustic" MMN (Näätänen et al., 1997; Shestakova et al, 2002). In linguistics, a basic distinction has traditionally been made between *phonetic* and *phonemic* (or *phonological*) contrasts, typically in reference to gradient as opposed to categorical differences (e.g., Chomsky & Halle, 1968). The variation in the MMN response observed in the present contexts and in earlier studies investigating cross-phonemic contrasts appears to be gradient rather than categorical. This suggests that this component might operate at a precategorization level, prior to or at least partly independently of the processes responsible for the much greater differences between subjects' responses in subcategorical versus cross-categorical contexts seen in other studies like that of Tilsen (2009), discussed earlier in the introduction. Therefore, the present results suggest that the neural processes critical in the determination of whether cross- or subphonemic distinctions are at issue occur later. Alternatively, it is possible that the production task used in Tilsen's study engaged resources that are relevant in perception, but nonetheless were not required in the particular tasks used in the present study.

In addition to the contrasts examined in the present study being subphonemic, another issue that is possibly relevant to the lack of a "phonetic" MMN here is the phonemic status of schwa itself. As noted earlier, schwa was chosen to serve as target vowel in this study because of its susceptibility to coarticulatory influence from nearby vowels, a property that has emerged in previous articulatory and acoustic studies (e.g., Alfonso & Baer, 1982; Fowler, 1981, respectively). The great variability in the production of English schwa has raised the question of whether this vowel may be completely underspecified except for a [+vocalic] or [−consonantal] feature when considered in phonological terms (e.g., see van Oostendorp, 2003) or may be altogether "targetless" when considered in articulatory terms. Browman and Goldstein (1992) investigated the latter possibility through experimental study of one speaker and concluded that at least for that speaker, schwa was not quite targetless, but rather had a weak target that was "completely predictable . . . [corresponding] to the mean tongue . . . position for all the full vowels" (p. 56). Even if not "targetless" in phonetic terms, it may be the case that schwa has a relatively weak phonological status accounting for the absence of a "phonetic" MMN in the present study.

A number of other factors may be relevant to the present study and hence should be noted here. The MMN response probably cannot be considered a reliable measure for listener sensitivity among individuals (for relevant work, see Kraus, Koch, McGee, Nicol, & Cunningham, 1999; Lang et al., 1995). Variation among individuals in the magnitude of the MMN response to particular stimuli in particular contexts, in addition to the variability associated with individuals' EEG data in general, may have influenced the outcomes reported in the present study. The relatively small number of subjects in our sensitive and insensitive subgroups

https://doi.org/10.1017/S0142716411000105 Published online by Cambridge University Press

may have amplified those issues in that portion of our analysis. Attention may also have played a role here: subjects were instructed not to attend to the sounds during EEG data collection, but they did devote attention to performing the behavioral task. Although this kind of procedure is typical for experiments targeting the MMN, it is possible that attentional factors were a stronger influence here than in most such studies, in view of the very small interstimulus differences that were characteristic of the Distance 3 task. It is also possible that the MMN is an incomplete index of listener sensitivity in the present context, particularly given the ambiguous phonological status of the target vowel schwa.

CONCLUSION

To the best of our knowledge, this study represents the first ERP study to investigate listeners' sensitivity to coarticulatory effects. Distance 1 VV coarticulatory effects (i.e., a vowel influencing another vowel across an intervening consonant) were associated with strong MMN-like patterns. Distance 2 effects (a vowel influencing another vowel across three intervening segments) were also associated with a highly significant MMN-like response, although not in the subgroup of subjects considered insensitive when tested behaviorally. At Distance 3 (VV effects across five intervening segments), the situation was quite different, with at best weakly significant ERP effects that were positive instead of negative, in contrast to the behavioral (d') results, which provided more straightforward evidence of some subjects' perceptual sensitivity. Although these results do not support some earlier findings that the MMN can provide a more sensitive measure than behavioral methods, the results at both Distances 1 and 2, particularly the latter, show that the MMN is sensitive even to quite subtle coarticulatory contrasts.

The MMN-like responses seen here were in general similar to those found in studies in which cross-phonemic contrasts were used, suggesting that these effects reflect a stage of processing at which a determination has not yet been made as to the phonemic status of the stimulus contrasts. Support for this is that no topographical evidence of a specifically "phonetic" MMN response (Näätänen et al., 1997; Shestakova et al., 2002) was seen here. Future studies exploring coarticulation in other contexts (i.e., segment to segment effects other than the influence of vowels on other vowels) and targeting other ERP components (e.g., the effect on sentence-level processing if coarticulatory effects are manipulated in various ways) should help to clarify many of the issues that have been raised here.

ACKNOWLEDGMENTS

We thank Tamara Swaab, Orhan Orgun, and Steve Luck for helpful comments and suggestions during the planning and execution of this study. An earlier version of this work was presented at the 2008 Western Conference on Linguistics held at UC Davis. We thank the attendees of that conference and two anonymous reviewers for valuable feedback concerning the presentation of our results. This project was supported in part by Grant NIH-NIDCD 2ROI-DC03099-11 (to D.C.).

https://doi.org/10.1017/S0142716411000105 Published online by Cambridge University Press

NOTES

1. Frenck-Mestre, Meunier, Espesser, Daffner, and Holcomb (2005) found effects as early as N1 for listeners hearing vowel contrasts that were not distinctive in their own language. However, this might not directly bear on the issues examined in the present study, because their study investigated *nonnative* as opposed to *subphonemic* contrasts.
2. Strictly speaking, the vowels in "a," "at," and "up" are not schwas when the words are spoken in isolation, of course. However, in the informal speech style the speakers were asked to adopt, both "a" and "at" were spoken with schwa-like vowels. The vowel [ˆ] was chosen as the Distance 3 vowel because of its acoustic similarity to schwa.
3. Recordings of different speakers were used just as multiple tokens in each condition were used, both for comparable reasons, so that in the event significant results were obtained, it would be much less likely that this was due to listeners relying on the characteristics of a particular speaker or a particular recording. The stimuli were taken from the recordings of two speakers who showed significant context-related formant frequency differences at the required distances, as measured by *t* tests on F1 and F2 for each speaker and for each distance condition. Out of a group of seven speakers whose production data were analyzed for possible use in this perception study, only two showed significant effects as far as at Distance 3. As discussed in Grosvald (2009) and Grosvald and Orgun (2009), if even a small proportion of listeners is sensitive to stronger than typical coarticulatory effects, this is still potentially relevant for linguistic phenomena such as language change.
4. Because both stimulus and response timing were recorded by the stimulus delivery software, it was straightforward to assign each response to a vowel stimulus, namely, the immediately preceding one. Although it is likely that subjects made occasional "late" responses (i.e., a response intended for one stimulus delayed until after presentation of the subsequent stimulus), participants' feedback indicated that they adjusted readily to the rhythm of the task.
5. The respective sets of electrodes corresponding to the six hemisphere and anteriority conditions (left, mid, right, anterior, central, posterior) were {FP1, AF3, F7, F3, FC5, T7, C3, CP5, P7, P3, PO3, O1}, {FPZ, FZ, FC1, FC2, CZ, CP1, CP2, PZ}, {FP2, AF4, F4, F8, FC6, C4, T8, CP6, P4, P8, PO4, O2}, {FP1, FPZ, FP2, AF3, AF4, F7, F3, FZ, F4, F8}, {T7, FC5, CP5, C3, FC1, CP1, CZ, FC2, CP2, C4, FC6, CP6, T8}, and {P7, P3, PO3, O1, PZ, P4, PO4, O2, P8}.
6. Because the correlation results were similar with or without those subjects' data, their data were included in the analysis.

REFERENCES

Aaltonen, O., Eerola, O., Lang, A. H., Uusipaikka, E., Tuomainen, J. (1994). Automatic discrimination of phonetically relevant and irrelevant vowel parameters as reflected by mismatch negativity. *Journal of the Acoustical Society of America, 96*, 1489–1493.

Abry, C., & Lallouache, M. T. (1995). Le MEM: Un modèle d'anticipation paramétrable par locuteur: Données sur l'arrondissement en français. *Bulletin du Laboratoire de la Communication Parlée, 3*, 85–99.

Alain, C., & Izenberg, A. (2003). Effects of attentional load on auditory scene analysis. *Journal of Cognitive Neuroscience, 15*, 1063–1073.

Alfonso, P. J., & Baer, T. (1982). Dynamics of vowel articulation. *Language and Speech, 25*, 151–173.

https://doi.org/10.1017/S0142716411000105 Published online by Cambridge University Press

Alho, K., Woods, D. L., Algazi, A., & Näätänen, R. (1992). Intermodal selective attention II: Effects of attentional load on processing auditory and visual stimuli in central space. *Electroencephalography and Clinical Neurophysiology, 82*, 356–368.

Beddor, P. S., Harnsberger, J. D., & Lindemann, S. (2002). Language-specific patterns of vowel-to-vowel coarticulation: Acoustic structures and their perceptual correlates. *Journal of Phonetics, 20*, 591–627.

Bell-Berti, F., and Harris, K. S. (1981). A temporal model of speech production. *Phonetica, 38*, 9–20.

Boersma, P., & Weenink, D. (2005). Praat: Doing phonetics by computer [Computer software]. Retrieved from http://www.praat.org

Boyce, S. E. (1990). Coarticulatory organization for lip rounding in Turkish and English. *Journal of the Acoustical Society of America, 88*, 2584–2595.

Browman, C., & Goldstein, L. (1992). "Targetless" schwa: An articulatory analysis. In G. J. Docherty & R. Ladd (Eds.), *Papers in laboratory phonology II. Gesture, segment, prosody* (pp. 26–56). Cambridge: Cambridge University Press.

Celsis, P., Boulanouar, K., Doyon, B., Ranjeva, J. P., Berry, I., Nespoulous, J. L., et al. (1999). Differential fMRI responses in the left posterior superior temporal gyrus and left supramarginal gyrus to habituation and change detection in syllables and tones. *NeuroImage, 9*, 135–144.

Cho, T. (1999). Effect of prosody on vowel-to-vowel coarticulation in English. In J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, & A. C. Bailey (Eds.), *Proceedings of the 14th International Congress of Phonetic Sciences* (pp. 459–462). San Francisco, CA: University of California.

Cho, T. (2004). Prosodically conditioned strengthening and vowel-to-vowel coarticulation in English. *Journal of Phonetics, 32*, 141–176.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper & Row.

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis [Computer software]. *Journal of Neuroscience Methods, 134*, 9–21.

Farnetani, E., & Recasens, D. (1999). Coarticulation models in recent speech production theories. In W. J. Hardcastle & N. Hewlett (Eds.), *Coarticulation: Theory, data and techniques* (pp. 31–65). Cambridge: Cambridge University Press.

Fletcher, J. (2004). An EMA/EPG study of vowel-to-vowel articulation across velars in Southern British English. *Clinical Linguistics & Phonetics, 18*, 577–592.

Fowler, C. A. (1980). Coarticulation and theories of extrinsic timing. *Journal of Phonetics, 8*, 113–133.

Fowler, C. A. (1981). Production and perception of coarticulation among stressed and unstressed vowels. *Journal of Speech and Hearing Research, 24*, 127–139.

Fowler, C. A. (1983). Converging sources of evidence on spoken and perceived rhythms in speech: Cyclic productions of vowels in monosyllabic stress feet. *Journal of Experimental Psychology: General, 112*, 386–412.

Fowler, C. A., & Saltzman, E. (1993). Coordination and coarticulation in speech production. *Language and Speech, 36*, 171–195.

Frenck-Mestre, C., Meunier, C., Espesser, R., Daffner, K., & Holcomb, P. (2005). Perceiving nonnative vowels: The effect of context on perception as evidenced by event-related brain potentials. *Journal of Speech, Language, and Hearing Research, 48*, 1–15.

Garrido, M., Kilner, J., Stephan, K., & Friston, K. (2009). The mismatch negativity: A review of underlying mechanisms. *Clinical Neurophysiology, 120*, 453–463.

Gay, T. (1977). Articulatory movements in VCV sequences. *Journal of the Acoustical Society of America, 62*, 183–193.

Giard, M. H., Perrin, F., Pernier, J., & Bouchet, P. (1990). Brain generators implicated in the processing of auditory stimulus deviance: A topographic event-related potential study. *Psychophysiology, 27*, 627–640.

Gomot, M., Giard, M.-H., Roux, S., Barthelemy, C., & Bruneau, N. (2000). Maturation of frontal and temporal components of mismatch negativity (MMN) in children. *NeuroReport, 14*, 3109–3112.

Gourevitch, V., & Galanter, E. (1967). A significance test for one-parameter isosensitivity functions. *Psychometrika, 32*, 25–33.

Grosvald, M. (2009). Interspeaker variation in the extent and perception of long-distance vowel-to-vowel coarticulation. *Journal of Phonetics, 37*, 173–188.

https://doi.org/10.1017/S0142716411000105 Published online by Cambridge University Press

Grosvald, M., & Orgun, C. O. (2009, January). *The role of the listener in sound change and the production and perception of sub-phonemic vowel contrasts*. Paper presented at the 7th Hawaii International Conference on Arts and Humanities, Honolulu.

Hardcastle, W. J., & Hewlett, N. (1999). *Coarticulation: Theory, data and techniques*. Cambridge: Cambridge University Press.

Hari, R., Hämäläinen, M., Ilmoniemi, R., Kaukoranta, E., Reinikainen, K., Salminen, J., et al. (1984). Responses of the primary auditory cortex to pitch changes in a sequence of tone pips: Neuromagnetic recordings in man. *Neuroscience Letters, 50*, 127–132.

Jääskeläinen, I. P., Ahveninen, J., Bonmassar, G., Dale, A. M., Ilmoniemi, R. J., Levänen, S., et al. (2004). Human posterior auditory cortex gates novel sounds to consciousness. *Proceedings of the National Academy of Science of the United States of America, 101*, 6809–6814.

Keating, P. (1988). Underspecification in phonetics. *Phonology, 5*, 275–292.

Keating, P. (1990a). Phonetic representations in a generative grammar. *Journal of Phonetics, 18*, 321–334.

Keating, P. (1990b). The window model of coarticulation: Articulatory evidence. In J. Kingston & M. E. Beckman (Eds.), *Papers in laboratory phonetics I: Between the grammar and the physics of speech* (pp. 451–470). Cambridge: Cambridge University Press.

Kekoni, J., Hämäläinen, H., Saarinen, M., Gröhn, J., Reinikainen, K., Lehtokoski, A., et al. (1997). Rate effect and mismatch responses in the somatosensory system: ERP-recordings in humans. *Biological Psychology, 46*, 125–142.

Kozhevnikov, V., & Chistovich, L. (1965). *Speech: Articulation and perception* (Translation 30, p. 543). Washington, DC: Joint Publications Research Service.

Krauel, K., Schott, P., Sojka, B., Pause, B. M., & Ferstl, R. (1999). Is there a mismatch negativity analogue in the olfactory event-related potential? *Journal of Psychophysiology, 13*, 49–55.

Kraus, N., Koch, D., McGee, T., Nicol, T. G., & Cunningham, J. (1999). Speech–sound discrimination in school-age children: Psychophysical and neurophysiologic measures. *Journal of Speech, Language, and Hearing Research, 42*, 1042–1060.

Kujala, T., Tervaniemi, M., & Schröger, E. (2007). The mismatch in cognitive and clinical neuroscience: Theoretical and methodological considerations. *Biological Psychology, 74*, 1–19.

Lang, H., Eerola, O., Korpilahti, P., Holopainen, I. E., Salo, S. K., & Aaltonen, O. (1995). Practical issues in the clinical application of mismatch negativity. *Ear and Hearing, 16*, 118–130.

Lang, H., Nyrke, T., Ek, M., Aaltonen, O., Raimo, I., & Näätänen, R. (1990). Pitch discrimination performance and auditory event-related potentials. In C. H. M. Brunia, A. W. K. Gaillard, A. Kok, G. Mulder, & M. N. Verbaten (Eds.), *Psychophysiological brain research* (Vol. 1, pp. 294–298). Tilburg, The Netherlands: Tilburg University Press.

Lehiste, I., & Shockey, L. (1972). On the perception of coarticulation effects in English VCV syllables. *Journal of Speech and Hearing Research, 15*, 500–506.

Luck, S. J. (2005). *An introduction to the event-related potential technique*. Cambridge, MA: MIT Press.

Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. New York: Cambridge University Press.

Magen, H. S. (1997). The extent of vowel-to-vowel coarticulation in English. *Journal of Phonetics, 25*, 187–205.

Manuel, S. Y. (1990). The role of contrast in limiting vowel-to-vowel coarticulation in different languages. *Haskins Laboratories Status Report on Speech Research, 103–104*, 1–20.

Manuel, S. Y., & Krakow, R. A. (1984). Universal and language particular aspects of vowel-to-vowel coarticulation. *Haskins Laboratories Status Report on Speech Research, 77–78*, 69–78.

Martin, J. G., & Bunnell, H. T. (1982). Perception of anticipatory coarticulation effects in vowel-stop consonant–vowel sequences. *Journal of Experimental Psychology: Human Perception and Performance, 8*, 473–488.

McLaughlin, J., Osterhout, L., & Kim, A. (2004). Neural correlates of second-language word learning: Minimal instruction produces rapid change. *Nature Neuroscience, 7*, 703–704.

Näätänen, R. (1979). Orienting and evoked potentials. In H. D. Kimmel, E. H. van Olst, & J. F. Orlebeke (Eds.), *The orienting reflex in humans* (pp. 61–75). Hillsdale, NJ: Erlbaum.

Näätänen, R. (1985). Selective attention and stimulus processing: Reflections in event-related potentials, magnetoencephalogram, and regional cerebral blood flow. In M. I. Posner & O. S. M. Marin (Eds.), *Attention and performance 1985* (Vol. 11, pp. 355–373). Hillsdale, NJ: Erlbaum.

https://doi.org/10.1017/S0142716411000105 Published online by Cambridge University Press

Näätänen, R. (1991). Mismatch negativity (MMN) outside strong attentional focus: A commentary on Woldorff et al. *Psychophysiology, 28*, 478–484.
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 and its magnetic equivalent. *Psychophysiology, 38*, 1–21.
Näätänen, R., Gaillard, A. W. K., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica, 42*, 313–329.
Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., et al. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature, 385*, 432–434.
Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology, 118*, 2544–2590.
Ohala, J. (1981). The listener as a source of sound change. In M. F. Miller (Ed.), *Papers from the parasession on language behavior*. Chicago: Chicago Linguistic Association.
Ohala, J. (1994). *Towards a universal, phonetically-based, theory of vowel harmony*. Paper presented at the Third International Conference on Spoken Language Processing (ICSLP 94), Yokohama, Japan, September 18–24. Retrieved from http://www.isca-speech.org/archive/icslp_1994
Öhman, S. E. G. (1966). Coarticulation in VCV utterances: Spectrographic measurements. *Journal of the Acoustical Society of America, 39*, 151–168.
Öhman, S. E. G. (1967). Numerical model of coarticulation. *Journal of the Acoustical Society of America, 41*, 310–320.
Osterhout, L., Poliakov, A., Inoue, K., McLaughlin, J., Valentine, G., Pitkanen, I., et al. (2008). Second-language learning and changes in the brain. *Journal of Neurolinguistics, 21*, 509–521.
Perkell, J. S., & Chiang, C. M. (1986). *Preliminary support for a "hybrid model" of anticipatory coarticulation*. Paper presented at the 12th International Conference of Acoustics.
Przezdziecki, M. (2000). Vowel harmony and vowel-to-vowel coarticulation in three dialects of Yoruba. *Working Papers of the Cornell Phonetics Laboratory, 13*, 105–124.
Recasens, D. (1989). Long range coarticulation effects for tongue dorsum contact in VCVCV sequences. *Speech Communication, 8*, 293–307.
Rinne, T., Alho, K., Ilmoniemi, R. J., Virtanen, J., & Näätänen, R. (2000). Separate time behaviors of the temporal and frontal mismatch negativity sources. *NeuroImage, 12*, 14–19.
Sallinen, M., Kaartinen, J., & Lyytinen, H. (1994). Is the appearance of mismatch negativity during stage 2 sleep related to the elicitation of K-complex? *Electroencephalography and Clinical Neurophysiology, 91*, 140–148.
Saltzman, E. L., & Munhall, K. G. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology, 1*, 333–382.
Scarborough, R. A. (2003, February 14–17). *Lexical confusability and degree of coarticulation*. Paper presented at the 29th Meeting of the Berkeley Linguistics Society.
Shestakova, A., Brattico, E., Huotilainen, M., Galunov, V., Soloviev, A., Sams, M., et al. (2002). Abstract phoneme representations in the left temporal cortex: Magnetic mismatch negativity study. *NeuroReport, 13*, 1813–1816.
Smith, C. L. (1992). *The timing of vowel and consonant gestures*. Unpublished doctoral dissertation, Yale University.
Smith, C. L. (1995). Prosodic patterns in the coordination of vowel and consonant gestures. In B. Copnnell & A. Arvaniti (Eds.), *Papers in laboratory phonology: Vol. 4. Phonology and phonetic evidence* (pp. 205–222). Cambridge: Cambridge University Press.
Stetson, R. (1951). *Motor phonetics: A study of speech movements in action* (2nd ed.). Amsterdam: North Holland.
Tales, A., Newton, P., Troscianko, T., & Butler, S. (1999). Mismatch negativity in the visual modality. *NeuroReport, 10*, 3363–3367.
Tervaniemi, M., Medvedev, S. V., Alho, K., Pakhomov, S. V., Roudas, M. S., van Zuijen, T. L., et al. (2000). Lateralized automatic auditory processing of phonetic versus musical information: A PET study. *Human Brain Map, 10*, 74–79.
Tiitinen, H., May, P., Reinikainen, K., & Näätänen, R. (1994). Attentive novelty detection in humans is governed by pre-attentive sensory memory. *Nature, 372*, 90–92.

https://doi.org/10.1017/S0142716411000105 Published online by Cambridge University Press

Tilsen, S. (2009). Subphonemic and cross-phonemic priming in vowel shadowing: Evidence for the involvement of exemplars in production. *Journal of Phonetics, 37*, 276–296.

Tremblay, K., Kraus, N., & McGee, T. (1998). The time course of auditory perceptual learning: Neurophysiological changes during speech-sound training. *NeuroReport, 9*, 3557–3560.

van Oostendorp, M. (2003). Schwa in phonological theory. In L. Cheng & R. Sybesma (Eds.), *Studies in generative grammar: Vol. 61. The second glot international state-of-the-article book. The latest in linguistics* (pp. 431–461). Berlin: Mouton de Gruyter.

Woldorff, M. G., Hackley, S. A., & Hillyard, S. A. (1991). The effects of channel-selective attention on the mismatch negativity wave elicited by deviant tones. *Psychophysiology, 28*, 30–42.

https://doi.org/10.1017/S0142716411000105 Published online by Cambridge University Press