Auditory Cortex Accesses Phonological Categories: An MEG Mismatch Study

Colin Phillips and Thomas Pellathy
University of Delaware

Alec Marantz, Elron Yellin, and Kenneth Wexler
Massachusetts Institute of Technology

David Poeppel
University of Maryland

Martha McGinnis
University of Calgary

Timothy Roberts
University of California-San Francisco

Abstract

■ The studies presented here use an adapted oddball paradigm to show evidence that representations of discrete phonological categories are available to the human auditory cortex. Brain activity was recorded using a 37-channel biomagnetometer while eight subjects listened passively to synthetic speech sounds. In the phonological condition, which contrasted stimuli from an acoustic /dæ/–/tæ/ continuum, a magnetic mismatch field (MMF) was elicited in a sequence of stimuli in which phonological categories occurred in a many-to-one ratio, but no acoustic many-to-one ratio was present. In order to isolate the contribution of phonological categories to the MMF responses, the acoustic parameter of voice onset time, which distinguished standard and deviant stimuli, was also varied within the standard and deviant categories. No MMF was elicited in the acoustic condition, in which the acoustic distribution of stimuli was identical to the first experiment, but the many-to-one distribution of phonological categories was removed. The design of these studies makes it possible to demonstrate the all-or-nothing property of phonological category membership. This approach contrasts with a number of previous studies of phonetic perception using the mismatch paradigm, which have demonstrated the graded property of enhanced acoustic discrimination at or near phonetic category boundaries. ■

INTRODUCTION

Distinguishing Phonetic and Phonological Representations

In natural language, the relationship between sound and meaning is mediated by a number of different levels of representation. Although the acoustic medium over which spoken language is transmitted is unmistakably analog, the expressive power of natural language derives from the ability to recursively combine discrete symbolic units to form vast numbers of words and sentences. Therefore, a critical first step in relating sound to meaning involves mapping analog speech sounds onto discrete phonological categories, the most basic symbolic units of language. It is out of these phonological categories that words are then built.

However, this mapping from acoustic to phonological representations is itself mediated by another level of representation, that of phonetics. Phonetic representations are language-specific, but unlike phonological representations they are not symbolic in nature. The distinction between phonetics and phonology has long been recognized in linguistics and speech science. In this paper we report studies that use magnetoencephalographic (MEG) recordings in a modified mismatch paradigm to identify auditory cortex activity specifically related to discrete, phonological representations (as opposed to nondiscrete phonetic representations). We begin by reviewing some of the key differences between phonological and phonetic representations.

Each phonetic category groups together a number of similar sounds, and members of the same phonetic category are less easily discriminated than members of different phonetic categories, as shown in classic work on speech perception of the 1950s and 1960s (Liberman, Harris, Kinney, & Lane, 1961). However, phonetic categories show category-internal structure. For example, speakers are able to discriminate different members of the same phonetic category—this is clearest for vowels (Pisoni, 1973; Stevens, Liberman, Studdert-Kennedy, & Ohman, 1969; Fry, Abramson, Eimas, & Liberman, 1962), but is also true to some extent for consonants (Carney, Widin, & Viemeister, 1977). Furthermore, speakers' judgments of category goodness reveal that phonetic categories have a prototype structure (Miller, 1994; Volaitis & Miller, 1992; Samuel, 1982). The number of distinct phonetic categories in a language typically exceeds the number of phonological categories that the language uses to store lexical forms (e.g., English distinguishes aspirated and unaspirated voiceless stops such as [p] and [pʰ] phonetically, but this distinction is irrelevant to the lexical representation of English words).

In contrast to the gradient nature of phonetic categories, phonological categories are symbolic in nature. As such, phonological categories are the discrete units that are combined into lexical forms and that participate in a variety of phonological processes. At the level of phonological categories, fine-grained distinctions among members of the same category are irrelevant. For example, when a speaker memorizes the word 'cat', she need only store the three phonological categories /k/, /æ/, and /t/ (or their constituent features), and does not need to store further details about the acoustic or phonetic properties of the word. Similarly, phonological rules and processes (such as assimilation, syllabification, and stress or tone assignment, cf. Kenstowicz, 1994) always apply equally to all members of a phonological category; there are no phonological processes that apply differently to prototype and nonprototype members of a category. At the level of phonology, therefore, membership in a category is an all-or-nothing property, and distinctions among category members are irrelevant. This all-or-nothing property of phonological categories is critical to the design of our studies below.

In the light of the distinction between phonetic and phonological categories, it is important to note that findings of 'categorical perception' of speech sounds most commonly involve phonetic categories, and do not directly reflect the discrete categories of phonology. Categorical perception refers to the nonmonotonic identification and discrimination functions obtained when speakers are asked to identify or discriminate speech sounds from a continuum that shows monotonic acoustic variation. For example, in a continuum that varies the same acoustic dimension of voice onset time (VOT) between endpoint stimuli /dæ/ and /tæ/ (as in 'dab' and 'tab'), most stimuli are reliably classified as belonging to one of the two categories, and there is only a very narrow window of uncertainty, typically centered around 30 msec VOT for English speakers. This results in a step-like identification function (see Figure 2 below). This nonmonotonic identification function is often accompanied by a nonmonotonic discrimination function, in which subjects are relatively poor at discriminating pairs of sounds drawn from the same phonetic category, but are much better at discriminating pairs of sounds drawn from different phonetic categories, even when the acoustic difference between the sounds in each pair is held constant (Liberman et al., 1961).

In other words, categorical perception phenomena involve better discrimination of between-category contrasts than within-category contrasts, a property of phonetic categories. Categorical perception of speech sounds is found in newborns and infants (Bertoncini, Bijeljac-Babic, Blumstein, & Mehler, 1987; Eimas, 1974; Eimas, Siqueland, Jusczyk, & Vigorito, 1971). Given that infants may be unable to use these discrimination abilities to support lexical storage until well into the second year of life (Barton, 1978; Stager & Werker, 1997), it is likely that the infants' categorical perception abilities reflect phonetic categories that precede the development of a phonological system. Furthermore, Werker and Logan (1985) have shown that at brief interstimulus intervals many English-speaking adults show categorical perception of the Hindi dental/retroflex stop contrast, even though this contrast is irrelevant to the speakers' native language phonology. Similarly, a number of nonhuman species have been shown to exhibit nonmonotonic discrimination of speech-like stimuli (Kluender, Diehl, & Killeen, 1987; Kuhl & Padden, 1983; Kuhl & Miller, 1978), and it can safely be assumed that these species are not associating phonological category labels with the stimuli that they perceive categorically. Evidence for categorical perception of various kinds of nonspeech stimuli in humans (e.g., Pastore et al., 1977; Pisoni, 1977; Pisoni, Carrell, & Gans, 1983; Miller, Weir, Pastore, Kelly, & Dooling, 1976) leads to the same conclusion: Nonmonotonic perception and phonological categories are independent phenomena.

The independence of categorical perception and phonological categories is further supported by phonological categories that are not associated with nonmonotonic discrimination functions. Discrimination of vowel sounds does not exhibit such sharp boundary effects as seen for stop consonants, and within-category discrimination is possible for both adults (Pisoni, 1973; Stevens et al., 1969; Fry et al., 1962) and infants (Swoboda, Morse, & Leavitt, 1976). Nevertheless, at the phonological level vowels form categories in just the same way as consonants. The fact that speakers easily recover words and phonemes from sinewave speech makes a similar point: Although speakers are able to associate phonological categories with the sinewave

combinations in order to recover words and sentences (Remez, Rubin, Burns, Pardo, & Lang, 1994; Remez, Rubin, Pisoni, & Carrell, 1981), human discrimination of the component sinewaves is roughly monotonic.

In sum, there is good reason to distinguish the acoustic and phonetic representations that underlie categorical perception from the discrete phonological category representations involved in lexical storage and phonological computation. Therefore, neuroscientific findings that demonstrate nonmonotonic sensitivity to an acoustic continuum most likely reflect phonetic representations. On the other hand, the defining property of phonological categories is that all within-category contrasts are lost: e.g., all different tokens of /d/ are treated by phonological processes as exactly the same.

Auditory Cortex and the Phonetics/Phonology Distinction

It is clear that the auditory cortex plays a critical role in spoken language processing. This would be true, even if the auditory cortex did not support any processes specific to speech, due to the importance of basic acoustic processing for higher-level speech processing. More interesting for our present concerns is the question of whether the auditory cortex supports representations and processes specific to phonetics and phonology. Existing work with clinical populations, functional neuroimaging, and electrophysiology provides a good deal of evidence that the auditory cortex supports phonetic processing, but provides less information on the question of whether the auditory cortex supports phonological representations.

Studies of Brain-Damaged Populations

The neuropsychological literature provides a good deal of support for the distinction between phonetic and phonological representations, based on the existence of distinct phonetic and phonological deficits, but provides less detailed information about which brain areas support phonetic and phonological processing (for a recent review, see Blumstein, 1995).

In speech production, phonological deficits lead to substitutions, deletions, or additions of phonological categories, but with appropriate phonetic implementation of the categories that are produced. Such deficits are found across all sub-types of aphasia (Blumstein, 1973). In speech perception, phonological deficits are seen in the misidentification of phonological categories, or the failure to discriminate phonological contrasts (Miceli, Caltagirone, Gainotti, & Payer-Rigo, 1978; Basso, Casati, & Vignolo, 1977; Blumstein, Cooper, Zurif, & Caramazza, 1977). As in the case of production, phonological impairments in perception are found across all sub-types of aphasia.

In contrast, phonetic impairments tend to be more restricted in aphasia. Errors in the detailed phonetic implementation of individual categories in production are predominantly found in anterior aphasic patients (Gandour & Dardarananda, 1984; Itoh et al., 1982; Blumstein, Cooper, Goodglass, Statlender, & Gottlieb, 1980). In the area of phonetic perception, studies of categorical perception in aphasia have found that aphasic patients show relatively normal nonmonotonic discrimination functions and relatively normal perceptual boundaries (Gandour & Dardarananda, 1982; Basso et al., 1977; Blumstein et al., 1977). Therefore, it is possible for patients to show phonological disruption without accompanying phonetic disruption.

Recently, direct recording and electrical interference studies of phonetic perception in patients undergoing presurgical mapping have shown that highly specific sites on the supratemporal plane and the superior temporal gyrus are involved in the encoding of phonetic information (Boatman, Lesser & Gordon, 1995; Steinschneider, Volkov, Noh, Garell, & Howard, 1999).

Functional Brain Imaging

A number of functional imaging studies have distinguished brain activation due to acoustic processing versus phonetic and phonological processing, and have implicated various parts of the superior temporal plane and the superior temporal gyrus in phonetic and phonological processing (e.g., Binder et al., 1997; Fiez et al., 1995; Zatorre, Evans, Meyer, & Gjedde, 1992; Zatorre, Meyer, Gjedde, & Evans, 1996). These findings provide good support for more detailed investigation of the role of the auditory cortex in specific phonetic and phonological processes; however, the tasks and the dependent measures available in these studies do not make it possible to distinguish the contributions of phonetics and phonology to the observed activation.

Electrophysiology

Although electrophysiological methods provide less precise localization information than hemodynamic techniques such as PET and fMRI, electrophysiological measures have proven to be particularly useful to investigations of phonetic processing in the brain. Using the detailed timing and amplitude information available in EEG and MEG recordings, investigators have attempted to match electrophysiological response profiles with the perceptual response profiles measured in behavioral tasks. A number of ERP and MEG studies have shown nonmonotonic patterns of brain responses elicited by speech sounds drawn from an acoustic continuum. One set of studies has shown nonmonotonic patterns in mid-latency auditory responses evoked by stimuli from a VOT continuum (Sharma & Dorman, 1999; Steinschneider et al., 1999; Simos et al., 1998;

Molfese, 1987), paralleling the nonmonotonic response patterns observed in behavioral tasks (Liberman et al., 1961). Another group of studies has shown that the amplitude of mismatch responses is greater to native-language phonetic contrasts than to nonnative contrasts or to contrasts between poor exemplars of phonetic categories (Sharma & Dorman, 1999; Winkler et al., 1999; Dehaene-Lambertz, 1997; Näätänen et al., 1997). Localization information, where available in electrophysiological studies, consistently points to the supratemporal auditory cortex as the origin of the nonmonotonic brain response patterns, consistent with the evidence from lesion studies, functional neuroimaging, and direct recordings discussed above.

Thus, there is a good deal of evidence that systems in the auditory cortex have access to at least certain aspects of phonetic category representations, as indicated by brain response patterns that parallel behavioral response patterns on phonetic tasks. However, existing evidence has left it an open question whether these systems in the auditory cortex can access representations of discrete phonological categories. This is perhaps not surprising, given the close relationship between phonetic and phonological category representations. The studies reported here examine whether phonological category representations are available to the left-hemisphere auditory cortex, using a modified version of an auditory mismatch paradigm.

Mismatch as a Measure of Discrimination

A widely used paradigm in auditory electrophysiological research is the mismatch paradigm. A sequence of standard and deviant (or oddball) stimuli is presented in a many-to-one ratio, while the subject either listens passively or attends to a secondary task (e.g., reading a book or watching a silent movie). Averaged responses to the deviant stimuli show a characteristic response component not observed in the responses to the standards, typically in the 150–300 msec latency range (Näätänen, Gaillard, & Mäntysalo, 1978; see Näätänen, 1992, 1995 for review). This component is known as the mismatch negativity (MMN) or its magnetic counterpart the mismatch field (MMF). Equivalent dipole modeling has localized the generator of the auditory mismatch response in the supratemporal auditory cortex (Alho et al., 1998; Sams, Kaukoranta, Hämäläinen, & Näätänen, 1991; Scherg, Vajsar, & Picton, 1989; Hari et al., 1984; for review of localization evidence see Alho, 1995). Much of the work in the mismatch paradigm has been devoted to investigating which kinds of auditory representations are accessible to the neural systems that generate the mismatch response. The interest in this kind of research is not to show that many levels of representation exist—this is known in advance—rather, the interest is in determining the flow of information involved in speech processing. Since the

auditory cortex is a relatively low-level component of the speech processing system, it is valuable to uncover which representations are already available at this level.

At the acoustic level it has been shown that contrasts on any of a wide variety of acoustic parameters elicit mismatch responses when presented in the requisite many-to-one ratio. In simple sequences of pure tones, contrasts in pitch, intensity, duration, and interstimulus interval all elicit a mismatch response, as do contrasts that are close to the threshold of perceptual discrimination (Kraus et al., 1993; Sams, Paavilainen, Alho, & Näätänen, 1985). Mismatch responses may be elicited by both attended and nonattended stimulus contrasts. Very many studies in this paradigm using ERP measures have focused on responses to nonattended contrasts, but a number of recent studies have demonstrated mismatch responses elicited by attended sound contrasts (e.g., Alho et al., 1998; Woldorff, Hillyard, Gallen, Hampson, & Bloom, 1998; Dehaene-Lambertz, 1997).

It is known from mismatch studies using contrasts between pure tones that larger acoustic differences along a single dimension yield higher amplitude mismatch responses (e.g., Aaltonen, Niemi, Nyrke, & Tuhkanen, 1987). Building upon this finding, a number of studies have used the mismatch paradigm to test for effects of phonetic representations on auditory cortex-evoked responses. All have focused on nonmonotonic discrimination functions as the defining characteristic of phonetic categories, and have examined whether the amplitude of mismatch responses is sensitive to the phonetic status of a speech sound contrast, i.e., whether mismatch response amplitudes show the same patterns observed in phonetic discrimination tasks. These studies have reached differing conclusions about the availability of phonetic representations to the generator of the mismatch response.

Sharma, Kraus, McGee, Carrell, and Nicol (1993) concluded that there is no effect of phonetic representations. They used two pairs of stimuli from a synthetic /da/–/ga/ continuum in their study, in which stimuli contrasted only in F2 and F3 onset values. In the two experimental conditions the acoustic difference between standards and deviants was the same, but in one of the two conditions there was also a phonetic contrast between the standard and the deviant. Neither this study nor the studies by Maiste, Wiens, Hunt, Scherg, and Picton (1995) or Sams, Aulanko, Aaltonen, and Näätänen (1990) found an increase in the mismatch response when the acoustic contrast was augmented by a phonetic contrast. Sharma et al. (1993) concluded from this that the generator of the mismatch response is not sensitive to phonetic contrasts, and can only access acoustic representations of speech sounds. However, as Maiste et al. (1995) point out, the negative result that mismatch responses to phonetic contrasts were no larger than mismatch responses to acoustic contrasts does not guarantee that the mismatch generator accesses purely acoustic representations. The

mismatch generator may be sensitive to phonetic representations, but the existence of contrasts at multiple levels of representation may fail to have an additive effect on mismatch amplitudes, or may be obscured by the greater strength of acoustic mismatch responses. Each of these studies focused on contrasts in stop consonants, and used a paradigm in which subjects listened to sequences of many hundreds of identical standards interrupted by occasional deviant stimuli.

On the other hand, a number of other studies report effects of phonetic representations in mismatch responses. Dehaene-Lambertz (1997) used stimuli from a /ba/ to /da/ to retroflex /Da/ continuum (Werker & Lalonde, 1988). Dehaene-Lambertz compared mismatch responses elicited by a /da/ versus /ba/ contrast (a native language contrast for the French-speaking subjects in this study), with different kinds of within-category contrasts: a /ba/ versus /ba/ contrast (a within-category contrast in all languages), and a /da/ versus /Da/ contrast (a within-category contrast in French, but a phonetic contrast in Hindi). A locally distributed mismatch response was elicited by the native /da/ versus /ba/ contrast, as evidenced by a significant location × condition (within- vs. across-category) interaction. Although the location × condition interaction did not reach significance for the comparison of the within-category condition with the no-contrast condition, there is evidence that a (smaller) mismatch response was elicited by the within-category contrast. An analysis that focused exclusively on the electrodes where the mismatch response was observed in the across-category condition showed a significant difference between the within-category condition and the no-contrast condition. Therefore, Dehaene-Lambertz's findings showed a smaller MMN to within-category contrasts and a larger contrast to across-category contrasts, i.e., the classic nonmonotonic discrimination profile. Although Dehaene-Lambertz tested a place of articulation contrast like the earlier studies, she used a paradigm in which stimuli were presented in small groups in an XXXY pattern, where the value of the standard (X) and the deviant (Y) varied from one group to the next. Subjects were also asked to make an explicit same/different judgment to the final member of each group of sounds. Therefore, as suggested by Dehaene-Lambertz, the greater perceptual salience of the phonological contrast in this design may be responsible for the increased MMN to phonetic contrasts. A recent study by Sharma and Dorman (1999) compared mismatch responses to two pairs of stimuli from a VOT continuum, and found that the across-category contrast (30 vs. 50 msec VOT) elicited a substantially stronger and more reliable mismatch response than the within-category contrast (40 vs. 60 msec VOT). Sharma and Dorman suggest that phonetic representations of voicing were accessible to the MMF generator in this study because voicing contrasts are established by very low-level processes in the auditory system.

Winkler et al. (1999) demonstrated a similar contrast between within-category and across-category contrasts. An elegant feature of this study is that it compares speakers of two languages: The across-category contrast for the Finnish speakers is a within-category contrast for the Hungarian speakers, and vice versa. Winkler et al. also focused on vowel contrasts, which are known to be readily discriminated both within- and across-categories. Clear mismatch responses were elicited by both kinds of contrasts in both language groups, but mismatch amplitudes were greater for the across-category contrasts. A similar result was reported by Näätänen et al. (1997), who compared the MMF responses of speakers of Finnish and speakers of Estonian to a variety of different vowel contrasts. MMF amplitudes differed between the two language groups when the oddball matched a vowel prototype in one language but not the other, with MMF amplitudes larger for the language group in which the deviant stimulus corresponded to a vowel prototype. In another study of Finnish vowels, Aaltonen, Eerola, Hellström, Uusipaikka, and Lang (1997) showed that MMN amplitudes for within-category contrasts are greater for contrasts involving nonprototype vowels than for contrasts involving a prototype vowel. This pattern of results matches the 'perceptual magnet effect,' a pattern of discrimination that Kuhl and her colleagues have argued to be specific to human phonetic processing (Kuhl, 1991).

Based on these studies, we may conclude that the auditory cortex generator of the mismatch response at least has access to representations of language-specific phonetic categories, but it is still unclear whether the auditory cortex has access to discrete phonological categories. This additional question can be addressed using a modified version of the mismatch paradigm.

Mismatch as a Measure of Identity

A study by Aulanko, Hari, Lounasmaa, Näätänen, and Sams (1993) adopted a different approach to distinguishing levels of representation in the mismatch paradigm, which we build on in our studies. Their study contrasted syllables that are perceived as /bæ/ and /gæ/, but used a number of different tokens for each category, varying in fundamental frequency (F0). The effect of this is that the categories /b/ and /g/ were presented in the requisite many-to-one ratio, although no individual stimulus was frequent. Deviant stimuli in this experiment elicited a typical mismatch response. Aulanko et al. correctly argue that this mismatch response could only have been elicited if subjects grouped together the different tokens of each phonetic category, and if the mismatch response generator was able to access these groupings. However, the mismatch response observed in this study could also be due to a simple acoustic contrast, since the F0 variation among standards and deviants was independent of the formant transitions that

distinguished /bæ/ from /gæ/, which did not show similar variation. Therefore, if the mismatch generator only tracked acoustic energy above 250 Hz in the stimuli, i.e., formants but not F0, then there would be an acoustically fixed standard/deviant contrast. The within-category differences were confined to an area of the acoustic spectrum that was irrelevant to phonological categorization. This interpretation is supported by other experiments that have shown that a mismatch response may be elicited by nonspeech sequences in which one feature dimension varies in a many-to-one ratio while other irrelevant stimulus dimensions vary randomly (Gomes, Ritter, & Vaughan, 1995; Winkler et al., 1990). Nevertheless, the approach by Aulanko et al. can be adopted in order to use mismatch as a measure of identity among the members of a phonological category.

In our experiments we also introduced variation in the standard and deviant stimuli, which were drawn from a synthetic /dæ–tæ/ VOT continuum. However, the within-standard and within-deviant variation involved exactly the same acoustic parameter that was used to differentiate the two phonological categories /d/ and /t/. Based on the results of a forced-choice identification pretest that identified each subject's perceptual boundary on the VOT continuum, a number of different sounds were chosen from both phonological categories for use in the mismatch experiments. In the phonological mismatch experiment, standards were randomly chosen tokens of one of the phonological categories, and deviants were randomly chosen tokens of the other phonological

category. As Figure 1 shows, the effect of this is to create a straightforward many-to-one ratio of standards to deviants at the phonological level, although there is no many-to-one ratio at the acoustic level. There is no fixed acoustic difference between standards and deviants, since VOT varies both within and between categories. Only when all of the different tokens of each category are treated as identical is it possible to find a consistent difference between standards and deviants. Note that the acoustic difference between categories (e.g., 40 – 24 msec = 16 msec) is smaller than the largest within-category acoustic difference (e.g., 64 – 40 msec = 24 msec). The many-to-one distribution of stimuli that is required to elicit an MMF is only present at the level of phonological categories.

We also tested the possibility that any mismatch response observed in the phonological mismatch experiment might be due to ad-hoc creation of 'long-VOT' and 'short-VOT' groupings of stimuli, requiring no access to phonological categories. In order to test this possibility we created an acoustic contrast experiment identical to the phonological contrast experiment, except that all VOTs were increased by around 20 msec, such that only two of the stimuli now fell on the [+voice] side of the subject's perceptual boundary. This preserved exactly the ratio of shorter-VOT to longer-VOT stimuli, but substantially changed the ratio among phonological categories. If the MMF simply reflects grouping of sounds at an acoustic level, then there should be no difference between the phonological and acoustic experiments; on

Figure 1. Design of phonological mismatch experiment, illustrating acoustic and phonological representation of sequence of stimuli: (a–b) phonological contrast experiment; (c–d) acoustic contrast experiment.
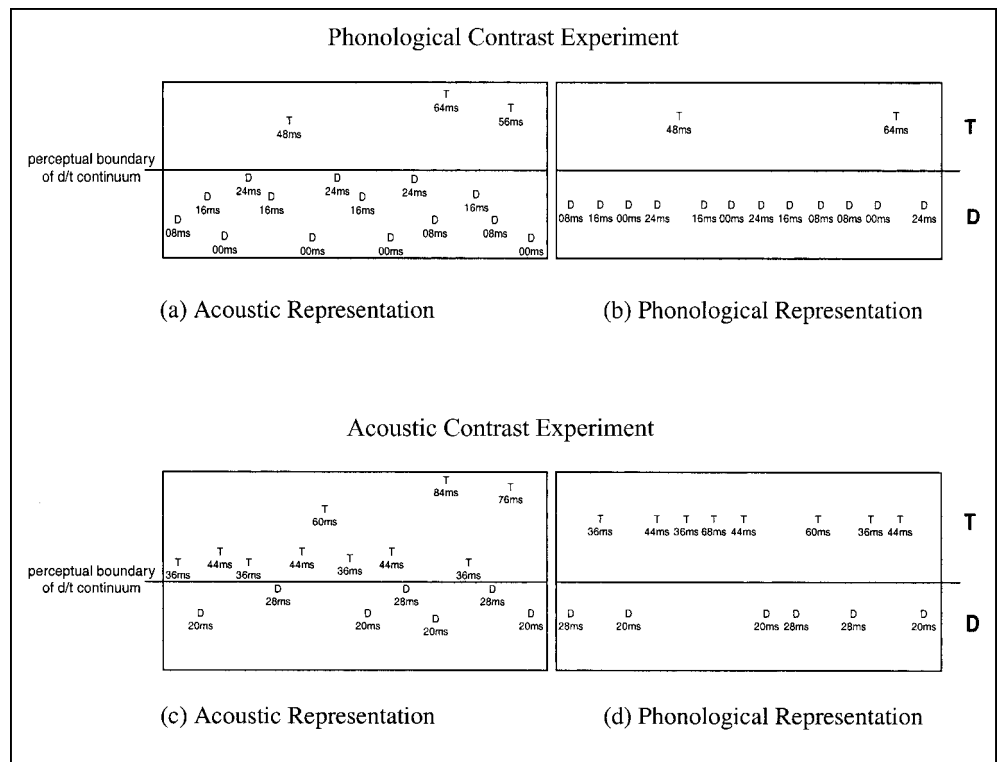
the other hand, if the MMF can access representations of phonological categories, then it should disappear entirely in the acoustic contrast experiment. Further details of the experimental design are presented below.

RESULTS

Identification Pretest

Figure 2a shows the aggregate results in the identification task pretest performed on all eight participants in the study, using 13 stimuli from the synthetic /dæ/–/tæ/ continuum. Identification responses show a classic step function, with a sharp shift in identification decisions between 20 msec VOT and 30 msec VOT, and with a mean perceptual boundary of around 25 msec. The response time data shown in Figure 2b also replicate classic findings by showing a sharp increase in response time for stimuli close to the perceptual boundary. Stimuli for the MEG study were selected for each individual subject based on the results of the pretest, in order to guarantee that the stimuli used would be consistently categorized as /dæ/ or /tæ/.

MEG Experiment 1: Phonological Contrast

If the auditory cortex generator of the MMF can access phonological category representations, we expect to find an MMF with a well-defined spatio-temporal distribution. First, the MMF should involve a reliable difference between responses to standard and deviant stimuli at a latency of around 200 msec. Second, this difference between standards and deviants should be most pronounced in the latency-window associated with the MMF, and should not be observed at other latencies. Third, the MMF should have a well-defined spatial distribution at the scalp, reflecting its origin in the auditory cortex. Specifically, a localizable MMF should

show a polarity reversal in the standard/deviant contrast at sensors on either side of the source of the MMF.

Figure 3a shows the grand average event-related field elicited across the 37-sensor array in response to the different tokens of the syllable /dæ/, with responses to standards and deviants superimposed. Responses to standards and deviants show a very similar M100 response, but diverge in the 150–250 msec interval. Responses to deviants are more negative at superior and anterior sensors (e.g., #23) and more positive at inferior and posterior sensors (e.g., #32), indicating the presence of a dipolar MMF. Figure 4 shows the grand average mismatch field obtained by subtracting responses to standards from responses to deviants. The amplitude of the difference wave is greatest at a latency of around 200 msec after stimulus onset, and there is a clear polarity inversion in the difference wave across the sensor array, indicating a source towards the middle of the array.

Statistical analyses of the waveform data confirm these observations. Mean magnetic field amplitudes across a series of time-intervals were calculated for each subject, phonological category (i.e., /dæ/ vs. /tæ/), stimulus-type (standard vs. deviant), and recording channel, and entered into a separate repeated measures ANOVA for each of a number of latency-intervals, with category, stimulus-type, and recording channel as within-subjects variables. Recording channels were included as a factor in the analysis because it is important to show a polarity reversal in the mismatch field across the sensor array (positioned over the left hemisphere auditory cortex) in order to localize the MMF to the auditory cortex. Standard corrections were applied to compensate for the violation of the sphericity assumption in the data (all values of epsilon reported reflect the more conservative Greenhouse–Geisser correction [Greenhouse & Geisser, 1959], unless noted otherwise). Results are summarized in Table 1. The highlighted cell indicates the stimulus-

Figure 2. Results from identification pretest using synthetic /dæ/–/tæ/ continuum: (a) identification judgments; (b) mean response times.
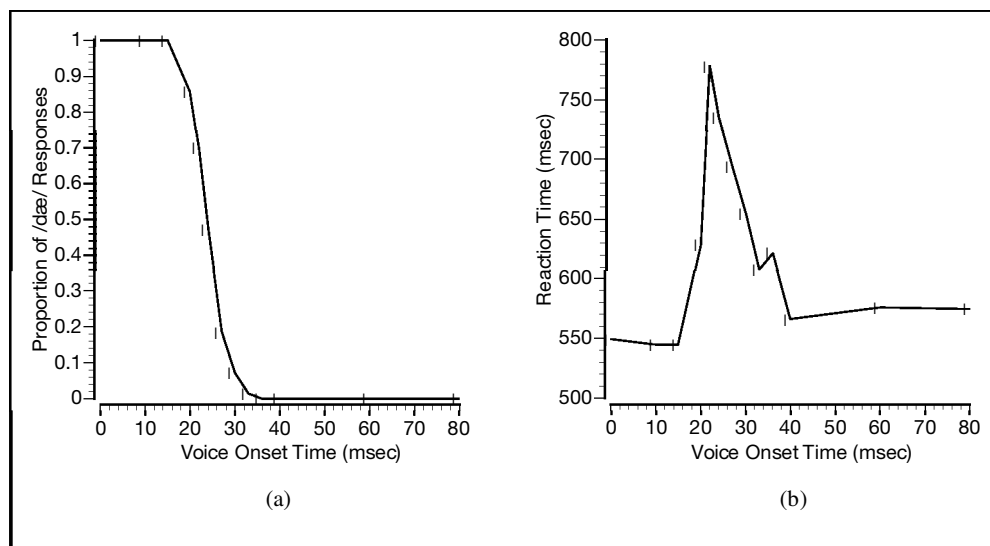
Figure 3. Grand average responses to standard and deviant /dæ/ across sensor array and at representative anterior/superior and posterior/inferior channels (eight subjects). A polarity reversal can be seen in the difference between standards and deviants from anterior/superior to inferior/posterior channels, indicating a source of the difference wave around the center of the sensor array.
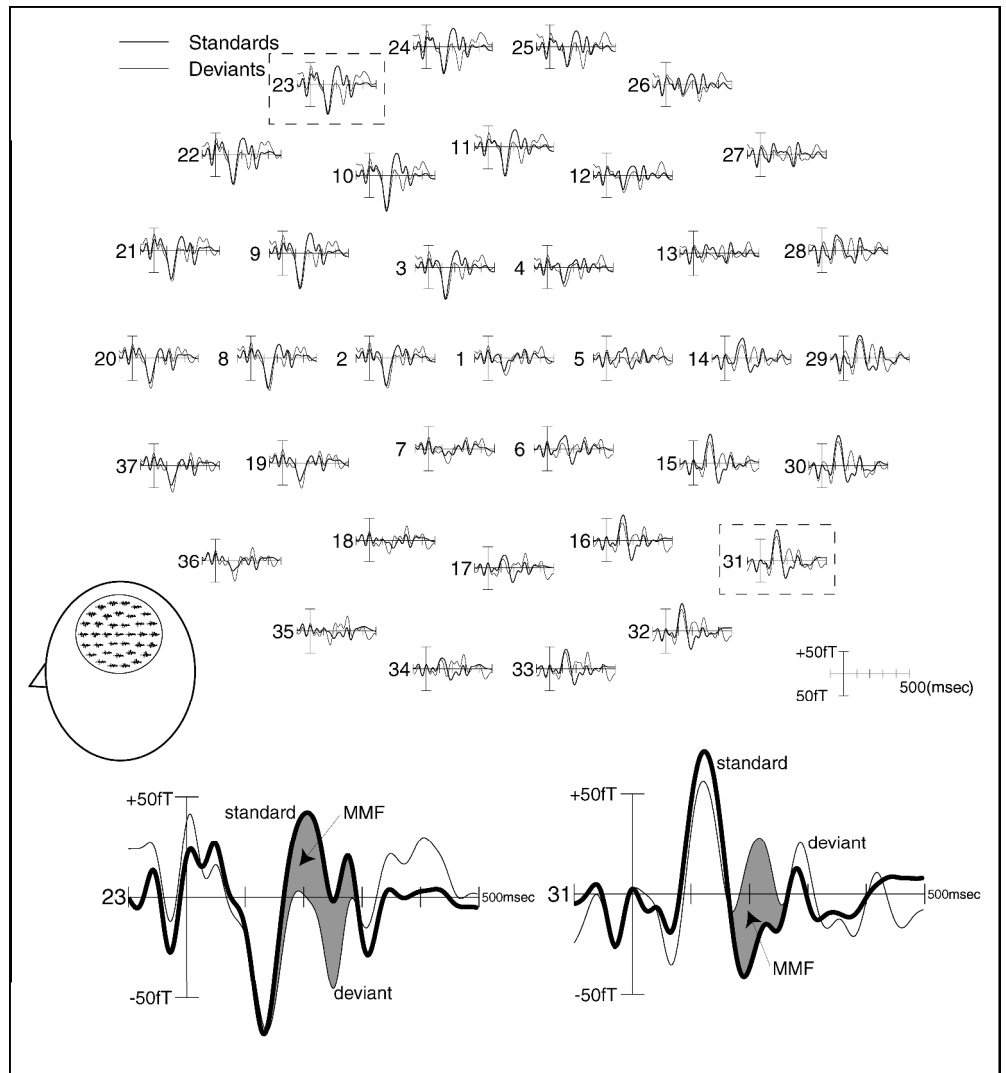
type sensor-location interaction that is expected if there is a difference between standards and deviants with a dipolar distribution, in the 150–210 msec latency window, which is when the MMF is expected. We report results from a number of different latency intervals, in order to demonstrate that the mismatch response was confined to the expected latency window.

Mismatch Response

The mismatch field in the 150–210 msec latency window is demonstrated by the significant stimulus-type × channel interaction. This interaction is the critical indicator of a mismatch field with a dipolar distribution. The dipolar distribution of the mismatch response is confirmed by separate ANOVAs run on contiguous subsets of channels, which show main effects of stimulus-type. In a superior/anterior group of nine channels (#3, #9–11, #21–25, see Figure 3) there was a main effect of

stimulus-type, $F(1, 7) = 7.87$, $p = .026$, with mean field strengths of 13.24 fT for standards and -2.21 fT for deviants. At a posterior/inferior group of nine channels (#6, #15–17, #30–34) the main effect of stimulus-type did not reach significance, $F(1, 7) = 2.52$, $p = .156$, although the mean field strength values are the inverse of the means observed in the superior/anterior channel set (standards: -5.94 fT; deviants: 8.49 fT), as predicted. The stimulus-type × channel interaction was not significant at any other latency window.

Other Main Effects and Interactions

There was a significant main effect of recording channel at the 0–90 msec latency window, the 90–150 msec latency window, and the 210–270 msec latency window. In each case this main effect was due to an auditory evoked response (e.g., M100) that showed a dipolar distribution across the sensor array.
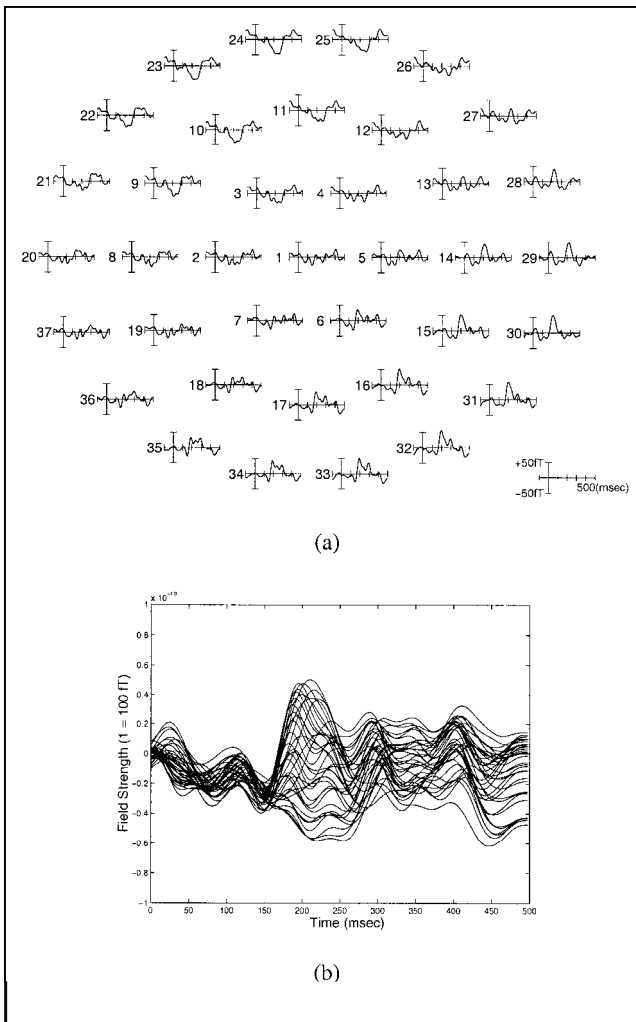
(a)

(b)

Figure 4. Grand average difference wave obtained by subtracting responses to standards from responses to deviants for /dæ/ stimuli (eight subjects); the sensor array display in (a) shows the distribution of the difference wave; the sensor overlay display in (b) shows that the difference wave is dipolar.

The main effect of category did not reach significance in any latency window, but it was marginally significant in the 150–210 msec latency window ($p = .058$) and the 210–270 msec latency window ($p = .091$). In both cases, this was due to a more positive mean response to /tæ/ than to /dæ/. Note that this more positive response does not imply a stronger auditory evoked response—a stronger response originating in the auditory cortex would show a dipolar distribution. The marginally significant category × channel interaction in the same 210–270 msec time window ($p = .10$) reflects the fact that the more positive response to /tæ/ was most pronounced at the anterior and superior recording channels.

There was a significant category × channel interaction at both the 90–150 msec latency window ($p < .05$) and the 150–210 msec latency window ($p < .05$), in both cases due to a higher amplitude response to /dæ/ than to /tæ/, with a dipolar distribution.

There was a significant or marginally significant category × stimulus-type interaction at the 0–90 msec latency window ($p = .06$), the 90–150 msec latency window ($p < .05$) and the 150–210 msec latency window ($p = .087$). In each case, this effect was due to the fact that mean field strengths were more positive to standards than to deviants for /dæ/, but more negative to standards than to deviants for /tæ/. At the 210–270 msec time interval only there was a marginally significant stimulus-type × phoneme × channel interaction ($p = .051$). This effect was due to differing distributions of a dipolar effect of stimulus-type for /dæ/ and /tæ/.

Localizations of M100 and MMF responses were calculated using an iterative least-squares single dipole modeling algorithm. Figure 5 shows the MRI-overlay of the localization of the peak of the MMF component elicited by /dæ/ deviants in one subject, a 30-year old woman. The localization is in the superior temporal plane of the left hemisphere, the site of the auditory cortex.

Table 1. Results of ANOVA, Experiment 1 (Phonological Contrast)

	df	0–90 msec	90–150 msec	150–210 msec	210–270 msec	270–330 msec
CAT	1, 7	$F = 3.03, p = .125$	$F = 0.79, p = .40$	$F = 5.14, p = .058$	$F = 3.85, p = .091$	$F = 0.04, p = .84$
ST	1, 7	$F = 1.41, p = .27$	$F = 1.09, p = .33$	$F = 0.05, p = .83$	$F < 0.01, p = .98$	$F < 0.01, p = .95$
CH	36, 7	$F = 3.34, p < .05,$ $\varepsilon = 0.077$	$F = 8.36, p < .01,$ $\varepsilon = 0.047$	$F = 0.15, p = .82,$ $\varepsilon = 0.044$	$F = 5.96, p < .01,$ $\varepsilon = 0.081$	$F = 1.61, p = .23,$ $\varepsilon = 0.059$
CAT × ST	1, 7	$F = 5.01, p = .06$	$F = 6.62, p < .05$	$F = 3.96, p = .087$	$F = 1.85, p = .22$	$F = 0.74, p = .42$
ST × CH	36, 7	$F = 0.93, p = .44,$ $\varepsilon = 0.079$	$F = 0.88, p = .48,$ $\varepsilon = 0.095$	$F = 4.55, p = .028,$ $\varepsilon = 0.058$	$F = 0.77, p = .49,$ $\varepsilon = 0.062$	$F = 1.94, p = .16,$ $\varepsilon = 0.075$
CAT × CH	36, 7	$F = 0.45, p = .67,$ $\varepsilon = 0.063$	$F = 3.86, p < .05,$ $\varepsilon = 0.064$	$F = 4.49, p = .027,$ $\varepsilon = 0.06$	$F = 2.53, p = .10,$ $\varepsilon = 0.069$	$F = 0.31, p = .82,$ $\varepsilon = 0.085$
CAT × ST × CH	36, 7	$F = 0.59, p = .61,$ $\varepsilon = 0.073$	$F = 0.45, p = .64,$ $\varepsilon = 0.055$	$F = 2.28, p = .12,$ $\varepsilon = 0.072$	$F = 3.8, p = .051,$ $\varepsilon = 0.051$	$F = 0.69, p = .54,$ $\varepsilon = 0.067$

CAT = category (/dæ, tæ/); ST = stimulus-type (standard, deviant); CH = recording channel. The highlighted cell indicates the critical stimulus-type × channel interaction that is expected if a mismatch response occurs.
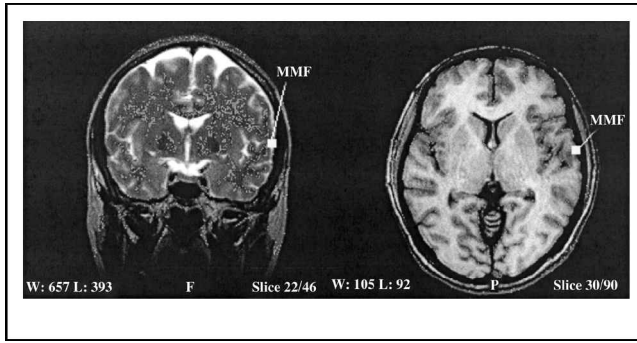
Figure 5. MRI overlay of single-dipole localization of /dæ/ mismatch field, 190 msec latency, 30-year-old female subject.

The fact that the MMF originated in the same region of the cortex as the auditory M100 response, as is typically found, can also be observed by comparing the scalp distribution of the grand average difference waves in Figure 4 with the distribution of the M100 in Figure 3.

Separate analyses of variance were run on responses to the individual phonemes /dæ/ and /tæ/, with stimulus-type (standard vs. deviant) and channel as within-subjects factors. Responses to /dæ/ show the stimulus-type × channel interaction characteristic of an MMF in the 150–210 msec time interval, $F(36, 7) = 6.38$, $p < .01$, $\varepsilon = 0.07$, and in the 210–270 msec time interval, $F(36, 7) = 4.17$, $p < .05$, $\varepsilon = 0.068$, and this interaction was not significant or marginally significant at any other time interval.

Comparison of responses to standard and deviant /tæ/ do not show a significant stimulus-type × channel interaction at the 150–210 msec time interval, $F(36, 7) = 1.95$, $p = .17$, $\varepsilon = 0.061$, or at any of the other intervals analyzed above. When the average field strength data are drawn from a narrower 170–190 msec time interval (corresponding to the peak of the difference wave), the stimulus-type × channel interaction approaches significance when the less conservative Huynh–Feldt

epsilon is used, $F(36, 7) = 2.17$, $p = .109$, $\varepsilon = 0.098$. However, comparisons of responses to /tæ/ deviants with /dæ/ standards, which were presented together in the experiment, show a clear MMF. There is a significant stimulus-type × channel interaction in the 150–210 msec interval, $F(36, 7) = 5.61$, $p = .013$, $\varepsilon = 0.061$. This interaction is not significant at any other time interval, although it is marginally significant in the 90–150 msec interval, $F(36, 7) = 2.73$, $p = .091$, $\varepsilon = 0.063$.

MEG Experiment 2: Acoustic Contrast

In the acoustic contrast experiment subjects listened to sequences of syllables that were identical to those of the main condition, except that all VOT values were increased by a constant amount (around 20 msec—precise values vary according to subject's perceptual boundary). The effect of this manipulation was that the relative acoustic distribution of VOT values was identical to the phonological contrast experiment, but the phonological distribution of stimuli was substantially changed. The perceptual category boundary now fell between the second and third stimulus of the 'short VOT' group.

Figure 6 shows side-by-side comparisons of the grand average difference wave in the phonological contrast experiment and the acoustic contrast experiment. As the figure clearly shows, the dipolar MMF observed in the phonological contrast experiment is not observed in the acoustic contrast experiment. This observation is confirmed by statistical analyses. A repeated measures ANOVA was run using the same latency intervals and factors used in the phonological contrast experiment. The only difference was that the levels of the category factor (/dæ/ and /tæ/ in the phonological experiment) were replaced with the categories 'low VOT' and 'high VOT.' Two separate ANOVAs were conducted, one for the high-VOT standards and deviants, excluding the low-VOT stimuli, and one for all stimuli combined. The combined data matches the range of conditions entered into the analysis reported for the phonological contrast

Figure 6. Comparison of difference waves in phonological and acoustic contrast experiments. (a) Grand average difference wave: /dæ–tæ/ combined. (b) Grand average difference wave: control combined.
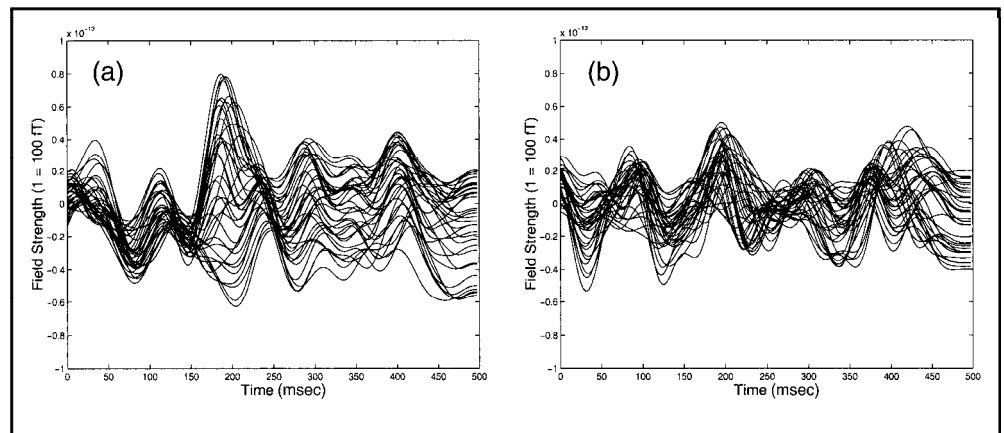
experiment above; however, it is the long-VOT deviant condition that is potentially most relevant, because this is the condition in which the phonological ratio of standards to deviants is close to 50:50, predicting the absence of an MMF.

High-VOT Conditions

In the analysis of the high-VOT conditions, no mismatch response was observed, as shown by the absence of a stimulus-type × channel interaction in the 150–210 msec latency window, $F(36, 5) = 1.088$, $p = .37$, $\varepsilon = 0.045$. In the same latency interval there was also a significant main effect of stimulus-type, $F(1, 5) = 6.90$, $p < .05$, due to more positive magnetic fields elicited by deviants than standards. In the 90–150 msec time interval there was a significant main effect of channel, $F(36, 5) = 5.63$, $p < .05$, $\varepsilon = 0.071$, due to the dipolar distribution of the M100 component.

Combined Control Conditions

In the analysis of the combined control conditions, there was no stimulus-type × channel interaction in the 150–210 msec latency window, indicating the absence of a mismatch response, $F(36, 5) = 0.313$, $p = .74$, $\varepsilon = 0.057$. The only effect that approached significance in this latency interval was a marginally significant stimulus-type × category interaction, $F(1, 5) = 5.24$, $p = .071$.

The only significant effect in the 0–90 msec time interval was a main effect of channel, $F(36, 5) = 7.00$, $p < .01$, $\varepsilon = 0.063$. Not surprisingly, a main effect of channel was also observed in the 90–150 msec time interval, $F(36, 5) = 7.59$, $p < .01$, $\varepsilon = 0.073$. These effects are both due to the dipolar distribution of the M100 and earlier response components. In the 90–150 msec interval there was also a marginally significant stimulus-type × category interaction, $F(1, 5) = 4.28$, $p = .093$, due to the fact that responses to standards were more positive than responses to deviants for stimuli in the short-VOT group, whereas the reverse was the case for stimuli in the long-VOT group. At the 210–270 msec time interval there was again a significant stimulus-type × category interaction, $F(1, 5) = 8.12$, $p < .05$, due to the same pattern of means described above. No other main effects or interactions were significant or marginally significant at any of the other time intervals used in the analysis of the combined data for the acoustic contrast experiment.

Further Comparisons

Further comparison of the phonological and acoustic contrast experiments reinforces the difference between the results of the two experiments. Comparison of the /dæ/-standard–/tæ/-deviant condition from the phonological contrast experiment and the low-VOT-

standard–high-VOT-deviant condition from the acoustic contrast experiment shows a clear contrast. As already seen above, analysis of /dæ/-standards and /tæ/-deviants shows a significant stimulus-type × channel interaction at the 150–210 msec interval, $F(36, 7) = 5.61$, $p = .013$, $\varepsilon = 0.061$. In contrast, analysis of low-VOT-standards and high-VOT-deviants shows no significant or marginally significant stimulus-type × channel interaction in the same interval, $F(36, 5) = 1.94$, $p = .19$, $\varepsilon = 0.058$. Since these standard-deviant sequences have an identical acoustic distribution of relative VOT values, and differ only in the ratios of the phonological categories /dæ/ and /tæ/, the difference in the results must be due to the presence or absence of a many-to-one contrast at the level of phonological categories.

Figure 7 provides a further comparison of the phonological and acoustic experiments. In order to focus on those components of the difference wave with a dipolar distribution across the sensor-array (i.e., activity with a probable source in the region of the auditory cortex), difference waves were first adjusted by subtracting the mean field strength across all 37 channels at each time interval from each channel. This correction preserves the distribution of field strengths across channels, but makes the mean field strength at each time interval equal to zero. Figure 7 shows the root mean squared magnetic field of the adjusted difference waves in the phonological and acoustic experiments. In the phonological experiment there is a clear MMF response at a latency of around 200 msec, but in the acoustic experiment the field strength of the difference wave does not rise above the level observed in the prestimulus interval. This contrast between the phonological and acoustic experiments shows the all-or-noth-
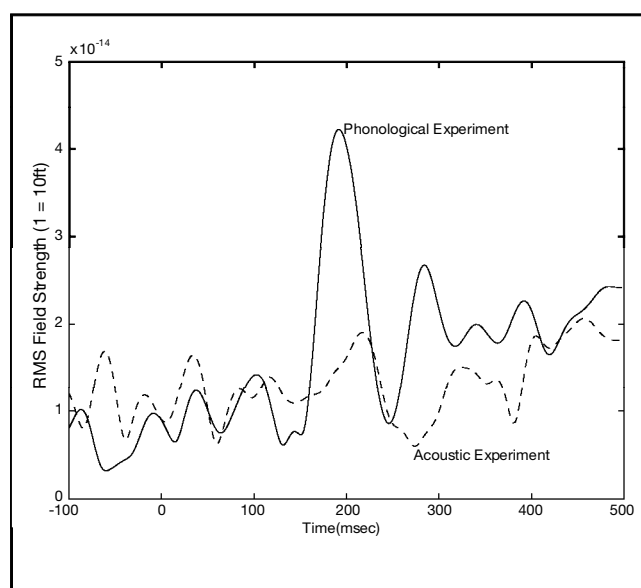

Figure 7. RMS of adjusted grand-average difference wave in main and control conditions.

ing property that is characteristic of phonological category membership.

DISCUSSION

Relationship to Electrophysiological Studies

In the mismatch studies described above, a many-to-one ratio among stimuli was confined to the level of phonological categories, by randomly selecting standards and deviants from groups of acoustically different tokens of the respective phonological categories. As a result, there was variation within the standard category and within the deviant category, and this variation occurred on exactly the same acoustic dimension that distinguished the two categories. The first experiment suggested that a many-to-one ratio at the level of phonological categories was sufficient to elicit a mismatch response. The absence of an MMF in the second experiment confirmed that phonological category representations were responsible for the MMF observed in the first experiment, since the second experiment preserved the acoustic distribution of VOT values in the first experiment, but eliminated the many-to-one ratio at the level of phonological categories.

The contrast between the two experiments makes it possible to rule out two alternative accounts of the mismatch response observed in the phonological contrast experiment. The MMF elicited by the /dæ/–/tæ/ contrast could not be due to the formation of ad-hoc groupings of acoustically similar 'short-VOT' and 'long-VOT' stimuli. If this was the cause of the MMF, then this would predict an identical result in the second experiment.

A second alternative account of the MMF in the first experiment is that deviant stimuli may be special not by virtue of their phonological category, but rather by virtue of how much each deviant stimulus differs from the immediately preceding stimulus. Adjacent pairs of standard and deviant stimuli in the experiments could differ by as little as 14–16 msec VOT, but they could also differ by as much as 72 msec VOT, with a mean difference of 40 msec VOT. On the other hand, adjacent pairs of standards differed by 0–24 msec VOT, with a mean difference of 12 msec VOT. It might be tempting to attribute the result of the first experiment to this acoustic difference between standards and deviants, but such an account would again predict identical results in the two experiments.

The contrast between the mismatch responses for /dæ/ and /tæ/ was not predicted. Comparison of standard and deviant /dæ/ stimuli showed a highly significant MMF. The same comparison for /tæ/ stimuli showed a lower-amplitude dipolar MMF that did not reach statistical significance, although the data showed a trend in the predicted direction. The difference between the two categories may be due to the fact that the /tæ/-standards elicited overall lower amplitude responses than the /dæ/-standards, leading to a worse signal-to-noise ratio for the /tæ/ conditions. This is partly expected based on existing studies that have shown lower amplitude mid-latency evoked responses to voiceless stops, relative to voiced stops (Sharma & Dorman, 1999; Gage, Poeppel, Roberts, & Hickok, 1998; Simos et al., 1998). It may also be due to the fact that the /tæ/-standards and /dæ/-deviants were presented in the second half of the phonological condition, which may have led to lower amplitude responses. However, even if the MMF response observed in the comparison of /tæ/-standards and /tæ/-deviants was reduced by such an effect, this does not compromise the main findings of the experiments. Both the /dæ/-deviants and the /tæ/-deviants show a clear MMF when compared to the /dæ/-standards. In addition, the contrast between the two experiments can only be accounted for in terms of phonological category representations.

The scalp distribution and source localization of the MMF suggests that it originates in the supratemporal auditory cortex, close to the generator of the auditory M100 response. This is similar to the source reported for mismatch responses to acoustic and phonetic contrasts in other studies.

These findings imply that the left-hemisphere auditory cortex has access to representations of the discrete phonological categories that are used to store words and that form the basis of further phonological processing. The current study is of course neutral on the issue of left hemisphere specialization for phonological processing, since we recorded only from the left hemisphere. Our results go beyond the body of evidence reviewed in the Introduction that shows that the auditory cortex supports representations of phonetic categories. Näätänen et al. (1997) showed that the amplitude of MMF responses to deviant vowels increase when the deviant vowel corresponds to a native language vowel-prototype. Winkler et al. (1999) showed that MMN amplitudes to across-category vowel contrasts are larger than to within-category contrasts. Sharma and Dorman (1999) and Dehaene-Lambertz (1997) showed that acoustic deviants in a stream of CV syllables elicit a large MMN when the deviant is drawn from a different phonetic category than the standards, but not when the standard and deviant are drawn from the same phonetic category. While these studies all clearly show effects of language-specific representations on mismatch responses, their results can be accounted for by language-specific acoustic or phonetic sensitivity, and do not necessarily implicate discrete phonological category representations. The experiments reported here do implicate phonological representations, by virtue of the fact that the generator of the MMF must treat the acoustically different tokens of each category as identical in order to detect the many-to-one ratio of standards to deviants.

Note that although the results of our experiments indicate that the left-hemisphere auditory cortex has access to symbolic representations of phonological categories, this does not entail that symbolic representations of phonological categories are stored or computed in the auditory cortex. It is quite feasible that the generator of the MMF accesses category representations that are stored in some other area of the brain. This possibility is especially feasible given the relatively late onset (150–200 msec) of the MMF response. This possibility is also supported by the finding that an auditory MMF can be elicited in a McGurk effect paradigm requiring integration of auditory and visual information (Sams et al., 1991).

Relation to Neurophysiological Mechanisms and Models

Looking beyond specific electrophysiological literature, it is important to ask whether our findings are expected based on existing results and theoretical models. We can ask this question separately about the timing and the location of the phonological mismatch response.

The fact that phonological category representations are available within 150–200 msec after stimulus onset is not surprising. A variety of sources of evidence suggest that lexical representations are available within around 250 msec of word onset. This evidence comes from the speed of speech shadowing (Marslen-Wilson, 1975), the onset of N400 responses to semantic anomaly (Helenius, Salmelin, Service, & Connolly, 1998; Kutas & Hillyard, 1984), and eye movements to named visual targets (Allopenna, Magnuson, & Tanenhaus, 1998). If words can be identified this quickly, then the phonological categories that they are built from should be identified rather earlier.

The finding that the auditory cortex accesses phonological categories is consistent with existing results and theoretical models, but it is not explicitly predicted by existing models, to our knowledge. There are many results that suggest that phonetic category representations may be supported by the auditory cortex—both in the form of theories of how phonetic category information is extracted from the acoustic input and in the form of neuroscientific results that implicate the auditory cortex in phonetic processes. None of these findings about phonetic categories in the auditory cortex guarantee that phonological categories will also be available to the auditory cortex, but they at least make it plausible.

Voicing distinctions among consonants have relatively straightforward acoustic cues, and hence there have been explicit models of how the phonetic categories voiced and voiceless could be extracted by a simple neural circuit. In particular, a number of investigators have explored the idea that the perceptual boundary for voicing reflects a constraint on temporal processing in the auditory system. At longer VOT values, the initial noise burst and the onset of voicing can be separately encoded, but at shorter VOTs (i.e., below 20–30 msec) the two acoustic events are too close to be encoded separately. A number of findings support this distinction between a 'single on' response for voiced stops and a 'double on' response for voiceless stops. This pattern has been observed in direct recordings from the chinchilla auditory nerve (Sinex & McDonald, 1988, 1989; Sinex, McDonald, & Mott, 1991) and from the primary auditory cortex of the monkey (Steinschneider, Schroeder, Arezzo, & Vaughan, 1996, 1994) and the cat (Eggermont, 1995), in scalp recordings from the guinea pig (McGee, Kraus, King, Nicol, & Carrell, 1996), scalp recordings from humans (Sharma & Dorman, 1999), and recently in direct recordings from the human auditory cortex (Steinschneider et al., 1999).

These findings strongly suggest that the phonetic category of voicing could be supported by the auditory cortex. However, single-on versus double-on encoding of voicing is a number of steps removed from the phonological categories /d/ and /t/. First, when voiceless stops are encoded as a double-on response pattern, within-category distinctions are preserved, in the form of the time lag between the two 'on' responses. As we have emphasized above, within-category differences are irrelevant to phonological categories. Second, phonetically voiced stops do not always correspond to phonologically voiced stops (and vice versa), due to language-particular processes such as the American English flapping rule or the syllable-final devoicing process found in many languages.

In sum, existing models and results are consistent with our findings, but due to the greater abstractness of phonological categories relative to phonetic categories, previous studies place relatively few constraints on when and where we should expect to find phonological category representations.

Relation to Sophistication of Auditory Cortex Representations

It is not news that phonological category representations are supported by the human brain—many years of research on phonology guarantee this. What is more surprising is that these relatively abstract, discrete linguistic categories are available to a part of the brain that is known to be involved in relatively low-level auditory processing. This finding raises questions about how auditory cortex achieves this degree of sophistication.

The functional organization of the human auditory cortex remains relatively poorly understood, despite extensive studies in a number of mammalian species (for reviews see Kaas & Hackett, 1998; Schreiner, 1992). The auditory cortex involves a number of different subareas that receive both thalamocortical and corticocortical inputs and participate in a variety of different

processes. This complexity raises the possibility that the auditory cortex may support sophisticated representations, but at present we know of few constraints on how sophisticated and abstract these representations may be.

In humans, the mismatch paradigm has been one of the main tools used to test the limits of sophistication and abstraction in the auditory cortex. Given the reasonable assumption that the mismatch generator cannot access all representations, investigators have asked what its limits are. To date, these limits have not been reached, and most auditory perceptual contrasts tested have elicited a mismatch response. A small number of studies have concluded that some representations are not supported by the auditory cortex (Ross, Tervaniemi, & Näätänen, 1996; Maiste et al., 1995; Sharma et al., 1993; Sams et al., 1990). However, in each of these studies the conclusion is based on the fact that the MMF is more sensitive than human perceivers, so that the MMF amplitude is just as large for hard-to-perceive contrasts as it is for easy-to-perceive contrasts. These findings are therefore all consistent with the possibility that the auditory cortex does indeed support the more abstract auditory representations, and thus do not impose constraints on the sophistication of the auditory cortex.

We can see at least three ways of accounting for the phonological sophistication of the auditory cortex demonstrated in our studies. The different accounts make contrasting predictions about the extent to which our findings will generalize to other categories.

First, our results may depend on the fact that voicing is an acoustically simple contrast. Due to the existence of a simple neural mechanism for identifying the phonetic categories of voiced and voiceless stops, the phonological categories may also be computed at a relatively low-level. This account predicts that our paradigm should fail when used with many other phonological categories and features. A similar account is suggested by Sharma and Dorman (1999) in order to explain the contrast between the results of their VOT mismatch study and an earlier place of articulation mismatch study (Sharma et al., 1993).

The second possibility is that phonological categories are supported by the auditory cortex because the auditory cortex is in fact quite sophisticated phonologically, and it is able to compute abstract categories. If this is the correct approach, then we should expect our finding to generalize to many other phonological categories and features, although there should still be clear limitations on the abstractness of the representations that the auditory cortex supports.

The third possibility, already mentioned above, is that the auditory cortex can access phonological category representations because it receives top-down information from other processing systems. If this is the case, then we should expect our findings to generalize yet further, with even fewer constraints on the abstractness

of the representations available to the auditory cortex. Further research is clearly needed in order to distinguish these alternatives.

CONCLUSION

In order to account for the fact that an MMF was elicited by the /dæ/–/tæ/ contrast in the phonological contrast experiment, despite the absence of any uniform acoustic contrast, we conclude that the left-hemisphere auditory cortex has access to representations of discrete phonological categories. This conclusion is supported by the absence of an MMF in the acoustic contrast experiment, which matched the acoustic distribution of stimuli in the first experiment, while removing the many-to-one distribution at the phonological level. These findings go beyond existing findings that show that the auditory cortex has access to language-specific phonetic representations. Whereas previous findings could be explained with reference to language-specific acoustic or phonetic representations, the studies reported here implicate phonological category representations, which form the basis of lexical storage and further linguistic computation. Nevertheless, we have expressed caution about what these findings show about the role of the auditory cortex in phonological processing. Although discrete phonological representations can certainly be accessed by the auditory cortex, the current results do not guarantee that it is the auditory cortex that computes or stores these representations.

METHODS

Subjects

Eight subjects participated as volunteers in the study (three males, five females) aged 21–50 (mean age 30 years). All subjects were right-handed native speakers of English, and had no hearing loss or neurological abnormalities. All eight subjects participated in the first study, and six of the eight subjects participated in the control study. All subjects gave informed consent, separately for the MEG and MRI components of the study.

Stimuli

All stimuli were selected from a synthesized /dæ/–/tæ/ continuum, created using the cascade vocal tract of the Klatt synthesizer (Klatt, 1980) for Macintosh (Sensimetrics, Cambridge, MA). Synthesis parameters were identical for all stimuli, except for VOT. All stimuli were 290 msec in duration. F0 decreased at a constant rate from 120 Hz at the beginning of the stimulus to 90 Hz after 290 msec. F1 increased from 300 to 600 Hz between 0 and 50 msec, and then to 690 Hz by 150

msec; F2 decreased from 1920 to 1770 Hz between 0 and 50 msec, and then to 1660 Hz by 150 msec; F3 decreased from 2780 to 2580 Hz between 0 and 50 msec, and then to 2490 Hz by 150 msec. All stimuli began with a 5-msec burst of frication at 30 dB, followed by aspiration set at 15 dB that lasted from 5 to 265 msec before decreasing to zero. Amplitude of voicing differed between stimuli: It began at 55 dB at the onset of voicing, then increased to 60 dB over 30 msec, remained constant until 165 msec, decreased to 58 dB by 230 msec, to 45 dB by 250 msec, and then to zero by 270 msec. Synthesis parameters and sound files for all stimuli are available on request from the first author.

Eight stimuli from the VOT continuum were selected for each of the two experiments. Specific stimuli were selected for each subject individually, based on the results of a behavioral pretest that determined the subject's perceptual boundary using a standard forced-choice identification paradigm. The pretest was conducted inside the magnetically shielded room, using the same stimulus delivery equipment as in the MEG experiments. Based on the pretest results, four evenly spaced stimuli were chosen from each of the categories /dæ/ and /tæ/. The difference between the longest VOT /dæ/ and the shortest VOT /tæ/ was twice the difference between individual stimuli within categories. For example, stimuli for a subject with a VOT boundary of 32 msec would be 0, 8, 16, 24, 40, 48, 56, and 64 msec. For the acoustic contrast condition, a further eight stimuli were selected, such that the acoustic spacing between stimuli remained exactly as in the main study, but now only two stimuli fell below the perceptual boundary. For the same subject with a 32-msec perceptual boundary, the eight stimuli used would be 20, 28, 36, 44, 60, 68, 76, and 84 msec.

Design

Stimulus presentation was controlled by a Macintosh Quadra computer running the Psyscope package (Cohen, MacWhinney, Flatt, & Provost, 1993). In the first half of each run, one phonological category was the standard, and the other category was the deviant; in the second half of the run, the roles of the two categories were reversed. This procedure was used in order to allow comparison of the same stimulus as standard and oddball, thereby eliminating any effect of acoustic differences between standards and deviants in the MMF analysis. Both standards and deviants were selected randomly from the four stimuli in each category, and deviants were selected with a probability of 12.5% Deviants never occurred in immediate succession. In both halves of the experiment 700 standards and 100 deviants were presented, for a total of 1600 stimuli. Interstimulus intervals were selected randomly from a range of values between 700 and 900 msec, in order to exclude artifacts of periodic origin, and to reduce the attenuation of responses caused by presentation of stimuli at fixed ISIs. Each run lasted approximately 30 min. Schematic diagrams of stimulus presentation sequences are shown in Figure 1.

Subjects were instructed to attend passively to the stimulus sequence, which was delivered monaurally to the right ear. No secondary task was used. Although many studies using the mismatch paradigm have presented sounds while subjects perform a secondary task, such as reading a book or watching a movie, a number of recent studies have shown that MMF responses may be elicited in the absence of a secondary task (e.g., Alho et al., 1998; Woldorff et al., 1998; Dehaene-Lambertz, 1997).

MEG Recordings

The recordings were performed in a magnetically shielded room using a 37-channel magnetometer with SQUID-based first-order gradiometer sensors (Magnes, Biomagnetic Technologies, San Diego, CA). The sensor-array was placed over the left temporal area, contralateral to the stimulus presentation, so as to optimally record the M100 elicited by a reference 1000 Hz tone presented at approximately 60 dB SPL. The position of a participant's head with respect to the sensors was recorded by a transceiver-based system that localized fiducial landmarks, thereby setting up a spatial reference frame for the MEG data. MEG epochs of duration 600 msec were collected around each stimulus (100 msec pretrigger). Recordings were made using a 1.0-Hz high-pass cutoff, and a sampling rate of 520 Hz with a 200-Hz bandwidth.

For each stimulus condition, all epochs were first averaged and then filtered using a 1–20-Hz digital band-pass filter. All further analyses were performed on the filtered data.

MRI Recordings

For each subject, high-resolution volumetric magnetic resonance images (SPGR sequence, $128 \times 128 \times 124$ matrix, resolution $\sim 1 \times 1 \times 1.5$ mm, TR = 36 msec, TE = 8 msec, flip = 70°) were acquired using a 1.5-T SIGNA magnetic resonance scanner (GE Medical Systems, Milwaukee, WI). The MR images were judged to be normal by a neuroradiologist.

Acknowledgments

REFERENCES

Aaltonen, O., Eerola, O., Hellström, A., Uusipaikka, E., & Lang, A. (1997). Perceptual magnet effect in the light of behavioral and psychophysiological data. *Journal of the Acoustical Society of America, 101,* 1090–1105.

Aaltonen, O., Niemi, P., Nyrke, T., & Tuhkanen, M. (1987). Event-related brain potentials and the perception of a phonetic continuum. *Biological Psychology, 24,* 197–207.

Alho, K. (1995). Cerebral generators of mismatch negativity (MMN) and its magnetic counterpart (MMNm) elicited by sound changes. *Ear & Hearing, 16,* 38–51.

Alho, K., Winkler, I., Escera, C., Huotilainen, M., Virtanen, J., Jääskeläinen, I., Pekkonen, E., & Ilmoniemi, R. (1998). Processing of novel sounds and frequency changes in the human auditory cortex: Magnetoencephalographic recordings. *Psychophysiology, 35,* 211–224.

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken *word* recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory & Language, 38,* 419–439.

Aulanko, R., Hari, R., Lounasmaa, O. V., Näätänen, R., & Sams, M. (1993). Phonetic invariance in the human auditory cortex. *NeuroReport, 4,* 1356–1358.

Barton, D. (1978). *The role of perception in the acquisition of phonology.* Unpublished doctoral dissertation, Stanford University.

Basso, A., Casati, G., & Vignolo, L. A. (1977). Phonemic identification deficits in aphasia. *Cortex, 13,* 84–95.

Bertoncini, J., Bijeljac-Babic, R., Blumstein, S. E., & Mehler, J. (1987). Discrimination in neonates of very short CVs. *Journal of the Acoustical Society of America, 82,* 31–37.

Binder, J. R., Frost, J. A., Hammeke, T. A., Cox, R. W., Rao, S. M., & Prieto, T. (1997). Human brain language areas identified by functional magnetic resonance imaging. *Journal of Neuroscience, 17,* 353–362.

Blumstein, S. E. (1973). *A phonological investigation of aphasic speech.* The Hague: Mouton.

Blumstein, S. E. (1995). The neurobiology of the sound structure of language. In M. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 915–929). Cambridge: MIT Press.

Blumstein, S. E., Cooper, W. E., Goodglass, H., Statlender, S., & Gottlieb, J. (1980). Production deficits in aphasia: A voice-onset time analysis. *Brain & Language, 9,* 153–170.

Blumstein, S. E., Cooper, W. E., Zurif, E. B., & Caramazza, A. (1977). The perception and production of voice-onset time in aphasia. *Neuropsychologia, 15,* 371–383.

Boatman, D., Lesser, R. P., & Gordon, B. (1995). Auditory speech processing in the left temporal lobe: An electrical interference study. *Brain & Language, 51,* 269–290.

Carney, A. E., Widin, G. P., & Viemeister, N. F. (1977). Noncategorical perception of stop consonants differing in VOT. *Journal of the Acoustical Society of America, 62,* 961–970.

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers, 25,* 257–271.

Dehaene-Lambertz, G. (1997). Electrophysiological correlates of categorical phoneme perception in adults. *NeuroReport, 8,* 919–924.

Eggermont, J. J. (1995). Representation of voice onset time continuum in primary auditory cortex of the cat. *Journal of the Acoustical Society of America, 98,* 911–920.

Eimas, P. D. (1974). Auditory and linguistic processing of cues for place of articulation by infants. *Perception and Psychophysics, 16,* 513–521.

Eimas, P., Siqueland, E., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science, 171,* 303–306.

Fiez, J. A., Raichle, M. E., Miezin, F. M., Petersen, S. E., Tallal, P., & Katz, W. F. (1995). PET studies of auditory and phonological processing: Effects of stimulus characteristics and task demands. *Journal of Cognitive Neuroscience, 7,* 357–375.

Fry, D. B., Abramson, A. S., Eimas, P. D., & Liberman, A. M. (1962). The identification and discrimination of synthetic vowels. *Language and Speech, 5,* 171–189.

Gage, N., Poeppel, D., Roberts, T., & Hickok, G. (1998). Auditory evoked M100 reflects onset acoustics of speech sounds. *Brain Research, 814,* 236–239.

Gandour, J., & Dardaradanda, R. (1982). Voice onset time in aphasia: Thai, I. Perception. *Brain & Language, 17,* 24–33.

Gandour, J., & Dardaradanda, R. (1984). Voice onset time in aphasia: Thai, II. Production. *Brain & Language, 18,* 389–410.

Gomes, H., Ritter, W., & Vaughan, H. (1995). The nature of preattentive storage in the auditory system. *Journal of Cognitive Neuroscience, 7,* 81–94.

Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika, 24,* 95–112.

Hari, R., Hämäläinen, M., Ilmoneimi, R., Kaukoranta, E., Reinikainen, K., Salminen, J., Alho, K., Näätänen, R., & Sams, M. (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.

Helenius, P., Salmelin, R., Service, E., & Connolly, J. F. (1998). Distinct time courses of word and context comprehension in the left temporal cortex. *Brain, 121,* 1133–1142.

Itoh, M., Sasanuma, S., Tatsumi, I., Murakami, S., Fukusako, Y., & Suzuki, T. (1982). Voice onset time characteristics in apraxia of speech. *Brain & Language, 17,* 193–210.

Kaas, J. H., & Hackett, T. A. (1998). Subdivisions of auditory cortex and levels of processing in primates. *Audiology and Neurootology, 3,* 73–85.

Kenstowicz, M. (1994). *Phonology in generative grammar.* Cambridge: Blackwell.

Klatt, D. (1980). Software for a cascade/parallel formant synthesizer. *Journal of the Acoustical Society of America, 67,* 971–995.

Kluender, K. R., Diehl, R. L., & Killeen, P. R. (1987). Japanese quail can learn phonetic categories. *Science, 237,* 1195–1197.

Kraus, N., McGee, T., Micco, A., Sharma, A., Carrell, T., & Nicol, T. (1993). Mismatch negativity in school-age children to speech stimuli that are just perceptibly different. *Electroencephalography and Clinical Neurophysiology, 88,* 123–130.

Kuhl, P. K. (1991). Human adults and infants show a 'perceptual magnet effect' for the prototypes of speech categories, monkeys do not. *Perception and Psychophysics, 50,* 93–107.

Kuhl, P. K., & Miller, J. D. (1978). Speech perception by the chinchilla: Voiced–voiceless distinction in alveolar plosive consonants. *Science, 190,* 69–72.

Kuhl, P. K., & Padden, D. M. (1983). Enhanced discriminability at the phonetic boundaries for the place feature in macaques. *Journal of the Acoustical Society of America, 73,* 1003–1010.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307,* 161–163.

Liberman, A. M., Harris, K. S., Kinney, J. A., & Lane, H. L. (1961). The discrimination of relative-onset time of the components

of certain speech and non-speech patterns. *Journal of Experimental Psychology, 61,* 379–388.

Maiste, A., Wiens, A., Hunt, M., Scherg, M., & Picton, T. (1995). Event-related potentials and the categorical perception of speech sounds. *Ear & Hearing, 16,* 67–89.

Marslen-Wilson, W. (1975). Sentence perception as an interactive parallel process. *Science, 189,* 226–228.

McGee, T., Kraus, N., King, C., Nicol, T., & Carrell, T. D. (1996). Acoustic elements of speechlike stimuli are reflected in surface recorded responses over the guinea pig temporal lobe. *Journal of the Acoustical Society of America, 99,* 3606–3614.

Miceli, G., Caltagirone, C., Gainotti, G., & Payer-Rigo, P. (1978). Discrimination of voice versus place contrasts in aphasia. *Brain & Language, 2,* 434–450.

Miller, J. (1994). On the internal structure of phonetic categories: A progress report. *Cognition, 50,* 271–285.

Miller, J. D., Weir, C. C., Pastore, L., Kelly, W. J., & Dooling, R. J. (1976). Discrimination and labeling of noise-buzz sequences with varying noise-lead times: An example of categorical perception. *Journal of the Acoustical Society of America, 60,* 410–417.

Molfese, D. (1987). Electrophysiological indices of categorical perception for speech. In S. Harnad (Ed.), *Categorical perception* (pp. 421–443). Cambridge, UK: Cambridge University Press.

Näätänen, R. (1992). *Attention and brain function.* Hillsdale, NJ: Erlbaum.

Näätänen, R. (1995). The mismatch negativity: A powerful tool for cognitive neuroscience. *Ear & Hearing, 16,* 6–18.

Näätänen, R., Gaillard, A. W. K., & Mäntysalo, S. (1978). Early selective attention effect on evoked potential reinterpreted. *Acta Psychologica, 42,* 313–329.

Näätänen, R., Lehtoskoskl, A., Lennes, M., Cheour, M., Huotilainen, M., Ilvonen, A., Vainio, M., Alku, P., Ilmoniemi, R., Luuk, A., Allik, J., Sinkkonen, J., & Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature, 385,* 432–434.

Pastore, R. E., Ahroon, W. A., Buffuto, K. A., Friedman, C. J., Puleo, J. S., & Fink, E. A. (1977). Common factor model of categorical perception. *Journal of Experimental Psychology: Human Perception and Performance, 4,* 686–696.

Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception and Psychophysics, 13,* 253–260.

Pisoni, D. B. (1977). Identification and discrimination of the relative onset of two component tones: Implications for voicing perception in stops. *Journal of the Acoustical Society of America, 61,* 1352–1361.

Pisoni, D. B., Carrell, T. D., & Gans, S. J. (1983). Perception of the duration of rapid spectrum changes: Evidence for context effects with speech and nonspeech signals. *Perception and Psychophysics, 34,* 314–322.

Remez, R. E., Rubin, P. E., Burns, S. M., Pardo, J. S., & Lang, J. M. (1994). On the perceptual organization of speech. *Psychological Review, 101,* 129–156.

Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. (1981). Speech perception without traditional cues. *Science, 212,* 947–950.

Ross, J., Tervaniemi, M., & Näätänen, R. (1996). Neural mechanisms of the octave illusion: Electrophysiological evidence for central origin. *NeuroReport, 8,* 303–306.

Sams, M., Aulanko, R., Aaltonen, O., & Näätänen, R. (1990). Event-related potentials to infrequent changes in synthesized phonetic stimuli. *Journal of Cognitive Neuroscience, 2,* 344–355.

Sams, M., Aulanko, R., Hämäläinen, M., Hari, R., Lounasmaa, O., Lu, S.-T., & Simola, J. (1991). Seeing speech: Visual information from lip movements modifies activity in the human auditory cortex. *Neuroscience Letters, 127,* 141–145.

Sams, M., Kaukoranta, E., Hämäläinen, M., & Näätänen, R. (1991). Cortical activity elicited by changes in auditory stimuli: Different sources for the magnetic N100m and mismatch responses. *Psychophysiology, 28,* 21–29.

Sams, M., Paavilainen, P., Alho, K., & Näätänen, R. (1985). Auditory frequency discrimination and event-related potentials. *Electroencephalography and Clinical Neurophysiology, 62,* 437–448.

Samuel, A. G. (1982). Phonetic prototypes. *Perception and Psychophysics, 31,* 307–314.

Scherg, M., Vajsar, J., & Picton, T. W. (1989). A source analysis of the late human auditory evoked potentials. *Journal of Cognitive Neuroscience, 1,* 336–355.

Schreiner, C. E. (1992). Functional organization of the auditory cortex: Maps and mechanisms. *Current Opinion in Neurobiology, 2,* 516–521.

Sharma, A., & Dorman, M. F. (1999). Cortical auditory evoked potential correlates of categorical perception of voice-onset time. *Journal of the Acoustical Society of America, 106,* 1078–1083.

Sharma, A., Kraus, N., McGee, T., Carrell, T., & Nicol, T. (1993). Acoustic versus phonetic representation of speech as reflected by the mismatch negativity event-related potential. *Electroencephalography and Clinical Neurophysiology, 88,* 64–71.

Simos, P. G., Diehl, R. L., Breier, J. I., Molis, M. R., Zouridakis, G., & Papanicolaou, A. C. (1998). MEG correlates of categorical perception of a voice onset time continuum in humans. *Cognitive Brain Research, 7,* 215–219.

Sinex, D. G., & McDonald, L. P. (1988). Average discharge rate representation of voice onset time in the chinchilla auditory nerve. *Journal of the Acoustical Society of America, 83,* 1817–1827.

Sinex, D. G., & McDonald, L. P. (1989). Synchronized discharge rate representation of voice onset time in the chinchilla auditory nerve. *Journal of the Acoustical Society of America, 85,* 1995–2004.

Sinex, D. G., McDonald, L. P., & Mott, J. B. (1991). Neural correlates of non-monotonic temporal acuity for voice onset time. *Journal of the Acoustical Society of America, 90,* 2441–2449.

Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature, 338,* 381–382.

Steinschneider, M., Schroeder, C. E., Arezzo, J. C., & Vaughan, H. G. (1994). Speech-evoked activity in primary auditory cortex: Effects of voice onset time. *Electroencephalography and Clinical Neurophysiology, 92,* 30–43.

Steinschneider, M., Schroeder, C. E., Arezzo, J. C., & Vaughan, H. G. (1996). Physiological correlates of the voice onset time boundary in primary auditory cortex of the awake monkey—temporal response patterns. *Brain and Language, 48,* 7326–340.

Steinschneider, M., Volkov, I. O., Noh, M. D., Garell, P. C., & Howard, M. A. (1999). Temporal encoding of the voice onset time phonetic parameter by field potentials recorded directly from human auditory cortex. *Journal of Neurophysiology, 82,* 2346–2357.

Stevens, K. N., Liberman, A.M., Studdert-Kennedy, M. G., & Ohman, S. E. G. (1969). Cross-language study of vowel perception. *Language and Speech, 12,* 1–23.

Swoboda, P., Morse, P. A., & Leavitt, L. A. (1976). Continuous vowel discrimination in normal and at-risk infants. *Child Development, 47,* 459–465.

Volaitis, L. E., & Miller, J. L. (1992). Phonetic prototypes: In-fluence of place of articulation and speaking rate on the internal structure of voicing categories. *Journal of the Acoustical Society of America, 77,* 1907–1912.

Werker, J. F., & Lalonde, C. E. (1988). Cross-language speech perception: Initial capabilities and developmental change. *Developmental Psychology, 24,* 672–683.

Werker, J. F., & Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Perception and Psychophysics, 37,* 35–44.

Winkler, I., Lehtokoski, A., Alku, P., Vaiiio, M., Czigler, I., Csepe, V., Aaltonen, O., Raimo, I., Alho, K., Lang, H., Ii-vonen, A., & Näätänen, R. (1999). Pre-attentive detection of vowel contrasts utilizes both phonetic and auditory memory representations. *Cognitive Brain Research, 7,* 357–369.

Winkler, I., Paavilainen, K., Alho, K., Reinikainen, K., Sams, M., & Näätänen, R. (1990). The effect of small variation of the frequent auditory stimulus on the event-related brain potential to the infrequent stimulus. *Psychophysiology, 27,* 228–235.

Woldorff, M., Hillyard, S., Gallen, C., Hampson, S., & Bloom, F. (1998). Magnetoencephalographic recordings demonstrate attentional modulation of mismatch-related neural activity in human auditory cortex. *Psychophysiology, 35,* 283–292.

Zatorre, R. J., Evans, A. C., Meyer, E., & Gjedde, A. (1992). Lateralization of phonetic and pitch discrimination in speech processing. *Science, 256,* 846–849.

Zatorre, R. J., Meyer, E., Gjedde, A., & Evans, A. C. (1996). PET studies of phonetic processing of speech: Review, replication and reanalysis. *Cerebral Cortex, 6,* 21–30.