



Regular article

Levels of representation in the electrophysiology of speech perception

Colin Phillips*

University of Maryland, Department of Linguistics, 1401 Marie Mount Hall, College Park, MD 20742, USA

Abstract

Mapping from acoustic signals to lexical representations is a complex process mediated by a number of different levels of representation. This paper reviews properties of the *phonetic* and *phonological* levels, and hypothesizes about how category structure is represented at each of these levels, and evaluates these hypotheses in light of relevant electrophysiological studies of phonetics and phonology. The paper examines evidence for two alternative views of how infant phonetic representations develop into adult representations, a *structure-changing* view and a *structure-adding* view, and suggests that each may be better suited to different kinds of phonetic categories. Electrophysiological results are beginning to provide information about phonological representations, but less is known about how the more abstract representations at this level could be coded in the brain. © 2001 Cognitive Science Society, Inc. All rights reserved.

Keywords: Phonetics; Speech perception; Electrophysiology; Mismatch negativity

1. Introduction

Speech perception involves a mapping from continuous acoustic waveforms onto the discrete phonological units used to store words in the mental lexicon. For example, when we hear the word *cat*, we map a complex and continuous pattern of vibration at the eardrum onto a phonological percept which has just three clearly distinct pieces: /k/, /æ/ and /t/. A great

* Tel.: +1-301-405-3082; fax: +1-301-405-7104.

E-mail address: colin@glue.umd.edu (C. Phillips).

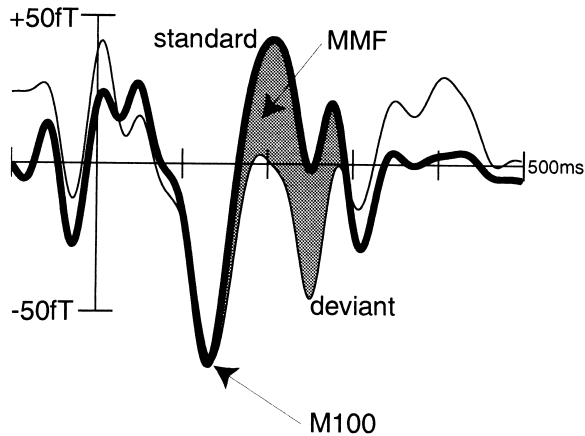


Fig. 1. Sample averaged MEG response to speech sounds, recorded in a mismatch paradigm. Both frequent ('standard') sounds and infrequent ('deviant') sounds elicit a similar M100 response. Responses to standards and deviants begin to diverge in the 150–200ms latency range. This divergence is known as the *Mismatch Negativity* (EEG) or the *Magnetic Mismatch Field* (MEG).

deal of evidence indicates that this mapping from sound to words is not a simple one-step mapping, but is instead mediated by a number of different levels of representation. This article reviews studies of how the brain supports the different levels of representation, with a focus on work using the electrophysiological measures electroencephalography (EEG) and magnetoencephalography (MEG), and how it relates to hypotheses derived from behavioral and theoretical research in phonetics and phonology.

EEG and MEG provide noninvasive measures of neuronal activity in the brain. Electrodes positioned on the scalp, or magnetic field sensors positioned close to the scalp, measure changes in scalp voltages or scalp magnetic fields, entirely passively and with millisecond resolution. These measures are direct, and provide excellent temporal resolution, but only modest localization information. Whereas the other papers in this special issue discuss findings involving a number of different brain regions, most of the existing electrophysiological findings about speech perception have focused on evoked responses occurring within 250 ms after a sound is presented, and generated in auditory cortex. Human auditory cortex is situated on the superior plane of the temporal lobe, that is, on the lower side of the Sylvian fissure. Human auditory cortex consists of a number of subareas, but most electrophysiological findings about speech perception do not reliably implicate specific subareas of auditory cortex.

A great deal of electrophysiological research on speech has focused on two evoked responses: the *auditory N100* and the *mismatch response* (see Fig. 1). The auditory N100, and its magnetic counterpart M100, are often referred to as *exogenous* responses, meaning that they are evoked by any acoustic stimulus with a well-defined onset, regardless of the listener's task or attentional state (Näätänen & Picton, 1987). However, the latency, amplitude and localization of the N100 vary reliably when certain acoustic and perceptual

parameters are varied, and there are reports of task-related modulation of M100 (e.g., Poeppel et al., 1996).

The auditory mismatch paradigm has been the most productive paradigm in the electrophysiological study of speech, and has revealed evidence of a number of different levels of representation. When a sequence of identical sounds, known as *standards*, is interrupted by infrequent *deviant* sounds, the deviant sounds elicit a characteristic response component known as the *Mismatch Negativity* (MMN) or *Magnetic Mismatch Field* (MMF; Näätänen, Gaillard & Mäntysalo, 1978; Näätänen & Winkler, 1999). The mismatch response typically occurs 150–250 ms after the onset of the deviant sound, and when localization information is available, it typically implicates supratemporal auditory cortex (Hari et al., 1984; Scherg, Vajsar, & Picton, 1989; Sams, Kaukoranta, Hämäläinen, & Näätänen, 1991; Alho, 1995).¹ It can only be seen by comparing responses to infrequent deviant sounds to responses to frequent standard sounds. It can be elicited even when the subject does not attend to the stimuli: studies in this paradigm often present sounds while subjects read a book or watch a silent movie. Hence, it is commonly used as a measure of preattentive processing. However, it is also elicited when subjects actively attend to the stimuli.

Importantly for research on speech, the amplitude of the mismatch response also tends to increase as the discriminability of the standard and deviant stimuli increases (Sams, Paavilainen, Alho, & Näätänen, 1985; Lang et al., 1990; Aaltonen, Tuomainen, Laine & Niemi, 1993; Tiitinen, May, Reinikainen, & Näätänen, 1994). For this reason a number of studies have examined whether mismatch response amplitudes track the discriminability profiles established in behavioral phonetics research, or are affected by the categorial status of pairs of sounds. The focus of this paper is on evidence for different levels of representation and how they might be encoded in the brain.

2. Multiple levels of representation

It is standard to distinguish at least the levels of *acoustics*, *phonetics*, and *phonology* in the representation of speech sounds, but there is also a good deal of evidence that these levels should be further subdivided.

At one end of the speech perception process, in the peripheral auditory system, is a fairly faithful analog representation of the *acoustics* of speech, which is most likely not modified by exposure to specific languages.

At the other end of the process are discrete, abstract *phonological* representations, which can be manipulated by symbolic processes. Phonological representations differ substantially across speakers of different languages, and cannot be derived by means of a simple transform from acoustic or phonetic representations (see Section 4).

What lies between these endpoints is less clear. There are at least one, and possibly multiple intermediate *phonetic* representations, which are also analog, but which organize speech sounds into linguistically-relevant categories. Linguistically-relevant acoustic distinctions are represented more strongly, and linguistically-irrelevant distinctions are represented less strongly. Whereas phonological categories resemble discrete, abstract categories like the integers ‘3’ and ‘42’, phonetic categories may be likened to gradable concept

categories like ‘bird’ or ‘tree’. Some key properties of phonetic representations are the following.

First, there are the phonetic representations found in young infants (below 6 months of age). Infants begin life as ‘universal listeners’, with the ability to discriminate speech contrasts that are phonetically relevant in any of the world’s languages (Streeter, 1976; Werker & Tees, 1984; Jusczyk, 1997). This ability alone could be accomplished by a rich representation of the acoustics of speech. More striking are demonstrations of categorical perception in young infants, that is, good discrimination of sounds drawn from separate phonetic categories, and poor discrimination of sounds drawn from the same phonetic category. Categorical perception has been demonstrated in infants for voicing contrasts (Eimas, Siqueland, Jusczyk, & Vigorito, 1971), place of articulation contrasts (Eimas, 1974), stop-glide (*ba* vs. *wa*) contrasts (Eimas & Miller, 1980), and r-l contrasts (Eimas, 1975). Therefore, infants bring to the language learning task representations which already divide up speech into categories resembling those found in adult languages.

Some investigators have argued that the innate representations that infants bring to the language learning task simply reflect the natural boundaries found in the sophisticated *acoustic* representations that humans share with other species (e.g., Kuhl, 1991b), based on the finding that nonhuman animals show categorical perception effects for certain speech contrasts (Kuhl & Miller, 1975; Kuhl & Padden, 1982, 1983; Kluender, Diehl, & Killeen, 1987). It is currently unclear whether infants’ initial phonetic representations go beyond what is found in nonhuman animals. One striking case of infant phonetic sophistication that goes beyond what has been observed in nonhuman animals is the finding that 6–8 month old Canadian infants group sounds from a synthetic labial-dental-retroflex (/ba-da-Da/) continuum in the same way as adult Hindi speakers (Werker & Lalonde, 1988).

Research on infant speech perception has also shown that native language phonetic categories emerge very early in life. By around 6 months of age infants show greater sensitivity to native-language vowel categories (Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992), and by around 10 months of age infants can only reliably discriminate native-language phonetic contrasts (Werker & Tees 1984, Werker & Tees 1999; Tsushima et al., 1994; Werker, 1995).

In addition to changes in category boundaries, reflecting the native language, a number of studies have also shown that development results in the formation of category ‘prototypes’ or ‘perceptual magnets’ (Samuel, 1982; Kuhl, 1991a; Volaitis & Miller, 1992; Miller, 1994). Kuhl has argued that early experience has the effect of explicitly *reshaping* perceptual space in such a way that the space is ‘compressed’ in the areas around the prototype sounds, and ‘stretched’ around the boundaries between categories. This has the effect that native-language contrasts show an enhanced representation, and non-native categories are less strongly represented, and predicts that development of language-specific representations ‘recursively alters the initial state of speech perception’ and ‘functionally erases’ non-native boundaries (Kuhl, 1994, Kuhl, 2000). This view of phonetic development has been likened to the development of self-organizing maps (Kohonen & Hari, 1999), a class of neural network models whose properties are well described.

Although a good deal is now known about how speech perception *behavior* changes in the first year of life and beyond, it is somewhat less clear how these changes are brought about

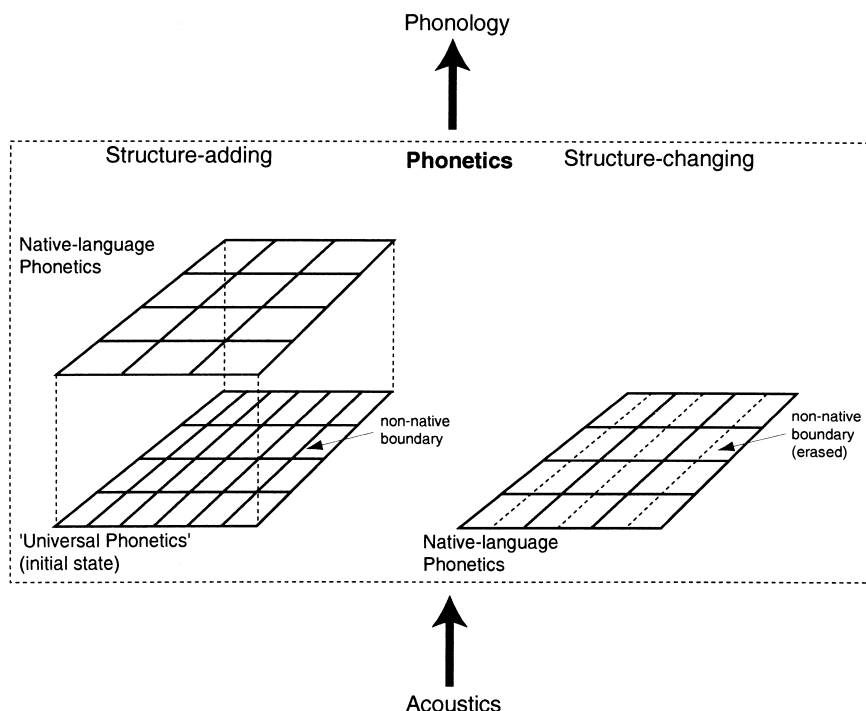


Fig. 2. Schematic illustration of alternative approaches to mature phonetic representations.

by changes in the brain representation of speech sounds. Does the change from the ‘infant perceptual space’ to the ‘adult perceptual space’ involve a literal reshaping of the infant representations, and the genuine loss of the representation of non-native boundaries, in what might be termed a *structure-changing* approach? This amounts to a direct translation of Kuhl’s claims about the reshaping of ‘perceptual space’ into claims about the reshaping of ‘representational space’, although Kuhl’s theoretical claims do not entail a commitment to this approach.² Alternatively, does the change in perception result from the formation of additional phonetic representations, which filter sounds according to native language categories, without explicitly modifying the infant’s phonetic representations, in what might be termed a *structure-adding* approach? Werker has argued for this view, emphasizing that ‘developmental change does not involve loss’ (Werker, 1994), and presenting evidence of preservation of certain non-native perceptual boundaries in adults (e.g., Werker & Logan, 1985).³ The alternatives are schematized in Fig. 2.

A number of studies have used electrophysiological measures to try to investigate brain representations of speech, with the aim of seeking corroborating evidence for different levels of representation, including preattentive levels which are hard to probe using behavioral measures. Section 3 argues that in order to represent the various analog categories of phonetics, different categories call for a variety of different representational formats. Section 4 argues that the discrete categories of phonology call for more uniform representations (at least, within the domain of segments and features).

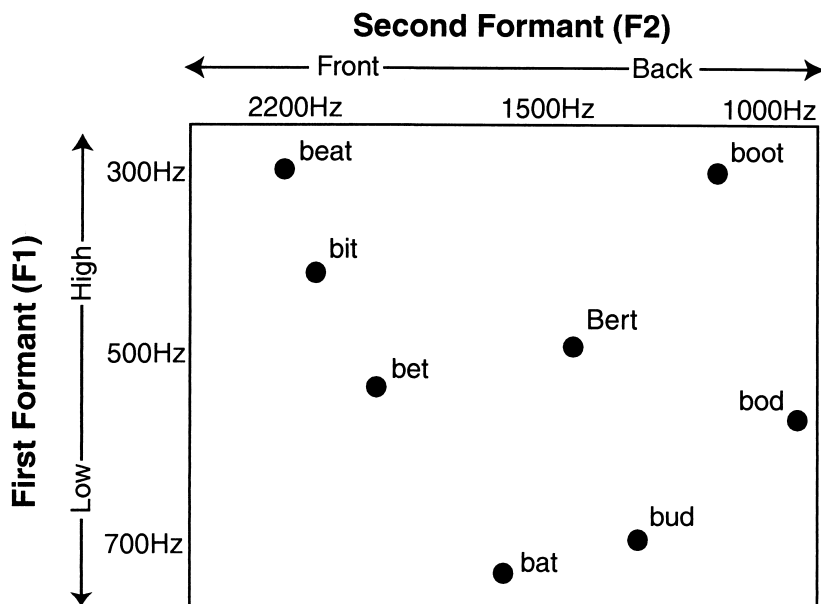


Fig. 3. 'Vowel space' of F1 and F2 coordinates for selected vowels (British English).

3. Phonetic representations

3.1. Vowels

Vowels are relatively easy to study, because their acoustic properties are well-described and easily manipulated. A good deal of information about vowel height, backness and rhotacism is carried in the values of the first three *formants* of the vowel (F1, F2, F3) respectively. The fundamental frequency (F0) of the vowel primarily conveys information about affect, focus and speaker identity, but is relatively unimportant in the identification of vowel categories.

The structure-changing approach can be clearly articulated—and has gained its best support—in the case of vowels and the liquids [r,l], that is, sounds with significant *formant* structure. This approach provides a clear hypothesis about the neural representation of vowel categories. The phonetics of vowels are assumed to be represented in an abstract 'map' (cf. Fig. 3), which may (but need not) be encoded spatially in a region of cortex. Native-language vowel prototypes form attractor basins in this map. Category boundaries correspond to points in the vowel space where the gravitational pull of a pair of category prototypes is equal and opposite. The boundaries of *non-native* categories are lost (or greatly attenuated). See Kuhl (1994, 2000) for a description of such a map of 'vowel space', although in terms which do not explicitly commit to a structure-changing approach.

Behavioral evidence on vowel perception is consistent with the structure-changing approach. Infants form 'perceptual magnets' around native language vowels as early as 6

months of age (Kuhl, 1991a; Kuhl et al., 1992), and there is no clear evidence that adolescents or adults retain the innate ‘universal’ vowel space after they have acquired their native language. Such findings raise the possibility that cortex supports a single phonetic ‘vowel map’, and accordingly a number of electrophysiological studies have searched for evidence for such a map in auditory cortex.

Some electrophysiological studies of vowels have shown that the M100 response reflects aspects of vowel category perception. Building upon the finding that M100 latency varies with changes in the pitch of a pure tone (Roberts & Poeppel, 1996), it has been shown that the latency of M100 responses to vowels is sensitive to variation in vowel category (/a, i, u/), specifically to variation in vowel height (F1), and is relatively insensitive to variation in speaker, that is, male versus female voice (Poeppel et al., 1997; Govindarajan, Phillips, Poeppel, Roberts, & Marantz, 1998). Studies of this kind show that the M100 is sensitive to certain important cues to vowel categorization, but do not show *how* vowel categories are represented.

In the mismatch paradigm, a number of recent studies have shown clear effects of the phonetic status of vowel categories on the amplitude of the mismatch response. This has been shown most strikingly using cross-language designs. Näätänen et al. (1997) and Winkler et al. (1999) compared mismatch responses elicited by vowel contrasts in speakers of Finnish, Estonian and Hungarian. In each group of speakers, vowel contrasts involving native language vowel prototypes elicited larger mismatch responses than contrasts involving non-native vowel prototypes, independent of the contribution of acoustic distance between vowels. Aaltonen, Eerola, Hellström, Uusipaikka and Lang (1997) report similar findings in a single-language study of Finnish vowels: perceptual discriminability predicts MMN strength (but cf. Sharma & Dorman, 1998 for conflicting evidence).

Note that although electrophysiological results such as these are compatible with the structure-changing approach, most are also compatible with the alternative structure-adding approach, since the structure-adding approach assumes richer representations than the structure-changing approach. Unfortunately, the best evidence for the structure-changing view would be *negative* results showing the unavailability of non-native categories.

The notion of a ‘vowel space’ can be clearly described at an abstract level, but more specific evidence on how this abstract space is encoded physiologically has been elusive. A natural hypothesis is that a vowel space like the one in Fig. 3 maps directly onto a two-dimensional region of cortex, in which cells are sensitive to specific coordinates of F1 or F2 values. Inspired by MEG studies showing evidence of ‘tonotopic’ maps in auditory cortex (Romani, Williamson & Kaufman, 1982; Pantev et al., 1988), and by the many well-known sensory maps in the visual and somatosensory systems, some researchers have investigated the possibility of a spatially represented vowel space. Although it is difficult to draw conclusions from negative findings, results so far have not been encouraging. Diesch, Eulitz, Hampson, and Ross (1996) examined single-dipole localizations of two responses (M100 and sustained field) evoked in auditory cortex by a number of German vowels (/a, i, u, æ, ø/). Although there was a tendency for responses to vowels to localize differently from responses to a 1 kHz pure tone, and a tendency for more acoustically distinct vowels to show more distinct localizations, no consistent evidence for a cortical vowel map was found.

In summary, electrophysiological recordings back up other kinds of evidence for the properties of phonetic vowel representations. The available evidence is consistent with the claim of the structure-changing approach, that native language vowel representations are the result of reshaping innate phonetic representations, but this evidence is also ultimately compatible with the alternative structure-adding approach. Meanwhile, there is no clear evidence to date for a spatial encoding of a ‘vowel space’.

3.2. Voicing

The different articulatory gestures which define different speech sounds produce diverse acoustic consequences, and these in turn demand diverse neural coding hypotheses, involving space, time or intensity. An important cue for voicing in syllable-initial stop consonants is *voice onset time* (VOT), the time lag between the consonantal release and the onset of voicing. Stops with shorter VOTs are perceived as voiced and stops with longer VOTs are perceived as voiceless, with a sharp perceptual boundary around 30 ms VOT (Liberman, Harris, Kinney, & Lane, 1961). Based on evidence for the same perceptual boundary in young infants (Eimas et al., 1971) and nonhuman mammals (Kuhl & Miller, 1975; Kuhl & Padden, 1982), it has been commonly assumed that the linguistic categories of voicing are built on top of pre-existing perceptual discontinuities.

A specific version of this hypothesis has been advanced by Steinschneider, Sinex and their colleagues (Sinex & McDonald, 1988; Sinex & McDonald, 1989; Sinex, 1993; Steinschneider, Schroeder, Arezzo, & Vaughan, 1994; Steinschneider, Schroeder, Arezzo, & Vaughan, 1996; Steinschneider, Volkov, Noh, Garrell, & Howard, 1999). According to this hypothesis, category boundaries for voicing are a natural consequence of the way in which specific groups of auditory neurons respond to events occurring in rapid succession. It is well-known that sensory neurons are often ‘tuned’ to specific stimulus features, to which they respond selectively. However, the act of responding to a preferred stimulus typically has the effect of *desensitizing* the neurons for a short period (the *inactivation period*), during which time the neurons cannot respond to new instances of the preferred stimulus. Steinschneider and Sinex suggest that the category boundary for VOT corresponds to the inactivation period for certain groups of auditory neurons which respond to both the noise-burst and the voicing onset. At VOTs above around 30 ms, the relevant cells can respond independently to the release and to the voicing onset, yielding a ‘double on’ response pattern. At shorter VOT values, on the other hand, the same cells respond only to the consonantal release, and cannot respond independently to the onset of voicing, yielding a ‘single on’ response pattern (see Fig. 4). Support for this idea comes from direct recordings in primary auditory cortex of macaques (Steinschneider et al., 1994, Steinschneider et al., 1996), and humans (Steinschneider et al., 1999), and the auditory nerve of chinchillas (Sinex & McDonald, 1988, Sinex & McDonald, 1989).

The ‘double on’ account of voicing suggests that voicing categories are based on low-level properties of the auditory system, and we might expect a mechanism based on such a basic property of neurons to show limited plasticity, as in the structure-adding approach. Some electrophysiological evidence supports this hypothesis. A striking prediction of the double-on hypothesis is that information about VOT is lost *asymmetrically*: responses to

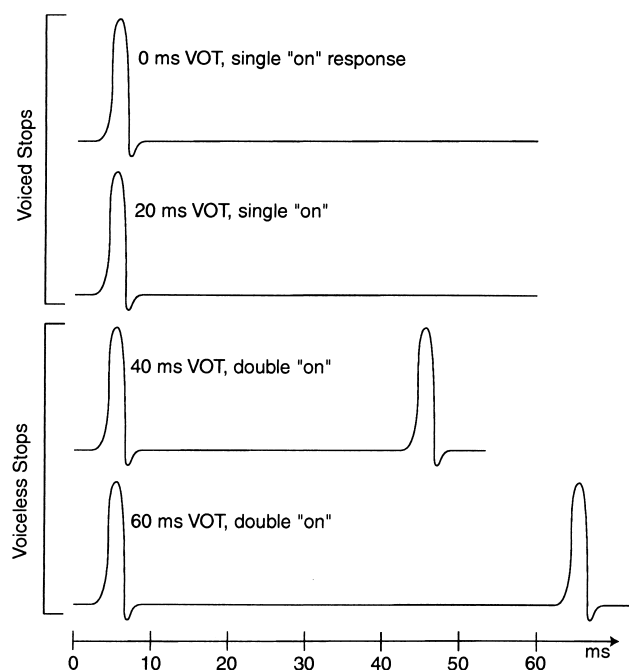


Fig. 4. Schematic of encoding of single versus double “on” encoding of voicing in stops.

different VOTs in the *voiced* range should be identical, and hence indistinguishable from one another, but responses to different VOTs in the *voiceless* range should provide a faithful encoding of VOT values.

Sharma and Dorman (1999) report that in responses to a 9-step VOT continuum shorter VOTs (<40 ms) elicit a single N100 response component (which varies with VOT in latency), whereas longer VOTs (>40 ms) elicit two distinct N100 components, only the second of which varies with VOT in latency. Phillips et al. (1995) report a very similar pattern of MEG responses elicited by a 9-step VOT continuum, except that M100 response latencies within the voiced range of the continuum do not vary with VOT (exactly as predicted by the Steinschneider/Sinex model).

Sharma and Dorman (1999) also compared a between-category VOT contrast (30ms vs. 50ms) to a within-category contrast (40 ms vs. 60 ms) in a mismatch paradigm. The between-category contrast elicited a significantly larger mismatch response than the within-category contrast, although the mismatch response to the within-category contrast did reach significance. This effect of phonetic category membership suggests that phonetic category representations are available to low-level cortical systems.

To date there have been no electrophysiological studies of voicing contrasts in subjects for whom they are non-native, paralleling similar studies in infants (Streeter, 1976).

These electrophysiological studies show that at least part of the representation of voicing categories may be achieved by low-level cortical mechanisms. However, there are a number of phonetic properties of voicing categories which do not follow from this hypothesis, and

so the representation of voicing probably cannot be entirely reduced to inactivation periods. First, the *asymmetric* loss of information about VOT that this hypothesis predicts does not correspond to existing behavioral results, which show within-category discrimination of both voiceless and voiced stops (Carney, Widin, & Viemeister, 1977; Samuel, 1982; Volaitis & Miller, 1992). Second, VOT boundary values vary across place of articulation (Summerfield, 1982), and across languages (Lisker & Abramson, 1964), facts which are not explained by this hypothesis. Finally, even if the contrast between one and two ‘on’ responses plays a crucial role in an early stage of voicing category identification, it is rather unlikely as a coding schema for the *phonological* representation of voicing, where we find processes in some languages which change voiced consonants into voiceless consonants in syllable-final position.

3.3. *Place of articulation*

Place of articulation distinguishes *labial* consonants /b, p, m/ from *alveolar* consonants /d, t, n/ and *velar* consonants /g, k, ɣ/. Place of articulation presents a rather different picture than vowels or voicing. The neural encoding hypotheses are less clear, but the evidence for multiple phonetic levels of representation is better.

Place of articulation has provided less specific coding hypotheses because the acoustic cues to place of articulation are less well understood. Formant transitions between consonants and neighboring vowels provide important cues to place of articulation, but the formant transition patterns associated with a given place of articulation differ according to the vowel context. Although there have been numerous proposals about how this variability may be overcome (e.g., Blumstein & Stevens, 1979, 1980; Kewley-Port, 1983; Sussman, Fruchter, Hilbert, & Sirosh, 1998), the nature of the solution is less clear than it is in the case of vowels or voicing. Also, whereas ‘vowel-space’ and ‘VOT-space’ are, at least in principle, continuous representational spaces which languages could choose to divide up in a variety of ways, ‘place-space’ is not. The physiology of the vocal tract makes a relatively limited set of place categories available, and cross-language variation is more constrained than variation in vowel categories. For example, articulatory constraints allow a number of different vowel category boundaries between /o/ and /e/, but impose a fixed articulatory boundary between stops that are articulated with the lips and stops that are articulated at the teeth or alveolar ridge.

Due to the articulatory constraints on place of articulation categories, it is feasible to describe an underlying universal set of place categories, which could be overlaid during development by a language specific representation of place categories, as expected under the structure-adding approach. Such a universal category inventory may not even be feasible for voicing or vowels, given the existence of overlapping categories across languages.

In fact, the best evidence for the structure-adding view comes from place of articulation categories, and involves demonstrations of *simultaneous* universal and native-language category representations. The classic finding that younger infants discriminate non-native place contrasts which older infants cannot discriminate (Werker & Tees, 1984) is compatible with both structure-changing and structure-adding views of phonetic representation, because this simply shows that important change occurs around 8–10 months of age. However,

studies by Werker and Logan (1985) show that many adult English speakers, who fail to discriminate an alveolar/retroflex place contrast at long interstimulus intervals (ISI), show categorical discrimination of these place categories at brief (<500 ms) ISIs. This suggests that both universal and language-specific representations are copresent, but that the lower-level representation is weaker, or decays faster.

Whereas mismatch studies of vowels and voicing have demonstrated clear effects of native-language phonetic categories, such evidence has been less forthcoming in studies of place contrasts. A number of studies have asked whether easily discriminable across-category contrasts yield larger amplitude mismatch responses than hard-to-discriminate within-category contrasts. A number of studies have answered this question in the negative (Sams, Aulanko, Aaltonen, & Näätänen, 1990; Sharma, Kraus, McGee, Carrell, & Nicol, 1993; Maiste, Wiens, Hunt, Scherg, & Picton, 1995), finding no special status for native language category contrasts. These results do not themselves support the structure-adding view, since they could be explained by purely acoustic representations.

More relevant for the structure-adding hypothesis is a series of recent ERP studies which have examined sensitivity to native and non-native contrasts along Werker & Lalonde's (Werker & Lalonde, 1988) synthetic labial-dental-retroflex ([ba-da-Da]) continuum (Dehaene-Lambertz, 1997; Rivera-Gaxiola, Csibra, Johnson, & Karmiloff-Smith, 2000a; Rivera-Gaxiola, Johnson, Csibra, & Karmiloff-Smith, 2000b; Tsui, Shi, Werker, & Stapells, 2000). The labial-dental contrast was native to the English and French-speaking participants, but the dental-retroflex contrast was a non-native contrast. These studies have examined whether similar mismatch responses are elicited by equally-sized acoustic contrasts that span (i) native phonetic boundaries ([ba-da]), (ii) non-native phonetic boundaries ([da-Da]), and (iii) no phonetic boundaries (e.g., [ba₁-ba₅]). Structure-changing approaches, which predict that non-native boundaries are genuinely lost, would be best supported by the finding that contrast (i) yields a much larger mismatch response than either (ii) or (iii). Structure-adding approaches, which predict that infants initial representations are preserved, would be best supported by similar mismatch responses to contrasts (i) and (ii), and weaker or absent mismatch responses to (iii).

The results of the different studies are not entirely in agreement, for reasons which may in part reflect the specifics of the technique used to elicit the MMN. The results of Dehaene-Lambertz (1997) are most clearly in accord with the predictions of the structure-changing approach, since a strong MMN was elicited only by the native phonetic contrast.⁴ Dehaene-Lambertz elicited the MMN by presenting short groups of sounds in XXXY sequences, in an active task requiring a same/different judgment to the final sound of each group. The active task and the interspersing of native and non-native contrasts may explain the predominance of native contrasts in the results of this study. A pair of studies by Rivera-Gaxiola and colleagues (2000a, b) are more compatible with the structure-adding approach. Although all three contrast types elicited some kind of response to the deviant sound, the mismatch responses to native and non-native contrasts were more similar to one another than to the response to the nonphonetic contrast. These studies used a more traditional passive mismatch paradigm. Finally, a recent study by Tsui and colleagues (2000) reports identical MMNs elicited by native and non-native contrasts, consistent with the

Table 1

Results of tests of the phonetic sensitivity of the auditory cortex mismatch response

Phonetic feature	Phonetic sensitivity	Reference
vowel backness	yes _{native}	Näätänen et al., 1997
	yes _{native}	Winkler et al., 1999
voicing	yes	Sharma & Dorman 1999
place of articulation	no	Sams et al., 1990
(b-d-g continuum)	no	Sharma et al., 1993
	no	Maiste et al., 1995
place of articulation	yes _{native}	Dehaene-Lambertz 1997
(b-d-D continuum)	yes _{non-native}	Rivera-Gaxiola et al., 2000a,b
	possible _{non-native}	Tsui et al., 2000

predictions of the structure-adding approach. However, since this study did not include a nonphonetic contrast, the results are equally compatible with a purely acoustic account. This study also used a passive oddball paradigm, but one in which the deviant sound was always the final sound in a block of 10 sounds.

In summary, results to-date on place-of-articulation contrasts provide some evidence for preservation of non-native contrasts, and hence provide some support for structure-adding approaches to phonetic representations. However, in light of apparent inconsistencies across studies, it is clear that more work needs to be done in this area.

3.4. Summary of results on different features

Electrophysiological studies of phonetic representations yield differing results across different phonetic categories. This picture is consistent with the diversity of acoustic cues to phonetic categories and the diversity of coding hypotheses. Table 1 summarizes the results of mismatch studies of a number of different phonetic contrasts. This table alone attests to the heterogeneity of phonetic representations.

To the extent that electrophysiological results support either the structure-changing or structure-adding approaches to phonetic representations, results do not *consistently* favor one approach over the other across all phonetic contrasts. Rather, it may be that the two hypotheses are both correct, for different areas of phonetics. We may tentatively suggest that the structure-adding hypothesis receives more support, and is more feasible, in areas where there are strong physiological constraints on the range of possible category boundaries across languages (e.g., place of articulation). On the other hand, the best support for the structure-changing hypothesis comes from areas of phonetics where there are fewer physiological constraints on the range of category boundaries, and the representational space is more plausibly continuous (e.g., vowels). Also, it should be noted that the clearest evidence for *structure* in infants' discrimination abilities comes from consonants, whereas the clearest evidence for 'perceptual magnets' in adults comes from vowels and liquids. Hence, the empirical evidence from behavioral studies already shows a separation between the two phonetic domains which best support the two hypotheses.

4. Phonological representations

Phonetics provides a number of detailed proposals about how representations of speech might be encoded in the brain, and a growing body of electrophysiological results provide evidence that bears upon these hypotheses. The diversity of phonetic coding hypotheses reflects the diversity of the acoustics of phonetic categories. In *phonology*, on the other hand, existing evidence suggests a need for representations which are fundamentally different from those found in phonetics, representations which are more uniform across the range of different phonological categories. In phonology, the coding hypotheses have proven harder to test, and existing neurolinguistic evidence is more limited. Although there is no complete agreement on the definition of phonological categories, I will review a number of properties of phonological categories, and comment on relevant neurolinguistic evidence. In general, the more abstract the property, the less neurolinguistic evidence is available.

A. Irrelevance of within-category variation. Phonological categories are discrete, symbolic units, that are used to store words in long-term memory. The within-category distinctions that are relevant to phonetic categories are *irrelevant* to phonological categories. A speaker's memory of the word CAT indicates that the word begins with a /k/, but does not encode details about whether it is a prototypical /k/, or whether it is coarticulated with the following vowel.

Some studies in the mismatch paradigm have used the irrelevance of within-category differences as a probe for phonological category representations. The logic of these studies is as follows. The mismatch response is elicited when infrequent *deviant* sounds interrupt a sequence of identical sounds, the *standards*, that is, when there is a *many-to-one* ratio among stimuli. By introducing within-category acoustic variation into a sequence of speech sounds, it is possible to restrict the many-to-one ratio required to elicit the mismatch response to a single level of representation. Phillips et al. (2000a) adopted this strategy to probe for phonological representations of /d/ and /t/, using sequences of syllables like the one shown in Fig. 5. There is no many-to-one ratio of standards to deviants at the acoustic and phonetic levels, due to the within-category variation in VOT. It is only at the phonological level that there is a many-to-one ratio of standards to deviants. A classic auditory cortex MMF was elicited, indicating that phonological representations of the categories /d/ and /t/ are available to auditory cortex. In a later study, Phillips, Pellathy, & Marantz (2000b) extended this approach to the level of the phonological feature [\pm voice], by contrasting multiple tokens of /pæ, tæ, kæ/ with multiple tokens of /bæ, dæ, gæ/. These mismatch studies depart from the standard practice of using a single acoustic stimulus as the standard sound; instead, standards and deviants are chosen randomly from a group of acoustically different sounds (see also Aulanko, Hari, Lounasmaa, Näätänen, & Sams, 1993).

A study by Dehaene-Lambertz and colleagues (Dehaene-Lambertz, Dupoux, & Gout, 2000) used a cross-language design to probe Japanese speakers' representations of syllable structure. Japanese syllables conform to a strict (C)V(N) structure, whereas French (like English) allows a much wider range of syllable-types than Japanese (e.g., *sprint*, CCCVCC). One consequence of this is that English loan words in Japanese are altered, by vowel insertion, so that they conform to the syllable structure of Japanese: for example, *brother* → *burazaa*; *volcano* → *borukano*. Another consequence is that Japanese speakers may hear

C. Neutralization. Phonological neutralization processes are processes which cause phonologically distinct categories to be realized as phonetically identical. A well-known neutralization process is the word-final devoicing of consonants found in many languages. For example, the German words *bunt* ('colorful') and *Bund* ('bundle') are both pronounced with word-final [t] in isolation, but show contrasting [t] and [d] pronunciations when suffixed with the vowel [ə]. Similarly, the American English flapping rule neutralizes the contrast between /t/ and /d/ when these consonants follow a stressed vowel and precede an unstressed vowel, for example, *bidder* ([bɪrə]) versus *bitter* ([bɪrə]).⁶ Since neutralization processes cause pairs of different phonological categories to be pronounced identically, no simple transform of acoustic space will allow successful recovery of neutralized phonological categories. Lexical information is necessary in order to achieve this.⁷

D. Dissimilation. Dissimilation processes in phonology create distinctions at the acoustic level when there are none at the phonological level. Consider, the adjectives *circular*, *columnar*, *lunar*, and *floral*, *annual*, *mental*, which show two different adjectival endings *-al* and *-ar*. The *-ar* ending is a vestige of a dissimilation process in Latin which blocked the occurrence of sequences of identical liquids ([.l.l.] or [.r.r.]), even when the liquids were not string-adjacent in a word (cf. Posner, 1961). Another example of dissimilation can be found in Bantu languages of Africa such as Kikuyu (Armstrong, 1967; Lombardi, 1995). The process known as Dahl's Law blocks sequences of voiceless consonants, such that the prefix *ko-* is pronounced as voiceless *ko-* in *ko-niina* ('to finish'), but as the voiced fricative [yo-*koora* ('to root out'). Thus, the sounds [k] and [ɣ] in these prefixes contrast acoustically and phonetically, but not phonologically.

E. Similarity versus Identity. One of the most compelling pieces of evidence for the abstractness of phonological representations comes from examples of dissimilation processes which distinguish *similarity* from *identity*. In a number of languages we find constraints which force neighboring segments (typically vowels in adjacent syllables) to have contrasting values of some feature, *unless* the two vowels are completely identical. For example, in Ngbaka (Wescott, 1965; Chomsky & Halle, 1968) a two-syllable word with the vowel sequence /i . . . u/ is impossible, because both vowels are high. However, the sequence /u . . . u/ is allowed in this language, despite the fact that both vowels are high, because the two vowels are identical. This apparently disjunctive constraint can be straightforwardly accounted for by representing a sequence of two identical vowels as a single discontinuous category (cf. Phillips, 1994 for further examples of this kind). Clearly, this requires representations of sound sequences which are far more abstract than simple transforms of the acoustic sound sequence.

With respect to neutralization processes, dissimilation processes and other aspects of basic segmental phonology, there is almost no relevant work in electrophysiology or other forms of neuro-imaging. This relative lack of attention to phonology (as opposed to phonetics) is unfortunate, since the decoding of the effects of phonological processes is clearly a critical stage in the mapping from acoustics onto word forms.

More generally, phonological representations and processes present a rather different kind of puzzle for neuroscience than does the phonetic aspect of speech perception. Phonological categories are symbolic units, which combine in more or less the same manner and undergo more or less the same kinds of processes (e.g., deletion, neutralization, assimilation, dissim-

ilation. . .). It is plausible that the neurophysiological coding of phonological representations is simpler and more uniform than the coding of phonetic categories, but we have little idea of what the coding of such symbolic categories might look like.

5. Conclusions

Recent years have seen a great increase in findings about how speech is encoded in the human brain, and electrophysiological techniques have played a central role in this. However, the extent of our understanding is still very limited. The greatest progress has been made in areas where there are straightforward hypotheses about acoustics-to-phonetics mappings. Nonlinearities in the acoustics-to-phonetics mapping can be correlated with nonlinearities in ERP or MEG response components. When the predicted nonlinearities are observed, they are taken as evidence for the involvement of phonetic representations in the generation of the electrophysiological response. This strategy has proven to be quite fruitful, and has made it possible to begin to use electrophysiological evidence to address basic questions about the nature and content of phonetic representations.

Models of the phonetics-to-phonology mapping are rather less well developed, and phonological representations are a good deal more abstract than phonetic representations. Therefore, it is not surprising that progress towards understanding how phonological categories are represented in the brain has been more limited. Some recent findings about the *failure* to distinguish members of the same phonological category implicate the contribution of phonological representations to mid-latency evoked responses, but provide no clues about how these representations are encoded. Since phonological representations are discrete, symbolic representations which are probably uniform across different phonological categories and features, we will need some highly explicit hypotheses about what these representations *might look like* before we can record people's brains in search of confirmation.

It is reasonable to ask whether electrophysiological findings to-date have provided any information that could not be learned from simpler (and cheaper) behavioral methods. It is true, and will likely continue to be true, that most representational hypotheses have been derived from theoretical and behavioral research, and cognitive neuroscience is faced with the subsidiary task of pinpointing these representations in space and time in the brain. One basic contribution of electrophysiological studies to date has been the demonstration of the phonetic sophistication of human auditory cortex. However, the spatial and temporal richness of electrophysiological data are beginning to add an extra dimension, and will continue to do so, provided that sufficiently detailed spatial and temporal hypotheses are formulated.

Notes

1. A number of studies of nonspeech sounds have pointed to sources for the mismatch response outside of auditory cortex (e.g., Giard, Perrin, Pernier, & Bouchet; 1990; King, McGee, Rubel, Nicol, & Kraus, 1995), and some studies of speech sound processing have observed different scalp topographies for different speech contrasts

(e.g., Dehaene-Lambertz & Baillet, 1998). It is therefore likely that the mismatch response is in fact a composite of activity in a number of areas. However, the implications of such spatial differences for speech perception are not yet well understood.

2. Although this is one possible implementation of Kuhl's theoretical claims, it is by no means a necessary consequence of the theory. Illustrations of the theory often emphasize the disappearance of category boundaries in a single 'perceptual space' (e.g., Kuhl, 2000: 105), but both the theory and the evidence for the theory leave it open whether there is a unique phonetic representational space underlying perceptual space, or whether perceptual space is the result of a series of independent phonetic representations or 'filters'.
3. Although Werker regularly characterizes developmental change as *functional reorganization*, this may be misleadingly understood to imply a structure-changing approach. In earlier work, Werker proposed that native-language category representations entirely reflect a *phonological* level of representation (Werker & Logan, 1985). However, this cannot be the case (cf. Werker & Pegg, 1992). Evidence for a separate native-language *phonetic* level comes from: (i) the existence of native-language phonetic prototypes, which encode within-category details that are irrelevant at the phonological level; (ii) the distinction recently shown in 14-month old infants between representations available for discrimination and for object-labeling (Stager & Werker, 1997).
4. The non-native and nonphonetic contrasts did elicit significant ERP responses in this study, but they were either later (non-native contrast) or less robust (nonphonetic contrast) than the response elicited by the native contrast.
5. A later response to the mismatching vowel was observed in Japanese speakers, in the 300–600 ms latency interval.
6. Even if there are subtle acoustic differences between the pronunciation of flapped /d/ and flapped /t/, such distinctions are not effective cues for listeners (Zue & Laferriere, 1979).
7. It has recently been suggested that devoicing involves *reduction* rather than *loss* of voicing contrasts (e.g., Port & O'Dell, 1985; Slowiaczek & Dinnsen, 1985; Port, 1996). However, Kim & Jongman (1996) provide unambiguous evidence for neutralization, based on the syllable-final *manner*-neutralization process found in Korean, where plain /t/, aspirated /t^h/, tense /t'/, affricate /c/ and fricative /s/ are all realized as plain (but often released) [t] in word-final position.

Acknowledgments

I thank three anonymous reviewers, Cindy Brown, Tom Pellathy and especially Bill Idsardi for valuable feedback on the material in this paper. Preparation of this paper was supported in part by grants from the McDonnell-Pew Program in Cognitive Neuroscience, Oak Ridge Associated Universities, and the National Science Foundation (BCS-0196004).

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., Tuomainen J., Laine M., & Niemi P. (1993). Cortical differences in tonal versus vowel processing as revealed by an ERP component called mismatch negativity (MMN). *Brain and Language*, *44*, 139–152.
- Alho K. (1995). Cerebral generators of Mismatch Negativity (MMN) and its magnetic counterpart (MMNm) elicited by sound changes. *Ear & Hearing*, *16*, 38–51.
- Armstrong L. (1967). *The phonetic and tonal structure of Kikuyu*. London: Dawsons.
- Aulanko R., Hari R., Lounasmaa O. V., Näätänen R., & Sams M. (1993). Phonetic invariance in the human auditory cortex. *Neuroreport*, *4*, 1356–1358.
- Blumstein S. E., & Stevens K. N. (1979). Acoustic invariance in speech production: evidence from measurements of the spectral characteristics of stop consonants. *Journal of the Acoustical Society of America*, *66*, 1001–1017.
- Blumstein S. E., & Stevens K. N. (1980). Perceptual invariance and onset spectra for stop consonants in different vowel environments. *Journal of the Acoustical Society of America*, *67*, 648–662.
- 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.
- Chomsky N., & Halle M. (1968). *The sound patterns of English*. New York: Harper & Row.
- Dehaene-Lambertz G. (1997). Electrophysiological correlates of categorical phoneme perception in adults. *Neuroreport*, *8*, 919–24.
- Dehaene-Lambertz G., & Baillet S. (1998). A phonological representation in the infant brain. *Neuroreport*, *9*, 1885–1888.
- Dehaene-Lambertz G., Dupoux E., & Gout A. (2000). Electrophysiological correlates of phonological processing: a cross-linguistic study. *Journal of Cognitive Neuroscience*, *12*, 635–647.
- Diesch E., Eulitz C., Hampson S., & Ross B. (1996). The neurotopography of vowels as mirrored by evoked magnetic field measurements. *Brain and Language*, *53*, 143–168.
- Dupoux E., Kakehi K., Hirose Y., Pallier C., & Mehler J. (1999). Epenthetic vowels in Japanese: a perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 1568–1578.
- Eimas P. (1974). Auditory and linguistic processing of cues for place of articulation by infants. *Perception & Psychophysics*, *16*, 513–521.
- Eimas P. (1975). Auditory and phonetic coding of the cues for speech: Discrimination of the [r-l] distinction by young infants. *Perception & Psychophysics*, *18*, 341–347.
- Eimas P., & Miller J. (1980). Contextual effects in infant speech perception. *Science*, *209*, 1140–1141.
- Eimas P., Siqueland E., Jusczyk P., & Vigorito J. (1971). Speech perception in infants. *Science*, *171*, 303–306.
- 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.
- Govindarajan, K., Phillips, C., Poeppel, D., Roberts, T., & Marantz, A. (1998). Latency of MEG M100 response indexes first formant frequency. Proceedings of the 1998 ASA Conference.
- 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.
- Jusczyk P. (1997). *The discovery of spoken language*. Cambridge, MA: MIT Press.
- Kewley-Port D. (1983). Time-varying features as correlates of place of articulation in stop consonants. *Journal of the Acoustical Society of America*, *73*, 322–335.
- Kim H., & Jongman A. (1996). Acoustic and perceptual evidence for complete neutralization of manner of articulation in Korean. *Journal of Phonetics*, *24*, 295–312.
- King C., McGee T., Rubel E. W., Nicol T., & Kraus N. (1995). Acoustic features and acoustic change are represented by different central pathways. *Hearing Research*, *85*, 45–52.
- Kluender K. R., Diehl R. L., & Killeen P. R. (1987). Japanese quail can learn phonetic categories. *Science*, *237*, 1195–1197.

- Kohonen T., & Hari R. (1999). Where the abstract feature maps of the brain might come from. *Trends in Neuroscience*, 22, 135–139.
- Kuhl P. K. (1991). Human adults and human infants show a “perceptual magnet effect” for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, 50, 93–107.
- Kuhl P. K. (1991). Perception cognition and the ontogenetic and phylogenetic emergence of human speech. In S. E. Brauth W. S. Hall R. J. Dooling (Eds.) *Plasticity of development* (pp. 73–106). Cambridge, MA: MIT Press.
- Kuhl P. K. (1994). Learning and representation in speech and language. *Current Opinion in Neurobiology*, 4, 812–822.
- Kuhl P. K. (2000). Language, mind, and brain: Experience alters perception. In M. Gazzaniga (Ed.), *The cognitive neurosciences 2nd ed.* (pp. 99–115). Cambridge, MA: MIT Press.
- Kuhl P. K., & Miller J. D. (1975). Speech perception by the chinchilla: voiced-voiceless distinction in alveolar plosive consonants. *Science*, 190, 69–72.
- Kuhl P. K., & Padden D. M. (1982). Enhanced discriminability at the phonetic boundaries for the voicing feature in macaques. *Perception & Psychophysics*, 32, 542–550.
- 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.
- Kuhl P. K., Williams K. A., Lacerda F., Stevens K. A., & Lindblom B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, 255, 606–608.
- Lang H. A., Nyrke T., Ek M., Aaltonen O., Raimo I., & Näätänen R. (1990). Pitch discrimination performance and auditory event-related potentials. In C. Brunia et al (Eds.) *Psychophysiological brain research, Vol. 1* (pp. 294–298). Tilburg: Tilburg University Press.
- Lieberman 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.
- Lisker L., & Abramson A. S. (1964). A cross-language study of voicing in initial stops: acoustical measurements. *Word*, 20, 384–422.
- Lombardi L. (1995). Dahl’s Law and privative [voice]. *Linguistic Inquiry*, 26, 365–372.
- Maiste A., Wiens A., Hunt M., Scherg M., & Picton T. (1995). Event-related potentials and the categorical perception of speech sounds. *Ear and Hearing*, 16, 67–89.
- Miller J. (1994). On the internal structure of phonetic categories: a progress report. *Cognition*, 50, 271–285.
- Näätänen R., Gaillard A. W. K., & Mäntysalo S. (1978). Early selective attention effect on evoked potential reinterpreted. *Acta Psychologica*, 42, 313–329.
- Näätänen R., Lehtoskoski 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.
- Näätänen R., & Picton T. W. (1987). The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. *Psychophysiology*, 24, 375–425.
- Näätänen R., & Winkler I. (1999). The concept of auditory stimulus representation in cognitive neuroscience. *Psychological Bulletin*, 125, 826–859.
- Pantev C., Hoke M., Lehnertz K., Lütkenhöner B., Anogianakis G., & Wittkowski W. (1988). Tonotopic organization of the human auditory cortex revealed by transient auditory evoked magnetic fields. *Electroencephalography and Clinical Neurophysiology*, 69, 160–170.
- Phillips C. (1994). Are feature hierarchies autosegmental hierarchies? *MIT Working Papers in Linguistics*, 21, 173–226.
- Phillips C., Marantz A., McGinnis M., Pesetsky D., Wexler K., Yellin A., Poeppel D., Roberts T., & Rowley H. (1995). Brain mechanisms of speech perception: a preliminary report. *MIT Working Papers in Linguistics*, 26, 125–163.
- Phillips C., Pellathy T., Marantz A., Yellin E., Wexler K., Poeppel D., McGinnis M., & Roberts T. (2000). Auditory cortex accesses phonological categories: an MEG mismatch study. *Journal of Cognitive Neuroscience*, 12.

- Phillips, C., Pellathy, T., & Marantz, A. (2000b). Phonological feature representations in auditory cortex. submitted manuscript.
- Poeppel D., Yellin E., Phillips C., Roberts T., Rowley H., Wexler K., & Marantz A. (1996). Task-induced asymmetry of the auditory evoked M100 neuromagnetic field elicited by speech sounds. *Cognitive Brain Research*, 4, 231–242.
- Poeppel D., Phillips C., Yellin E., Rowley H., Roberts T., & Marantz A. (1997). Processing of vowels in supratemporal auditory cortex. *Neuroscience Letters*, 221, 145–148.
- Port R. F. (1996). The discreteness of phonetic elements and formal linguistics: response to A. Manaster Ramer. *Journal of Phonetics*, 24, 491–511.
- Port R. F., & O'Dell M. (1985). Neutralization of syllable-final voicing in German. *Journal of Phonetics*, 13, 455–471.
- Posner R. R. (1961). *Consonantal dissimilation in the romance languages*. Oxford: Blackwell. (Publications of the Philological Society, 19).
- Rivera-Gaxiola M., Csibra G., Johnson M., & Karmiloff-Smith A. (2000). Electrophysiological correlates of cross-linguistic speech perception in native English speakers. *Behavioural Brain Research*, 111, 13–23.
- Rivera-Gaxiola M., Johnson M., Csibra G., & Karmiloff-Smith A. (2000). Electrophysiological correlates of category goodness. *Behavioral Brain Research*, 112, 1–11.
- Roberts T., & Poeppel D. (1996). Latency of auditory evoked M100 as a function of tone frequency. *Neuroreport*, 7, 1138–1140.
- Romani G. L., Williamson S. J., & Kaufman L. (1982). Tonotopic organization of the human auditory cortex. *Science*, 216, 1339–1340.
- Sams M., Paavilainen P., Alho K., & Näätänen R. (1985). Auditory frequency discrimination and event-related potentials. *Electroencephalography & Clinical Neurophysiology*, 62, 437–448.
- 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., 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.
- 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.
- 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.
- Sharma A., & Dorman M. F. (1998). Exploration of the perceptual magnet effect using the mismatch negativity auditory evoked potential. *Journal of the Acoustical Society of America*, 104, 511–517.
- 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.
- 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. (1993). Simulation of neural responses that underlie speech discrimination. In F. Eeckman (Ed.), *Neural systems: Analysis and modeling* (pp. 307–313). Norwell, MA: Kluwer.
- Slowiaczek L. M., & Dinnsen D. A. (1985). On the neutralizing status of Polish word-final devoicing. *Journal of Phonetics*, 13, 325–341.
- Stager C. L., & Werker J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature*, 388, 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., Arezzo J., & Vaughan H. (1996). Physiological correlates of the voice onset time boundary in primary auditory cortex of the awake monkey—temporal response patterns. *Brain & Language*, 48, 326–340.
- 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.
- Streeter L. (1976). Language perception of 2-month old infants shows effects of both innate mechanisms and experience. *Nature*, 259, 39–41.
- Summerfield Q. (1982). Differences between spectral dependences in auditory and phonetic temporal processing—relevance to the perception of voicing in initial stops. *Journal of the Acoustical Society of America*, 72, 51–61.
- Sussman H. M., Fruchter D., Hilbert J., & Sirosh J. (1998). Linear correlates in the speech signal: the orderly output constraint. *Behavioral and Brain Sciences*, 21, 241–299.
- 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.
- Tsui, V., Shi, R., Werker, J., & Stapells, D. (2000). MMN measures of categorical perception of native and non-native consonant contrasts. Paper presented at the 2nd International Congress on Mismatch Negativity and its Clinical Applications, Barcelona, Spain.
- Tsushima, T., Takizawa, O., Sasaki, M., Shiraki, S., Nishi, K., Kohno, M., Menyuk, P., & Best, C. (1994). Discrimination of English /r-l/ and /w-y/ by Japanese infants 6–12 months: Language-specific developmental changes in speech perception abilities. Paper presented at the International Conference on Spoken Language. Yokohama, Japan
- Volaitis L. E., & Miller J. L. (1992). Phonetic prototypes: influence 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. (1994). Cross-language speech perception: developmental change does not involve loss. In H. C. Nusbaum J. Goodman (Eds.) *The development of speech perception: The transition from speech sounds to spoken words* (pp. 93–120). Cambridge, MA: MIT Press.
- Werker J. F. (1995). Exploring developmental changes in cross-language speech perception. In L. Gleitman M. Liberman (Eds.) *An invitation to cognitive science, part 1: Language* (pp. 87–106). Cambridge, MA: MIT Press.
- 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. (1985). Cross-language evidence for three factors in speech perception. *Perception and Psychophysics*, 37, 35–44.
- Werker J. F., & Pegg J. E. (1992). Infant speech perception and phonological acquisition. In C. Ferguson L. Menn C. Stoel-Gammon (Eds.) *Phonological development: Models, research, implications* (pp. 285–311). Timonium, MD: York Press.
- Werker J. F., & Tees R. C. (1984). Cross-language speech perception: evidence for perceptual reorganization during the first year of life. *Infant Behavioral Development*, 7, 49–63.
- Werker J. F., & Tees R. C. (1999). Influences on infant speech processing: toward a new synthesis. *Annual Review of Psychology*, 50, 509–535.
- Wescott R. W. (1965). Review of J.M.C. Thomas: Phonologie, morphologie, syntaxe. *Language*, 41, 346–347.
- Winkler I., Lehtokoski A., Alku P., Vainio M., Czigler I., Csepe V., Aaltonen O., Raimo I., Alho K., Lang H., Iivonen A., & Näätänen R. (1999). Pre-attentive detection of vowel contrasts utilizes both phonetic and auditory memory representations. *Cognitive Brain Research*, 7, 357–369.
- Zue V., & Laferriere M. (1979). Acoustic study of medial /t,d/ in American English. *Journal of the Acoustical Society of America*, 66, 1039–1050.