Levels of Representation in the Electrophysiology of Speech Perception

Colin Phillips, University of Delaware

Abstract

Mapping from acoustic signals to lexical representations is a complex process mediated by a number of different levels of representation. This paper reviews certain properties of the *phonetic* and *phonological* levels of representation and hypotheses about how category structure is represented at each of these levels, and evaluates these in light of relevant electrophysiological studies of phonetics and phonology. The phonetic level is characterized by diverse (and competing) hypotheses about how categories could be represented in the brain, and there is a growing body of electrophysiological work which can be brought to bear on these hypotheses. Existing electrophysiological evidence is consistent with behavioral evidence for the heterogeneity of phonetic representations. Furthermore, some evidence supports the existence of multiple phonetic levels of representation. Meanwhile, rather less is known about how representations at the phonological level could be coded in the brain. These representations are more abstract than phonetic representations, and are probably more homogeneous than phonetic representations.

1. Introduction[1]

At a minimum, the process of speech perception involves a mapping from continuous acoustic waveforms onto the discrete phonological units used to store words in the mental lexicon. To take a simple example, when we hear the word *cat*, a complex and continuous pattern of vibration hits the eardrum, but this gives rise to a phonological percept which has just three clearly distinct pieces: /k/, /æ/ and /t/. A great deal of evidence from phonetics and phonology, language acquisition and psychophysical research 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, and should thus be seen as a multi-step process. 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

---

[1] I am grateful to Cindy Brown, Tom Pellathy and especially Bill Idsardi for helpful discussion of the material in this paper. Preparation of this paper was supported in part by grants from the McDonnell-Pew Program in Cognitive Neuroscience and Oak Ridge Associated Universities. Address correspondence to Colin Phillips, Department of Linguistics, University of Delaware, 46 East Delaware Avenue, Newark, DE 19716, email: colin@udel.edu.

magnetoencephalography (MEG), and discusses the relationship between this work and specific proposals about how the brain encodes information about speech. The article emphasizes the connection between neurolinguistic findings and hypotheses derived from behavioral and theoretical research on phonetics and phonology.

EEG and MEG provide non-invasive 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 provide only modest localization information compared to hemodynamic measures such as PET or fMRI. Furthermore, they require synchronous activity of many thousands of neurons in order for a response to be detected at the scalp. In practice, however, most existing hypotheses about the localization of speech representations in the brain are either coarse-grained enough to be tested with MEG or EEG, or too fine-grained to be tested with any of the available non-invasive methods.

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 generated in auditory cortex between 50 ms and 250 ms after a sound is presented. Evoked responses are temporary increases in brain activity which are time-locked to stimulus presentation, and which therefore survive averaging of a large number of presentations (typically 100 or more) of each stimulus. Human auditory cortex is situated on the superior plane of the temporal lobe, i.e. on the lower side of the Sylvian fissure. To date, electrophysiological data have provided a rich set of findings about representations of speech in auditory cortex, but have revealed relatively little about the role of other brain areas in speech perception, although results using other techniques clearly implicate additional areas in speech perception. Human auditory cortex consists of a number of sub-areas, but most electrophysiological findings about speech perception do not reliably implicate specific sub-areas of auditory cortex.

A great deal of electrophysiological research on speech has focused on two evoked responses: the *N100* and the *mismatch response* (see Figure 1). The auditory N100 (the 'N' stands for the negative voltage deflection characteristic of this response), and its magnetic counterpart M100, are generally referred to as exogenous response components, 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 phonetic/phonological 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; Figure 1) (Näätänen, Gaillard, & Mäntysalo, 1978; for review see Näätänen, 1992; Näätänen & Winkler, 1999). The mismatch response typically occurs 150-250 ms after the onset of the deviant sound, and is known to originate in supratemporal auditory cortex (Alho *et al.*, 1998; Hari *et al.*, 1984; Sams *et al.*,1991; Scherg, Vajsar, & Picton, 1989; for review of localization evidence see Alho, 1995). The mismatch response 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 (Aaltonen *et al.*, 1993; Lang *et al.*, 1990; Tiitinen *et al.*, 1994), and for this reason a number of studies have examined whether mismatch response amplitudes are affected by the categorial status of pairs of sounds, and whether mismatch amplitudes track the discriminability profiles established in behavioral phonetics research.

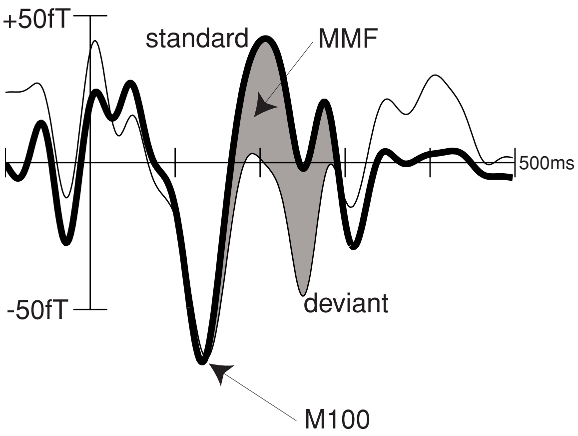


Figure 1: Sample averaged MEG response to speech sounds, measured at a single recording site in a mismatch paradigm. Both frequent (‘standard’) sounds and infrequent (‘deviant’) sounds elicit a similar M100 response. Responses to standards and deviants begin to diverge in the 150-200ms latency range. This divergence is known as the *Mismatch Negativity* (EEG) or the *Magnetic Mismatch Field* (MMF).

The focus of this paper is on evidence for different levels of representation and how they might be encoded in the brain. Section 2 reviews the basic levels of phonetic/phonological representation. Section 3 discusses coding hypotheses and relevant electrophysiological evidence for a number of different phonetic properties. Section 4 reviews reasons why phonological representations are fundamentally different, and more abstract, than phonetic representations, and reviews some electrophysiological evidence in this area. Section 5 concludes.

2. Evidence for 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.

The representation of speech carried by the peripheral auditory system provides a relatively faithful representation of the *acoustics* of speech. This is an analog representation of speech, and it is probably not modified by exposure to specific languages.

Phonetic representations of speech are also analog, but they more closely approximate an organization of speech sounds into linguistically-relevant categories. Linguistically-relevant acoustic distinctions are represented more strongly, and linguistically-irrelevant distinctions are represented less strongly. Phonetic representations are almost certainly modified by exposure to specific languages, such that the same speech sounds will be represented differently by speakers of different languages at this level. Nevertheless it is commonly assumed that phonetic representations can be derived from acoustic representations by means of relatively simple non-linear transform functions. For this reason, acoustic phonetic representations are often viewed as variously 'stretched' and 'compressed' versions of acoustic space (Kuhl, 1994). Additionally, phonetic representations may be modality-specific, distinguishing acoustic-phonetic and articulatory-phonetic representations.

Phonological representations of speech are discrete rather than analog. They are built from categories such as phonemes, phonological features, syllables, and feet, and they can be manipulated by symbolic processes which conform to general schemata (e.g. $[+x\ +x]\qquad[+x\ –x]$, or x --> y / z___, where x, y, and z are individual phonological features or sets of features). Phonological representations differ substantially across speakers of different languages, and almost certainly cannot be derived by means of a simple transform from acoustic or phonetic representations. Phonological representations are most likely modality neutral, in that they are deployed in speaking, listening, and reading.

Some of the fundamental questions in studying how the brain supports phonetics and phonology include the following. How are category representations encoded? How are these representations shaped by experience? How uniform is encoding and learning across different categories? Answers to these questions may be quite different in phonetics and phonology. In *phonetics*, it is likely that different phonetic categories are encoded and shaped by experience in a variety of different ways. Therefore, what we learn about the neural encoding of one phonetic category may not generalize to other phonetic categories. In *phonology*, on the other hand, it is more plausible to assume that different phonological categories and features are encoded in similar ways. Therefore, it is more likely that findings about how one phonological category is encoded can be extended to the encoding of other phonological categories.

3. Phonetic Representations

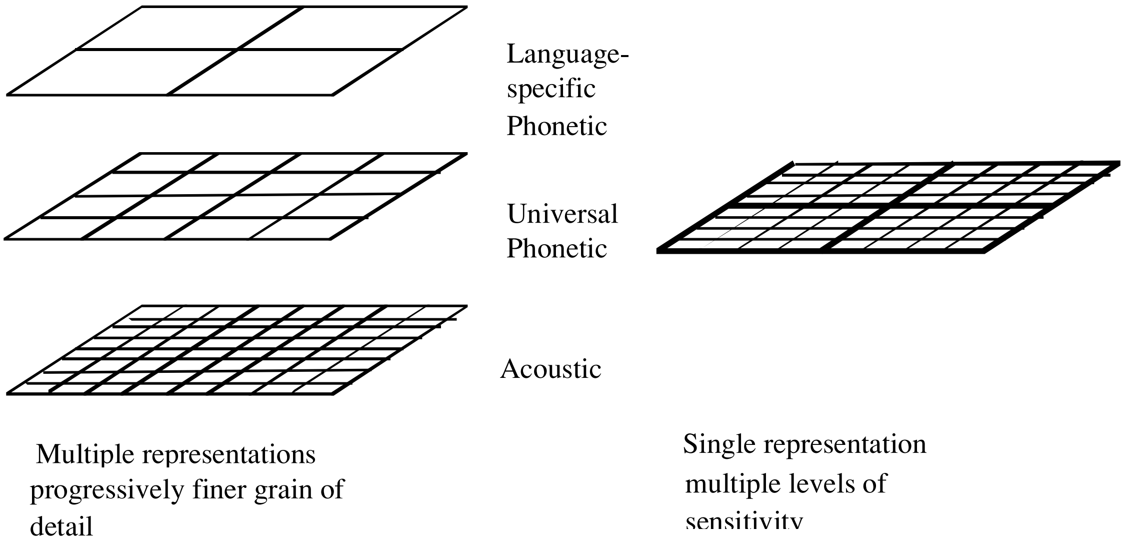
Evidence for the distinction between acoustic and phonetic representations of speech comes from a wide variety of findings about the identification and discrimination of speech sounds. Classic results on *categorical perception* show that when speakers are asked to classify or discriminate sounds from an acoustic continuum (e.g. [i] to [u], or [bæ] to [pæ]), responses vary across the continuum. Sounds from the same phonetic category are less easily discriminated from one another than sounds from different phonetic categories, and identification judgments typically yield a step-like function, with consistent judgments for most members of the continuum and only a narrow window of uncertainty at the boundary between the phonetic categories ('categorical perception'). Nevertheless, the fact that some within-category discrimination is retained, both for vowels (Fry, Abramson, Eimas, & Liberman, 1962; Pisoni, 1973; Stevens, Liberman, Studdert-Kennedy, & Ohman, 1969) and to a lesser extent for consonants (Carney, Widin, & Viemeister, 1977), suggests that phonetic representations do not consist of discrete categories, but are instead a non-linear transform of acoustic representations. Similar conclusions can be drawn from the finding that phonetic categories show a prototype structure (Miller, 1994; Samuel, 1982; Volaitis & Miller, 1992).

Work on the development of native language phonetic representations indicates that infants begin life as 'universal listeners', with the ability to perceive speech categorially (Eimas *et al*, 1971) and the ability to discriminate speech contrasts that are phonetically relevant in any of the world's languages (see Jusczyk, 1997 for review). However, effects of exposure to specific languages appear early – by around 6 months of age infants show greater sensitivity to native-language vowel categories (Kuhl *et al.*, 1992), and by around 10 months of age infants can only reliably discriminate contrasts from their native language (Werker, 1995; Werker &

Tees 1984, 1999). The evidence from infancy thus suggests that there are innate representations, at least at the phonetic level.

The change in perceptual abilities over the first year of life suggests a contrast between innate universal phonetic representations and the native-language phonetic representations which develop in the second half of the first year. However, it is unclear whether these are two independent representations which co-exist in mature speakers, or whether native-language phonetic representations are a reshaped version of the initial universal phonetic representations, which *replaces* the initial state. Both of these alternatives have been proposed, and there is interesting evidence in favor of both positions. Werker has argued that the development of adult speech perception abilities does not reflect a loss of innate representations, but rather reflects the development of *additional* language-specific representations (Werker, 1994; Werker & Tees, 1999). On the other hand, Kuhl has argued that early experience has the effect of explicitly *reshaping* the innate representations, ‘stretching’ and ‘compressing’ phonetic space in such a way that native-language categories show an enhanced representation, and non-native categories are less strongly represented (Kuhl, 1994).

Kuhl’s 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. In self-organizing maps, repeated presentation of a given stimulus causes cells tuned to that stimulus to increase their sensitivity to the stimulus, and also causes neighboring cells to respond more similarly. The effect of this is the formation of attractor basins, or, in Kuhl’s terminology, ‘perceptual magnets’. Applied to the development of speech perception, the suggestion is that repeated presentation of native language phonetic prototypes causes the innate phonetic maps in the infant brain to be reorganized, such that the representations of the prototypes are strengthened, and such that they exert a gravitational pull on similar stimuli. The effect of this is that discrimination of pairs of sounds close to a prototype is difficult, because sounds tend to be assimilated to the ‘perceptual magnets’, whereas discrimination of pairs of non-prototypes is easier. Under this view, the mechanisms driving the development of adult representations are made highly explicit, but it is also predicted that development of language-specific representations substantially erases the infant's universal phonetic representations.



Multiple representations progressively finer grain of detail

Single representation multiple levels of sensitivity

Figure 2. Alternative views of mature phonetic representations.

Schematic diagrams of the two approaches to mature phonetic representations are shown in Figure 2.[2] In what follows, we examine some of the evidence in favor of each of these positions, with a particular focus on relevant electrophysiological evidence.

VOWELS

Kuhl's approach has been most clearly articulated for the case of vowels and the liquids ([r, l]), and it provides a clear hypothesis about the neural representation of vowel categories. The phonetics of vowels are assumed to be represented in an abstract 'map' (Kuhl, 1994). This map may be encoded spatially in a region of cortex, but this need not be the case. Native-language vowel prototypes form a series of attractor basins in this map, such that sounds in the

---

[2] Although Werker regularly uses the expression *functional reorganization* to describe her approach, this may be misleadingly understood to mean that innate representations are *modified*, rather than *augmented* by additional representations of native language phonetic categories, as Werker in fact proposes (Werker, 1994, 1995).

acoustic neighborhood of the prototypes are hard to distinguish from the prototype itself. 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, on the other hand, are not represented – those boundaries are lost (or at least greatly attenuated) over the course of development.

Vowels are relatively easy to study, because their acoustic properties are well-described and easily manipulated. A good deal of information about vowel height and backness is carried in the values of the first three *formants* of the vowel (F1, F2, F3; formants are bands of increased energy in the spectrum of the vowel). The fundamental frequency (F0) of the vowel conveys important information about affect, focus and speaker identity, but is relatively unimportant in the identification of vowel categories. There are competing accounts of what are the critical cues for vowel category identification, but there is broad agreement that the relationship among the lower formants (F1-F3) is of central importance (e.g. Syrdal & Gopal, 1986).

Behavioral evidence on vowel perception is consistent with Kuhl's view of adult phonetic representations as a reshaped version of innate phonetic representations, rather than as an additional level of representation. Infants form 'perceptual magnets' around native language vowels as early as 6 months of age (Kuhl, 1991; Kuhl *et al.*, 1992), and there is no clear evidence that adolescents or adults retain an innate 'universal' vowel space after they have acquired their native language. These 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 M100 responses elicited by vowels have shown that this response component reflect 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), Poeppel *et al.* (1997) showed that the latency of M100 responses to vowels is sensitive to variation in vowel category (/a, i, u/), and is relatively insensitive to variation in speaker, i.e. male vs. female voice. A further study showed that the M100 latency variation is primarily due to variation in F1, a good predictor of the height of a vowel (Govindarajan *et al*., 1998). Studies of this kind show that the M100 is sensitive to certain important cues to vowel categorization, but do not speak to the question of 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, languages which are related but which show slightly different vowel categories. In each group of speakers, vowel contrasts involving pairs of native language vowel prototypes elicited larger mismatch responses than contrasts involving vowels which did not correspond to native language prototypes. This phonetic effect was independent of the effect of acoustic distance between vowels on the size of the mismatch response. Aaltonen *et al.* (1997) report similar findings in a single-language study of Finnish vowels: the amplitude of mismatch responses correlates with the discriminability profiles predicted by Kuhl's Perceptual Magnet theory (but cf. Sharma & Dorman, 1998 for conflicting evidence).

Note that although electrophysiological results such as these are compatible with Kuhl's single-level approach, most are also compatible with Werker's alternative multiple-level approach. This is perhaps not surprising, since the multiple-level approach assumes richer representations than the single-level approach. The main contrast between the predictions of the two approaches involves the availability to adults of latent non-native category representations. Unfortunately, this means that the best evidence for the single-level view would be *negative* results showing the unavailability of non-native categories.

The notion of a 'vowel space' can been clearly described at an abstract level, and we find substantial behavioral and electrophysiological support for certain properties of this space. However, more specific evidence on *how* this abstract space is encoded physiologically has been elusive. A natural hypothesis would be that the vowel space directly maps onto a two-dimensional region of cortex, in which, for example, cells are sensitive to specific coordinates of

F1 or F2 values. Inspired by reports that MEG recordings show evidence of spatial ‘tonotopic’ maps in auditory cortex (Romani, Williamson & Kaufman, 1982; Pantev *et al.*, 1988), plus numerous 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 always difficult to draw conclusions from negative findings, results so far have not been encouraging. Diesch *et al.* (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 1kHz 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, then, brain recordings back up other kinds of evidence for the properties of phonetic vowel representations. The available evidence is consistent with Kuhl’s claim that native language vowel representations are the result of reshaping innate phonetic representations, but this evidence is mostly also ultimately compatible with Werker’s multiple-level alternative. Meanwhile, there is no clear evidence to date for a spatial encoding of a ‘vowel space’.

VOICING

The articulatory gestures which distinguish speech sounds have a wide variety of acoustic consequences. As a result, individual phonetic features show a great deal of diversity in their primary acoustic correlates. This in turn has the effect that a plausible neural coding hypothesis for one phonetic feature is inappropriate for other features. As is well-known, 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. Classic studies show that stops with shorter VOTs are perceived as voiced and stops with longer VOTs are perceived as voiceless, with a sharp perceptual boundary around 30ms VOT (Liberman *et al.*, 1961). Based on the finding that VOT continua are perceived categorially by young infants (Eimas *et al.*, 1971) and by non-human mammals (Kuhl & Miller, 1978), 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 in a number of papers (Steinschneider *et al*, 1994, 1996, 1999; Sinex & MacDonald, 1988, 1989; Sinex, 1993). 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. The neurons respond strongly each time an appropriate stimulus is presented. However, the act of responding to a preferred stimulus typically has the effect of *desensitizing* the neurons for a short period (known as 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 length of the inactivation period for certain auditory neurons which respond to both the noise-burst and the voicing onset. When VOT is longer than around 30ms, 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. Direct recordings from cells in primary auditory cortex show precisely this pattern of responses to syllables from a VOT continuum in the macaque (Steinschneider *et al.*, 1994, 1996) and in humans (Steinschneider *et al.*, 1999). Similar response patterns have been recorded in the auditory nerve of the chinchilla (Sinex & MacDonald, 1988, 1989).

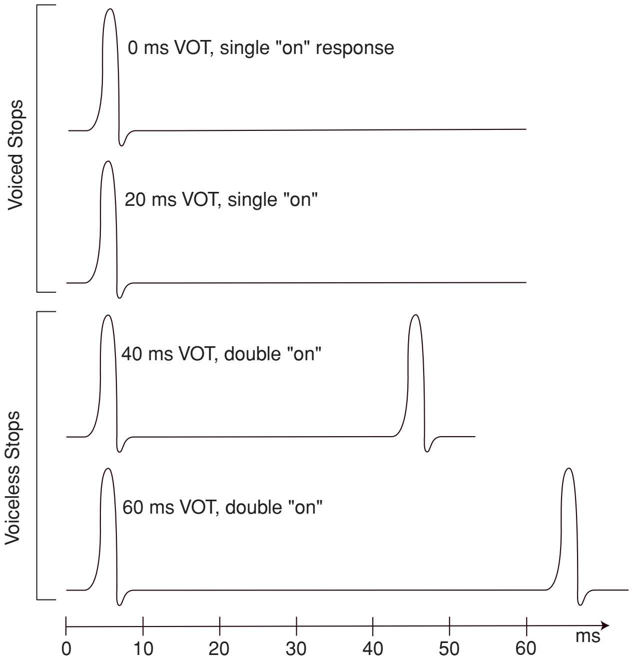


Figure 3: Schematic diagram of neural encoding of voiced vs. voiceless stops as single vs. double “on” response.

The ‘double on’ account of voicing is a highly explicit neural coding hypothesis, which suggests that voicing categories are based on low-level properties of the auditory system. We might expect a mechanism based on such a basic property of neurons to show very limited plasticity. There is some electrophysiological evidence from humans in support of this hypothesis. A striking prediction of the double-on hypothesis is that information about VOT is lost *asymmetrically*: responses to 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 (cf. Figure 3).

A series of electrophysiological studies have sought evidence for similar asymmetric representations of VOT information in evoked responses, and results have been promising. Sharma & Dorman (1999) report that in responses to a 9-step VOT continuum shorter VOTs (<40ms) elicit a single N100 response component (which covaries with VOT in latency), whereas longer VOTs (>40ms) elicit two distinct N100 components, only the second of which covaries with VOT in latency. Phillips *et al.* (1995) report pattern of MEG responses elicited by a 9-step VOT continuum which is very similar to this, except that M100 response latencies within the voiced range of the continuum do not covary with VOT (as predicted by the Steinschneider/Sinex model).

Sharma & Dorman (1999) also compared a between-category VOT contrast (30ms vs. 50ms) to a within-category contrast (40ms vs. 60ms) in a mismatch paradigm. The between-category contrast elicited a significantly larger mismatch response than the within-category contrast. This effect of phonetic category membership suggests that phonetic category representations are available to low-level cortical systems. However, the result in fact goes beyond what is predicted by the Steinschneider/Sinex encoding hypothesis, because this hypothesis preserves within-category distinctions among stimuli with long VOTs.

These electrophysiological studies show first that the ‘double on’ representation of VOT may be present in humans, and second, that the representation of voicing categories may be achieved by low-level cortical mechanisms. The correlation between direct recordings and electrophysiological results is impressive. However, there are a number of phonetic properties of voicing categories which do not obviously follow from this hypothesis, and so it is unlikely that the phonetics of voicing could 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. Evidence for preservation of within-category VOT information (e.g. prototype effects) comes from both voiceless and voiced stops (Samuel, 1982; Volaitis & Miller, 1992). Second, VOT boundary values vary across place of articulation (labials show shorter boundaries, velars show longer boundaries, cf. Summerfield, 1982), and may vary 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 which (for example) change voiced consonants into voiceless consonants in syllable-final position.

PLACE OF ARTICULATION

A third feature which we will discuss is Place of Articulation, which distinguishes (among other things) *labial* consonants /b, p, m/ from *alveolar* consonants /d, t, n/ and *velar* consonants /g, k, /. Place of articulation presents a rather different picture than vowels or voicing. In this instance we find that the neural encoding hypotheses are even less clear, but the evidence for multiple levels of phonetic representation is better.

Two reasons why place of articulation has provided less specific coding hypotheses may be the following. First, the acoustic cues to place of articulation are less well understood. The most widely examined acoustic cue to place of articulation – at least for stop consonants – involves the distribution of spectral energy in the formant transitions between consonants and neighboring vowels. However, 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 *et al.* 1998), the nature of the solution is less clear than it is in the case of vowels or voicing.

Second, 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 although there is scope for cross-language variation, this variation is more constrained than variation in vowel categories. For example, whereas articulatory constraints allow a number of different vowel category boundaries between /o/ and /e/, articulatory constraints impose a fixed articulatory boundary between stops that are articulated with the lips and stops that are articulated with 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 predicted by Werker’s multiple-level view. It is less clear whether whether such a universal category inventory is feasible for voicing or vowels, given the existence of overlapping categories across languages. In fact, the best evidence for Werker's view comes from place of articulation categories. Recall that the best evidence for multiple phonetic levels of representation would be a demonstration that universal and language-specific category representations are present *simultaneously* in adult speakers of a language.

The classic developmental finding that younger infants (up to 8 months) perceive non-native place contrasts which older infants (10+ months) cannot perceive (Werker & Tees, 1984) is compatible with both reorganization and overlay views of phonetic representation, because this simply shows that important change occurs around 8-10 months of age. However, studies by Werker & Logan (1985) on adults shows that many adult English speakers, who fail to discriminate an alveolar/retroflex place contrast when sounds are presented at long interstimulus intervals (ISI), show categorial discrimination of these place categories when the pairs of sounds are presented at brief (<500ms) ISIs. This suggests that both universal and language-specific representations are co-present, but that the lower-level representation decays faster.

Available electrophysiological evidence is also compatible with this view. 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. Based on the premise that mismatch response amplitude tends to increase as the discriminability of the tested sounds increases, a number of studies have asked whether easily discriminable between-category place contrasts yield larger amplitude mismatch responses than hard-to-discriminate within-category place contrasts. A number of studies have answered this question in the negative (Sams *et al.*, 1990; Sharma *et al.*, 1993; Maiste *et al.*, 1995), finding no special status for native language category contrasts. These results do not themselves support Werker's view, since they could be explained by purely acoustic representations. More relevant for Werker's hypothesis is a study by Dehaene-Lambertz (1997), which compared mismatch responses elicited in French speakers by a native language contrast (/d/ vs. /b/) a non-native contrast (dental /d/ vs. retroflex

/D/), and a contrast that is never linguistically relevant in any language (ba1 vs. ba5), using stimuli from a 17-step synthetic continuum. Dehaene-Lambertz reports that the native-language contrast (/ba/ vs. /da/) elicited a clear mismatch response, starting around 200ms, and that no other contrast elicited a mismatch response as early as this. Interestingly, however, Dehaene-Lambertz also reports that at a later time window (around 370ms) the non-native Hindi contrast elicited a significant mismatch responce, whereas the contrast that is never linguistically relevant did not. This suggests that universal phonetic representations contribute to adult electrophysiological responses, above and beyond the contribution of pure acoustics.[3]

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.

Phonetic feature	**Phonetic sensitivity?**	**Reference**
vowel (backness)	yes	Näätänen *et al.*, 1997 Winkler *et al.*, 1999
voicing	yes	Sharma & Dorman 1999
place of articulation	no	Sams *et al.*, 1990 Sharma *et al.*, 1993 Maiste *et al.*, 1995
place of articulation	yes	Dehaene-Lambertz 1997

Table 1: Results of tests of phonetic sensitivity of auditory cortex mismatch response.

---

[3] The experiment design used by Dehaene-Lambertz may explain the contrast between the phonetic sensitivity revealed in her study and the apparent lack of phonetic sensitivity found in earlier mismatch studies of place of articulation. Whereas most studies present a long sequence of sounds contrasting a single standard-deviant pair in a passive listening task, Dehaene-Lambertz presented short sequences of sounds in XXXY sequences in an active task in which subjects had to judge whether the final sound matched the rest of the sequence.

As we have seen, to the extent that existing electrophysiological results support either Kuhl's reshaping view or Werker's augmentation view of phonetic representations, results do not *consistently* favor one view 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 Werker's hypothesis receives more support (and is more feasible) in areas of phonetics 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 Kuhl's 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).

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 support for these hypotheses. The fact that different results provide support for a variety of mutually incompatible proposals may simply reflect the fact that phonetic categories are a motley bunch. The categories involve different acoustic parameters, they may be learned differently, and therefore it would not be surprising to find that they are represented differently.

In *phonology*, on the other hand, existing evidence indicates the need for representations which are fundamentally different from those found in phonetics, and for 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 rather 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. First, phonological categories are discrete, symbolic units. They are the 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 good instance of /k/, or the degree to which the /k/ is coarticulated with the following vowel. Such details are specified independently.

Some studies in the mismatch paradigm have used the irrelevance of within-category differences as a probe for phonolological 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*, i.e. 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* (in press) used this strategy to probe for phonological representations of /d/ and /t/. Using syllables from a synthetic VOT continuum, they presented sequences like the one in Figure 4. 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 *et al*., 1993).

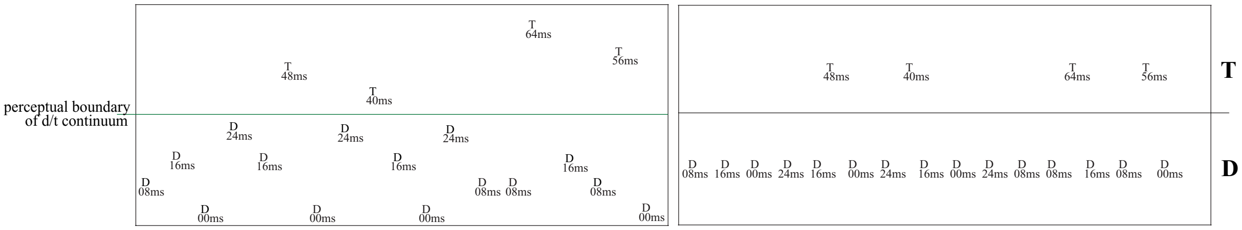


Figure 4: Design of Phonological Mismatch Study (Phillips *et al*., in press). Left: acoustic representation of sound sequence; right: phonological representation.

In sequences like this, there is both within-category and between-category variation in VOT. Therefore, there is no many-to-one ratio of standards to deviants at the acoustic and phonetic levels. It is only at the phonological level that there is a many-to-one ratio of standards to deviants. Based on the fact that a classic auditory cortex MMF was elicited, Phillips *et al.* conclude that phonological representations of the categories /d/ and /t/ must be available to the auditory cortex generator of the mismatch response. In a later study using whole-head MEG recordings, Phillips *et al.* (2000b) extended this approach to the level of the phonological features *voiced* and *voiceless*. Voiceless stimuli consisted of multiple tokens of each of /pæ, tæ, kæ/, and voiced stimuli consisted of multiple tokens of each of /bæ, dæ, gæ/. This contrast also elicited a mismatch response, indicating that the representation of the phonological feature [±voice] is available to auditory cortex.

An elegant study by Dehaene-Lambertz *et al.* (in press) used a cross-language design to probe Japanese speakers' representations of syllable structure. Japanese syllables conform to a strict (C)V(N) structure: the vowel nucleus of a syllable may be preceded by at most one consonant, and followed by at most one nasal consonant. In contrast, French (like English) allows a much wider range of syllable-types than Japanese, including syllables as complex as 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: e.g., *brother*   *burazaa*; *volcano*   *borukano*. Another consequence is that when Japanese speakers listen to words which do not conform to the syllable-structure of Japanese, they may perceive the words as if they did conform to the syllable-structure of Japanese, i.e. with 'phantom' vowels. Accordingly, Dupoux *et al.* (1999) showed that Japanese speakers perceive the nonsense-word *ebzo* as *ebuzo*. The same speakers also have difficulty discriminating *ebzo* from *ebuzo*, whereas speakers of French obviously have no such difficulty. This cross-language contrast presumably reflects the fact that speakers represent words using the phonological syllable-templates made available by their native language. Dehaene-Lambertz *et al.* (in press) replicated this finding in an ERP mismatch paradigm. Speakers of French and speakers of Japanese listened to sequences of pseudo-words presented in an XXXY pattern, such as *ebzo, ebzo, ebzo, ebuzo*, in which the final pseudo-word either matched the preceding words or contained an additional vowel. Contrasting final vowels elicited a clear mismatch response in French speakers, but not in Japanese speakers, suggesting that the mismatch response is sensitive to native language syllabic representations.

Interestingly, however, Japanese speakers did show a later mismatch-related response to the *ebzo*/*ebuzo* contrast, beginning in the 300-400ms latency window. This effect is similar to the late response to non-native contrasts observed in Dehaene-Lambertz' earlier study of cross-language place-of-articulation contrasts (Dehaene-Lambertz, 1997).

Note that although these mismatch studies present evidence that phonological categories are available to auditory cortex, they give no indication of *how* the categories are encoded.

B. ACOUSTIC DIVERSITY OF CATEGORIES. A second property of phonological categories, which shows that they are more abstract than phonetic categories, is the fact that phonological categories often group together sounds which are acoustically and phonetically quite diverse. For example, although VOT provides a good characterization of the difference between voiced and voiceless *stop* consonants (b/p, d/t, g/k), the contrast between voiced and voiceless *fricatives* (z/s, v/f, zh/sh, j/ch) is acoustically rather different, and the same is true for voicing contrasts in vowels and nasals, in languages where these are phonologically contrastive. Despite this acoustic variation, phonological devoicing processes consistently affect all categories alike. In addition to variability in the implementation of features *across* categories, there is substantial variation in the implementation of a given feature *within* individual categories. Voice onset time is a fine cue for the identification of syllable-initial /t/, but is not a useful cue for syllable-final /t/. To date, there have been no electrophysiological studies of sounds which are phonologically uniform but which show this degree of acoustic diversity.

A similar argument can be made with place of articulation categories. It is well-known that the formant transitions that provide the strongest cues to place-of-articulation vary substantially depending on the vowel context in which a consonant appears. There have been many attempts to identify acoustically invariant cues which characterize place categories across speakers, manners of articulation, and vowel contexts (e.g., Blumstein & Stevens, 1979, 1980; Kewley-Port, 1983). A promising recent proposal by Sussman and colleagues suggests that *locus* *equations* (linear regressions of coordinates defined by F2 values at vowel onset and vowel nucleus) are the invariant cues to place-of-articulation (Sussman *et al*., 1991, 1998). Despite the numerous successes of locus equations, they fail to group together velar consonants when they are palatalized before high front vowels (as in *kit*) with their non-palatalized counterparts found in other vowel contexts (as in *cat* or *cut*), although these sounds clearly fall together phonologically.

C. NEUTRALIZATION. Further evidence that the abstractness of phonological categories goes beyond mere grouping of acoustically different sounds comes from phonological *neutralization* processes. These are processes which cause phonologically distinct categories to be realized as phonetically *identical.* A well-known neutralization processes is the word-final devoicing of consonants found in many languages. In German, for example, the words *bunt* (‘colorful’) and *Bund* (‘federation’) are both pronounced with a final [t] when spoken in isolation or in phrase-final position. When suffixes are added to the word *Bund*, such that the /d/ is no longer word-final, the consonant is voiced, as in *Bundestag* (‘Federal Parliament’). 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, e.g. *bidder* ([b]) vs. *bitter* ([b]).[4] 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.[5] An additional, and very commonplace example of neutralization involves the effect of stress on the pronunciation of vowels. Unstressed vowels in English are generally pronounced

---

[4] 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).

[5] There has been recent controversy over the question of whether neutralization processes do in fact involve absolute loss of phonetic contrast. It has 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 strong evidence for genuine neutralization, based on the syllable-final manner-neutralization process found in Korean, which causes all of plain /t/, aspirated /t^h/, tense /t’/ and fricative /s/ to be realized as plain (but released) [t] in word-final position. This neutralization involves unambiguous *loss* of manner distinctions.

as the neutral vowel schwa ([ ]), leading to alternations in vowel pronunciation, as in the underlined vowels in *similar* and *similarity* ([ ] vs. [æ]).

D. DISSIMILATION. Fourth, 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 ' in '*o-koora* ('to root out').

E. SIMILARITY VS. IDENTITY. Fifth, one of the most compelling pieces of evidence for the abstractness of phonological representations involves various 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 in phonological theories which allows a sequence of two identical vowels to be represented as a single category, which is realized discontinuously (cf. Phillips, 1994 for further examples of this kind). Clearly, though, this requires representations of sound sequences which are a good deal more abstract than simple transforms of the linear 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. As we have seen, a striking property of phonetic categories is their diversity, which is reflected in the variation in relevant electrophysiological results, and in the diverse range of possible coding hypotheses. By contrast, most evidence suggests that phonological categories are substantially more homogeneous. They are symbolic units, which all combine in more or less the same manner and undergo more or less the same kinds of processes (e.g. deletion, neutralization, assimilation, dissimilation…). Therefore it is plausible that the neurophysiological coding of phonological representations is simpler and more uniform than the coding of phonetic categories.

The view of phonological representations that I am presenting here is not one that all phonologists would wholeheartedly endorse. In recent years in particular, many phonologists have been impressed by the fact that different phonological categories undergo different kinds of processes. Furthermore, it is common to try to account for these differences among categories in terms of the ‘phonetic naturalness’ of phonological processes. However, these caveats do not undermine the fact that phonological representations clearly are different – and more abstract – than phonetic representations.

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. Non-linearities in the acoustics-to-phonetics mapping can be correlated with non-linearities in ERP or MEG response components. When the predicted non-linearities 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 lent a good deal of support to the notion that phonetic representations are supported by supratemporal auditory cortex. Existing results are also fully consistent with the notion that phonetic representations are rather diverse, with different encodings for different categories. It is almost certain that there are additional brain regions which support phonetic representations and processes, based on evidence from functional brain imaging and deficit-lesion studies, but the electrophysiological literature has offered little to-date on these other regions.

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 is beginning to add an extra dimension. For example, results which implicate multiple levels of representation can challenge existing views of how these different levels interact in speech processing. An example is the finding by Dehaene-Lambertz, in two separate studies of different phenomena (Dehaene-Lambertz, 1997; Dehaene-Lambertz *et al.*, in press), that non-native contrasts are represented, but that these representations are accessed more slowly than native-language contrasts. This finding supports models in which universal and language-particular representations are co-present (Werker, 1994), but entails a rethinking of the time course of when these representations are available (cf. Avery & Idsardi, 2000).

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 involvement of phonological representations in 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.

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.

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.

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.

Avery, P. & Idsardi, W. (2000). Issues in the acquisition of contrast. Paper presented at the First North American Phonology Conference, Montreal.

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., Dupoux, E. & Gout, A. (in press). Electrophysiological correlates of phonological processing: A cross-linguistic study. *Journal of Cognitive Neuroscience.*

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

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.

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 Phonetic*s, **2**4, 295–312.

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. (1994). Learning and representation in speech and language. *Current Opinion in Neurobiology, 4*, 812-822.

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., 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.

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.

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. (1992). *Attention and Brain Function*. Hillsdale, NJ: Erlbaum.

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.

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. (in press). Auditory cortex accesses phonological categories: an MEG mismatch study. *Journal of Cognitive Neuroscience.*

Phillips, C., Pellathy, T. & Marantz, A. (submitted). Biomagnetic evidence for phonological natural classes as features. Manuscript, U. of Delaware & MIT.

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

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: reponse 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* (Publications of the Philological Society, 19). Oxford: Blackwell.

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., 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., Hämälainen, 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*. Norwell, MA: Kluwer, pp.307-313.

Slowiaczek, L. M. & Dinnsen, D. A. (1985). On the neutralizing status of Polish word-final devoicing. *Journal of Phonetics, 13*, 325-341.

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.

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.

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., McCaffrey, H. A. & Matthews, S. A. (1991). An investigation of locus equations as a source of relational invariance for stop place categorization. *Journal of the Acoustical Society of America, 90*, 1309-1325.

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.

Syrdal, A. & Gopal, H. S. (1986). A perceptual model of vowel recognition based on the auditory representation of American English vowels. *Journal of the Acoustical Society of America, 79*, 1086–1100.

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.

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*. Cambridge, MA: MIT Press, pp.93-120.

Werker, J. F. (1995). Exploring developmental changes in cross-language speech perception. In *An Invitation to Cognitive Science*. Part 1: *Language*, ed. L. Gleitman & M. Liberman, pp.87-106. Cambridge, MA: MIT Press.

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

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: Le Parler Ngbaka de Bokanga: Phonologie, morphologie, syntaxe. *Language* 41, 346-347.

Winkler, I., Lehtokoski, A., Alku, P., Vaiiio, 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.

colin@udel.edu
www.ling.udel.edu/colin