SOME MEG CORRELATES FOR DISTINCTIVE FEATURES

William J. Idsardi

Department of Linguistics and Program in Neuroscience and Cognitive Science,
University of Maryland, College Park MD, USA
idsardi@umd.edu

ABSTRACT

This presentation reviews the use of distinctive features for the mental representation of speech sounds, briefly considering three bases for feature definition: articulatory, auditory and translational. We then review several recent neuroimaging studies examining distinctive features using magnetoencephalography (MEG). Although this area of research is still relatively new, we already have interesting findings regarding vowel place, nasality and consonant voicing. Although this research is not yet definitive, some refinements of these experiments can be expected to yield important results for feature theory, and more generally for our understanding of the neural computations that underlie the transformations between articulatory and auditory space necessary to produce and perceive speech.

Keywords: distinctive features, vowel height, nasality, voice onset time, magnetoencephalography

1. INTRODUCTION

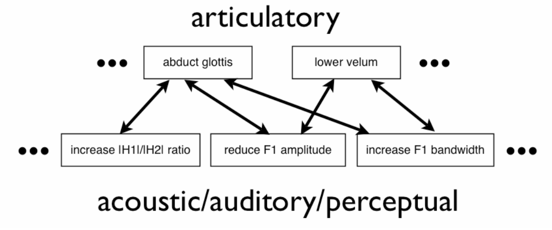
Our goal is to understand the neuro-biology of the human speech system to a degree akin to the current understanding of echolocation in the barn owl [36], [9]. Study of barn owl echolocation has more or less achieved Marr's [17] objectives of a coherent account across all three levels of description: computational, algorithmic and implementation. Of course this is still a very distant goal for speech science, but the use of new brain-imaging techniques, such as magnetoencephalography (MEG) promises to bring important new evidence to old problems. I will briefly review some positions on the role of distinctive features in the representation of speech sounds, and consider some MEG studies relevant to these questions. As should be expected at this early stage, the MEG findings are not definitive, but this is a growing area of research that will become increasingly important to both experimental work and theorizing about the mental representation of speech sounds.

2. THREE VIEWS OF FEATURES

In this presentation we will generally ignore the contentious issue of the ontological status of phonological features; that is, the question of whether features are simply a convenient classificatory system for sets of whole-sound fundamental speech units (phones or phonemes) as implied by IPA chart, or whether the features are the fundamental units themselves [12], [11]. We do wish to briefly review a separate issue, the question of the type of information to be used to define the features. There are three main views on this question of feature definition, namely that the primary definitions are (1) articulatory, or (2) auditory (or perceptual), or (3) translational, providing the basis of the mapping between articulation and perception.

As has long been recognized, there is a complicated, many-to-many relationship between the motor speech actions and their acoustic and auditory consequences; a small fragment of this mapping is shown in Fig. 1.

Figure 1: Many-to-many relationships (after [34]).



Different researchers have emphasized different aspects of this problem. In the next three sections, we give quick summaries of the three positions.

2.1. Articulation-oriented theories

The most prominent of the articulation-oriented approaches to the mental representation of speech sounds is the motor theory, developed at Haskins Laboratories starting in the 1950's (see for example the collected papers in [15]). A more recent summary of this perspective and connecting it with the neuro-biological discovery of mirror neurons is given in [8:241]: "For example, in the case of speech, we hypothesize that gestural task-space, i.e., the space of articulatory constriction variables, is the domain of confluence for perception-oriented and action-oriented information during the perception and production of speech. Such a task-space would provide an informational medium in which gestures become the objects shared by speech production and perception, ... It is possible that a mirror neuron system could be the neurophysiological instantiation of this cognitive coupling, though the identification of such a specific system in humans remains an outstanding challenge."

Likewise Halle [11:6-7] is explicit about the change in his views over the years: "At an earlier time I believed that features in memory were directly connected to the acoustic side … I now believe that there is a direct connection only between features in memory and the articulatory actions to which they give rise, and that the acoustic aspects of features play a secondary role in phonology."

In this view, the defining characteristics of features are given by the upper set of boxes in Fig. 1 (whether this be individual motor commands, such as "denervate the posterior cricoarytenoid" or gestural constellations of commands, such as "adduct the glottis"). The rest of Fig. 1 is then extra-linguistic and the conversion in the listener from acoustic events back to their articulatory progenitors must occur very early in the perceptual system, and ideally this mapping should be relatively domain-general and not specific to speech.

2.2. Perception-oriented theories

The reverse view is perception-oriented, giving primacy to the lower set of boxes in Fig. 1. This view has many proponents in the speech perception literature, but there is also important evidence from speech production studies. Their basic idea is that there are auditory targets for speech, and that the speaker's production system needs to match those targets through a feedback loop. Bite-block studies [6], [7] have shown that speakers can rapidly and accurately accommodate for unusual motor situations, and they are able to match the acoustic targets through unusual vocal tract configurations (though there are some limits to such accommodations [16]). [7] remark: "The formant patterns of the bite-block vowels were found to approximate those of the naturally spoken vowels. Measurements derived from lateral view still x-ray films showed that the bite blocks induce drastic articulatory reorganization. … A computer simulation of our speakers' compensatory strategy revealed that they behaved optimally according to acoustic theory. These findings suggest that a vowel target is coded neurophysiologically in terms of acoustically significant area-function…"

More refined versions of auditory-oriented theories, such as [10], [20], [22], initially train a forward model through auditory feedback, a process often compared to infant babbling. After training, the forward model then has some relative independence from auditory feedback, as is consistent with the relatively slow, drifting degradation of speech production observed in the post-lingually deaf, e.g. [14]. Unfortunately, the auditory system cannot rely on acoustic information as simple as that suggested in the lower boxes in Fig. 1, as even young children are apparently able to correct for different vocal tract sizes and morphology [18], for example matching not raw formant values, but rather transformed formant values appropriate to their own body size.

2.3. Translational theories

The translational approach is best exemplified by [12]. In this approach distinctive features have dual definitions, both articulatory and auditory, and it is the features themselves that provide the fundamental connection between action (articulation) and perception (audition). In terms of Fig. 1, the features are (some of) the lines connecting the top boxes with the lower boxes. Ideally, if such a theory was correct, we should find that features have limited motor implementations, and limited acoustic cues [33]. For example, the feature [+round] would define the connection between the motor gesture of lip-rounding (i.e. the enervation of the orbicularis oris) and a particular perceptual pattern, perhaps the down-sweep in frequencies across the whole spectral range (as is observed in the Doppler effect). In this view, the mapping between articulation and audition is linguistic in nature, and thus could be speech-specific, though evolution will tend to repurpose pre-existing equipment. [25] considers such theories in light of recent advances in the neuro-biology of speech.

2.4. Summary

The caricatures of the three models just presented are meant to highlight the difference amongst the points of view. Comprehensive study of speech production and perception leads to narrowing of the gaps between the theories. For example, analysis-by-synthesis models [32], [35] incorporate submodules similar in function to those in the sophisticated auditory-oriented models [10], [22].

3. MAGNETOENCEPHALOGRAPHY

Magnetoencephalography (MEG) is a non-invasive brain-imaging technique that measures the magnetic field generated by the electrical activity of the brain by using an array of super-conducting magnetic detectors. MEG provides excellent temporal resolution (~1ms) and reasonably good spatial resolution. A picture of the whole-head system at the Cognitive Neuroscience of Language Lab at the University of Maryland is shown in Fig. 2. The magnetic field from each detector is recorded digitally and can be analyzed to assess a number of different characteristics of the signals. Three of these measures are particularly relevant for current speech research: (1) the M100 latency, (2) the mismatch field, and (3) localization patterns.

Figure 2: KIT-UMD MEG scanner.

3.1. M100 Latency

Presenting a subject with a perceptual stimulus, such as a speech sound, will evoke a complex series of brain responses. For auditory events, there is usually a relatively strong response that occurs approximately 100 ms after the auditory event. This response is known as the M100 ("M" for magnetic, "100" denoting the approximate latency of the response). Although the morphology of the M100 response is still unknown, nevertheless it can be used to show that the brain differentially encodes two stimuli, showing that some aspect of this difference is represented in the primary auditory areas.

3.2. Mismatch fields

Most M100 studies study the brain response evoked by a single stimulus (albeit averaged over a large number of trials). A more complicated type of study contains two kinds of stimuli, one of which is presented more often to the subject (the "standard"), the other is presented more rarely (the "deviant"). The idea is that subject will construct some kind of summary representation of the set of standards in short-term memory, and that there will be a brain response (showing "violation of expectation") when a deviant is encountered (the "mismatch field"). Different sets of standards can be tested, and in theory this technique would allow a more-or-less direct test of phonological natural classes. That is, we would expect [p t k] to be a reasonable set of standards, to which [b d g] would be deviant, but we would not expect [p d k] to be an effective standard set.

3.3. Localization

MEG data can also be used to reconstruct the location of the source of the observed magnetic field. Speech studies using this method include [3], [4], [21]. Since this session also includes a report from Carsten Eulitz [2], I will not discuss this analysis technique any further.

3.4. Summary and discussion

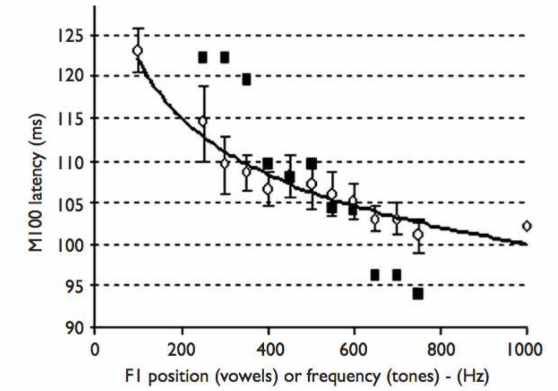
MEG is certainly not a panacea for neurological investigation of speech, but it is an important new tool that speech researchers should be aware of. The M100 latency acts as a kind of very early reaction time, offering some insight into what is represented in the early stages of auditory and linguistic processing. Mismatch field studies, though longer and more complicated to set up, offer the possibility of testing simple phonological natural classes.

4. VOWEL HEIGHT

Early auditory MEG experiments examined brain responses to sinusoidal tones. Poeppel et al [26] examined MEG responses to three different vowels ([a], [i], [u]) and showed that manipulating F1 changed the latency of the M100 response. In particular, [a] showed a shorter M100 latency, while [i] and [u] showed longer M100 latencies (the values for [i] vs. [u] were not significantly different).

Subsequent studies [28] show that this is not a simple acoustic effect. As shown in Fig. 3, the vowel response does not follow the sinusoidal tone response, but shows three distinct areas of response. Taking the sinusoidal response as the standard, the first three vowel points show a longer latency than expected, and the last three have a shorter latency than expected. The middle five points follow the sinusoidal response fairly well, but these points also showed diminished identification accuracy in a behavioral test. One interpretation, then, is that the middle five points do not categorize as either English /a/ or /u/. Whatever the correct interpretation may be, the experimental result is clear: the speech response is differentiated from the sinusoidal response. Of course we don't yet know if this diagnoses a feature such as [high] or [low]; to do that we would need at least to see parallel results from an /æ/-/i/ study as well.

Figure 3: (from [28]) M100 latency for vowel continuum (solid squares) and sinusoidal tones (open circles, with 1/f curve fit).



Work in progress [19] shows some additional evidence for higher-order invariants, in this case formant ratios. The hypothesis is that since listeners need to be able to normalize formant values across different speakers [18], and since they are able to do this quite rapidly, they need a relatively easy, on-line normalization procedure. Since F3 acts as a relatively good indicator of overall vocal tract length, we conjecture that listeners might use F3 as the basis for a formant ratio representation of vowel space (log F1/F3 x log F2/F3). If that is the case, then the previous results still follow because manipulating F1 also changed the F1/F3 ratio. To test this hypothesis, we held F1 and F2 constant and moved F3; the tested formant values are given in Table 1. The high and low F3 variants were obtained by moving F3 equal distances in mels above and below standard values for F3 in the two vowels tested ([ɛ] and [ə]) [23].

Table 1: Formant values (in Hz).

Vowel	F1	F2	F3
/ɛ/ (low F3)	580	1712	2156
/ɛ/ (high F3)	580	1712	3247
/ə/ (low F3)	500	1500	2040
/ə/ (high F3)	500	1500	3179

The two vowel types (/ɛ/ and /ə/) elicited different results. Changes in F3 for /ɛ/ did affect the M100 latency. The larger F1/F3 ratio (the lower F3; mean latency = 134ms) elicited a shorter M100 peak latency than the smaller F1/F3 ratio (the higher F3; mean latency = 138ms). However, changes in F3 for /ə/ did not significantly change the M100 latency.

Taken in conjunction with the previous study, we could conjecture that the lack of an effect for /ə/ is due to the lack of other phonemic central vowels in English. We are conducting a follow-up study with /o/ to see if results with back vowels pattern similarly to front vowels. Of course, other evidence could come from cross-linguistic investigation of languages with richer central vowel inventories, such as Vietnamese.

4.1. Summary and discussion

The M100 latency results show that some acoustic correlate related to vowel height is being neurally encoded in an interesting way. We have strong evidence that the encoding is not the same as that for sinusoidal tones, and we have some preliminary evidence to show that there may be a higher-order encoding of F1, perhaps encoded in some way similar to formant ratios.

5. NASALITY

There are fewer MEG studies investigating nasality or acoustic attributes potentially related to nasality (see, e.g. [27]). Since nasal vowels show an increase in the F1 bandwidth, studies manipulating bandwidth are potentially relevant. Two studies on bandpass filtered white noise [30], [31] showed that as the bandwidth of the filter increased, the amplitude of M100 response decreased, but there was no significant change in the M100 latency. So one potential prediction is that nasal vowels might show decreased M100 amplitude relative to the corresponding oral vowels.

Flagg et al. [5] conducted a very different test of phonological activity. They cross-spliced oral and nasal vowels with following oral and nasal consonants yielding congruent ([ab], [ãm]) and incongruent ([ãb], [am]) sequences. The incongruent sequences violate English phonotactic patterns. The M100 response latency evoked by the consonant following the vowel was longer in the incongruent sequences. But all of the sequences are linguistically possible in other languages (for example Indonesian shows perseveratory nasalization instead of the anticipatory nasalization in English). Thus, a tempting interpretation of the results is that “the temporal disparity in evoked neuromagnetic activity in English speakers reflects a role for language-specific phonological knowledge at the earliest stages of auditory processing.” [5]. Ideally we would replicate this with other phonotactic violations, and would show that the reverse sequences in Indonesian display the same effect (congruent: [mã], [ba], incongruent: [ma], [bã]).

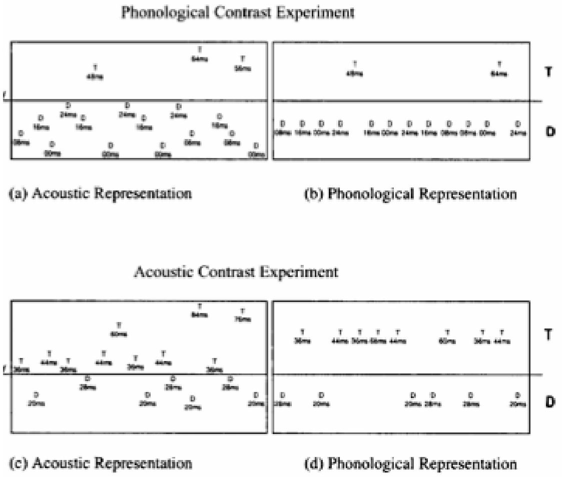
5.1. Summary and discussion

Even though there are few MEG studies relevant to nasalization, [5] represents an important advance in MEG techniques for speech. Their method allows sequences of sounds to be examined and shows that at least some incongruent sequences result in a longer latency in the M100 component. Since all languages have legal and illegal phonotactic patterns, this promises to be a particularly fruitful area of research over the next few years.

6. VOICING

Voicing distinctions (as implemented in voice onset time) are some of the most-studied aspects of speech production and perception. Phillips et al [24] used a very clever mismatch design to test phonological knowledge of voicing, see Fig. 4. A synthetic VOT continuum was created for the phonological contrast condition. In this condition, the standards were drawn from a legitimate phonological category of English (e.g. /d/) even though there was substantial acoustic variation in both the standards and the deviants. Nevertheless, a mismatch field was induced by the deviant stimuli in this condition. The other (acoustic) condition was created by adding approximately 20ms of voice onset time to all the stimuli. The set of standards in this condition was not a coherent phonological category of English, but spanned across two categories (e.g. /d/ and /t/). In this condition the deviants did *not* induce a comparable mismatch field. We can conclude that the subjects were unable to induce an ad-hoc category for the standards in the acoustic condition, and that the existing VOT boundaries for the subjects were predictive of whether the mismatch field was induced. The techniques employed in this study are important as they demonstrate how phonetic variation within a phonological category can be incorporated into the stimuli and still induce a categorical MEG response.

Figure 4: (from [24]) Design of phonological mismatch experiment


Phonological Contrast Experiment
T
D
(a) Acoustic Representation
(b) Phonological Representation
Acoustic Contrast Experiment
T
D
(c) Acoustic Representation
(d) Phonological Representation

Employing this same technique, Kazanina et al [13] examine prevoicing in Korean and Russian, a difference that is phonemic in Russian, but only allophonic in Korean (due to inter-sonorant voicing). The Russian speakers showed a significant mismatch field response to the deviant stimuli, whereas the Korean speakers did not. Note that Korean does have two clusters of tokens for plain stops phonetically: one with pre-voicing, and one without. But this level of phonetic details seems to be inaccessible to the speakers in constructing the short-term memory summary pattern necessary to detect the deviant stimuli. Thus, the oddball paradigm seems particularly useful at distinguishing between phonemic and allophonic distinctions.

6.1. Summary and discussion

The mismatch field studies on voice onset time show that MEG experiments can incorporate significant phonetic variation while still inducing categorical, phonemic responses, and can even distinguish phonemic from allophonic categorization.

7. CONCLUSION

Although MEG work on speech is still in its infancy, it promises to allow us to examine neural responses to speech in new ways and to reveal new aspects of speech perception relevant to age-old questions about the mental representation of speech. Latency of the M100 response can be modulated by changes in vowel formants and by the legality of phonotactic sequences. Oddball paradigms inducing mismatch fields are a subtle technique for testing which groups of sounds can form a coherent standard, and thereby test which sets of sounds form a phonemically natural class in a language. It is a very safe bet to predict that we will see an increasing number of MEG speech experiments testing a wider variety of contrasts across different languages.

8. ACKNOWLEDGEMENTS

Thanks to Colin Phillips, David Poeppel, Phil Monahan, Diogo Almeida, So-One Hwang and Elissa Flagg for many helpful discussions.

9. REFERENCES

[1] Arbib, M.A. (ed) 2006. *Action to Language via the Mirror Neuron System.* Cambridge: Cambridge University Press.

[2] Eulitz, C. 2007. This volume.

[3] Eulitz, C., Lahiri, A. 2004. Neurobiological evidence for abstract phonological representations in the mental lexicon during speech recognition. *J. Cog. Neurosci.* 16, 577-83.

[4] Eulitz, C., Obleser, J., Lahiri, A. 2004. Intrasubject replication of brain magnetic activity during the processing of speech sounds. *Cog. Brain Res.* 19, 82-91.

[5] Flagg, E.J., Oram Cardy, J.E., Roberts, T.P. 2006. MEG detects neural consequences of anomalous nasalization in vowel-consonant pairs. *Neurosci. Lett.* 397, 263-8.

[6] Fowler, C.A, Turvey, M.T. 1981. Immediate compensation in bite-block speech. *Phonetica* 37, 306-26.

[7] Gay, T., Lindblom, B., Lubker, J. 1981. Production of bite-block vowels: acoustic equivalence by selective compensation. *J. Acoust. Soc. Am.* 69, 802-10.

[8] Goldstein, L., Byrd, D., Saltzman, E. 2006. The role of vocal tract gestural action units in understanding the evolution of phonology. In: [1], 215-249.

[9] Grothe, B. 2003. New roles for synaptic inhibition in sound localization. *Nat. Rev. Neurosci*. 4, 540-50.

[10] Guenther, F.H. 2006. Cortical interactions underlying the production of speech sounds. *J. Comm. Dis.* 39, 350-365.

[11] Halle, M. 2002. *From Memory to Speech and Back.* Berlin: Mouton de Gruyter.

[12] Jakobson, R., Fant, G., Halle, M. 1951. *Preliminaries to Speech Analysis*. Cambridge MA: MIT Press.

[13] Kazanina, N., Phillips, C., Idsardi, W.J. 2006. The influence of meaning on the perception of speech sounds. *Proc. Nat. Acad. Sci.* 103, 11381-6.

[14] Lane, H, Perkell, J.S. 2006. Control of voice-onset time in the absence of hearing: a review. *J. Speech Lang. Hear. Res.* 48, 1334-43

[15] Liberman, A. 1996. *Speech: A Special Code.* Cambridge MA: MIT Press.

[16] McFarland DH, Baum SR. 1995. Incomplete compensation to articulatory perturbation. *J. Acoust. Soc. Am.* 97, 1865-73.

[17] Marr, D. 1982. *Vision*. San Francisco: Freeman.

[18] Ménard, L., Schwartz, J.-L., Boë, L.-J., Aubin, J. 2007. Articulatory–acoustic relationships during vocal tract growth for French vowels: Analysis of real data and simulations with an articulatory model. *J. Phon.* 35, 1-19.

[19] Monahan, P., Idsardi, W.J. 2007. Submitted.

[20] Nieto-Castanon, A., Guenther, F.H., Perkell, J.S., Curtin, H. 2005. A modeling investigation of articulatory variability and acoustic stability during American English /r/ production. *J. Acoust. Soc. Am.* 117, 3196-3212.

[21] Obleser, J., Lahiri, A., Eulitz, C. 2004. Magnetic brain response mirrors extraction of phonological features from spoken vowels. *J. Cog. Neurosci*. 16, 31-9.

[22] Perkell, J.S., Matthies, M.L., Svirsky, M.A., Jordan, M. I. 1995. Goal-based speech motor control: A theoretical framework and some preliminary data. *J. Phon.* 23, 23-35.

[23] Peterson, G.E., Barney, H.L. 1952. Control methods used in a study of the vowels. *J. Acoust. Soc. Am.* 24, 175-184.

[24] Phillips, C., Pellathy, T., Marantz, A., Yellin, E., Wexler, K., Poeppel, D., McGinnis, M., Roberts, T. 2001. Auditory cortex accesses phonological categories: an MEG mismatch study. *J. Cog. Neurosci.* 12, 1038-55.

[25] Poeppel, D., Idsardi, W.J., van Wassenhove, V. In press. Speech perception at the interface of neurobiology and linguistics. *Proc. of the Royal Society of London (B).*

[26] Poeppel, D., Phillips, C., Yellin, E., Rowley, H.A., Roberts, T.P., Marantz, A. 1997. Processing of vowels in supratemporal auditory cortex. *Neurosci. Lett.* 221, 145-8.

[27] Pruthi, T. 2007. Analysis, vocal-tract modeling and automatic detection of vowel nasalization. Ph.D. thesis, U. Maryland.

[28] Roberts, T.P., Flagg, E.J., Gage, N.M. 2004. Vowel categorization induces departure of M100 latency from acoustic prediction. *Neuroreport* 15, 1679-82.

[29] Skipper, J.I., Nusbaum, H.C., Small, S.L. 2006. Lending a helping hand to hearing: another motor theory of speech perception. In: [1], 250-285.

[30] Soeta, Y., Nakagawa, S., Tonoike, M. 2005. Auditory evoked magnetic fields in relation to bandwidth variations of bandpass noise. *Hear. Res.* 202, 47-54.

[31] Soeta Y, Nakagawa S, Matsuoka K. 2006. The effect of center frequency and bandwidth on the auditory evoked magnetic field. *Hear. Res.* 218, 64-71.

[32] Stevens, K.N., Halle, M. 1967. Remarks on analysis by synthensis and distinctive features. In: Wathen-Dunn, W. (ed) *Models for the Perception of Speech and Visual Form.* Cambridge MA: MIT Press, 88-102.

[33] Stevens, K.N. 1989. On the quantal nature of speech. *J. Phon.* 17, 3-46.

[34] Stevens, K.N. 1999. Articulatory-acoustic-auditory relationships. In: Hardcastle, W., Laver, J. (eds), *The Handbook of Phonetic Science*. Oxford: Blackwell, 462-506.

[35] Stevens, K.N. 2002. Toward a model for lexical access based on acoustic landmarks and distinctive features. *J. Acoust. Soc. Am.* 111, 1872-91.

[36] Young, D. 1989. *Nerve Cells and Animal Behaviour.* Cambridge: Cambridge University Press.