Neuromagnetic evidence for a featural distinction of English consonants: Sensor- and source-space data

Mathias Scharinger\textsuperscript{a}, Jennifer Merickel\textsuperscript{b}, Joshua Riley\textsuperscript{a}, William J. Idsardi\textsuperscript{a}

\textsuperscript{a}Department of Linguistics, University of Maryland, College Park, MD, USA

\textsuperscript{b}Department of Linguistics, University of Rochester, Rochester, NY, USA

Corresponding author:

Dr. Mathias Scharinger
Department of Linguistics
University of Maryland
College Park, MD 20742-7505
Tel.: +1-(301) 405-3127
Fax: +1-(301) 405-7104
Email: mts@umd.edu
ABSTRACT

Speech sounds can be classified on the basis of their underlying articulators or on the basis of the acoustic characteristics resulting from particular articulatory positions. Research in speech perception suggests that distinctive features are based on both articulatory and acoustic information. In recent years, neuroelectric and neuromagnetic investigations provided evidence for the brain’s early sensitivity to distinctive feature, particularly for place of articulation distinctions. Here, we compare English consonants in a Mismatch Field design across two broad and distinct places of articulation - LABIAL and CORONAL - and provide further evidence that early evoked auditory responses are sensitive to these features. We further add to the findings of asymmetric consonant processing. Labial glides (Experiment 1) and fricatives (Experiment 2) elicited larger Mismatch responses than their coronal counterparts. Further, their M100 dipoles differed along the anterior/posterior dimension in the auditory cortex that has previously been found to spatially reflect place of articulation differences. Intriguingly, consonants produced at more “frontal” positions in the vocal tract showed dipoles located at more anterior positions in auditory cortex. Conversely, consonants produced further back in the vocal tract resulted in more posterior dipoles. The results are discussed with respect to acoustic and articulatory bases of featural speech sound classifications and featural underspecification.

KEY WORDS

Auditory Cortex, Magnetoencephalography, Auditory Evoked Fields, Mismatch Fields, Place of Articulation, Asymmetric Speech Processing, Underspecification
Introduction

Speech sounds can be classified on the basis of their active articulators, pertaining to the aspects of how and where they are produced in the vocal tract. In this respect, manner of articulation (MoA) describes the way by which speech sounds are produced, i.e. with or without constriction in the oral cavity, while place of articulation (PoA) indicates where such a constriction occurs (cf. Ladefoged, 2001; Reetz & Jongman, 2008; Stevens, 1998). For instance, speech sounds produced at or in the vicinity of the lips are referred to as labial sounds (e.g. English [v], [w]), while sounds produced between the alveolar ridge and the palatum are labeled palato-alveolar, palatal, or more generally, coronal (e.g. English [ʒ], [j]). The differences between [v] and [w], and [ʒ] and [j] relate to differences between fricatives with a narrow constriction ([v], [ʒ]) and glides with a vowel-like configuration ([w], [j]). Naturally, these articulator configurations have particular acoustic consequences and characteristics. Together with specific articulator configurations, acoustic cues (“landmarks”, cf. Stevens & Blumstein, 1978; Stevens, 2002) have been taken to describe and classify sets of phonetic and phonological features. Not surprisingly, feature theories differ as to whether the corresponding features are predominantly determined by their underlying articulators (Chomsky & Halle, 1968; Halle, 1983; Jakobson, Fant, & Halle, 1952) or by their concomitant acoustic properties (e.g. Stevens, 2002). The Featurally Underspecified Lexicon model (FUL, Lahiri & Reetz, 2002) employs abstract features which may be related to either articulatory or acoustic bases, but not in a necessarily isomorphic way. For instance, the PoA features labial, coronal and dorsal stem from articulatory positions (lips, tongue corona, and tongue back). However, their usage for both vowels and consonants implies a more abstract and articulator-independent classification. Yet, PoA feature combinations for vowels and consonants differ: A consonant cannot be labial and coronal or labial and
DORSAL at the same time. In contrast, vowels can be LABIAL and simultaneously either CORONAL or DORSAL, since vocalic labiality rather expresses a secondary place of articulation. Note that these assumptions bear on articulatory and acoustic bases: The three-way PoA distinction for consonants reflects their relative positioning in the oral tract, where labial precedes coronal, and coronal precedes dorsal. Due to the complex acoustic structure of labial consonants (Stevens, 1998), consonantal PoA cannot be related to one acoustic cue (such as the second formant frequency, F2). The two-way PoA for vowels, distinguishing between coronal (front) and dorsal (back), reflects the positional involvement of the tongue. The primary acoustic cue is F2. Labiality as secondary PoA indicates the involvement of the lips. Again, it cannot be attributed a single acoustic cue (such as F1 or F2). Thus, PoA distinctions in FUL are based on abstract features. These features have a relation to their articulatory and acoustic bases, but it is not isomorphic or linear (cf. Stevens, 1989). Finally, FUL assumes coronal underspecification. Coronal underspecification describes the assumption that the long-term representation of coronal consonants and vowels has no underlying PoA feature (Lahiri & Reetz, 2002, 2010). Amongst other phenomena, coronal underspecification accounts for asymmetric assimilation effects, whereby underspecified coronal [n] in *lean bacon* assimilates to [m] due to the following labial consonant [b], but specified labial [m] in *cream dessert* does not assimilate to [n]. Although coronals are underlyingly underspecified, their acoustic-phonetic surface forms naturally provide cues for their PoA.

In this article, we are interested in neuromagnetic responses to the perception of consonantal PoA distinctions LABIAL and CORONAL as formulated in the FUL framework. The opposition of labial and coronal is especially interesting, since the two features cannot be teased apart by a single acoustic parameter, and since coronal is claimed to be unspecified in the long-term representation of consonants and vowels. From an articulatory perspective, labial precedes
coronal in the vocal tract and clearly bears more visual salience. We also chose to look at
different consonantal MoA for the labial-coronal contrast. If the co-occurrence distinction
between consonants and vowels is categorical, we expect to find very similar results for the PoA
oppositions for different consonantal MoA. As for methodology, we opted for the Mismatch
Negativity paradigm (or its magnetic equivalent, the Mismatch Field, MMF; Näätänen, 2001;
Näätänen & Alho, 1997; Näätänen, et al., 1997), a technique previously used to test the mapping
of acoustic-phonetic surface features to underlying language specific long-term representations
(Eulitz & Lahiri, 2004). A crucial advantage of Magnetoencephalography (MEG) is that the
recording also allows for source estimations of the underlying brain activity, a measure
previously used for testing the neuronal basis of PoA distinctions (Obleser, Lahiri, & Eulitz,

**Testing features in the brain**

Recent behavioral and neurophysiological studies support the proposal of abstract features as
units of perception and representation in long-term memory (Lahiri & Reetz, 2002, 2010;
Poeppel, Idsardi, & van Wassenhove, 2008; Reetz, 1998, 2000; Stevens, 2005). Coronal
underspecification has been successfully tested in both consonants (Friedrich, Lahiri, & Eulitz,
2008; Friedrich, 2005; Ghini, 2001; Lahiri & Marslen-Wilson, 1991; Wheeldon & Waksler,
2004) and vowels (Eulitz & Lahiri, 2004; Lahiri & Reetz, 2002, 2010). MMN/MMF designs for
the investigation of coronal underspecification rely on the creation of a central sound
representation through the repetitious auditory presentation of the standard. This representation is
assumed to reflect language-specific long-term memory representations of single segments
(Näätänen, et al., 1997), affixes (Shtyrov & Pulvermüller, 2002a), and words (Pulvermüller, et
al., 2001; Pulvermüller, Shtyrov, Kujala, & Näätänen, 2004; Shtyrov & Pulvermüller, 2002b). The infrequently occurring deviant, on the other hand, provides a more surface-faithful representation, involving an acoustic and feature-based mismatch to the standard representation (Eulitz & Lahiri, 2004; Winkler, et al., 1999). While the MMN as an automatic change detection of the brain is elicited whenever the deviant acoustically differs from the standard, the MMN is additionally modulated by the featural properties of standard and deviant. In Eulitz & Lahiri (2004), the coronal deviant vowel [ø] preceded by the dorsal standard vowel [o] elicited an earlier and larger MMN response than the dorsal deviant [o] preceded by the coronal vowel [ø]. Eulitz & Lahiri assumed that in the first case, the dorsal standard [o] activated its fully specified PoA representation. The coronal deviant supplied the mismatching feature, resulting in a larger MMN than in the second case, where the coronal standard activated an underspecified representation, for which the dorsal deviant did not supply a mismatching feature. This principle should hold in general: Coronal deviants differing solely along the PoA dimension from their standards should always elicit larger MMN effects compared to their non-coronal counterparts.

While there is evidence that the MMN is also modulated by lexical effects (Menning, et al., 2005; Pulvermüller & Shtyrov, 2006) – thus lending support to the assumption that coronal underspecification is indeed a lexical principle – earlier evoked auditory components are less readily usable to test underspecification. Nevertheless, these components (N100m/M100 and M50, cf. Gage, Poeppel, Roberts, & Hickok, 1998; Poeppel, et al., 1996; Roberts, Ferrari, & Poeppel, 1998; Roberts & Poeppel, 1996; Tavabi, Obleser, Dobel, & Pantev, 2007) were used in neurophysiological investigations of acoustic-phonetic and phonological features (Obleser, Elbert, Lahiri, & Eulitz, 2003; Obleser, Lahiri, et al., 2003, 2004; Tavabi, et al., 2007). We focus on the M100, which is a component of the evoked magnetic field response, with latencies at around 100 ms post stimulus onset. It is considered an indicator of basic speech processing in the
auditory cortices (Diesch, Eulitz, Hampson, & Ross, 1996; Shestakova, Brattico, Soloviev, Klucharev, & Huotilainen, 2004) and reliably elicited by tones, vowels, syllables, and word onsets. The latency and amplitude of the M100 has been shown to correlate with the fundamental frequency of tones (Roberts & Poeppel, 1996; Woods, Alain, Covarrubias, & Zaidel, 1995) and the first formant frequency of vowels (Poeppel & Marantz, 2000; Roberts, et al., 1998). Recent work suggests a more specific sensitivity to vowel formants ratios in dense vowel spaces (Monahan & Idsardi, in press). Larger ratios between the first and third vowel formant (F1/F3) elicited earlier M100 latencies. With respect to PoA, Obleser and colleagues investigated the equivalent current dipoles (ECDs) underlying the M100 response to vowels, consonants, and consonant-vowel syllables. Obleser, Lahiri, et al., 2004, tested 7 naturally spoken German vowels and found that dorsal vowels elicited ECDs with a more posterior location in the auditory cortex than coronal vowels. Intriguingly, the spatial differences only held for the coronal/dorsal, not for the labial-coronal/coronal comparison. This suggests that the ECD separation is particularly sensitive to mutually exclusive PoA features (such as coronal and dorsal), mapping to an acoustic difference in the second formant (F2). The ECD distinction in the anterior-posterior dimension for PoA differences was replicated with CV syllables containing either a dorsal ([o]) or a coronal vowel ([ø]), irrespective of the initial consonant, which was a labial, coronal, or dorsal stop ([b], [d], [g], Obleser, Lahiri, et al., 2003). Consonantal differences were reflected in ECD source orientations, again distinguishing between coronal and dorsal. Due to potentially mutual influences between the consonant and vowel in this study, coronal ([t]) and dorsal ([k]) stop consonants were separately analyzed as single segments. In order to demonstrate that the previous ECD location pattern did not solely result from acoustic spectro-temporal characteristics, Obleser, Scott, & Eulitz, 2006 used dorsal and coronal stops in an intelligible and in an unintelligible condition. The unintelligible condition was matched to the intelligible
condition for spectro-temporal complexity and diversity. Intriguingly, dorsal stops elicited more posterior dipoles, in accordance with previous findings, but only in the intelligible condition. Further, the spatial dipole pattern was not affected by the co-occurring voicing feature, distinguishing between [t,k] and [d,g]. In line with the observation that the M100 can be modulated by top-down processes (Näätänen & Picton, 1987; Näätänen & Winkler, 1999; Sanders, Newport, & Neville, 2002), Obleser, Elbert, & Eulitz, 2004 showed that the absolute ECD dipole locations differed when attention was shifted away from the linguistic stimulus characteristics, while ECDs to dorsal vowels still were significantly more posterior than ECDs to coronal vowels. Thus, there is good evidence that the spatial alignment of the evoked auditory M100 follows featurally PoA distinctions, especially if these are mutually exclusive, which is true for DORSAL and CORONAL for both vowels and consonants. Further, it could be the case that the specific spatial arrangement of dorsal and coronal dipoles actually parallels relative articulatory positions. Independent evidence for such an assumption comes from studies showing an articulatory-based alignment of centers of activation in the somatosensory cortex (Picard & Olivier, 1983; Tanriverdi, Al-Jehani, Poulin, & Olivier, 2009). The relative positions of activation on the sensory strip appear to parallel the classic place of articulation distinctions (Ladefoged, 2001), starting at the lips and ending at the tongue root. Given the long-debated acoustic invariance problem for speech perception (Blumstein & Stevens, 1981; Lahiri, Gewirth, & Blumstein, 1984; Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967), relative articulatory positions appear to be a more robust top-down cue for phoneme distinctions and directly relate to the quantal theory of speech perception and production (cf. Discussion; Stevens, 1989).
Place of articulation differences in English consonants

On the basis of previous findings regarding featural speech sound representations, we are interested in the PoA distinction between labial and coronal consonants in American English. This distinction is interesting because it refers to mutually exclusive PoAs for consonants, but not for vowels, following the assumption of FUL (Lahiri & Reetz, 2002). With respect to the two auditory evoked components – MMF and M100 – we expect coronal and labial consonants to yield asymmetries in the MMF and separable ECD dipole locations underlying the M100. In particular, if the relative articulator hypothesis is correct, we expect the consonantal labial dipoles to be more anterior than the coronal dipoles, since labial consonants are produced further “front” in the mouth than coronals.

We opted for labial and coronal glides ([w], [j]; Experiment 1) and fricatives ([v], [ʒ]; Experiment 2), for a couple of reasons. First, comparing PoA differences across different MoAs enables us to test whether PoA effects are in fact independent of MoA within the consonants. Second, our selection lets us compare more vowel-like consonants (glides) with less vowel-like consonants (fricatives). If the co-occurrence restriction of LABIAL and either DORSAL and CORONAL is categorical, we expect to find effects comparable to previous consonant studies, but not to vowel studies. Finally, LABIAL and CORONAL oppositions in fricatives have not been systematically tested with neuromagnetic methods before. Table 1 illustrates the experimental material for Experiment 1 and 2.

Table 1 approximately here
The respective consonants were embedded in V_V contexts, using the vowel [a] in all cases. We opted for naturally spoken /VCV/ stimuli in order to ensure that the stimuli resembled natural speech as closely as possible. Further, since phonetic research has shown that the initial vowel-consonant transition is shaped by the final vowel in VCV syllables (Oehman, 1966), we made sure that the same vowel preceded and followed the consonant of interest in our stimuli. Thus, a constant vowel environment ensures that potential co-articulatory effects on the initial vowel can be attributed to the medial consonant of interest, and that differences between responses to initial vowels are the result of the subsequent consonant rather than the categorical identity of either of the vowels.

**Experiment 1**

**Material**

For Experiment 1, we chose the labio-velar and palatal glides [w] and [j], and matched the PoA to the labial and palato-alveolar fricatives [v] and [ʒ] in Experiment 2. We assign the labio-velars and labio-dentals to the macro-category LABIAL, and likewise, the palatals and palatal-alveolars to CORONAL.

The consonants of interest were embedded in VCV sequences and spoken by a phonetically trained native speaker of English, recorded on a computer with a sampling rate of 44.1 kHz and an amplitude resolution of 16 bits, and further processed in PRAAT (Boersma & Weenink, 2009). Care was taken that the phonetically trained speaker avoided a velar constriction for [w].

We used 20 different exemplars of each sequence, in order to obtain stimulus material with substantial natural variation (cf. Phillips, et al., 2000). From these 20 exemplars, we chose 10 for
Each stimulus sequence with pitch, loudness, and intensity characteristics in a relative narrow, but nevertheless variable range.

In order to assess the co-articulatory influences of the respective medial consonants onto the first vowel [a] of each stimulus, we measured formant frequencies (F1, F2) at the approximate middle of the initial vowels (at 100 ms). Figure 1 illustrates the location of the initial vowels in the F2/F1 vowel space.

Figure 1 approximately here

Statistically, there were no effects for the first vowel formant, F1 (all Fs < 3, n.s., based on Linear Mixed Model analyses, cf. Baayen, 2008; Baayen, Davidson, & Bates, 2008). However, the second vowel formant, F2, differed across place ($F(1,36) = 121.17, p < .001$) and manner of articulation ($F(1,36) = 13.87, p < .001$) of the preceding consonant. Coronality of the consonant led to a fronting of the initial vowel, i.e. to an increase in its F2 value. Furthermore, there was a significant interaction of place and manner ($F(1,36) = 10.21, p < .01$), reflecting stronger fronting by coronal glides than by coronal fricatives.

Behavioral task

The stimulus material was tested in a same-different AX discrimination task in which subjects heard every possible opposition of the categorical contrasts (3 x 4) for each exemplar (10) in two repetitions, yielding a total of 240 pairs. 12 native speaker of American English (6 female; mean age 23) participated in this task, but neither in Experiment 1 nor in Experiment 2.
Subjects’ accuracy as measured in $d'$ (Macmillan & Creelman, 2005) showed acceptable values in all conditions ($d'>1$), implying that there were no a priori psycho-acoustic differences between the stimulus categories.

**Stimulus organization and presentation**

Stimuli were organized in a classic odd-ball paradigm. In Experiment 1, the VCV sequences [aja] and [awa] were distributed over two blocks in which they occurred in either standard (p=0.875, N=700) or deviant position (p=0.125, N=100). The number of standards between two deviants varied randomly in each block and for each subject. The total number of stimulus presentations per block was 800. The inter-stimulus interval varied randomly between 1200 and 1400 ms. Each block lasted for about 20 minutes. The stimuli presentation was done using the software package PRESENTATION (Neurobehavioral Systems).

**Subjects and procedure**

14 participants (9 males, mean age 23.6, SD 6.4), graduate and undergraduate students of the University of Maryland, College Park without hearing or neurological impairments participated for class credit or monetary compensation. They were all native speakers of American English and had not participated in the behavioral same-different task. Participants gave their informed consent and were tested for their handedness, using the Edinburgh Handedness Inventory (Oldfield, 1971). The cut-off (exclusion) criterion was 80%, but no subject was excluded for this reason, all being strongly enough right-handed. Prior to the main experiment, participants’ head shapes were digitized, using a POLHEMUS 3 Space Fast Track system. Together with
localization data as measured from two pre-auricular, and three pre-frontal electrodes, these data allowed us to perform dipole localization analyses, as reported in the Analysis section.

**MEG recording**

For the MEG recording, participants lay supine in a magnetically shielded chamber with their heads in a whole-head device of 157 axial gradiometers (Kanazawa Institute of Technology, Kanazawa, Japan). Magnetic field recording were done with a sampling rate of 500 Hz, a low-pass filter of 200 Hz, and a notch filter of 60 Hz. Before the main experiment, participants were screened with a 2-tone perception task during which they were instructed to silently count a total of 300 high (1000 Hz) and low (250 Hz) sinusoidal tones, occurring in (pseudo)-random succession. The scalp distribution of the resulting averaged evoked M100 field was consistent with the typical M100 source in the supra-temporal auditory cortex (Diesch et al., 1996). Only participants with a reliable bilateral M100 response were included in further analyses. One participant was excluded on this criterion. Based on the topography of the M100, we selected 10 channels per participant in each hemisphere (5 source, 5 sink). The selection was based on the strongest average field of the M100 peak.

The main experiment consisted of a passive oddball paradigm as described above. Participants were tested in two blocks that differed as to whether [aja] or [awa] occurred in standard or deviant position. If [aja] was standard in block 1, it was deviant in block 2. The block order was counter-balanced across subjects.

In order to reduce eye movements and to keep participants awake, they were presented a silent movie during the passive listening task, which was projected onto a screen approximately
15 cm in front of them. Furthermore, the experimenter guaranteed a short break every 10 minutes. The total experiment lasted for about an hour.

**Data Analysis**

MEG raw data were filtered with respect to environmental and scanner noise (de Cheveigné & Simon, 2007, 2008). Trial epochs with a length of 800 ms (100 ms pre-stimulus interval) were averaged for each condition, using the MegLaboratory Software (Kanazawa Institute of Technology, Japan). Artifact rejection was done by visual inspection. The cut-off criteria were amplitudes higher than 3 pico-Tesla (pT) or more than 3 consecutive eye blinks within one epoch. Due to excessive noise and artifacts in the raw data, one participant was excluded from further analyses, since more than 15% of standard and deviant epochs had to be excluded. Averaged data were base-line corrected and band-pass filtered (0.03-30 Hz). Amplitudes and peak latencies were calculated on the basis of pre-defined time frames, using the selected left- and right-hemispheric channels from the 2-tone screening. The time frames were selected based on visual inspection of the grand averaged data across all subjects. These time windows were in accordance with the usual time course of MEG components (Ackermann, Hertrich, Mathiak, & Lutzenberger, 2001). The first window covered the prominent M100 peaks of the initial vowel (between and 75 and 175 ms). The second window (225 – 325 ms) included the M100 of the consonant, while the third window covered the area of the mismatch negativity (350 – 450 ms, ca. 150-250 ms after consonant onset).

The time course of these windows is illustrated in Figure 2.
Mixed Model analyses with subject as random factor (MM, Baayen, 2008; Baayen et al., 2008) were calculated with mean magnetic field strengths in the four time windows as independent variable based on the RMS values for the selected left- and right-hemispheric channels.

MMs comprised the within-subject factors position (standard/deviant), word (aja/awa) and hemisphere (left/right). Due to the few-to-many design of an oddball paradigm, we randomly selected the same amount of standard epochs than deviant epochs for this analysis, in order to guarantee that the RMS values did not involve unequal variance within the factor position. In the RMS analysis across our four time windows, any difference in mismatch responses between the two stimuli sequences should produce significant interactions of word x position. We used the same factors for the analyses with RMS peak latency and peak amplitude as dependent variables.

**Dipole fitting**

Equivalent current dipole (ECD) fitting was based on 32 channels surrounding the selected 10 channels from the 2-tone screening and followed the procedure of Obleser et al. (2004). ECD modeling used a sphere model that was fitted to the head shape of each participant. Left and right hemispheric dipoles were modeled separately (Sarvas, 1987). Source parameters of the vowel and consonant M100 as well as for the deviants’ mismatch field were determined based on the median of the five best ECD solutions on the rising slope of each of the components, covering a time range of 10 ms. No solution after the respective peaks was included (cf. Scherg, Vajsar, & Picton, 1990). Only solutions with a fit better than 90% were used for the fitting calculation.
For the MM analyses, we used the dependent measures dipole intensity, orientation (in the sagittal and axial dimension) and location in the medial-lateral, anterior-posterior, and inferior-superior dimension. Within-subject independent factors comprised word and position.

Results
Generally, effects were larger in the left than in the right hemisphere. MMFs as significant differences between standard and deviant responses were robust between 350 and 450 ms ([aja]: $t = 3.69, p < 0.01$; [awa]: $t = 12.98, p < 0.001$). In the following, we report significant main effects and interactions involving the factors position and word for all three time frames.

75-175 ms:
No amplitude effects were found in this time frame. The latency analysis revealed an interaction of word x position ($F(1,932) = 13.58, p < .001$), differing across hemispheres, as seen in the three-way interaction hemisphere x word x position ($F(1,932) = 4.03, p < .05$). Based on RMS peaks, the standards [aja] and [awa] peaked at the same time, while the deviant [awa] elicited a significantly earlier M100 ($t = 3.65, p < .01$).

225-325 ms:
RMS amplitudes differed between standards and deviants, as seen in the main effect of position ($F(1,77) = 16.21, p < .001$). This effect was replicated in the peak amplitude analysis ($F(1,74) = 8.14, p < .01$). The latencies of the RMS peaks differed between position and hemispheres
(F(1,74) = 6.13, p < .05), reflecting left-hemispheric earlier deviant latencies, and right-hemispheric earlier standard latencies.

350-450 ms:

The RMS analysis revealed a significant interaction of word x position (F(1,77) = 3.90, p < .05). The difference between standard and deviant was greater for [awa] than for [aja]. Latencies did not differ.

Figure 3 approximately here

Dipoles

Consonant and vowel M100 dipole parameters were analyzed in MMs, separately for each hemisphere. In the left hemisphere, the intensity analysis showed a trend for a word effect in the same direction as seen in the amplitude analyses. The deviant [awa] produced a stronger mismatch response than the deviant [aja] (F(1,7) = 3.82, p = .11). For the M100 consonant dipole, intensities differed left-hemispherically between standard and deviant (F(1,24) = 11.28, p < .01), and for the M100 vowel dipole, intensities differed on the left (F(1,29) = 23.40, p < .001) as well as on the right hemisphere (F(1,10) = 9.10, p < .05). The dipole latency analysis showed a position effect for the consonant M100 only (F(1,24) = 7.26, p < .01). Deviant responses peaked earlier than standard responses.
M100 consonant dipole orientations differed in the left-hemispheric sagittal plane depending on position (standard, deviant) and word ([aja], [awa]), as seen in a significant interaction (position x word: $F(1,24) = 7.02, p < .05$). This was driven by the standard [awa] that showed a more vertical orientation than the other stimuli. Further, the M100 vowel dipole orientation differed across position ($F(1,29) = 11.52, p < .01$). Standards were oriented more horizontally, pointing anterior. Finally, dipole locations differed in the anterior-posterior dimension of the left hemisphere. There was a trend of an interaction of position x word for the M100 consonant dipole ($F(1,24) = 2.67, p = .11$), driven by a more anterior location of the deviant [awa], compared to the deviant [aja] ($t = 3.98, p < .05$). This interaction was significant for the M100 vowel dipole ($F(1,29) = 6.84, p < .05$), reflecting again a more anterior location of the deviant [awa].

**Discussion**

The MMF results of Experiment 1 with the English glides [w] and [j] yielded an interesting amplitude asymmetry, but in the opposite direction than predicted by the FUL model. Crucially, while FUL would predict that underspecified [j] elicits a larger MMF, we found that in fact, specified [w] elicited larger MMF amplitudes, while latencies did not differ. There could be at least two reasons for this outcome. First, in choosing VCV stimuli without lexical representations, the MMF pattern may just reflect a form-based, but not a lexical property of the corresponding glide. The assumption of underspecification, on the other hand, is clearly a lexical principle, such that the technique employed here may have been not sensitive enough. This holds even more so since it appears that there are qualitative differences in the mismatching consonant features. In the direction from coronal standards to labial deviants, the deviant feature [labial]
was more salient than in the reverse case, i.e. in the direction from labial standards to coronal deviants. This salience refers to the acoustic complexity of labial sounds, but at the same time to the underlying articulators (the lips) that are ‘more front’ in the vocal tract. The particular salience of labials relates to its stronger visual cues (this is a prominent topic, particularly in the realm of language acquisition, cf. among others Boysson-Bardies & Vihman, 1991; Dorman & Loizou, 1996; MacNeilage, Davis, & Matyear, 1997; Vihman, 1986; Winters, 1999, see General Discussion).

With respect to the M100 dipoles, the results of Experiment 1 are in line with the findings of Obleser and colleagues (Obleser, Elbert, et al., 2004; Obleser, Lahiri, et al., 2003, 2004; Obleser, et al., 2006). As in their studies, dipole localizations in Experiment 1 differed along the anterior-posterior axis. Deviants with labial [w] produced cortical activation with a more anterior dipole than deviant segments with coronal [j]. In addition, the consonant dipole orientation for [awa] was more vertical. Note that the particular alignment of centers of activation (together with the previous findings) can be interpreted as paralleling somatosensory representations. Places of articulation involving the lips (hence, being further “front” in the vocal tract) resulted in more frontal ECDs, compared to places of articulation involving the coronal area of the tongue that had more posterior ECDs. While we are aware that we would need a third point on this anterior-posterior line to backup our claim, the relative positions of the dipoles in comparison with the Obleser data are suggestive in that relative and broad articulatory locations (undoubtedly with very characteristic acoustic consequences) are paralleled in cortical areas.

The reason why we found the dipole location effects only for deviants is likely to be rooted in the predictability of the standards. Most instances of standard presentations were encountered in circumstances under which that particular standard could be predicted. Thus, it is not
surprising that this would be accompanied by a substantial suppression in evoked activity, thereby weakening the overall levels for the standard cases, and possibly also blurring the dipole distances.

Taken together, the results of Experiment 1 provide evidence for a featural PoA sensitivity of early evoked auditory responses on the basis of acoustic, but also on the basis of relational articulatory information. If the distinction between labial and coronal consonants is based on robust acoustic and articulatory cues and in fact a salient distinction during early stages of speech perception, we expect to find a similar pattern of responses with consonants of a different MoA. For that reason, we used the fricatives [v] and [ʒ] in Experiment 2 with the same design and setup as in Experiment 1. Note that if the source- and sensor-space patterns of Experiment 1 were driven solely by acoustic or articulatory properties of our stimuli, we would expect to find a different outcome in Experiment 2. PoA distinctions of fricatives are less based on formant values, but rather on spectral peak locations and spectral moments (Jongman, Wayland, & Wong, 2000). This can be seen by the coarticulatory influences of the fricatives onto their preceding vowels of our stimuli. These influences were more pronounced for the glides [w] and [j] that have a more vowel-like formant structure, than for the fricatives [v] and [ʒ] without such structure. Note also that on a narrower articulatory classification, [w] and [v] and [j] and [ʒ] slightly differ from each other with regard to PoA, while their relations within the oral tract (front vs. less front) are still preserved.
**Experiment 2**

*Material, subjects, and procedure*

In Experiment 2, the VCV sequences [ava] and [aʒa] were distributed over standard and deviant positions as described in Experiment 1.

15 participants (7 female, mean age 21.4, SD 2.6), graduate and undergraduate students of the University of Maryland in College Park, without hearing or neurological impairments, participated for class credit or monetary compensation. They were all native speakers of American English and had neither participated in Experiment 1 nor in the behavioral same-different task. Participants gave their informed consent and were tested for their handedness, using the Edinburgh Handedness Inventory (Oldfield, 1971). The cut-off (exclusion) criterion was at 80%, but all participants performed above this threshold. Two subjects were excluded from further analyses, since their 2-tones screening test did not allow for a reliable M100 localization. They additionally had excessive proportions of artifacts in the denoised data.

The experimental procedure, MEG recordings, data analyses, and dipole fittings were identical to Experiment 1.

**Results**

The analysis windows for the amplitude and latency data corresponded to those of Experiment 1 (cf. Figure 2). Again, effects were stronger in the left than in the right hemisphere.

Dipole analyses were based on the M100 of the vowel and the consonant. MMFs as significant differences between deviants and standards of the same VCV sequence were reliable
between 350 and 450 ms post stimulus onset ([ava]: $t = 32.26, p < 0.001$; [aŋa]: $t = 10.61, p < 0.01$). This window was used for the MMF analysis.

75-175 ms

There were no RMS amplitude or latency effects in this time window.

225-325 ms

The RMS amplitude analysis indicated higher deviant than standard amplitudes ($F(1,84) = 16.35, p < .001$), while the RMS latencies did not differ. The RMS amplitude effect was replicated in the peak amplitude analysis ($position: F(1,78) = 16.00, p < .001$).

350-450 ms

The RMS amplitude analysis replicated the mean amplitude findings. Crucially, [ava] produced higher amplitudes ($F(1,84) = 5.40, p < .05$), which was restricted to the left hemisphere ($F(1,84) = 5.52, p < .05$). Further, [ava] produced a greater mismatch than [aŋa] ($word \times position: F(1,84) = 4.58, p < .05$), and more so in the left than the right hemisphere. Peak amplitudes showed a similar pattern, except that the interaction $word \times position$ just failed to reach significance ($F(1,83) = 3.09, p = 0.08$). There were no RMS latency effects (cf. Figure 4).
Dipoles

Dipole parameters were investigated in MMs separately for each time frame and hemisphere. Intensities differed between standards and deviants left- \( (F(1,31) = 12.84, p < .01) \) and right-hemispheric \( (F(1,18) = 23.59, p < .01) \) for the consonant M100, and left-hemispheric only for the vowel M100 \( (F(1,25) = 13.51, p < .01) \).

Word differences were seen as trends and corresponded to the analyses in sensor space, yet were not significant. Deviants led to earlier consonantal dipole latencies left \( (F(1,31) = 12.50, p < .01) \) and right-hemispheric \( (F(1,18) = 4.97, p < .05) \). In the anterior-posterior dimension, dipole locations showed trends for [ava] to be more anterior than [aʒa]. This was true for the consonantal M100 dipole \( (F(1,26) = 1.86, p = .18) \) and the vowel M100 dipole \( (F(1,21) = 3.38, p = .07) \) on the left hemisphere. The consonantal dipole furthermore showed that the trend of a more anterior [ava] dipole held only for the deviant position \( (t = 3.68, p = .08) \). Consonantal dipole orientations in the sagittal plane especially differed between standard and deviant in the left hemisphere \( (F(1,26) = 13.08, p < .01) \). Compared to standard dipoles, deviant dipoles were oriented less vertical, pointing more into the anterior direction.

Combined analyses

In order to assess the place of articulation asymmetry across Experiment 1 and 2, we performed a combined RMS amplitude and dipole location analysis. The factor *place* in this analysis...
comprised the two values coronal (for the stimuli [aja] and [aʒa]) and labial (for [awa] and [ava]). Additionally, we used the between-experiment factor manner (glide, fricative).

The amplitude analysis showed an interaction of place x position for the MMF ($F(1,160) = 8.34, p < .01$), yielding higher amplitudes for labial deviants. There were no significant effects or interactions with manner (all $Fs < 1$, n.s.).

The left-hemispheric dipole location analysis showed a significant interaction of position x place for the consonantal dipole ($F(1,49) = 4.24, p < .05$). Crucially, labial deviants were located at more anterior locations compared to all other stimuli (cf. Figure 6). These results were independent of MoA, as revealed by a lack of significant main effects or interactions involving the factor manner (all $Fs < 1$, n.s.).

Interestingly, the vowel dipole showed a main effect for place ($F(1,48) = 8.20, p < .01$, cf. Figure 5). Again, the labial-coronal relation was preserved: If the vowel was followed by a labial consonant, its ECD was at more anterior locations than if it was followed by a coronal consonant. There was a trend for an interaction with position ($F(1,48) = 2.76, p = 0.09$), showing that the location difference between labials and coronals in the anterior-posterior dimension was greater for deviants than for standards. Again, this pattern of results was independent of MoA.

---

Figure 5 approximately here
Discussion

The most important result of Experiment 2 is that it replicated the sensor- and source-space pattern of Experiment 1. The labial deviant [ava] elicited larger MMF responses than the deviant [a3a]. While this outcome is again not in accordance with the predictions of FUL, we take it as supporting evidence for our alternative view according to which labiality is a more salient deviant feature and consequently results in a stronger MMF response.

In source space, ECDs for coronal [ʒ] and labial [v] differed on the anterior-posterior axis, although there was only a trend for [v] to be located at a more anterior location than [ʒ]. Combined ECD analyses of Experiment 1 and 2 showed that the difference in dipole location along the anterior/posterior axis was robust: Labials elicited dipoles with more anterior locations than coronals. This spatial location was independent of MoA: Anterior-posterior locations did not differ between [w] and [v] or [j] and [ʒ]. Thus, the combined analyses of Experiment 1 and 2 strengthen the assumption that the relative articulator positions, described by the labels labial and coronal, are paralleled in the auditory cortex: The further front the place of articulation, the more anterior the center of activity. Intriguingly, the combined analysis showed the same effect for the vowel ECD location. This suggests coarticulatory effects, whereby the acoustic cues for labiality or coronality spread onto the initial vowel (cf. Figure 1). Note that this vowel was consistently [a], such that any response differences must be attributed to the following consonants.

The repercussions of these findings and potential subsequent investigations are considered in the General Discussion.
General Discussion

The main result of our experiments is that the classification of speech sounds according to their place of articulation is reflected in early pre-attentive brain magnetic measures as well as in the spatial alignment of the corresponding centers of activities in the auditory cortex. Since we found the same pattern of results for different manners of articulation, and since the ECD separation follows broad articulatory positions within the oral tract, the brain’s sensitivity to featural long-term representations seems to be based both on acoustic and articulatory information. The PoAs referring to LABIAL and CORONAL can be considered “preferred” articulator positions within which acoustic correlates provide robust cues for feature-based speech sound distinctions. This line of argumentation follows the assumption of a “quantal nature of speech production and perception” (Stevens, 1989; Stevens & Blumstein, 1978; Stevens, 1972; Stevens, 1981). Steven’s theory provides a phonetic basis for distinctive features, resulting from non-linearities in the relation of articulator positions and the acoustic signal produced by the vocal tract. Quantal regions are considered regions in which articulator configurations produce stable acoustic resonance frequencies over a wide range of constriction positions. As suggested by our ECD findings, the auditory cortex is sensitive to these quantal regions in that dipole locations along the anterior/posterior axis differed between labial and coronal, but not between finer-grained articulatory positions. If we take previous ECD findings into consideration (Obleser, Elbert, et al., 2004; Obleser, Lahiri, et al., 2003, 2004; Obleser, et al., 2006), the relative positioning of dipoles elicited by labials, coronals, and dorsals is such that labials precede coronals, and coronals precede dorsals. Again, we are aware that this linear hypothesis has to be put to test in an experiment with all three PoAs, but we think that this account is a sensible conjecture, since the sequence labial-coronal-dorsal describes a relative order of articulatory positions commonly used for phonetic descriptions (cf. Ladefoged, 2001). Further, the alignment of lip and tongue
positions on the sensory strip in the somatosensory cortex (Picard & Olivier, 1983; Tanriverdi, et al., 2009) is reminiscent of the proposed alignment of articulatory positions in the auditory cortex. We hypothesize that the somatosensory sequence is paralleled in the auditory cortex, an assumption that is also reasonable with respect to the relative vicinity of the two cortices.

Of course, ECD locations in the auditory cortex should be mainly driven by acoustic properties of speech sounds. In this respect, previous research has established tonotopicity in the auditory cortex (cf. Diesch & Luce, 1997; Huotilainen, et al., 1995; Langner, Sams, Heil, & Schulze, 1997; Pantev, et al., 1988; Pantev, Hoke, Lutkenhoner, & Lehnertz, 1989; Tiitinen, et al., 1993). Pantev et al. (1989) showed that M100 dipoles elicited by tones with a high (subjective) pitch were located more medial as opposed to low pitch tones, which were located more lateral in the axial plane. Similar results were provided by Huotilainen et al. (1995) for magnetic responses (M100 and sustained fields) from the supratemporal plane of the auditory cortex. Langner et al. (1997) differentiated between a tonotopic gradient in the lateral-medial plane and a periodotopy in the anterior-posterior plane. Further research showed that more complex (and more speech-like) stimuli showed yielded dipole arrangements which differed in other dimensions than the tonotopic axial plane resulting from pure tones (Diesch & Luce, 2000; Mäkelä, Alku, & Tiitinen, 2003; Obleser, Elbert, et al., 2003; Obleser, Lahiri, et al., 2003, 2004; Shestakova, et al., 2002; Shestakova, et al., 2004). The results of Diesch & Luce (2000) demonstrated that vowels with high frequency formants elicited magnetic responses with dipoles located more to the anterior than dipoles of vowels with low frequency formants. This is reminiscent of the findings by Obleser and colleagues (Obleser, Lahiri, et al., 2003, 2004), where dorsal – which was found to elicit more anterior ECDs – corresponds to high F2 values. Interestingly, Mäkelä et al. (2003) could show that labial vowels led to more anterior dipole localizations than non-labial vowels in their experiment. Acoustically, their vowels primarily differed in the dimension of the first
formant, which suggests that the acoustic distinction between labial and coronal might based on the ratio of F2 and F1. Note that the lower frequency bands may provide cues for PoA distinctions even in fricatives, where the most salient cues are spectral peak locations and spectral moments in higher frequency bands (Jongman, et al., 2000). Formant structures of glides, on the other hand, resemble those of their closest vowels (cf. Stevens, 1998), such that we were able to use the F2/F1 ratio as estimator for the consonant dipole location in the auditory cortex. We tested the possible correlation of F2/F1 ratio and dipole location along the anterior/posterior axis in a linear regression analysis and found that in fact, the ratio between F2 and F1 positively correlated with the anterior-posterior location of the corresponding dipole (r = 0.38, p < 0.05). This is not a contradiction to our positional hypothesis, whereby relations between articulatory locations are reflected in the auditory cortex. It just so happens that the F2/F1 ratio appears to be the most robust acoustic cue for the quantal articulator positions LABIAL and CORONAL. Importantly, this cue is crucially dependent on linguistic intelligibility, as suggested by previous work (Obleser, et al., 2006), which strengthens the view that this acoustic-to-articulator mapping (i.e. F2/F1 to PoA) is a solid basis for distinctive feature theory.

Our view that the featural basis is both acoustic and articulatory information is bolstered by our coarticulatory finding, as seen in the vowel M100 dipole locations, which followed the same pattern as the consonants. As seen in Figure 1, fricatives and especially glides exerted a measurable coarticulatory influence on the initial vowel of our stimuli. This influence is expressible in slight changes of F2 and F1, but did not change the vowel category (i.e. the [a] did not change to [e]). Obviously, this slight acoustic change transported knowledge about the corresponding articulator movements responsible for the production of the upcoming consonants. We argue that therefore, the dipole locations are determined by top-down knowledge of quantal
PoA features and bottom-up information of robust acoustic cues, resulting from the corresponding articulator configurations.

Regarding lexical feature specifications, our experiment provided no direct evidence for featural underspecification as proposed by Lahiri & Reetz, 2002. In sensor space, we found that the MMF in our MEG study was consistently stronger if the deviants had a labial place of articulation. In addition, the contrast between a coronal standard and a labial deviant led to a stronger brain magnetic response than vice versa. Since coronals are assumed to be underspecified, while labials are specified, our outcome is inconsistent with the Eulitz & Lahiri (2004) findings and proposed explanation. It is worth noting that we used different segment types (consonants instead of vowels) and a different language (English instead of German), limiting the strict equivalence of our results. Further, since labials are more marked in English than coronals, our MMF results are much better explained by assuming that more marked and/or specified segments are more distinguishable, since discrimination accuracy is known to correlate quite well with MMN amplitudes (Näätänen, Paavilainen, Rinne, & Alho, 2007). Our alternative interpretation stresses the relative salience of [labial], potentially overriding the mismatch effects for coronal deviants. We conjectured that the labial, i.e. “more front” PoA bears greater perceptual salience due to the availability of a visual cue, i.e. the lips. Labial saliency has been found to play a crucial role in language acquisition (Boysson-Bardies & Vihman, 1991; Vihman, 1986) as well as audio-visual speech processing (Dorman & Loizou, 1996; Hume, Johnson, Seo, Tserdanelis, & Winters, 1999; Winters, 1999). Thus, it seems that a change towards a place of articulation with greater salience results in enhanced neural activity.

Regarding latencies, our data are less consistent and showed no effects for PoA. We attribute this to the selection of naturally spoken VCV stimuli, where acoustic and co-articulatory cues
might have become available at variable times so that we cannot provide a concise account of the
time course of feature processing and integration. Interestingly, the coronal fricative showed
earlier latencies in the consonantal M100 and mismatch field time windows, which could have
resulted from an earlier availability of spectral energy at the consonant onset.

Taken together, our data indicate that the brain uses feature-based categories to structure the
phoneme inventory of a given language (here: American English). The responses in source and
sensor space seemed to be driven by both acoustic and articulatory information. We suggest that
the features labial and coronal circumscribe particular articulatory configurations that result in
robust acoustic cues over a range of constriction types and manners of articulation. Our MMF
results – though not directly supporting underspecification theory – provide evidence for the
assumption that LABIAL and CORONAL are mutually exclusive PoAs for consonants. This
constraint is categorical, since we observed the same pattern of results in more vowel-like
consonants (glides, Experiment 1) and less vowel-like consonants (fricatives, Experiment 2). The
corresponding ECDs of their evoked auditory M100 responses predominantly differed along the
anterior/posterior axis in the auditory cortex. If considering a third point in this dimension –
dorsal from Obleser et al. (2004) – the ECD alignment reflects the broad articulator sequence and
parallels the alignment of centers of activation in the somatosenory cortex. Regarding the quantal
theory of speech production and perception, the brain keeps robust and distinctive information
about articulator configurations spatially apart during early stages of speech perception.
References


Näätänen, R. (2001). The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology, 38*, 1-21.


ACKNOWLEDGMENTS

We would like to thank David Poeppel, Diogo Almeida, and Philip Monahan for helpful comments on earlier versions of this manuscript.
Figure captions

Figure 1
First and second formant values at 100 ms of the initial vowel [a] of each stimulus, depending on the preceding consonant context. Note that the F1 axis shows only a small portion of the full F1 range, which is indicated in the upper half of the figure.

Figure 2
Time course of the MEG responses to the VCV stimuli. The onset of deviance in each stimulus is at the offset of the initial vowel (~ 200 ms). The mismatch field is expected between 150 and 250 ms post onset of deviance, i.e. between 350 and 450 ms post stimulus onset.

Figure 3
Dipole locations in the anterior-posterior and inferior-superior plane in the auditory cortex, based on the consonant M100. Grand averages for standards and deviants [aja] and [awa] are plotted in the lower panel.
Figure 4

Dipole locations in the anterior-posterior and inferior-superior plane in the auditory cortex, based on the consonant M100. Grand averages for standards and deviants [aʒja] and [ava] are plotted in the lower panel.

Figure 5

Dipole locations in the anterior-posterior dimension for the consonant M100, pooled data from Experiment 1 & 2. Labial deviants were located more anterior than coronal deviants. Based on Obleser et al. (2004), we hypothesize that dorsal (i.e. velar, uvular etc.) consonants would be about 2.5 mm more posterior in relation to coronal consonants.