Predicting Speech: Neural Correlates of Voicing Mismatch using MEG

Philip J. **Monahan**[1], So-One K. **Hwang**[1], William J. **Idsardi**[1,2]
Department of Linguistics[1] and Neuroscience and Cognitive Science Program[2]
University of Maryland

Address for Correspondence:

Philip J. Monahan
Cognitive Neuroscience of Language Laboratory
Department of Linguistics
1401 Marie Mount Hall
University of Maryland
College Park, MD 20742-7505
Office: (301) 405-8140
E-Mail: pmonahan@umd.edu

Running Header: Predicting Speech

Word Count: 4628
Figures: 3
Tables: 1

Abstract

Constraints on how speech sounds are sequenced in spoken language comprehension and production constitute an important part of one's phonological knowledge. Understanding how this knowledge is deployed - and how it might influence early auditory processing - is critical for determining the mechanisms involved in speech perception. We report MEG results that suggest early auditory cortical processing is sensitive to violations of a universal phonological constraint, specifically that syllable final obstruent consonant clusters must agree in voicing. By 150 ms after the second (violating) obstruent, we find a reliable difference in the areal response amplitude (RMS) of the MEG temporal wave-form to syllables in which the final obstruents disagree in voicing. These findings suggest that listeners make highly specific, knowledge-based predictions about rather abstract anticipated properties of the upcoming speech signal and violations of these predictions are evident in early cortical processing.

Keywords

Speech, Language, Auditory Cortex, Magnetoencephalography, MEG, Perception

1. Introduction

Languages place constraints on acceptable and unacceptable sequences of sounds. For example, in English, [tr] is an acceptable consonant sequence at the beginning of a word, whereas [rt] is not; at the ends of words, the situation is reversed, [rt] is acceptable, but [tr] is not. Some of these constraints are specific to individual languages, while others are universally attested, that is, they seem to occur in every extant human language. Many studies, both behavioral and neurophysiological, have demonstrated the role of language-specific phonetic inventories in speech perception (Hacquard et al., 2007; Kazanina et al., 2006; Kuhl, 2004; Kuhl et al., 1992; Näätänen et al., 1997; Werker and Tees, 1984), but relatively few have discussed the role and time-course that the knowledge of phonological constraints play in the auditory mapping between acoustic input and linguistic representations. In this paper, we present magnetoencephalographic (MEG) data that suggests that auditory cortex is sensitive to violations of phonological constraints as early as 150 ms post-onset of the violating segment. We take this evidence to suggest that listeners generate predictions about relatively abstract properties of the speech signal based on higher-order knowledge (i.e., knowledge of constraints on sound sequences), and this knowledge constrains early auditory cortical processing of speech sounds.

Several earlier behavioral studies have demonstrated that psychophysical measures index such abstract phonological knowledge and have suggested that listeners are sensitive to violations of language-specific phonological constraints (Fowler and Brown, 2000; Gaskell and Snoeren, 2008; Gow, 2001; Lahiri and Marslen-Wilson, 1991). A few of these experiments have exploited the anticipatory nasalization rule in English, which accounts for the fact that typically oral vowels are produced with additional air-flow through the nasal cavity, and hence, become “nasalized vowels”, when they precede nasal consonants (i.e., [m n ŋ]). One question that could be asked regarding these structures is whether listeners can use the information on the vowel (namely, that it is nasalized) to predict that the upcoming segment is a nasal consonant in real-time speech perception. Behaviorally, it has been demonstrated using a consonant identification task that native speakers of English encounter most difficulty when they perceive a nasal consonant following an oral vowel (Fowler and Brown, 2000), and moreover, the presence of nasalization on the vowel leads to more nasal responses in a gating task prior to the perception of the critical consonant (i.e., the nasal consonant) (Lahiri and Marslen-Wilson, 1991). These findings suggest that listeners, to some extent, are able to use this language specific phonetic information to predict the nature of the upcoming speech signal.

More recently, electrophysiological techniques (MEG/EEG) have been employed to ask questions about the neural time-course of these violations of phonological expectation (Flagg et al., 2005; Mitterer and Blomert, 2003; Tavabi et al., 2009). Flagg, et al. (2005) exploited the pattern described above where pre-nasal vowels are nasalized in English and measured the electrophysiological latencies of the response peaks in MEG to congruent (i.e., [aba], [ãma]) and incongruent (i.e., [ãba], [ama]) VCV sequences. Overall, latencies in the time-window of 50-100 ms post-onset of the consonant were shorter for the congruent as opposed to incongruent sound sequences. In a different experiment, Mitterer & Blomert (2003) found that violations of expected nasal place assimilation patterns between words elicited a larger MMN amplitude (approximately 100-200 ms after onset of the consonant) than adherence to assimilation patterns in Dutch listeners. These previous electrophysiology results (Flagg et al., 2005; Mitterer and Blomert, 2003) suggest that although cortical responses that reflect phonological processing are early, the types of responses measured (latency versus amplitude) and the time-window in which differences were found varied. To date, the precise nature of the time course of the role of phonological knowledge in speech perception remains poorly understood.

In the present study, we use a pervasive cross-linguistic generalization as our case study: syllable-final obstruent consonant clusters must agree in their specification for voicing (i.e., whether or not the vocal folds vibrate during the production of the speech sound). That is, in all known languages, word-final obstruent consonant phonetic sequences that agree in voicing, such as [dz] (both voiced) and [ts] (both voiceless), are acceptable, while those that disagree are not: *[ds] ([d]: voiced; [s]: voiceless) and *[tz] ([t]: voiceless; [z]: voiced). Traditional linguistic analyses have taken this constraint to be the product of a phonological rule of assimilation. Typically, the feature [-voice] on voiceless consonants spreads to the word final consonant (Chomsky and Halle, 1968).

In a previous behavioral study (Hwang, S.O., Monahan, P.J., Idsardi, W.J., submitted), participants were presented with congruent (e.g., [udz], [uts]) and incongruent (e.g., [uds], [utz]) tokens and asked to respond whether they perceived [z] or [s]. In both the reaction time and accuracy data, we found that English listeners are sensitive to these violations, but only when the first consonant of the sequence is voiced (e.g., [d]). That is, we found an asymmetric pattern of results: the congruent pair did not behave alike ([udz] was faster and more accurate than [uts]), and more interestingly, the incongruent pair also did not behave alike ([uds] was slower and less accurate than [utz]). Moreover, there was no difference between the grammatically acceptable [uts] and the grammatically unacceptable [utz]. We suggest that this particular pattern of results is anticipated if one assumes that phonological representations can be underspecified. That is, predictable distinctive features (e.g., [-voice], [coronal], etc.) are not a part of a sound's representation in long-term memory (Archangeli, 1988; Lahiri and Reetz, 2002; Lahiri, 2007). To illustrate, because the nasal consonant /n/ often undergoes place assimilation (/in+kʌmplit/ [/iŋkʌmplit] 'incomplete') with the following sound (i.e., it takes on the place of articulation of the following consonant), it has been hypothesized that [coronal] is not specified for /n/. Therefore, the long-term memory representation for the phonological segment /n/ does not contain a specification for where in the mouth it is articulated because its place is predictable provided a phonological context. Only the features that are unpredictable count toward a particular sound's long-term representation. Returning to our previous behavioral results, the particular pattern is predicted if we combine the proposals that the feature [voice] is specified for only voiced consonants (e.g., [d], [z]) (Lombardi, 1995; Mester and Itô, 1989) with the hypothesis that only represented features can be used as the basis to make predictions about the upcoming speech signal (Lahiri and Reetz, 2002). More concretely, the feature [voice] on /d/ can be exploited to predict that the following sound is going to be voiced (i.e., [z]), consistent with the universal constraint. The prediction is met when the listener encounters /z/ and is violated when the listener encounters /s/. On the other hand, when a listener encounters /t/, where there is no feature for voicing in the long-term representation, and therefore, no prediction can be generated regarding whether the following sound is going to be voiced or voiceless. Asymmetric results of this sort are not uncommon, even using electrophysiological techniques, and are typically taken to support underspecified long-term representations (Eulitz and Lahiri, 2004; Friedrich et al., 2006).

Using MEG, we examine the time course of the role of phonological knowledge in speech perception for this specific and putatively universal constraint (syllable final obstruent clusters must agree in voicing) and aim to better understand whether neurophysiological measures reflect a sensitivity to constraints on how speech sounds sequence in early auditory perception. Such findings contribute to a better understanding of the time course and nature of the mechanisms involved in mapping acoustic signals onto linguistic representations (Hickok and Poeppel, 2007; Obleser and Eisner, 2009; Poeppel et al., 2008; Poeppel and Monahan, 2008).

2. Results

Given that we were interested in the response to the final fricative, and to eliminate the possibility that any differences found in the RMS of the MEG temporal waveform could be attributed to low-level acoustic properties of the stimulus, we compared only items that had the same final consonant (e.g., [uts] with [uds]; [udz] with [utz]). For each participant in each condition, we calculated the root mean square (RMS) amplitude of the MEG temporal waveforms from ten left hemisphere channels, selected on the basis of an auditory localizer pretest.

INSERT FIGURE 1 ABOUT HERE

Subsequently, we calculated the grand average RMS across participants (n=14) for each condition. For the conditions of interest, paired two-tailed t-tests were performed on the grand averaged RMS waveforms for specific time-windows. Two factors led to our choice of the 150-400 ms time window as the initial window of comparison. First, visual inspection of the grand averaged RMS waveforms suggested a large difference in this time window. Second, early auditory responses are known to occur before 150 ms (Roberts et al., 2000) and higher-order lexical-semantic effects are traditionally seen by 350 to 400 ms (Lau et al., 2008). We hypothesize that whatever effects we might find would likely be due to mechanisms that operate between initial auditory processing and lexical access. Therefore, the time window of 150–400 ms provides the temporal boundary conditions within which we were looking for a systematic response modulation of the MEG temporal signal.

Collapsing across place of articulation of the stop consonant, we found a significant difference between UTS (UTS: [uts],[ups],[uks]; C1: unvoiced, C2: voiced) and UDS (UTS: [uds],[ubs],[ugs]; C1: voiced, C2: unvoiced) in the time-window of 150-400 ms post onset of the fricative ($t = -2.68$; $p < 0.02$).

INSERT FIGURE 2 ABOUT HERE

For the stimuli in which the final consonant was voiceless (UTS/UDS), we found a significant effect for coronals ($t=-3.01$; $p < 0.02$), a marginal effect for labials ($t=-1.93$; $p = 0.08$) and no effect between the dorsal tokens ($t=-0.95$; $p = 0.36$) in the time window of 150-400 ms. In order to more precisely determine the time course of these effects, we tested smaller time windows by dividing the original time window in half: 150-275 ms and 275-400 ms. For the earlier time window (150-275 ms), we found a significant difference when collapsing across the place of articulation ($t=-2.29$; $p < 0.05$). Analyzing the individual places of articulation independently, we found marginal effects for the labial ($t=-1.77$; $p = 0.10$), coronal ($t=-1.75$; $p = 0.10$) and dorsal ($t=-1.93$; $p = 0.08$) pairs. In the later time window (275-400 ms), we again found a difference when collapsing across the place of articulation (t: -2.64; $p < 0.05$). Moreover, we found a reliable effect for the coronal pair (t=-3.20; p < 0.01) and no effect for the labial (t=-1.69; p = 0.12) and dorsal pairs (t=-0.15; p = 0.88). We did not, however, find a reliable difference between UTZ and UDZ in the larger time window (150-400 ms: t=0.27; *p* = 0.79) or in either of the smaller time windows (150-275 ms: t=-0.62; *p* =0.54; 275-400 ms: t=1.33; *p* = 0.21). No other differences were found in any other time-window.

3. Discussion

In the current experiment, we tested a cross-linguistically attested phonological constraint that requires syllable final consonant clusters to agree in voicing. To the best of our knowledge, this process has yet to be exploited in the understanding of the neural mechanisms underlying phonological processes. Analyzing the grand-averaged RMS of the MEG temporal waveforms, we found a reliable difference between congruent (i.e., UTS) and incongruent (i.e., UDS) syllables as early as 150 ms post-onset of the violating segment, in this case, the fricative (i.e., [s]). We take the differences between UTS and UDS to suggest that listeners can exploit their knowledge of phonological processes and representation to constrain early perceptual parses of the sensory input, as well as using this detailed knowledge to serve as the basis for generating hypotheses and predictions about the nature of the upcoming speech signal.

Consistent with an underspecification view of perception, we hypothesize that only features that are specified in the long-term representation for phonological segments can form the basis for the generation of predictions regarding the nature of the upcoming speech stimulus (Lahiri and Reetz, 2002; Lahiri, 2007). In this particular case, we adopt the proposal that [-voice] is not represented for voiceless consonants while [+voice] is represented for voiced consonants (Lombardi, 1995; Mester and Itô, 1989). Therefore, only sounds that contain [voice] (i.e., voiced consonants) can be used to predict the phonological quality of the upcoming sound, in particular, that the next consonant in the syllable must also be specified for [voice]. Sounds that do not have a specification for [voice] (i.e., voiceless consonants) cannot be used to predict whether the next consonant in the syllable is also voiceless. Consequently, from our data, it appears that when a listener unexpectedly encounters a voiceless consonant followed a consonant specified for [voice], they are surprised, and this violation of expectation is indexed by early cortical processing. When a listener encounters a sound that is not specified for [voice], no predictions can be made, and thus, we found no difference between UTZ and UTS. Moreover, it appears that this knowledge of phonological representations and constraints which act as the basis for these online predictions are reflected in early cortical processes.

Unlike our previous behavioral study (Hwang, S.O., Monahan, P.J., Idsardi, W.J., submitted), where two effects were found: processing facilitation for UDZ and processing difficulty for UDS (while UTS and UTZ showed no differences), the present MEG study only showed a difference between UDS in comparison with UTS. While we did not predict a lack of difference between UTZ and UDZ (it should be noted that the magnitude of the effect size in both the reaction time and accuracy data was significantly smaller for UDZ versus UTZ as opposed to UTS versus UDS in the previous behavioral experiment), this asymmetric result does, however, allow us to eliminate a possible alternative explanation for our findings in which the phonetic quality of the obstruent consonant (i.e., D versus T) is driving the differences in the RMS of the MEG temporal waveform. If this alternative explanation were true, then we would have expected to also find a difference between UTZ and UDZ. We can thus be more confident that the difference we did find is attributable to a violation of expectation and not low-level acoustic properties of the stimulus.

The time course of our effects are consistent with the previous electrophysiological results (Flagg et al., 2005; Mitterer and Blomert, 2003) that investigated the role of phonological knowledge in speech perception. Flagg et al. (2005) tested violations of the language specific constraint that pre-nasal vowels are nasalized in English. They found a reliable difference in the latency of the M50 in the MEG waveform to the consonant for the tokens [aba] compared with [ãba]. They failed to find a difference when the consonant was the nasal [m], however (i.e., [ãma] compared with [ama]). Given the complex nature of the evoked magnetic waveform to these stimuli, we are less confident that the component they identified was, in fact, the M50. The time-course of these effects, however, suggests that differences are evident in early cortical processing. Mitterer & Blomert (2003) used an MMN paradigm with Dutch speakers and found that unviable phonological assimilations (/n/ becoming [m] before /s/) elicited a mismatch negativity in a passive oddball paradigm, while viable phonological assimilations (/n/ becoming [m] before /b/) did not. The effects in Mitterer & Blomert (2003) were seen in the traditional MMN time window (~250 ms post-onset), again, suggesting that early cortical processes are sensitive to violations of a phonological constraint. Unlike the Flagg et al. (2005) results, the phonological process used by Mitterer & Blomert (2003) is generally cross-linguistically attested. That is, coronal nasal consonants (i.e., /n/) usually undergo assimilation to the place of articulation of a neighboring consonant, though the directionality of assimilation often differs on a language-by-language basis. The cross-linguistic nature of this assimilation pattern is evident by the fact that they also tested German listeners on the Dutch contrast and found similar results.

3.1 Conclusion.

Convergent with other behavioral findings, these MEG results suggest that listeners make use of their knowledge of phonological constraints regarding sound sequences to predict the phonetic quality of the upcoming sound. Moreover, violations of these expectations are seen in early auditory cortical processes, as indexed by the RMS of the MEG temporal waveform, in particular, by 150 ms post-onset of the violating segment. These results provide further support to the idea that listeners actively utilize their phonological knowledge to parse the speech signal and make predictions regarding the phonetic quality of the upcoming segment.

4. Experimental Procedure

4.1 Participants.

Fourteen monolingual native speakers of American English participated in this study (11 female; mean age: 20.8 years) and were included in the analysis. An additional six participants took part but were excluded from the analysis (3 for poor magnetic field contours; 2 for poor M100 responses to the vowel; 1 for poor M100 to the auditory localizer pretest). These are metrics that we use to assure that the participant is providing a reliable auditory response to the speech tokens, specifically, and auditory tokens, generally. Each participant tested strongly right handed on the Edinburgh Handedness Survey (Oldfield,

1971) and provided written informed consent. Subjects received either course credit or $10 for their participation.

4.2 Materials.

A male native speaker of American English recorded natural utterances of the English non-words [ups], [uts], [uks] and [ubz], [udz], [ugz]. These recordings were edited with Praat (Boersma, 2001) to create tokens with voicing agreement (e.g., [uts], [udz]) and voicing disagreement (e.g., [utz], [uds]). The tokens with voicing agreement were spliced together from two tokens of the same type (e.g., [ut] was spliced from [uts] and combined with [s] which was spliced from a different token of [uts]). The tokens with voicing disagreement were cross-spliced together from two natural tokens (e.g., [ut] was spliced from [uts] and combined with [z] which was spliced from [udz]). This was done for each place of articulation. There were a total of twelve items in this study. In the end, all sounds were edited to eliminate any response bias toward edited speech stimuli. Additionally, the stimuli were edited such that each segment was 100 ms in duration (300 ms total for each VCC syllable) to eliminate any response bias based on stimulus length alone.

INSERT TABLE 1 ABOUT HERE

The stop bursts were removed and each token was gradually ramped so that the vowel had a 20 ms fade-in and the final fricative had a 20 ms fade-out.

INSERT FIGURE 3 ABOUT HERE

4.3 Procedure.

Magnetoencephalographic recordings were acquired using a 157-channel whole-head axial gradiometer MEG system (Kanazawa Institute of Technology, Kanazawa, Japan). Participants lay supine in a dimly lit magnetically shielded room. Auditory stimuli were delivered binaurally via Etymotic ER3A insert earphones. Earphones were calibrated to have a flat frequency response between 50 Hz and 3100 Hz within the shielded room. Participants were first presented with an auditory localizer pretest to ensure adequate positioning of the head within the scanner. The pretest involved passive listening to a series of 250 Hz and 1 KHz sinusoids.

For the experiment, stimulus presentation included 150 randomized trials of each of the twelve tokens using Presentation® (Neurobehavioral Systems, Inc.) software. The inter-stimulus interval (ISI) varied pseudo-randomly between 450 ms and 1450 ms. Participants were asked to respond to a distracter 1 KHz sinusoid tone and listen passively to the speech stimuli. The neuromagnetic signal was sampled at 500 Hz with an online 200 Hz LPF and 60 Hz notch filter. Offline, the data were noise reduced using a multi-shift PCA noise reduction algorithm (Cheveigné and Simon, 2007) and was band-pass filtered by a Hamming-window digital filter with frequency cut-offs at 0.03 Hz and 30 Hz. Stimulus-related epochs of 1100 ms (500 ms pre-trigger/fricative (i.e., [s z])) were averaged according to stimulus type. A portion of the epoched window (-400 to -300 ms) where no stimulus was present (-200 ms was onset of vowel) was used to baseline correct the averaged file. We were interested in whether the presence of a voiced obstruent (e.g., [d]) caused difficulty in the processing of a voiceless fricative (e.g., [s]) and vice versa. Therefore, the trigger was placed at the onset of the fricative (200 ms after the start of the token). Ten left hemisphere channels that best correlated with the neuromagnetic signal (five channels from the source and five channels from the sink of the dipole) were selected for statistical analysis based on the auditory pretest. The channels were selected on a participant-by-participant basis, but the selected channels were the same for all conditions for each individual subject. The RMS of the ten left hemisphere channels was calculated for each subject for each condition. Subsequently, the RMS temporal waveforms were averaged across participants for each condition, and a paired t-test was performed on the RMS of the planned comparisons for particular time windows.

References

Archangeli, D., 1988. Aspects of underspecification theory. Phonology. 5, 183-207.

Boersma, P., 2001. Praat, a system for doing phonetics by computer. Glot International. 5, 341-345.

Cheveigné, A.d., Simon, J.Z., 2007. Denoising based on time-shift pca. J Neurosci Methods. 165, 297-305.

Chomsky, N., Halle, M., 1968. The sound pattern of English. Vol., Harper and Row, New York.

Eulitz, C., Lahiri, A., 2004. Neurobiological evidence for abstract phonological representations in the mental lexicon during speech recognition. J Cogn Neurosci. 16, 577-583.

Flagg, E.J., Cardy, J.E.O., Roberts, T.P.L., 2005. MEG detects neural consequences of anomalous nasalization in vowel-consonant pairs. Neurosci Lett. 397, 263-269.

Fowler, C.A., Brown, J.M., 2000. Perceptual parsing of acoustic consequences of velum lowering from information for vowels. Percept Psychophys. 62, 21-32.

Friedrich, C.K., Eulitz, C., Lahiri, A., 2006. Not every pseudoword disrupts word recognition. Behav Brain Funct. 2, 36.

Gaskell, M.G., Snoeren, N.D., 2008. The impact of strong assimilation on the perception of connected speech. J Exp Psychol Hum Percept Perform. 34, 1632-1647.

Gow, D.W., 2001. Assimilation and anticipation in continuous spoken word recognition. J Mem Lang. 45, 133-159.

Hacquard, V., Walter, M.A., Marantz, A., 2007. The effects of inventory on vowel perception in french and spanish: An MEG study. Brain Lang. 100, 295-300.

Hickok, G., Poeppel, D., 2007. The cortical organization of speech processing. Nat Rev Neurosci. 8, 393-402.

Kazanina, N., Phillips, C., Idsardi, W.J., 2006. The influence of meaning on the perception of speech sounds. Proc Natl Acad Sci U S A. 103, 11381-11386.

Kuhl, P., 2004. Early language acquisition: Cracking the speech code. Nat Neurosci. 5, 831-843.

Kuhl, P.K., Williams, K.A., Lacerda, F., Stevens, K.N., Linblom, B., 1992. Linguistic experience alters phonetic perception in infants by 6 months of age. Science. 255, 606-608.

Lahiri, A., Marslen-Wilson, W., 1991. The mental representation of lexical form: A phonological approach to the recognition lexicon. Cognition. 38, 245-294.

Lahiri, A., Reetz, H., 2002. Underspecified recognition. In Laboratory phonology 7. C. Gussenhoven, N. Warner, eds. Mouton de Gruyter, Berlin, pp. 637-675.

Lahiri, A., 2007. Non-equivalence between phonology and phonetics. Proceedings of ICPhS XVI. 31-34.

Lau, E.F., Phillips, C., Poeppel, D., 2008. A cortical network for semantics: (de)constructing the n400. Nat Rev Neurosci. 9, 920-933.

Lombardi, L., 1995. Dahl's law and privative voice. Linguist Inq. 26, 356-372.

Mester, R.A., Itô, J., 1989. Feature predictability and underspecification: Palatal prosody in Japanese mimetics. Language. 65, 258-293.

Mitterer, H., Blomert, L., 2003. Coping with phonological assimilation in speech perception: Evidence for early composition. Percept Psychophys. 65, 956-969.

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Houtllainene, M., Iivonen, A., Vainio, M., Alku, P., Iimoniemi, R.J., Luuk, A., Allik, J., Sinkkonen, J., Alho, K., 1997. Language-specific phoneme representations revealed by electric and magnetic brain responses. Nature. 385, 432-434.

Obleser, J., Eisner, F., 2009. Pre-lexical abstraction of speech in the auditory cortex. Trends Cogn Sci. 13, 14-19.

Oldfield, R.C., 1971. Assessment and analysis of handedness: Edinburgh inventory. Neuropsychologia. 9, 97-113.

Poeppel, D., Idsardi, W.J., Wassenhove, V.v., 2008. Speech perception at the interface of neurobiology and linguistics: Prospects and problems. Philos Trans R Soc Lond B Biol Sci. 363, 1071-1086.

Poeppel, D., Monahan, P.J., 2008. Speech perception: Cognitive foundations and cortical implementation. Curr Dir Psychol Sci. 17, 80-85.

Roberts, T.P.L., Ferrari, P., Stufflebeam, S.M., Poeppel, D., 2000. Latency of the auditory evoked neuromagnetic field components: Stimulus dependence and insights toward perception. J Clin Neurophysiol. 17, 114-129.

Tavabi, K., Elling, L., Dobel, C., Pantev, C., Zwisterlood, P., 2009. Effects of place of articulation changes on auditory neural activity: A magnetoencephalography approach. PLoS ONE. 4, e4452.

Werker, J.F., Tees, R.C., 1984. Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. Infant Behav Dev. 7, 49-63.

TABLE 1
Title: List of conditions in the experiment by place of articulation.
A male native speaker of American English recorded natural utterances of the acceptable English non-words. Each segment was 100 ms in duration so that the total duration was 300 ms. The final consonant of each recorded token was removed and spliced together to create tokens with voicing agreement (attested) and voicing disagreement (unattested).

FIGURE 1
Title: Sample MEG waveform from a representative subject for one condition.
The waveform plot is an overlay of ten left-hemisphere sensors selected on the basis of an auditory localizer pretest. Bold line is the root mean square (RMS) of the ten channels. Magnetic field contour plots from an individual subject for four different time points are presented below the waveform plot (-100 ms: 100 ms post-onset of the vowel; 0 ms: 100 ms post-onset of first consonant; 100 ms: 100 ms post-onset of the second consonant; 200 ms: 200 ms post-onset of the second consonant). Dark gray refers to sink of magnetic field, while light gray refers to the source of the magnetic field.

FIGURE 2
Title: Overlay of the grand average MEG RMS temporal waveforms.
Conditions UDS (dotted line) and UTS (solid line) collapsed across place of articulation. Shaded area on the waveform plot designates region of significant difference between the two conditions. Acoustic waveform overlay on temporal waveform plot denotes stimulus presentation relative to neuromagnetic signal.

FIGURE 3
Title: Spectrograms and waveforms for four conditions in the experiment.
The conditions on the left (A and B) are cross-linguistically attested (the two consonants agree in voicing). These tokens were created by splicing (A) [ud] with [z] from two tokens of [udz] and (B) [ut] with [s] from two tokens of [uts]. The conditions on the right (C and D) are unattested in the world's languages (the two consonants disagree in voicing). These tokens were created by cross-splicing (C) [ud] with [s] from a token of [udz] and a token of [uts], respectively and (D) [ut] with [z] from two tokens of [uts] and [udz] respectively. The amount of voicing is evident by the low frequency energy in the spectrograms and the periodicity in the waveforms between 0.1 and 0.2 sec in (A) and (C) and the respective absence in (B) and (D).

Acknowledgements
We would like to thank Jeffrey Walker for invaluable lab assistance and David Poeppel for comments on various aspects of this project. This work was supported by NIH R01 05660 to David Poeppel and William J. Idsardi.

Figure 1

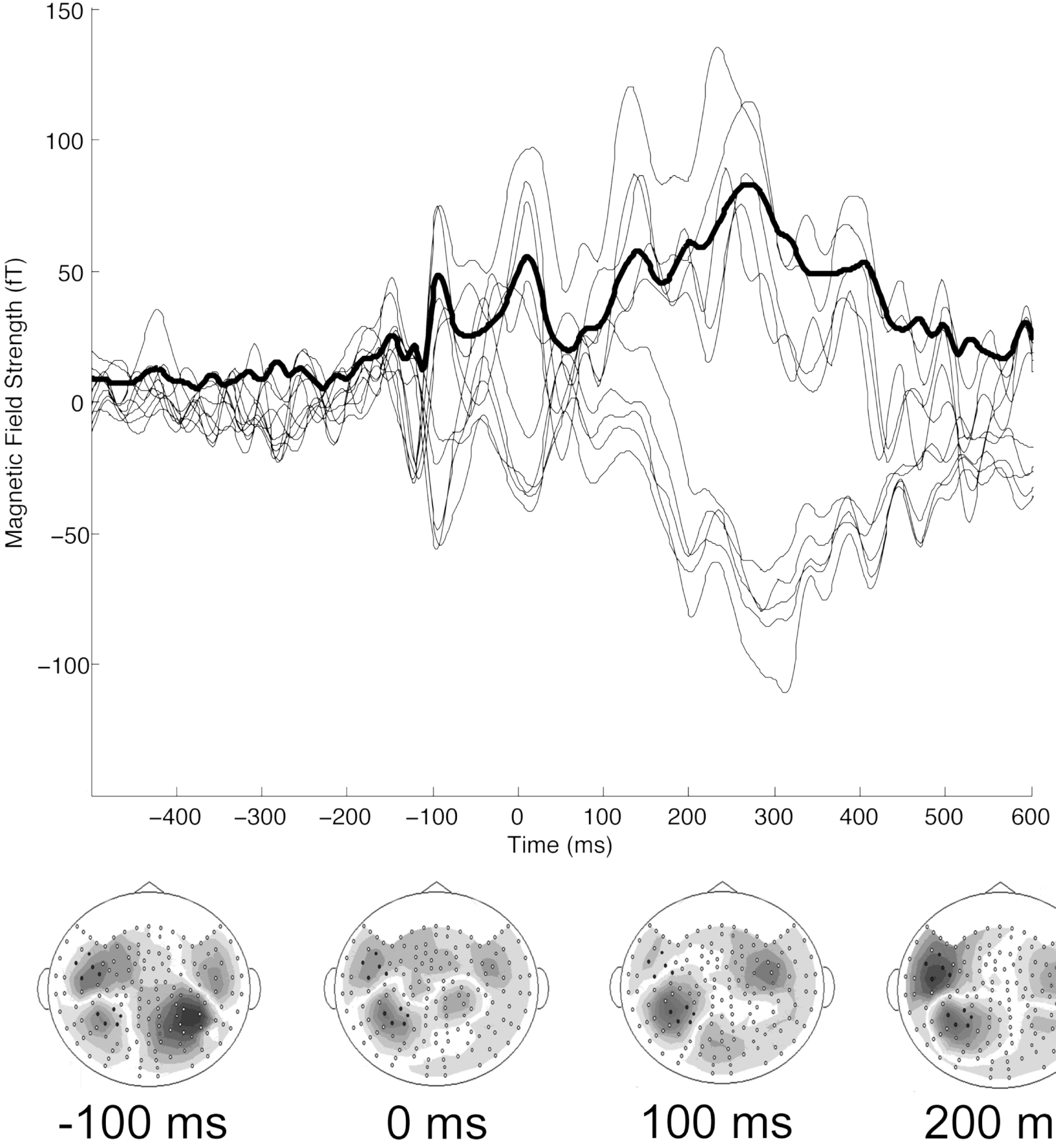
150
100
50
0
−50
−100
Magnetic Field Strength (fT)
−400
−300
−200
−100
0
100
200
300
400
500
600
Time (ms)
-100 ms
0 ms
100 ms
200 ms

Figure 2

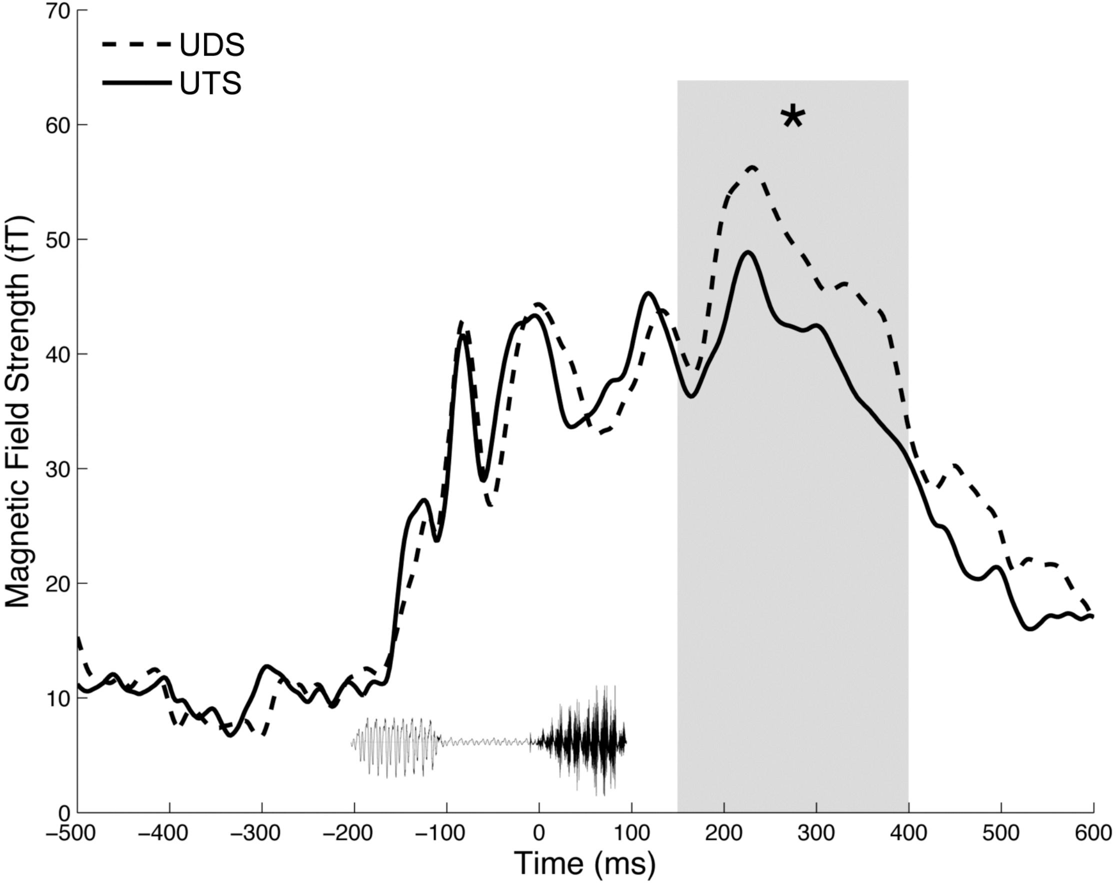
UDS
UTS
*
Magnetic Field Strength (fT)
70
60
50
40
30
20
10
0
−500
−400
−300
−200
−100
0
100
200
300
400
500
600
Time (ms)

Figure 3

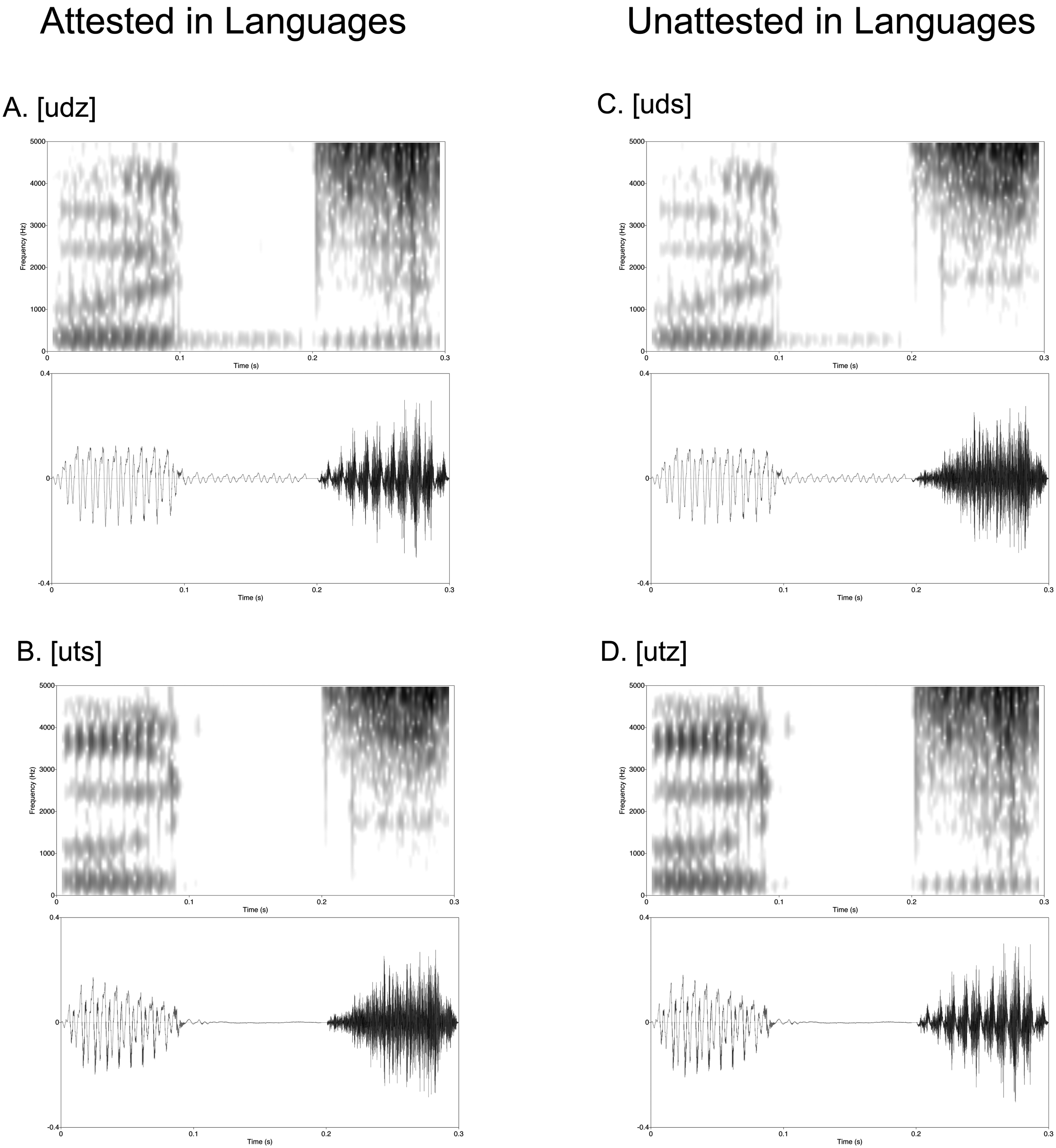
Attested in Languages
Unattested in Languages
A. [udz]
B. [uts]
C. [uds]
D. [utz]
Frequency (Hz)
5000
4000
3000
2000
1000
0
Time (s)
0
0.1
0.2
0.3
0.4
-0.4

Table 1

Title: List of conditions in the experiment by place of articulation.

	Attested		Unattested	
Condition	UTS	UDZ	UTZ	UDS
Labial	[ups]	[ubz]	[upz]	[ubs]
Coronal	[uts]	[udz]	[utz]	[uds]
Dorsal	[uks]	[ugz]	[ukz]	[ugs]