Published in final edited form as:

Neuroimage. 2011 June 15; 56(4): 2329–2338. doi:10.1016/j.neuroimage.2011.04.007.

You had me at “Hello”: Rapid extraction of dialect information from spoken words

Mathias Scharinger^{a,c}, Philip J. Monahan^b, and William J. Idsardi^a

^aDepartment of Linguistics, University of Maryland, College Park, MD, USA

^bBasque Center on Cognition, Brain and Language (BCBL), Donostia-San Sebastián, Spain

^cMax Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

Abstract

Research on the neuronal underpinnings of speaker identity recognition has identified voice-selective areas in the human brain with evolutionary homologues in non-human primates who have comparable areas for processing species-specific calls. Most studies have focused on estimating the extent and location of these areas. In contrast, relatively few experiments have investigated the time-course of speaker identity, and in particular, dialect processing and identification by electro- or neuromagnetic means. We show here that dialect extraction occurs speaker-independently, pre-attentively and categorically. We used Standard American English and African-American English exemplars of ‘Hello’ in a magnetoencephalographic (MEG) Mismatch Negativity (MMN) experiment. The MMN as an automatic change detection response of the brain reflected dialect differences that were not entirely reducible to acoustic differences between the pronunciations of ‘Hello’. Source analyses of the M100, an auditory evoked response to the vowels suggested additional processing in voice-selective areas whenever a dialect change was detected. These findings are not only relevant for the cognitive neuroscience of language, but also for the social sciences concerned with dialect and race perception.

Keywords

Speaker identity; dialect perception; cerebral voice areas; Mismatch Negativity (MMN); Magnetoencephalography (MEG)

1. Introduction

Little is known about how the brain encodes linguistic dialect information, in particular which speech attributes may cue the recognition of social group affiliations of a speaker during word recognition. While a majority of previous neurophysiological auditory research has focused on voice perception and speaker identity (e.g., Belin et al., 2004; von Kriegstein & Giraud, 2004), several studies have examined the processing of social information carried by speech vocalizations, and in particular, the extraction of speaker-related attributes, such

© 2010 Elsevier Inc. All rights reserved.

Corresponding author: Mathias Scharinger (PhD), Max Planck Institute for Human Cognitive & Brain Sciences, Stephanstr. 1A - 04103 Leipzig, Germany, Tel.: +49 (0) 341-9940-2482, mscharinger@cbs.mpg.de.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

as gender (Mullennix, Johnson, Topcu-Durgun, & Farnsworth, 1995), perceived age (Hartman, 1979; Hartman & Danhauer, 1976), or estimates of body size (Patterson, Smith, Dinther, & Walters, 2008; Smith, Patterson, Turner, Kawahara, & Irino, 2005; van Dommelen & Moxness, 1995). Brain imaging studies also support a dissociation of voice-related properties ('who') and linguistic information ('what') in the speech signal (Formisano, De Martino, Bonte, & Goebel, 2008; Knösche, Lattner, Maess, Schauer, & Friederici, 2002; Stevens, 2004). So-called 'voice selective' areas have been identified along the upper bank of the superior temporal sulcus (STS; Figure 1; area 1), particularly at medial and anterior locations (Belin, Fecteau, & Bédard, 2004; Belin, Zatorre, Lafaille, Ahad, & Pike, 2000), showing increased activity in response to speaker voice characteristics. Neuronal correlates of speaker identity seem to lateralize to right STS (von Kriegstein, Eger, Kleinschmidt, & Giraud, 2003; von Kriegstein & Giraud, 2004). This area showed habituation to repeated voices and strong activation in response to unfamiliar voices (Belin & Zatorre, 2003). Neuronal voice processing is often compared to face perception in that vocal emotions similarly activate the amygdala (Phillips, et al., 1998, Figure 1; area 2) and voice-selective areas show species-specificity effects in both human (Fecteau, Armony, Joannette, & Belin, 2004) and non-human primates (Belin, 2006; Petkov, et al., 2008; Rendall, Owren, & Rodman, 1998; Rendall, Rodman, & Emond, 1996). Conspecific calls lead to greater activation of these areas than calls from other species. Furthermore, voice selective areas in STS occupy brain regions that also subserve multisensory integration of visual (faces) and auditory (voices) information during speech perception (Beauchamp, Argall, Bodurka, Duyn, & Martin, 2004; Ghazanfar, Chandrasekaran, & Logothetis, 2008). Finally, brain responses to voice characteristics are as similarly fast and categorical as brain responses to facial properties.

Electrophysiological experiments using the Mismatch Negativity (MMN, Mismatch Magnetic Field, MMF) response (Näätänen, 1992, 2001; Näätänen, Paavilainen, Rinne, & Alho, 2007) provided evidence that speaker information is extracted from the speech signal within the first 200 ms of the stimulus (Beauchemin, et al., 2006; Knösche, et al., 2002; Titova & Näätänen, 2001). In the auditory domain, the Mismatch Negativity is an automatic, pre-attentive auditory change detection response of the brain observed when there is a detectable change in an otherwise unchanging stimulus (see Pazo-Alvarez, Cadaveira, & Amenedo, 2003 for a review in the visual domain). It localizes to the superior temporal plane in primary auditory cortex (or areas immediately adjacent to primary auditory cortex) (Figure 1; area 3; cf. Alain, Cortese, & Picton, 1998; Csépe, Karmos, & Molnár, 1987; Hari et al., 1984; Javitt, Schroeder, Steinschneider, Arezzo, & Vaughan, 1992; Javitt, Steinschneider, Schroeder, Vaughan, & Arezzo, 1994; Näätänen, 2001; Näätänen & Alho, 1997; see Näätänen & Alho, 1995 for a review) and is generally unaffected by attentional or task demands required of the participants (Alho et al., 1998; Näätänen, 1992; Ritter et al., 1992; Ritter & Ruchkin, 1992). Detectable changes in stimulus frequency, intensity and duration have all been found to elicit an MMN (Gomes, Ritter, & Vaughan, 1995; Sams, Paavilainen, Alho, & Näätänen, 1985; Winkler et al., 1990). In the context of speech perception, it has been elicited by both phonetic distributions of sound categories (Phillips et al., 2000; Sharma & Dorman, 1999; Sharma, Kraus, McGee, Carrell, & Nicol, 1993), as well as phonological categories (Kazanina, Phillips, & Idsardi, 2006; Näätänen et al., 1997; Sharma & Dorman, 2000). Jääskeläinen, et al., (2004) suggests that the MMN might arise as an artifact of the subtraction of the repeated standard from the infrequent deviant, and consequently, is simply the N1p, i.e., there are no dedicated change-detection neuronal populations (cf., Näätänen, 2001; Näätänen, et al. 2007). Näätänen, et al. (2005), however, reviews compelling evidence against this position, e.g., the existence of MMN responses in the absence of auditory stimulation (Yabe, Tervaniemi, Reinikainen, & Näätänen, 1997; Yabe et al., 1998; see also Maess, Jacobsen, Schröger, & Friederici, 2007 for MEG evidence in support of an independent MMN component).

Comparatively fewer studies have looked at MMN modulations by speaker-specific properties. Titova & Näätänen (2001) used standard and deviant stimuli that differed in speaker voice (male/female) and found that differences in gender elicit a robust MMN. Beauchemin et al. (2006) showed that the MMN varies as a function of speaker familiarity. In their experiment, MMN responses were stronger if the deviant was spoken in a familiar voice than if it was spoken in an unfamiliar voice. Although these findings suggest that voice recognition is a pre-attentive process that can be well assessed by electrophysiological measures, to date, no MMN study has contrasted voices from different dialectal affiliations (but see Conrey, Potts, & Niedzielski, 2005 for dialectal influences on phoneme perception).

To this end, the current MMN study uses word stimuli from two different American English dialects, Standard American English (SAE) and African-American English (AAE). In a previous behavioral study (Purnell, Idsardi, & Baugh, 1999), the multi-dialectal J. Baugh conducted an apartment search over the telephone in different neighborhoods within the San Francisco, CA metropolitan area. Phone calls leading to successful appointments depended on a match of his dialect (SAE vs. AAE) with the predominant population group in the respective neighborhood. These results were taken as evidence that information correlated with social background can be extracted from speaker voice, and dialects can be differentiated with very little speech input. Perceptual identification experiments using the single word 'Hello' showed statistically significant abilities to identify racially affiliated dialects from very short (500 ms or less) samples of speech.

2. Materials and Methods

2.1. Motivation

We used the utterance-initial 'Hello' tokens from Purnell, Idsardi, & Baugh (1999), pronounced in the SAE and AAE dialects and investigated the MMN response to the respective dialect change. Our hypothesis is that the MMN is modulated by dialectal categories similar to the way it is modulated by phoneme categories (Näätänen, et al., 1997). For this purpose, it is ideal to use the same speaker for both dialects; consequently, potential differences cannot be attributed to speaker variance or persistent morphological properties of an individual's vocal tract, but should solely stem from indexical acoustic properties of dialect representations. We further opted to use magnetoencephalography (MEG) since it offers the best non-invasive compromise between temporal and spatial resolution of brain responses (Kutas, Petten, & Kluender, 2006).

Auditory stimuli elicit distinct cortical brain responses with characteristic topographies and latencies (Ackermann, Hertrich, Mathiak, & Lutzenberger, 2001). We are interested in two of these auditory evoked responses. The first is the N1/N1m (M100), a prominent peak around 100 ms after stimulus onset (Näätänen & Picton, 1987; Poeppel & Marantz, 2000). The electrical N1 in EEG is a negative-going potential, with a primary subcomponent whose neuronal generators localize to A1 (Picton, Woods, Baribeau-Braun, & Healey, 1976). It is an exogenous response evoked by any auditory stimulus with a clear onset and is found regardless of task or attentional state (see Näätänen & Picton, 1987 for a review). Its MEG counterpart, i.e., the N1m or M100, appears to be the magnetic equivalent of the primary subcomponent that localizes to A1 in supratemporal auditory cortex (Eulitz, Diesch, Pantev, Hampson, & Elbert, 1995; Hari, Aittoniemi, Järvinen, Katila, & Varpula, 1980; Virtanen, Ahveninen, Ilmoniemi, Näätänen, & Pekkonen, 1998). The M100 is an automatic, auditory evoked neuromagnetic component that peaks roughly 100 ms post-onset of a stimulus (Roberts, Ferrari, Stufflebeam, & Poeppel, 2000) and localizes to planum temporale, a region of auditory cortex posterior to Heschl's Gyrus. In good signal-to-noise conditions, its localization reflects the tonotopic organization of primary auditory cortex (Lütkenhöner & Steinsträter, 1998). Equivalent Current Dipole (ECD) modeling of the sources of the M100

has also revealed an *ampliotopic* organization (Pantev, Hoke, Lehnertz, & Lütkenhöner, 1989; cf., Vasama, J. P. Mäkelä, Tissari, & Hämäläinen, 1995). Its evoked latency is modulated as a function of stimulus frequency in response to sinusoidal stimuli (Roberts & Poeppel, 1996) and predictably varies with the first formant (F1) in vowels (Diesch, Eulitz, Hampson, & Ross, 1996; Poeppel et al., 1997; Roberts, Flagg, & Gage, 2004; Tiitinen, A. M. Mäkelä, Mäkinen, May, & Alku, 2005).

The second response we are interested in is the MMN, as introduced above. Note that one interpretation of the MMN is such that the standard activates a memory trace with which the deviant is compared (see Näätänen, 1992). The more variability that occurs within the standards, the more likely that listeners construct an abstraction from the actual acoustic stimulus (Philips et al., 2000). Thus, variable standard tokens are commonly used to engage phonetic and phonological effects that involve more than just acoustic stimulus information. Furthermore, the MMN response is modulated by top-down information pertaining to language-specific long-term representations of speech sounds, such as the inventory of speech sound categories (Näätänen, et al., 1997), affixes (Shtyrov & Pulvermüller, 2002), or words (Pulvermüller, Shtyrov, Kujala, & Näätänen, 2004).

In order to dissociate low-level acoustic effects from phonetic effects that eventually mediate dialect perception, we used two types of deviants in our passive-oddball paradigm that followed variable standards from both dialects ('Hello' in SAE or AAE, N=10). In the across-dialect conditions, SAE standards were followed by AAE deviants, and vice versa, AAE standards were followed by SAE deviants. For the within-dialect conditions, we chose stimuli from the set of natural productions that best approximated the extent of the acoustic differences between across-category stimuli (i.e. SAE and AAE 'Hello'), but stemmed from the same dialect. This was possible since the speaker produced highly variable 'Hello' tokens in both dialects. Thus, for SAE standards, we used a sham SAE deviant that was not used in any other block. Similarly, for AAE standards, we used a sham AAE deviant with the same characteristics. We were therefore able to compare between purely acoustic standard-deviant differences and differences that were accompanied by dialectal differences, mediated through phonetic distinctions. We expect that MMN responses to across-dialect contrasts to be larger than MMN responses to within-dialect contrasts, since across-dialect contrasts involve concomitant higher-level distinctions, with the same logic as has been argued for MMN modulation by native sound inventories (Näätänen, et al., 1997).

2.2. Materials

Ten variable tokens of 'Hello', pronounced in Standard American English (SAE) and African-American Vernacular English (AAE) served as standards and deviants. All stimuli were normalized for intensity using PRAAT (Boersma & Weenink, 2009), had average durations of 500 ms (SD = 34.4) and were presented to participants at a comfortable listening level (~70 dB SPL). Acoustic properties of the 'Hello' tokens are summarized in Table 1 and Figure 2 A.

The dialect tokens differed primarily in the acoustic properties of the vowels [ɛ] and [o]. In particular, SAE [ɛ] is produced slightly further forward in the mouth than its AAE counterpart (Figure 2 B). In order to differentiate between a pure acoustic and a true categorical effect of dialect in MMN responses, we chose a 'sham' deviant from the standards in each dialect that acoustically differed from the rest of the standards to a similar extent than the SAE tokens differed from the AAE tokens on average (Figure 2 C). The logic of this selection is that if the MMN is elicited by a pure acoustic across-dialect contrast between SAE and AAE stimuli, it should be similar to an MMN elicited by a within-dialect contrast between sham deviant and the rest of same-dialect stimuli. In contrast, if the MMN is additionally modulated by dialectal contrasts between the across-dialect standard and

deviants – mediated through the phonetic difference – we expect it to be larger than in sham deviant within-dialect contrasts. Acoustic differences were calculated based on the first three formant/resonant frequencies (F1, F2, F3) of the two vowels [ɛ] and [o] (Figure 2 C).

2.3. Methods

SAE and AAE ‘Hello’ tokens were arranged in a passive oddball-paradigm (Näätänen, Pakarinen, Rinne, & Takegata, 2004). A classical passive oddball setup intersperses one type of rare deviants among frequently occurring standards. In our design, we used two different kinds of deviants: Real deviants (across-dialect) and sham deviants (within-dialect), as described above. In our experiment, both types of deviants occurred after a minimum of 3 consecutive standards with a probability of 0.125.

All standards and deviants were separated by inter-stimulus intervals that varied randomly between 600 and 1200 ms. SAE and AAE stimuli were arranged in two passive oddball blocks (conditions). In condition 1, standards were SAE stimuli and real deviants AAE stimuli, while the sham deviant was the aforementioned SAE exemplar. In condition 2, standards and the sham deviant were AAE exemplars, whereas real deviants were SAE tokens (see Figure 3). Standards and real deviants comprised all ten variable ‘Hello’ tokens of each dialect (see Phillips, et al., 2000). Note that the stimulus variation among the standards was common to both the sham deviant and real deviant conditions.

2.4. Subjects and procedure

Fifteen native speakers of Standard American English (SAE), students from the University of Maryland (4 males, mean age 22.9, SD 5.7), participated for class credit or monetary compensation. They were all from the Mid-Atlantic region and had some exposure to AAE dialects. Participants reported no history of hearing or neurological problems and tested strongly right-handed on the Edinburgh Handedness Inventory (Oldfield, 1971). All participants provided informed written consent approved by the University of Maryland institutional review board.

During MEG recording and stimulus presentation, participants lay supine in a magnetically-shielded chamber. Magnetic fields were recorded by a whole-head, 157 axial-gradiometer MEG system (Kanazawa Institute of Technology, Kanazawa, Japan). The recording sampling rate was 500 Hz. Auditory stimuli were delivered binaurally via Etymotic ER3A insert earphones calibrated to have a flat frequency response between 50 Hz and 3100 Hz within the magnetically shielded room. In order to accurately locate responses from auditory cortex, a two-tone screening experiment preceded the main study. Participants were instructed to silently count high (1000 Hz) and low (250 Hz) sinusoidal tones (~300 total) presented in pseudo-random order. 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). All participants showed a reliable bilateral M100 response with a source/sink reversal between anterior and posterior channels in the left and right hemisphere.

During the main experiment, participants passively listened to ‘Hello’ tokens presented in blocks (average duration: 15 minutes; Figure 3). Block order was counterbalanced between subjects, and there was always a short break after each block. Participants viewed a silent movie during the passive listening task in order to reduce excessive eye movements and maintain an awake state (Tervaniemi et al., 1999). The movie was projected onto a screen approximately 15 cm above the participants. Together with the head shape construction and the tone pre-test, the entire experiment lasted approximately 90 minutes.

2.5. Data Analysis

Environmental and scanner noise was removed using a multi-shift PCA noise reduction algorithm (de Cheveigné & Simon, 2007, 2008). Subsequently, epochs were averaged with a 200 ms pre-stimulus and 600 ms post-stimulus interval. Individual trials were rejected if amplitudes were higher than 3 pT or contained more than 3 consecutive eye blinks. No more than 15% of standards or deviants were excluded. Averaged data were baseline-corrected over the 200 ms pre-stimulus interval and band-pass filtered by a Hamming-window digital filter with frequency cut-offs at 1 Hz and 30 Hz.

2.5.1 MMN Amplitude Analysis—For the MMN amplitude analysis, the 10 strongest channels from the tone pre-test were selected. Subsequently, we calculated root-mean squared (RMS) values for magnetic field strengths over the selected 10 left and right-hemispheric channels. We selected two 50 ms time windows of interest in which standard and deviant responses differed for at least 25 ms. The early window (180–230 ms) corresponds to responses to the vowel [ε] while the later one (360–410 ms) represents responses to the vowel [o] of our ‘Hello’ tokens. RMS amplitudes were subsequently analyzed separately for each time-window of interest in Mixed Effect Models with *subject* as random effect (Baayen, 2008; Pinheiro & Bates, 2000). Due to the few-to-many design of the oddball paradigm, we randomly selected the same amount of standard epochs as deviant epochs (N=100) to guarantee that the RMS values were calculated from samples with equal variance. Note that the first three standards of each block and the first standard after a deviant were not used for the averaging of epochs.

2.5.2 M100 Source Analysis—Although not particularly prominent in the grand-average displayed in Figure 4, M100 responses occurred before the MMN and had the expected topographies (Poeppel & Marantz, 2000, Figure 4 left panel). M100 sources were determined on the basis of equivalent current dipoles (ECDs) as described in Obleser, Lahiri, & Eulitz (2004). First, an orthogonal left-handed head frame was defined with lateral-medial, anterior-posterior, and superior-inferior coordinates (cf. Figure 1 and 5). Then, a sphere, whose center position and radius were calculated in head frame coordinates, was fit for each participant covering the entire surface of his/her digitized head-shape. A single ECD model in a spherical volume conductor was used for source modeling analysis of the neuromagnetic data (Diesch & Luce, 1997). From a total of 16 anterior and 16 posterior channels in each hemisphere, dipole solutions only comprised source solutions with a goodness-of-fit greater than 90%. Additionally, they had to be located at least 25 mm from the mid-point of the lateral-medial axis. This led to an average exclusion rate of 15% per subject. The selection of 16 anterior and 16 posterior channels was based on the initial selection of 10 channels obtained from the tone data. The extension of the channel selection was necessary in order to balance between an over- and under-fitting of the solution to the inverse problem (Sarvas, 1987). ECD models were calculated on the rising slope of the M100 of [ε] (120–170 ms) and [o] (300–350 ms). Dipole solutions only comprised ECDs before the respective RMS peak in order to avoid tapping into changed source configuration after the peak (see Scherg, Vajsar, & Picton, 1990). ECDs were fitted separately for the left and right hemisphere, using a moving dipole model and Sarvas’ Law for the forward model (Sarvas, 1987). The resulting ECD coordinates in lateral-medial, anterior-posterior, and inferior-posterior dimensions together with dipole orientations in the axial and sagittal plane were analyzed in Mixed Effect models (Baayen, 2008).

3. Results

3.1. Amplitudes

MMN amplitudes were determined as the average RMS amplitudes in two time windows of interest (180–230 and 360–410 ms). They were calculated for standards, real deviants (across-dialects) and sham deviants (within-dialects). The within-dialect analysis comprised RMS amplitudes for standards and sham deviants, while the across-dialect analysis used RMS amplitudes for standards and real deviants. Both analyses comprised the fixed effects *Hemisphere* (left/right), *Time* (early/late), *Position* (standard/deviant), and *Dialect* (AAE/SAE) in a full-factorial design. There were no significant effects in the within-dialect analysis (all $F_s < 3$, n.s.). In contrast, in the across-dialect analysis, we found main effects for *Hemisphere* ($F(1,196) = 4.51$, $p < 0.05$), *Time* ($F(1,196) = 21.22$, $p < 0.001$), *Position* ($F(1,196) = 6.66$, $p < 0.05$) as well as *Dialect* ($F(1,196) = 4.40$, $p < 0.05$). These effects reflected that amplitudes were larger in the left hemisphere, in the late time window, for deviants, and for AAE stimuli, compared to the respective alternatives. We further found a significant interaction of *Time* $\times$ *Position* $\times$ *Dialect*, which reflected that the MMNs (as difference between deviants and standards) to SAE stimuli were larger than MMNs to AAE stimuli in the early time window (i.e. as a response to [ε]). In contrast, MMNs to AAE stimuli were larger to SAE stimuli in the late time window (i.e. as a response to [o], Figure 4).

The pattern reflected by the three-way interaction of *Time* $\times$ *Position* $\times$ *Dialect* was strongly left-lateralized, as suggested by a significant four-way interaction of *Time* $\times$ *Position* $\times$ *Dialect* $\times$ *Hemisphere* ($F(1,196) = 4.41$, $p < 0.05$).

Post-hoc comparisons showed that standard-deviant differences in the left hemisphere were significant for SAE stimuli in the early window ($t = 1.98$, $p < 0.05$) and for AAE stimuli in the late window ($t = 3.39$, $p < 0.001$), while AAE stimuli did not show differences in the first time window ($t = 0.09$, $p = 0.68$), and SAE stimuli did not show differences in the second time window ($t = 1.11$, $p = 0.27$). In the right hemisphere, standard-deviant differences were significant only for AAE stimuli in the late time window ($t = 2.25$, $p < 0.05$).

In order to show that the left-hemisphere pattern was, in fact, dependent on the deviant type, we calculated a combined analysis with the additional effect *Deviant type* (sham/real). This analysis revealed a main effect of *Deviant type* ($F(1,196) = 5.25$, $p < 0.05$). Overall, amplitudes were larger in the real deviant condition than in the sham deviant condition. Furthermore, there was a significant interaction of *Time* $\times$ *Position* $\times$ *Dialect* $\times$ *Deviant type* ($F(1,196) = 6.30$, $p < 0.05$), showing that the observed left-lateralized pattern reflected by the *Time* $\times$ *Position* $\times$ *Dialect* interaction indeed depended on the deviant type, i.e. was due to the real deviant, not the sham deviant.

3.2. Dipole sources

For the M100 corresponding to the first vowel [ε] and the second vowel [o] in ‘Hello’, we analyzed dipole coordinates and orientations in Mixed Effect Models with the fixed effects *Position* (standard/deviant), *Dialect* (AAE/SAE) and *Hemisphere* (left/right) in a full-factorial design. Analyses were computed separately for [ε] and [o] (Figure 5). Although non-averaged raw dipole locations showed rather wide-spread variation in the sagittal (posterior-anterior) and axial (inferior-posterior) dimension (Figure 6), averaged dipole locations for [ε] differed in anterior-posterior location between sources from AAE stimuli and sources from SAE stimuli ($F(1,75) = 7.11$, $p < 0.05$). SAE sources were located anterior to AAE sources. This effect was restricted to deviants (*Position* $\times$ *Dialect*: $F(1,75) = 8.77$, $p < 0.01$) and corresponded to differences in dipole orientation between the two dialects

($F(1,75) = 6.56, p < 0.05$). In the sagittal plane, SAE sources were oriented more toward anterior regions than AAE sources. This effect was left-dominant (*Dialect* $\times$ *Hemisphere*: $F(1,75) = 5.72, p < 0.05$).

In contrast, deviant dipoles in the right hemisphere were oriented more vertically for SAE exemplars than for AAE exemplars ($t = 2.50, p < 0.05$). Along the lateral-medial axis, there was a trend for *position* ($F(1,75) = 3.08, p = 0.08$), reflecting more medial sources for standards than for deviants. This effect interacted with hemisphere and was stronger in the left than in the right hemisphere ($F(1,75) = 10.86, p < 0.01$).

For [o], the spatial locations of the ECDs differed in the superior-inferior dimension: deviants had more inferior locations than standards ($F(1,70) = 4.90, p < 0.05$; Figure 3). The significant interaction *Position* $\times$ *Dialect* ($F(1,70) = 4.79, p < 0.05$) showed that locations of the ECDs of SAE and AAE deviants as well as AAE standards were all positioned inferior to SAE standards. These source location differences had corresponding orientation differences. In the sagittal plane, standards were oriented more toward inferior locations ($F(1,70) = 5.81, p < 0.01$). There was also a hemisphere effect on dipole orientation ($F(1,70) = 10.05, p < 0.01$), reflecting more inferior orientations in the right than in the left hemisphere. Finally, we found an interaction of *Position* $\times$ *Dialect* $\times$ *Hemisphere* ($F(1,70) = 6.52, p < 0.05$). This interaction was mainly driven by differences in dipole orientation between SAE and AAE deviant sources in the right hemisphere ($t = 2.59, p < 0.05$). Note that the inferior-posterior location difference between deviant and standard could also be observed for [ε], but was not significant ($F(1,75) = 1.12, n.s.$). A pooled analysis across [ε] and [o] showed that deviants were always located inferior to standards ($F(1,167) = 6.74, p < 0.05$).

As for [ε], dipole locations differed along the lateral-medial axis. ECDs for standards were located lateral to ECDs for deviants ($F(1,70) = 4.47, p < 0.05$). Again, this effect was larger in the left than right hemisphere, as seen in a significant interaction of *position* $\times$ *hemisphere* ($F(1,70) = 5.07, p < 0.05$).

4. Discussion

Our passive oddball experiment using AAE and SAE tokens of ‘Hello’ is the first neuromagnetic evidence for an early extraction of phonetic information that allows for categorical dialect perception. We found dialect-specific MMNs as a response to the first ([ε]) and second ([o]) vowel of our stimuli that were driven by the dialect-dependent phonetic differences between the ‘Hello’ tokens. The comparison between sham deviants (within-category) and real deviants (across-category) showed that MMN effects were only reliable across dialect, i.e. when the phonetic differences were accompanied by dialect differences. Although the acoustic distance to the respective standards was similarly matched for sham and real deviants, sham deviants did not elicit significant MMNs. We therefore conclude that the MMN pattern in the across-dialect conditions is based on the acoustic standard-deviant difference and an additional top-down modulation that results from long-term representations of dialectal knowledge, similar to the long-term representations for words and speech sound inventories. We propose that dialect differences between standards and deviants result in enhanced MMNs, parallel to previous interpretations of MMN data showing that phonemic differences between standards and deviants increased the MMN response (Näätänen, et al., 1997). Thus, it seems that dialect extraction from auditory speech input occurs relatively quickly and does not require attention (Beauchemin, et al., 2006; Knösche, et al., 2002; Titova & Näätänen, 2001), a finding that is consistent with recent neurophysiological data on social perception (Cunningham, et al., 2004; Hart, et al., 2000; Ito & Bartholow, 2009). Could we interpret

our results solely on the basis of acoustic standard-deviant differences? As may be noted by looking at the acoustic vowel differences (illustrated in Figure 2 C), these differences seem to be well aligned with the corresponding MMN responses (Figure 4) in that larger acoustic distances result in larger MMN amplitudes. However, by using variable acoustic tokens as standards in our experiment, at some level of processing, there should have been an abstraction from low-level acoustic stimulus differences, as suggested by Philips et al. (2000). We conjecture that multiple acoustic standard tokens evoked a phonetically based memory trace against which the deviant was compared. Note further that our crucial comparison is between the sham and the real deviant with similar acoustic distance to the respective standards. We propose that the larger effect for the real deviant is due to the extraction of a salient phonetic property (frontness/backness) that cues the categorical dialect difference. Our M100 source localization results support this interpretation, since we found a displacement for the dipole sources in response to the first vowel along the posterior-anterior dimension. There is independent evidence that this dimension codes phonetic differences along the front-back dimension (Obleser, Elbert, Lahiri, & Eulitz, 2003; Obleser, Lahiri, et al., 2004).

One potential alternative explanation for our results is that the tokens produced in the SAE dialect are more familiar to our SAE speaking participants than tokens produced in the AAE dialect. Effects of familiarity have been previously found in oddball and mismatch negativity designs for both linguistic and non-linguistic stimuli. Jacobsen, et al. (2005) reported larger MMN amplitudes when deviants occurred embedded within a series of non-linguistic familiar standards (e.g., relatively common environmental sounds) as compared to unfamiliar standards (i.e., the same environmental sounds played backward). Kirmse, Jacobsen, & Schröger (2009) found an enhanced N1, P2 and additional P250 component in the processing of a non-linguistic, yet familiar sound (i.e., a common animal sound) compared to acoustically matched, spectrally complex, unfamiliar sound outside the attentional focus of participants. Using more linguistically-relevant stimuli, Jacobsen, et al. (2004) presented German and Hungarian participants with auditory words in German and Hungarian in an oddball mismatch design. For the German participants, the Hungarian words were phonotactically legal, but non-lexical and semantically meaningless. The corollary was true of the German words for the Hungarian participants. While an MMN was elicited in all cases, i.e., unfamiliar deviants (e.g., German words for Hungarian participants) embedded within a context of familiar standards (e.g., German words for German participants) and vice versa, the MMN was larger when the standards were familiar to the participants than when they were unfamiliar. The magnitude of the MMN has also been shown to be modulated by whether the deviant is produced by a voice familiar to the participant. Beauchemin, et al. (2006) found larger MMN responses to a familiar deviant [a], e.g., produced by a close friend or relative, compared to an unfamiliar deviant [a], e.g., produced by an unfamiliar talker, infrequently occurring in a string of standard tokens of /a/ produced by a unique, unfamiliar talker. In the current experiment, the MMN to SAE deviants was larger than the MMN to AAE deviants in the earlier time-window, consistent with the results from Beauchemin, et al. (2006), whereby familiar tokens (SAE deviants with SAE participants) elicit a larger response than unfamiliar tokens (AAE deviants with SAE participants). Additionally, when AAE was the standard, sham deviants (AAE, but acoustically distant) elicited smaller MMNs than real deviants (SAE), again suggesting that familiarity might play a role in the explanation of the current data. There are, however, a few caveats with respect to this particular explanation: First, the directionality of the magnitude of the MMN is reversed in the later, second time-window (360–410 ms), i.e., AAE deviants evoked a larger MMN than SAE deviants. Second, because tokens were produced by the same talker, unfamiliar to participants, the results cannot be due to familiarity or personal significance (see Roye, Jacobsen, & Schröger, 2007; Roye, Schröger, Jacobsen, & Gruber, 2010) of the talker in a manner similar to Beauchemin, et al. (2006). Instead, familiarity

would have to be posited at a higher-level of abstraction, e.g., dialect/phonetic, a hypothesis consistent with our conclusions but requires additional empirical testing. Finally, when SAE was the standard, we might expect the AAE deviants to elicit a smaller response than the sham-SAE deviants, since the SAE deviants would be more familiar to participants. This was not the case. Consequently, while we cannot rule-out a role for familiarity, it cannot account for the entire pattern of data reported here and requires further research.

Another important result is that categorical dialect distinctions do not seem to require differences in speaker – the same bidialectal speaker produced the SAE and AAE stimuli used in our study. This suggests that auditory speech input is analyzed not only with regard to speaker voice properties, but also with regard to which social group these speaker voice properties belong, perhaps moment-by-moment, as in the case of dialect code-switching. If, in fact, the latter knowledge is determined by exposure to one or the other dialect, we can account for the vowel-specific MMN pattern in our study: SAE, but not AAE deviants, elicited an early MMN which we attributed to the first vowel [ɛ] in ‘Hello’. In contrast, AAE, but not SAE deviants elicited a late MMN that we attributed to the second vowel [o]. It seems plausible that on the basis of our predominantly SAE participants, this observation corresponds to an in-group/out-group effect (Hart, et al., 2000). This effect manifested itself in an early MMN if the deviant was an in-group exemplar, followed by a series of out-group exemplars, but not if the deviant was an out-group-exemplar, followed by a series of in-group exemplars. Our SAE participants obviously needed only very little information from the SAE deviant in order to show a categorical mismatch response. On the other hand, participants needed more acoustic information if the deviants were from the AAE (out-group) category. As a result, the mismatch response was delayed until the following salient cue, the second vowel. While this explanation can only be a speculative hypothesis at the moment, it makes very interesting predictions for future research. There are, however, further indications in our study for the validity of the in-group/out-group hypothesis. First, while the early MMN as response to the SAE vowel [ɛ] was restricted to the left hemisphere, the later MMN on the AAE vowel [o] was significant in both hemispheres, thus, crucially involved right-hemispheric sources. It has been shown previously that voice-selective areas and neuronal correlates of speaker identity are predominantly found in the right STS (von Kriegstein, et al., 2003; von Kriegstein & Giraud, 2004). It is therefore possible that the AAE deviants involved the recruitment of additional neuronal sources in the right hemisphere in order to process speaker identity, and on top of that, speaker-related dialect. Our source dipole analyses support this hypothesis. We found that source dipoles for [o] were at more inferior locations if measured from deviants than if measured from standards. The inferior-superior displacement of source dipoles has been previously associated with the additional processing in STS areas (Obleser, Elbert, & Eulitz, 2004). Therefore, more inferior dipole locations may indicate the additional processing of speaker identity in STS. The fact that we observe this shift for deviants suggests that the change from standard to deviant is accompanied by the processing of speaker identity and speaker dialect. This is further supported by the pattern of source dipole orientations for [ɛ], but particularly for [o]. In the sagittal plane, AAE deviant dipoles elicited by [o] were oriented more vertically (i.e. were more aligned with the inferior-superior axis) than were SAE deviant dipoles. In contrast, AAE deviant dipoles elicited by [ɛ] were oriented less vertically than their SAE counterparts. Both effects were restricted to the right hemisphere, suggesting that the processing of SAE [ɛ] or AAE [o] in deviant position additionally involves voice- (or, by extension, dialect) selective areas. Therefore, these vowels were selectively salient for one or the other dialect, which was also reflected in the MMN amplitude pattern. Note that the dipole location difference between SAE and AAE [ɛ] along the anterior-posterior axis is reminiscent of a finding by Obleser and colleagues (Obleser, Elbert, Lahiri, & Eulitz, 2003; Obleser, Lahiri, et al., 2004). They observed that M100 sources were located at more anterior positions in auditory cortex if they were elicited by speech sounds produced further

front in the mouth. Here, we see that SAE [ɛ] elicited more anterior source dipoles than AAE [ɛ], reflecting the main acoustic differences between the two vowels. Crucially, SAE [ɛ] is fronted relative to AAE [ɛ], i.e. produced further front in the mouth than AAE [ɛ] (Figure 2 & 5).

The linguistic differences in dialect, especially in this experiment, are clearly phonetic in nature, and it is these phonetic differences that reliably cue the racial identity of a speaker (see Purnell, et al. 1999). Advances within the cognitive neurosciences have facilitated studying the biological and neurophysiological underpinnings of race perception (see Ito & Bartholow, 2009 for a review), although race perception mediated through speaker voice has not yet been explored to the same degree as racial perception mediated through face perception, presumably due to the less arbitrary, biologically-based racially correlated differences in facial pigmentation and morphology.

5. Conclusion

On the basis of our findings, we propose that cues for dialectal background information are rapidly extracted from spoken words and processed in the same way as speaker voice information. Bottom-up acoustic information is then integrated with top-down knowledge of speaker and dialect categories in long-term memory. Note that attribution of a speaker to a dialectal community is necessarily mediated by speaker identity. This identity is partially defined by the dialectal background. Thus, our speaker takes different identities depending on whether he used his SAE or AAE guise. Consequently, it is reasonable to assume that neuronal processing of dialect categories shares or is equivalent to voice selective areas (Belin, 2006; Belin, et al., 2004; Belin & Grosbras, 2010; Belin & Zatorre, 2003; Belin, et al., 2000; von Kriegstein, et al., 2003; von Kriegstein & Giraud, 2004). Given that these areas have a long evolutionary history (Belin, 2006; Fecteau, et al., 2004; Petkov, et al., 2008), it is not surprising that they analyze auditory speech signals very robustly and very quickly. Despite the evolutionary and developmental (Belin & Grosbras, 2010) specificity of cerebral voice areas, enabling adults to quickly extract speaker identity, dialectal and racial information from spoken words, it is important to note that these early pre-attentive categorical responses can still be controlled (Wheeler & Fiske, 2005). Thus, human speakers must rely on later, higher-order, conscious cognitive faculties to eliminate any potential prejudice resulting from initial racial identification.

Acknowledgments

We are especially indebted to John Baugh and Thomas Purnell who provided us with the stimulus materials for our experiment and for their helpful discussion. We further thank David Poeppel, Thomas Bever, David Kleinschmidt, Alan Mishler, and the audience of the 2010 CNS Annual Meeting for invaluable comments on this research. The research for this study was funded by NIH grant 7ROIDC005660-07 to W.J.I. & David Poeppel and carried out while M.S. was affiliated with the University of Maryland, College Park.

References

- Ackermann H, Hertrich I, Mathiak K, Lutzenberger W. Contralaterality of cortical auditory processing at the level of the M50/M100 complex and the mismatch field: A whole-head magnetoencephalography study. *Neuroreport*. 2001; 12:1683–1687. [PubMed: 11409739]
- Alain C, Cortese F, Picton TW. Event-related brain activity associated with auditory pattern processing. *Neuroreport*. 1998; 9:3537–3541. [PubMed: 9855313]
- Alho K, Winkler I, Escera C, Huottilainen M, Virtanen J, Jääskeläinen IP, Pekkonen E, et al. Processing of novel sounds and frequency changes in the human auditory cortex: magnetoencephalographic recordings. *Psychophysiology*. 1998; 35:211–224. [PubMed: 9529947]

- Baayen, H. *Analyzing Linguistic Data: A Practical Introduction to Statistics using R*. Cambridge: Cambridge University Press; 2008.
- Baugh J. Racial identification by speech. *Am. Speech*. 2000; 75:362–364.
- Beauchamp MS, Argall BD, Bodurka J, Duyn JH, Martin A. Unraveling multisensory integration: Patchy organization within human STS multisensory cortex. *Nat. Neurosci*. 2004; 7:1190–1192. [PubMed: 15475952]
- Beauchemin M, Beaumont Ld, Vannasing P, Turcotte A, Arcand C, Belin P, Lassonde M. Electrophysical markers of voice familiarity. *Eur. J. Neurosci*. 2006; 23:3081–3086. [PubMed: 16819998]
- Belin P. Voice processing in human and non-human primates. *Philos. Trans. R. Soc. Lond. B Biol. Sci*. 2006; 361:2091–2107. [PubMed: 17118926]
- Belin P, Fecteau S, Bédard C. Thinking the voice: Neural correlates of voice perception. *Trends Cogn. Sci*. 2004; 8:129–135. [PubMed: 15301753]
- Belin P, Grosbras M-H. Before speech: Cerebral voice processing in infants. *Neuron*. 2010; 65:733–735. [PubMed: 20346748]
- Belin P, Zatorre RJ. Adaptation to speaker's voice in right anterior temporal lobe. *NeuroReport*. 2003; 14:2105–2105. [PubMed: 14600506]
- Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B. Voice-selective areas in human auditory cortex. *Nature*. 2000; 403:309–312. [PubMed: 10659849]
- Boersma, P.; Weenink, D. PRAAT: Doing Phonetics by Computer (ver. 5.1.0). Amsterdam: Institut for Phonetic Sciences; 2009.
- Conrey B, Potts GF, Niedzielski NA. Effects of dialect on merger perception: ERP and behavioural correlates. *Brain Lang*. 2005; 95:435–449. [PubMed: 16099494]
- Csépe V, Karmos G, Molnár M. Evoked potential correlates of stimulus deviance during wakefulness and sleep in cat--animal model of mismatch negativity. *Clin. Neurophysiol*. 1987; 66:571–578.
- Cunningham WA, Johnson MK, Raye CL, Gatenby JC, Gore JC, Banaji MR. Separable neural components in the processing of black and white faces. *Psychol. Sci*. 2004; 15:806–806. [PubMed: 15563325]
- de Cheveigné, Ad; Simon, JZ. Denoising based on time-shift PCA. *J. Neurosci. Methods*. 2007; 165:297–305. [PubMed: 17624443]
- de Cheveigné, Ad; Simon, JZ. Sensor noise suppression. *J. Neurosci. Methods*. 2008; 168:195–202. [PubMed: 17963844]
- Diesch E, Luce T. Magnetic fields elicited by tones and vowel formants reveal tonotopy and nonlinear summation of cortical activation. *Psychophysiology*. 1997; 34:501–510. [PubMed: 9299904]
- Diesch E, Eulitz C, Hampson S, Ross B. The neurotopography of vowels as mirrored by evoked magnetic field measurements. *Brain Lang*. 1996; 53:143–168. [PubMed: 8726531]
- Eulitz C, Diesch E, Pantev C, Hampson S, Elbert T. Magnetic and electric brain activity evoked by the processing of tone and vowel stimuli. *J. Neurosci*. 1995; 15:2748–2755. [PubMed: 7722626]
- Fecteau S, Armony JL, Joannette Y, Belin P. Is voice processing species-specific in human auditory cortex? An fMRI study. *Neuroimage*. 2004; 23:840–848. [PubMed: 15528084]
- Formisano E, De Martino F, Bonte ML, Goebel R. "Who" is saying "what"? Brain-based decoding of human voice and speech. *Science*. 2008; 322:970–973. [PubMed: 18988858]
- Ghazanfar AA, Chandrasekaran C, Logothetis NK. Interactions between the Superior Temporal Sulcus and Auditory Cortex mediate dynamic face/voice integration in Rhesus Monkeys. *J. Neurosci*. 2008; 28:4457–4469. [PubMed: 18434524]
- Gomes H, Ritter W, Vaughan HG. The Nature of preattentive storage in the auditory system. *J. Cog. Neurosci*. 1995; 7:81–94.
- Hari R, Aittoniemi K, Järvinen ML, Katila T, Varpula T. Auditory evoked transient and sustained magnetic fields of the human brain. Localization of neural generators. *Exp. Brain Res*. 1980; 40:237–240. [PubMed: 7428878]
- Hari R, Hämäläinen M, Ilmoniemi R, Kaukoranta E, Reinikainen K, Salminen J, Alho K, et al. Responses of the primary auditory cortex to pitch changes in a sequence of tone pips: neuromagnetic recordings in man. *Neurosci. Lett*. 1984; 50:127–132. [PubMed: 6493619]

- Hart AJ, Whalen PJ, Shin LM, McInerney SC, Fischer H, Rauch SL. Differential response in the human amygdala to racial outgroup vs ingroup face stimuli. *NeuroReport*. 2000; 11:2351–2351. [PubMed: 10943684]
- Hartman DE. The perceptual identity and characteristics of aging in normal male adult speakers. *J. Commun. Disord*. 1979; 12:53–61. [PubMed: 422747]
- Hartman DE, Danhauer JL. Perceptual features of speech for males in four perceived age decades. *J. Acoust. Soc. Am*. 1976; 59:713–713. [PubMed: 1254800]
- Ito TA, Bartholow BD. The neural correlates of race. *Trends Cogn. Sci*. 2009; 13:524–531. [PubMed: 19896410]
- Jääskeläinen IP, Ahveninen J, Bonmassar G, Dale AM, Ilmoniemi RJ, Levänen S, Lin FH, May P, Melcher J, Stufflebeam S, et al. Human posterior auditory cortex gates novel sounds to consciousness. *Proc. Natl. Acad. Sci. U. S. A*. 2004; 101:6809–6814. [PubMed: 15096618]
- Jacobsen T, Horváth J, Schröger E, Lattner S, Widmann A, Winkler I. Pre-attentive auditory processing of lexicality. *Brain Lang*. 2004; 88:54–67. [PubMed: 14698731]
- Jacobsen T, Schröger E, Winkler I, Horváth J. Familiarity affects the processing of task-irrelevant auditory deviance. *J Cogn Neurosci*. 2005; 17:1704–1713. [PubMed: 16269107]
- Javitt DC, Schroeder CE, Steinschneider M, Arezzo JC, Vaughan HG. Demonstration of mismatch negativity in the monkey. *Clin. Neurophysiol*. 1992; 83:87–90.
- Javitt DC, Steinschneider M, Schroeder CE, Vaughan HG, Arezzo JC. Detection of stimulus deviance within primate primary auditory cortex: intracortical mechanisms of mismatch negativity (MMN) generation. *Brain Res*. 1994; 667:192–200. [PubMed: 7697356]
- Kazanina N, Phillips C, Idsardi W. The influence of meaning on the perception of speech sounds. *Proc. Natl. Acad. Sci. U. S. A*. 2006; 103:11381–11386. [PubMed: 16849423]
- Kirmse U, Jacobsen T, Schröger E. Familiarity affects environmental sound processing outside the focus of attention: An event-related potential study. *Clin Neurophysiol*. 2009; 120:887–896. [PubMed: 19345610]
- Knösche TR, Lattner S, Maess B, Schauer M, Friederici AD. Early parallel processing of auditory word and voice information. *Neuroimage*. 2002; 17:1493–1503. [PubMed: 12414288]
- Kutas, M.; Petten, Cv; Kluender, R. Psycholinguistics Electrified II (1994–2005). In: Taxler, M.; Gernsbacher, MA., editors. *Handbook of Psycholinguistics*. New York: Academic Press; 2006. p. 659–724.
- Lütkenhöner B, Steinsträter O. High-precision neuromagnetic study of the functional organization of the human auditory cortex. *Audiol. Neurotol*. 1998; 3:191–213. [PubMed: 9575385]
- Maess B, Jacobsen T, Schröger E, Friederici AD. Localizing pre-attentive auditory memory-based comparison: magnetic mismatch negativity to pitch change. *NeuroImage*. 2007; 37:561–571. [PubMed: 17596966]
- Mullennix JW, Johnson KA, Topcu-Durgun M, Farnsworth LM. The perceptual representation of voice gender. *J. Acoust. Soc. Am*. 1995; 98:3080–3080. [PubMed: 8550934]
- Näätänen, R. *Attention and Brain Function*. Hillsdale, NJ: Erlbaum; 1992.
- Näätänen R. The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology*. 2001; 38:1–21. [PubMed: 11321610]
- Näätänen R, Alho K. Generators of electrical and magnetic mismatch responses in humans. *Brain Topography*. 1995; 7(4):315–320. [PubMed: 7577329]
- Näätänen R, Alho K. Mismatch negativity (MMN) - the measure for central sound representation accuracy. *Audiol. Neurotol*. 1997; 2:341–353. [PubMed: 9390839]
- Näätänen R, Lehtokoski A, Lennes M, Cheour M, Huottilainen M, Ilvonen A, Vainio M, Alku P, Ilmoniemi RJ, Luuk A, Allik J, Sinkkonen J, Alho K. Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*. 1997; 385:432–434. [PubMed: 9009189]
- Näätänen R, Paavilainen P, Rinne T, Alho K. The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clin. Neurophysiol*. 2007; 118:2544–2590. [PubMed: 17931964]

- Näätänen R, Pakarinen S, Rinne T, Takegata R. The mismatch negativity (MMN): Towards the optimal paradigm. *Clin. Neurophysiol.* 2004; 115:140–144. [PubMed: 14706481]
- Näätänen R, Picton T. The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. *Psychophysiology.* 1987; 24:375–425. [PubMed: 3615753]
- Obleser J, Elbert T, Eulitz C. Attentional influences on functional mapping of speech sounds in human auditory cortex. *BMC Neurosci.* 2004a; 5:1–24.
- Obleser J, Lahiri A, Eulitz C. Magnetic brain response mirrors extraction of phonological features from spoken vowels. *J. Cogn. Neurosci.* 2004b; 16:31–39. [PubMed: 15006034]
- Obleser J, Elbert T, Lahiri A, Eulitz C. Cortical representation of vowels reflects acoustic dissimilarity determined by formant frequencies. *Brain Res. Cogn. Brain Res.* 2003; 15:207–213. [PubMed: 12527095]
- Oldfield RC. The assessment and analysis of handedness: The Edinburgh Inventory. *Neuropsychologia.* 1971; 9:97–113. [PubMed: 5146491]
- Pantev C, Hoke M, Lehnertz K, Lütkenhöner B. Neuromagnetic evidence of an amplitopic organization of the human auditory cortex. *Clin. Neurophysiol.* 1989; 72:225–231.
- Patterson, RD.; Smith, DRR.; Dinther, R.; Walters, TC. Size information in the production and perception of communication sounds. In: Yost, WA.; Popper, AN.; Fay, RR., editors. *Auditory Perception of Sound Sources*. New York: Springer; 2008. p. 43-75.
- Pazo-Alvarez P, Cadaveira F, Amenedo E. MMN in the visual modality: a review. *Biol. Psychol.* 2003; 63:199–236. [PubMed: 12853168]
- Petkov CI, Kayser C, Steudel T, Whittingstall K, Augath M, Logothetis NK. A voice region in the monkey brain. *Nat. Neurosci.* 2008; 11:367–374. [PubMed: 18264095]
- Phillips C, Pellathy T, Marantz A, Yellin E, Wexler K, Poeppel D, McGinnis M, Roberts T. Auditory cortex accesses phonological categories: An MEG mismatch study. *J. Cogn. Neurosci.* 2000; 12:1038–1105. [PubMed: 11177423]
- Phillips ML, Young AW, Scott SK, Calder AJ, Andrew C, Giampietro V, Williams SC, Bullmore ET, Brammer M, Gray JA. Neural responses to facial and vocal expressions of fear and disgust. *Proc. R. Soc. Lond. B Biol. Sci.* 1998; 265:1809–1809.
- Picton TW, Woods DL, Baribeau-Braun J, Healey TM. Evoked potential audiometry. *J. Otolaryngol.* 1976; 62:90–119. [PubMed: 1030745]
- Pinheiro, JC.; Bates, DM. *Mixed-effects models in S and S-PLUS*. Springer Verlag: 2000.
- Poeppel D, Phillips C, Yellin E, Rowley HA, Roberts TPL, Marantz A. Processing of vowels in supratemporal auditory cortex. *Neurosci. Lett.* 1997; 221:145–148. [PubMed: 9121685]
- Poeppel, D.; Marantz, A. Cognitive neuroscience of speech processing. In: Marantz, A.; Miyashita, Y.; O'Neil, W., editors. *Image, Language, Brain*. Cambridge, MA: MIT Press; 2000. p. 29-49.
- Pulvermüller F, Kujala T, Shtyrov Y, Simola J, Tiitinen H, Alku P, Alho K, Martinkauppi S, Illmoniemi RJ, Näätänen R. Memory traces for words as revealed by the mismatch negativity. *Neuroimage.* 2001; 14:607–616. [PubMed: 11506534]
- Pulvermüller F, Shtyrov Y, Kujala T, Näätänen R. Word-specific cortical activity as revealed by the mismatch negativity. *Psychophysiology.* 2004; 41:106–112. [PubMed: 14693005]
- Purnell T, Idsardi WJ, Baugh J. Perceptual and phonetic experiments on American English dialect identification. *J. Lang. Soc. Psychol.* 1999; 18:10–30.
- Rendall D, Owren MJ, Rodman PS. The role of vocal tract filtering in identity cueing in rhesus monkey (*Macaca mulatta*) vocalizations. *J. Acoust. Soc. Am.* 1998; 103:602–602. [PubMed: 9440345]
- Rendall D, Rodman PS, Emond RE. Vocal recognition of individuals and kin in free-ranging rhesus monkeys. *Anim. Behav.* 1996; 51:1007–1016.
- Ritter W, Paavilainen P, Lavikainen J, Reinikainen K, Alho K, Sams M, Näätänen R. Event-related potentials to repetition and change of auditory stimuli. *Clin. Neurophysiol.* 1992; 83:306–321.
- Ritter W, Ruchkin DS. A review of event-related potential components discovered in the context of studying P3. *Annals of the New York Academy of Sciences.* 1992; 658:1–32. [PubMed: 1497254]

- Roberts TPL, Ferrari P, Stufflebeam SM, Poeppel D. Latency of the auditory evoked neuromagnetic field components: stimulus dependence and insights toward perception. *J. Clin. Neurophysiol.* 2000; 17:114–129. [PubMed: 10831104]
- Roberts TPL, Flagge EJ, Gage NM. Vowel categorization induces departure of M100 latency from acoustic prediction. *Neuroreport.* 2004; 15:1679–1682. [PubMed: 15232306]
- Roberts TPL, Poeppel D. Latency of auditory evoked M100 as a function of tone frequency. *Neuroreport.* 1996; 7:1138–1140. [PubMed: 8817518]
- Roye A, Jacobsen T, Schröger E. Personal significance is encoded automatically by the human brain: An event-related potential study with ringtones. *Eur. J. Neurosci.* 2007; 26:784–790. [PubMed: 17634070]
- Roye A, Schröger E, Jacobsen T, Gruber T. Is my mobile ringing? Evidence for rapid processing of a personally significant sound in humans. *J. Neurosci.* 2010; 30:7310–7313. [PubMed: 20505097]
- Sams M, Paavilainen P, Alho K, Näätänen R. Auditory frequency discrimination and event-related potentials. *Clin. Neurophysiol.* 1985; 62:437–448.
- Sharma A, Dorman MF. Cortical auditory evoked potential correlates of categorical perception of voice-onset time. *J. Acoust. Soc. Am.* 1999; 106:1078–1083. [PubMed: 10462812]
- Sharma A, Dorman MF. Neurophysiologic correlates of cross-language phonetic perception. *J. Acoust. Soc. Am.* 2000; 107:2697–2703. [PubMed: 10830391]
- Sharma A, Kraus N, McGee T, Carrell T, Nicol T. Acoustic versus phonetic representation of speech as reflected by the mismatch negativity event-related potential. *Clin. Neurophysiol.* 1993; 88:64–71.
- Shtyrov Y, Pulvermüller F. Memory traces for inflectional affixes as shown by mismatch negativity. *Eur. J. Neurosci.* 2002; 15:1085–1091. [PubMed: 11918667]
- Smith DRR, Patterson RD, Turner R, Kawahara H, Irino T. The processing and perception of size information in speech sounds. *J. Acoust. Soc. Am.* 2005; 117:305–318. [PubMed: 15704423]
- Stevens AA. Dissociating the cortical basis of memory for voices, words and tones. *Brain Res. Cogn. Brain Res.* 2004; 18:162–171. [PubMed: 14736575]
- Tervaniemi M, Kujala A, Alho K, Virtanen J, Ilmoniemi RJ, Näätänen R. Functional Specialization of the Human Auditory Cortex in Processing Phonetic and Musical Sounds: A Magnetoencephalographic (MEG) Study. *NeuroImage.* 1999; 9:330–336. [PubMed: 10075902]
- Tiitinen H, Mäkelä AM, Mäkinen V, May PJC, Alku P. Disentangling the effects of phonation and articulation: Hemispheric asymmetries in the auditory N1m response of the human brain. *BMC Neurosci.* 2005; 6:62. [PubMed: 16225699]
- Titova N, Näätänen R. Preattentive voice discrimination by the human brain as indexed by the mismatch negativity. *Neurosci. Lett.* 2001; 308:63–65. [PubMed: 11445287]
- van Dommelen WA, Moxness BH. Acoustic parameters in speaker height and weight identification: Sex-specific behaviour. *Lang. Speech.* 1995; 38:267–267. [PubMed: 8816083]
- Vasama JP, Mäkelä JP, Tissari SO, Hämäläinen MS. Effects of intensity variation on human auditory evoked magnetic fields. *Acta Oto-Laryngologica.* 1995; 115:616–621. [PubMed: 8928632]
- Virtanen J, Ahveninen J, Ilmoniemi RJ, Näätänen R, Pekkonen E. Replicability of MEG and EEG measures of the auditory N1/N1m-response. *Clin. Neurophysiol.* 1998; 108:291–298.
- von Kriegstein K, Eger E, Kleinschmidt A, Giraud AL. Modulation of neural responses to speech by directing attention to voices or verbal content. *Brain Res. Cogn. Brain Res.* 2003; 17:48–55. [PubMed: 12763191]
- von Kriegstein K, Giraud AL. Distinct functional substrates along the right superior temporal sulcus for the processing of voices. *Neuroimage.* 2004; 22:948–955. [PubMed: 15193626]
- Wheeler ME, Fiske ST. Controlling racial prejudice. *Psychol. Sci.* 2005; 16:56–63. [PubMed: 15660852]
- Winkler I, Paavilainen P, Alho K, Reinikainen K, Sams M, Näätänen R. The effect of small variation of the frequent auditory stimulus on the event-related brain potential to the infrequent stimulus. *Psychophysiology.* 1990; 27:228–235. [PubMed: 2247552]
- Yabe H, Tervaniemi M, Reinikainen K, Näätänen R. Temporal window of integration revealed by MMN to sound omission. *Neuroreport.* 1997; 8:1971–1974. [PubMed: 9223087]

Figure 1. Brain topography (left hemisphere). Relevant areas discussed in the text are indexed in the order of appearance.

A. Average segment durations in 'Hello'

B. F1/F2 differences of [ε]

C. Average differences between deviants and standards

Real deviant vs. standards

Sham deviant vs. standards

Figure 2.

Acoustic properties of 'Hello' tokens. Panel A shows the average duration of segments across all stimuli. Panel B illustrates the fronting effect of SAE [ε] compared to AAE [ε]. It is produced slightly more front in the mouth. Panel C displays the first three resonant frequency contrasts between real deviants and standards, and sham deviants and standards.

Experimental design

Real deviant (across dialect)

Sham deviant (within dialect)

Figure 3. Design of MMN experiment. Standards and deviants are presented in two conditions: (1) real deviants (across-dialect) and (2) sham deviants (within-dialect, see Text for details).

Figure 4.

Representative topographies of the M100, the first, and the second MMN (left panel) with channel selection indicated with black dots. Grand mean of RMS amplitudes for SAE and AAE stimuli in the across-dialect condition are shown in the top middle and top right panel, while responses for SAE and AAE stimuli in the within-dialect condition are displayed in the bottom middle and bottom right panel. Deviant responses are given in red, standard responses in blue.

Dipole source locations

Figure 5. Dipole source location in the sagittal plane. The left graph shows the source differences between SAE and AAE [ε]. The right graph displays the differences between standard and deviant locations for the vowel [o] across dialects. Both graphs are based on averages across hemispheres. Whiskers indicate one standard error of the mean.

Figure 6.

Individual ECD coordinates along the anterior-posterior and the inferior-superior axes. The left panel shows ECDs in the left hemisphere for the dialects AAE and SAE, the vowels “o” [o] and “e” [ɛ] in standard (blue) and deviant (red) position. Correspondingly, the right panel illustrates dipoles in the right hemisphere.

Table 1

Averaged segment properties of AAE and SAE ‘Hello’ tokens. Standard errors of the mean are given in parentheses.

Dialect	Segment	Length [ms]	Intensity [dB]	Pitch [Hz]
AAE	[h]	117 (6.86)	47 (3.62)	100 (2.51)
	[e]	74 (1.80)	70 (1.00)	106 (1.44)
	[l]	62 (3.61)	74 (0.27)	105 (2.08)
	[o]	270 (11.41)	70 (0.32)	100 (1.90)
SAE	[h]	100 (8.21)	45 (2.64)	98 (2.63)
	[e]	66 (2.42)	68 (0.92)	110 (3.18)
	[l]	60 (2.99)	74 (0.28)	104 (1.49)
	[o]	257 (7.25)	70 (0.19)	110 (1.12)