A R T I C L E   I N F O

Article history:
Received 29 October 2010
Accepted 4 April 2011
Available online 12 April 2011

Keywords:
Speaker identity
Dialect perception
Cerebral voice areas
Mismatch Negativity (MMN)
Magnetoencephalography (MEG)

A B S T R A C T

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 dialectal 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.

© 2011 Elsevier Inc. All rights reserved.

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 and 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 as gender (Mullennix et al., 1995), perceived age (Hartman, 1979; Hartman and Danhauer, 1976), or estimates of body size (Patterson et al., 2008; Smith et al., 2005; van Dommelen and Moxness, 1995). Brain imaging studies also support a dissociation of voice-related properties (‘who’) and linguistic information (‘what’) in the speech signal (Formisano et al., 2008; Knösche et al., 2002; Stevens, 2004). So-called ‘voice selective’ areas have been identified along the upper bank of the superior temporal sulcus (STS; Fig. 1; area 1), particularly at medial and anterior locations (Belin et al., 2004; Belin et al., 2000), showing increased activity in response to speaker voice characteristics. Neuronal correlates of speaker identity seem to lateralize to right STS (von Kriegstein et al., 2003; von Kriegstein and Giraud, 2004). This area showed habituation to repeated voices and strong activation in response to unfamiliar voices (Belin and Zatorre, 2003). Neuronal voice processing is often compared to face perception in that vocal emotions similarly activate the amygdala (Phillips et al., 1998, Fig. 1; area 2) and voice-selective areas show species-specificity effects in both human (Fecteau et al., 2004) and non-human primates (Belin, 2006; Petkov et al., 2008; Randall et al., 1998; Randall et al., 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 et al., 2004; Ghazanfar et al., 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 et al., 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 and 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 unchanged stimulus (see Pazó-Alvarez et al., 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) (Fig. 1; area 3; cf. Alain et al., 1998; Csépe et al., 1987; Hari et al., 1984; Javitt et al., 1992; Javitt et al., 1994; Näätänen, 2001; Näätänen and Alho, 1997; see Näätänen and Alho,

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

Mathias Scharinger a,c,⁎, Philip J. Monahan b, William J. Idsardi a

a Department of Linguistics, University of Maryland, College Park, MD, USA
b Basque Center on Cognition, Brain and Language (BCBL), Donostia-San Sebastián, Spain
c Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

⁎ Corresponding author at: Max Planck Institute for Human Cognitive & Brains Sciences, Stephanstr. 1A-04103 Leipzig, Germany.
E-mail address: mscharinger@cbs.mpg.de (M. Scharinger).

1053-8119/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2011.04.007

Available online 12 April 2011

Accepted 4 April 2011

Revised 22 March 2011

Article history:

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

youhadmeat”Hello”:Rapidextractionofdialectinformationfromspokenwords
showed statistically significant abilities to identify racially affiliated dialects from very short (500 ms or less) samples of speech.

Materials and methods

Motivation

We used the utterance-initial ‘Hello’ tokens from Purnell et al. (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 et al., 2006).

Auditory stimuli elicit distinct cortical brain responses with characteristic topographies and latencies (Ackermann et al., 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 and Picton, 1987; Poeppel and Marantz, 2000). The electrical N1 in EEG is a negative-going potential, with a stimulus onset (Näätänen and Picton, 1987; Poeppel and Marantz, 2000).

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 et al., 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 apartment search over the telephone in different neighborhoods (Purnell et al., 1999), the multi-dialectal J. Baugh conducted an exploration of primary auditory cortex posterior to Heschl’s Gyrus. In good signal-to-noise conditions, its localization reflects the tonotopic organization of primary auditory cortex (Lütkenhöner and Steinsträter, 1998). Equivalent current dipole (ECD) modeling of the sources of the M100 has also revealed an ampliotopic organization (Pantev et al., 1989; cf., Vasamä et al., 1995). Its evoked latency is modulated as a function of stimulus frequency in response to sinusoidal stimuli (Roberts and Poeppel, 1996) and predictably varies with the first formant (F1) in vowels (Diesch et al., 1996; Poeppel et al., 1997; Roberts et al., 2004; Tiitinen et al., 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 (Phillips 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 (Shytrov and Pulvermüller, 2002), or words (Pulvermüller et al., 2001, 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 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).

Materials

Twenty 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 and 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 Fig. 2A.

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 (Fig. 2B). 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 (Fig. 2C). 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, and F3) of the two vowels [ɛ] and [o] (Fig. 2C).

Methods

SAE and AAE ‘Hello’ tokens were arranged in a passive oddball-paradigm (Näätänen et al., 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 Fig. 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.

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 in 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 supratemporal 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 min; Fig. 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

<table>
<thead>
<tr>
<th>Dialect</th>
<th>Segment</th>
<th>Length [ms]</th>
<th>Intensity [dB]</th>
<th>Pitch [Hz]</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAE</td>
<td>[h]</td>
<td>117 (6.86)</td>
<td>47 (3.62)</td>
<td>100 (2.51)</td>
</tr>
<tr>
<td></td>
<td>[ɛ]</td>
<td>74 (1.80)</td>
<td>70 (1.00)</td>
<td>106 (1.44)</td>
</tr>
<tr>
<td></td>
<td>[l]</td>
<td>62 (3.61)</td>
<td>74 (0.27)</td>
<td>105 (2.08)</td>
</tr>
<tr>
<td></td>
<td>[o]</td>
<td>270 (11.41)</td>
<td>70 (0.32)</td>
<td>100 (1.90)</td>
</tr>
<tr>
<td>SAE</td>
<td>[h]</td>
<td>100 (8.21)</td>
<td>45 (2.64)</td>
<td>98 (2.63)</td>
</tr>
<tr>
<td></td>
<td>[ɛ]</td>
<td>66 (2.42)</td>
<td>68 (0.92)</td>
<td>110 (3.18)</td>
</tr>
<tr>
<td></td>
<td>[l]</td>
<td>60 (2.99)</td>
<td>74 (0.28)</td>
<td>104 (1.49)</td>
</tr>
<tr>
<td></td>
<td>[o]</td>
<td>257 (7.25)</td>
<td>70 (0.19)</td>
<td>110 (1.12)</td>
</tr>
</tbody>
</table>
and the tone pre-test, the entire experiment lasted approximately 90 min.

Data analysis

Environmental and scanner noise was removed using a multi-shift PCA noise reduction algorithm (de Cheveigné and 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.

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

Fig. 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.

C. Average differences between deviants and standards

2332

M. Scharinger et al. / NeuroImage 56 (2011) 2329–2338

Fig. 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).
subsequently analyzed separately for each time-window of interest in Mixed Effect Models with subject as random effect (Baayen, 2008; Pinheiro and 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.

**M100 source analysis**

Although not particularly prominent in the grand-average displayed in Fig. 4, M100 responses occurred before the MMN and had the expected topographies (Poeppel and Marantz, 2000, Fig. 4 left panel). M100 sources were determined on the basis of equivalent current dipoles (ECDs) as described in Obleser et al. (2004a, 2004b). First, an orthogonal left-handed head frame was defined with lateral–medial, anterior–posterior, and superior–inferior coordinates (cf. Figs. 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 and 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 [c] (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 et al., 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).

**Results**

**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 Fs < 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 × Position × 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 [ɛ]). In contrast, MMNs to AAE stimuli were larger to SAE stimuli in the late time window (i.e. as a response to [o], Fig. 4).

The pattern reflected by the three-way interaction of Time × Position × Dialect was strongly left-lateralized, as suggested by a significant four-way interaction of Time × Position × Dialect × 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 × Position × Dialect × Deviant type (F(1,196) = 6.30, p < 0.05), showing that the observed left-lateralized pattern reflected by the Time × Position × Dialect interaction indeed depended on the deviant type, i.e. was due to the real deviant, not the sham deviant.

**Dipole sources**

For the M100 corresponding to the first vowel [ɛ] 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 [ɛ] and [o] (Fig. 5). Although non-averaged raw dipole locations showed rather wide-spread variation in the sagittal (posterior–anterior) and axial (inferior–superior) dimension (Fig. 6), averaged dipole locations for [ɛ] 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 × Dialect: F(1,75) = 8.77, p < 0.01) and

---

**Fig. 5.** Dipole source location in the sagittal plane. The left graph shows the source differences between SAE and AAE [ɛ]. 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.

**Fig. 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 “ɛ” [ɛ] in standard (blue, triangles) and deviant (red, filled circles) position. Correspondingly, the right panel illustrates dipoles in the right hemisphere.
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 ($\text{Dialect} \times \text{Hemisphere}: F(1,75) = 5.72, p < 0.05$).

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 $\text{o}$, the spatial locations of the ECDs differed in the superior–inferior dimension: deviants had more inferior locations than standards ($F(1,70) = 4.90, p < 0.05$; Fig. 3). The significant interaction $\text{Position} \times \text{Dialect} (F(1,70) = 4.79, p < 0.05)$ 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 $\text{Position} \times \text{Dialect} \times \text{Hemisphere} (F(1,70) = 6.52, p < 0.05)$. 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 $\text{e}$, but was not significant ($F(1,75) = 1.12, n.s.$). A pooled analysis across $\text{e}$ and $\text{o}$ showed that deviants were always located inferior to standards ($F(1,167) = 6.74, p < 0.05$).

As for $\text{e}$, 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 $\text{Position} \times \text{Hemisphere} (F(1,70) = 5.07, p < 0.05)$.

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 ($\text{[e]}$) and second ($\text{[o]}$) 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 these difference 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 (Nätänen 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 and Nätänen, 2001), a finding that is consistent with recent neurophysiological data on social perception (Cunningham et al., 2004; Hart et al., 2000; Ito and 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 Fig. 2C), these differences seem to be well aligned with the corresponding MMN responses (Fig. 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 Phillips 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 et al., 2003; Obleser et al., 2004a, 2004b).

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 et al. (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
voice- (or, by extension, dialect) selective areas. Therefore, these vowels effects were restricted to the right hemisphere, suggesting that the by \[\text{than were SAE deviant dipoles. In contrast, AAE deviant dipoles elicited more vertically (i.e. were more aligned with the inferior}

pattern of source dipole orientations for \[\text{fraction relative to AAE}\ [\text{e}, i.e. produced further front in the mouth than AAE}\ [\text{e}\] (Figs. 2 and 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 and 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.

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 and Grosbras, 2010; Belin and Zatorre, 2003; Belin et al., 2000; von Kriegstein et al., 2003; von Kriegstein and 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 and 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 and 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 7RO1DC005660-07 to W.J.L. and David Poeppel and carried out while M.S. was affiliated with the University of Maryland, College Park.

References


