

Published in final edited form as:

Brain Lang. 2014 March ; 130: 42–49. doi:10.1016/j.bandl.2014.01.007.

Neurophysiology of spectrotemporal cue organization of spoken language in auditory memory

Aaron C. Moberly, Jyoti Bhat, D. Bradley Welling, and Antoine J. Shahin

The Ohio State University Wexner Medical Center Department of Otolaryngology – Head & Neck Surgery

Abstract

Listeners assign different weights to spectral dynamics, such as formant rise time (FRT), and temporal dynamics, such as amplitude rise time (ART), during phonetic judgments. We examined the neurophysiological basis of FRT and ART weighting in the /ba/-/wa/ contrast.

Electroencephalography was recorded for thirteen adult English speakers during a mismatch negativity (MMN) design using synthetic stimuli: a /ba/ with /ba/-like FRT and ART; a /wa/ with /w a/-like FRT and ART; and a /ba/^{wa} with /ba/-like FRT and /wa/-like ART. We hypothesized that because of stronger reliance on FRT, subjects would encode a stronger memory trace and exhibit larger MMN during the FRT than the ART contrast. Results supported this hypothesis. The effect was most robust in the later portion of MMN. Findings suggest that MMN is generated by multiple sources, differentially reflecting acoustic change detection (earlier MMN, bottom-up process) and perceptual weighting of ART and FRT (later MMN, top-down process).

Keywords

Amplitude rise time; Auditory evoked potentials; Formant rise time; Mismatch negativity; Speech perception

1. Introduction

To perceive spoken language, the brain perceptually organizes representations of spectrotemporal bits of the speech signal, termed acoustic “cues,” into a coherent phonetic code and ultimately a speech percept. To do so, normal-hearing (NH) listeners assign different weights to correctly integrate multiple properties of the speech signal and then assign phonetic labels (Bailey and Summerfield, 1980; Best et al., 1981). Adult listeners of a

© 2014 Elsevier Inc. All rights reserved

Correspondence concerning this article should be addressed to Aaron C. Moberly, Department of Otolaryngology, The Ohio State University, 915 Olentangy River Rd., Suite 4000, Columbus, OH 43212, Aaron.Moberly@osumc.edu, Phone: 614-293-6926, Fax: 614-293-9698.

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.

Conflicts of Interest and Source of Funding: The authors declare no conflict of interest. This work was supported by an NIH/NIDCD grant (AJS, R03-DC011168).

native language generally use the same cue-weighting strategies in which they assign similar perceptual weights to a given acoustic cue (Nittrouer and Miller, 1997; Ohde and Haley, 1997). This is true even when cues are equally discriminable and informative (Holt and Lotto, 2006), most likely because those strategies allow for the most accurate and efficient speech perception in their native language (Best, 1994; Jusczyk et al., 1995; Nittrouer, 2005).

Acoustic cue-weighting studies, such as those addressing the /ba/-/wa/ amplitude and formant transition distinction (Nittrouer et al., 2013; Nittrouer and Studdert-Kennedy, 1986; Walsh and Diehl, 1991), have mainly used psychoacoustic methods to understand weighting strategies in adult native English speakers. These studies showed that regardless of the rate of amplitude rise time (ART) or formant rise time (FRT), individuals overwhelmingly use FRT to categorize the /ba/-/wa/ consonant-vowels (CVs). However, psychoacoustics alone cannot inform us of the neural time-course of ascension of activity along the auditory pathway. For example, it is not clear from the above behavioral studies whether acoustic cue weighting commences at the level of obligatory (P1-N1-P2) auditory evoked potentials (AEPs) or whether this process is subsequent and is exhibited at higher levels of processing. The motivation for this study was to further understand the neural underpinning of acoustic cue weighting for ART and FRT. Because the ART has been shown to be important in facilitating speech perception and phonetic identification and discrimination in children (Carpenter and Shahin, 2013; Goswami et al., 2011) and in populations where speech is delivered through artificial prostheses (e.g., cochlear implant listeners), it is important to understand the brain-behavior relationship of ART and FRT cues in auditory memory. We start with native adult English speakers so we can better understand normal, typically-developed neural mechanisms. Ultimately, this approach will generate a framework from which to approach perceptual organization strategies of these cues in children and in CI users.

This problem was recently addressed by Carpenter and Shahin (2013). Using a passive listening protocol they examined the ART and FRT auditory cortex (AC) representations recorded in electroencephalography (EEG). Electrophysiological testing provides a means to assess the timing and strength of neural activity and thus can be informative in revealing weighting dynamics of acoustic cues. The authors used modified natural stimuli consisting of /ba/ and /wa/ stimuli as well as a /ba^{wa} stimulus defined as a /ba/ with a superimposed /wa/ envelope. They found that the amplitude of the N1-P2 AEP observed at the vertex (channel Cz) reflected the fidelity of formant (FRT) representations (/ba/ and /ba^{wa} N1-P2 amplitudes were similar, but /ba^{wa} and /wa/ N1-P2 amplitudes were different) while the N1-P2 amplitude at channel C4, to the right of the vertex, reflected the fidelity of ART representations (opposite the effect observed at Cz). Furthermore, they showed that ART representations (N1-P2 at channel C4) exist before age 4 years; however, N1-P2 starts to emerge at the vertex after age 4-5 years. This morphological shift coincides with a shift in children's weighting strategy whereby they learn to predominately use FRT in labeling the /ba/-/wa/ CVs. In short, the Carpenter and Shahin (2013) study revealed that both FRT and ART are encoded at the AC (N1-P2 obligatory AEPs), but are differentially exhibited at Cz and C4 (likely involving different neural generators). However, their findings could not explain why the N1-P2 at the vertex indexing FRT is favored over the

lateral N1-P2 indexing ART during phonetic judgments. Thus, in view of Carpenter and Shahin's (2013) finding, we reasoned that the perceptual organization of acoustic cues must take place subsequent to the obligatory N1-P2 AEP, likely involving higher level processes reflecting top-down maintenance of auditory memory.

To this end, the current undertaking builds on the study by Carpenter and Shahin (2013) by expanding the approach to the use of the mismatch negativity (MMN) to assess the neural time-course of acoustic cue weighting. The MMN is an AEP characterized by a fronto-central negativity elicited by any acoustically discriminable change (deviant or oddball stimulus) within a regular (standard stimulus) stream of stimuli (Näätänen, 2001; Picton et al., 2000). It usually peaks about 150-350 ms following deviant stimulus onset. The MMN is believed to represent the brain response to a violation in the matching of a stimulus representation to a sensory-memory trace formed in short- or long-term auditory memory (Näätänen et al., 2011; Pulvermüller and Shtyrov, 2006). The resulting MMN magnitude evoked by the different cue manipulations can inform us of how well one cue, compared with the other cue, is represented neurophysiologically in auditory memory.

Individuals listened to standard and oddball speech sounds consisting of synthetic /ba/ and /wa/ CVs and a synthetic /ba/ stimulus superimposed with the slowly rising /wa/ envelope (termed /ba/^{wa}). An MMN evoked during a /ba/ and /ba/^{wa} contrast should isolate an ART effect, as both stimuli possess the same FRT. On the other hand, an MMN evoked during a /wa/ and /ba/^{wa} contrast should isolate an FRT effect, as both possess the same ART. We hypothesized that because adults have reliably been shown to weight the spectral cue more heavily than amplitude cue in the /ba/-/wa/ contrast (Carpenter and Shahin, 2013; Nittrouer et al., 2013; Nittrouer and Studdert-Kennedy, 1986; Walsh and Diehl, 1991), subjects should exhibit greater MMN during the FRT than the ART contrast.

2. Materials and Methods

2.1. Participants

Subjects were recruited by posted flyers and from a subject pool of the Department of Otolaryngology, and they were paid for their participation. Thirteen adult, native English speakers (age range 19-42 years, mean 25.2 years; 8 female) with no known hearing problems or history of language impairment were recruited for the study. Prior to testing, participants responded to a questionnaire inquiring about handedness (Edinburgh Handedness Inventory), language experience (other than English), years of education, and history of hearing or speech problems. All subjects were right-handed. Most participants (8 of 13) had no previous experience with a second language, three claimed a moderate proficiency in a foreign language, and two claimed a poor proficiency in a foreign language. Average number of years of school was 15.2 (range 12 to 18). Pure-tone thresholds for the frequencies of .25 to 4 kHz were less than 25 dB hearing level for all participants based on a hearing test within the previous two years. Seven participants had a history of music training (average duration 5.6 years). No subject was on medication for psychological or neurological conditions at the time of testing. Subjects were tested at The Auditory Neuroscience Lab, Eye and Ear Institute, The Ohio State University. Informed consent was

obtained from all subjects in accordance with the ethical guidelines of the Institutional Review Board of The Ohio State University.

2.2. Stimuli

For evaluating acoustic cue weighting in this study, the /ba/-/wa/ contrast was chosen. Both spectral change (in the form of formant rise time, FRT) and amplitude change (in the form of amplitude rise time, ART) could serve as robust cues to the /ba/ versus /wa/ contrast and should both be acoustically salient to normal-hearing individuals. For /ba/, both formant and amplitude rise to steady state quickly (short FRT and ART), whereas for /wa/, both FRT and ART are long. Three synthetic versions of /ba/ and /wa/ from Nittrouer et al. (2013) were used in this experiment and are shown in Figure 1. The first was a synthetic /ba/ stimulus with a /ba/-like FRT (30 ms) and ART (10 ms) and will be termed the /ba/. The second stimulus had a /wa/-like FRT (110 ms) and ART (70 ms) and will be termed the /wa/. The third stimulus had a /ba/-like FRT (30 ms) and a /w α/-like ART (70 ms), termed /ba/^{wa}. In natural speech, fundamental frequency may differ between the two consonantal contexts and could serve as a confounding variable. Using synthetic stimuli eliminates that possibility because fundamental frequency can be held constant across stimuli. Thus, this particular contrast and these synthetic stimuli provide an ideal opportunity for examining the MMN responses to changes in spectral structure and temporal structure. Stimuli were created using a Klatt synthesizer (Sensyn) with a sampling rate of 10 kHz. All three tokens were 370 ms in duration, with a constant fundamental frequency of 100 Hz. Starting and steady-state frequencies of the first two formants were the same for all stimuli, but the time to reach steady-state frequencies varied. F1 started at 450 Hz and rose to 760 Hz at steady state. F2 started at 800 Hz and rose to 1150 Hz at steady state. F3 was kept constant at 2400 Hz. Figure 1 shows waveforms and spectrograms for the /ba/, /ba/^{wa}, and /wa/ stimuli. More details about stimulus generation can be found in Nittrouer et al. (2013).

2.3. Procedure

Participants underwent EEG testing during a passive oddball auditory task involving presentation of a series of “standard” stimuli interrupted by occasional “deviant” stimuli. The standard and the deviant stimuli were pseudo-randomly presented using Presentation software (Neurobehavioral systems, Albany, CA). The stimuli were delivered using free-field stimulation with two Tannoy Precision 8D (TANNOY, Scotland, UK) speakers 1.5 meters from the participant at 45 degrees off center. Loudness was originally calibrated at 70 dB at subject distance, but was adjusted ($< \pm 5$ dB) to the participants’ comfort level and kept constant across the entire experiment. Continuous EEG data was recorded using a 64-channel cap (10-20 system, Ag-AgCl electrodes, 512 A/D conversion rate, BioSemi ActiveTwo system, Amsterdam, Netherlands) in a sound-attenuated room, with Common Mode Sense (CMS) and Driven Right Leg (DRL) passive electrodes serving as grounds.

The task involved eight oddball blocks, consisting of two identical blocks for each of four conditions, and one control block of /ba/ only stimuli. In each of the oddball blocks, the /ba/^{wa} served as either a standard or deviant stimulus. Thus, there were four conditions: (1) standard /ba/, deviant /ba/^{wa}, an ART contrast; (2) standard /ba/^{wa}, deviant /ba/, an ART

contrast; (3) standard /wa/, deviant /ba/^{wa}, an FRT contrast; and (4) standard /ba/^{wa}, deviant /wa/, an FRT contrast. Note that the blocks were set up to incorporate a “flip-flop” design in which a stimulus acted as the standard in one block and as the deviant in another block. This “flip-flop” design (also known as a “counterbalanced oddball paradigm”) for a given contrast eliminated any responses due only to differences in the obligatory activity (i.e., N1-P2, sustained field) responses (Hall, 2007; Wunderlich and Cone-Wesson, 2001).

In each of the 8 oddball blocks, participants were presented with 300 stimulus trials containing 15% deviants (45 trials) and 85% standards (255 trials). For the control block, participants were presented with 200 stimulus trials of the /ba/ only stimulus. The InterStimulus Interval for all blocks was set to 1000 ms. The nine blocks were randomized across subjects. During testing, a pseudo-random sequence of stimuli was presented, with at least 3 standard stimuli presented before each deviant stimulus. Throughout testing, participants watched a silent movie of their choice on a 24-inch LCD computer monitor placed 1 meter in front of them and were instructed to ignore the auditory stimuli. Time of testing was 10 minutes per block with one-minute breaks between blocks. The entire testing session lasted under 2 hours.

2.4. Data Analysis

Using EEGLAB (Delorme and Makeig, 2004) and in-house MATLAB (The MathWorks, Inc., Natick, MA) code, continuous EEG files for each of the nine blocks were combined to generate one grand continuous file and then epoched from -100ms to +500ms around the stimulus marker. The prestimulus interval (-100ms to 0ms) was used to baseline the epoched data. This baseline-corrected file was subjected to independent component analysis (ICA), in which 64 ICA components were generated. Topographies and waveforms of the ICA components were visually inspected and components representing ocular artifacts were rejected (2/64 components per subject, except for one subject in which 3 additional components were rejected due to frontal muscle activity). Subsequently, trials containing amplitudes of $\pm 150 \mu\text{V}$ or greater in any channel were rejected. The data were average-referenced and bandpass filtered between 0.1 and 30 Hz using a zero-phase butterworth filter. Trials were separated to generate a set of standard trials and deviant trials for each of the ART and FRT conditions. Auditory evoked potentials (AEPs) for each standard and deviant condition were computed by averaging all trials separately for each condition, producing one standard and deviant pair for each participant, channel, and condition. The mean group number of trials included for each condition was as follows: for the ART contrasts, 515 standard trials and 94 deviant trials; for the FRT contrasts, 497 standard trials and 90 deviant trials.

Analysis was limited to the mean AEP waveforms of the fronto-central electrodes Fz, F3, F4, FC1, FC2, Cz, C3, and C4, which were computed for each of the standard and deviant pair, condition, and subject. For each individual, the standard and deviant waveforms (collapsed across the “flip-flop” conditions) were submitted to two-tailed sliding window *t*-tests to assess whether a statistically significant ($p < 0.05$) MMN response had been elicited. This method was similar to techniques used by Kraus et al. (1995) to identify MMN in individual subjects and also similar to that of Bishop and Hardiman (2010) who performed a

t-test on single-trial analysis of difference waveforms. The test was confined to the period between 150 ms and 350 ms, and was done by sliding a 15 ms segment every 1 sample point (~ 2 ms) and performing the *t*-test between the deviant and standard waveforms. In other words the *t*-tests for each window contrasted 8 time (sample) points between the standard and deviant waveforms. The MMN onset was taken as the latency at which the *t*-test (Bonferroni corrected for the number of executions) became significant, and the MMN offset was taken as the latency at which the *t*-test was no longer significant. Thus, the MMN duration was determined to be the duration between the onset and offset time points, as long as a region of negativity was visibly confirmed on the MMN (deviant minus standard) waveform. The MMN peak values were determined to be the time-points at which the largest negative deflection in the difference waveform coincided with a significant *p*-value. Because this method could conceivably lead to multiple noncontiguous regions of significant negativities in the difference waveforms, we supplemented this analysis with a topographic analysis. That is, topographies were evaluated for the MMN amplitude peaks noted on the MMN waveforms within the statistically significant MMN time periods. Individual topographic plots were examined visually to verify or rule out the presence of MMN, defined as a significant fronto-central to mastoid negativity. Therefore, the combination of a significant difference between the standard and deviant waveform amplitude on the sliding *t*-test, a visible negativity on the difference (deviant minus standard) waveform, and a confirmatory topography was used to verify or exclude the presence of a MMN response for each individual subject. If a subject's responses did not meet the *t*-test, visible waveform negativity, and topographic criteria, the MMN response was noted to be absent. If the *t*-test, visible difference waveform negativity, and topographic criteria revealed more than one peak as consistent with an MMN response, the peak with the most MMN-like topographic response was taken as the true MMN response. This process ensured a conservative verification and measurement of true MMN responses.

When analyzing MMN responses, no consensus exists as to the best approach for quantifying MMN responses (Hall, 2007). Therefore, multiple approaches were chosen to evaluate MMN responses. First, individual MMN response peak amplitude and peak latency were identified and used in statistical analysis. Second, we examined the area under the curve of the MMN. MMN area has been used as a measure of MMN response magnitude in previous studies (Kraus et al., 1995; Tremblay et al., 1997, 1998; Ylinen et al., 2009). Because the MMN response may have a variable latency and duration with a shallow peak, the MMN area is a useful way to assess MMN dynamics. The MMN area was calculated as the area under the curve for the difference waveform (deviant minus standard) from the MMN onset to the MMN offset, based on the results of the sliding *t*-test.

The ninth control condition of /ba/ only stimuli was used to evaluate if spuriously significant negativity would be found by performing the same analysis on the average of 100 permutations, each of which had 85% randomly labeled "standard" and 15% randomly labeled "deviant" stimuli in this series of identical stimuli. This would help validate the analyses used for identifying true MMN responses. This analysis did not show an MMN-like response, validating that the negativity observed for the deviant stimuli in the main analysis represented a true effect.

2.5. Statistical Analysis

First, we implemented paired *t*-tests to examine whether there was a significant difference in the average MMN response peak amplitude or latency, as well as the MMN area, distinguishing the FRT and ART contrast. For any individual and condition in which a MMN response was noted to be absent, the MMN amplitude was determined to be the potential value for that individual/condition waveform corresponding to the group average MMN latency. Similarly, when MMN was noted to be absent, the MMN area was determined to be the area under the curve for that individual/condition waveform corresponding to the beginning and end latency points of the group average MMN latencies.

Second, we evaluated the influence of FRT and ART manipulations on the successive MMN areas under the curve over the time period spanning 150 to 350 ms. We applied a two-way analysis of variance (ANOVA, general Linear Model of Statistic v. 9.1, StatSoft, OK) to the data with MMN area as the dependent variable, here defined across the entire period from 100 to 350 ms as the area under the curve for deviant minus standard waveforms. The first factor evaluated was cue, either the FRT or the ART. The second factor evaluated was time window, breaking the 150 to 350 ms time period into 50 ms windows. This was performed to determine at which time period the MMN response was greatest overall, as well as at which time period the largest difference in magnitude of MMN was seen between the FRT and ART conditions. The MMN spans a long period and is believed to be a superposition of several sources in auditory and non-auditory cortices. Thus, we justify this analysis as a way to determine whether an MMN difference in weighting strategy is due to engagement of more neural sources that unfold with time. That is, if the MMN difference was confined to a particular segment of the MMN, we could conclude that additional sources are associated with the ART-FRT perceptual distinction. All tests were two-tailed and corrected for sphericity violations (Green-house Geisser) where appropriate (following a significant value for Mauchly sphericity test). Post-hoc testing used Scheffe's test. Precise values from statistical tests are reported for $p < 0.05$. Outcomes are reported as not significant when $p > 0.05$.

3. Results

We would like to point out from the outset that a negative potential representing the MMN response was found for every individual (100%) for the FRT “flip-flop” averaged waveforms, and an MMN wave was found for 11 out of 13 (85%) individuals for the ART “flip-flop” averaged waveforms. The total MMN response area for the individual FRT waveforms was greater than that for the ART waveforms in 9 out of 13 individuals (69%). Notice that the area difference between percept types is not robust, despite the fact that it reached significance. However, this contrast becomes much more robust when the area is confined to a specific time window (see below) providing support to our interpretation for how the cues are perceptually organized in auditory memory (see Discussion).

Figure 2 depicts group AEP waveforms at the fronto-central sites (averaged across electrodes). The standard, deviant, and difference waveforms (revealing the MMN) are shown for the FRT contrast (Figure 2A) and ART contrast (Figure 2B). Topographic maps at MMN peak latencies (noted by black arrow on the waveform plot) are shown to the right.

Based on prior evidence demonstrating that normal-hearing adults assign greater weight to FRT versus ART (Carpenter and Shahin, 2013; Nittrouer et al., 2013; Nittrouer and Studdert-Kennedy, 1986; Walsh and Diehl, 1991) in phonetic judgments, we hypothesized that the MMN response for the FRT contrast would be greater than the MMN response for the ART contrast. Indeed, this was found to be the case as evidenced by a larger MMN peak amplitude occurring for the FRT over the ART contrast ($t_{(12)} = 3.33, p < 0.006$) (Figure 2C). MMN peak latency was found to be later for the FRT contrast than the ART contrast ($t_{(12)} = -4.15, p < 0.002$) (Figure 2D). In addition to the larger MMN peak amplitude occurring for FRT over ART, the MMN area was larger for the FRT contrast than the ART contrast ($t_{(12)} = 2.80, p < 0.02$; 9 out of 13 subjects exhibited this effect) (Figure 2E).

Subsequently, we evaluated the relationship between magnitude of MMN responses (areas under the curve) and time-window. An ANOVA on the MMN area (factors: cue and time window) revealed a main effect of cue ($F_{(1,12)} = 5.97, p < 0.04$), which was due, as hypothesized and consistent with the MMN peak results above, to a larger MMN area occurring for the FRT cue over the ART cue. The ANOVA also revealed a main effect approaching significance for window ($F_{(3,36)} = 3.08, p < 0.08$), which was due to the maximum MMN response occurring in window 2 (201-250 ms), with a trend towards a significant difference compared with window 4 (301-350 ms, Scheffe's test, $p = 0.05$). An interaction was seen between the variables cue and window ($F_{(3,36)} = 5.32, p < 0.004$), which was attributed to a significant difference between the MMN for FRT and ART (Scheffe's test, $p < 0.001$) occurring during the 251-300 ms time window, but not for the other windows (Figure 3).

In summary, greater MMN responses, both in terms of MMN peak amplitude and MMN area, were seen for the FRT cue contrast, compared to the ART cue contrast. However, additional analysis revealed that this difference was most robust (reached significance) for the later portion of the MMN (251-300 ms), despite the fact that the MMN on average reached its maximum earlier (201-250 ms).

Supplemental Signal-to-Noise Ratio Analysis

One concern when comparing standards and deviants waveforms is the difference in number of trials (signal-to-noise ratios). We addressed this point as follows: 1) We randomly selected standard trials that were equal in number to all deviant trials for that particular condition (ART, FRT) and subject. 2) This process was repeated 500 times. 3) We averaged the trials of these 500 permutations into one AEP. 4) We contrasted the original standard trial AEPs with the permuted ones using cross correlation and t-tests. These tests revealed that the two signals were identical as clearly shown in the supplementary Figure. The cross correlation between the two signals at zero lag was 100% ($r = 1$).

4. Discussion

We aimed in the present study to elucidate the neural time-course underlying amplitude (ART) and spectral (FRT) cue weighting strategies during the /ba/-/wa/ contrast. To do so, we compared the MMN responses evoked during the spectral and temporal contrasts of manipulated /ba/ and /wa/ CVs during a passive oddball listening task in adult native

English listeners. In agreement with our hypothesis, this study revealed a larger MMN response to the heavily-weighted FRT contrast compared to the ART contrast. However, in a subsequent analysis we revealed that the effect was most prominent in the later portion of the MMN. We should note that it is unlikely that this effect was due to differences in the obligatory N1-P2 responses, because their influence was excluded by the use of a “flip-flop” protocol. Our results motivate several interpretations about the neurophysiological underpinnings of the perceptual organization of spectrotemporal cues in spoken language processing.

In light of the N1-P2 AEP results of Carpenter and Shahin (2013), the current MMN results suggest that perceptual organization of speech cues, the next step of auditory processing, occurs subsequent to initial encoding (N1-P2) at the AC. Because individuals exhibited the MMN response for both the ART and FRT contrasts, clearly both cues are encoded in auditory memory as well. In fact, MMN magnitudes of both distinction tasks not only were maximally exhibited during the 201-250 ms MMN peak window but also did not differ significantly from one another during that time period. However, the finding that the ART-FRT MMN difference reached significance following its peak value, in the 251-300 ms period, suggests that a subsequent process is triggered that guided the perceptual outcome. That is, the earlier process is associated with acoustic change detection, while the latter process determined which cue to use during phonetic classification. Specifically, this late MMN difference may represent a process in which representations of ART and FRT are perceptually organized so that the ART representation becomes irrelevant (i.e., pushed to the background in auditory memory), while that of the FRT is made relevant (i.e., pushed to the foreground in auditory memory). This explains why the later MMN peak was inhibited during the ART contrast and augmented during the FRT contrast.

The finding that multiple generators reflect different neural processes is not surprising given that different acoustic and/or phonetic features are processed in different neural networks in the primary and non-primary AC (i.e., planum temporale or PT). Thus, the MMN morphology will behave differently depending on the cue being manipulated (Ahlo 1995; Deouell and Bentin, 1998; Jancke et al., 2002; Sophie et al., 2005; Zaehle et al., 2008). Jancke et al. (2002), using fMRI, found bilateral activations of medial PT for CVs vs. vowels, implicating the PT in the analysis of voice-onset-time (VOT) and formant transitions. However, when narrowing the scope of the contrast by comparing voiceless vs. voiced CVs (differing in VOT length), they revealed a stronger activation in the left, than right, medial PT, emphasizing the significance of the left PT in coding VOT and highlighting the acoustic-feature specialization of auditory networks. Along the same lines, in an EEG study, Deouell and Bentin (1998) examined the MMN of non-speech stimuli evoked by deviancies in intensity, frequency, stimulus-onset asynchrony, and location. In their thoughtful design, they adjusted the individual magnitude of deviance such that the detection rates were similar within dimensions and within subjects to allow for across dimension and across subject MMN comparisons. They found that MMN to frequency deviancy, especially at frontal sites, was larger than MMNs for the other contrasts. They attributed this effect to 1) a possible commencement of additional MMN frontal-generators associated with involuntary attention switching process or 2) to a difference in underlying source orientation, which also may imply additional generators. Similarly, results reported in

Näätänen et al. (2011) suggested multiple generators of the MMN response, attributing a portion of the MMN process to an attention-switching response mediating selectivity. Contrary to this account, Garagnani et al. (2009) found that magnetic MMN responses to words and nonwords reflected involved distinct memory circuits but were relatively immune to attention variations. Others suggested that multiple traces for MMN generation may be maintained simultaneously, even in the absence of a task requiring attention (Praamstra and Stegeman, 1992). In this study we did not manipulate attention and thus we cannot rule out an attention switching effect despite the passive design. It is possible that the deviant cues commanded (automatically) the attention of subjects to a greater extent during the FRT vs. ART contrast.

The current study also raises the question of whether the spectrotemporal weighting effect observed here is a strategy that can be generalized across other phonetic categories in adult native English speakers or is specific to certain categorizations (e.g., the /ba-/wa/ contrast). We believe the latter. That is, weighting shifts that favor one cue over the other are dictated by the type of phoneme and context of the speech discourse. This is evident in studies that utilized the MMN to study spectral and durational cue weighting in phonetic categorization. Tuomainen and van der Lely (2007) manipulated the duration and frequency characteristics of a syllable final stop voicing ([bot] vs. [bod]) contrast, for which the duration cue plays a more prominent role in perception of these syllables. They found that the duration manipulation resulted in larger MMN amplitude than the frequency manipulation. In another study, Ylinen et al. (2009) examined the effects of auditory training on cue weighting in Finnish second-language users of English. They used the /i/ and /I/ vowel contrast, in which the duration cue is central to this distinction in the Finnish language, as opposed to frequency in the English language. MMN responses were larger to a non-native spectral cue contrast after training, reflecting a shift in cue weighting. Finally, Lipski et al. (2012) showed a correspondence between the MMN response amplitude and behavior. Specifically, they found a weaker MMN response for Spanish listeners compared with Dutch listeners for spectrally-cued contrasts in the /ɤ/ versus /ɑ/ contrast, in accordance with weaker spectral weighting by the Spanish listeners during a phonetic categorization task of their native language. Taken together, and combined with the current findings, the above suggests that spectrotemporal cue weighting favors the cue that leads to the most efficient perceptual outcome in the language of interest, and this process is automatic, as revealed by the MMN response.

What distinguishes our design from other MMN studies addressing temporal cue weighting (Lipski et al., 2012; Tuomainen and van der Lely, 2007; Ylinen et al., 2009) is that we used an amplitude envelope cue, ART, as opposed to a durational cue. The ART has been shown to be important in facilitating speech perception and phonetic identification and discrimination in children and in cochlear implant listeners (Carpenter and Shahin, 2013; Goswami et al., 2011). Thus, by understanding the perceptual organization (MMN behavior) of ART and FRT cues in auditory memory in native adult English speakers, we should be better situated to address perceptual organization strategies of these cues in children and in CI users. It is conceivable that children and CI users would make greater use of ART cues, and this might be reflected in greater MMN responses to the ART cue over the FRT cue. Additionally, clinical intervention may benefit from auditory training protocols that

encourage cue-weighting strategies that lead to the most efficient and accurate speech perception. Changes in the MMN response could serve as a marker of effective auditory training.

We should emphasize that our results do not suggest that individuals become less sensitive to ART cues. The ART contrast is robustly represented in the MMN response and in the obligatory N1-P2 AEPs (Carpenter and Shahin, 2013). Our results only provide evidence that phonetic classification during the /ba/-/wa/ contrast is dependent on how spectrotemporal cues are organized in auditory memory and not the brain's sensitivity to these cues. Most likely, if individuals are instructed to identify ART differences, i.e., during a discrimination task, they should be capable of such a task. Furthermore, previous studies have clearly shown that temporal cues, such as duration cues, can evoke larger MMN than spectral cues when perception is more contingent on these cues (Lipski et al., 2012; Tuomainen and van der Lely, 2007; Ylinen et al., 2009). It is possible that the brain's reliance on ART may also be augmented if the FRT is degraded. Further studies with stimuli of degraded spectral quality (such as in noise-vocoded speech and speech encoded by cochlear implants) can test whether perceptual strategies of cue weighting, and thus MMN magnitude, shift in favor of ART under these circumstances.

5. Conclusions

In summary, our findings provide evidence that the perceptual weighting of ART and FRT cues distinguishing the /ba/-/wa/ CVs is reflected neurophysiologically, as evidenced by the behavior of the MMN. We are currently investigating the current experimental design in individuals with cochlear implants (CIs) who have previously undergone behavioral testing for cue weighting for the same /ba/-/wa/ contrast and have shown variable weighting of ART and FRT (Moberly et al., in press). The spectral quality of speech encoded by CIs is highly degraded, and it is thought that CI listeners may rely more on envelope cues (ART) than spectral cues (FRT). Hence, we expect to find larger MMNs during the ART contrast than the FRT contrast in CI users.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This study was supported by a NIH/NIDCD award (AJS, R03-DC011168, 2009). The authors would like to thank Dr. Susan Nittrouer, Dr. Joanna Lowenstein, and Dr. Eric Tarr for their assistance in stimulus synthesis. We also would like to thank Dr. Kelly Tremblay for her insightful comments on the manuscript.

References

- Alho K. Cerebral generators of mismatch negativity (MMN) and its magnetic counterpart (MMNm) elicited by sound changes. *Ear. Hear.* 1995; 16:38–51. [PubMed: 7774768]
- Bailey P, Summerfield Q. Information in speech: observations on the perception of [s]-stop clusters. *J. Exp. Psychol. Hum. Percept. Perform.* 1980; 6:356–363.
- Best, CT. The emergence of native-language phonological influences in infants: A perceptual assimilation model. In: Goodman, JC.; Nusbaum, HC., editors. *The Development of Speech*

- Perception: The Transition from Speech Sounds to Spoken Words. MIT Press; Cambridge: 1994. p. 167-224.
- Best CT, Morriongiello BA, Robson RC. Perceptual equivalence of acoustic cues in speech and nonspeech perception. *Percept. Psychophys.* 1981; 29:191–211. [PubMed: 7267271]
- Bishop DVM, Hardiman MJ. Measurement of mismatch negativity in individuals: A study using single-trial analysis. *Psychophys.* 2010; 47:697–705.
- Carpenter AL, Shahin AJ. Development of the N1-P2 auditory evoked response to amplitude rise time and rate of formant transition of speech sounds. *Neurosci. Lett.* 2013; 544:56–61. [PubMed: 23570734]
- Delorme A, Makeig S. EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods.* 2004; 134:9–21. [PubMed: 15102499]
- Deouell LY, Bentin S. Variable cerebral responses to equally distinct deviance in four auditory dimensions: A mismatch negativity study. *Psychophys.* 1998; 35:745–754.
- Garagnani M, Shtyrov Y, Pulvermuller F. Effects of attention on what is known and what is not: MEG evidence for functionally discrete memory circuits. *Front. Hum. Neurosc.* 2009; 3:1–12.
- Goswami U, Fosker T, Huss M, Mead N, Szucs D. Rise time and formant transition duration in the discrimination of speech sounds: the Ba-Wa distinction in developmental dyslexia. *Dev. Sci.* 2011; 14:34–43. [PubMed: 21159086]
- Hall, JW. Mismatch negativity (MMN) response. In: Hall, JW., editor. *New Handbook of Auditory Evoked Responses*. Pearson Education; Boston: 2007. p. 548-580.
- Holt LL, Lotto AJ. Cue weighting in auditory categorization: Implications for first and second language acquisition. *J. Acoust. Soc. Am.* 2006; 119:3059–3071. [PubMed: 16708961]
- Jancke L, Wustenberg T, Scheich H, Heinze HJ. Phonetic perception and the temporal cortex. *Neuroimage.* 2002; 15:733–746. [PubMed: 11906217]
- Jusczyk, PW.; Hohne, EA.; Mandel, DR. Picking up regularities in the sound structure of the native language. In: Strange, W., editor. *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. York Press; Baltimore: 1995. p. 91-119.
- Kraus N, McGee T, Carrell TD, Sharma A. Neurophysiologic bases of speech discrimination. *Ear. Hear.* 1995; 16:19–37. [PubMed: 7774767]
- Lipski SC, Escudero P, Benders T. Language experience modulates weighting of acoustic cues for vowel perception: An event-related potential study. *Psychophys.* 2012; 49:638–650.
- Moberly, AC.; Lowenstein, JH.; Tarr, E.; Caldwell-Tarr, A.; Welling, DB.; Shahin, AJ.; Nittrouer, S. Do adults with cochlear implants rely on different acoustic cues for phoneme perception than adults with normal hearing?. *J. Speech. Lang. Hear. Res.* in press
- Molholm S, Martinez A, Ritter W, Javitt DC, Foxe JJ. The neural circuitry of preattentive auditory change-detection: an fMRI study of pitch and duration mismatch negativity generators. *Cereb. Cortex.* 2005; 15:545–551. [PubMed: 15342438]
- Näätänen R. The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophys.* 2001; 38:1–21.
- Näätänen R, Kujala T, Winkler I. Auditory processing that leads to conscious perception: A unique window to central auditory processing opened by the mismatch negativity and related responses. *Psychophys.* 2011; 48:4–22.
- Nittrouer S. Age-related differences in weighting and masking of two cues to word-final stop voicing in noise. *J. Acoust. Soc. Am.* 2005; 118:1072–1088. [PubMed: 16158662]
- Nittrouer S, Lowenstein JH, Tarr E. Amplitude rise time does not cue the /ba/-/wa/ contrast for adults or children. *J. Speech Lang. Hear. Res.* 2013; 56:427–440. [PubMed: 22992704]
- Nittrouer S, Miller ME. Predicting developmental shifts in perceptual weighting schemes. *J. Acoust. Soc. Am.* 1997; 101:2253–2266. [PubMed: 9104027]
- Nittrouer S, Studdert-Kennedy M. The stop-glide distinction: Acoustic analysis and perceptual effect of variation in syllable amplitude envelope for initial /b/ and /w/. *J. Acoust. Soc. Am.* 1986; 80:1026–1029. [PubMed: 3771922]

- Ohde RN, Haley KL. Stop-consonant and vowel perception in 3- and 4-year-old children. *J. Acoust. Soc. Am.* 1997; 102:3711–3722. [PubMed: 9407663]
- Picton TW, Alain C, Otten L, Ritter W, Achim A. Mismatch negativity: different water in the same river. *Audiol. Neurootol.* 2000; 5:111–139. [PubMed: 10859408]
- Praamstra P, Stegeman DF. On the possibility of independent activation of bilateral mismatch negativity (MMN) generators. *Electroencephalogr. Clin. Neurophysiol.* 1992; 82:67–80. [PubMed: 1370146]
- Pulvermüller F, Shtyrov Y. Language outside the focus of attention: The mismatch negativity as a tool for studying higher cognitive processes. *Prog. Neurobiol.* 2006; 79:49–71. [PubMed: 16814448]
- Tremblay K, Kraus N, Carrell TD, McGee T. Central auditory system plasticity: Generalization to novel stimuli following listening training. *J. Acoust. Soc. Am.* 1997; 102:3762–3773. [PubMed: 9407668]
- Tremblay K, Kraus N, McGee T. The time course of auditory perceptual learning: Neurophysiological changes during speech-sound training. *Neuroreport.* 1998; 9:3557–3560. [PubMed: 9858359]
- Tuomainen, O.; van der Lely, H. Processing of acoustic cues for voicing in English: A MMN study; *Proc. 16th Int. Cong. Phon. Sci.*; 2007. p. 813-816.
- Walsh MA, Diehl RL. Formant transition duration and amplitude rise time as cues to the stop/glide distinction. *Q. J. Exp. Psychol.* 1991; 43:603–620.
- Wunderlich JL, Cone-Wesson BK. Effects of stimulus frequency and complexity on the mismatch negativity and other components of the cortical auditory-evoked potential. *J. Acoust. Soc. Am.* 2001; 109:1526–1537. [PubMed: 11325124]
- Ylinen S, Uther M, Latvala A, Vepsäläinen S, Iverson P, Akahane-Yamada R, et al. Training the brain to weight speech cues differently: A study of Finnish second-language users of English. *J. Cognitive Neurosci.* 2009; 22:1319–1332.
- Zaehle T, Geiser E, Alter K, Jancke L, Meyer M. Segmental processing in the human auditory dorsal stream. *Brain Res.* 2008; 1220:179–190. [PubMed: 18096139]

Highlights

- English listeners weight formant more than envelope dynamics in phonetic labeling.
- Mismatch negativity (MMN) reflects spectrotemporal cue weighting of phonemes.
- MMN consists of generators reflecting change detection and perceptual organization.

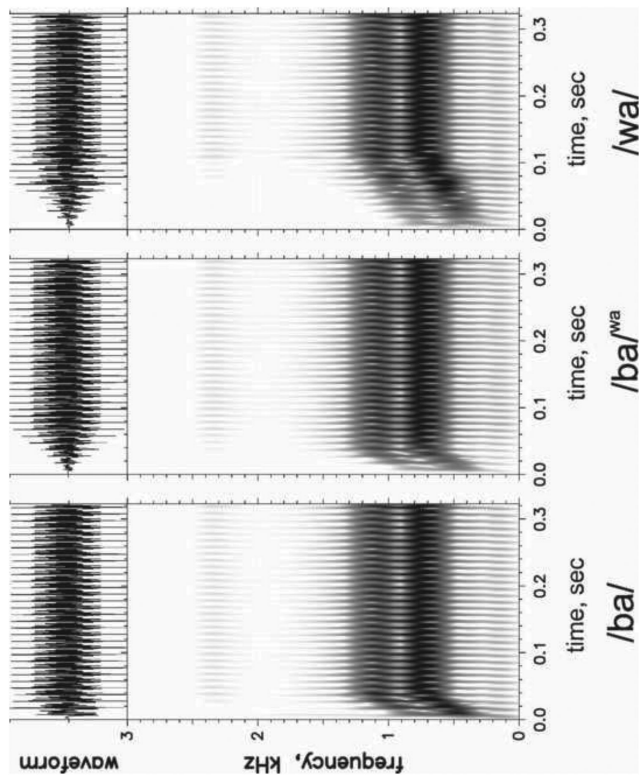


Figure 1. Waveforms (above) and spectrograms (below) for synthetic /ba/, /ba^{wa}, and /wa/ stimuli.

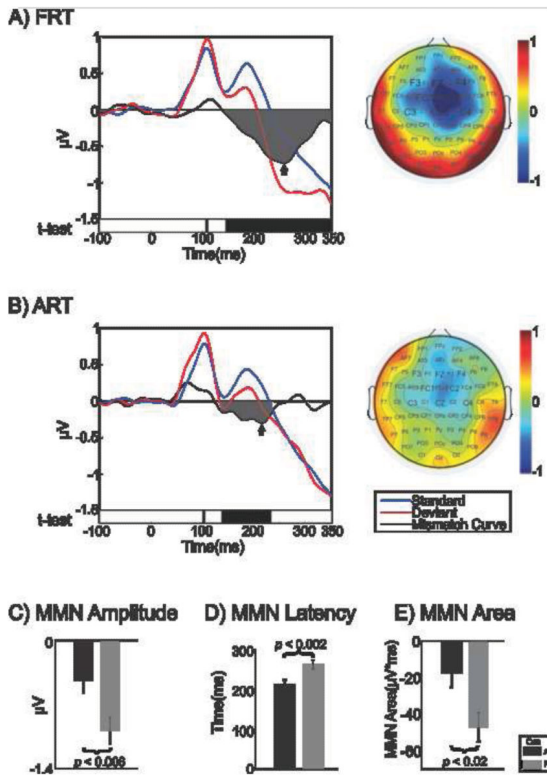


Figure 2. Group average responses to **A)** spectral contrast (FRT), and **B)** amplitude contrast (ART), collapsed across “flip-flop” conditions. Auditory evoked potential waveforms are shown in the left panel; blue waveform represents response to standard stimuli, red waveform represents response to deviant stimuli, and black waveform represents the difference waveform (deviant minus standard) revealing the MMN. The time periods at which the deviant waveform amplitude was significantly different from the standard on *t*-test ($p < .05$) are shown as black bars below the waveforms. Topographic plot of the best MMN peak with corresponding time point, noted by arrow on waveform panel, are shown in the right panel. MMN area is shaded in gray in left panel. Graph **C)** shows the group mean (and standard error) MMN response peak amplitudes to the two cue contrasts. Graph **D)** shows the group mean (and standard error) MMN response peak latencies to the two cue contrasts. Graph **E)** shows the group mean (and standard error) MMN response areas to the two cue contrasts. Negative labels on y-axis for areas signify that MMN is a negative response. MMN: mismatch negativity

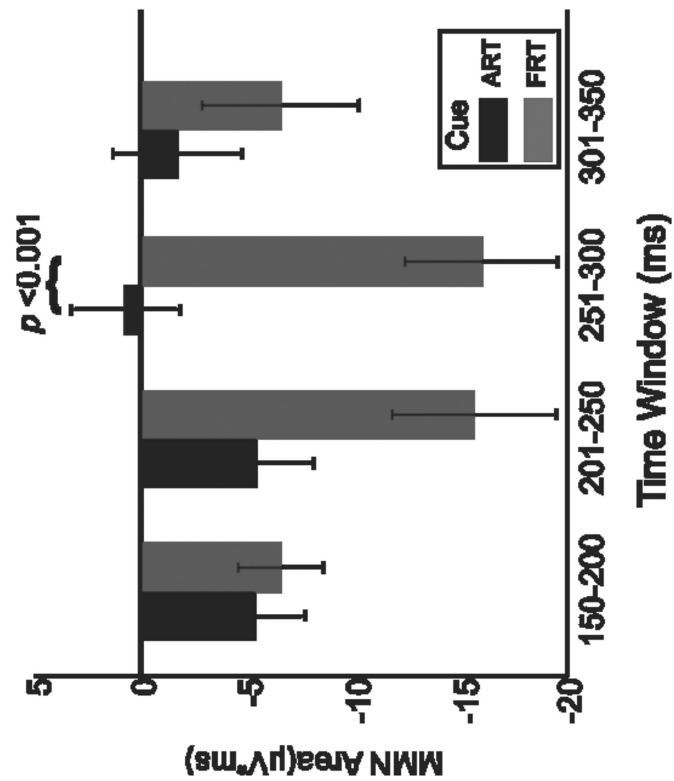


Figure 3. Bar graph showing MMN response areas (calculated as “area under the curve” for the difference waveform, deviant minus standard) for the cue contrasts, ART versus FRT, and the 50-ms time windows from 150 to 350 ms. Black bars represent response to ART contrast; gray bars represent response to FRT contrast (error bars represent standard error). Negative labels on y-axis for areas signify that MMN is a negative response. MMN: mismatch negativity