

Published in final edited form as:

Brain Res. 2010 November 11; 1360: 89–105. doi:10.1016/j.brainres.2010.08.092.

Perception of a Japanese Vowel Length Contrast by Japanese and American English listeners: Behavioral and Electrophysiological Measures

Miwako Hisagi^{1,2}, Valerie L. Shafer¹, Winifred Strange¹, and Elyse S. Sussman³

¹The City University of New York-Graduate School and University Center, PhD. Program in Speech, Language, and Hearing Sciences, 365 Fifth Avenue, New York, New York 10016-4309

² Massachusetts Institute of Technology, Speech Communication Group, Research Laboratory of Electronics, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139

³ Department of Neuroscience, Albert Einstein College of Medicine, 1410 Pelham Parkway S, Bronx, New York 10461

Abstract

This study examined the role of automatic selective perceptual processes in native and non-native listeners' perception of a Japanese vowel length contrast (*tado* vs. *taado*), using multiple, natural-speech tokens of each category as stimuli in a “categorical oddball” design. Mismatch Negativity (MMN) was used to index discrimination of the temporally-cued vowel contrast by naïve adult American listeners and by a native Japanese-speaking control group in two experiments in which attention to the auditory input was manipulated: in Exp 1 (Visual-Attend), listeners silently counted deviants in a simultaneously-presented visual categorical oddball shape discrimination task; in Exp 2 (Auditory-Attend), listeners attended to the auditory input and implicitly counted target deviants. MMN results showed effects of language experience and attentional focus: MMN amplitudes were smaller for American compared to Japanese listeners in the Visual-Attend Condition and for the American listeners in the Visual compared to Auditory-Attend condition. Subtle differences in topography were also seen, specifically in that the Japanese group showed more robust responses than the American listener's at left hemisphere scalp sites that probably index activity from the superior temporal gyrus. Follow-up behavioral discrimination tests showed that Americans discriminated the contrast well above chance, but more poorly than did Japanese listeners. This pattern of electrophysiological and behavioral results supports the conclusion that early experience with phonetic contrasts of a language results in changes in neural representations in auditory cortex that allow for more robust automatic, phonetic processing of native-language speech input.

Keywords

Event-related potentials; Mismatch Negativity (MMN); Japanese; temporal-cues; speech perception; attention

Correspondence: Address, Miwako Hisagi, Massachusetts Institute of Technology, Room 36-581 Speech Communication Group, Research Laboratory of Electronics, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, hisagi@mit.edu; phone/fax number: 617-258-9255.

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.

1. Introduction

Previous cross-language studies have shown that phonetic segments that are contrasted in the non-native language (L2) can cause significant perceptual difficulty if the contrasts are not distinguished in the listener's native language (L1), and the listener has not had considerable experience with that language (see Best, 1995; Strange, 1995). For example, Japanese listeners have difficulty perceiving the English [r/l] contrast (e.g., rock vs. lock) that is not phonologically contrastive in Japanese. One interpretation of this finding is that the highly over-learned, automatic processes by which adult listeners differentiate native phonemic categories interfere with the perception of phonetic categories that are not present in the native language (Strange & Shafer, 2008). Speech perception studies support this claim by showing that L2 learners are often focusing on the incorrect acoustic parameters (derived from their L1) when discriminating or categorizing L2 phonemes (MacKain, Best, & Strange, 1981; Iverson, Hazan, & Bannister, 2005).

Research in speech perception has often made reference to the notion that native-language speech processing is “automatic” (Jusczyk, 1997) and that the development of speech perception results in changes in selective attention (e.g., Werker & Curtain, 2005; Jusczyk, 1997). The Automatic Selective Perception (ASP) model proposed in Strange and Shafer (2008) characterizes first-language speech processing in adults as reflecting automatic Selective Perceptual Routines for the detection of language-specific, phonologically relevant, information in speech signals. In the ASP model, selective perceptual processing of L1 phonetic segments and sequences is automatic due to extensive L1 experience. In contrast, non-native listeners may require attentional resources to discriminate the same phonetic contrasts. One focus of this model is to predict levels of difficulty in cross-language speech perception as a function of stimulus and task factors, as well as L1/L2 phonological and phonetic similarities and differences. They agree with other theorists (e.g. Flege, 1995; Best, 1995) who suggest that perceptual difficulties partially underlie production difficulties of L2 learners and explain why L2 learners can neither produce non-accented speech nor discriminate phonetic differences in the L2 at native levels of accuracy and speed, especially in conditions of high cognitive load.

The ASP model predicts that speech perception of non-native contrasts is dependent on task demands that determine the degree of attentional focus that is placed on the phonetic details of the stimuli. In addition, the number of stimulus properties specifying the contrast will influence success in perceiving the contrast. Behavioral studies have shown that non-native listeners can recover the phonetically-relevant information specified by multiple spectral and temporal parameters when the stimuli are relatively simple (e.g., citation-form (di)syllables) and when listeners' attention is focused (through instructions or training) on the most relevant acoustic-phonetic parameters (Hisagi & Strange, in press; Strange & Dittmann, 1984) However, non-native listeners may require more cognitive resources to access phonetically-relevant information for non-native phonemic contrasts in online perception processing. Thus, when the stimulus materials are more complex (as in sentence-length utterances or when the context varies from trial to trial) and/or when the behavioral task focuses on other levels of processing (as in lexical decision or “semantic goodness” tasks), perceptual performance may suffer. Without focused attention on the relevant cues differentiating the non-native phones, the non-native listeners may not categorize stimuli accurately. Differences between L1 and non-native phonology in how acoustic-phonetic parameters specify contrasts in different contexts predict differences in the amount of focused attention needed in processing.

In current models of perception, highly salient “objects” in relation to their context attract attention and do not need focused attention for perception (Koch, 2004). These salient

objects are hypothesized to be explicitly represented in sensory cortex (Koch, 2004, p. 167). Explicit representation in terms of neuronal connectivity may be the result of determination by genes and developmental processes that have evolved for some types of information. These representations can also be formed by frequently repeated experience that is characterized as over-learning by Crick and Koch (1990). The language-specific Selective Perception Routines (SPRs) described by Strange and Shafer (2008), are likely to be over-learned in this sense. Thus, learning L1 SPRs will result in explicit cortical representation that leads to increased salience of the relevant phonetic cues. Configurations of acoustic cues that are novel, however, will not be explicitly represented at the level of primary auditory cortex. Thus, attention may be needed to facilitate the perception of novel phonemes. Contrasting phonemes that are distinct in a non-native language may be assimilated into a single category in the L1 of a listener. Therefore, selective attention will be required to target the relevant acoustic-phonetic cues that differentiate the contrast.

1.1. Electrophysiological Measures of Auditory Processing

Event-related brain potentials (ERPs) index automatic and attentive processes associated with the perception of speech contrasts and generated in different areas of cortex. ERP measures can provide information regarding processes preceding behavioral output and in the absence of focused attention (e.g., Näätänen, 1988). Specifically, the ERP component mismatch negativity (MMN) serves as an index of passive (automatic) discrimination of an infrequent sound change within a sequence of auditory stimuli (Näätänen, Tervaniemi, Sussman, Paavilainen, & Winkler, 2001). MMN is modality-specific, its maxima for auditory stimuli best seen at fronto-central electrode sites, consistent with neural generators within auditory cortices. MMN reflects an early change detection process. The MMN component peaks approximately 150 ms from detection of the deviation. At this level of processing, the deviance detection process occurs regardless of the direction of attention (Sussman, 2007). The MMN component is modulated by the ease of sound change detection. The easier the change is to detect, the earlier in latency and larger in amplitude is the MMN (Menning, Imaizumi, Zwislerlood, & Pantev, 2002; Nenonen, Shestakova, Huotilainen, & Näätänen, 2003; Nenonen, Shestakova, Huotilainen, & Näätänen, 2005; Dehaene-Lambertz, Dupoux, & Gout, 2000; Ylinen, Shestakova, Huotilainen, Alku, & Näätänen, 2006).

MMN can also be modulated by attention. For example, Sussman and colleagues showed increased MMN to a Finnish vowel duration contrast, “tuli (fire) vs. tuuli (wind)” with attention to the targets (Sussman, Kujala, Halmetoja, Lyytinen, Alku, & Näätänen, 2004). Attention effects, however, may only occur for difficult discriminations (e.g., Gomes, Sussman, Ritter, Kurtzberg, Cowan, & Vaughan, 1999).

Support for the hypothesis that L1 speech perception is highly automatic is found in a study using MMN in a dichotic listening paradigm (Szymanski, Yund, & Woods, 1999). Szymanski et al. suggested that pre-attentive discrimination of speech sounds, as indexed by MMN, is influenced by phonemic status. They found that a greater degree of phonemic (category) difference for a stimulus pair led to a larger amplitude MMN compared to a stimulus pair that was less different in phonemic structure but more different in acoustic structure. Important to the current study, they also found that selective attention to acoustic structure (i.e., by instructing the listener to identify intensity deviants) when the acoustic element was irrelevant to phonemic perception, increased the MMN amplitude to the target intensity cue, but had no effect on the phonemic level of processing. The MMN amplitude evoked by the phonemic difference was not significantly different when attended compared to ignored.

The status of a speech contrast in a listener's L1 has been shown to affect the amplitude and latency of the MMN (e.g., Näätänen, Lehtokoski, Lennes, Cheour, Huotilainen, Iivonen, Vainio, Alku, Ilmoniemi, Luuk, Allik, Sinkkonen, & Alho, 1997; Shafer, Schwartz, & Kurtzberg, 2004; Menning et al., 2002; Nenonen et al., 2003; Peltola, Kujala, Tuomainen, Ek, Aaltonen, & Näätänen, 2003). MMN is generally larger and/or earlier in listeners for whom a contrast is native compared to those for whom the contrast is non-native. These findings support the notion that L1 learning leads to automatic pre-attentive processing of L1 phonological contrasts. In contrast, perception of non-native contrasts is less automatic and, thus, we hypothesize that manipulating attention should influence brain indices of early cortical processing of these contrasts.

Source localization models suggest that major sources of the change-detection system indexed by MMN are found bilaterally in auditory cortex (Alho, Winkler, Escera, Huotilainen, Virtanen, Jaaskelainen, Pekkonen, & Ilmoniemi, 1998). Processing speech often leads to increased contribution of left-hemisphere auditory cortex sources (Näätänen, Paavilainen, Rinne, & Alho, 2007, for review). A recent study using fMRI supports the suggestion that left hemisphere auditory cortex plays a special role in speech processing (Specht, Osnes & Hugdahl, 2009). Specht et al. demonstrated that the middle regions of the left superior temporal sulcus (mid-STS) showed particular sensitivity to perception of an auditory stimulus as speech. The right temporal sites showed sensitivities to perception of stimuli either as speech or as music. These findings suggest that the over-learned representations of native language SPRs are likely to have stronger localization to left auditory cortex. The right auditory cortex representations may reflect representation of the acoustic and prosodic (i.e., melody and timing) features of speech, such as in the case of non-native phones, as first proposed by Näätänen et al. (1997).

ERPs also index specific attention-based processes. N2b and P3b components, index sound change detection, but are elicited only when the sounds are attended (Näätänen, 1990; Sussman et al., 2004). The N2b is an increased negativity at superior sites to the deviant stimulus that serves as an attentional target; it generally peaks around 200 ms and follows MMN in time (Novak, Ritter, & Vaughan, & Wiznitzer, 1990). N2b is largely non-modality specific with widespread generators resulting in a centro-parietal distribution on the scalp (Perrault & Picton, 1984). The P3b is a large posterior-parietal positivity following the N2b, and is also not modality specific. Its peak is generally associated with response times (200 and 600 ms following stimulus onset), reflecting conscious discrimination of an event. P3b is sometimes associated with context updating of working memory representations (Näätänen, 1990; Donchin & Coles, 1988) but a more recent view is that P3b reflects the decision-making process (Nieuwenhuis, Aston-Jones, & Cohen, 2005; Verleger, Jaśkowski & Wascher, 2005; Sussman & Steinschneider, 2009). Thus, together, the ERP components can be used to evaluate the difficulty of discriminating a speech contrast at different processing stages; at early, automatic stages (indexed by MMN), as well as during active discrimination of contrast differences that are volitionally detected (indexed by N2b/P3b). The earlier MMN component, with modality specific generators, reflects early (acoustic-phonetic) stages of speech processing to be compared with later attention-based components that integrate information from earlier stages (Picton, 1992). Thus, the elicitation of these components can distinguish neural processing associated with long-term language learning.

1.2. The Present Study

The present study examined whether selective attention to a non-phonemic (non-native) vowel duration contrast found in Japanese (but not English) would result in improvements in discrimination as indexed by ERP discriminative responses for American English (AE) listeners. Japanese (JP) listeners were expected to be highly automatic in discriminating vowel duration contrasts because these contrasts are phonemic in Japanese. American

English listeners, however, were expected to show less robust indices of automatic perception of vowel duration contrasts because duration does not serve as a primary phonemic cue in English. Previous behavioral studies have shown that AE listeners are often poorer than JP listeners in discriminating these contrasts (Enomoto, 1992; Toda, 1994; Muroi, 1995; Hirata, 2004a, 2004b, Hirata, 1990; Tajima, Kato, Rothwell, & Munhall, 2003; Hisagi and Strange, in press).

Previous studies have found smaller MMNs to segment duration contrasts for non-native as opposed to native listeners in tasks where participants ignore the auditory modality and attend to a muted video or read a book (Menning et al., 2002; Nenonen et al., 2003; Nenonen et al., 2005; Dehaene-Lambertz et al., 2000; Ylinen et al., 2006). Attention, however, was not manipulated in these studies. Furthermore, most of these ERP studies examined speech contrasts using synthetic speech or single instances of each vowel category (but see Dehaene-Lambertz et al., 2000). Thus, only physical differences among stimuli cued the phonetic contrast. An additional goal of the present study was to examine phonetic category discrimination of the vowel contrast using a *categorical or name-identity* discrimination task in which multiple exemplars of each category are presented (e.g., Shestakova, Brattico, Huotilainen, Galunov, Soloviev, Sams, Ilmoniemi, & Näätänen, 2002).

In this discrimination task, listeners are required to decide whether two physically different segments belong to the same or different phonetic categories without requiring the use of unfamiliar response labels. The present study used a “categorical oddball” discrimination task in which four natural-speech exemplars of each phonetic segment were presented as the “standard” and “deviant” categories. Nonsense words in both Japanese and English (Tado versus Taado) were chosen in order to control for possible lexical effects on brain and behavioral indices of discrimination.

Two experiments were conducted, one in which native and non-native listeners ignored the speech stimuli and attended to a visual oddball counting task (Exp 1) and a second in which they selectively attended to the speech contrast via an oddball counting task (Exp 2). Different groups of AE and JP listeners were used in the two experiments to preclude learning effects across attention conditions.

We predicted that in the Visual-Attend Condition (Exp 1), AE listeners would show absent or smaller MMNs than the JP group because processing of non-native contrasts is less automatic. In the Auditory-Attend Condition (Exp 2), it was predicted that AE listeners would show more robust MMNs because they could make use of attentional resources. However, their MMNs might be smaller in amplitude and/or later in latency than for the JP listeners. Furthermore the increase in MMN amplitudes might be more apparent over the right than the left hemisphere when AE listeners are attending, because they will be focusing on the acoustic rather than the linguistic structure of the stimuli. Little effect of attention was expected for the JP listeners because processing of these L1 contrasts should be highly automatic. In addition, the N2b and P3b components were expected to be elicited to the infrequent stimuli in both AE and JP listeners when they served as auditory targets. Of particular interest, was whether there would be evidence of greater improvement in discrimination of vowel duration with selective attention for the AE than for the JP group in the Auditory Attend compared to the Visual Attend Experiment. We also predicted that a behavioral discrimination task of vowel duration, given after the ERP measurements in the Visual Attend Experiment, would produce better than chance-level performance by the non-native group and ceiling performance by the native group. Behavioral discrimination of the contrasts would provide further information regarding the relationship between MMN and behavioral measures.

2 Results

2.1 Experiment 1: Visual-Attend Condition

2.1.1 Behavioral Results—Conditions are named according to whether the deviant was longer than the target (deviant increment) or was shorter (deviant decrement). For the behavioral responses on the visual counting task, 0-2 errors per block counted as “good” performance. The mean scores in Table 1A describe the number of good blocks per condition out of 14 possible. A Median test was conducted on the number of good blocks for the visual counting task to examine whether participants in the two groups were equally focused on the visual task. Three of the four comparisons showed no significant differences between groups (means greater than 12 good blocks, range 4 to 14). The AE group performed slightly worse on several blocks in the deviant increment condition where they were counting deviant hexagons among pentagons (AE mean = 11.92 (SD=2.35), range 5 to 14; JP mean = 13.58 (SD=0.52), range 4 to 14, see Table 1). Thus, most participants in both groups performed the task well (i.e., within two of the actual number of deviants in a block) for the majority of blocks. Participants also reported that identifying the visual deviants was challenging.

On the vowel-duration detection button-press task, following the ERP study, a Median test revealed significant differences between the AE and JP groups for deviant increment ($p < 0.001$), but not for the deviant decrement condition (see Table 1B). In the deviant increment condition, the JP group was more accurate than the AE group. Reaction Time (RT) was also examined by using a Median test and ANOVA (see Table 1C). There was no significant overall difference between the two groups ($p > 0.05$), but both groups were approximately 70 ms slower when identifying the deviant decrements than the deviant increments (paired t -test: $p < 0.001$).

2.1.2 Electrophysiological Results –Visual Attend Task—The grand averages of the ERPs to the standard and deviant at FZ and the subtraction waveforms at FZ and LM are shown for each language group in Fig. 1. This figure reveals a greater negativity to the deviant compared to the standard for both stimulus contrasts and for both language groups. Fig. 2 shows Global Field Power (standard deviation across all sites) comparing the standard and deviant for the deviant decrement (left) and deviant GFP for 40 ms intervals from 120 to 800 ms indicate significant differences between the standard and deviants from 120 to 140 ms, 200 to 240 ms and 320 to 800 ms for the deviant-increment condition and from 120-140 and 320-800 ms for the deviant-decrement condition ($df = 23$, $t > 2.03$). Comparisons of GFPs for the short and long standard conditions and short and long deviant conditions revealed no significant differences, except for the 200-240 ms interval (standards: $t = 2.32$, $p < 0.05$; deviants, $t = -1.97$, $p = 0.056$) (see Fig. 2). (Note that comparing the short standard and short deviant shows significant differences from 120-140 and from 320-800 ms and comparing the long standard to the long deviant reveals significant differences from 120-240 and 320-800 ms.) Forty-ms time intervals between 120 and 400 ms were selected for the following analyses examining group differences at specific sites. The later time intervals (400 to 800 ms) were not analyzed further because we were interested in the processes indexed by the earlier negativity (presumably MMN) and not the later negativity.

Statistical analyses were carried out for each contrast order separately (i.e., deviant increment, and deviant decrement) to determine whether these apparent differences between the deviant and the standard were significant for each language group.

Table 2 summarizes the results of the four-way ANOVAs (stimulus (standard, deviant) \times site (frontal, central) \times hemisphere (left, midline right) \times time (seven 40-ms intervals) for the two orders. Both groups of participants showed significantly greater negativity of the

deviant compared to the standard beginning around 160 ms for both groups and orders and ending around 300 ms for the deviant increment condition for both groups and after 300 ms for the deviant decrement condition (see Table 2 for latency range).

Four-way ANOVAs with language group as the between-subject factor and site (frontal, central), hemisphere (left, midline, right), and time (seven 40-ms intervals) as the within-subject factors were carried out using the subtraction waveforms (deviant minus standard). The deviant increment condition showed no main effect of group or interactions with group. Significant interactions of Hemisphere \times Time [F (10, 220) = 3.87, $p = 0.0008$] and Site \times Hemisphere \times Time [F (10, 220) = 3.30, $p = 0.008$] were observed. The frontal midline site was more negative than the central midline site by about 0.5 μ V, from 204 to 240 ms.

An analysis at the mastoid sites (LM and RM) revealed a significant hemisphere by time by Language Group interaction [F (5,110) = 2.97, $p = 0.015$]. A follow-up of this interaction revealed that the difference is found at the right mastoid (Time \times Language Group [F (5, 110) = 3.84, $p = 0.015$]. The JP group showed a larger positivity than the AE group from 200 to 240 ms time intervals ($M = 0.90$ vs. 0.44μ V, $SD = 0.88$ vs. 0.98μ V, respectively), whereas the positivity was smaller for the JP than AE group from 280-320 ms ($M = 0.02$ vs. 0.71μ V, $SD = 0.66$ vs. 0.91μ V). However, posthoc pairwise comparisons between the groups for each time interval did not show significant differences.

For the deviant decrement condition, significant interactions of Time \times Language Group [F (5, 110) = 2.93, $p = 0.016$], Hemisphere \times Time [F (10, 220) = 4.67, $p = 0.002$] and Hemisphere \times Time \times Language Group [F (10, 220) = 3.17, $p = 0.018$] were observed. Two-way analyses examining hemisphere and language group for each time separately reveal that there was a significant Group \times Hemisphere interaction for the time interval 284-320 ms [F (2, 44) = 3.27, $p = 0.05$]. The JP group showed a larger MMN than the AE group, especially at the left and midline sites.

For the mastoids, a significant Time by Language Group interaction was observed [F (5,110) = 2.68, $p = 0.025$]. Follow-up ANOVA's for each time interval showed a main effect of Language Group approaching significance for the 240-280 ms interval [F (1,22) = 3.47 $p = 0.075$]. The JP group tended to have greater positivity at the mastoids than the AE group.

In sum, the JP group showed a significantly larger MMN than the AE group, specifically at the left and midline sites for the deviant decrement, and at the right mastoid for the deviant increment.

2.2. Experiment 2: Auditory Attend Task

2.2.1 Behavioral Results—The same analysis and criteria were used as in Exp 1. Table 3A shows the results for the behavior on the counting task during the ERP recording. The JP group appeared to perform slightly better than the AE group overall; however, Median tests showed no significant language-group difference for any condition ($p > 0.05$).

Table 3B shows the descriptive statistics for performance on the button-press vowel duration task that followed the ERP experiment. Median tests revealed that there were significant differences between the AE and JP groups for both conditions (tado: $p < 0.0001$; taado: $p < 0.05$); in both comparisons, the JP group was more accurate than the AE group. Reaction Time (RT) was also examined using Median tests and an ANOVA (see Table 3C). There was no overall significant difference in RT between the two groups ($p > 0.05$), but again both groups were significantly slower at identifying deviant decrements, than deviant increments by 50 to 60 ms (paired t-test: $p = 0.001$).

2.2.2 Electrophysiological results—The grand averages of the ERPs to the standard and deviant at FZ and the subtraction waveforms at FZ and LM are shown for each language group in Fig. 3. Two-tailed *t*-tests using the 24 participants comparing Global Field Power (GFP) for 40 ms intervals from 120 to 400 ms indicate significant differences between the standard and deviants from 200 to 240 and 280 to 800 ms for the deviant-increment condition and from 320-800 ms for the deviant-decrement condition ($df=23$, $t > 2.03$), as shown in Fig. 4. Comparisons of GFPs for the short and long standard waveforms and short and long deviant waveforms revealed no significant differences, except for the deviant waveforms in the 200-240 ms interval ($t = -2.28$, $p = 0.027$) (see Fig. 4). (Note that comparing the short standard and short deviant shows significant differences from 320-400 ms, and approaching significance from 160-200 ms, and comparing the long standard to the long deviant reveals significant differences from 320-400 ms.). Forty-ms time intervals between 164 and 400 ms were selected to include these intervals for the ANOVAs examining specific sites for the MMN and from 400 to 600 ms to examine the P3b component.

Table 4 summarizes the results of the four-way ANOVAs (stimulus (standard, deviant) \times site (frontal, central) \times hemisphere (left, midline right) \times time (seven 40-ms intervals) for the two orders. Both groups of participants showed significantly greater negativity of the deviant compared to the standard beginning around 164 ms and ending around 240 ms for the deviant increment, and beginning around 244 ms and ending around 340 ms for deviant decrement (see Table 4 for latency range).

Four-way ANOVAs with group as the between-subject factor and site (frontal, central), hemisphere (left, midline, right), and time (seven 40-ms intervals) as the within-subject factors were carried out using the subtraction waveforms (deviant minus standard). The deviant increment condition revealed significant interactions of Site \times Language Group [$F(1, 22) = 5.79$, $p = 0.025$] and Hemisphere \times Language Group [$F(2, 44) = 3.50$, $p = 0.039$]. The AE group showed the greatest negativity at central sites while the JP group showed the greatest negativity at frontal sites; additionally the JP group generally showed greater negativity at the left than the midline or right sites, whereas the AE group showed equivalent negativity across these sites. However, post-hoc pairwise comparisons of the groups for each site and hemisphere were not significant. Significant two-way interactions of Site \times Hemisphere [$F(2, 44) = 6.59$, $p = 0.003$] and Site \times Time [$F(5, 110) = 3.89$, $p = 0.003$] and a three-way interaction of Site \times Hemisphere \times Time [$F(10, 220) = 2.59$, $p = 0.05$] were also found, but are not examined further in this paper because they did not include group as a factor.

The deviant decrement condition revealed a significant interaction of Time \times Language Group [$F(5, 110) = 2.51$, $p = 0.03$]: the AE group showed a slightly larger MMN than the JP group in the intervals 284-320 ms and 324-360 ms although post-hoc comparisons of the language groups for each time-interval were not significant. Significant interactions of Site \times Time [$F(5, 110) = 3.86$, $p = 0.003$] and Hemisphere \times Time [$F(10, 220) = 2.48$, $p = 0.0078$] were seen, but will not be examined further because they did not include a group difference.

Analysis for the mastoids was also carried out to more clearly determine whether the groups differed in MMN amplitude. The MMN, but not the N2b, inverts in polarity, so an increase in amplitude at frontocentral sites due to N2b would not be reflected at the mastoids. The ANOVA examining the mastoids alone revealed no significant interactions including language group for deviant increment ($F < 1.73$, $p > 0.13$) or deviant decrement ($F < 0.60$, $p > 0.44$).

A three-way analysis (language group (AE vs. JP), hemisphere (left as P3 vs. midline as PZ vs. right as P4, and time in five 40 ms time windows between 400 and 600 ms)) of the subtraction waveforms to examine the P3b component revealed for the deviant increment condition, a significant interactions of Stimulus \times Hemisphere [F (2, 44) = 10.752, p = 0.0002], and Stimulus \times Time [F (4, 88) = 6.118, p = 0.0002]. For the deviant decrement condition, significant interactions of Stimulus \times Hemisphere [F (2, 44) = 5.273, p = 0.0089] and Stimulus \times Time [F (4, 88) = 27.75, p = 0.0000] were seen. P3b was largest for the midline, PZ site. Neither order revealed significant differences between AE and JP groups.

2.3 Experiment 1 vs. Experiment 2

The results from the Visual-Attend and Auditory-Attend Conditions were compared directly to determine whether attention to the target auditory stimuli led to larger MMNs for the AE or JP group. The subtraction waveforms (deviant – standard) were used for all comparisons and the two orders were collapsed.

Five-way ANOVAs with Task (Visual vs. Auditory), Language group (AE vs. JP), Site (frontal vs. central), Hemisphere (left vs. midline vs. right), and Time (Time 1: 204-240 ms; Time 2: 244-280 ms) were carried out. Interactions of Time \times Task \times Language Group [F (1, 44) = 5.20, p = 0.03], and Site \times Time \times Task [F (1, 44) = 5.33, p = 0.03] were observed. The AE group in the Auditory-Attend Condition showed a larger negativity than the AE group in the Visual-Attend Condition. The JP groups in the two tasks showed little difference in amplitude for Time 1 or Time 2, and even showed a tendency for the negativity to be slightly larger for the JP group in the Visual-Attend Condition for Time 2. The Site \times Time \times Task interaction was the result of an increase in negativity for both frontal and central sites for the early time interval and for the central sites in the latter time interval (244-288 ms), as shown in Table 5. Examination of the means for the language groups separately shows that the AE participants show greater negativity in the Auditory Task for both time intervals and both sites. Many of the JP participants, however, showed equal or larger negativity for the Visual than the Auditory Task at both frontal and central sites from 244-288 ms.

We also examined how many participants in each group showed robust MMNs. For the Auditory Task, 9/12 AE listeners and 10/12 JP listeners showed MMN amplitudes at frontal or central sites greater than $-0.6 \mu\text{V}$. For the Visual Task, 8/12 JP, but only 4/12 AE listeners showed amplitudes greater than $-0.6 \mu\text{V}$. All but one JP listener had amplitudes greater than $-0.3 \mu\text{V}$. In contrast four AE listeners in the Visual Task had no evidence of MMN, that is, either positive amplitudes or amplitudes within $0.15 \mu\text{V}$ of zero.

2.4. MMN Topography

Fig. 5 displays voltage maps at the peak of the MMN (deviant - standard) to the Deviant Increments and Deviant Decrements for the AE and JP groups in the Visual and Auditory Tasks. In all cases, a large negativity (blue) is found at fronto-central sites that inverts in polarity (red) at inferior sites.

We calculated the Global Dissimilarity Index (GDI) between the topographies by calculating the square root of the mean of the squared differences between all corresponding electrodes after normalizing the data by dividing the mean voltage by its own GFP (i.e., standard deviation) (Lehman & Skrandies, 1980) (see Fig. 6). Greater dissimilarity is seen as values approaching 2, whereas greater similarity is shown by values closer to zero. Note that GDI is related to Pearson's correlation coefficient (r) by $2*(1-r)$. The pairwise comparisons between the AE and JP for each task and order (decrement and increment) and between the tasks for each language group and order reveal highly similar topographies to the Deviant Increment

for all group and tasks near the MMN peak latency ($GDI < 0.5$; $r > 0.75$). GDI of 0.5 is equivalent to $r = 0.75$ with confidence intervals of 0.62 to 0.84 ($df = 64$). GDIs less than 1.44 (equivalent to greater than $r = 0.29$) have an upper bound for the confidence interval of 1.87. Any value above 1.87 indicates that the points are not positively correlated. For the Deviant Decrement, the topographies around the MMN peak showed GDIs less than 0.5 for all but the comparison between the AE and JP groups during the Visual Task where the GDI at the peak was 0.6 ($r = 0.7$). In the later time intervals after 400 ms, the GDI indicates similarity in topography ($GDI < 1$, $r > 0.5$) for the AE and JP group only during the Auditory Task. This indicates that the P3b topography was highly similar for the two groups.

3. Discussion

A major goal of this study was to examine whether selective attention to a vowel duration contrast would show improved discrimination for non-native listeners for whom the contrast was non-phonemic. As predicted, native listeners showed little difference in processing the vowel duration contrast related to attentional level, presumably because processing of this phonemic duration contrast is highly automatic. Furthermore, the non-native AE group selectively attending to the vowel duration contrast showed a larger negativity than when ignoring the auditory information and performing a visual attention task. In terms of the number of participants in each group showing robust MMNs, at least three quarters showed MMN amplitudes greater than $0.6 \mu V$ for all but the AE participants in the task where the attention was focused on visual stimuli. For this group, only four of the 12 participants showed robust MMNs, suggesting automatic perception of the vowel duration differences. These findings will be discussed below in relation to the previous literature and our proposed model of native versus non-native speech perception.

3.1. Cross-linguistic differences in speech perception

Our study adds to a growing number of investigations that have demonstrated larger MMNs to a phoneme contrast for native compared to non-native listeners (Näätänen, et al., 2007, for review). These include a number that have focused specifically on duration (or quantity) as a phonemic feature (Nenonen et al., 2003; Nenonen et al., 2005; Menning et al., 2002; Kirmse, Ylinen, Tervaniemi, Vainio, Schröger, and Jacobsen, 2008; Tervaniemi, Kruck, Baene, Schröger, Alter, & Friederici, 2009). All of these studies used a passive oddball design, in which the participant is asked to ignore the auditory modality and watch a video with the sound muted. This task is intended to direct attentional resources away from the speech information, and thus allows observation of pre-attentive automatic processing. However, because performance on the video-watching task is not measured, it cannot be ascertained to what extent attention is drawn away from the speech stimuli for the native versus non-native groups. Thus, it is possible that listeners for whom the contrast is native are extending some resources to processing the speech, leading to a larger (or earlier) MMN. To minimize this possibility and to provide for a measure of performance on the distracting task, we chose to have participants perform a visual oddball task that required sustained attention. Both AE and JP listeners showed accurate counting of visual oddballs, suggesting that they were focusing attentional resources away from the auditory modality. The JP group actually showed slightly better performance on the visual oddballs, indicating that they were not extending more resources to the auditory modality than the AE group. Thus, our results can be interpreted as a strong indicator that the amplitude difference in the MMN between the AE and JP groups indicates higher automaticity for the JP than the AE group. Given that MMN reflects processing at a level that minimizes sensory-processing load (Horvath et al., 2008), the JP group will have more cognitive resources for focusing on other levels of linguistic processing when listening to their native language.

3.2 Automatic Selective Perception

Our results are compatible with a model that suggests that native language speech processing is highly automatic. This view has been implicit in most developmental and L2 models of speech perception, but generally has not been clearly articulated. Developmental models describe the acquisition of first language speech categories as learning to selectively attend to the relevant cues in the speech signal (e.g., Jusczyk, 1997) or making a neural commitment to first language speech categories (e.g., Kuhl, 2008). Explicitly postulating that first language speech perception is a matter of automatizing Selective Perceptual Routines allows us to ask at what level of processing this automatization occurs. Kuhl (2008) argues that infants show neural commitment to their native language early in development, but the notion of “neural commitment” is not defined. In our view, neural commitment is developing automatic Selective Perceptual Routines so that they are optimized for processing the first language. In a neurophysiological model, this consists of creating explicit representations (instantiated in neural connectivity) of the set of relevant acoustic features in auditory cortex. By doing this, first language speech perception becomes highly automatic, efficient, and requires few cognitive resources (low cognitive load). The drawback to this automatization is that speech perception of non-native contrasts that relies on phonetic cues that are not used or used differently in the L1 requires increased cognitive resources, because selecting the relevant L2 cues is not automatic and because L1 Selective Perceptual Routines may actively interfere with detecting L2 cues (e.g., Iverson et al., 2005).

Using event-related potentials to test this hypothesis allowed us to ascertain that the automaticity or “neural commitment” to L1 speech perception is registered at the level of auditory cortex indexed by MMN. It also revealed that selectively attending to the relevant contrast (in this case, vocalic duration) improves the deviance detection process. This improvement may be the result of sharpening the representation of the standards or the deviants, or both. Other studies have shown that attention can lead to increases in MMN to speech, but it appears that this may only be the case for more difficult discriminations (e.g., Sussman et al., 2004; Gomes et al., 2000). For the Japanese listeners, there was no attentional advantage, and thus no increase in MMN was found when actively discriminating the speech sounds, indicating that detection of this contrast is fully automated for them (Sussman et al., 2009).

On the other hand, AE listeners appeared to be making greater use of controlled processes to identify deviant targets than JP listeners. This was indicated by the more central contribution of the negativity (consistent with N2b) evoked by the deviant during the Auditory-Attend Task for AE compared to JP listeners. The finding that amplitudes at the mastoid sites were not different between the language groups in the Auditory Attend Task, however, suggests that the increased negativity found at frontal central sites with attention was not entirely a function of the controlled processes indexed by N2b and that attention enhanced the processes indexed by MMN. The AE listeners' performance, on average, was not quite as good as the JP group in behaviorally identifying the targets, though most participants were within two of the actual correct count on over 75% of the blocks. The amplitude of the P3b did not distinguish the groups at this level of accuracy in target identification, which may suggest that they were able to sufficiently update their memory representations to allow for accurate target identification (Horvath et al., 2008). A number of studies have shown that increasing task difficulty or stimulus variability leads to greater decrements in performance for non-native compared to native listeners of speech duration contrasts (e.g., Enomoto, 1992; Toda, 1994; Muroi, 1995; Hirata, 2004a, 2004b; Hirata, 1990; Tajima et al., 2003; Hisagi & Strange, in press). Dehaene-Lambertz and colleagues (2000) showed that JP listeners had poor discrimination (no MMN/N2b or P3b) of a French phonemic contrast in a study where they used multiple exemplars for the standard and deviants that included differences in speaker voice (male vs. female). There was some variability in our stimuli, with the use of

four different exemplars from each phoneme category, but all were produced by the same speaker. The robust amplitude of the ERP components in the Auditory Attend Task (MMN/N2b, and P3b) evoked by deviants for both the AE and JP groups suggests that this level of variability was not detrimental to instantiating in memory a standard template and determining that the four deviant exemplars were different (see Shestakova et al, 2002). Increasing the task difficulty (e.g., by increasing stimulus variability, or adding noise) would be expected to affect non-native listeners to a greater extent and require greater attentional resources (and, presumably, engagement of parietal attentional networks) for successful target attention compared to native listeners.

In sum, our findings support the claim that L1 speech perception is fully automated, and that this automatization is implemented in auditory cortex, as it can be indexed by MMN. Our results also suggest that attentional focus to speech contrasts can result in improved deviance-detection by listeners for whom the contrast is non-native and non-phonemic.

3.3 Perceptual Asymmetry Effects

An asymmetry in the latency of the behavioral responses was seen between the different conditions (deviant increment versus deviant decrement). The latter condition led longer RTs for both language groups. Comparisons of GFP indicated slightly earlier peaks (corresponding to the P2) for the short stimulus (Tado) than the long stimulus (Taado), which led to differences in the onset of the MMN for the deviant increment and deviant decrement conditions. Similar asymmetries have been observed in other studies (Shafer et al., 2004; Friedrich, Weber, and Friedrich, 2004; Kirmse et al., 2008). Comparing the longer deviant to itself in the standard order (and visa versa for the longer deviant) leads to no differences in the onset of the MMN, indicating that the durational differences between the stimuli in the blocks led to the apparent timing difference on MMN. These asymmetries are likely to be related to refractoriness of neurons indexed by P1 and N1 (Näätänen, 1990). Whatever the neurophysiological explanation for this asymmetry, the later RTs in the behavioral task suggest that duration decrements are more difficult to process than duration increments. In our study, we showed statistically significant differences between the AE and JP groups in the Visual Attention Task for the harder discrimination (deviant decrement). It is possible that the deviant increment was sufficiently easy that group differences become less apparent. In other words, the duration increment may have been sufficiently salient (and thus explicitly represented) that it allowed for automatic recognition by many of the AE listeners (see, Koch, 2005). In contrast, the more difficult order (deviant decrement) is explicitly represented only for the JP group who have overlearned the relevant cues for making this discrimination at an automatic level (Crick & Koch, 1991).

3.4 Neural sources for native-language speech perception

A number of other studies have also shown MMN enhancement to native-language phonetic contrasts compared to non-native (or non-distinctive) phonetic contrasts specifically over left hemisphere fronto-central sites (e.g., Kasai, Yamada, Kamio, Nakagome, Iwanami, Fukuda, Itoh, Koshida, Yumoto, Iramina, Kato, & Ueno, 2001; Kirmse et al., 2008). This pattern is best explained by neural generators in left hemisphere auditory cortex leading to more robust deviance detection. Kirmse et al. (2008) showed enhanced MMN amplitudes of the left frontal sites to vowel duration differences in pseudo words for Finnish but not German listeners. Finnish makes use of duration as a primary phonemic cue, whereas, German uses length differences as a secondary cue accompanying spectral differences between vowels or serving as a cue for stress. Similarly, vowel duration can serve as a secondary cue in English, for voicing for stop consonants (vowels are lengthened preceding voiced final consonants), Borden, Harris, & Raphael, 2003). Our results are in agreement with Kirmse et al. (2008), in suggesting that increased automaticity in processing L1

phonetic cues is implemented to a greater extent in left compared to right auditory cortex. This implies that during first language development, explicit representations, in terms of changes in neural connectivity, are initiated for L1 speech categories (i.e., phonemes) in left hemisphere auditory cortex. This does not preclude changes in connectivity in right hemisphere regions, but suggests that for some phonetic cues, the changes may be greater in the left hemisphere. The topographical comparisons (GDI) revealed that JP and AE groups had greater topographical differences for the more difficult contrast (deviant decrement). This finding is consonant with the suggestion that the JP group has increased left hemisphere involvement in processing the contrast. It will be necessary to use methods with better spatial resolution (e.g., functional Magnetic Resonance Imaging) to confirm whether the findings at the left frontal-central sites are best explained by sources in the left hemisphere.

4. Experimental Procedures

4.1. Stimulus Materials

Auditory stimuli—The stimuli were recorded by a native female speaker of Japanese (Tokyo dialect), in sentence context to maintain the naturalness of each token. The output was stored as wave files on a PC1. An acoustic analysis of the four extracted target nonsense words for each category (Tado and Taado) revealed that the duration ratio of the long to short target vowels (/aa/ in /taado/ vs /a/ in /tado/) was 1.61 on average, with durations of individual tokens varying from 83 ms to 88 ms for /a/ (mean: 86 ms) and from 128 ms to 148 ms for /aa/ (mean: 138 ms). The mean ratio of long vs short total word duration (from onset of /t/ to offset of /o/) was 1.21. The mean for /tado/ was 206 ms (range: 202-214 ms) and for /taado/ was 250 ms (range: 237-264 ms). In general, these differed minimally and overlapped on other cues (intensity and fundamental frequency) for the short and long stimuli and were comparable to duration ratios for real Japanese words, as ascertained by pilot measurements of the words [kado] ‘corner’ vs [kaado] ‘card’.

Familiarization task materials—The familiarization task materials for AE participants were nonsense words, /kæpi/ and /kɛpi/ produced by a native female speaker of AE (New York dialect) in the same manner as the JP stimuli. For the JP participants, familiarization words were nonsense words, /kuto/ and /keto/.

Visual stimuli—Four different sizes of pentagon and hexagon shapes (orange with black background) were created for the construction of the oddball visual discrimination test.

4.2 Participants

Twenty-four native speakers of Japanese, who had lived in the United States less than 36 months, served as the native-language group. Twelve participated in each Experiment (Exp 1: two males and ten females; Exp 2: two males, ten females). The JP participants were native speakers of the Tokyo (Kanto) dialect, who had had standard English education in Japan with minimal conversational experience with native English speakers. Twenty-four native speakers of American English, who had never studied Japanese and had no strong second language background, participated as non-native listeners (Exp 1: four males, eight females; Exp 2: four males, eight females). All but one participant (American male) were right-handed. Each participant was asked to fill out a consent form and a language background questionnaire before the experiment. For the AE participants, foreign language experience with Spanish or French in junior-high, high school and/or college was accepted because these languages do not use temporal cues to distinguish vowels. All participants

¹Please see Hisagi (2007) and Hisagi & Strange, (in press) for the details on the stimulus selection process.

were between 21 and 40 years of age with normal hearing (at 500, 1000, 2000, 4000 Hz at 20 dB HL). All participants were paid to participate in the experiment.

4.3 Paradigms and Procedures

All testing took place in a 9' × 10' soundproof, electrically-shielded booth. Using E-prime to control the experiment, stimuli were presented over speakers at 75 dB SPL. Participants were seated in a comfortable chair. A consonant contrast was also tested in this session but is not reported in this paper (see Hisagi, 2007). Thus, each participant was tested for approximately 3.5 hours (for both vowel and consonant duration contrasts) with frequent breaks.

ERP tasks—A categorial oddball paradigm was used with standards presented on 85% and deviants on 15% of the trials. Words occurred with inter-stimulus intervals (ISI) of 800 ms. In both experiments, participants received two conditions: the long stimulus as deviant (deviant increment) and the short stimulus as deviant (deviant decrement) in counterbalanced order. Trains of stimuli were presented in fourteen blocks for each of the 4 conditions (two of which were consonants, which are not discussed in this paper). Each block had a different number of standards and deviants (average: 100 trials per block). Participants received a randomized order of blocks for each condition. Each block lasted approximately 1-2 minutes. Each condition took approximately 30 minutes. The total number of trials for each condition was 1400 (1190 standards + 210 deviants).

For Exp 1, participants performed a visual oddball task and were instructed to ignore the auditory speech sounds, silently count visual deviants and report the number to the researcher at the end of each block. The shape categories (pentagon versus hexagon) both served as standards and deviants in two different conditions. The ISI for the shapes (780 ms) was different from that of the vowels in order to make sure that the onset of the auditory and visual stimuli was not synchronized.

For Exp 2, participants performed an auditory oddball task in which they were instructed to count the short vowel deviants in the deviant decrement condition, and the long-vowel deviants, in the deviant increment condition. The visual stimuli were also presented, but with no instructions other than to look at the screen while listening to the auditory targets.

At the beginning of the experiment, detailed oral instructions about the visual shapes or the speech targets (using the familiarization words) were given to make the participants aware of what they needed to do. Both auditory and visual stimuli were presented.

Behavioral Task—After the end of the ERP session, a ten-minute behavioral discrimination task using the same auditory stimuli was given to assess participants' behavioral performance. Participants were asked to press a button when they heard a deviant stimulus. In this task one block of 100 trials (85 standards and 15 deviants) was presented for each order. The order of short versus long stimuli as deviants was counterbalanced. The stimulus order and ISI of the vowel stimuli were identical to those of the electrophysiological task.

4.4 Electroencephalogram (EEG) Recording

A Geodesic net of 65 electrodes (wrapped in sponges) was placed on the participant's scalp, after soaking in saline solution for five minutes. The Vertex served as the reference during data collection. Vertical and horizontal eye movements were monitored from frontal electrodes Fp1 (left) and Fp2 (right) and electrodes placed below each eye. Impedance was

maintained below 40 k Ω , which is acceptable for the high-impedance Geodesic amplifiers (200 M Ω ; Ferree, Luu, Russell & Tucker, 2001).

The EEG was amplified with a bandpass of 0.1 to 30 Hz, using Geodesic Amplifiers. A Geodesic software system (NetStation version 3.0) in continuous mode was used to acquire the data at a sampling rate of 250 Hz per channel for later off-line processing. During data acquisition, all channels were observed by an experimenter to monitor each participant's state, artifacts due to electrical interference, defective electrode contacts, and/or excessive muscle movement.

4.5 Data Analysis

The continuous EEG was processed off-line, using a lowpass filter of 20 Hz. The EEG was segmented into epochs with an analysis time of 1000 ms post-stimulus and a 100 ms pre-stimulus baseline. The data were transferred into an EEGLAB program (Matlab-based; toolbox: Delorme & Makeig, 2004) to perform an Independent Component Analysis (ICA) (Bell and Sejnowski, 1995; Glass, Frishkoff, Frank, Davery, Dien & Maloney, 2004) for eye blink correction. The algorithm decomposes the signal into components and removes those components highly correlated with an eye-blink component ($r > 0.9$), and then recomposes the signal minus these components. The corrected data were then baseline corrected and examined for artifacts, using Netstation software. An epoch for a channel was marked bad if the fast average amplitude exceeded 200 μ V; if the differential amplitude exceeded 100 μ V; or if it had zero variance. A channel was marked as bad if greater than 20 % of the total epochs were marked as bad. An epoch was rejected if more than 3 channels for that epoch were marked as bad. The data had no more than 10% of trials lost on average due to artifacts. Channels with high artifact rates on greater than 20% of trials were replaced by spline interpolation using adjacent sites. ERP averages were calculated for each stimulus type (standard, deviant) and baseline corrected using the 100-ms pre-stimulus activity. ERP averages were re-referenced to an average-reference. Finally, the difference waveforms were created by subtracting the standard waveform from the deviant waveform for each participant.

Global Field Power (GFP) was used to help select the time intervals used in the ANOVAs (Lehmann and Skrandies, 1980). GFP is the standard deviation calculated from the mean of all 65 channels as a function of time. A peak of GFP may reflect a maximum of underlying dipolar brain activity that contributes to the surface potential field and is useful for identifying time intervals for analyzing a component of interest (e.g., Shafer, Ponton, Datta, Morr, & Schwartz, 2007). Two-tailed t-tests ($p < 0.05$) were used to determine the relevant latency range for ANOVAs planned to examine specific sites (see Figs. 2 and 4).

For each stimulus condition (tado-as-standard paired with taado-as-deviant, and taado-as-standard paired with tado-as-deviant²), four-way ANOVAs with stimulus (standard vs. deviant), site (frontal: F3, FZ, F4 vs. central: C3, CZ, C4), hemisphere (left vs. midline vs. right), and time in 40 ms time windows (120-160; 160-200; 200-240; 240-280; 280-320; 320-360 ms) were carried out to establish whether there were significant differences in stimulus conditions that were consistent with MMN timing and topography for each group separately. A significance level of 0.01 is used for this set of analyses to minimize spurious findings. Hemisphere was included as a factor in the analyses because previous research suggests that the greatest language-related differences might be seen at left hemisphere sites (Näätänen, et al., 2007). Time was included to determine the temporal extent of the MMN.

²Comparing the tado as standard to tado as deviant and taado as standard to taado as deviant leads to the same pattern of results.

After establishing whether MMNs were present for the two groups, additional ANOVAs using subtraction waveforms (deviant – standard) were undertaken to directly compare the groups and included site, hemisphere and time, where appropriate. A significance level of 0.05 was used for these comparisons because they directly test the questions of interest. Stepdown analyses followed up on significant interactions, and Tukey's Honestly Significantly Different (HSD) post-hoc tests were used for post-hoc pairwise comparisons. The Greenhouse-Geisser correction was applied to correct for violations of sphericity when necessary.

It was expected that N2b and P3b components would be observed in Exp 2 (Auditory-Attend task) following or partially overlapping the MMN. The N2b was expected to be seen as greater negativity of the deviant than standard; it is largest at superior central sites (top, middle of the head) and does not invert in polarity at the mastoids (behind the ears, and thus beneath the temporal lobes). Thus, comparing the amplitude for the two groups at the mastoids would reveal whether there are differences in MMN alone. An ANOVA with group, site (left mastoid (LM), right mastoid (RM)) and time (six times) was carried out to examine these sites.

To examine the P3b amplitude and latency, three-way ANOVAs of language group (AE vs. JP), hemisphere (P3 (left parietal) vs. PZ (midline parietal) vs. P4 (right parietal), and time in 40 ms time windows between 400-600 ms (400-440; 440-480; 480-520; 520-560; 560-600) were conducted.

For the behavioral tasks (counting and button-press responses), the number of correct responses was calculated for the analyses. Median and Wilcoxin Matched Pairs tests were used to compare differences across groups and within groups respectively (Siegel and Castellan, 1988). For the button-press task, reaction time (RT), measured from the onset of the deviant stimuli, was also examined, using repeated measures ANOVAs. RT (in ms) was analyzed using the correct responses only.

5. Conclusion

The present study supports the claim that, at the level of the detection of relevant acoustic-phonetic information for categorizing phonemic segments, L1 speech perception is fully automatic and that focal attention has little or no effect on differentiation of native phonemes, at least when the stimulus materials are simple. Left hemisphere auditory cortex appears to be the locus for neural instantiation of L1 Selective Perception Routines. In contrast, perception of non-native contrasts by adults was less automatic and required more attentional focus for successful performance. This increase in cognitive load was shown to result in poorer performance compared to native speakers. However, it remains an open question to what extent L2 Selective Perception Routines become automatic with immersion experience or with intensive training. Paradigms that assess brain responses will be important for evaluating changes in automaticity of perception with language learning in adulthood.

Acknowledgments

This publication was made possible by grant number 01-113 BCS from the National Science Foundation (NSF) Dissertation Enhancement Award to Miwako Hisagi, Valerie Shafer and Winifred Strange, and grant number HD-46193 from National Institute of Child Health and Human Development (NICHD) at the National Institutes of Health (NIH) to Valerie Shafer, DC-00323 from National Institute of Deafness and Other Communicative Disorders (NIDCD) of NIH to Winifred Strange, and DC-004263 from The National Institute on Deafness and Other Communicative Disorders (NIDCD) of NIH to Elyse Sussman. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of NSF or NIH.

Literature references

- Alho K, Winkler I, Escera C, Huotilainen M, Virtanen J, Jaaskelainen IP, Pekkonen E, Ilmoniemi RJ. Processing of novel sounds and frequency changes in the human auditory cortex: magnetoencephalographic recordings. *Psychophysiology*. 1998; 35(2):211–24. [PubMed: 9529947]
- Bell AJ, Sejnowski TJ. An information maximisation approach to blind separation and blind deconvolution. *Neural Computation*. 1995; 7(6):1129–1159. [PubMed: 7584893]
- Best, CT. A direct realist view of cross-language speech perception. In: Strange, W., editor. *Speech perception and linguistic experience: Issues in cross-language research*. York Press; Baltimore: 1995. p. 171-204.
- Borden, GJ.; Harris, KS.; Raphael, LJ. *Speech Science Primer: Physiology, Acoustics, and Perception of Speech*. 4th. Lippincott Williams & Wilkins; New York: 2003.
- Crick F, Koch C. Towards a neurobiological theory of consciousness. *Seminars in Neuroscience*. 1990; 2:263–275.
- Dehaene-Lambertz G, Dupoux E, Gout A. Electrophysiological correlates of phonological processing: A Cross-linguistic Study. *Journal of Cognitive Neuroscience*. 2000; 12:635–647. [PubMed: 10936916]
- Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*. 2004; 134:9–21. [PubMed: 15102499]
- Donchin E, Coles MGH. Is the P300 component a manifestation of cognitive updating? *The Behavioral and Brain Sciences*. 1988; 11:357–427.
- Enomoto K. Interlanguage Phonology: The Perceptual Development of Durational Contrasts by English-Speaking Learners of JP. *Edinburgh working papers in applied linguistics*. 1992; 3:25–35.
- Ferree TC, Luu PL, Russell J, Tucker DM. Scalp-electrode impedance, infection risk, and EEG data quality. *Clinical Neurophysiology*. 2001; 112(3):536–544. [PubMed: 11222977]
- Flege, J. Second language speech learning: Theory, findings and problems. In: Strange, W., editor. *Speech perception and linguistic experience: Theoretical and methodological issues*. York Press; Timonium, MD: 1995. p. 233-277.
- Friedrich M, Weber C, Friedrich AD. Electrophysiological evidence for delayed mismatch response in infants at-risk for specific language impairment. *Psychophysiology*. 2004; 41(5):772–782. [PubMed: 15318883]
- Glass, K.; Frishkoff, GA.; Frank, RM.; Davery, C.; Dien, J.; Maloney, AD. A framework for evaluating ICA methods of artifact removal from multichannel EEG. In: Puntinet, CG.; Prieto, A., editors. *ICA 2004, LNCS 3195*. Springer-Verlag; Berlin & Heidelberg: 2004. p. 1033-1040.
- Gomes H H, Sussman E, Ritter W, Kurtzberg D, Cowan N, Vaughan HG Jr. Electrophysiological evidence of developmental changes in the duration of auditory sensory memory. *Dev Psychol*. 1999; 35:294–302. [PubMed: 9923483]
- Hirata Y. Perception of Geminated Stops in Japanese Word and Sentence Levels By English-speaking Learners of Japanese Language. *Nihon Onsei Gakkai*. 1990:4–10.
- Hirata, Y. *Journal of the Acoustical Society of America*. Vol. 116. 2004a. Training native English speakers to perceive Japanese length contrasts in word versus sentence contexts; p. 2384-2394.
- Hirata Y. Effects of speaking rate on the vowel length distinction in Japanese. *Journal of Phonetics*. 2004b; 32(4):565–589.
- Hisagi, M. Dissertation. *Speech and Hearing Sciences, The City University of New York – Graduate Center*; 2007. Perception of Japanese temporally-cued phonetic contrasts by Japanese and American English Listeners: Behavioral and electrophysiological measures.
- Hisagi M, Strange W. Perception of Japanese Temporally-Cued Contrasts by American English Listeners. *Language and Speech*. in press.
- Horváth J, Roeber U, Bendixen A, Schröger E. Specific or general? The nature of attention set changes triggered by distracting auditory events. *Brain Research*. 2008; 1229:193–203. [PubMed: 18634759]

- Iverson P, Hazan V, Bannister K. Phonetic training with acoustic cue manipulations: a comparison of methods for teaching English /r/-/l/ to Japanese adults. *Journal of the Acoustical Society of America*. 2005; 118(5):3267–78. [PubMed: 16334698]
- Jusczyk, PW. *The Discovery of Spoken Language*. Cambridge: MIT Press; 1997.
- Kasai K, Yamada H, Kamio S, Nakagome K, Iwanami A, Fukuda M, Itoh K, Koshida I, Yumoto M, Iramina K, Kato N, Ueno S. Brain lateralization for mismatch response to across- and within-category change of vowels. *Neuroreport*. 2001; 12:2467–2471. [PubMed: 11496131]
- Kirmse U, Ylinen S, Tervaniemi M, Vainio M, Schröger E, Jacobsen T. Modulation of the mismatch negativity (MMN) to vowel duration changes in native speakers of Finnish and German as a result of language experience. *Int J Psychophysiol*. 2008 Feb; 67(2):131–43. [PubMed: 18160160]
- Kuhl, PK. *Philosophical Transactions of the Royal Society B*. Vol. 363. 2008. Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e); p. 979-1000.
- Koch, Christof. *The Quest for Consciousness: A Neurobiological Approach*. Roberts and Company Publishers; Englewood, Colorado: 2005.
- Lehmann D, Skrandies W. Reference-free identification of components of checkerboard-evoked multichannel potential fields. *Electroencephalography and Clinical Neurophysiology*. 1980; 48:609–621. [PubMed: 6155251]
- MacKain KS, Best CT, Strange W. Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics*. 1981; 2:369–390.
- Menning H, Imaizumi S, Zwitserlood P, Pantev C. Plasticity of the human auditory cortex induced by discrimination learning of non-native, mora-timed contrasts of the JP language. *Learning Memory*. 2002; 9:253–267. [PubMed: 12359835]
- Muroi K. Problems of Perception and Production of JP Morae –The Case of Native English Speakers. (written in Japanese). *Sophia Linguistica*. 1995; 38:41–60.
- Näätänen R. Implications of ERP data for psychological theories of attention. *Biological Psychology*. 1988; 26(1-3):117–163. [PubMed: 3061477]
- Näätänen R. The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive functions. *The Behavioral and Brain Sciences*. 1990; 13:201–288.
- Näätänen R, Lehtokoski A, Lennes M, Cheour M, Huotilainen M, Iivonen 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, Tervaniemi M, Sussman E, Paavilainen, Winkler I. “Primitive intelligence” in the auditory cortex. *Trends in Neuroscience*. 2001; 24:283–288.
- Näätänen R, Paavilainen P, Rinne T, Alho K. The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*. 2007; 118:2544–2590. [PubMed: 17931964]
- Nenonen S, Shestakova A, Huotilainen M, Näätänen R. Linguistic relevance of duration within the native language determines the accuracy of speech-sound duration processing. *Cognitive Brain Res*. 2003; 16:492–495.
- Nenonen S, Shestakova A, Huotilainen M, Näätänen R. Speech-sound duration processing in a second language is specific to phonetic categories. *Brain and Language*. 2005; 92:26–32. [PubMed: 15582033]
- Nieuwenhuis S, Aston-Jones G, Cohen JD. Decision making, the P3, and the locus coeruleus norepinephrine system. *Psychol Bull*. 2005; 131:510–532. [PubMed: 16060800]
- Novak GP, Ritter W, Vaughan HG, Wiznitzer ML. Differentiation of negative event-related potentials in an auditory discrimination task. *Electroencephalography and Clinical Neurophysiology*. 1990; 75:255–275. [PubMed: 1691075]
- Peltola MS, Kujala T, Tuomainen J, Ek M, Aaltonen O, Näätänen R. Native and foreign vowel discrimination as indexed by the mismatch negativity (MMN) response. *Neuroscience Letters*. 2003; 352:25–28.

- Perrault N, Picton T. Event-related potentials recorded from the scalp and nasopharynx. I. N1 and P2. *Electroencephalography and Clinical Neurophysiology*. 1984; 59:177–194. [PubMed: 6203709]
- Picton T. The P300 wave of the human event-related potential. *Journal of Clinical Neurophysiology*. 1992; 9
- Shafer LV, Schwartz RG, Kurtzberg D. Language-specific memory traces of consonants in the brain. *Cognitive Brain Res*. 2004; 18:242–254.
- Shafer LV, Ponton C, Datta H, Morr ML, Schwartz RG. Neurophysiological indices of attention to speech in children with specific language impairment. *Clinical Neurophysiology*. 2007
- Shestakova A, Brattico E, Huotilainen M, Galunov V, Soloviev A, Sams M, Ilmoniemi RJ, Näätänen R. Abstract phoneme representations in the left temporal cortex: magnetic mismatch negativity study. *NeuroReport*. 2002; 13:1813–1816. [PubMed: 12395130]
- Specht K, Osnes B, Hugdahl K. Detection of Differential Speech-Specific Processes in the Temporal Lobe Using fMRI and a Dynamic “Sound Morphing” Technique. *Human Brain Mapping*. 2009; 30:3436–3444. [PubMed: 19347876]
- Strange, W. Cross-language studies of speech perception: A historical review. In: Strange, W., editor. *Speech perception and linguistic experience: Issues in cross-language research*. York Press; Baltimore: 1995. p. 3-54.
- Strange W, Dittman S. Effects of discrimination training on the perception of /r-/l/ by Japanese adults learning English. *Perception & Psychophysics*. 1984; 36:131–145. [PubMed: 6514522]
- Strange, W.; Shafer, VL. Speech perception in late second language learners: the re-education of selective perception. In: Zampini, M.; Hansen, J., editors. *Phonology and Second Language Acquisition*. Cambridge University Press; 2008. p. 153-191.
- Sussman E, Kujala T, Halmetoja J, Lyytinen H, Alku P, Näätänen R. Automatic and controlled processing of acoustic and phonetic contrasts. *Hearing Research*. 2004; 190:128–140. [PubMed: 15051135]
- Sussman E. A new view on the MMN and attention debate: Auditory context effects. *Journal of Psychophysiology*. 2007; 21(3-4):164–175.
- Sussman E, Steinschneider M. Attention effects on auditory scene analysis in children. *Neuropsychologia*. 2009; 47:771–785. [PubMed: 19124031]
- Szymanski, MD.; Yund, WE.; Woods, DL. *Journal of the Acoustical Society of America*. Vol. 106. 1999. Phonemes, intensity and attention: Differential effects on the mismatch negativity (MMN); p. 3492-3505.
- Tajima K, Kato H, Rothwell A, Munhall KG. Perception of phonemic length contrasts in Japanese by native and non-native listeners. 15th ICPHS Barcelona. 2003
- Tervaniemi M, Kruck S, Baene WD, Schröger E, Alter K, Friederici AD. Top-down modulation of auditory processing: Effects of sound context, musical expertise and attentional focus. *European Journal of Neuroscience*. 2009; 30:1636–1642. [PubMed: 19821835]
- Toda T. Interlanguage Phonology: Acquisition of Timing Control in JP. *ARAL (Australian Review of Applied Linguistics)*. 1994; 17:51–76.
- Verleger R, Jaśkowski P, Wascher E. Evidence for an integrative role of P3b in linking reaction to perception. *J Psychophysiol*. 2005; 19:165–181.
- Werker JF, Curtin S. PRIMIR: A Developmental Framework of Infant Speech Processing. *Language Learning and Development*. 2005; 1(2):197–234.
- Ylinen S, Shestakova A, Huotilainen M, Alku P, Näätänen R. Mismatch negativity (MMN) elicited by changes in phoneme length: A cross-linguistic study. *Brain Research*. 2006; 1072:175–185. [PubMed: 16426584]

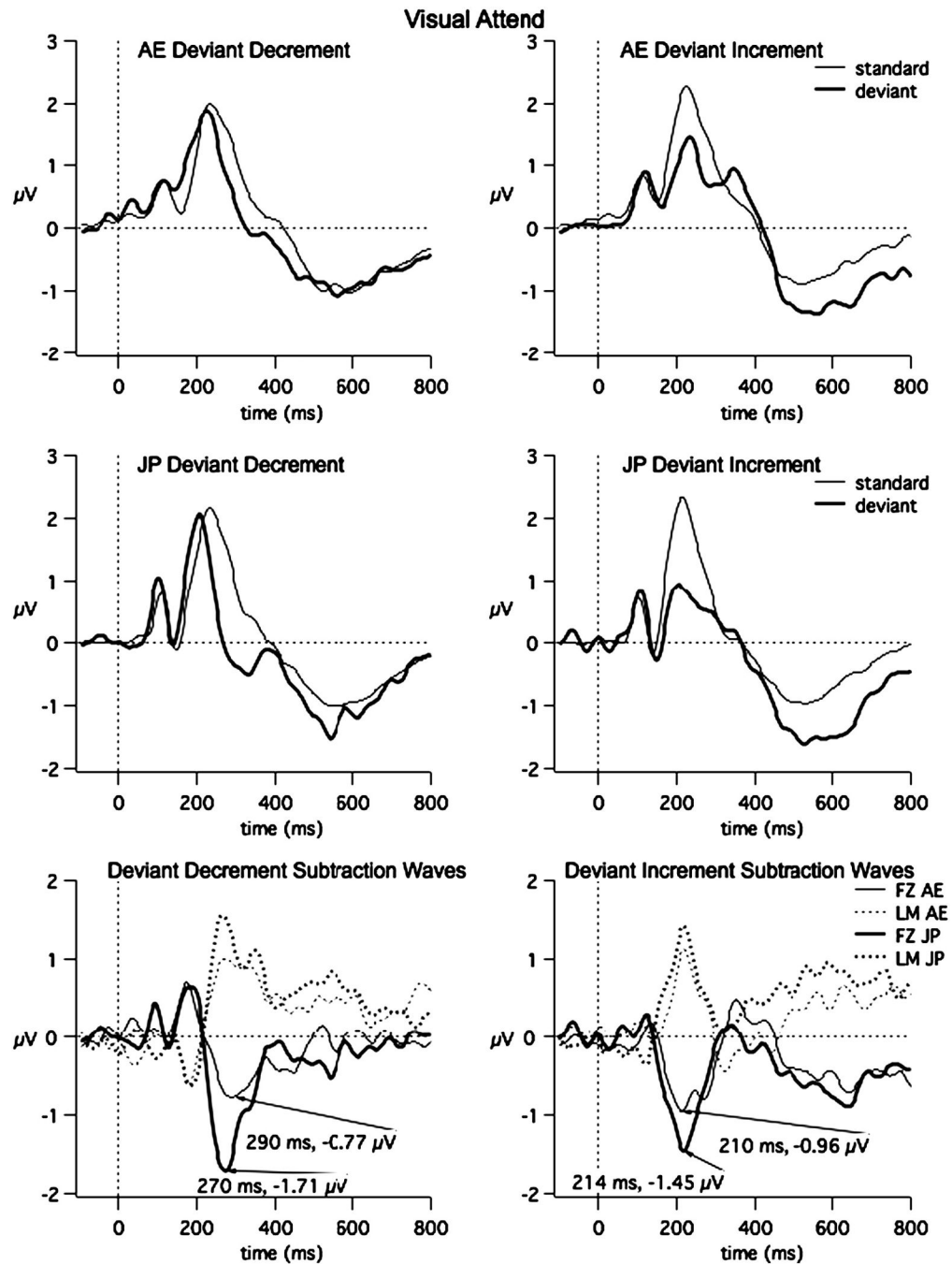


Fig. 1. Visual-Attend Grand Average data (FZ): the grand average at the mid-frontal site (FZ) for the standard, deviant and deviant-standard and at LM for the deviant-standard for all four contrasts and for AE (top a) and JP (bottom b) are shown above. Latency and amplitude are shown for peak negativities of the subtraction waves. It illustrates cross-group responses.

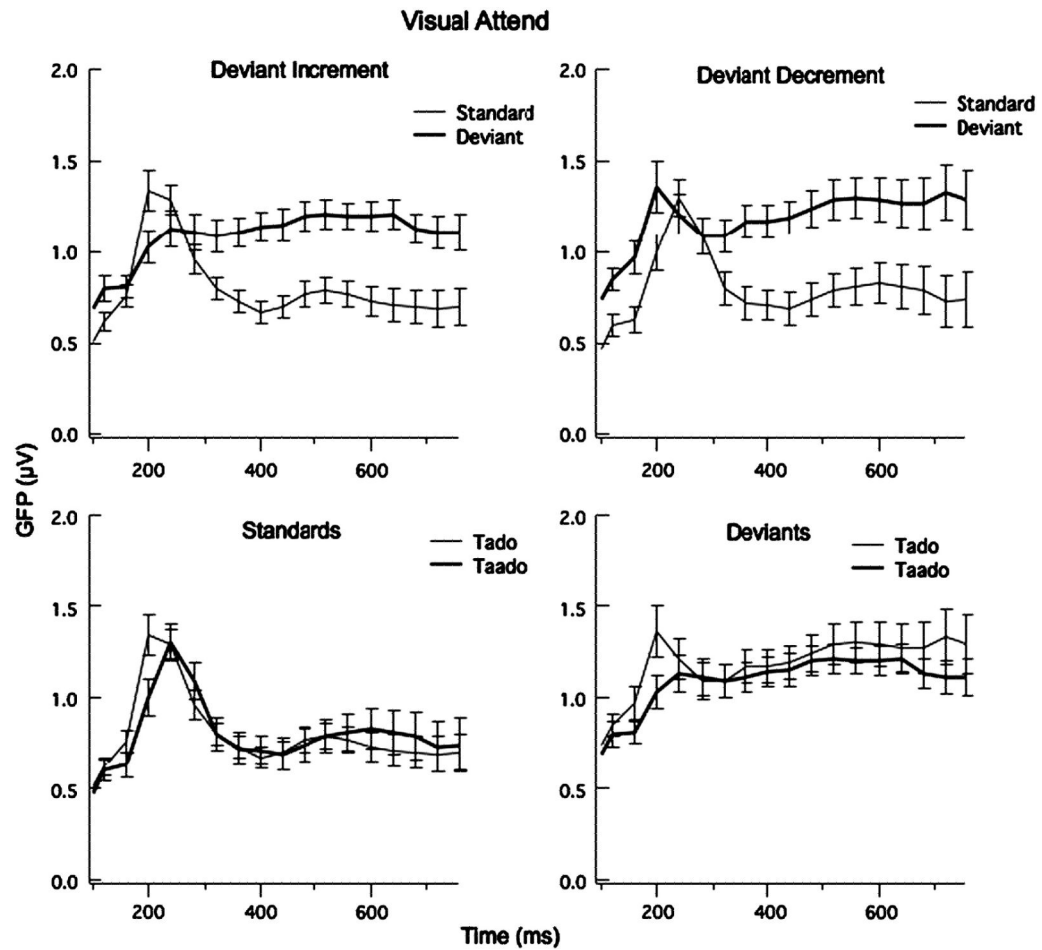


Fig. 2. Mean Global Field Power and Standard Error bars for the 24 participants for the Visual Attend Task comparing the Standard and Deviant stimuli in the Deviant Increment (left graph) and Deviant Decrement (right graph) conditions. The bottom graphs compare the short and long standards and short and long deviants.

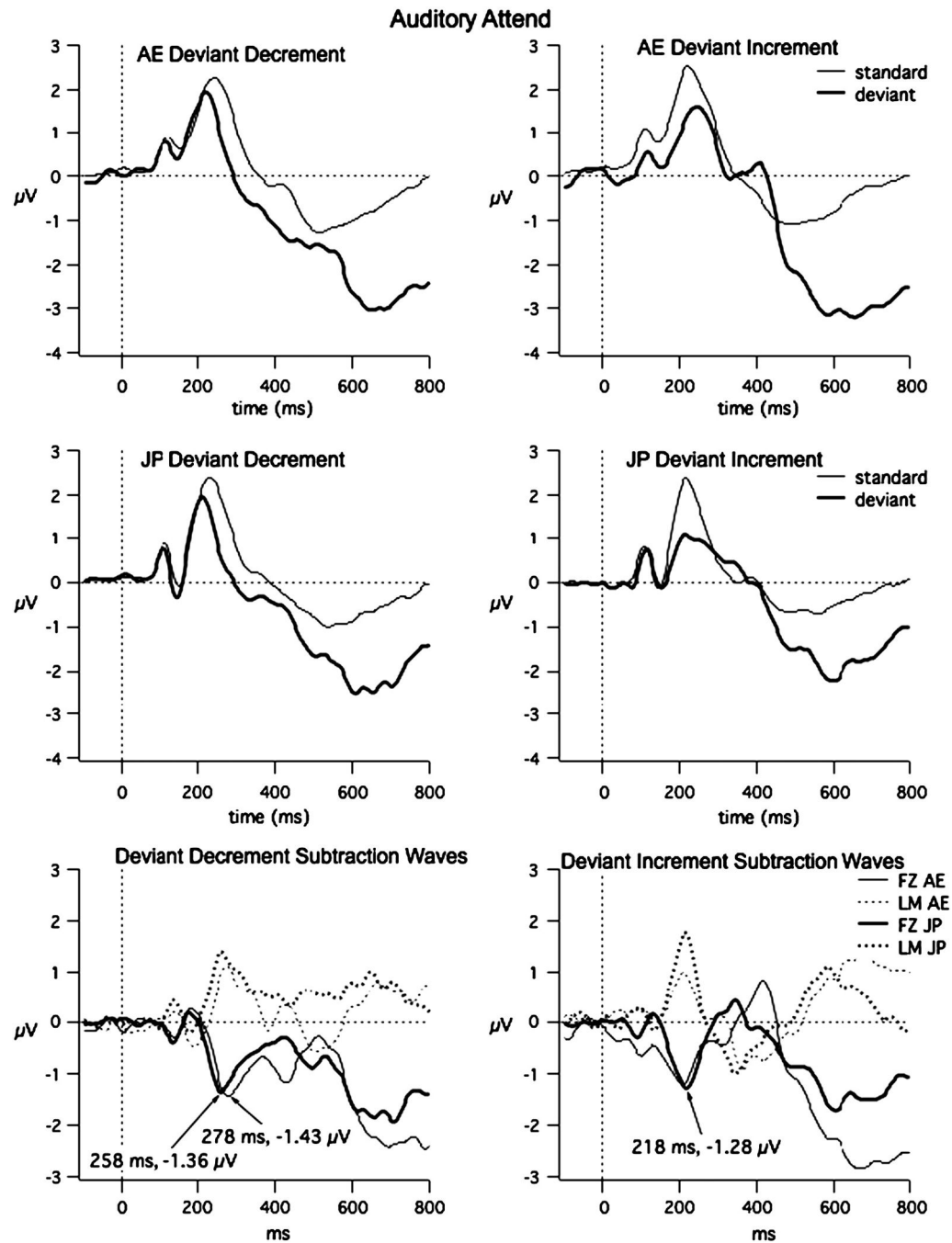


Fig. 3. Auditory-Attend Grand Average data (FZ): the grand average at the mid-frontal site (FZ) for the standard, deviant and deviant-standard and at LM for the deviant-standard for all four contrasts and for AE (top a) and JP (bottom b) are shown above. Latency and amplitude are shown for peak negativities of the subtraction waves. It illustrates cross-group responses.

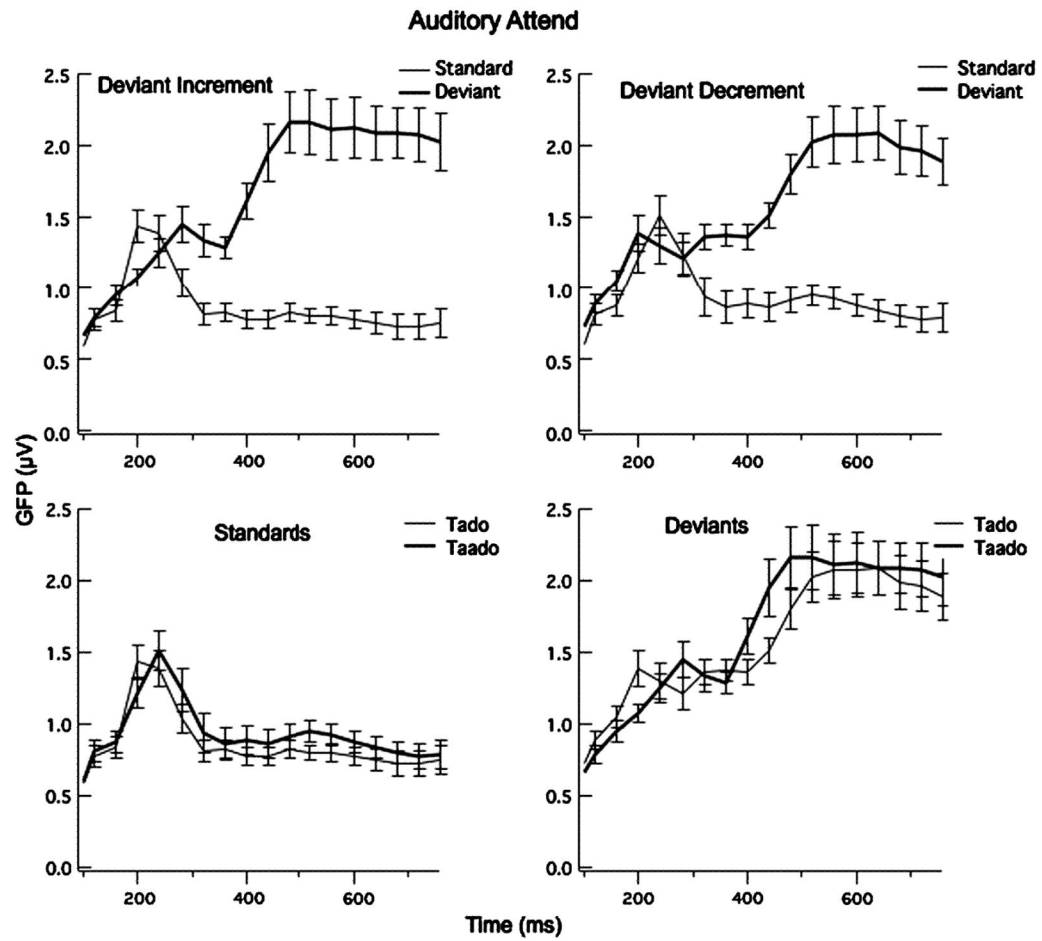


Fig. 4. Mean Global Field Power and Standard Error bars for the 24 participants for the Auditory Attend Task comparing the Standard and Deviant stimuli in the Deviant Increment (left graph) and Deviant Decrement (right graph) conditions. The bottom graphs compare the short and long standards and short and long deviants.

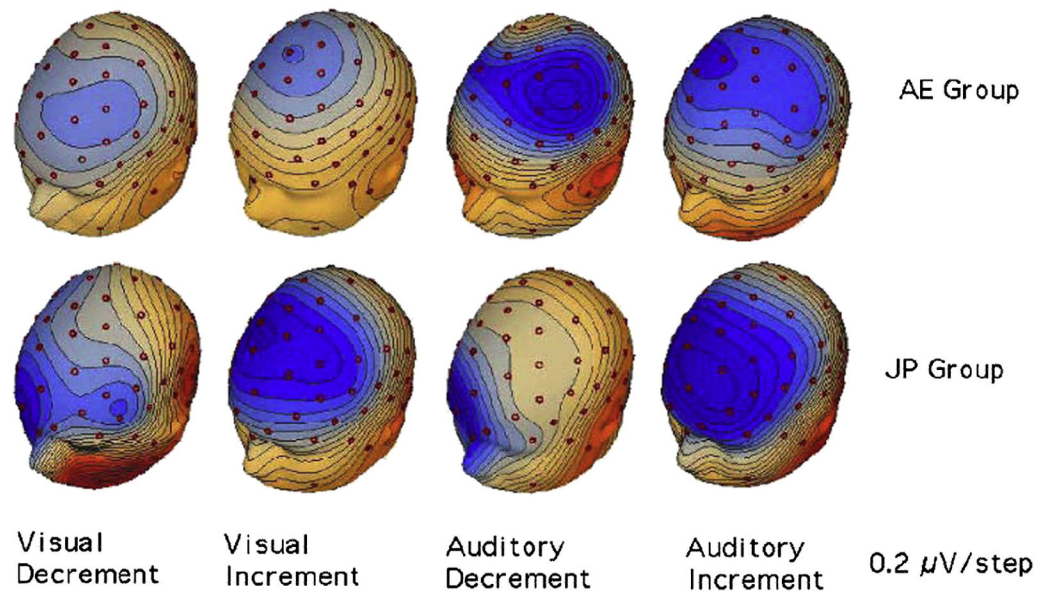


Fig. 5. Voltage Maps at the MMN peak (approximately 270 ms for the deviant decrement and 220 ms for the deviant increment) for the two language groups in the two tasks. The AE group in the Auditory-Attend Task showed a larger negativity than the AE group in the Visual-Attend Task while the JP groups in the two tasks showed little difference. Blue = negative, Red = positive.

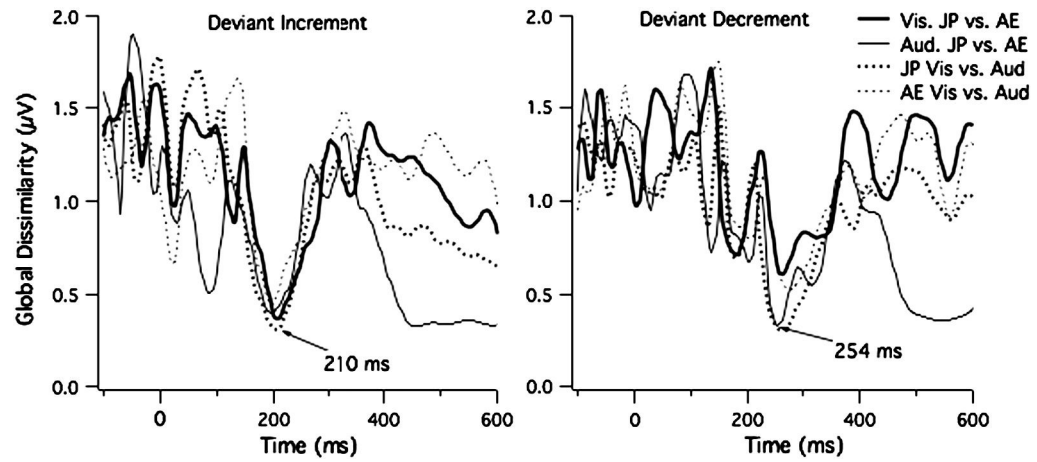


Fig. 6. Global Dissimilarity Index (GDI) for each time point for the Deviant Increment and Deviant Decrement. Higher Values indicate greater dissimilarity in topography. Values < 1 are equivalent to Pearson's $r > 0.5$. The lowest values are found near the MMN peak, indicated similar topographies across groups and tasks for the MMN.

Table 1

Visual-Attend Condition: behavioral data showing the mean number correct, median number correct, and the standard deviation for American (AE) and Japanese (JP) groups.

	JP	AE	JP	AE
	tado	tado	taado	taado
<i>A. Counting visual targets (14=perfect score)</i>				
Mean	13.58	11.92	13.58	12.50
Median	14.00	12.50	14.00	14.00
Std. deviation	0.52	2.35	1.17	2.43
<i>B. Button-press auditory targets (15=perfect score)</i>				
Behavior				
Mean	14.42	11.58	14.33	14.33
Median	14.00	11.50	15.00	15.00
Std. deviation	0.52	2.97	1.07	0.99
<i>C. Button-press auditory targets (reaction time)</i>				
Reaction time (ms)				
Mean	229.33	227.83	298.75	284.20
Median	228.50	198.00	295.00	269.50
Std. deviation	95.80	91.40	77.96	89.25

Table 2

Statistical results of the four-way ANOVAs in the Visual-Attend Task of Experiment 1 for the two orders ([tado]-as standard (Tado) and [taado]-as standard (Taado)). The last column displays the latency range showing significant differences between the standard and deviant intervals.

Group	Effect	F	df	p	MMN range
<i>Tado</i>					
JP	Stimulus	16.10	1, 11	0.002	
JP	Stimulus × Time	14.12	5, 55	0.000	164–288 ms
AE	Stimulus × Time	5.69	5, 55	0.000	164–288 ms
<i>Taado</i>					
JP	Stimulus	15.23	1, 11	0.002	
JP	Stimulus × Time	43.31	5, 55	0.000	164–200; 244–360 ms
AE	Stimulus × Time	10.40	5, 55	0.001	164–320 ms

Table 3

Auditory-Attend Condition: behavioral data showed the mean number correct, median number correct, and the standard deviation.

	JP	AE	JP	AE
	tado	tado	taado	taado
<i>A. Counting Auditory targets (14=perfect score)</i>				
Mean	12.75	11.00	12.67	11.75
Median	14.00	12.50	13.00	12.00
Std. deviation	2.09	3.38	1.44	2.63
<i>B. Button-press Auditory targets (15=perfect score)</i>				
Behavior				
Mean	14.50	11.58	14.50	12.50
Median	15.00	13.00	14.50	14.00
Std. deviation	0.67	2.94	0.52	3.29
<i>C. Button-press Auditory targets (reaction time)</i>				
Reaction time (ms)				
Mean	232.33	232.42	289.17	279.75
Median	242.00	184.00	283.50	269.50
Std. deviation	84.44	100.18	90.25	84.94

Statistical results of the four-way ANOVAs in the Auditory-Attend Task of Experiment 2 for the two orders ([tado]-as standard (Tado) and [taado]-as standard (Taado)). The last column displays the latency range showing significant differences between the standard and deviant intervals.

Table 4

Group	Effect	F	df	p	MMN range
<i>Tado</i>					
JP	Stimulus × Time	21.26	5, 55	0.000	164–240 ms
AE	Stimulus × Time	4.50	5, 55	0.001	164–240 ms
<i>Taado</i>					
JP	Stimulus × Time	16.96	5, 55	0.000	244–320 ms
AE	Stimulus × Time	17.24	5, 55	0.000	244–360 ms

Table 5

Means and SDs (in parentheses) for the frontal and central sites for time 1 and time 2 in the two tasks. (Note: frontal time 1 = FT1; frontal time 2 = FT2; central time 1 = CT1; central time 2 = CT2).

Task		FT1	FT2	CT1	CT2
Auditory	All	-0.7 (-0.36)	-0.66 (-0.59)	-0.5 (-0.4)	-0.7 (-0.56)
Auditory	AE	-0.54 (-0.35)	-0.62 (-0.74)	-0.45 (-0.49)	-0.77 (-0.72)
Auditory	JP	-0.87 (-0.31)	-0.7 (-0.41)	-0.55 (-0.31)	-0.64 (-0.35)
Visual	All	-0.5 (-0.49)	-0.67 (-0.68)	-0.38 (-0.42)	-0.61 (-0.57)
Visual	AE	-0.33 (-0.45)	-0.4 (-0.69)	-0.26 (-0.41)	-0.42 (-0.61)
Visual	JP	-0.68 (-0.48)	-0.94 (-0.57)	-0.51 (-0.41)	-0.79 (-0.48)