

## Research report

## Second-language learning effects on automaticity of speech processing of Japanese phonetic contrasts: An MEG study



Miwako Hisagi<sup>a,b,\*</sup>, Valerie L. Shafer<sup>b</sup>, Shigeru Miyagawa<sup>a</sup>, Hadas Kotek<sup>a</sup>, Ayaka Sugawara<sup>a</sup>, Dimitrios Pantazis<sup>c</sup>

<sup>a</sup> Massachusetts Institute of Technology, Department of Linguistics and Philosophy, 77 Massachusetts Avenue, Cambridge, MA 02139, United States

<sup>b</sup> The City University of New York-Graduate School and University Center, PhD. Program in Speech and Hearing Sciences, 365 Fifth Avenue, New York, NY 10016-4309, United States

<sup>c</sup> Massachusetts Institute of Technology, McGovern Institute for Brain Research, 77 Massachusetts Avenue, Cambridge, MA 02139, United States

## ARTICLE INFO

## Keywords:

Event-related potentials (ERPs)  
Mismatch negativity (MMN)  
Mismatch field (MMF)  
Japanese phonology  
Speech perception, Speech discrimination  
Primary auditory cortex

## ABSTRACT

We examined discrimination of a second-language (L2) vowel duration contrast in English learners of Japanese (JP) with different amounts of experience using the magnetoencephalography mismatch field (MMF) component. Twelve L2 learners were tested before and after a second semester of college-level JP; half attended a regular rate course and half an accelerated course with more hours per week. Results showed no significant change in MMF for either the regular or accelerated learning group from beginning to end of the course. We also compared these groups against nine L2 learners who had completed four semesters of college-level JP. These 4-semester learners did not significantly differ from 2-semester learners, in that only a difference in hemisphere activation (interacting with time) between the two groups approached significance. These findings suggest that targeted training of L2 phonology may be necessary to allow for changes in processing of L2 speech contrasts at an early, automatic level.

### 1. Introduction

Studies of second language (L2) learning have firmly established that late (adult) L2 learners generally show poorer perception and production of L2 phonological contrasts than early (child) L2 learners, even after considerable experience and seemingly high proficiency (Levy and Strange, 2008). Most studies focusing on adult L2 learners have shown that increased L2 experience can result in improved L2 phonology, but rarely to a native-like level (Flege, 2003).

Strange (2011; Strange and Shafer, 2008) proposed the Automatic Selective Perception (ASP) Model, to help explain the challenges in obtaining native-like speech perception of L2 speech sounds that are not closely matched to the first language (L1). In the ASP model, adult L1 learners are characterized as having developed highly automatic selective perception routines (SPRs) for detecting the most reliable acoustic-phonetic cues for differentiating L1 phonemes in variable phonetic and prosodic contexts. Automaticity of these SPRs accounts for the relative ease with which L1 listeners recover phonemic identity of L1 speech sounds.

Strange (2011) hypothesized that non-native listeners need to use focused attention to identify the relevant L2 phonetic cues, even when

the stimuli and task are relatively simple. Without focused attention, non-native listeners will fall back on L1 SPRs, which may not be optimal for identifying L2 phonological categories. Support for the ASP model comes from behavioral studies showing that L1 speech perception is highly robust under conditions of high cognitive load (increased task difficulty and stimulus complexity) and non-optimal listening conditions (e.g., background noise, Tabri et al., 2011; Bradlow and Alexander, 2007); in contrast, non-native and L2 learners show marked decrements in speech perception performance as cognitive load is increased (Strange, 2011).

Neurophysiological measures of non-native speech processing support the claim that native language speech processing is fairly automatic (e.g., Hisagi et al., 2010). In particular, change detection measures, such as mismatch-negativity (MMN) and its magnetic counterpart, mismatch field (MMF) serve as useful indices of processing that occurs when attention is directed away from a speech signal (Näätänen et al., 2007). The MMF/MMN have major sources in auditory cortex, and a large body of literature has demonstrated that these measures can be used to examine speech processing in relation to language experience (e.g., Näätänen, 1990; Alho et al., 1994; Alho et al., 1998; Gomes et al., 1999; Winkler et al., 2003; Sussman, 2007).

\* Correspondence to: Department of Linguistics, University of Connecticut, Oak Hall, Room 339, 365 Fairfield Way, Unit 1145, Storrs, CT 06269-1145, United States.  
E-mail address: [hisagi@mit.edu](mailto:hisagi@mit.edu) (M. Hisagi).

Specifically, MMF/MMN is larger to a speech contrast if that contrast is phonological in listeners' native language (e.g., Dehaene-Lambertz et al., 2000; Menning et al., 2002; Nenonen et al., 2003, 2005; Shafer et al., 2004; Näätänen et al., 2007). In these studies of speech processing, the MMF/MMN has generally been recorded in a passive task, in which attention is directed away from the speech modality via a muted video. Thus, presence of a significant MMF/MMN in this passive paradigm has been interpreted to indicate automatic change detection. However, a few studies have found absence of a significant MMN to native language consonant contrasts (e.g., Shafer et al., 2004; Hisagi et al., 2015). This finding suggests that automaticity of SPRs should be considered a relative measure (more versus less) rather than absolute. Irrespective of which definition is correct, MMF/MMN can serve to evaluate how L2 learning affects processing of an L2 contrast when attention is directed away from the speech contrast.

### 1.1. Training studies

An important question is what conditions facilitate learning of L2 contrasts in late learners of an L2? Traditional classroom learning may be insufficient to induce learning of L2 phonological categories to a near-native level. But some improvement in L2 perception seems to occur, given that correlations are observed between years of experience with an L2 and L2 speech perception (e.g., Werker and Tees, 1984; Flege, 2009; Flege and MacKay, 2011). However, other studies suggest little improvement in L2 speech perception as a result of classroom immersion (Peltola et al., 2003, 2007). To our knowledge, the only study that has asked whether traditional learning in a college classroom leads to changes in L2 speech perception was undertaken by Hisagi et al. (2014). In this study, 12 American-English (AE) learners of Japanese who had had one-semester classroom experience were compared to 12 naïve American-English listeners and 12 native Japanese listeners from Hisagi et al. (2010). This study showed a larger neural discriminative response (MMN) to a Japanese vowel duration contrast compared to the naïve AE group, but of smaller amplitude than found for a native Japanese group. This finding suggested that as little as four hours of classroom experience per week over a 16-week semester could lead to enhancement of L2 speech perception.

Training studies that directly target speech perception have identified some factors that result in better L2 speech perception, such as variability in the phonetic information representing a phoneme category during training (e.g., Pisoni and Lively, 1995; Clopper and Pisoni, 2004; Hirata et al., 2007). The few studies that have examined whether training leads to increased MMF/MMN amplitude have provided mixed results (Menning et al., 2002; Peltola et al., 2005; Ylinen et al., 2010). Menning et al. (2002) found increased MMF amplitude and improved behavioral discrimination in German listeners after 10 training sessions of 1.5 h each, targeting discrimination of pairs of words differing by one mora (e.g., anni vs. ani). The study by Zhang et al. (2009) found that training of Japanese learners on the English /l/ versus /r/ contrast led to more robust MMF over the left hemisphere after 12 training sessions over a 4-week period. Thus, targeted training does seem to have a clear benefit for L2 speech perception learning.

Despite demonstrations of success with targeted training, it appears that most students learning a second language in a classroom context do not receive targeted training as used in experimental research. In the classroom experience, students are presented with multiple levels of language (e.g., lexical knowledge, sentence structure, language use, written forms) and the phonetic characteristics of phonological forms may not receive special attention. It is likely that phonetic training is not included because it requires considerable time, and schools chose to focus the most time on vocabulary and grammar. Nevertheless, experience with the spoken language presented by a teacher who is a native speaker of the L2 may allow for learning of L2 phonetics.

### 1.2. The present study

The goal of the current study was to examine whether intensive focus on the core elements (vocabulary and grammar) in a university L2 classroom setting would incidentally lead to more robust processing of an L2 phonological contrasts, as indexed by the neural measure MMF. Regular classroom teaching of an L2 at the college level often consists of approximately four hours of classroom experience per week. It is possible that the intensity of L2 learning may influence the development of L2 SPRs. The question that will be addressed in this paper is whether doubling the contact classroom hours for learning Japanese (JP) by American English (AE) listeners to eight hours can lead to evidence of more robust L2 speech processing. Increased robustness is inferred by an increase in MMF in a passive paradigm in which attention is directed away from the speech contrast.

The target group was American-English (AE) learners of Japanese (JP) and the target L2 contrast was the JP temporally-cued contrast of vowels tado vs. taado used in Hisagi et al. (2010, 2014). AE and JP phonemic inventories differ greatly in the relevance of spectral compared to temporal cues in differentiating categories. The AE vowel system primarily uses spectrally-cued distinctions (formant frequency values and formant trajectories), while temporal (duration) differences in vocalic nuclei play only a secondary role for some AE vowel contrasts (Bohn and Flege, 1990; Hisagi et al., 2008). In contrast, the Japanese vowel system makes fairly gross spectral distinctions among vowels (only five vowel categories), but also includes a primary cue difference in duration that is phonemic for both consonants and vowels (e.g., [hato] 'pigeon' vs. [haato] 'heart').

A number of behavioral and ERP studies demonstrate the naïve AE listeners show poorer perception/processing of JP vowel and consonant duration (Hirata, 2004a, 2004b; Tajima et al., 2003; Hirata et al., 2007; Tajima et al. 2008; Hirata and Kelly, 2010; Hisagi and Strange, 2011). In particular, naïve AE listeners showed smaller MMN to a vowel duration difference in the Japanese nonsense-word forms "tado" versus "taado" when attention was directed away from the auditory stimulus (Hisagi et al., 2010). The current study uses MMF to examine whether L2 classroom experiences leads to more robust processing of this vowel duration difference.

The main comparison was between L2 learners in a regular, second-semester classroom experience (JP II group) and those in an accelerated-rate learning course that doubles classroom time and covers 30% more material (MIT Very-Fast-Track; VFT group). Thus, the two groups differ in both the amount of material covered and in the number of hours per week learning the material. However, in terms of time elapsed (number of months) for experience with the L2 material, both groups are exposed to the material across a 9-month time span.

A second comparison was between second semester learners (both VFT and JP II) and fourth semester learners (JP IV) to examine whether the additional two semesters of experience (equivalent to 9 months) resulted in differences in processing of L2 speech contrasts.

We hypothesized that greater L2 experience would result in more robust processing of L2 vowel contrasts. Thus, we predicted an increase in MMF amplitude and decrease in MMF latency with increasing experience. We also hypothesized that greater intensity and amount of experience (in the VFT program) would result in more robust L2 speech processing than observed for students matched for number of hours in the regular-rate learning class and equal or greater robustness than for students matched for proficiency from the regular-rate class, if intensity of experience is important. Thus, we predicted greater MMF amplitude and/or decreased latency for the VFT learners compared to regular rate learners with two semesters of JP groups (JP II). In addition, we expected MMF amplitude/latency for the VFT group to be greater than or equal to that of the regular-rate learners with four semesters of JP (JP IV) groups. We also hypothesized that the MMF group differences would be found (or be greater) in the left hemisphere, since several studies suggest that the effect of language experience is

greater at left hemisphere sites (Shafer et al., 2004; Näätänen et al., 2007; Zhang et al., 2009).

## 2. Results

Fig. 1 shows the MMF responses of all the magnetometers of representative subjects from the three groups. All three groups showed a negative deflection peaking between 200 and 300 ms, which is consistent with the expected timing of the MMF. Fig. 2 shows MMF signals for the VFT and JPPII groups before and after JP course administration (Course Time 1 vs. 2). Course effects were generally not very pronounced, but there appeared to be weak MMF increases after the course administration in the right hemisphere for both groups, and decreases in the left hemisphere for JPPII group. Fig. 3 shows comparison of MMF signals of the JPPIV group versus the VFT and JPPII groups (separately for Course Time 1 and Course Time 2). The JPPIV group was measured only once at the end of a fourth semester JP course. JPPIV responses appear greater than VFT and JPPII for both hemispheres and for both course times.

Statistical results from mixed Analyses of Variance (ANOVAs) are reported below. We also provided all non-significant results as Supplemental tables.

### 2.1. VFT vs. JPPII groups

A four-way (Group×Course×Time×Hemisphere×Interval) ANOVA examined the MMF strength between 160 and 400 ms at both the right and left temporal sites (see Fig. 1). The time-period between 160 and 400 ms was separated into six, 40 ms intervals. This analysis revealed a main effect of Interval: [F (5, 50)=14.85,  $p < .001$ ; partial eta-squared =.598] and two-way interactions of Course Time× Interval: [F (5, 50)=2.90,  $p=.044$ ; partial eta-squared =.225] and Hemisphere× Interval: [F (5, 50)=3.24,  $p=.044$ ; partial eta-squared =.245]. We detected no significant Group main effect, or significant interactions with the Group factor (note that the Group×Course Time×Hemisphere×Interval interaction was [F (5, 50)=2.25,  $p=.109$ ]). Tukey's post-hoc tests followed up the Course Time×Interval interaction, comparing course time for each interval; these post-hoc tests showed that the MMF from Course Time 1 was significantly larger than the MMF from Course Time 2 for the interval 360–400 ms ( $p=.026$ ). Examination of Fig. 2 illustrates this pattern for both the VFT and JPPII groups; the post-hoc tests following up the Hemisphere × Interval interaction compared the two hemispheres for each interval and revealed that the left hemisphere MMF was larger than the right MMF for the interval 240–400 ms ( $p < .02$ ). Supplemental Fig. 1 shows the topography, illustrating the difference between Course Time 1 and Course Time 2 in three different time points, 250 ms, 350 ms, and 450 ms. There was no significant activity after 400 ms.

It was important to understand whether an absence of a difference between the JPPII and VFT groups was due to the participants reaching some ceiling of performance. Thus, we combined the VFT and JPPII groups into one group who had been exposed to JP across two semesters. This new group was then compared the JPPIV group, who had four semesters of classroom experience. We predicted that four semesters of JP would allow for more robust processing of the JP vowel contrast and this would be seen as a larger MMF. We compared MMF responses separately for the VFT/JPPII groups versus JPPIV for Course Time 1 in one analysis and for the Course Time 2 in a second analysis (since the JPPIV group was only measured once at the end of the 4th JP semester). Please see Supplemental table 1a for all the results.

### 2.2. JPPIV vs. VFT/ JPPII (for course time 1)

The 3-way (Group×Hemisphere×Interval) ANOVA analyses revealed a main effect of Interval [F (5, 95)=27.51,  $p < .001$ ; Partial eta-squared=.591], but there were no significant interactions. Fig. 3

shows that the VFT and JPPII groups had greater MMF in the left than the right hemisphere, while the JPPIV group had nearly the same MMF in both hemispheres. Please see Supplemental table 1b for all the results.

### 2.3. JPPIV vs. VFT/JPPII (for course time 2)

A main effect of Interval [F (5, 95)=28.004,  $p < .001$ ; Partial eta-squared=.596] was found, but there were no significant interactions. Specifically, the MMF was largest in the Interval 200–280 ms. Please see Supplemental table 1c for all the results.

## 3. Discussion

The current study examined the nature of changes in L2 speech perception relative to amount and intensity of experience. We had predicted that experience with Japanese would lead to more robust processing of phonological contrasts at an automatic level. Our findings showed a significant effect of learning via an interaction of hemisphere in comparing the second and fourth semester groups, but no increase in robustness of neural discrimination following the second semester of experience for either the regular program or the VFT learners. Indeed, the MMF showed a decrease in MMF from the first to the second semester for the 360–400 ms interval. These findings suggest that the increased experience from the beginning to the end of a second semester of Japanese did not have a discernable effect on the speech discrimination process indexed by the MMF. Furthermore, the additional classroom experience (3rd and 4th semester) also showed little effect, with only an interaction with hemisphere and interval revealing a difference from the two-semester learners. This lack of a clear difference in MMF was unexpected for the JPPIV versus JPPII groups, since the JPPIV group clearly had more experience with Japanese (four semesters) compared to the JPPII group (two semesters). We had predicted that the VFT group after the second semester might show greater MMF than the JPPII group and that they might also show similar MMF amplitude to the JPPIV group because of the increased intensity of JP experience. This pattern, however, was not found, with both groups showing similar MMF. An additional prediction was that the MMF would be larger over left than right sites, because of the linguistic nature of the stimulus contrast. Our findings were consistent with this hypothesis for the JPPII and VFT groups; however, for the JPPIV the tendency was for the right hemisphere to show greater activation than

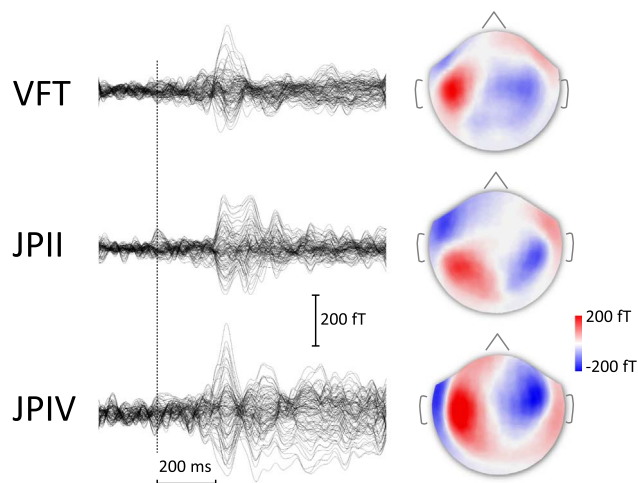
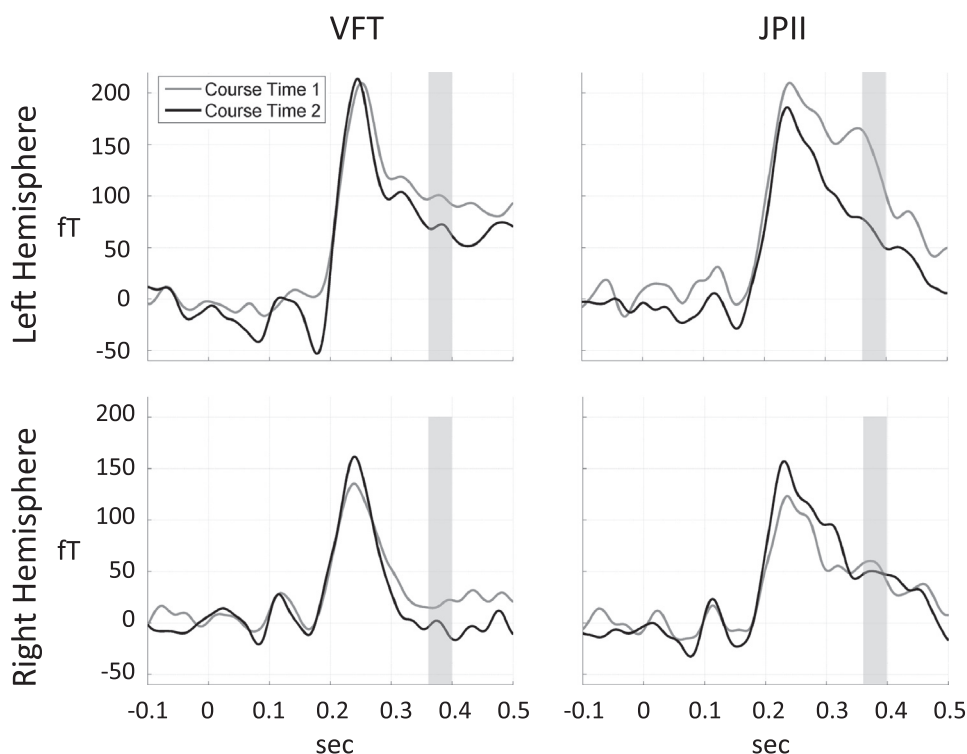
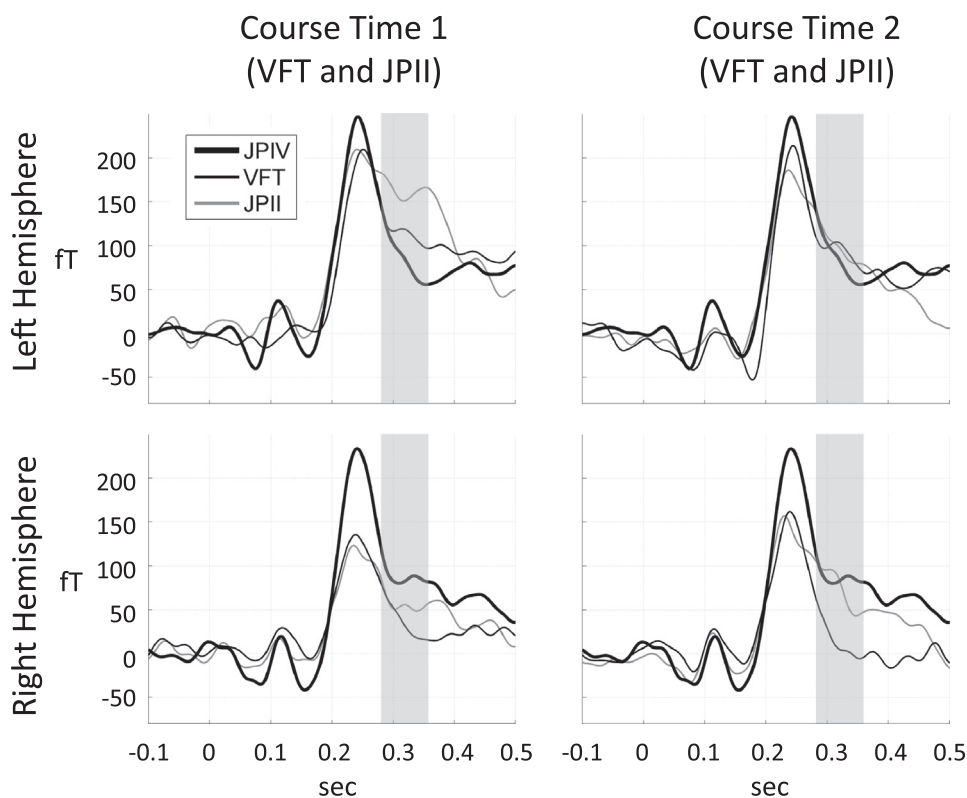


Fig. 1. Example MMF signals from representative subjects from the VFT, JPPII, and JPPIV groups. Left: MMF signals from all magnetometers overlaid on the same axis for each subject. MMF signals have consistent temporal signatures for all individuals, with a maximum magnetic deflection at 240 ms. Right: Magnetic field topography at 240 ms. Though variable across subjects, all topographies show signals originating bilaterally from the temporal cortex.



**Fig. 2.** MMF signals for the VFT and JPII groups before (Course Time 1) and after (Course Time 2) JP course administration. The shaded region indicates a significant Course Time×Interval interaction [ $F(5, 50)=2.90, p=.022$ ], with post-hoc tests revealing the MMF from Course Time 1 was significantly larger than the MMF from Course Time 2 for the interval 360–400 ms ( $p=.026$ ).



**Fig. 3.** Comparison of MMF signals of JPIV group versus VFT and JPII groups (separately for Course Time 1 and Course Time 2). JPIV group was measured only once at the end of a fourth semester JP course. The shaded region indicates a significant main effect of Hemisphere [ $F(1, 19)=5.38, p=.032$  for the 280–320 ms interval; and  $F(1, 19)=5.21, p=.034$  for the 320–360 ms interval] with responses greater over the left than the right hemisphere. While JPIV responses were overall greater than VFT and JPII for both hemispheres and course times, pairwise comparisons did not show any significant differences across groups.



the left hemisphere. We followed up the interaction that approached significance, but no pairwise comparisons reached significance. Thus, these findings taken together indicate that we cannot make any general claims about L2 learning of phonological patterns primarily affecting the left hemisphere auditory regions.

In general, this pattern of findings suggests more time is necessary to see changes in L2 phonological discrimination as measured at the level of the MMF. Below we discuss these findings in relation to the prior literature.

### 3.1. Changes in robustness of L2 speech perception

Our main goal was to examine whether increased experience with an L2 could lead to increased robustness in phonological discrimination at a largely attention-independent level of processing, even when the L2 was learned past puberty (e.g., [Strange and Shafer, 2008](#)). We also predicted that the intensity of learning in the VFT program might lead to a larger increase in MMF activation for the VFT group compared to the standard classroom experience (i.e., JPII). We based this prediction on the argument that adult learners of an L2 might need increased intensity of experience to overcome L1 SPRs. Research on other species suggest that intense training may be necessary to learn new sensory routines in the case that there was a sensitive period or critical period for learning this information ([Knudsen, 2004](#)).

We did not find a larger increase in MMF for the VFT group. In fact, we observed a small decrease in MMF for both groups after the second-semester training. It is possible that the decrease in MMF amplitude was due to the latency of MMF shifting earlier for some, but not all participants. This would lead to increased variability in MMF peak latency and an apparent decrease in MMN peak amplitude in a fixed time interval. Identifying a peak latency of MMF could help verify whether this suggestion is possible. We chose to test intervals rather than peak latencies because individual often do not show a clear peak, and thus examining amplitude across successive time intervals is a more objective method for determining when in time MMF was largest.

An alternative explanation for the decrease in MMF is that some other process overlaps with those indexed by the MMF in such a way to be seen as a decrease in MMF activation. Even though we attempted to better control for attention than other studies that have used a passive oddball design, it is possible that some attentional resources were allocated to processing the speech sounds and that greater experience with Japanese led to participants allocating more (or fewer) resources to processing these stimuli. It will be necessary to compare processing of these speech stimuli in an experiment in which attention is manipulated to and away from the speech sounds, as we have done using EEG ([Hisagi et al., 2010](#)) to understand whether this is a plausible explanation.

We did not find a clear difference in MMF between the fourth semester (JPIV) and the second-semester learners, although hemispheric contribution was different. It is possible that an absence of a clear difference between the second and fourth semester groups is because this comparison was between groups, and there might be individual differences in the participants that overshadowed learning effects. The within-group comparison of the before- and after-second-semester experience, however, controls for individual variability. These findings, in conjunction with [Hisagi et al. \(2014\)](#) suggest that the L2 phonological learning trajectory may not be linear in shape. Listeners may initially make gains (during the first semester), but then show a plateau, or even a dip in learning (second semester), followed by an additional gain at some later time. It will be important for future studies to sample the same group of participants across a larger time span (e.g., first semester through 8th semester) to map out the learning trajectory for this type of phonological information.

The current study examined students who were enrolled in JP courses that focused on teaching multiple levels of the L2 (phonology, semantics, syntax, pragmatics, writing). Thus, classroom time was not

focused exclusively on teaching JP phonology. It is also the case that classes often do not focus on training the phonetic realization of a L2 phonology. It is possible that the lack of increase in MMF can be attributed to the absence of targeted phonetic/phonological training. Several studies have shown that targeted training can increase the MMN/MMF in naïve or L2 listeners (e.g., [Menning et al., 2002](#); [Zhang et al., 2009](#)). For example, [Zhang et al. \(2009\)](#) found that a targeted phonetic training course with twelve sessions over four weeks resulted in increased MMF in Japanese L2 learners of the English /l/ versus /r/ contrast. This training was carried out with Japanese listeners who had nine years of classroom English in Japan, with little spoken English experience. The students in the VFT and JPII programs may need such targeted training to make equivalent rapid gains in L2 speech perception. Alternatively, it may be that the Japanese listeners in the [Zhang et al. \(2009\)](#) study had so little experience with native-like spoken English that the four weeks experience in the training course led to a significant change at the level reflected by the MMF. Gains in robustness of discrimination might slow down after this first spurt of learning. Our previous EEG study with same paradigm ([Hisagi et al., 2014](#)) suggested a strong effect of one-semester learning in the classroom. In the Zhang study, even though the Japanese learners of English made great gains, their perception of the /l/ versus /r/ contrast was reported to still differ from that of native English listeners ([Zhang et al., 2009](#)). A future study will need to directly contrast classroom learning and targeted L2 phonological training across time to identify the factors that allow for increased robustness in L2 discrimination.

An alternative explanation for the absence of an increase in the MMF is that familiarity with the stimuli acquired during the first session in some way attenuated any gains in discriminating this contrast in the second session. In this case, increased automaticity could be masked by some type of familiarity effect. However, we do not know of any other studies that support this explanation. Even if this were the correct explanation, the VFT group did not show greater MMF activation compared to the JPII group (assuming familiarity would be similar across groups). Thus, the conclusion that the greater intensity of experience did not lead to more robust processing holds, at least for the one-semester time span tested here. In addition, the fourth semester group was tested just once, but their MMF activation was not significantly different from the second-semester group, other than the hemispheric difference.

Future studies will need to use different stimulus sets for the first and second testing to determine whether familiarity with the stimuli influenced the findings and to examine long-term benefits. In addition, it will be important to examine whether the VFT group shows an effect of experience after more time has elapsed. Specifically, it may take more time to consolidate the learning. Finally, it will be important to examine whether the intensity of more targeted experience in conjunction with more traditional methods of learning an L2 can induce an over-learning effect for L2 phonemes. The findings of [Zhang et al. \(2009\)](#) suggest that this is a real possibility.

### 3.2. Psychoacoustic salience effect

A second possible explanation for no difference between the JPII and VFT groups is that the vowel duration contrast was sufficiently psycho-acoustically salient to allow a robust MMF at the first testing session. [Burnham \(1986\)](#) suggested that more psycho-acoustically salient contrasts, such as a temporal voice cue, are generally learned earlier in L1 development, are easier L2 contrasts to perceive, and are easier to improve and maintain after L2 training. In contrast, less psycho-acoustically salient contrasts, such as spectral place cues, are learned later in development and are harder to learn in an L2, even with targeted training. The /l/ versus /r/ distinction examined in the [Zhang et al. \(2009\)](#) study is considered an extremely difficult contrast for Japanese L2 learners of English; it is possible this greater difficulty allowed for greater gains in learning compared to the vowel duration

difference of the current study. Thus, a smaller gain, or no gain in perception from before to after the second-semester course might have occurred for both JPII and VFT groups.

Strange (2011) suggested that non-native listeners might need focused attention for identifying relevant L2 phonetic cues even in the case of a fairly easy contrast, such as the vowel duration contrast used here. The finding that non-native listeners (Hisagi et al., 2010), first semester learners (Hisagi et al., 2014) and the participants in this study all show evidence of neural discrimination suggests that this vowel contrast is fairly easy. However, the amplitude was smaller than found for native listeners (Hisagi et al., 2010), even though MMN was observed in non-native listeners to this vowel duration contrast in our previous study. In addition, behavioral discrimination was worse for non-native than native listeners (Hisagi et al., 2010). Thus, merely presence versus absence of an MMF/MMN to a phonetic contrast cannot be used to predict behavioral discrimination. Rather, the amplitude of the MMF/MMN for an experimental versus a control group must be compared. The data from Hisagi et al. (2010) supported Strange's model because MMN amplitude was small for the non-native group relative to the native control group.

In the current study the fourth semester students (JPIV) serve as a control group, and pattern of findings indicate very little change from the second to the fourth semester. As we suggested above, testing at multiple time points during the learning process will be necessary to understand the shape of the learning function. In addition, L2 speech perception often suffers in noisy conditions (Shafer and Strange, 2008). Thus, using less intense vowels (i.e., lowering the stimulus intensity) or including background noise may serve as a better test of whether this type of phonological contrast can become native-like.

### 3.3. Hemisphere effects

Our study showed a stronger effect of MMF over left hemisphere than the right hemisphere, as we predicted. However, there was no training effect (i.e., no improvement) at the left site as observed in the studies by Peltola et al. (2003, 2007). In fact, the only hemisphere effect that might be attributed to L2 experience was more bilateral MMF activation for the fourth compared to the second semester students at the onset of the second semester. The pairwise comparison between each hemisphere between the Course Time 2, semester 2 students and the JPIV group did not reveal a significant difference, although the ANOVA revealed a significant group by hemisphere by interval interaction that approached significance. The follow-up comparisons did show that only the second semester students at this first testing session showed greater left than right activation. Testing after completing the second semester led to no difference in MMF amplitude from the fourth semester group. So it seems that experience during the second semester led to a more bilateral effect for all students. However, this increased bilateral effect may be attributable to a reduced left hemisphere response, rather than an increase in the right hemisphere. This finding will need to be replicated to have confidence in it, particularly since we were comparing two different groups, each with small numbers.

### 3.4. Limitations

It is important to point out some limitations in this study, namely, a relatively small sample size (N=6 for JPII, N=6 for VFT and N=9 for JPIV) and the use of one stimulus contrast. Enrollment in VFT classes was limited to up to ten students and for the JPII class up to 15 per year. Participation in the current study was voluntary and not all students were interested in participating. In two years, we succeeded in recruiting six JPII and six VFT students who were tested before and after the second semester course, but power was low for comparing JPII to VFT. However, the comparison of all twelve participants before and after the second semester experiences allows for greater con-

fidence. We also included the JPIV group as a control group to determine whether the MMF activation was greater for an additional two semesters of experience. It would be useful to also include an L1 JP group and a naïve American-English group in a future study, as additional control groups, since they represent the extreme cases (see Hisagi et al. (2010)).

We chose to examine only one contrast type because we wanted to limit the whole experiment time to 1.5 h so as not to discourage the students from remaining in the study for the second testing session. Future studies will be needed to examine other contrasts to determine how the inherent difficulty of the contrast type affects learning. For example, in another study we have examined consonant duration cues (miʃi versus miʃʃi) and found that these are more difficult to discriminate for L2 learners and also show less robust MMN for native JP listeners (Hisagi et al., 2015). It is possible that these contrast types will allow for greater gains in learning. One final limitation is that we did not have behavioral discrimination tasks using these vowel stimuli, in part because we expected fairly good perception with attention, and because of the time limitations mentioned above. Despite these limitations, our findings suggest that a second semester of classroom experience, without some targeted training, is insufficient to induce robust L2 phoneme categories for the vowel duration contrasts of interest.

Lastly, we conducted our study with MEG rather than EEG for two reasons: a) The MEG system was offered for free at MIT at the time of the study because the McGovern Institute at MIT provided free access during the first 2 years of the lab operation, b) we made considerable effort to recruit students (our sample size was extremely limited, restricted to a few students attending Japanese L2 courses at MIT) and MEG was much more convenient and appreciated by our subjects because we did not have the prolonged EEG setup time required to apply EEG sensors, gel, and other inconveniences. However, in a future study it would be interesting to examine how the MEG findings compare to EEG, since our original studies used EEG (Hisagi et al., 2010, 2014).

## 4. Conclusion

This study revealed the changes in robustness (automaticity) of JP L2 speech processing appear to need more than two semesters of classroom experience and that the accelerated learning program does not improve L2 speech perception for these vowel contrasts, even though vocabulary and grammatical learning was accelerated. It will be important to further test this claim with a larger group of participants, considering that the current result was based on a small group of participants. If this finding is replicated, specifically that little to no gain in L2 speech processing (as measured by MMF) was observed over the brief period for either the standard or accelerated learning program, it indicates that learning of L2 speech patterns is not facilitated by the current classroom curriculum. It will be important to examine whether adding targeted training of L2 speech patterns to teaching programs will allow for improved L2 speech perception over this short time span (one or two semesters).

### 4.1. Methods

The MIT regular JP program consists of 4 h per week of classroom time, over a semester (13 week) period. All classes are taught by native speakers of Japanese. The VFT Program replaces the second semester of a regular JP program and any student with one semester of Japanese can register. The course is accelerated to cover material in semesters 2, 3 and 4, with students meeting 8 h per week. In all programs, students can perform drills from internet access. English is kept to a minimum. In the VFT program, successful students achieve Intermediate-Mid to Intermediate-High language proficiency scores on the American Council on the Teaching of Foreign Languages (ACTFL) proficiency

scale. This is the equivalent proficiency-level found for most two-year (four semesters) programs of Japanese. Oral Proficiency Interview (OPI) tests of students following one semester of VFT indicates an Intermediate-Mid level of proficiency, which is comparable to students in two years of the regular track. Students in the regular track generally do not obtain this level of proficiency until 3 or 4 semesters of JP.

Students in the VFT track initially take the first semester of Japanese in the regular track; the VFT begins in the second semester. The selection of the regular vs. VFT track is typically based on individual time constraints. Most importantly, the VFT students are not selected based on superior performance in the first semester class.

#### 4.2. Subjects

Native speakers of AE who have been studying JP at the college level in regular-rate courses (a second semester (JPII) group (1M5F, mean 19y) and a fourth semester (JPIV) group) (6M3F, mean 22y), and an accelerated-rate course (VFT group) (3M3F, mean 19y) were recruited from MIT. Six students in the VFT and six in the JPII group (matched for scores and aptitude) were tested at two time points, beginning (Course Time 1: Before) and end (Course Time 2: After) semester 2 classes. Nine regular class students (JPIV group) were tested at the end of semester 4 (i.e., potential equivalent proficiency level with the VFT group at Time 2). All subjects were between the age of 18 and 29 years, with normal hearing (20 dB HL at 500–4000 Hz) and all but one (who received a B grade) had received a grade of A in the prior language class. All L2 subjects were from the foreign language departments at MIT and were compensated for participation.

#### 4.3. Experimental design

Overall materials and methods are the same as in Hisagi et al. (2010), for greater detail see also Hisagi (2007). The long vowel nonsense word form /taado/ served as the standard (1190 standards: 85%) and the short vowel nonsense word /tado/ as the deviant stimulus (210 deviants: 15%), which was the hardest contrast in the previous study (Hisagi et al., 2010). There were 14 blocks and each block had a different number of standards and deviants (average of 100 trials per block). Each block lasted about 1–2 min. We created four natural tokens of each category from a female Japanese speaker of Tokyo dialect. The average duration ratio of the long to short target vowels (/aa/ in /taado/ vs. /a/ in /tado/) was 1.61, 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). Intensity was normalized.

A visual attention task served to draw attention away from the auditory stimuli. Visual stimuli consisted of four different sizes of pentagon and hexagon shapes (orange with black background), and participants were instructed to perform an oddball visual discrimination task. The visual stimuli were presented using Psychtoolbox ([www.psychtoolbox.org](http://www.psychtoolbox.org); Brainard, 1997).

An inter-stimulus interval (ISI) for the auditory stimuli was 800 ms. The ISI for the shapes as described below was 780 ms and differed from that of the auditory words in order to make sure that the presentation of auditory and visual stimuli would not inadvertently be in synchrony. Note that the shapes were delivered independently on a laptop computer and served in a distractor task to pull attention away from the auditory modality.

#### 4.4. Magnetoencephalography (MEG) recordings

MEG data was recorded using an Elekta Triux system (306-channel probe unit with 204 planar gradiometer sensors and 102 magnet-

ometer sensors) at a sampling rate of 1,000 Hz, filtered between .03 and 330 Hz. The location of the head was measured at the beginning of each recording session by activating a set of 5 head position indicator coils placed over the head. Prior to the MEG recording, a 3D digitizer (Fastrak, Polhemus, Colchester, Vermont, USA) was used to register the locations of 3 anatomical landmarks (right and left preauricular points and the nasion) with respect to the 5 head position indicator coils.

#### 4.5. Extraction of MMF signals

Raw data was pre-processed with the Maxfilter software (Elekta, Stockholm) to compensate for head movements and perform noise reduction with spatiotemporal filters (Taulu et al., 2004; Taulu and Simola, 2006). We used default parameters (harmonic expansion origin in head frame=[0 0 40] mm; expansion limit for internal multipole base=8; expansion limit for external multipole base=3; bad channels automatically excluded from harmonic expansions=7 sd. above average; temporal correlation limit=.98; buffer length=10 s). Intuitively, the software first applied a spatial filter that separated the signal data from spatial patterns emanating from distant noise sources outside the sensor helmet. It then applied a temporal filter that discarded components of the signal data with time series strongly correlated with the ones from the noise data. The resulting filtered data were subsequently analyzed with Brainstorm (Tadel et al., 2011). Blink artifacts were removed by performing principal component analysis on data contaminated with blinks (estimated with EOG recordings) and projecting away from the first principal component. We then extracted peri-stimulus data from –200 ms to +800 ms with respect to each auditory stimulus.

Every trial was baseline-corrected to remove the mean (–200 to 0 ms) from each channel. Evoked response fields (ERFs) were computed by averaging the trials separately for the standard /taado/ and deviant /tado/ stimuli. Mismatch field (MMF) responses were then computed for each subject by subtracting the standard ERFs from the deviant ERFs (examples for individual subjects shown in Fig. 1).

Since we did not have subject-specific structural MR scans to accurately localize signals on the auditory cortex, we proceeded by analyzing the channel recordings with maximal MMF signals. Given the variable MMF topographies, we selected separately for each subject magnetometer sensors with the strongest MMF deflections between 220 and 260 ms post-stimulus. Selection was constrained to sensors in proximity to temporal cortex and included sensors measuring both incoming and outgoing magnetic fields with respect to the head. MMF measurement signals were averaged across the selected sensors separately for the right and left hemispheres, after accounting for the differences in field direction by flipping the sign of the sensors measuring incoming magnetic fields. Thus, for each subject and each hemisphere we estimated an overall MMF effect by averaging the contribution of sensors over the temporal cortex with the stronger individual MMF deflections. Figs. 2 and 3 show the subject-averaged MMF effect for different groups and course times.

#### 4.6. ANOVA analysis

The extracted averaged MMF waveforms were subjected to a repeated measures analysis of variance (ANOVA). We first compared the VFT and JPII groups with a 4-way ANOVA that had a between-subjects factor Group (VFT, JPII) and within-subjects factors Course Time (Time 1: before course; Time 2: after course), Hemisphere (left, right) and Interval (six 40-ms intervals: 160–200; 200–240; 240–280; 280–320; 320–360; 360–400 ms).

In follow-up analyses, we combined VFT and JPII into a common group and compared it against the JPIV group. The comparison was performed separately for Course Time 1 and then for Course Time 2 signals for the VFT and JPII groups. The 3-way ANOVA analysis used a



between-subjects factor Group (with 2 levels: VFT & JPII vs. JP/IV), and within-subjects factors Hemisphere, and Interval.

We included Hemisphere as a factor because we predicted that changes might be larger over left than right sites, given the linguistic nature of the stimuli. We included Interval as a factor because increased experience might lead to earlier MMF, in addition to increased MMF power. Interval values were within the range 160–400 ms, wide enough to capture the dynamics of the MMF signals. A significance level of .05 was used for all tests.

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. Significant effect size (Partial eta-squared) was also provided where it was a significant p-value.

## Acknowledgements

This research was supported by the Toshiba Foundation (to Shigeru Miyagawa) and by the National Science Foundation - Division of Behavioral and Cognitive Sciences (NSF-BCS) (Grant number 01-113 BCS) Dissertation Enhancement Award #0518989 (to Miwako Hisagi, Valerie Shafer and Winifred Strange).

## Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.brainres.2016.10.004.

## References

- Alho, K., Woods, D.L., Algazi, A., 1994. Processing of auditory stimuli during auditory and visual attention as revealed by event-related potentials. *Psychophysiology* 31 (5), 469–479.
- Alho, K., Winkler, I., Escera, C., Huottilainen, M., Virtanen, J., Jaaskelainen, I.P., Pekkonen, E., Ilmoniemi, R.J., 1998. Processing of novel sounds and frequency changes in the human auditory cortex: magnetoencephalographic recordings. *Psychophysiology* 35, 211–224.
- Bohn, O.S., Flege, J., 1990. Interlingual identification and the role of foreign language experience in l2 vowel perception. *Appl. Psycholinguist.* 11, 303–328.
- Bradlow, A.R., Alexander, J.A., 2007. Semantic-contextual and acoustic-phonetic cues for english sentence-in-noise recognition by native and non-native listeners. *J. Acoust. Soc. Am.* 121 (4), 2339–2349.
- Brainard, D.H., 1997. The psychophysics toolbox. *Spat. Vis.* 10, 433–436.
- Clopper, C.G., Pisoni, D.B., 2004. Effects of talker variability on perceptual learning of dialects. *Lang. Speech* 47 (3), 207–238. <http://dx.doi.org/10.1177/00238309040470030101>.
- Dehaene-Lambertz, G., Dupoux, E., Gout, A., 2000. Electrophysiological correlates of phonological processing: a cross-linguistic study. *J. Cogn. Neurosci.* 12, 635–647.
- Flege, J., MacKay, I., 2011. What accounts for “age” effects on overall degree of foreign accent? In: Wrembel, M., Kul, M., K. Dziubalska-Kolaczyk, (Eds.), *Achievements and Perspectives in the Acquisition of Second Language Speech: New Sounds 2010*. Peter Lang, Bern, Switzerland, vol. 2, pp. 65–82.
- Flege, J.E., 2003. Assessing constraints on second-language segmental production and perception. In: Meyer, A., Schiller, N. (Eds.), *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities*. Mouton de Gruyter, Berlin, 319–345.
- Gomes, H., Sussman, E., Ritter, W., Kurtzberg, D., Cowan, N., Vaughan, H.G., Jr., 1999. Electrophysiological evidence of developmental changes in the duration of auditory sensory memory. *Dev. Psychol.* 35, 294–302.
- Hisagi, M., Nishi, K., Strange, W., 2008. Acoustic properties of Japanese and English vowels: Effects of Speaking Style and Consonantal Context. In: M. Endo-Hudson, Jun, S.-A., P. Sells, P.M. Clancy, S. Iwasaki, and S.O., Sohn. (Eds.), *Japanese and Korean Linguistics (Proceeding)*, volume 13. Stanford: CSLI, pp. 29–40.
- Hisagi, M., Shafer, V.L., Strange, W., Sussman, E., 2010. Perception of a Japanese vowel length contrast by Japanese and American English listeners: behavioral and electrophysiological measures. *Brain Res.* 1360, 89–105.
- Hisagi, M., Miyagawa, S., Shafer, V.L., Kotek, H., Sugawara, A., Pantazis, D., 2014. Perception of Japanese vowel duration contrasts by L1 and L2 learners of Japanese: an EEG study. In: MITWPL #73 Proceedings of FAJL 7, pp. 45–55.
- Hisagi, M., Shafer, V.L., Strange, W., Sussman, E., 2015. Perception of Japanese consonant length contrasts by Japanese and American English Listeners: Behavioral and electrophysiological measures. *Brain Res.* 1626, 218–231.
- Hisagi, M., 2007. Perception of Japanese temporally-cued phonetic contrasts by Japanese and American English Listeners: behavioral and electrophysiological measures. (Dissertation), Speech and Hearing Sciences, The City University of New York – Graduate Center.
- Flege, J., 2009. Give input a chance. In: Piske, T., Young-Scholten, M. (Eds.), *Input Matters in SLA. Multilingual Matters*, Bristol, 175–190.
- Hirata, Y., Whitehurst, E., Cullings, E., 2007. Training native English speakers to identify Japanese vowel length with sentences at varied speaking rates. *J. Acoust. Soc. Am.* 121, 3837–3845.
- Hisagi, M., Strange, W., 2011. Perception of Japanese Temporally-Cued Contrasts by American English Listeners. *Language and Speech*
- Knudsen, E.I., 2004. Sensitive periods in the development of the brain and behavior. *J. Cogn. Neurosci.* 16 (8), 1412–1425.
- Levy, E., Strange, W., 2008. Perception of French vowels by American English adults with and without French language experience. *J. Phon.* 36, 141–157.
- Menning, H., Imaizumi, S., Zwitserlood, P., Pantev, C., 2002. Plasticity of the human auditory cortex induced by discrimination learning of non-native, mora-timed contrasts of the JP language. *Learn. Mem.* 9, 253–267.
- Näätänen, R., 1990. The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive functions. *Behav. Brain Sci.* 13, 201–288.
- Näätänen, R., Paavilainen, P., Rinne, T., Alho, K., 2007. The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clin. Neurophysiol.* 118, 2544–2590.
- Nenonen, S., Shestakova, A., Huottilainen, M., Näätänen, R., 2003. Linguistic relevance of duration within the native language determines the accuracy of speech-sound duration processing. *Cogn. Brain Res.* 16, 492–495.
- Nenonen, S., Shestakova, A., Huottilainen, M., Näätänen, R., 2005. Speech-sound duration processing in a second language is specific to phonetic categories. *Brain Lang.* 92, 26–32.
- Peltola, M., Tuomainen, O., Koskinen, M., Aaltonen, O., 2007. The effect of language immersion education on the preattentive perception of native and non-native vowel contrasts. *J. Psycholinguist. Res.* 36 (1), 15–23.
- Peltola, M.S., Kujala, T., Tuomainen, J., Ek, M., Aaltonen, O., Näätänen, R., 2003. Native and foreign vowel discrimination as indexed by the mismatch negativity (MMN) response. *Neurosci. Lett.* 352, 25–28.
- Peltola, M.S., Kuntola, M., Tamminen, H., Hämäläinen, H., Aaltonen, O., 2005. Early exposure to non-native language alters preattentive perception. *Neurosci. Lett.* 388, 121–125.
- Pisoni, D.B., Lively, S.E., 1995. Variability and invariance in speech perception: A new look at some old problems in perceptual learning. In: Strange, W. (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research* (pp. 433–459). York Press, Baltimore, MD.
- Shafer, L.V., Schwartz, R.G., Kurtzberg, D., 2004. Language-specific memory traces of consonants in the brain. *Cogn. Brain Res.* 18, 242–254.
- Strange, W., 2011. Automatic Selective Perception (ASP) of First and Second Language Speech: a Working Model. *J. Phon.* <http://dx.doi.org/10.1016/j.wocn.2010.09.001>.
- Strange, W., Shafer, V.L., 2008. Speech perception in late second language learners: the re-education of selective perception. In: Hansen, J.G., Zampini, M.L. (Eds.), *Phonology and Second Language Acquisition* John Benjamins. Cambridge University Press, Amsterdam, 153–191.
- Sussman, E., 2007. A new view on the MMN and attention debate: auditory context effects. *J. Psychophysiol.* 21 (3–4), 164–175.
- Tabri, D., Abou Chacra, K.M.S., Pring, T., 2011. Speech perception in noise by monolingual, bilingual and trilingual listeners. *Int. J. Lang. Commun. Disord.* 46, 411–422.
- Tadel, F., Baillet, S., Mosher, J.C., Pantazis, D., Leahy, R.M., 2011. Brainstorm: a user-friendly application for meg/eeg analysis. *Comput. Intell. Neurosci.* 2011. <http://dx.doi.org/10.1155/2011/879716>.
- Tajima, K., Kato, H., Rothwell, A., Akahane-Yamada, R., Munhall, K.G., 2008. Training English listeners to perceive phonemic length contrasts in Japanese. *J. Acoust. Soc. Am.* 123, 397–413.
- Tajima, K., Kato, H., Rothwell, A., Munhall, K.G., 2003. Perception of phonemic length contrasts in Japanese by native and non-native listeners. In: *Proceedings of the 15th ICPHS, Barcelona*.
- Taulu, S., Simola, J., 2006. Spatiotemporal signal space separation method for rejecting nearby interference in MEG measurements. *Phys. Med. Biol.* 51, 1759–1768.
- Taulu, S., Kajola, M., Simola, J., 2004. Suppression of interference and artifacts by the signal space separation method. *Brain Topogr.* 16 (4), 269–275.
- Winkler, I., Kujala, T., Alku, P., Näätänen, R., 2003. Language context and phonetic change detection. *Cogn. Brain Res.* 17, 833–844.
- Werker, J.F., Tees, R.C., 1984. Cross-language speech perception: evidence for perceptual reorganization during the first year of life. *Infant Behav. Dev.* 7, 49–63.
- Ylinen, S., Uther, M., Latvala, A., Vepsäläinen, S., Iverson, P., Akahane-Yamada, R., Näätänen, R., 2010. Training the brain to weight speech cues differently: a study of Finnish second-language users of English. *J. Cogn. Neurosci.* 22 (6), 1319–1332.
- Zhang, Y., Kuhl, P.K., Imada, T., Iverson, P., Pruitt, E.B., Kawakatsu, M., Tohkura, Y., Nemoto, I., 2009. Neural signatures of phonetic learning in adulthood: a magnetoencephalography study. *NeuroImage* 46, 226–240.