

Asian Journal of
Biological
Sciences

Source Generator of Changing from Level Tone to Contour Tone Indexed by Mismatch Negativity

Wichian Sittiprapaporn

Faculty of Medicine, Mahasarakham University, Maha Sarakham, 44000, Thailand

ABSTRACT

The main objective of the work reported here was to investigate the pre-attentive processing of changing from level tone to contour tone and vice versa. The long-to-short duration change with level-to-falling/rising and falling/rising-to-level tone changes perception elicited mismatch negativity (MMN) between 196-220 msec with reference to the standard-stimulus event-related potentials (ERPs). The long-to-short duration and level-to-falling/rising tone changes elicited a strong MMN bilaterally for both native and nonnative speakers of Thai, unlike short-to-long duration and falling/rising-to-level tone changes. Source localization analyses performed using LORETA-Key demonstrates that sources were obtained in the Middle Temporal Gyrus (MTG) of the left hemisphere and in the Superior Temporal Gyrus (STG) of the right hemisphere for both groups. The present study demonstrated that the grand enterprise of mapping language onto the human brain can be vitally enhanced by MMN studies.

Key words: Event-related potential, sound, level tone, contour tone, mismatch negativity

INTRODUCTION

When engaged in a conversation, listeners tune in to the relevant stream of speech and filter out irrelevant speech input that may be present in the same environment. Nonetheless, attention might be involuntarily diverted to meaningful items coming from an ignored stream, like in the well-known own-name effect. This brings up the question of to what extent speech is processed in the ignored streams. A fundamental, yet controversial issue is whether the human brain contains neural circuits that are uniquely engaged in the early, pre-attentive stage of speech processing (Naatanen, 1992). While it seems indisputable that language is subserved by Left-hemisphere (LH) and Right-hemisphere (RH) are lateralized for speech, language or something else. Hypotheses proposed to account for functional hemispheric asymmetries can generally be classified as either cue dependent i.e., basic neural mechanism underlie processing of complex auditory stimuli regardless of linguistic relevance (Ivry and Roberson, 1998) or task dependent, i.e., specialized neural mechanisms exist that are activated only by speech (Imaizumi *et al.*, 1998).

In earlier studies (Gandour, 1998; Gandour *et al.*, 2000; Hsieh *et al.*, 2001), Chinese and English listeners did not show the same left-hemisphere lateralization as Thai listeners when making perceptual judgments of Thai tones. In addition, Chinese and English listeners were asked to make perceptual judgments of Chinese tones, consonants and vowels. Chinese listeners showed left-hemisphere lateralization for both suprasegmental and segmental phonological units (Hsieh *et al.*, 2001). These earlier studies suggest that functional circuits engage in early, pre-attentive speech perception of either suprasegmental or segmental units in tone languages. Previous event-related potential (ERP) studies at a phonetic level demonstrated that the Mismatch

Negativity (MMN) was enhanced in Finnish subjects by their first-language (Finnish) phoneme prototype rather than a non-prototype (Estonian) (Naatanen, 1992) and that the MMN for a vowel contrast in Finnish was not generated in native Hungarian speakers with no knowledge of Finnish (Winkler *et al.*, 1999), implying that the MMN reflects language-specific memory traces formed by early and extensive exposure to a first language. However, language-specific word-related MMN/MMF components at acoustic and phonetic levels remain to be investigated in future studies. The differences between these studies provide the impetus for future investigations of duration processing and temporal integration differences across language groups.

The present study proposes a task-dependent hypotheses in which the phonological units of segmental (e.g., consonant and vowel) and suprasegmental (e.g., prosodic) are processed in discrete asymmetry of the human brain. Although, the left hemisphere is selectively employed for processing linguistic information irrespectively of acoustic cues or subtype of phonological unit, the right hemisphere is employed for prosody-specific cues (Imaizumi *et al.*, 1998). The propose of the present study is, thus, to use both an auditory MMN component of ERP recording and the low resolution electromagnetic tomography (LORETA) techniques to measure the degree of cortical activation and to localize the brain area contributing to the scalp recorded auditory MMN component, respectively, during the passive oddball paradigm. In addition, the importance of the development of a MMN using a passive oddball paradigm for the neurophysiological study of pre-attentive processing in both native and nonnative speakers of Thai is emphasized. Thus, the specific objectives of the work reported here is to investigate the pre-attentive processing of changing from level tone to contour tone and vice versa.

MATERIALS AND METHODS

Subjects: Twenty-two healthy right-handed adults with normal hearing and no known neurological disorders volunteered for participation: eleven Native Speakers (NS) of Thai, aged 23-33 (mean 24.2; five females) and eleven Nonnative Speakers (NonS), aged 28-32 (mean 31.2; seven females). The approval of the institutional committee on human research and written consent from each subject were obtained.

Stimuli and procedures: Stimuli consisted of four pairs of monosyllabic, Thai words. Speech stimuli were digitally generated and edited to have equal peak energy level in decibels SPL with the remaining data within each of the stimuli scaled accordingly using the Cool Edit Pro v. 2.0 (Syntrillium Software Corporation). The sound pressure levels of speech stimuli were then measured at the output of the earphones (E-A-RTONE 3A, 50 Ω) in dBA using a Brüel and Kjaer 2230 sound-level meter. Pairs were designed to have similar long vowel duration. Three different stimuli were synthetically generated:

- **Stimuli 1:** /k^haam/-level tone
- **Stimuli 2:** /k^haam/-falling
- **Stimuli 3:** /k^haam/-rising

Five NS listened to the synthesized words and evaluated them all as natural sounding. The standard (S)/deviant (D) pairs for each experiment which was randomized across subjects, were shown:

- **Experiment 1: Standard (1)-Deviant (2):** (Stimuli 1: /k^haam/-level tone)-(Stimuli 2: /k^haam/-falling)
- **Experiment 2: Standard (2)-Deviant (1):** (Stimuli 2: /k^haam/-falling)-(Stimuli 1: /k^haam/-level tone)
- **Experiment 3: Standard (1)-Deviant (3):** (Stimuli 1: /k^haam/-level tone)-(Stimuli 3: /k^haam/-rising)
- **Experiment 4: Standard (3)-Deviant (1):** (Stimuli 3: /k^haam/-rising)-(Stimuli 1: /k^haam/-level tone)

The sounds were presented binaurally via headphones (Telephonic TDH-39-P) at 85 dB. The Inter-stimulus Interval (ISI) was 1.25 second (offset-onset). Deviant stimuli appeared randomly among the standards at 10% probability. Each experiment included 125 trials (10% D). The stimuli were binaurally delivered using SuperLab software (Cedrus Corporation, San Pedro, USA) via headphones (Telephonic TDH-39-P). EEG signal recording was time-locked to the onset of a word. Subjects were instructed not to pay attention to the stimuli presented via headphones but rather to concentrate on a self-selected silent, subtitled movie.

Electroencephalographic recording: For EEG/ERP recording, the standard 20 locations of the 10-20 system, EEG is recorded via an electro-cap (Electro-cap International) from 20 active electrodes (Fp1, Fp2, F7, F3, Fz, F4, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1, Oz, O2) positioned according to the 10-20 International System of Electrode Placement, plus Oz and Ground are applied, pre-mounted in an elastic Electro-Cap. Reference electrodes are manually applied to left and right mandibles, where the Fp1 and Fp2 electrodes are used for ocular artifact detection. Horizontal eye movements are monitored with electrodes at the left and right outer canthi and vertical eye movements are monitored at Fp1 and Fp2. EEG is amplified with a gain of 30,000 and filtered with a bandpass of 0.1-30 Hz. EEGs are acquired as continuous signals and are subsequently segmented into epochs of 1 sec (a 100 msec pre-stimulus baseline and a 900 msec post-stimulus epoch).

EEG data processing: The recordings were filtered and carefully inspected for eye movement and muscle artifacts. ERPs were obtained by averaging epoch which started 100 msec before the stimulus onset and ended 900 msec thereafter; the -100 to 0 msec interval was used as a baseline. Epochs with voltage variation exceeding $\pm 100 \mu\text{V}$ at any EEG channel were rejected from further analysis. The MMN was obtained by subtracting the response to the standard from that to the deviant stimulus. All responses were recalculated offline against average reference for further analysis.

Spatial analysis: The average MMN latency was defined as a moment of the global field power with an epoch of 40 msec time window related stable scalp-potential topography (Pascual-Marqui *et al.*, 1994). In the next step, low-resolution electromagnetic tomography (LORETA) is applied to estimate the current source density distribution in the brain which contributes to the electrical scalp field (Pascual-Marqui *et al.*, 1994). Maps are computed with LORETA. LORETA computes the smoothest of all possible source configurations throughout the brain volume by minimizing the total squared Laplacian of source strengths.

Data analysis: During the auditory stimulation, electric activity of the subjects' brain was continuously recorded. The MMN was obtained by subtracting the response to the standard from

that to the deviant stimulus. The statistical significance of MMN was tested with one sample t-test. An across-experiment ANOVA was carried out so as to make cross-linguistic comparisons.

RESULTS

The grand-averaged ERPs show that both level-to-contour and contour-to-level tone changes perception elicited MMN between 196-220 msec with reference to the standard-stimulus ERPs. An ANOVA comparing MMN amplitudes of the S and D yield a main effect of conditions in Experiments 1 ($F_{(3,40)} = 10.65$, $p < 0.0001$) and 3 ($F_{(3,40)} = 6.64$, $p < 0.0001$). In Experiment 2 and 4, however, the S-D differences were not significant (e.g., $F_{(3,40)} = 1.44$, $p = 0.1808$ in Experiment 2, NS, $F_{(3,40)} = 1.11$, $p = 0.2948$ in Experiment 4, NS, for the main effect of conditions). The result showed that level-to-falling and level-to-rising tone changes elicited a strong MMN bilaterally for NS and NonS, unlike falling-to-level and rising-to-level tone changes (Table 1). Furthermore, an across-experiment ANOVA demonstrated an interaction and main effects. The significant difference in MMN amplitudes was observed between groups across experiments ($F_{(7,80)} = 46.66$, $p < 0.0001$).

Source localization analyses were performed using LORETA-Key (Pascual-Marqui *et al.*, 1994). Table 2 demonstrates the xyz-values in Talairach space as calculated with LORETA in the time window 196-220 msec. In all experiments, a single source was estimated to be located in the Middle

Table 1: Amplitude (μV) of MMN elicited by a level-to-contour tones perception in NS and NonS

Tone changes	Native speaker (NS)	Nonnative speaker (NonS)
Mid-to-falling	-3.03 $\pm$ 0.74	-2.41 $\pm$ 0.94
Falling-to-mid	-0.40 $\pm$ 0.55	-2.77 $\pm$ 0.62
Mid-to-rising	-1.62 $\pm$ 0.42	-2.22 $\pm$ 0.59
Rising-to-mid	-0.20 $\pm$ 0.41	-1.58 $\pm$ 0.56

Values are Mean $\pm$ SD

Table 2: Stereotaxic coordinates of activation foci during the level-to-contour tone changes perception

	Coordinates (mm)				
Tone change	x	y	z	BA	t-value
Native speaker (NS)					
Mid-to-falling	-52	-60	15 ^a	22	3.73
	53	-60	15 ^b	22	3.42
Falling-to-mid	-52	-67	8 ^c	39	3.48
	60	-39	15 ^b	22	2.12
Mid-to-rising	-59	-32	8 ^a	42	3.67
	53	-53	15 ^b	39	3.20
Rising-to-mid	-52	-67	8 ^c	39	2.59
	46	-67	8 ^d	37	1.11
Non-native speaker (NonS)					
Mid-to-falling	-52	-67	8 ^c	39	4.58
	53	-60	8 ^d	39	3.90
Falling-to-mid	-52	-67	8 ^c	39	4.52
	60	-32	6 ^d	21	4.44
Mid-to-rising	-59	-32	1 ^c	21	4.40
	46	-67	8 ^d	37	1.06
Rising-to-mid	-52	-67	8 ^c	39	4.28
	53	-60	8 ^d	39	3.83

^aLeft superior temporal gyrus, ^bRight superior temporal gyrus, ^cLeft middle temporal gyrus, ^dRight middle temporal gyrus

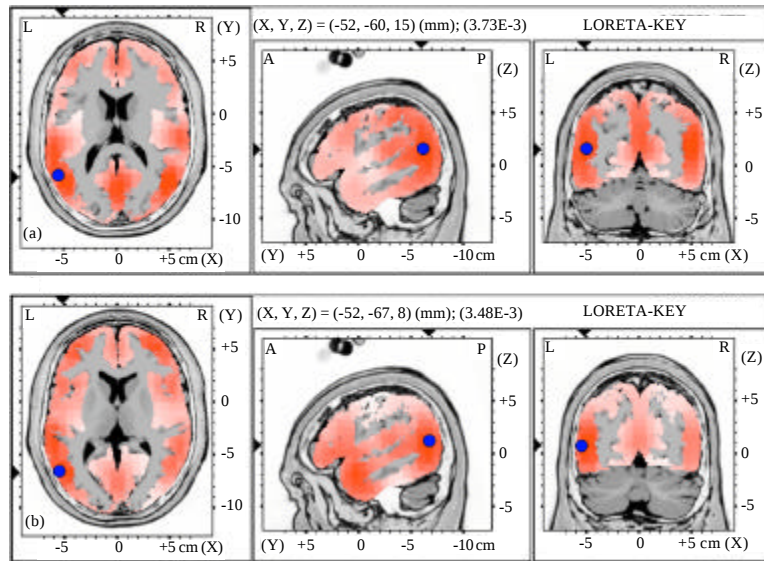


Fig. 1(a-b): Graphical representation of the low-resolution electromagnetic tomography (LORETA) t-statistic comparing the event-related potentials for mismatch negativity (MMN) responses at the time point of the individual peak over Fz for (a) Level-to-falling and (b) Falling-to-level tones change occupied by the long vowel duration of native speaker (NS) activated in left hemisphere (LH). Red color indicates local maxima of increased electrical activity for across- and within-category of vowel change responses in an axial, a sagittal and a coronal slice through the reference brain. Blue dots mark the center of significantly increased electric activity

Temporal Gyrus (MTG) of each hemisphere for NonS group. In Experiment 1, 3 and 4, a single source was estimated to be located in the Middle Temporal Gyrus (MTG), for Experiment 4 and the Superior Temporal Gyrus (STG), for Experiment 1 and 3, of each hemisphere for NS group. In Experiment 2, sources were obtained in the MTG of LH and in the Superior Temporal Gyrus (STG) of RH for NonS group (Table 2, Fig. 1-8).

DISCUSSION

The long-to-short duration change with level-to-falling/rising and falling/rising-to-level tone changes perception elicited MMN between 196-220 msec with reference to the standard-stimulus ERPs. The long-to-short duration and level-to-falling/rising tone changes elicited a strong MMN bilaterally for both native and nonnative speakers of Thai, unlike short-to-long duration and falling/rising-to-level tone changes. Source localization analyses performed using LORETA-Key demonstrates that sources were obtained in the Middle temporal Gyrus (MTG) of the left hemisphere and in the Superior Temporal Gyrus (STG) of the right hemisphere for both groups.

The Main finding is that for both groups, vowel-duration shortening and falling tone elicited prominent MMN components, whereas vowel-duration lengthening and level tone. The same results were obtained in MEG studies using both tone and Japanese words (Inouchi *et al.*, 2002, 2003). The present findings are in accord with previous vowel experiment that showed that shortening elicited a larger MMN than lengthening (Inouchi *et al.*, 2002; Jaramillo *et al.*, 1990).

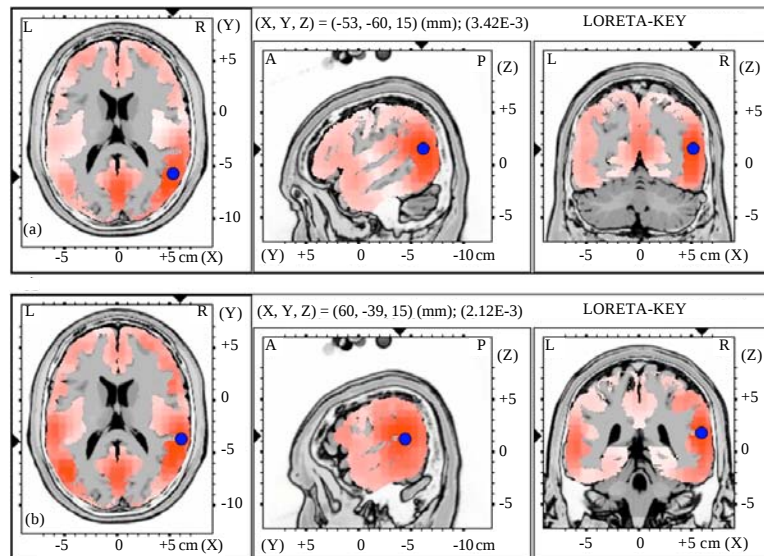


Fig. 2(a-b): Graphical representation of the low-resolution electromagnetic tomography (LORETA) t-statistic comparing the event-related potentials for MMN responses at the time point of the individual peak over Fz for (a) Level-to-falling and (b) Falling-to-level tone changes occupied by the long vowel duration of NS activated in RH

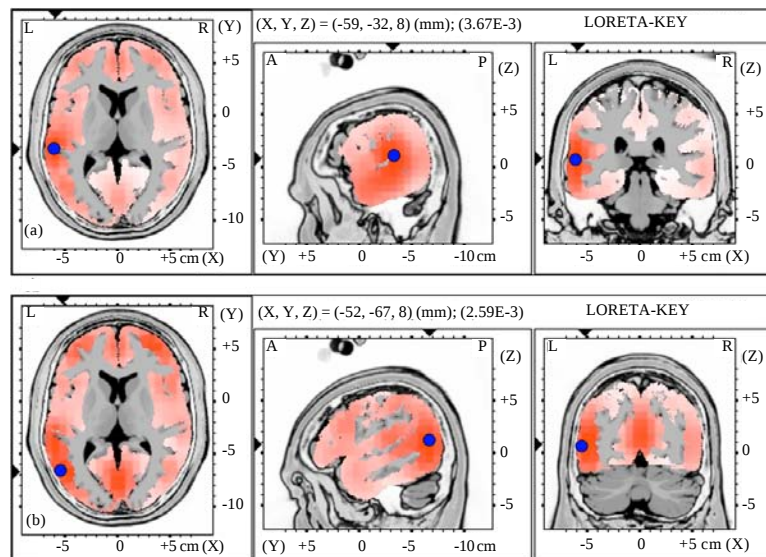


Fig. 3(a-b): Graphical representation of the low-resolution electromagnetic tomography (LORETA) t-statistic comparing the event-related potentials for MMN responses at the time point of the individual peak over Fz for (a) Level-to-rising and (b) Rising-to-level tone changes occupied by the long vowel duration of NS activated in LH

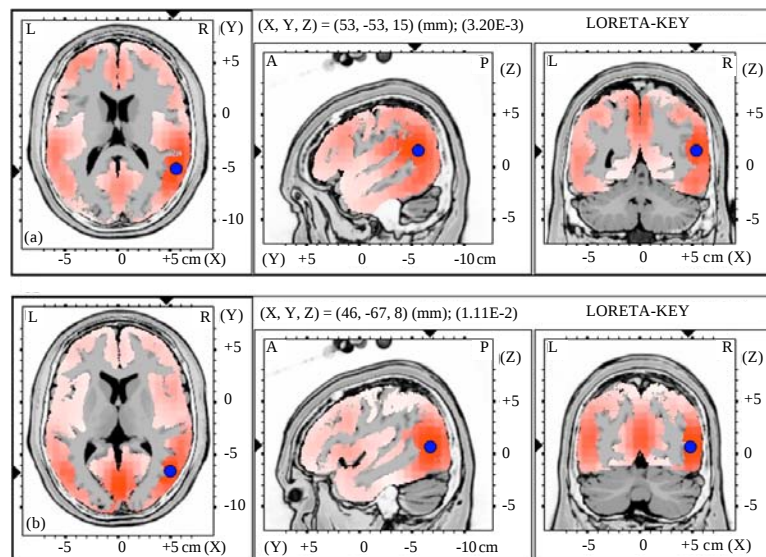


Fig. 4(a-b): Graphical representation of the low-resolution electromagnetic tomography (LORETA) t-statistic comparing the event-related potentials for MMN responses at the time point of the individual peak over Fz for (a) Level-to-rising and (b) Rising-to-level tone changes occupied by the long vowel duration of NS activated in RH

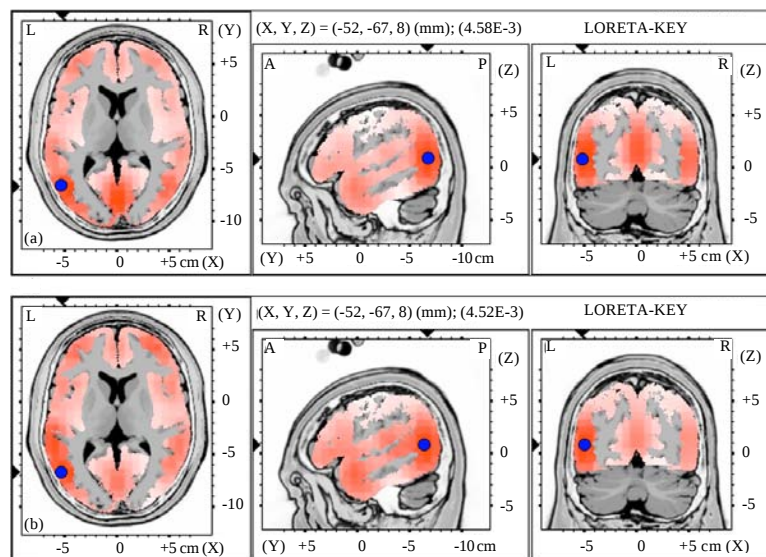


Fig. 5(a-b): Graphical representation of the low-resolution electromagnetic tomography (LORETA) t-statistic comparing the event-related potentials for MMN responses at the time point of the individual peak over Fz for (a) Level-to-falling and (b) Falling-to-level tone changes occupied by the long vowel duration of NonS activated in LH

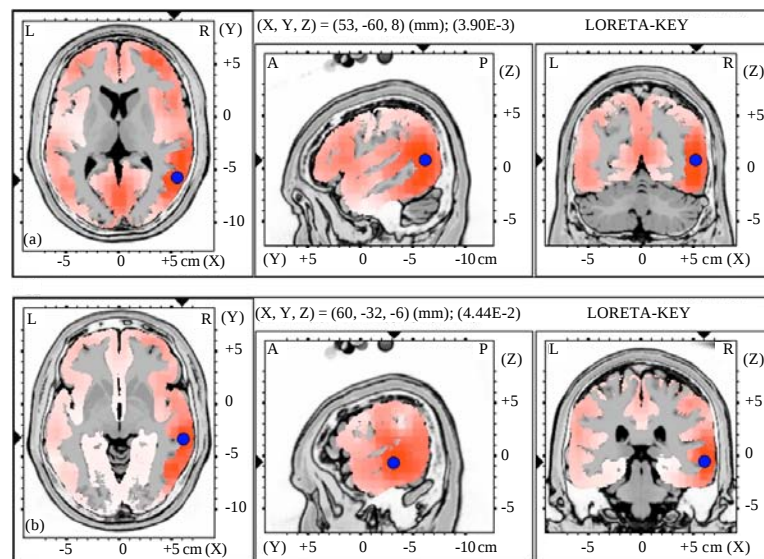


Fig. 6(a-b): Graphical representation of the low-resolution electromagnetic tomography (LORETA) t-statistic comparing the event-related potentials for MMN responses at the time point of the individual peak over Fz for (a) Level-to-falling and (b) Falling-to-level tone changes occupied by the long vowel duration of NonS activated in RH

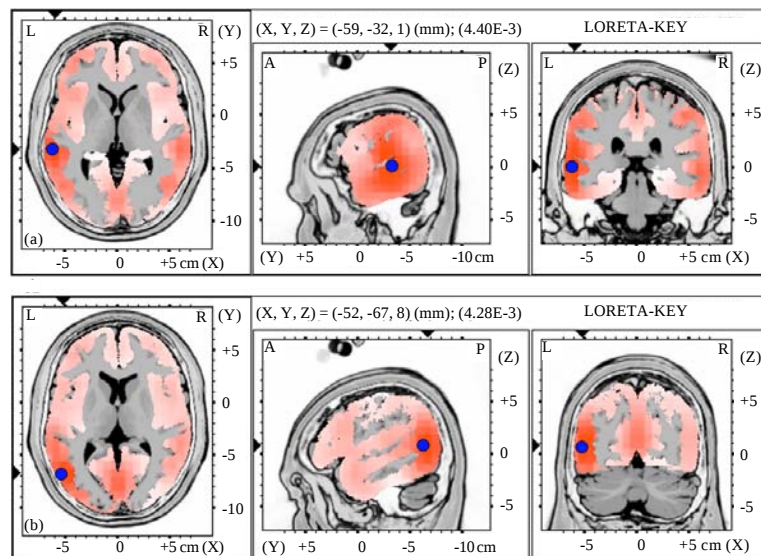


Fig. 7(a-b): Graphical representation of the low-resolution electromagnetic tomography (LORETA) t-statistic comparing the event-related potentials for MMN responses at the time point of the individual peak over Fz for (a) Level-to-rising and (b) Rising-to-level tone changes occupied by the long vowel duration of NonS activated in LH

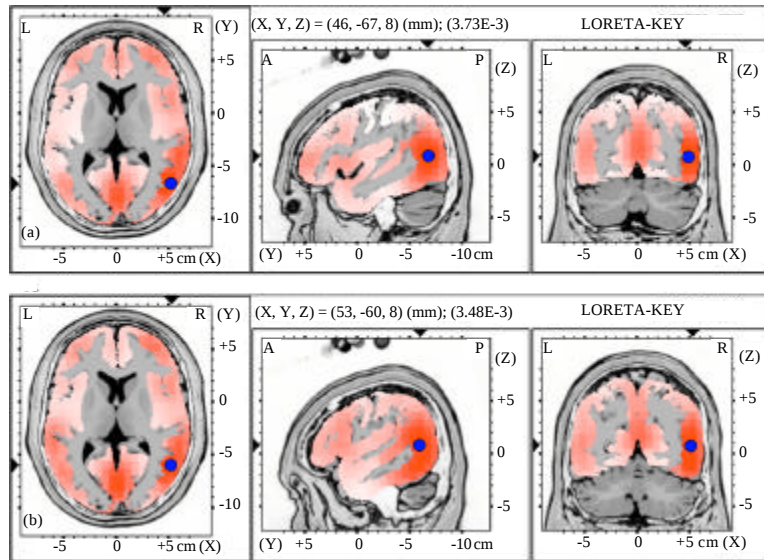


Fig. 8(a-b): Graphical representation of the low-resolution electromagnetic tomography (LORETA) t-statistic comparing the event-related potentials for MMN responses at the time point of the individual peak over Fz for (a) Level-to-rising and (b) Rising-to-level tone changes occupied by the long vowel duration of NonS activated in RH

This suggests that at the point of stimulus disparity or thereafter, change detection of vowel lengthening is somehow compromised. It is reasonable to speculate that the continued auditory processing required for the longer vowel interferes with or masks the detection mechanism underlying the MMN (Inouchi *et al.*, 2002). In contrast, the current findings contradict previous tone studies that reported a clear MMN elicited by both duration increments and decrements (Naatanen *et al.*, 1989) and a larger MMN elicited by increments than decrements (Jaramillo *et al.*, 1990). However, the stimulus and presentation between these studies differed considerably; notably the referred to tone study contained stimuli of short durations compared to our longer ‘word length’ stimuli and this might explain the disparity.

The present results parallel the finding in previous studies (Inouchi *et al.*, 2002, 2003). The present study, the detection of vowel duration and tone changes was most likely acoustically driven rather than semantically driven, such that the stimuli were processed without any access to semantic information. The acoustic aspect in the absence of phonetic or higher-order properties may account for why NonS had similar neuronal responses to NS subjects. Additionally, the length of our stimuli may have contributed to the lack of difference in MMN between NS and NonS. Temporal integration of acoustic energy is therefore purposed to take place in the time range of the initial 200 msec of the stimulus (Naatanen, 1992). After such time, auditory processes have integrated the sensory components as a unitary event in sensory memory (Naatanen, 2001). Nenonen *et al.* (2003) who found cross-linguistic differences for processing shortened durations used deviants that were shortened from 200-150 msec, a deviation that may have straddled the integration threshold for one or both groups. Therefore, our stimuli (e.g., Thai words) were considerably longer and most likely were integrated in sensory memory. The differences between these studies provide the impetus for future investigations of duration processing and temporal integration differences across language groups.

The MMN component was also found to be more sensitive to tone falling than rising and leveling for both groups. To the extent that both shortening-duration and falling tone conditions reduce acoustic energy in the deviant compared to the standard, similar sensitivity in the MMN is not surprising. The delay in the MMN for falling tone, compared to that of shortening duration, may be due to the incremental spectral deviation, as opposed to the rapid change (omission) in the shortened stimulus (Inouchi *et al.*, 2002).

The source analyses of the MMN components suggest that shortened-duration and tone changes may elicit activity in the Superior Temporal Sulcus (STS) of each hemisphere for both NS and NonS. Source modeling using a single equivalent dipole approach has well-recognized spatial limitations, perhaps accounting for the discrepancy between current findings and previous reports of MMN/MMF generators in the planum temporale. However, while there existed no hemispheric differences in the current study, the present studies are in line with previous f-MRI (Binder *et al.*, 1997) and MEG (Inouchi *et al.*, 2002) reports that found the left posterior superior temporal gyrus to be activated by the pre-attentive detection of acoustic changes in non-speech (tones) and speech (CV syllables). However, a previous study (Inouchi *et al.*, 2002) purposed that it may be that duration changes in tones activated more inferior brain regions centered in STS in contrast to spectral changes centered in the lateral fissure but such inferences are beyond the scope of the study, since only one subject in each group had MR imaging.

CONCLUSION

The MMN component is sensitive to vowel shortening rather than lengthening and to pitch falling rather than rising and leveling. Automatic detection of changes in vowel shortening and pitch falling is a useful index of language universal auditory memory traces. While the right hemisphere is predominant in the perception of the non-native speech sounds, the left hemisphere is predominant in the perception of the native speech sounds. The present study has also demonstrated that the grand enterprise of mapping language onto the human brain can be vitally enhanced by MMN studies.

REFERENCES

- Binder, J.R., J.A. Frost, T.A. Hammeke, R.W. Cox, S.M. Rao and T. Prieto, 1997. Human brain language areas identified by functional magnetic resonance imaging. *J. Neurosci.*, 17: 353-362.
- Gandour, J., 1998. Aphasia in Tone Languages. In: *Aphasia in Atypical Populations*, Coppers, P., A. Basso and Y. Lebrun (Eds.). Lawrence Erlbaum, Hillsdale, NJ., USA., pp: 117-141.
- Gandour, J., D. Wong, L. Hsieh, B. Weinzaepfel, D. van Lacker and G.D. Hutchins, 2000. A crosslinguistic PET study of tone perception. *J. Cognitive Neurosci.*, 12: 207-222.
- Hsieh, L., J. Gandour, D. Wong and G.D. Hutchins, 2001. Functional heterogeneity of inferior frontal gyrus is shaped by linguistic experience. *Brain Lang.*, 76: 227-252.
- Imaizumi, S., K. Mori, S. Kiritani, H. Hosoi and M. Tonoike, 1998. Task-dependent laterality for cue decoding during spoken language processing. *Neuroreport*, 9: 899-903.
- Inouchi, M., M. Kubota, P. Ferrari and T.P.L. Roberts, 2002. Neuromagnetic auditory cortex responses to duration and pitch changes in tones: Cross-linguistic comparisons of human subjects in directions of acoustic changes. *Neurosci. Lett.*, 331: 138-142.
- Inouchi, M., M. Kubota, P. Ferrari and T.P.L. Roberts, 2003. Magnetic mismatch fields elicited by vowel duration and pitch changes in Japanese words in humans: Comparison between native-and non-speakers of Japanese. *Neurosci. Lett.*, 353: 165-168.

- Ivry, R. and L. Roberson, 1998. *The Two Sides of Perception*. MIT Press, Cambridge, MA., USA.
- Jaramillo, M., P. Alku and P. Paavilainen, 1990. An Event-Related Potential (ERP) study of duration changes in speech and non-speech sounds. *Neuroreport*, 10: 3301-3305.
- Naatanen, R., P. Paavilainen and K. Reinikainen, 1989. Do event-related potentials to infrequent decrements in duration of auditory stimuli demonstrate a memory trace in man? *Neurosci. Lett.*, 107: 347-352.
- Naatanen, R., 1992. *Attention and Brain Function*. Lawrence Erlbaum, Hillsdale, USA.
- Naatanen, R., 2001. The perception of speech sounds by the human brain as reflected by the Mismatch Negativity (MMN) and its Magnetic Equivalent (MMNm). *Psychophysiology*, 38: 1-21.
- Nenonen, S., A. Shestakova, M. Huutilainen and R. Naatanen, 2003. Linguistic relevance of duration within the native language determines the accuracy of speech-sound duration processing. *Cognitive Brain Res.*, 16: 492-495.
- Pascual-Marqui, R.D., C.M. Michel and D. Lehmann, 1994. Low resolution electromagnetic tomography: A new method for localizing electrical activity in the brain. *Int. J. Psychophysiol.*, 18: 49-65.
- Winkler, I., A. Lehtokoski, P. Alku, M. Vainio and I. Czigler *et al.*, 1999. Pre-attentive detection of vowel contrasts utilizes both phonetic and auditory memory representations. *Cognitive Brain Res.*, 7: 357-369.