Songklanakarin J. Sci. Technol.
38 (6), 599-603, Nov. - Dec. 2016

http://www.sjst.psu.ac.th

Original Article

Effects of prosody on spoken Thai word perception in pre-attentive brain processing: a pilot study

Kittipun Arunphalungsanti* and Chailerd Pichitpornchai

Department of Physiology, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok Noi, Bangkok, 10700 Thailand.

Received: 27 October 2015; Accepted: 15 February 2016

Abstract

This study aimed to investigate the effect of the unfamiliar stressed prosody on spoken Thai word perception in the pre-attentive processing of the brain evaluated by the N2a and brain wave oscillatory activity. EEG recording was obtained from eleven participants, who were instructed to ignore the sound stimuli while watching silent movies. Results showed that prosody of unfamiliar stress word perception elicited N2a component and the quantitative EEG analysis found that theta and delta wave powers were principally generated in the frontal area.

It was possible that the unfamiliar prosody with different frequencies, duration and intensity of the sound of Thai words induced highly selective attention and retrieval of information from the episodic memory of the pre-attentive stage of speech perception. This brain electrical activity evidence could be used for further study in the development of valuable clinical tests to evaluate the frontal lobe function in speech perception.

Keywords: prosody, speech perception, pre-attentive, N2a, quantitative EEG

1. Introduction

Thai language is classified as a “tonal language” by using the 5 tones or 5 different frequencies of sound resulting in speech and different meanings for different words and tones (Honbolygó *et al.*, 2004; Kaan *et al.*, 2008). This is different from those classified as “stress languages” such as English which do not use tones but use the stress of words to distinguish their meanings and use emotive expression in speech communication. The variation in “frequency” of tonal and stress languages is called “Suprasegment” or the prosody of speech (Chandrasekaran *et al.*, 2007; Honbolygó *et al.*, 2004). The melody of speech is a result of the fluctuations in pitch, rhythm, and stress as found in stress and tonal languages (Chandrasekaran *et al.*, 2007; Honbolygó *et al.*, 2004; Kaan *et al.*, 2008).

* Corresponding author.
Email address: kitctu@gmail.com

Lexical stress or *word stress* is the stress placed on syllables within words. It is the linguistic unit expressed from the environment. On the other hand, the stress placed on words within sentences is called *sentence stress* or *prosodic stress* (Terken, 1991). The differences in emphasis are categorized into four varied forms: the length, the volume, the frequency (pitch), and the quality of voice related to physical properties including duration, intensity, and fundamental frequency of the speech (Fry, 1958). The stressed words have higher fundamental frequency, higher intensity, and longer duration than unstressed words (Fry, 1958).

Sound discrimination was studied in subjects who were told to ignore the sound stimulus (Jemel *et al.*, 2002; Näätänen *et al.*, 2004; Rinne *et al.*, 1999). Their brain activity in the pre-attentive stage elicited the N2a component or Mismatch Negativity component (MMN). The N2a component was a subcomponent of N200 which was the anterior cortical distribution evoked while ignoring any deviant stimulus (Patel and Azzam, 2005). The MMN was a negative component calculated by subtracting the event-related potential (ERP) components of deviant stimuli from that of the standard stimuli. It was found at 100-250 ms in the fronto-central scalp areas (Fz) and is generated by the supratemporal and frontal processing areas (Jemel *et al.*, 2002; Näätänen *et al.*, 2004; Rinne *et al.*, 1999). The MMN can be used as an index for sensory (echoic) memory, pre-attentive processing, discrimination accuracy of sound features, brain plasticity, and the effects of training or experience of sound such as language and music (Kujala *et al.*, 2007). Previous studies performing the spectral analysis of quantitative EEG found that the MMN was related to the spectral amplitude peak at the 3–9 Hz which is the range of theta and alpha oscillatory activities in normal subjects (Fuentemilla *et al.*, 2008; Hsiao *et al.*, 2009; Javitt *et al.*, 2000).

The stimulation of word stress in unattended condition associated with the brain processing in the pre-attentive stage (Ylinen *et al.*, 2009; Zora, 2011). A previous study examined the effect of English stress words in both words and pseudo-words (familiar and unfamiliar sounds) stimuli on brain processing by evaluating the MMN, showing that the MMN amplitude was increased when subjects heard unfamiliar words and unfamiliar word-stress pattern (Ylinen *et al.*, 2009). The effect of word stress patterns in English showed that word and pseudo-word stimuli elicited a biphasic MMN at 110-160 ms and 200-300 ms, respectively (Zora, 2011). Each stimulus was determined by differences of sound intensity, fundamental frequency, and combinations of intensity and frequency (Zora, 2011). As a result, MMN was the important cue for stress perception (Ylinen *et al.*, 2009; Zora, 2011). However, most studies of MMN characters were performed by using English as the sound stimuli. The present study aimed to investigate the effect of unfamiliar stressed prosody in Thai spoken word levels (which are different in frequency, duration, and amplitude) on the N2a component and brain wave oscillatory activity in native Thai speakers during the pre-attentive stages.

2. Materials and Methods

2.1 Participants

Eleven right-handed participants (handedness assessed by Edinburgh inventory test (Oldfield, 1971); 6 women and 5 men, aged 20-30 years, with normal hearing threshold, and a mean age of (± SD) 26.18±1.83 years, were recruited from the Faculty of Medicine, Siriraj Hospital, Mahidol University. All the participants had no history or presence of neurological illness, drug addiction, musical training (i.e., formal musical training within the past 7 years and 4 years or less total musical training), language-related disorders, or neuromuscular disorder of hands and fingers.

Prior to the experiment, each participant signed an informed consent form explaining the study details and stating that participants were free to leave the study at any time, and for any reason. The protocol was approved by the ethics committee of Siriraj Institutional Review Board, Faculty of Medicine, Siriraj Hospital, Mahidol University (Certificate of Approval number: Si314/2013).

2.2 Stimuli

The sound stimuli were recorded at the Educational Technology Studio, Faculty of Medicine, Siriraj Hospital, Mahidol University. Words were verbalized by a female adult Thai native speaker, because a female voice is more complex and clearer than a male voice, due to differences in the length and shape of the vocal cords and larynx (Belin *et al.*, 2004). The audio files were adjusted to be at the intensity level and frequency of 60-80 dB SPL and word duration was adjusted by Adobe audition software (Adobe Systems Incorporated, USA) and analyzed by Praat software (Phonetic Sciences Department, University of Amsterdam, Netherlands).

The stimuli were Thai monosyllabic words with corresponding meanings. A single word was recorded twice as a stressed word and an unstressed word. The components of stressed and unstressed words comprising of the frequency (Hz), amplitude (dB), and duration (ms) were analyzed and averaged by Praat software (Phonetic Sciences Department, University of Amsterdam, Netherlands). The mean (±SD) of frequency (Hz), amplitude (dB), and duration (ms) of the stressed words characters were 182.755±16.670 Hz, 76.952±2.096 dB, and 656.817±41.471 ms, and those of the unstressed words are 173.221±13.319 Hz, 74.192±1.560 dB, and 606.365±43.758 ms, respectively. All 3 components of stressed words were significantly higher than those of unstressed words. The audio files were sequenced by the Stim^2 software (Compumedics Neuroscan, USA) using the Gentask to set the inter-stimulus interval at 2,000 ms. The standard stimuli were the 280 unstressed words (80%) and deviant stimuli were the 70 stressed words (20%).

2.3 Experimental procedure

The subjects sat in front of a computer screen at 60 cm distance. To avoid recording artifacts, the subjects were asked to fix their eye focus at a red-cross on the center of the screen at all times and tried to avoid eye blinking or any head movement. Sound stimuli were presented in an auditory passive oddball paradigm involved in presenting a regular train of frequent (standard) and rare (deviant) stimuli which were binaurally presented via a pair of earphones. The right and left acoustic channels of the inserted earphones were calibrated for equal and comfortable loudness (60-80 dB SPL). The subjects were instructed not to attend to the sound stimuli but they had to watch the silent nature documentary without any subtitles. Meanwhile, the EEG was recorded.

2.4 EEG recording and analysis

The EEG was recorded by using the Neuroscan and Quick cap with 32 channels (Ag / Ag-Cl electrode) (Compumedics Neuroscan, USA). Electrodes were attached on the subject's scalp according to the international 10-20 system connected to the head-box of the SynAmps amplifier (Compumedics Neuroscan, USA). The configuration of the electrode positions was pre-defined according to the SynAmps Digital. The montage was referenced to both mastoid processes. Electrode impedance was kept below 10 kÙ. Filters were high pass 0.01 Hz and low pass 30 Hz.

The horizontal and vertical electro-oculograms were recorded and any trial with voltage about 100 mV would be removed by artifact rejection. ERPs were obtained by cutting epoch, which started about 100 ms before the stimulus onset and ended at 800 ms after recording. The -100 to 0 ms interval was used as a baseline by baseline correction. The averaged ERP contained at least 50% accepted deviant trials. The MMN was obtained by subtracting the response between the standard and the deviant stimuli. The quantitative EEG analysis of MMN was achieved by Fast Fourier transform (FFT) for spectral analysis evaluating the brain wave in each frequency band and presented as power in each wave.

The statistical tests of MMN and power of brain waves were analyzed using paired t-test in all the subjects whose data were collected from Fz, F3, and F4 electrodes.

3. Results and Discussion

3.1 ERP results

The grand-average ERPs of Fz, F3 and F4 electrode sites showed that the deviant stimulus elicited larger N2a between 100 and 200 ms than the standard stimuli. ERP and the MMN amplitude of the Fz electrode site were slightly larger than at the F3 and F4 electrode sites (Figure 1).

The N2a amplitude elicited by deviant stimulus was significantly larger than those elicited by standard stimuli at the Fz, F3 and F4 electrode sites. The N2a amplitude elicited by deviant stimulus at the Fz electrode sites was significantly larger than that at the F3 electrode sites (Table 1). The scalp distribution of the N2a of deviant stimulus was larger than the standard stimuli at 100-183 ms (Figure 2). The MMN component was not significantly different among Fz, F3 and F4 electrode sites (Table 2).

The main results showed that the prosody unfamiliar stressed word elicited N2a amplitude peaking at 100-200 ms which was larger than the prosody familiar stressed word reflecting the outcome of a comparison process between the presentation of the current event and a presentation of the regularities in the event history (Karanasiou *et al.*, 2010). The N2a relates to the modulation of the early stages of response preparation and selection (Kopp *et al.*, 1996). The subtraction of ERPs elicited from deviant and standard stimuli reflected the MMN amplitude. The MMN amplitude at frontrocentral area was slightly higher than those at the F3 and F4 electrode sites. The change of MMN amplitude was consistent with the previous study in the perception of stress pattern of English words, which is a stress language (Ylinen *et al.*, 2009; Zora, 2011). It suggests that the native Thai speakers

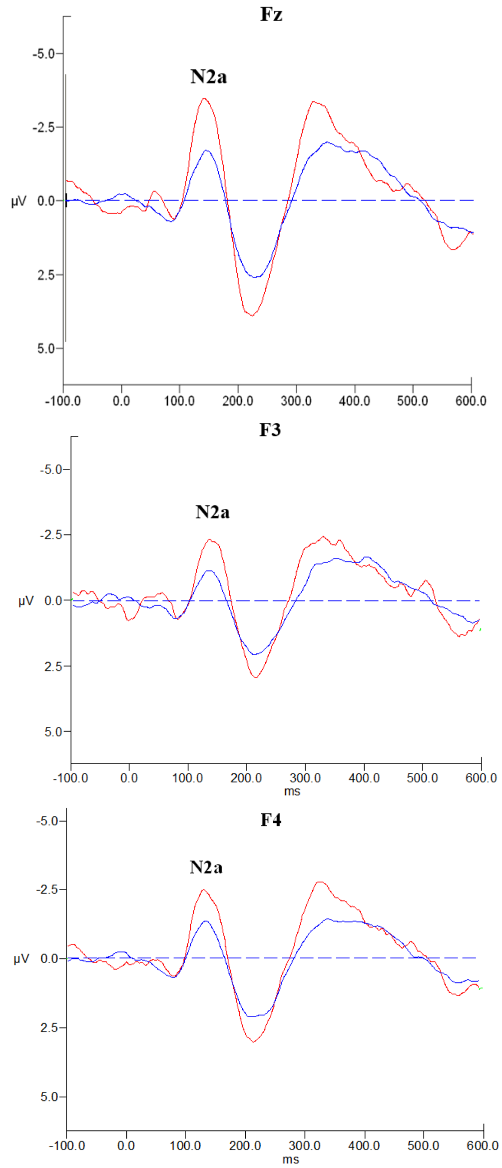

Figure 1. Grand average event-related potentials (n=11) elicited by deviant and standard stimuli at the Fz, F3 and F4 electrodes sites. (red line: deviant stimuli, blue line: standard stimuli)

Table 1. Mean ± SEM of the N2a amplitude at the Fz, F3 and F4 electrodes sites (n=11).

Electrodes	Mean ± SEM (µV)	
	Deviant stimuli	Frequent stimuli
Fz	-3.54 ± 0.479	-1.83 ± 0.313*
F3	-2.86 ± 0.425†	-1.51 ± 0.193*
F4	-3.15 ± 0.368	-1.64 ± 0.290*

* p-valued < 0.05, Compared between deviant and frequent stimuli

† p-valued < 0.05, Compared between electrode sites of deviant stimuli

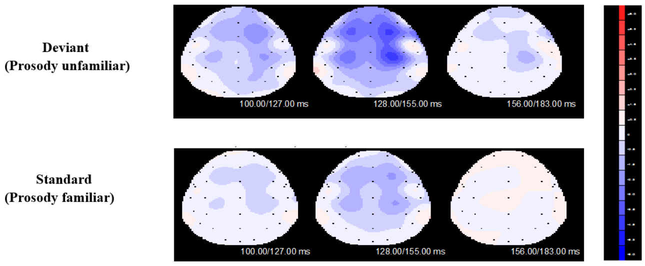

Figure 2. Scalp distribution of the N2a at time window 100-183 ms elicited by deviant and standard stimuli (n=11).

Table 2. Mean ± SEM of the MMN amplitude at the Fz, F3 and F4 electrodes sites (n=11).

Electrodes	Mean ± SEM (µV)
Fz	-2.15 ± 0.504
F3	-1.67 ± 0.387
F4	-1.80 ± 0.416

who are familiar to using the tones in Thai language can perceive and discriminate the prosody with stress comparable to the native stress language speakers at an early stage of speech perception. It also reflects that the frontrocentral area plays a role in discriminating unfamiliar prosody stress in Thai words of native Thai speakers with differences in frequency, intensity, and duration at the pre-attentive stage.

3.2 Quantitative EEG findings

The quantitative EEG analysis found that the mean power of the theta wave was higher than that of the delta, alpha and beta waves from all stimuli at all electrode sites. The mean power of the theta wave (2.85-7.12 Hz) from deviant stimulus was significantly higher than that elicited by standard stimuli over the F3 electrode site and theta wave from deviant stimulus at the Fz electrode site was significantly higher than those at the F3 and F4 electrode sites (Table 3). The scalp distribution of the power of the theta wave elicited both from the deviant and from the standard stimuli was highly distributed at the frontal area. The theta wave power distribution of deviant stimuli was slightly higher than that of the standard stimuli (Figure 3). The mean power of delta (0-2.85 Hz) from deviant stimulus at the Fz electrode site was significantly higher than that at the F3 electrode site (Table 4). Moreover, the mean power of delta from deviant stimulus was also slightly higher than that elicited by standard stimuli at all electrode sites.

The results of quantitative EEG analysis showed a significant increase in frontal theta and delta wave power correlated with the unfamiliar prosody perception. A previous review of theta wave suggested that the theta wave activity was related to the episodic long-term and working memory (Sauseng *et al.*, 2010). The cortical theta activity was increased during encoding of information into episodic long-term memory, during retrieval of information from episodic memory, and also during the information encoding and retention (Sauseng *et al.*, 2010). Increasing mean power of delta was observed in frontal area elicited from auditory target stimuli during oddball experiments reflecting signal detection and decision making (Başar *et al.*, 1999). This result suggests that the increase in mean power of delta is a key mechanism of selective attention to rhythmic auditory streams (Lakatos *et al.*, 2008).

Table 3. Mean ± SEM of the power of Theta wave at the Fz, F3 and F4 electrodes sites (n=11).

Electrodes	Mean ± SEM (μV^2)	
	Deviant stimuli	Standard stimuli
Fz	14.22 ± 2.194	12.60 ± 1.383
F3	11.15 ± 1.571†	9.92 ± 1.198*
F4	11.69 ± 2.094†	10.32 ± 1.325

* p-valued < 0.05, Compared between deviant and frequent stimuli

† p-valued < 0.05, Compared between electrode sites of deviant stimuli

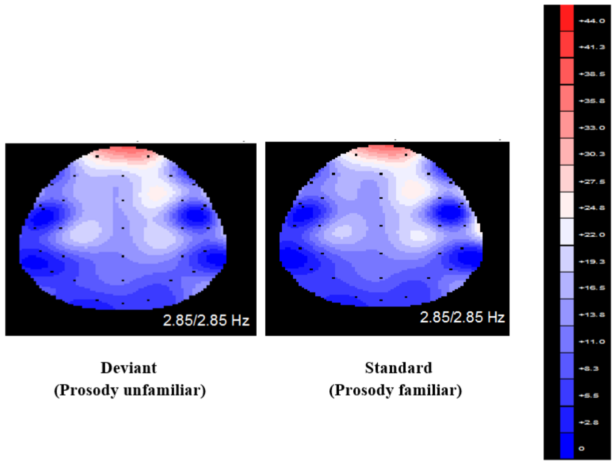

Figure 3. Scalp distribution of the theta wave spectrum (n=11) elicited by deviant and standard stimuli.

Table 4. Mean ± SEM of the power of delta wave at the Fz, F3 and F4 electrodes sites (n=11).

Electrodes	Mean ± SEM (μV^2)	
	Deviant stimuli	Standard stimuli
Fz	12.15± 2.312	11.33± 1.833
F3	9.59± 1.592†	8.56± 1.248
F4	10.60± 2.427	9.32± 1.692

†p-valued < 0.05, Compared between electrode sites of deviant stimuli

This study provides the brain electrical evidence, ERP component and EEG data in native Thai speakers that can be discriminated by the prosody with stress in Thai language at the pre-attentive stage. Further study should be done to obtain information of the change in ERP component in native speakers of stress language.

4. Conclusions

The prosody of unfamiliar stress in Thai word perception elicited the N2a component. As assessed by quantitative EEG analysis, the N2a was principally generated as theta and the delta waves in the frontal area. These data showed the role of the frontal area in the auditory pre-attentive processing of prosody with stress in Thai word perception in native Thai speakers. Further study needs to clarify whether frontal lobe lesions affect the sound discrimination at pre-attentive stage. In the future, the evaluation of prosody with stress in word perception may provide another valuable test to evaluate the frontal lobe function in patients with frontal lobe lesion.

Acknowledgements

This study was supported by the Siriraj Research Fund type 1 under Grant number (IO) R015631049. Moreover, I would like to express my deep appreciation to Assoc. Prof. Supatra Lohsiriwat for her great advice in manuscript preparation.

References

Başar, E., Başar-Eroğlu, C., Karakaş, S., and Schürmann, M. 1999. Are cognitive processes manifested in event-related gamma, alpha, theta and delta oscillations in the EEG?. Neuroscience Letters. 259(3), 165-8.

Belin, P., Fecteau, S., and Bédard, C. 2004. Thinking the voice: neural correlates of voice perception. Trends in Cognitive Sciences. 8(3), 129-35.

Chandrasekaran, B., Krishnan, A., and Gandour, J.T. 2007. Mismatch negativity to pitch contours is influenced by language experience. Brain Research. 1128(1), 148-56.

Fry, D.B. 1958. Experiments in the Perception of Stress. Language and Speech. 1(2), 126-152.

Fuentemilla, L., Marco-Pallarés, J., Münte, T.F., and Grau, C. 2008. Theta EEG oscillatory activity and auditory change detection. Brain Research. 1220, 93-101.

Honbolygó, F., Csépe, V., and Ragó, A. 2004. Suprasegmental speech cues are automatically processed by the human brain: a mismatch negativity study. Neuroscience Letters. 363(1), 84-8.

Hsiao, F.J., Wu, Z.A., Ho, L.T., and Lin, Y.Y. 2009. Theta oscillation during auditory change detection: An MEG study. Biological Psychology. 81(1), 58-66.

Javitt, D.C., Shelley, A., and Ritter, W. 2000. Associated deficits in mismatch negativity generation and tone matching in schizophrenia. Clinical Neurophysiology. 111(10), 1733-7.

Jemel, B., Achenbach, C., Müller, B.W., Röpcke, B., and Oades, R.D. 2002. Mismatch negativity results from bilateral asymmetric dipole sources in the frontal and temporal lobes. Brain Topography. 15(1), 13-27.

Kaan, E., Barkley, C.M., Bao, M., and Wayland, R. 2008. Thai lexical tone perception in native speakers of Thai, English and Mandarin Chinese: an event-related potentials training study. BMC Neuroscience. 9, 53.

Karanasiou, I.S. *et al.* 2010. Mismatch task conditions and error related ERPs. Behavioral and Brain Functions. 6, 14.

Kopp, B., Rist, F., and Mattler, U. 1996. N200 in the flanker task as a neurobehavioral tool for investigating executive control. Psychophysiology. 33(3), 282-94.

Kujala, T., Tervaniemi, M., and Schröger, E. 2007. The mismatch negativity in cognitive and clinical neuroscience: theoretical and methodological considerations. Biological Psychology. 74(1), 1-19.

Lakatos, P., Karmos, G., Mehta, A.D., Ulbert, I., and Schroeder, C.E. 2008. Entrainment of neuronal oscillations as a mechanism of attentional selection. Science. 320 (5872), 110-3.

Näätänen, R., Pakarinen, S., Rinne, T., and Takegata, R. 2004. The mismatch negativity (MMN): towards the optimal paradigm. Clinical Neurophysiology. 115(1), 140-4.

Oldfield, R.C. 1971. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia. 9(1), 97-113.

Patel, S.H. and Azzam, P.N. 2005. Characterization of N200 and P300: selected studies of the Event-Related Potential. International Journal of Medical Sciences. 2(4), 147-54.

Rinne, T. *et al.* 1999. Analysis of speech sounds is left-hemisphere predominant at 100-150ms after sound onset. Neuroreport. 10(5), 1113-7.

Sauseng, P., Griesmayr, B., Freunberger, R., and Klimesch, W. 2010. Control mechanisms in working memory: a possible function of EEG theta oscillations. Neuroscience and Biobehavioral Reviews. 34(7), 1015-22.

Terken, J. 1991. Fundamental frequency and perceived prominence of accented syllables. Journal of the Acoustical Society of America. 89(4 Pt 1), 1768-76.

Ylinen, S., Strelnikov, K., Huotilainen, M., and Näätänen, R. 2009. Effects of prosodic familiarity on the automatic processing of words in the human brain. International Journal of Psychophysiology. 73(3), 362-8.

Zora, H. 2011. Effects the suprasegmental features on the processing of the human brain: evidence from Mismatch negativity (MMN), Center for languages and literature, Lund university, Sweden, pp. 1-37.