

Auditory Pre-attentive Processing of Segmental and Suprasegmental (Speech Prosody) Phonological Units of Lexical Tones

Wichian Sittiprapaporn, Chittin Chindaduangratn and Naiphinich Kotchabhakdi

Clinical and Research Electroneurophysiology Laboratory
Neuro-Behavioural Biology Center
Institute of Science and Technology for Research and Development
Mahidol University
wichian_s@thailand.com

Abstract

Electrophysiological approach was used to examine effects on the amplitude and latency of segmental and suprasegmental (speech prosody) phonological units MMNs in response to consonant-vowel (CV) syllables in lexical tones. There was significant difference in amplitude and latency for MMNs between phonological units. Our result suggests that the MMN responses to the suprasegmental (speech prosody) phonological unit were greater than the segmental phonological unit. These data are relevance to a growing number of researches of MMN effects in response to changes of speech prosody in normal adults.

1. Introduction

The cortical mechanisms of speech processing are functionally asymmetric in the human brain: the left hemisphere is predominant in the perception and production of speech, whereas the right one is specialized in processing its prosodic and emotional component [1]. Numerous studies demonstrated that the left-hemispheric predominance in speech processing was represented already at a preattentive processing level to vowels [2-6] and consonant-vowel (CV) syllable [7-8]. These studies utilized mismatch negativity (MMN) as an index of preattentive processing of speech sounds. The MMN is an ERP component elicited by deviant stimuli sequences of repetitive auditory stimuli [9]. The MMN component appears as a frontocentrally negative wave usually peaking between 100 and 300 ms from the onset of stimulus deviation. The MMN, with its major source of activity in the supratemporal auditory cortex, can be used to investigate neural processing of speech and languages [9].

Recently, MMN has also been used to investigate the discrimination of speech and of complex nonspeech sounds [10] indicating that the right hemisphere is predominant in the perception of slow acoustic transitions, whereas neither hemisphere clearly dominates the discrimination of nonspeech sounds with fast acoustic transitions. This indicates that the perception of speech stimuli with similarly rapid acoustic transitions was dominated by the left hemisphere, which may be explained by the presence of acoustic templates (long-term memory traces) for speech sounds formed in this hemisphere [10]. However, the above-mentioned study utilized complex nonspeech sounds but did not address the issue whether the CV syllables with prosody are specific to the level of perception of suprasegmental phonological unit (speech prosody) change. To address this issue, we investigated the segmental and suprasegmental (speech prosody) phonological units in CV syllables by means of MMN paradigm.

2. Materials and methods

2.1. Subjects

Twenty right-handed, Thai subjects (13 females; aged 18-35 years) participated in the study. All subjects had normal hearing sensitivity and gave their written informed consent before participation in the study. This study has been approved by the Ethical Committee on Human Rights Related to Human Experimentation of Mahidol University, Thailand.

2.2. Language

Thai has five contrastive lexical tones [11], traditionally labeled mid (ˊ), low (ˋ), falling (ˊˋ), high (ˊˊ), and rising (ˊˋˊ): for example, /k^haa/ 'stuck' /k^haa/ 'galangal' /k^haa/ 'kill' /k^haa/ 'trade' /k^haa/ 'leg'. The midtone can be described phonologically as midlevel with a final drop, low tone as low-falling, falling tone as high-falling, high tone as high-rising, and rising tone as low-rising. The primary acoustic correlated of Thai tones is voice fundamental frequency [12].

2.3. Stimuli

Two pairs of speech stimuli of central Thai, each consisting of CV syllable was prepared to elicit MMN in response to segmental and supra-segmental (speech prosody) phonological units changes (In segmental condition: standard, with natural speech of /ti/; deviant with /pi/; in suprasegmental condition: standard, with low-pass filtered speech of /ti/; deviant with /pi/). Syllable rhymes in given pairs were not different. All stimuli were identical at their suprasegmental (i.e., tone) unit, which were always "mid" tone, thus eliminating any effects due to differences in frequency of occurrence of tones. In the filtered stimuli, the aim was to eliminate semantic and segmental phonetic (i.e., consonant, vowel) information while at the same time preserving suprasegmental (duration, loudness level, pitch contour) information [12]. Stimulus duration was unaffected by filtering; digital editing was used to equalize loudness levels between filtered and unfiltered stimuli. As judged by all subjects (not used in MMN study), none of the filtered stimuli were recognizable as Thai words. All stimuli were spoken by a native central Thai speaker with 500 ms in duration and digitally edited to have an equal peak energy level in dB SPL using the Cool Edit Pro v. 2.0 (Syntrillium Software Corporation). The sounds were presented binaurally via headphones (Telephonic TDH-39-P) at a comfortable listening level of ~85 dB (determined using a Brüel & Kjaer 2230 sound level meter).

2.4. Acoustic stimulation

(1) The natural speech sound /pi/ deviant (10%) was presented among the natural speech sound /ti/ standard (90%) and (2) the

low-pass filtered speech sound /pi/ deviant (10%) was presented among the low-pass filtered speech sound /ti/ standard (90%) in random order (except that each deviant stimulus was preceded by at least one standard stimulus). The stimuli were binaurally delivered at comfortable sound level through earphones. The inter-stimulus interval (ISI) was 1.25 second (offset-onset).

2.5. Electroencephalographic recording

Subjects were seated in an electrically and acoustically shielded chamber, instructed to read a book of their own choices and to ignore any auditory signals. During the auditory stimulation, electric activity of the subjects' brain was continuously recorded with 21 active electrodes positioned according to the 10-20 International System of Electro-cap and referred to linked mastoids with an electrode between Fz and Fpz connected to ground. A biologic Brain Atlas system amplified (Band-pass 0.01-100 Hz), analog-digital converted (128 samples/s/channel) and stored the data. Averaged responses were digitally filtered with a bandpass of 1-30 Hz.

2.6. EEG data processing

ERPs were obtained by averaging epoch, which started 100 ms before the stimulus onset and ended 900 ms thereafter; the $-100 - 0$ ms interval was used as a baseline. Epochs with voltage variation exceeding $\pm 100 \mu V$ at any EEG channels were rejected from further analysis. The MMN was obtained by subtracting the response to the standard from that to the deviant stimulus. For each experiment subject, the averaged MMN responses contained at least 125 accepted deviant trials. The latency mismatch responses were analyzed and defined as a moment of the peak global field power (GFP) with an epoch of 100-ms time window [13].

2.7. Statistical analysis

The statistical significance of MMN (deviant-minus-standard difference) was tested with one sample t -test. This was done by comparing the mean MMN amplitude against a hypothetical zero at the frontal (Fz) electrode site, where the MMN is most prominent.

3. Results

The grandmean standard, deviant and difference waveforms vs. average reference at the frontal (Fz) electrode site for segmental and suprasegmental (speech prosody) phonological units were shown. Grand averages were obtained for the standard, and the deviant, using linked mastoids as a reference.

3.1. ERP waveform analysis

Grand mean amplitudes and GFP for segmental and suprasegmental (speech prosody) phonological units were shown in Fig. 1 and Fig. 2 and the mean amplitudes, GFP and latency were also listed in Table 1.

Table 1: Mean amplitudes and latencies of MMN responses elicited by stimulus change in segmental and in suprasegmental (speech prosody) phonological units incorporating natural speech and low-pass filtered speech changes, respectively.

Stimulus Type	Mean Amplitudes	Mean GFP	Latency
Segmental	-2.16 (± 0.69)	1.32 (± 0.50)	156 (± 2.57)
Supra-segmental	-2.32 (± 0.39)	1.89 (± 0.27)	152 (± 2.77)

Note: Standard deviations of respective values are in brackets.

The t -test showed that the mean amplitude averaged over a 40 ms time window around the peak latency of MMN revealed a significant affects on condition (segmental condition: $t(10) = 2.16$; $p < 0.05$; suprasegmental condition: $t(10) = 2.32$; $p < 0.05$). The mean amplitude of segmental condition was $-2.16 \mu V$, S.D. = ± 8.69 whereas the suprasegmental condition was $-2.32 \mu V$, S.D. = ± 5.39 , respectively (Fig. 1). In addition, separate analyses of latency variance for each condition showed that the segmental condition was 156.86 (± 2.57) whereas the suprasegmental was 152.00 (± 2.77), respectively.

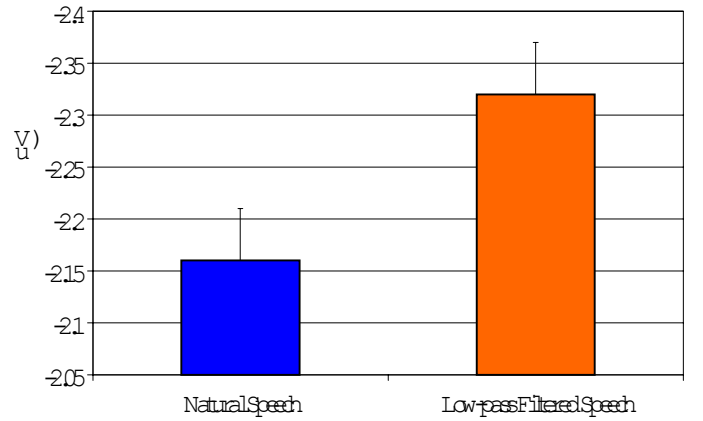


Figure 1. Mean amplitudes and standard deviation of MMN responses elicited by natural speech and low-pass filtered speech.

3.2. Global field power (GFP) analysis

There were significant effects on each condition (e.g., segmental and suprasegmental phonological units) in GFP amplitude values at the latency of the difference waveforms, showing amplitude variability across segmental and suprasegmental phonological units at the frontal (Fz) electrode site. These GFP amplitude values are corresponded to the mean voltage of the 40 ms intervals (so the peak plus minus 20 ms), centered at the corresponding peak latencies of the left and right frontal electrodes in the grand-mean difference waveforms. The t -test showed that the mean GFP averaged over a 40 ms time window around the peak latency of MMN revealed a significant affects on condition (segmental condition: $t(10) = 1.32$, S.D. = ± 0.50 ; $p < 0.05$; suprasegmental condition: $t(10) = 1.89$, S.D. = ± 0.27 ; $p < 0.05$) (Fig. 2 and Table 1).

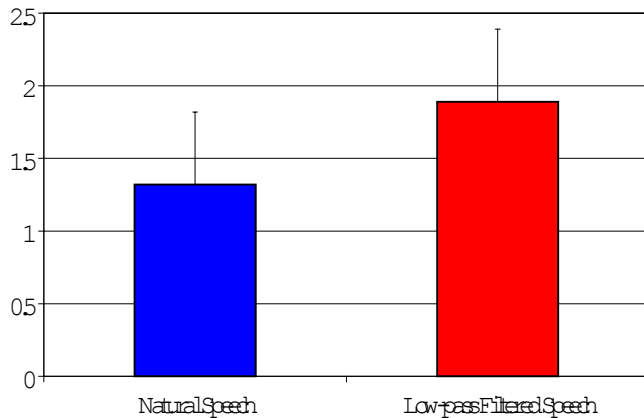


Figure 2. Mean GFP and standard deviation of MMN responses elicited by natural speech and low-pass filtered speech.

4. Discussion

The main finding of our study indicates that the MMNs elicited by suprasegmental (speech prosody) phonological unit changes were significantly larger than the segmental phonological unit changes. These results indicate that there was an early discrimination of suprasegmental (speech prosody) phonological unit carried out in speech perception. This predominant contribution was presumably obtained at the preattentive level of the suprasegmental phonological unit or speech prosody perception in CV syllables, indicating that the preattentive discrimination of speech prosody in CV syllables is based on the hemispheric neural networks. The present finding is thus in line with those of previous study using nonspeech complex stimuli incorporating fast and slow acoustic sounds [10] that reported the dipole moments of the MMNm elicited by nonspeech complex stimuli incorporating slow acoustic changes were significantly larger than the speech stimuli. This presumably implies that the brain extracts the speech-specific, probably phonetic features from speech sounds already at this early, preattentive stage of speech prosody processing. Additionally, the locus of the phoneme traces for speech prosody is probably indicated by the locus of MMN generation that is supported by the feature-specific loci of MMN generation in the auditory cortex. Our results could demonstrate the brain extracts of mismatch response to the speech prosody change perception in CV syllables, which has not been fully addressed in previous studies.

In relation to the hemispheric dominance in speech perception, our result was compatible with the early left anterior negativity (ELAN) recorded ERP component occurring between 100 and 300 ms. The ELAN has been observed for either the processing of phrase structure violations [14] or the processing of function words as compared to open class words [15] suggesting that the processing of word category information appears to be processed particularly early as compared to agreement and subcategorization information. Our result clearly supported the idea that the early speech sound change perception in CV syllable is rather independent of the participants' conscious expectancies and can therefore be claimed to be automatic in nature. In view of the rationale underlying our experiment paradigm, this data pattern leads to the conclusion that the process reflected in the ELAN component is fairly automatic in nature and the early anterior negativity reflects a highly automatic, early process during the phoneme of speech prosody perception.

However, our preliminary study did not investigate the hemispheric predominance in mismatch response to speech prosody in CV syllables. Further studies of hemispheric predominance of speech prosody in CV syllables are required to support a definite conclusion in whether the right hemispheric predominance of mismatch response to speech prosody is based on the across-

category change detection of the phoneme rather than merely reflecting the processing of language-related stimuli.

5. Conclusion

The present study was conducted to determine the perception of speech prosody in CV syllables. The present study indicated MMN responses to suprasegmental (speech prosody) phonological unit were larger than to segmental phonological unit emphasizing the role of brain extracts in the auditory preattentive speech prosody perception in CV syllables.

6. Acknowledgement

This study was supported by the University Development Communities (UDC) grant from Mahidol University, Thailand.

7. References

- [1] Kupfermann, I. 1991. Localization of higher cognitive and affective functions: The association cortices. In *Principles of neural science* (E. Kandel, J. Schwartz, and T. Jessel, Eds.), 823-838. Elsevier, New York.
- [2] Tervaniemi, M., Kujala, A., Alho, K., Virtanen, J., Ilmoniemi, R.J., and Näätänen, R. 1999. Functional specialization of the human auditory cortex in processing phonetic and musical sounds: A magnetoencephalographic (MEG) study. *Neuroimage* 9, 330-336.
- [3] Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huottilainen, M., Iivonen, A., Vainio, M., Alku, P., Ilmoniemi, R., Luuk, A., Allik, J., Sinkkonen, J., and Alho, K. 1997. Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature* 385, 432-434.
- [4] Rinne, T., Alho, K., Alku, P. 1999. Analysis of speech sounds is left-hemisphere predominant at 100-150 ms after sound onset. *NeuroReport* 10, 1113-1117.
- [5] Koyama, S., Gunji, A., Yabe, H., 2000. Hemispheric lateralization in an analysis of speech sounds: left hemispheric dominance replicated in Japanese subjects *Cogn Brain Res* 10, 119-124.
- [6] Kasai, K., Yamada, H., Kamio, S., Nakagome, K., Iwanami, A., Fukuda, M., Itoh, K., Koshida, I., Yumoto, M., Iramina, K., Kato, N., and Ueno, S. 2001. Brain lateralization for mismatch response to across- and within-category change of vowels. *Neuroreport* 12, 2467-2471.
- [7] Alho, K., Connolly, J.F., Cheour, M., Lehtokoski, A., Huottilainen, M., Virtanen, J., Aulanko, R., and Ilmoniemi, R.J. 1998. Hemispheric lateralization in preattentive processing of speech sounds. *Neurosci Lett* 258, 9-12.
- [8] Sittiprapaporn, W., Chindaduangratn, C., Tervaniemi, M., and Khotchabhakdi, N. 2003. Preattentive processing of lexical tone perception by the human brain as indexed by the mismatch negativity paradigm. *Ann N Y Acad Sci.* 999, 199-203.
- [9] Näätänen, R., Gaillard, A.W., and Mäntysalo, S. 1978. Early selective effect on evoked potential reinterpreted. *Acta Psychol.* 42, 313-329.
- [10] Shtyrov, Y., Kujala, T., Palva, S., Ilmoniemi, R.J., and Näätänen, R. 2000. Discrimination of Speech and of Complex Nonspeech Sounds of Different Temporal Structure in the Left and Right Cerebral Hemispheres. *NeuroImage.* 12, 657-663.

- [11] Gandour, J. 1975. On the representation of tone in Siamese. In J. G. Harris & J. R. Chamberlain (Eds.), *Studies in Thai linguistics in honor of William J. Gedney*, 170-195, Bangkok: Central Institute of English Language.
- [12] Gandour, J., Wong, D., and Hutchins, G. 1998. Pitch processing in the human brain is influenced by language experience. *NeuroReport*. 9, 2115-2119.
- [13] Lehmann, D. 1987. *Principles of spatial analysis*. Amsterdam: Elsevier, 309-354.
- [14] Friederic, AD. 1995. The time course of syntactic activation during language processing: A model based on neuro-psychological and neurophysiological data. *Brain and Lang.* 50, 259-281.
- [15] Nobre, A. and McCarthy, D. 1994 "Language-related ERPs: Scalp distribution and modulation by word type and semantic priming" *J Cog. Neurosci.* 6, 233-255.