2004 International Symposium on Nonlinear
Theory and its Applications (NOLTA2004)
Fukuoka, Japan, Nov. 29 - Dec. 3, 2004

Neuronal Correlates of Human Auditory Grouping: An ERP Study

Shin'ichiro Kanoh†, Ryoko Futami† and Nozomu Hoshimiya‡

†Graduate School of Engineering, Tohoku University, Aoba-yama 6-6-05, Sendai, 980-8579 Japan
‡Tohoku Gakuin University, 1-3-1 Tsuchitoi, Sendai, 980-8511 Japan
Email: kanoh@ecei.tohoku.ac.jp

Abstract—The human sequential grouping which organizes parts of tones into a group was examined by the mismatch negativity (MMN), a component of event-related potentials that reveals the sensory memory process. The sequential grouping is accomplished by the combinations of some factors, e.g. temporal and frequency proximity principles. Experimental results showed that the sequential grouping of presented tones was achieved on the auditory pre-attentive sensory memory process, and MMN amplitudes correlated to the temporal configurations of tones to be grouped. The investigation of MMN properties could reveal the nature of auditory sequential grouping.

1. Introduction

The auditory sensory memory is a memory system that stores auditory information within a short period (a few seconds). The basic information processing is applied to the stored information during the auditory sensory memory process [1].

Sequential grouping [2], which binds some successive information into one entity, is one of such a primitive processing. This function is one of the bases of higher-order auditory processing, like the segmentation and "cocktail-party effect" of speech sounds.

By the auditory psychophysical studies, it is found that the sequential grouping is accomplished by the combinations of some factors [3], e.g.

- Proximity: sounds that are near to one another in time (temporal proximity) or in frequency (frequency proximity) will group with one another.
- Good continuation: sounds with good continuation (e.g. frequency change) will tend to bind together.

But no physiological evidence to validate these factors on sequential grouping has been reported. Observing the neuronal correlate of sequential grouping will be a help to determine the principle of auditory information processing and to hypothesize the neuronal representation of temporal tone sequence, e.g. how the order and frequency of the successive tones are encoded.

In this study, the sequential grouping in human auditory cortex was examined by the mismatch negativity (MMN) [4], a component of event-related potentials (ERPs) that revealed sensory memory process. It was shown that the sequential grouping of successive tones was achieved during the auditory sensory memory process, and the properties of the sequential grouping were revealed by the MMN activities. These results will help to find out the cortical representation of auditory stimuli and will contribute to a computational theory of auditory information processing.

2. Working hypothesis

The mismatch negativity (MMN), one of the event-related potential (ERP) components, is a measure of pre-attentive sensory memory properties [4]. It is elicited when some of the frequently presented ("standard") stimuli are randomly replaced by the infrequent ("deviant") stimuli with some different attributes from the standard ones (oddball paradigm). MMN is generated by comparing the memory trace of the standard stimuli and the neuronal representation of incoming deviant stimulus.

Consider the experiment with oddball tone sequence in which each of the stimuli consists of several tones, and only the frequency of the last tone (deviated tone) is different between the standard and the deviant stimulus.

Let us assume that the neuronal activations of each grouped auditory stimuli would be bound together as a unitary one on the auditory cortex.

Incorporating this assumption, when the last tone is grouped with other tones, MMN will be elicited by comparing two unitary neuronal representations on the standard and the deviant stimuli, both of which encode the same set of grouped tones including the deviated tone.

In this case, when the number of tones in the grouped unitary entity was more, relative change of attributes in the corresponding neuronal representations on the standard and the deviant stimulus would be smaller.

The important property of MMN is that the MMN magnitude increases as a function of the extent of deviancy (difference between the attributes in the standard and the deviant stimuli) [5, 6, 7]. A relative difference between the neuronal representations of two sensory events in the sensory memory can be evaluated by MMN measurements.

Therefore, the present study was examined under the working hypothesis: "when the number of tones grouped together with the deviated tone increases, the elicited MMN magnitude will decrease."

In the present study, the two experiments were executed to determine whether the sequential grouping was achieved in the sensory memory process, and to find out what "rules" for sequential grouping are applied to various configurations of tone series.

3. Experiment 1: Temporal proximity factor on sequential grouping

At first, the sequential grouping of multiple tones by the temporal proximity factor was evaluated by the elicited MMN amplitude.

3.1. Method

Six male volunteers with normal hearing ability participated in the experiment as subjects. Auditory stimuli were applied to both ears by headphones. Each subject in an electro-magnetically shielded room was instructed to read a self-selected book and to ignore the presented auditory stimuli (reading condition). The EEG signal was recorded with Ag–AgCl electrodes at two locations (Fz, F4), and referred to linked earlobe (A1, A2) with a forehead ground.

Each stimulus consisted of several tone bursts (20 ms, 60 dB SL, rise/fall time 5 ms), i.e. the first tone, the last tone, and the inserted tone(s) which was (were) inserted between the first tone and the last tone. Total duration of the stimulus was 200 ms. The intervals between the last inserted tone and the last tone and between inserted tones were set to 20 ms. Inter-stimulus interval (onset to onset) was set to 600 ms. In the standard stimulus ($p = 0.9$), frequencies of the first, inserted and the last tones were 1000 Hz, 1000 Hz and 2000 Hz, respectively. And in the deviant stimulus ($p = 0.1$), only the frequency of the last tone was deviated to 1000 Hz.

The number of inserted tones (N_t) was set to 0, 1, 2 and 3, and was fixed within a single session. The relationship between N_t and elicited MMN by the infrequent change of the frequency of the last tone was observed. In this experiment, it is intended to evaluate the number of tones that are grouped with the last tone.

After the application of low-pass filter (30 Hz) to the averaged waveform, the amplitude of the negative peak in the difference waveform (deviant − standard) whose latency was 100 – 250 ms was evaluated. The data were statistically tested using one-way analysis of variance (ANOVA) with the factors of stimulus type ($p < 0.05$) to determine whether the ERPs associated with the standard stimuli were significantly different from those associated with the deviant stimuli at the MMN peak amplitude.

3.2. Results and Discussion

Figure 1 shows the difference waveforms of averaged responses to the standard and the deviant stimuli in Experiment 1 (Subject A, electrode Fz). From this subject, MMN was elicited when the number of inserted tones N_t was 0, 1 and 2, but no significant MMN was observed on $N_t = 3$.

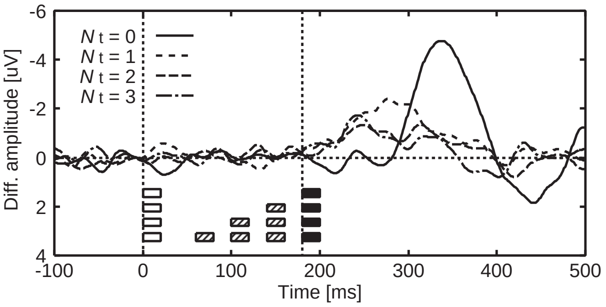

Figure 1: [Experiment 1] Averaged difference waveforms (deviant – standard). (Subject A, electrode Fz)

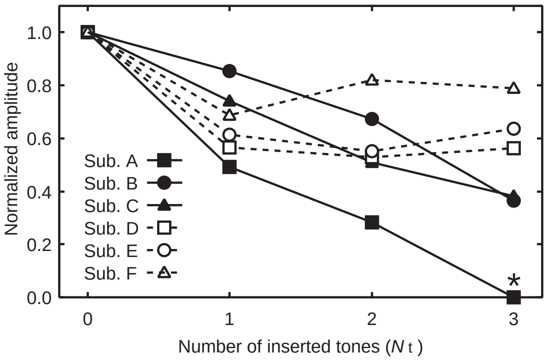

Figure 2: [Experiment 1] Relationship between N_t and normalized MMN amplitude on all subjects (electrode Fz). From Subject A, no significant MMN was observed on $N_t = 3$ (marked *).

The normalized amplitudes of the observed MMN on all subjects are shown in Figure 2. On all subjects, the elicited MMN amplitude taken from Fz was larger than that from F4. The properties of the observed MMN amplitudes could be divided into the following two groups:

- Group 1: the observed MMN amplitude was smaller for larger N_t (Subjects A, B, C).
- Group 2: the MMN amplitudes on $N_t = 0$ was the largest, and almost the same amplitudes were observed on $N_t = 1$, 2 and 3 (Subjects D, E, F).

This result could be explained on the assumption of the working hypothesis described in Section 2 as follows. On Group 1, the MMN amplitude decreased monotonically when N_t increased. This result could be explained by the hypothesis if all the inserted tones were grouped together with the last tone by the factor of temporal proximity during each of the experiments. And on Group 2, the largest MMN amplitude was elicited on $N_t = 0$, but no clear differences of MMN amplitudes were observed on $N_t \neq 0$. If the subjects in Group 2 tended to group only the last two tones (the last inserted tone and the last tone), such result would agree with the present hypothesis.

Therefore, the different phenomena on Group 1 and Group 2 can be explained by the different width of the time window for grouping based on temporal proximity: longer than 140 ms (Group 1), 60 ms to 80 ms (Group 2).

From Figure 1, it was also found that both peak and onset latencies were shorter when N_t was larger. And the duration of MMN was shorter when $N_t = 0$ (120 ms), but was longer when $N_t = 1, 2$ and 3 (200 ~ 250 ms). These results could be observed on the other subjects. Such an ERP component with shorter latency might be related to components other than the MMN.

4. Experiment 2: Frequency proximity factor on sequential grouping

Next, the sequential grouping of multiple tones by frequency proximity factor, and the effect of interplay with the other grouping factors were evaluated by the elicited MMN amplitude.

4.1. Method

Five male volunteers with normal hearing ability participate in the experiment as subjects. Auditory stimuli were applied to both ears by headphones.

Each stimulus consisted of the first tone (duration 40 ms), the second tone (20 ms) and the third tone (40 ms), each of which was 60 dB SL tone burst (rise/fall time 5 ms). Total duration of the stimulus was 200 ms. The interval between the second tone and the third tone was set to 30 ms. Inter-stimulus interval (onset to onset) was set to 600 ms. In the standard stimulus ($p = 0.9$), the frequencies of the first and the third tone (f_1, f_3) were 1000 Hz and 2000 Hz, respectively. In the deviant stimulus ($p = 0.1$), only the frequency of the third tone f_3 was deviated to 1800 Hz.

The frequency of the second tone (f_2) was set to 1000, 1200, 1500 and 2000 Hz, and was fixed within a single session. The relationship between f_2 and elicited MMN by the infrequent change of the frequency of the third tone was observed. Other experimental procedures were the same as in Experiment 1.

4.2. Results and Discussion

Figure 3 shows the difference waveforms of averaged responses to the standard and the deviant stimuli in Experiment 2 (Subject A, electrode Fz). From this subject, significant MMN was elicited on each condition.

The normalized amplitudes of the observed MMN for all subjects are shown in Figure 4. On all subjects, the elicited MMN amplitude taken from Fz was larger than that from F4. The elicited MMN amplitude was the largest when f_2 was 1000 Hz on all subjects, but the tendency of MMN amplitudes could be divided into the following two groups:

- Group 1: the smallest amplitude was observed when f_2 was 1500 Hz, and the amplitudes on $f_2 = 1200$ and 2000 Hz were almost the same (Subjects A, B).
- Group 2: the smallest MMN amplitude was observed when $f_2 = 1200$ Hz, but there was no significant difference between $f_2 = 1200$, 1500 and 2000 Hz (Subjects D, E, F).

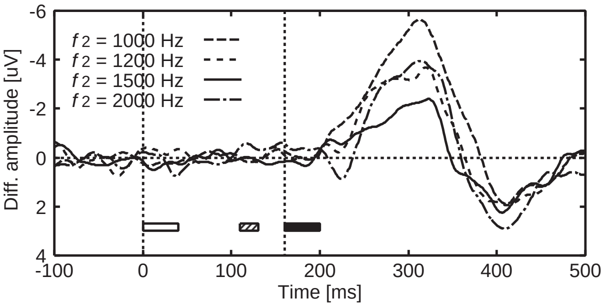

Figure 3: [Experiment 2] Averaged difference waveforms (deviant – standard) (Subject A, electrode Fz)

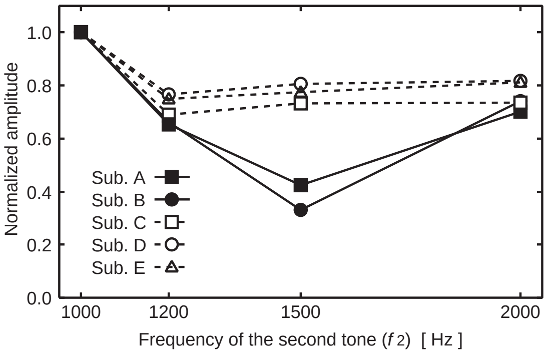

Figure 4: [Experiment 2] Relationship between f_2 and normalized MMN amplitude on all subjects. (electrode Fz)

Taking only the factor of frequency proximity into account, it was expected that the second tone tended to organize together with the first tone when $f_2 = f_1$ (1000 Hz). But when $f_2 = f_3$ (2000 Hz in the standard stimulus), the expected partner to be grouped with would be the third tone.

On the other hand, by considering only the factor of temporal proximity, the second tone tended to be grouped together with the third tone rather than the first tone in this case (intervals between the second and the first tone and between the second and the third tone were 70 ms and 30 ms, respectively).

In the former case, the two processes of sequential grouping based on temporal and frequency proximity factors contradicted each other, because the temporal position of the second tone was closer to the third tone than the first tone.

In such a situation, it might be possible that the two factors are combined (competition or cooperation) to solve the contradiction. By assuming the working hypothesis on this study, MMN amplitude is expected to decrease monotonically when f_2 increased from 1000 Hz to 2000 Hz, because the second tone tends to be grouped together with the third

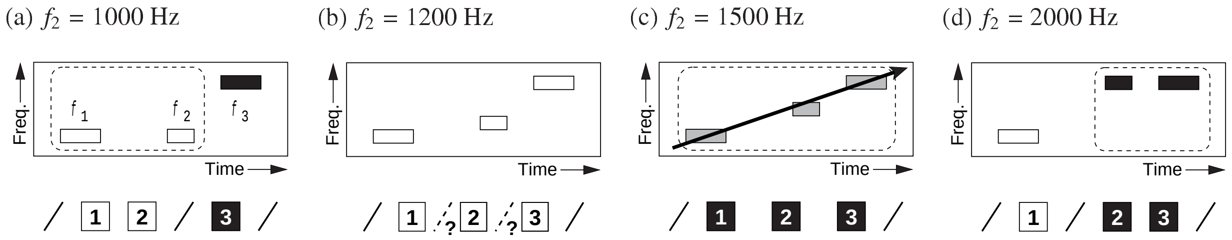
(a) f_2 = 1000 Hz (b) f_2 = 1200 Hz (c) f_2 = 1500 Hz (d) f_2 = 2000 Hz

Figure 5: [Experiment 2] The schematic view of the standard tone, and the expected combination of sequential grouping on each condition. See text in details. (“1”, “2” and “3” denote the first tone, the second tone and the third tone, respectively).

tone easier when f_2 approaches to the frequency of the third tone f_3.

But on the two subjects (Subjects A, B) out of five, the minimum MMN amplitude was observed when f_2 = 1500 Hz. This phenomenon could be explained by introducing the “good continuation” as one of the factors in addition to the factors of temporal/frequency proximity for sequential grouping of successive tones.

Figure 5 is the schematic diagram on the present hypothesis. When f_2 was 1000 Hz and 2000 Hz, the second tone tended to be grouped together with the first and the third tone, respectively. And when f_2 was 1500 Hz, the frequency change of the first tone (1000 Hz), the second tone and the third tone (2000 Hz) was roughly regular and continuous. On such a condition, incorporating the factor of good continuation, these three tones were organized together as a tone stream, or an unitary entity.

Therefore, when f_2 was 1000 Hz, 1500 Hz and 2000 Hz, the number of tones in one group including the third tone (deviated tone) was one, three and two, respectively. So the results on the subjects in Group 1 can be explained by assuming the working hypothesis on MMN amplitude. The results on the subjects in Group 2 are thought to be valid with the lack of the factor of good continuation.

From the experiments in the present study, no direct evidence on the neuronal process of the sequential grouping of auditory information was given. But these results could be explained by introducing the working hypothesis shown in Section 2. Similar experiments with variable temporal positions of the second tone will be needed for investigating the plausibility of the hypothesis. And by evaluating the frequency of the second tone in various temporal positions on which minimum MMN amplitude is observed, the physiological scale of tone frequency (c.f. the Mel scale) in the auditory cortex might be determined. The investigation of the neuronal or physiological basis of the present experiments is left for further study.

5. Conclusion

The neuronal process of sequential grouping of auditory information in human auditory sensory memory was investigated. In the two experiments, elicited MMN magnitudes were changed according to the configuration (the number of tones, frequency) of successive tone sequences. From this result, it was shown that the sequential grouping of successive tones was one of the processes of the auditory sensory memory. And it was also shown that the relationship between the configuration of tones and the elicited MMN magnitudes met the working hypothesis, on which the expected MMN magnitudes were assumed for temporal grouping factors (temporal/frequency proximities and good continuation) which was proposed by the auditory psychophysiology. This results showed that the factors (rules) of the grouping could be revealed by the mismatch negativity (MMN) component. These results will be helpful to investigate the cortical representation in the auditory system and the computational theory of auditory information processing.

This study was approved by the Ethics Committee on Clinical Investigation, Graduate School of Engineering, Tohoku University, and was performed in accordance with the policy of the Declaration of Helsinki.

References

[1] R. Näätänen, M. Tervaniemi, E. Sussman, P. Paavilainen, I. Winkler, “’Primitive intelligence’ in the auditory cortex,” *Trends in Neurosciences*, **24**, 5, pp.283–288, 2001.

[2] C. J. Darwin, “Auditory grouping,” *Trends in Cognitive Sciences*, **1**, 9, pp.327–333, 1997.

[3] A. S. Bregman, *Auditory scene analysis*, MIT Press, Cambridge, MA. 1990.

[4] R. Näätänen, *Attention and Brain Function*, Erlbaum, Hillsdale, NJ, 1992.

[5] H. Tiitinen, P. May, K. Reinikainen, R. Näätänen, “Attentive novelty detection in humans is governed by pre-attentive sensory memory,” *Nature*, **372**, pp.90–92, 1994.

[6] S. Kanoh, T. Itoh, K. Kazami, R. Futami, N. Hoshimiya, “A study on mismatch response to a change in tone duration by EEG and MEG,” *Japanese Journal of Medical Electronics and Biological Engineering* (in Japanese), **34**, 4, pp.128–136, 1996.

[7] S. Kanoh, T. Arai, R. Futami, N. Hoshimiya, “Properties of auditory temporal integration revealed by mismatch negativity,” Proceedings of the 23rd Annual International Conference of IEEE EMBS, 2.7.3.4, 2001.