Sequential Grouping of Tone Sequence as Reflected by the Mismatch Negativity

Shin'ichiro Kanoh, Ryoko Futami, Nozomu Hoshimiya

Graduate School of Engineering, Tohoku University
Aoba-yama 05, Sendai, 980–8579 Japan

E-mail: kanoh@ecei.tohoku.ac.jp

Abstract: The human sequential grouping which organized parts of tones into a group was examined by the mismatch negativity (MMN), a component of event–related potentials which revealed sensory memory process. The sequential grouping is accomplished by the competition and cooperation 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 sensory memory process, and MMN amplitudes correlated to the temporal configurations of tones to be grouped. It was also shown that the investigation of MMN properties could elicit the nature of auditory sequential grouping.

Keywords: auditory sensory memory, mismatch negativity (MMN), sequential grouping, auditory scene analysis

1 Introduction

In the process of human auditory information processing, it is generally thought that many of the presented information is handled not serially, but in parallel using "buffer" memory.

The auditory sensory memory is the memory system to store auditory information within a short period (a few seconds). In this system, the successive auditory information is integrated into meaningful entities and is represented as an unitary integrated event (temporal integration)[1, 2]. It has been shown by the authors that the temporal window for this integration process is 300 – 400 ms[3].

During the auditory sensory memory process, the primitive information processing is applied to the stored information[4]. Sequential grouping, which binds some successive information into one entity, is one of such primitive processing. Like the perceptual grouping in visual system, sequential grouping is accomplished by the competition and cooperation of some factors, e.g. temporal and frequency proximity factors[5, 6, 7].

In this study, the sequential grouping in human auditory cortex was examined by the mismatch negativity (MMN) [8], i.e. a component of event–related potentials (ERPs) which revealed sensory memory process. By the electroencephalogram (EEG) measurement on which the configuration of tone stimulus consisted of some tone bursts was changed, the relationship between the amplitude of elicited MMN and the expected grouping of these tones was investigated. 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 was revealed by the MMN activities. These results will help to find out the cortical representation in the auditory system and the computational theory of auditory information processing.

2 Background and working hypothesis

2.1 Mismatch negativity

The mismatch negativity (MMN), one of the event–related potential (ERP) components, is elicited when some of the frequent ("standard") stimuli are randomly replaced by the infrequent ("deviant") stimuli with some different attributes to the standard ones (oddball paradigm) [8].

MMN is a measure of pre–attentive sensory memory properties and is quite related to the auditory discrimination and storage functions[9, 10]. The important property for the present study is that the MMN magnitude is larger when the difference of the attribute in the deviant stimulus from that in the standard one is larger.

2.2 Sequential grouping

The sequential grouping, which organizes or binds some successive information into one entity, is one of such a primitive processing. This function in auditory modality is one of the basis on segmentation of speech sounds.

Like the perceptual grouping in visual system, sequential grouping is accomplished by the combination of various factors[5, 6, 7]. Some of the factors are as follows.

- Proximity: sounds which 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.

2.3 Working hypothesis

Consider the experiment with oddball tone sequence in which each of the stimulus consists of several tones, and only the frequency of the last tone (deviated tone) is different in standard and deviant stimulus. When the sequential grouping is occurred together with the deviated tone during the experiment, MMN will be elicited by comparing two unitary neuronal representations on standard and 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 standard and deviant stimulus would be smaller.

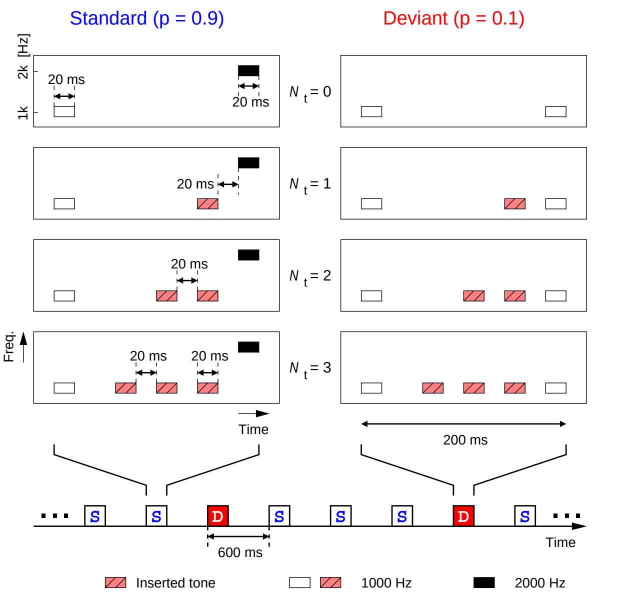


Figure 1. [Experiment 1] Schematic diagram of stimulus train on each condition.

As described in Section 2.1, MMN magnitude is smaller by less different attribute in standard and deviant stimulus.

So, the present study was examined under the working hypothesis: "when the number of tones grouped together with the deviated tone is larger, the elicited MMN magnitude will be smaller."

3 Experiment 1: The number of tones and sequential grouping

At first, the relationship between the number of tones and the elicited MMN amplitude was examined.

In this experiment, oddball sequence in which each stimulus consisted of several tones (the first tone, inserted tone(s) and the last tone) was used. The relationship between the number of inserted tones and elicited MMN amplitude was examined, and the tendency of sequential grouping by the factor of temporal proximity was discussed.

3.1 Method

Six male healthy volunteers with normal hearing ability took part in the experiment as subjects. Auditory stimuli were applied to the both ears by headphones. Each subject in an electro–magnetically shielded room was required to read a self–selected book and was instructed to ignore the presented auditory stimuli (reading condition).

The schematic diagram of stimulus train is shown in Figure 1. Each stimulus consisted of several tone bursts (20 ms, 60 dB SL, rise/fall time 5 ms): 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, and the intervals between the last inserted tone and the last tone and between inserted tones were set to 20 ms.

In standard stimulus ($p = 0.9$), frequencies of the first, inserted and the last tones were 1000 Hz, 1000 Hz, 2000 Hz, respectively. And in deviant stimulus ($p = 0.1$), only the frequency of the last tone was changed to 1000 Hz. Inter–stimulus interval (onset to onset) was set to 600 ms.

The number of inserted tones (N_t) was set to 0, 1, 2, 3, and the relationship between N_t and elicited MMN by the infrequent change of the frequency of the last tone was observed. N_t was fixed within a single session.

The EEG was recorded with Ag–AgCl electrodes from two channels (Fz, F4). The vertical electro–oculogram (EOG) was also recorded from the outer canthus of the left eye. Two electrodes attached on A1 and A2 were short–circuited and were taken as the reference. EEG was amplified (gain $\times 1000$) and recorded with a band-pass filter setting of 0.1 – 100 Hz and analyzed from 100 ms before to 500 ms after the onset of the first tone (sampling rate 1000 Hz with anti–aliasing filter of 200 Hz).

EEG epochs in which recorded amplitudes exceeded $\pm 100\mu$V were automatically rejected. The responses to the standard stimuli just after the deviant ones were excluded from the analysis. The accepted EEG epochs were averaged.

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 analyzed using one–way analysis of variance (ANOVA) for repeated measures with factors stimulus type (standard and deviant, $p < 0.05$).

3.2 Results

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

The normalized amplitudes of the observed MMN for all subjects are shown in Figure 3. 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 became smaller for larger N_t (Subjects A, B, C).
- Group 2: the MMN amplitudes on N_t =1, 2, 3 were almost the same (Subjects D, E, F).

3.3 Discussion

Although the interval between the last inserted tone and the last tone was fixed to 20 ms, the dependency of MMN amplitude to the number of inserted tones (N_t) was observed. This result could be explained on the assumption of the working hypothesis described in Section 2.3 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 and the last tone were grouped together by the factor of temporal proximity during the experiment (Figure 4(a)). And on

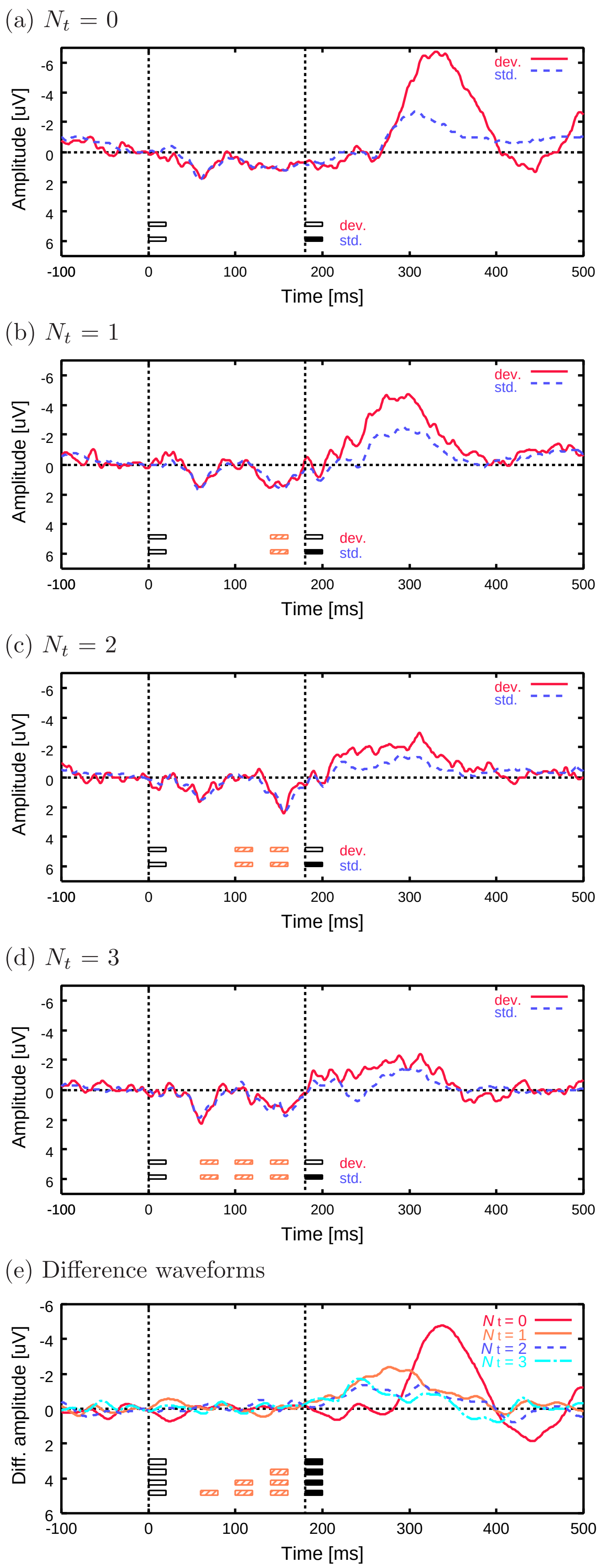


Figure 2. [Experiment 1] Averaged standard (dashed) and deviant (solid) waveforms (a–d), and their difference waveforms after low–pass filtering of 30 Hz (e). (Subject A, electrode Fz)

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 two tones, i.e. last inserted tone and the last tone, such result would agree with the present hypothesis (Figure 4(b)).

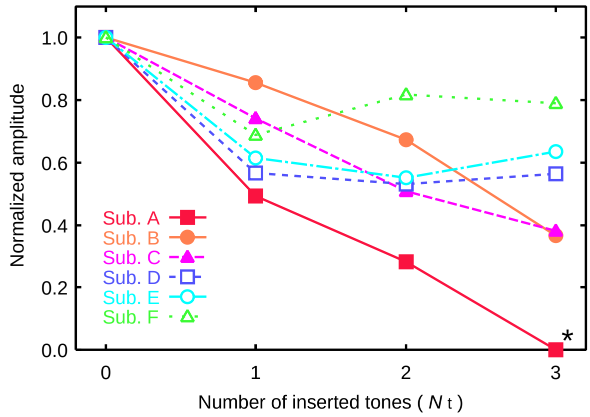


Figure 3. [Experiment 1] Relationship between N_t and normalized MMN amplitude on all subjects (electrode Fz). From Subject A, no significant MMN was observed on $N_t = 3$ (marked *).

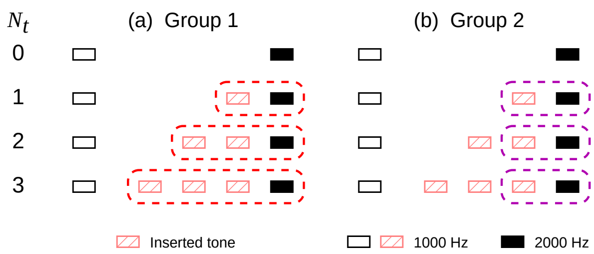


Figure 4. [Experiment 1] Schematic diagram on possible hypothesis. Temporal configuration of standard stimulus on each condition is shown.

From Figure 2, it was found from Subject A, that the latency (peak latency and starting time) was shorter when N_t was larger. And the period of MMN was shorter (120 ms) when $N_t = 0$, but the longer MMN period was observed on $N_t = 1, 2, 3$ (200 – 250 ms). These results could be observed from the other subjects. Such a ERP component with shorter latency might be related to other components other than the MMN.

4 Experiment 2: Frequency of tones and sequential grouping

Next, the relationship between the tone frequency and sequential grouping was evaluated by the elicited MMN amplitude.

In this experiment, the oddball sequence in which each stimulus consisted of three tones (the first tone, inserted tone and the last tone) was used, and the relationship between frequency of the inserted tone and elicited MMN amplitude was examined.

As described in Section 2.2, temporal and frequency proximity are known as the factor of proximity for auditory sequential grouping. By changing the frequency of the inserted tone, it was intended to control the tendency of sequential grouping, i.e. with which the inserted tone was tended to be grouped.

4.1 Method

Five male healthy volunteers with normal hearing ability took part in the experiment as subjects. Auditory stimuli were applied to the both ears by headphones. Each subject was required to read a self–selected book and was instructed to ignore the presented auditory stimuli (reading condition) in an electro–magnetically shielded room.

The schematic diagram of stimulus train is shown in Figure 5. Each stimulus consisted of the first tone (duration 40 ms), inserted tone (20 ms) and the last 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, and the interval between inserted tone and the last tone was set to 30 ms.

The frequencies of the first and the last tone in standard stimulus ($p = 0.9$) were 1000 Hz, 2000 Hz, respectively. In deviant stimulus ($p = 0.1$), only the frequency of the last tone was deviated to 1800 Hz. Inter–stimulus interval (onset to onset) was set to 600 ms.

The frequency of of inserted tone (f_i) was set to 1000, 1200, 1500, 2000 Hz, and the relationship between f_i and elicited MMN by the infrequent change of the frequency of the last tone was observed. f_i was fixed within a single session. The procedure of EEG recording and analysis was the same as in Experiment 1.

4.2 Results

Figure 6 shows the averaged responses for standard and deviant stimuli and their difference waveforms in Experiment 2 (Subject A, electrode Fz). From this subject, significant MMN was elicited on each condition of inserted tone f_i.

The normalized amplitudes of the observed MMN for all subjects are shown in Figure 7. 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_i was 1000 Hz on all subjects, but the tendency could be divided into the following two groups:

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

4.3 Discussion

In Experiment 2, the frequency of the inserted tone f_i was changed to control the contribution of the frequency proximity factor to sequence grouping. Taking only the factor of frequency proximity into account, it was expected that the inserted tone tended to organize together with the first tone when f_i was 1000 Hz (the frequency of the first tone). But when f_i was 2000 Hz (the frequency of the last tone in standard stimulus), the expected partner to be grouped with would be the last tone.

In the former case, the two processes of sequential grouping based on temporal and frequency proximity factors contradicted in each other, because the temporal

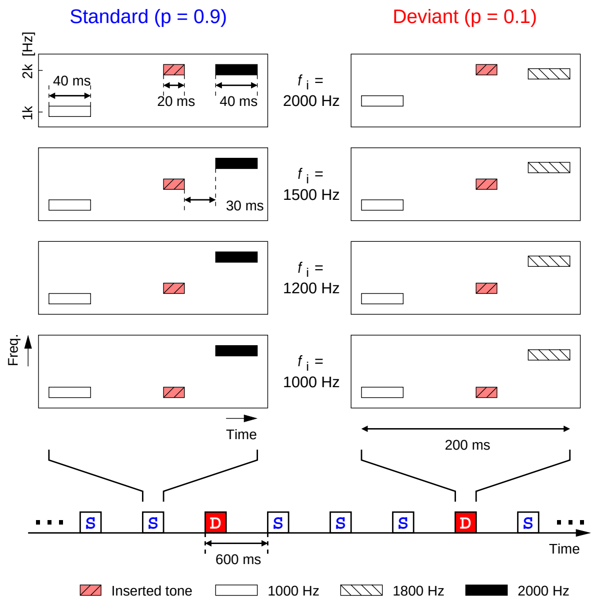


Figure 5. [Experiment 2] Schematic diagram of stimulus train on each condition.

position of the inserted tone was closer to the last tone than the first tone (intervals between inserted tone and the first, the last tone were 70 ms, 30 ms, respectively). Thinking only the factor of temporal proximity, inserted tone tended to be grouped together with the last tone rather than the first tone in this case.

In such a case, it seems to 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 be decreased monotonically when f_i increased from 1000 Hz to 2000 Hz, because inserted tone tends to be grouped together with the last tone easier when f_i approaches to the frequency of the last tone.

But on the two subjects (Subjects A, B) out of five, the minimum MMN amplitude was observed when $f_i = 1500$ 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.

Table 1 is the schematic diagram on the present hypothesis. When f_i was 1000 Hz and 2000 Hz, inserted tone tended to be grouped together with the first and the last tone, respectively. And when f_i was 1500 Hz, the frequency change of the first tone (1000 Hz), inserted tone and the last tone (2000 Hz) was roughly regular and continuous. In such a case, assuming the factor of good continuation, these three tones were organized together as a tone stream, or an unitary entity.

Therefore, when f_i was 1000 Hz, 1500 Hz, 2000 Hz, the number of tones in one group including the last 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

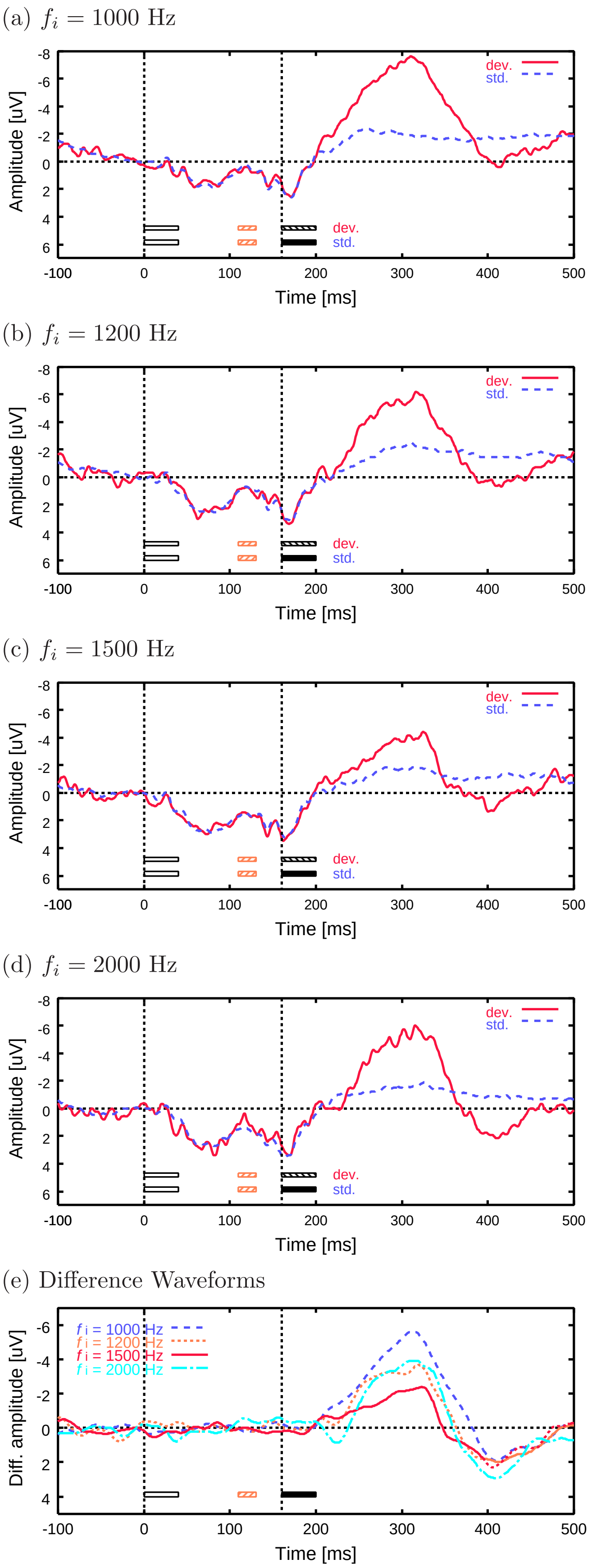


Figure 6. [Experiment 2] Averaged standard (dashed) and deviant (solid) waveforms (a–d), and their difference waveforms after low–pass filtering of 30 Hz (e). (Subject A, electrode Fz)

could be explained by introducing the working hypothesis shown in Section 2.3. Similar experiments with variable temporal positions of inserted tone will be needed

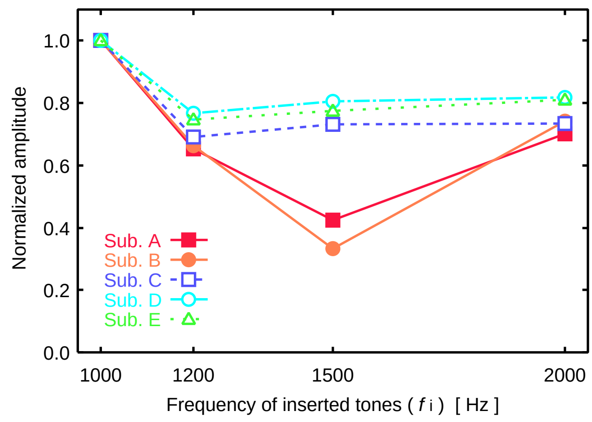


Figure 7. [Experiment 2] Relationship between f_i and normalized MMN amplitude on all subjects. (electrode Fz)

for investigating the plausibility of the hypothesis. And by evaluating the frequency of inserted tone in various temporal positions on which minimum MMN amplitude is observed, the physiological scale of tone frequency (c.f. the Mel scale[(11)]) in the auditory cortex might be determined. The investigation of neuronal or physiological basis of the present experiments is left for the further study.

5 Conclusion

The neuronal process of sequential grouping of auditory information in human auditory sensory memory was investigated. It was shown that the sequential grouping of successive tones was one of the processes of the auditory sensory memory, and 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.

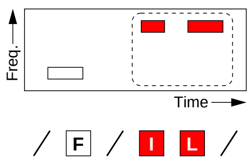
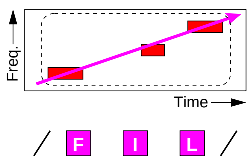
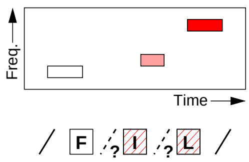
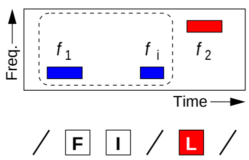
References

(1) M. Tervaniemi, J. Saarinen, P. Paavilainen, N. Danilova & R. Näätänen, “Temporal integration of auditory information in sensory memory as reflected by the mismatch negativity,” *Biological Psychology*, **38**, pp.157–167 (1994).

(2) H. Yabe, M. Tervaniemi, K. Reinikainen & R. Näätänen, “Temporal window of integration revealed by MMN to sound omission,” *NeuroReport*, **8**, pp.1971–1974 (1997).

(3) S. Kanoh, T. Arai, R. Futami & N. Hoshimiya, “Properties of auditory temporal integration revealed by mismatch negativity,” Proceedings of the 23th Annual International Conference of IEEE EMBS, 2.7.3.4. (2001).

Table 1. [Experiment 2] Schematic diagram on possible hypothesis. The three abstract relative degrees of factors for grouping inserted tone together with the last tone: temporal proximity ("Temp."), frequency proximity ("Freq."), good continuation ("Cont."), the expected number of tones in the group including the last tone ("Number of Grouped Tones"), and the expected MMN amplitude on the assumption of the working hypothesis ("MMN Amplitude") were shown for each stimulus condition. f_f, f_i and f_l denote the frequencies of the first tone, inserted tone and the last tone, respectively. The expected combination of sequential grouping was also shown under each schematic view of the standard tone ("F", "I" and "L" denote the first tone, inserted tone and the last tone, respectively).

f_f	f_i	f_l	Schematic View of Standard Stimulus	Factors for Grouping			Number of Grouped Tones	MMN Amplitude
				Temp.	Freq.	Cont.		
1000 Hz	2000 Hz	2000 Hz (Std.) 1800 Hz (Dev.)		+	++	+	2	++
	1500 Hz			+	+	+++	3	+
	1200 Hz			+	−	++	1∼2	+++
	1000 Hz			+	−−	+	1	++++

(4) 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).

(5) A. S. Bregman, *Auditory scene analysis*, MIT Press, Cambridge, MA (1990).

(6) B. C. J. Moore, *An Introduction to the Psychology of Hearing*, 3rd Ed., Academic Press, London (1989).

(7) D. Deutsch (Ed.), *The Psychology of Music*, Academic Press, New York, NY (1982).

(8) R. Näätänen, *Attention and Brain Function*, Erlbaum, Hillsdale, NJ (1992).

(9) 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).

(10) 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).

(11) S. S. Stevens & J. Volkmann, "The relation of pitch to frequency: a revised scale," *American Journal of Psychology*, **53**, pp.329–353 (1940).