Auditory imagery mismatch negativity elicited in musicians

Masato Yumoto,[1,CA] Maki Matsuda,[4] Kenji Itoh,[2] Akira Uno,[5] Shotaro Karino,[3] Osamu Saitoh,[6] Yuu Kaneko,[7] Yutaka Yatomi[1] and Kimitaka Kaga[3]

[1]Departments of Laboratory Medicine; [2]Speech and Cognitive Science; [3]Otolaryngology, Faculty of Medicine, University of Tokyo, Tokyo; [4]Department of Musicology, Graduate School of Music, Tokyo National University of Fine Arts and Music, Tokyo; [5]Graduate School of Comprehensive Sciences, University of Tsukuba, Tsukuba; [6]Departments of Psychiatry; [7]Neurosurgery, National Center of Neurology and Psychiatry, Tokyo, Japan

[CA]Corresponding Author: yumoto-tky@umin.ac.jp

A mismatch between auditory sensation and expectant imagery of syllables elicited a possible equivalent of mismatch negativity in a previous study. The purpose of this study was to verify whether auditory imagery from musical notation could also mediate such imagery-based mismatch negativity. Neuromagnetic recording was obtained from eight musicians, who were instructed to identify unpredictably occurring pitch mismatches between a random tone sequence and a visually presented musical score. The difference between incongruent and congruent responses showed a magnetic distribution consistent with two frontal-negative current dipoles bilaterally located in the vicinity of Heschl's gyrus, peaking at approximately 150 ms in latency. This imagery-based mismatch negativity may represent an early neural process of deviance detection between the sensory input and expectant imagery. *NeuroReport* 16:1175–1178 © 2005 Lippincott Williams & Wilkins.

Key words: Absolute pitch; Audiovisual associative correlation; Auditory cortex; Auditory imagery; Human; Magnetoencephalography; Mismatch negativity; Music reading; Musical score; Musical syntax

INTRODUCTION

Infrequent ('deviant') sounds occurring in a sequence of repetitive frequent ('standard') sounds elicit an event-related brain response called the mismatch negativity (MMN) even in the absence of attention to the sound sequence. MMN represents a neural process of mismatch detection between the deviant auditory input and a sensory memory trace developed by the standard stimuli [1,2]. MMN has been so firmly established in human neurophysiology that researchers' interest extends to a new paradigm for efficient measurement [3,4] and functional evaluation of auditory sensory memory for clinical application [5,6]. At the same time, a recently emerging line of research suggests that a sensory memory trace developed by stimulus repetition or regularities [7] in the auditory past may not be necessarily essential for MMN elicitation.

In an event-related potential study, Widmann *et al.* [8] demonstrated that audiovisual incongruence infrequently occurred between an auditory two-tone sequence and a visual two-symbol sequence elicited 'MMN-like' activity after training the participants to learn the following rule: the higher and lower tones in pitch are predicted by the higher and lower symbols in a vertical position, respectively. They claimed that a very strong relationship between the visual and the corresponding auditory stimuli mediated either by the brief training or by very simple rules was essential for this kind of audiovisual incongruence responses. On the other hand, a similar audiovisual incongruence response was also detected in a symbol-to-sound mismatch task mimicking a real life situation of reading [9]. In the task, a text of nonsense syllables was presented as a predictor of a simultaneously presented sequence of spoken syllables reading out the text with infrequent errors. The participants were instructed to imagine upcoming auditory syllables by referring to the visual text and to pay attention to the unpredictable occurrence of audiovisual mismatches. A difference in magnetoencephalographic (MEG) responses to matched and mismatched syllables showed a spatiotemporal profile indistinguishable from that of MMN in the superior temporal plane.

Recent neuroimaging studies suggested that auditory perception and imagery share common neural substrates, including the primary auditory cortex [10,11]. These findings led to the hypothesis that memory resources for auditory sensation and imagery may also be shared and the imagery memory trace can function as a template for MMN elicitation. The audiovisual experiments described in the previous paragraph suggest that a cross-modal associative correlation, which is strong enough to provoke auditory imagery from visual symbols, is an important factor to elicit this type of MMN. In contrast to well established sensory MMN (sMMN), which is a product of dissonance with the sensory memory trace, the variant which is produced by dissonance with 'expectant imagery' [12] is hereafter referred to as 'imagery MMN' (iMMN) in this paper.

As syllabograms have auditory and visual modalities, so do musical notes on a staff for trained musicians [13]. When we read a book silently, our eyes travel line by line over a page to convert visual symbols into auditory information, which is not vocalized, but is merely sounded out in our

minds [14]. Such auditory imagery is also provoked by music reading in well trained musicians [15,16]. The purpose of the present study was to verify whether a nonlinguistic indexical link, such as that between a musical score and corresponding tones, also produces iMMN in trained musicians. The experimental paradigm used in the present study was adopted from the previous linguistic version of our study [9]. To focus on auditory imagery predominantly from musical notation, a background musical context, which may drive participants' expectancies toward musical intuitions (i.e. what the next note should be on a theoretical or experiential basis of tonal music) [17], was minimized in the present study by applying random tone series in a chromatic scale as stimuli.

MATERIALS AND METHODS

Eight right-handed musical experts (age: 22–45 years; six women and two men), who were recruited from Tokyo National University of Fine Arts and Music (Tokyo, Japan) as paid volunteers, participated in this study. All the participants were self-reported possessors of absolute pitch, with no history of neurological or audiological disorders. All of them gave written, informed consent before the experiments, and the procedures used in this study were approved by the Ethics Committee of the University of Tokyo.

The auditory stimuli consisted of 12 oboe tones, which were sampled from a software synthesizer VSC-MP1 (Roland, Osaka, Japan), on the notes from C_4 (262 Hz) to B_4 (494 Hz) in semitone steps. Using the sound editor Audition (Adobe Systems, San Jose, California, USA), the duration of each tone was adjusted to 900 ms including 10 and 100 ms linear rising and falling, respectively. One out of the 12 tones was presented per second with equal probability, in random order. The study consisted of 10 blocks with intermissions of 15 s in between. At the beginning of each block, the whole sequence of tones in the block was projected onto a screen in front of the participants as a musical score of 112 notes on seven consecutive staves (Fig. 1). The auditory sequence started 3 s after presentation of the score in each block. Auditory and visual sequences were matched, except for audiovisual pitch mismatches, which were implanted to occur at random with a probability of 17%. Mismatched notes in visual and auditory modalities were randomly chosen from the 12 notes with equal probability, with the stipulation that a pitch interval of audiovisual mismatches should be a whole (diatonic) tone step.

The participants were instructed to imagine the upcoming tone by referring to a visually presented note sequence and to pay attention to the infrequent occurrence of audiovisual mismatches. The auditory stimuli were sequenced by the STIM2 system (Compumedics Neuroscan, El Paso, Texas, USA) and were delivered binaurally to the participant's ears at 60 dB sound pressure level through ER-3A earphones (Etymotic Research, Elk Grove Village, Illinois, USA) at a stimulus onset asynchrony of 1 s. The MEG signals were recorded in a magnetically shielded room using VectorView (Elekta Neuromag, Helsinki, Finland), which has 204 first-order planar gradiometers at 102 measuring sites on a helmet-shaped surface that covers the entire scalp. Auditory stimulus-triggered epochs of 450 ms duration (including a 50-ms prestimulus baseline) were filtered online with a band-pass of 1–200 Hz and recorded at a sampling rate of
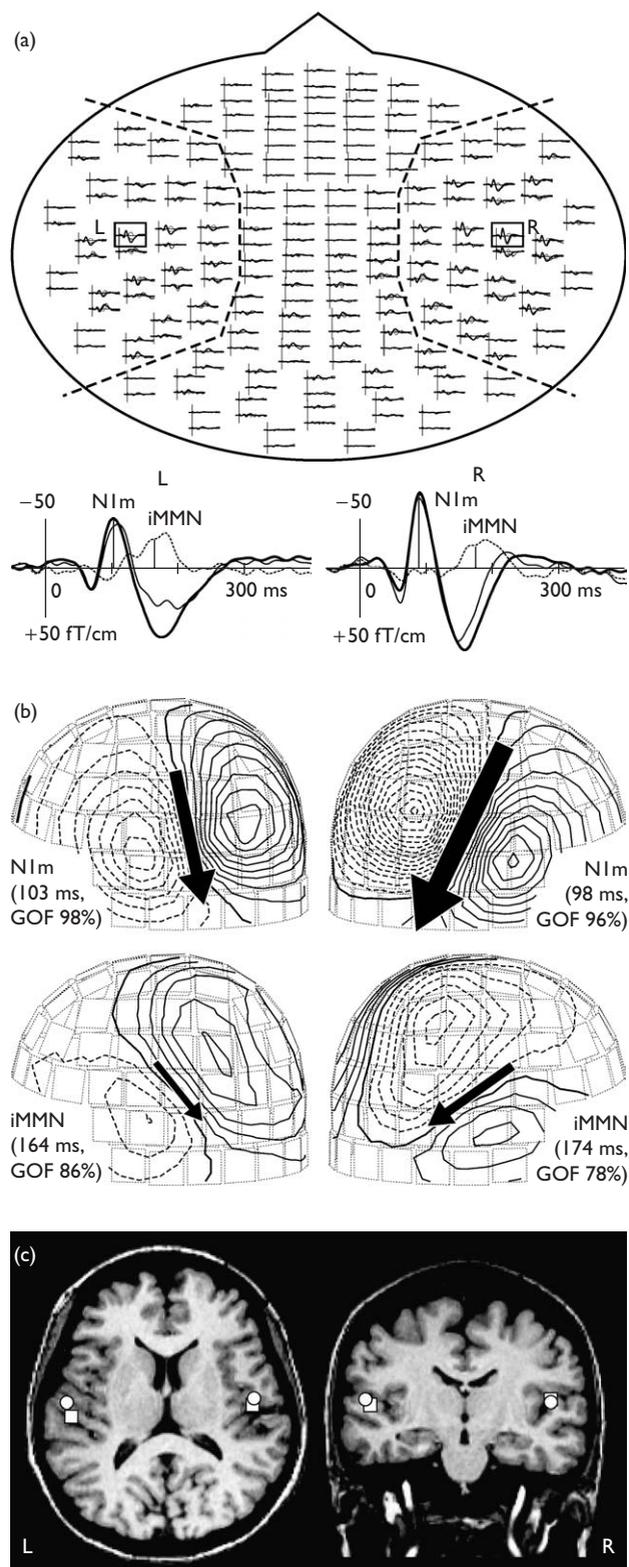

Fig. I. A musical score, which is read from left to right and from top to bottom, was projected onto a screen and the image was centered in each participant's visual field, so that notes on the seven staves spanned approximately $\pm 13°$ and $\pm 10°$ in horizontal and vertical visual angles, respectively. A participant's vision was corrected with nonmagnetic glasses, if necessary.

600 Hz. Epochs with artifacts exceeding 3 pT/cm in any MEG channel or with an electrooculogram (EOG) response exceeding 150 μV were discarded. EOG was also used to monitor the participant's performance in note tracking along the auditory stream during measurement. The MEG responses to auditory stimuli, both matched and mismatched with the visual sequence, were selectively averaged together for analysis. After MEG measurement, audiovisual mismatch detection ability of each participant was assessed by conducting a behavioral test, in which the participants were instructed to mark the notes on the score that did not match the auditory sequence, using another preliminary block of the study.

To isolate the activity that occurred when an audiovisual mismatch was observed, the difference in waveforms was calculated by subtracting the averaged responses to matched tones from the averaged responses to mismatched tones. The averaged and difference waveforms were filtered offline with a low-pass at 40 Hz, and the baseline for the waveforms was defined as the mean amplitude between -50 and 0 ms relative to tone onset. The peak latency of the main component (N1m) to matched tones was determined for each hemisphere by the time point at which the root mean square (RMS) of the predefined perisylvian channels (Fig. 2a) reached the maximum between 70 and 140 ms after the auditory stimulus onset. The latency of the earliest prominent peak (iMMN) in the difference waveforms was determined by the same procedure, except for the time window of 0–200 ms. Paired *t*-tests were performed on the mean perisylvian RMS values of the responses to matched and mismatched tones within the time window of 20 ms centered at the peak latency of iMMN.

The sources of each component were modeled separately as a single equivalent current dipole (ECD) for each hemisphere. The ECDs were calculated at the peak latencies from the same perisylvian channels, independently for N1m and iMMN in each participant. The estimated ECDs were described in a head-based coordinate system. With a positive direction to the right, the *x*-axis passes through the two preauricular points that are digitized before data acquisition. The *y*-axis passes through the nasion and is

normal to the x-axis. The z-axis points upward according to the right-handed rule and is normal to the xy-plane. For coregistration of MEG data with the individual anatomy, magnetic resonance imaging scans were obtained from all the participants. The ECD parameters for N1m and iMMN were compared by paired t-tests.

RESULTS

The preliminary behavioral test of audiovisual mismatch detection revealed 100% accuracy in every participant. EOG data showed synchronized tracking of notes with the auditory sequence during measurement in all the participants. Every participant showed the N1m components peaking at latencies of approximately 100 ms within perisylvian areas (Fig. 2a). Difference waveforms, which were calculated by the subtraction of congruent from incongruent responses, showed prominent deflections (iMMN) peaking at latencies of approximately 150 ms within the perisylvian areas in both hemispheres. None of the participants showed the RMS peak at the ends of the analysis windows (i.e. 70 and 140 ms for N1m; 0 and 200 ms for iMMN), validating the setting of the windows for N1m and iMMN. The mean RMS amplitude values of matched and mismatched responses showed a significant difference in each hemisphere ($p < 0.01$). The magnetic distribution of the iMMN component across channels in each hemisphere was similar to that of the N1m component in congruent conditions (Fig. 2b). The ECDs for iMMN were localized in the vicinity of Heschl's gyrus on the superior temporal plane with a goodness of fit (GOF) above 70% in all the participants. The ECD parameters for N1m and iMMN reached a statistical significance ($p < 0.05$) only for the latency (left: $p < 0.002$; right: $p < 0.001$), posteroanterior ECD location (left: $p < 0.005$; right: $p = 0.95$) and the GOF (left: $p < 0.005$; right: $p < 0.005$) (Fig. 2c, Table 1).

DISCUSSION

The notation system of Western music was utilized as a visual predictor of auditory pitch sequence in this non-linguistic version of the symbol-to-sound mismatch task. Contrary to a large majority of neurofunctional studies on music perception, musical intuitions mediated by syntax on the basis of tonality and metrical dynamics [17] were minimized as potential contaminators of auditory imagery predominantly from visual prediction, by applying stimuli of a random tone series with constant tempo. The random occurrence of audiovisual mismatches was counterbalanced across 12 notes in an octave with equal probability. A difference in waveforms detected in the present study was, therefore, considered as a pure product of audiovisual incongruity independent of the auditory past.

As we reported in a previous study using a linguistic version of the symbol-to-sound mismatch task [9], a differential component with the similar temporal profile to sMMN was extracted again in the present study using a music version that was symmetrically designed to the linguistic version of the task. The spatial location of ECDs for iMMN was also replicated in the perisylvian areas within or in the close vicinity of Heschl's gyrus. The resemblance of iMMN to sMMN in the temporal and spatial

Fig. 2. Magnetoencephalographic responses in a representative participant. The 44 perisylvian channels in each hemisphere are bordered with dashed lines (**a**, top). The largest responses in each hemisphere are enlarged: matched (thick lines), mismatched (thin lines), and difference waves (dotted lines) (a, bottom). Isofield contour maps for N1m (**b**, top) and imagery-based mismatch negativity (iMMN) (b, bottom) are shown for each hemisphere. Outflux (solid lines) and influx (dashed lines) are stepped by 20 fT. Black arrows represent equivalent current dipoles (ECDs). ECD locations for N1m (square) and iMMN (circle) on the magnetic resonance image (**c**).

Table 1. Mean (SEM) of the ECD parameters for N1m and iMMN in each hemisphere (*n*=8).

ECD Parameters	Left		Right	
	N1m	iMMN	N1m	iMMN
Peak latency (ms)	96.6 (3.0)	148.4 (10.3)	95.3 (2.1)	153.5 (9.9)
Location x (mm)	−51.4 (2.6)	−51.5 (3.6)	50.1 (1.6)	49.1 (2.8)
y (mm)	5.5 (3.2)	13.3 (4.1)	16.5 (3.8)	16.2 (5.2)
z (mm)	53.5 (2.2)	56.3 (3.5)	54.3 (2.6)	54.6 (2.7)
Moment (nAm)	21.8 (3.2)	14.9 (2.4)	20.7 (3.2)	16.5 (4.2)
Goodness of fit (%)	94.7 (0.8)	81.6 (3.1)	92.0 (1.8)	81.3 (2.4)

profiles strongly suggests that iMMN is an equivalent of sMMN, although the experimental paradigm and modality of attention required for measurement are different. The most probable explanation for iMMN elicitation is, as was the case with the previous linguistic version of the study, that musical notation provokes auditory expectant imagery and its storehouse may function as a template for iMMN elicitation against audiovisual mismatches.

It is worthwhile to discuss the present study from the view point of music perception. It is debatable whether iMMN detected in the present study involves solely nonverbal imagery of notes. Possessors of absolute pitch obligatorily label tones with pitch names [18]. It is probable that the participants might listen to tones with pitch names in their minds and this labeling may be involved in the cortical activity found in the present study [19]. The ECDs for iMMN were localized significantly anterior to those for N1m in only the left hemisphere. This laterality may be in line with a left hemispheric functional [20] and anatomical [21,22] bias for possessors of absolute pitch. On the other hand, the early right anterior negativity is known to be elicited to the violation of musical intuitions of harmony progression in the frontal lobe [23]. In the present study, however, no prominent activity in the frontal areas was detected in this latency range. This may be attributable to single-tone sequences without global tonality used in the present study [16].

The results of the present study were consistent with the previous studies on auditory imagery [10,11]. Perception and expectant imagery in the auditory domain may share neural substrates at an early stage of cortical sensory processing in the auditory cortex. It is controversial whether the primary auditory cortex is involved in imagery [24,25]. Although some experimental conditions may not explicitly activate primary auditory areas, the present study suggests that those areas including auditory sensory memory at least prepare for the upcoming auditory input relevant to immediate imagery on the basis of the memory system.

CONCLUSION

A visually presented musical score provokes auditory imagery of upcoming tones, which can be detected as iMMN localized in the superior temporal plane in musicians. iMMN may represent the neural process of sensory deviance detection against an imagery memory trace in the presyntactic stage, whether in a linguistic or nonlinguistic context. The concept of iMMN proposed in the present study may provide an objective measure of the distance between sensation and expectant imagery, which is otherwise concealed in our minds.

REFERENCES

1. Alho K. Cerebral generators of mismatch negativity (MMN) and its magnetic counterpart (MMNm) elicited by sound changes. *Ear Hear* 1995; **16**:38–51.
2. Näätänen R, Tervaniemi M, Sussman E, Paavilainen P, Winkler I. 'Primitive intelligence' in the auditory cortex. *Trends Neurosci* 2001; **24**:283–288.
3. Näätänen R, Pakarinen S, Rinne T, Takegata R. The mismatch negativity (MMN): towards the optimal paradigm. *Clin Neurophysiol* 2004; **115**:140–144.
4. Karino S, Yumoto M, Itoh K, Yamakawa K, Mizuochi T, Kaga K. A modified parallel paradigm for clinical evaluation of auditory echoic memory. *Neuroreport* 2005; **16**:683–687.
5. Näätänen R. Mismatch negativity: clinical research and possible applications. *Int J Psychophysiol* 2003; **48**:179–188.
6. Kasai K, Yamada H, Kamio S, Nakagome K, Iwanami A, Fukuda M *et al*. Do high or low doses of anxiolytics and hypnotics affect mismatch negativity in schizophrenic subjects? An EEG and MEG study. *Clin Neurophysiol* 2002; **113**:141–150.
7. Tervaniemi M, Maury S, Näätänen R. Neural representations of abstract stimulus features in the human brain as reflected by the mismatch negativity. *Neuroreport* 1994; **5**:844–846.
8. Widmann A, Kujala T, Tervaniemi M, Kujala A, Schröger E. From symbols to sounds: visual symbolic information activates sound representations. *Psychophysiology* 2004; **41**:709–715.
9. Yumoto M, Uno A, Itoh K, Karino S, Saitoh O, Kaneko Y *et al*. Audiovisual phonological mismatch produces early negativity in auditory cortex. *Neuroreport* 2005; **16**:803–806.
10. Kraemer DJ, Macrae CN, Green AE, Kelley WM. Musical imagery: sound of silence activates auditory cortex. *Nature* 2005; **434**:158.
11. Dierks T, Linden DE, Jandl M, Formisano E, Goebel R, Lanfermann H *et al*. Activation of Heschl's gyrus during auditory hallucinations. *Neuron* 1999; **22**:615–621.
12. Janata P. Neurophysiological mechanisms underlying auditory image formation in music. In: Godøy RI, Jørgensen H (eds). *Musical Imagery*. Lisse: Swets & Zeitlinger; 2001, pp. 27–42.
13. Schürmann M, Raij T, Fujiki N, Hari R. Mind's ear in a musician: where and when in the brain. *Neuroimage* 2002; **16**:434–440.
14. Reisberg D. *Auditory Imagery*. Hillsdale: Lawrence Erlbaum Associates; 1992.
15. Brodsky W, Henik A, Rubinstein BS, Zorman M. Auditory imagery from musical notation in expert musicians. *Percept Psychophys* 2003; **65**:602–612.
16. Schön D, Besson M. Visually induced auditory expectancy in music reading: a behavioral and electrophysiological study. *J Cogn Neurosci* 2005; **17**:694–705.
17. Lerdahl F, Jackendoff R. *A Generative Theory of Tonal Music*. Massachusetts: MIT Press; 1983.
18. Marin OSM, Perry DW. Neurological aspects of music perception and performance. In: Deutsch D (ed.) *The Psychology of Music*. San Diego: Academic Press; 1999, pp. 653–724.
19. Hirose H, Kubota M, Kimura I, Yumoto M, Sakakihara Y. N100m in adults possessing absolute pitch. *Neuroreport* 2004; **15**:1383–1386.
20. Ohnishi T, Matsuda H, Asada T, Aruga M, Hirakata M, Nishikawa M *et al*. Functional anatomy of musical perception in musicians. *Cereb Cortex* 2001; **11**:754–760.
21. Schlaug G, Jäncke L, Huang Y, Steinmetz H. *In vivo* evidence of structural brain asymmetry in musicians. *Science* 1995; **267**:699–701.
22. Keenan JP, Thangaraj V, Halpern AR, Schlaug G. Absolute pitch and planum temporale. *Neuroimage* 2001; **14**:1402–1408.
23. Maess B, Koelsch S, Gunter TC, Friederici AD. Musical syntax is processed in Broca's area: an MEG study. *Nat Neurosci* 2001; **4**:540–545.
24. Jäncke L, Shah NJ. Hearing syllables by seeing visual stimuli. *Eur J Neurosci* 2004; **19**:2603–2608.
25. Halpern AR, Zatorre RJ, Bouffard M, Johnson JA. Behavioral and neural correlates of perceived and imagined musical timbre. *Neuropsychologia* 2004; **42**:1281–1292.