Published in final edited form as:

Neurosci Lett. 2008 May 2; 436(1): 85–89. doi:10.1016/j.neulet.2008.02.066.

Units of sound representation and temporal integration: A mismatch negativity study

Attila Océák^a, István Winkler^{a,b,*}, and Elyse Sussman^c

^a Institute for Psychology, Hungarian Academy of Sciences, Budapest, Hungary

^b Institute of Psychology, University of Szeged, Hungary

^c Albert Einstein College of Medicine, Bronx, New York, United States

Abstract

Two acoustic events occurring successively within 200 ms are processed as a single event when the first event predicts the occurrence of the second, but are processed as two separate events when the two events can also occur independently of each other and thus the second event provides new information. However, if the two events are carried by the same stimulus, they are always processed as a single event. This was shown by studies using the mismatch negativity (MMN) event-related potential (ERP). The current study was aimed at investigating the acoustic parameters that determine the integration of successive events within the putative temporal window of integration (TWI). The results demonstrate that temporal grouping (achieved here by presenting sounds in pairs) of the acoustic events within the TWI creates strong unitization, which is not broken up by higher level contingencies of the sound sequence, such as the predictability of the second successive event.

Keywords

Auditory perception; Auditory event detection; Sound representation; Temporal window of integration (TWI); Event-related brain potentials (ERP); Mismatch negativity (MMN)

In our everyday environment, sounds are usually perceived within sequential patterns. For example, a sequence of notes played on piano usually results in the perception of musical phrases and tunes. Finding structure within sound sequences is an essential function of the auditory system in transforming acoustic information into neural representations that support higher level mental operations. Sequential grouping of sounds is not a single operation; rather, it involves several processes working on different timescales and based on different principles [3,7]. One of the strongest forms of sequential grouping is termed temporal integration and it is assumed to underlie perceptual phenomena, such as loudness summation, concurrent masking, perception of short durations, etc. (for a review, see [1]). It is hypothesized that neural effects of the acoustic input are integrated within approximately 170 ms from sound onset (or

*Corresponding author at: Institute for Psychology, Hungarian Academy of Sciences, H-1394 Budapest, P.O. Box 398, Hungary. Tel.: +36 1 3542296; fax: +36 1 3542416. E-mail address: iwinkler@cogpsyphy.hu (I. Winkler). URL: <http://www.cogpsyphy.hu/~iwinkler> (I. Winkler).

Publisher's Disclaimer: This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues. Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit: <http://www.elsevier.com/copyright>

from some large spectral change within a longer sound; slightly differing estimates have been obtained with different methods; see, e.g., [1,16–18]) by an algorithm favoring the more recent input [19,20]. For example, perceived sound intensity is based on the sound energy arriving within the temporal window of integration (TWI) and, therefore, longer duration (but still within the TWI) can compensate for lower signal amplitude.

Effects of temporal integration have also been observed in studies of event-related brain potentials (ERP), especially in those investigating the mismatch negativity (MMN; [4,5]) ERP component. MMN is elicited when a sound violates some acoustic regularity of the preceding sequence (e.g., deviant stimuli in an oddball paradigm). MMN indexes auditory memory representations [6] reflecting the outcome of temporal integration as well as other early steps of auditory information processing [9]. Three prominent temporal integration effects on MMN have been demonstrated. First, masking has parallel effects on sound recognition and the MMN amplitude [15]. Second, omitting a sound from a sequence elicits MMN when the stimulus onset asynchrony (SOA; onset-to-onset interval) is shorter than the duration of the TWI, but not when it is longer [16–18]. Third, when two successive deviations follow each other within the TWI period, usually only a single MMN response is elicited [2,12]. The latter phenomenon has been termed auditory event synthesis, because it shows that multiple auditory events occurring within the TWI are treated as a single event. However, two exceptions to the event synthesis rule have been observed. (1) When the two successive deviations violate two different acoustic regularities, two successive MMNs are elicited even within the TWI [12,13]. This finding indicated that the function of the MMN-generating process is related to the memory representations describing the violated acoustic regularities (i.e., these representations must be updated, when the incoming sound does not fit to them). (2) When the sound sequence includes both single and double deviants (i.e., some deviants were followed immediately by a second deviant, whereas others were not), then double deviants delivered within the TWI elicited two successive MMNs [10,11]. This finding demonstrated that auditory event synthesis is not a destructive process: the second deviation within the TWI is not lost. In sequences in which deviants always occur in pairs (i.e., a deviant is always followed by a second deviant, after which the regular sounds return), the second deviant does not carry new information, since the first deviant fully predicts it. In contrast, when single deviants also occur within the sound sequence, the second deviant carries new information, since a deviant is not always followed by a second deviant. However, this exception to auditory event synthesis requires that successive deviations are carried by separate sounds [12]. When the two deviations occur in succession on a single sound, only one MMN is elicited even when single deviants are included in the sound sequence [2]. Thus, there appears to be a contradiction between the results of event synthesis experiments testing the processing of successive within-TWI deviations in the presence of single deviants. The current experiment was designed to resolve this issue.

Two specific hypotheses have been tested to resolve the contrasting findings.

1. The carrier of the deviations may determine integration or segregation. Thus, two deviations occurring on the same sound would be more strongly integrated than the same successive deviations occurring with the same timing but carried by separate sounds. Such a strong low-level form of sequential grouping could overrule the higher order principle based on the relative information carried by the two deviations. To test this hypothesis, sounds were delivered in pairs in the current study, to strengthen the relationship between the successive deviances but still allow them to occur on separate sounds. These conditions were compared to the isochronous presentation of stimulus deviance replicating the conditions used in previous studies. If temporal grouping results in strong integration of the deviant event, then we should expect that successive deviations occurring within sound pairs of less than 200 ms duration will

elicit one MMN even in the presence of single deviants, mimicking the results obtained when the two deviations were carried by the same sound [2].

2. Continuation of the first deviance through the second deviation, as occurs when they are presented on the same sound, may link the two deviations strongly together. In previous experiments, double deviants occurring on a single sound differed from double deviants occurring on two successive sounds in that at the time when the second deviation occurred, the sound was also deviant in terms of the first deviation. For example, Winkler et al. [14] presented sounds that started as simple sinusoid tone bursts but ended with a frequency glide commencing at 150 ms after the onset of the sound. Double deviants differed from the standard in both sound intensity (10 dB up or down from the standard) and the direction of the frequency glide. Double deviants had the deviant intensity all through the sound, including the interval of the deviant frequency glide. Thus, during the second deviation (the glide part), the deviant concurrently differed from the standard in two stimulus features (intensity and glide direction). In contrast, experiments in which successive deviants were carried by separate sounds presented two single-feature deviants in a row (e.g., a frequency deviant having the standard intensity followed by an intensity deviant having the standard frequency—see, e.g., [12]). Therefore, in the current study, we tested the number of MMNs elicited when the second of two successive deviant sounds carried both deviations.

Finally, we tested the combination of the two manipulations (paired presentation with the continuation of the first deviation through the second deviant). The results, by separating out the possible distinguishing factors of the previous studies, will determine what properties of the acoustic signal lead to the integration of separate deviations to a single event when they occur within the TWI.

Twelve healthy adult volunteers (6 males, age from 19 to 27, mean age 21.7 years) were studied in a sound-attenuated and electrically shielded experimental chamber at the Institute for Psychology, Budapest. Participants gave written informed consent after the procedures of the experiment were explained to them. Experiments were conducted in accordance with the Declaration of Helsinki and with the permission of the Ethical Committee of the Institute for Psychology, Hungarian Academy of Sciences. Participants were instructed to watch a self-selected subtitled movie without sound and to ignore the sounds delivered to their ears through headphones (Sennheiser HD 430). Due to high number of electrical artefacts, two subjects' data was discarded from the analysis.

Stimulus sequences were composed of four different sinusoid tones of 50 ms duration, which included 7.5 ms rise and 7.5 ms fall times. The “standard” tone had a frequency of 440 Hz and an intensity of 50 dB above the individual's hearing threshold (above hearing level—AHL). Individual hearing thresholds were established for the standard tones using Békessy's audiometry method (the staircase method). Frequency-only deviants had a frequency of 470 Hz, whereas intensity-only deviants had an intensity of 30 dB (AHL). Both-features deviants had the combination of the two deviant features.

Fig. 1 shows the four experimental conditions. The Isochronous Separate Deviations (ISD) condition was similar to the sequences that have been previously tested in studies presenting double deviants on separate successive sounds. This was the baseline condition to which the effects of the two experimental manipulations were compared to (1st row in Fig. 1). The ISD condition presented 4800 stimuli in two stimulus blocks of 2400 stimuli each. Each stimulus block included 90 single frequency-only deviants, 90 single intensity-only deviants, and 90 double deviants in which a frequency-only deviant was followed immediately by an intensity-only deviant. That is, altogether 15% of the tones were deviants, half of which appeared in

double, the other half in single deviants. The order of standard and deviant stimuli was randomized with the constraint that two deviant tones of any type were separated by at least two standard tones (double-deviants counted as a single deviant event). The SOA was uniformly 120 ms.

The Isochronous Combined Deviations (ICD) condition (2nd row in Fig. 1) differed from the ISD condition in that the second of the two successive deviants was a “both-features” deviant stimulus, deviating from the standard tone in both frequency and intensity. This condition modeled the continuation of the first deviation throughout the second deviation in double deviants when presented within single sounds.

The Paired Separate Deviations (PSD) condition (3rd row in Fig. 1) was similar to the ISD condition except that the tones were presented in pairs. An inter-pair interval of 400 ms was used, creating pairs with an onset-to-onset interval of 570 ms (120 ms within pair SOA + 50 ms duration of the second tone + 400 ms between-pair interval). Thus the within-pair SOA was identical to the uniform SOA of the ISD and ICD conditions. Double deviants occurred within pairs. In both double and single deviants, frequency deviance always occurred on the first tone of a pair, and intensity deviance always on the second tone of a pair. The PSD condition was presented in three stimulus blocks of 1600 stimuli each. Within each stimulus block, there were 60 double-deviant pairs, 60 pairs with a frequency-only deviant in the first position, and 60 pairs with an intensity-only deviant in the second position. Similarly to the isochronous conditions, the overall deviant-stimulus probability was 15% with half of the deviants appearing in double deviants, the other half as single deviants. The order of the standard and different deviant pairs was randomized with the constraint that two deviant pairs of any type were separated by at least one standard pair.

The Paired Combined Deviance (PCD) condition (4th row in Fig. 1) was similar to the PSD condition except that the second tone of the double-deviant pairs was deviant in both frequency and intensity.

Thus, the two possible variables (isochronous vs. paired presentation and separated vs. combined deviance) were tested in a 2×2 design. The effect of the temporal structure of the stimulation can be assessed by comparing the results of the two isochronous conditions (ISD and ICD) with those of the two paired conditions (PSD and PCD). The effects of the relation between the first and the second deviant can be assessed by comparing the results of the two separated deviations conditions (ISD and PSD) with those of the two combined deviations conditions (ICD and PCD).

In addition to the experimental stimulus blocks, two types of control stimulus blocks, 400 tones each, were also presented to subjects. The two control stimulus blocks presented the two double-deviant stimulus pairs (frequency-only-deviant followed by an intensity-only-deviant and frequency-only-deviant followed by a both-features deviant), alone, with the same timing as that used in the paired conditions (400 ms between-pair interval). These stimulus blocks served to delineate the deviation-specific responses elicited by the corresponding deviant pairs. Similar controls could not be constructed for the isochronous stimulus blocks, because repeating two different tones (the first and the second deviant of the double-deviants) with a uniform SOA creates a sequence of two regularly alternating tones. Responses obtained from such stimulus blocks would not be comparable with those obtained for the deviants of the oddball stimulus blocks. Therefore, for the isochronous conditions, responses elicited by the standard tones were used as controls for the deviants (as was also done in all previous studies of this type). The order of the various stimulus blocks was balanced both within and across participants.

Electroencephalogram (EEG) was recorded (SYNAMP amplifiers, NeuroScan Inc.) from scalp locations at Fz, Cz, Pz, F3, F4, C3, C4, P3, P4 (10–20 system) and from the left and right mastoids (Lm and Rm, respectively) with Ag–AgCl electrodes and referred to the common reference electrode placed on the tip of the nose. Horizontal and vertical electrooculograms (HEOG and VEOG, respectively) were measured with separate bipolar electrode montages: HEOG was recorded between electrodes placed lateral to the outer canthi of the two eyes, VEOG between electrodes below and above right eye. After amplification and on-line filtering (low pass frequency 40 Hz, -24 dB/octave), the EEG signals were digitized with a sampling frequency of 250 Hz. The stored EEG signals were off-line band-pass filtered between 1 and 20 Hz (-48 dB/octave) and epochs of 600 ms duration (including 100 ms pre-stimulus interval) were extracted for each tone in the isochronous and each tone-pair in the paired conditions. Epochs with an amplitude change exceeding $100 \mu\text{V}$ at any EEG or EOG channel were omitted from further processing. The remaining epochs were separately averaged for the different conditions and stimulus types and baseline-corrected using the average voltage in the initial 100 ms period of the epochs.

Deviance-related ERP responses were assessed by subtracting from the deviant-stimulus response the corresponding control-stimulus response (the standard-stimulus response for the isochronous and the control-pairs response for the paired conditions). MMN peak latencies were determined separately for each subject and condition within the expected MMN time intervals (between 100 and 200 ms from deviation onset): in the 100–200 and 220–320 ms windows from the onset of the first tone of a double deviant. MMN amplitudes were measured separately for each subject and condition as the mean voltage of 20-ms long windows centered on the negative peak(s) found in the expected MMN intervals in the group-averaged difference waveforms. MMN elicitation in these intervals was tested with Student's t test. The effects of the experimental conditions on the MMN peak latency and amplitude were tested by two-way repeated measures ANOVAs with factors of Temporal Structure [Isochronous vs. Paired] $\times$ Second Deviant Type [Single-feature vs. Both-features]. All significant effects are discussed and effect size (η^2) is reported.

Fig. 2 shows the frontal (Fz) and left mastoid responses recorded for control and double-deviant stimuli in the four experimental conditions, together with the corresponding deviant-minus-control difference waveforms. Two successive negative difference peaks were elicited by the isochronous conditions (both ISD and ICD), whereas only one such peak was elicited in the paired conditions. The mastoid responses show slight polarity inversions coinciding with the frontally negative peaks. Therefore, these peaks have been identified as MMN responses elicited by the deviant stimuli. Student's t tests showed that each observed peak in the group-averaged difference waveforms reflected a statistically significant MMN response. Group-averaged MMN peak latencies and amplitudes as well as the results of Student's t tests are given in Table 1. Additionally, for the paired conditions, difference amplitudes were measured in the latency range centered on the peak of the difference between single intensity (or both-features) deviants and standard pairs. This allowed the testing of possible MMN elicitation by the second deviant of double-deviant pairs (PSD and PCD conditions), which did not produce a separate difference peak. The peak latencies of the single-deviant MMNs and the corresponding mean amplitudes in the paired conditions (measured from 20-ms long windows centered on the single-deviant MMN peaks) are also given in Table 1. Neither of these measures differed significantly from zero.

Peak latencies of the first MMN were longer in the paired than in the isochronous conditions ($F(1,9) = 6.79, p < .05, \eta^2 = 0.43$). Peak latencies of the single frequency-only (first) deviants showed the same result: significantly longer MMN peak latencies for paired compared with isochronously presented tones ($F(1,9) = 5.40, p < .05, \eta^2 = 0.38$). No significant difference was found between the second MMN peak latencies between the two isochronous conditions

(in which two successive MMNs were elicited). No differences were found between the amplitudes of the first MMN responses between the four experimental conditions ($F(1,9) = 0.40$ and 1.60 for the Temporal Structure and the Second Deviant Type factors, respectively). Finally, for the isochronous conditions (since no second MMN was elicited in the paired conditions), there was no significant differences between the second-MMN amplitudes ($F(1,9) = 0.87$ for the Second Deviant Type factor) or between the amplitudes of the first and second MMNs ($F(1,9) = 2.54$ for the 1st vs. 2nd MMN factor).

In summary, two successive MMN components were elicited by double deviants with isochronous presentation, whereas only one MMN was elicited in the paired conditions. Thus we found that temporal grouping (paired stimuli) within the TWI period creates a sound unit. Events within the unit are processed together even when higher level rules suggest separate processing of the events. The grouping demonstrated by the current and previous studies of auditory event synthesis may be a type of “default” auditory process. That is, when participants have no task related to the sounds, the extraction of auditory regularities and the detection of events violating them, are processed in terms of these units. However, this type of grouping does not result in the loss of information (unlike some stronger forms of temporal integration, see [1]), because when participants are directed to detect the second of two successive deviants, they are able to do so [11]. In everyday life, a large part of the sound input remains outside the focus of attention. Thus, the formation of sound units that may provide a sufficiently informative description of the unattended sounds likely reduces the amount of capacities engaged by the irrelevant information.

The latency of the MMNs elicited by first (frequency) deviants was longer in the paired than in the isochronous conditions. Since pairs were preceded by longer SOAs than single sounds in the isochronous conditions, this result replicates those of Schröger & Winkler [8], who suggested that longer pre-deviant SOA deteriorates the representation of the standard, thus making it more difficult to detect deviance. Alternatively, it is also possible that when the auditory system is set for processing stimulus pairs or longer stimuli, it requires a longer sample of the auditory input before deviance can be detected.

Acknowledgements

This study was supported by the Hungarian National Research fund (OTKA T048383, I.W.) and the National Institutes of Health (DC 004263, E.S.).

References

1. Cowan N. On short and long auditory stores. *Psychol Bull* 1984;96:341–370. [PubMed: 6385047]
2. Czigler I, Winkler I. Preattentive auditory change detection relies on unitary sensory memory representations. *NeuroReport* 1996;7:2413–2417. [PubMed: 8981394]
3. Köhler, W. *Gestalt Psychology*. Liveright; New York: 1947.
4. Näätänen R, Gaillard AWK, Mäntysalo S. Early selective attention effect on evoked potential reinterpreted. *Acta Psychol* 1978;42:313–329.
5. Näätänen R, Paavilainen P, Rinne T, Alho K. The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clin Neurophysiol* 2007;118:2544–2590. [PubMed: 17931964]
6. Näätänen R, Winkler I. The concept of auditory stimulus representation in cognitive neuroscience. *Psychol Bull* 1999;125:826–859. [PubMed: 10589304]
7. Poeppel D. The analysis of speech in different temporal integration windows: cerebral lateralization as ‘asymmetric sampling in time’. *Speech Commun* 2003;41:245–255.
8. Schröger E, Winkler I. Presentation rate and magnitude of stimulus deviance effects on pre attentive change detection. *Neurosci Lett* 1995;193:185–188. [PubMed: 7478179]

9. Sussman ES. Integration and segregation in auditory scene analysis. *J Acoust Soc Am* 2005;117:1285–1298. [PubMed: 15807017]
10. Sussman E, Winkler I. Dynamic sensory updating in the auditory system. *Cogn Brain Res* 2001;12:431–439.
11. Sussman E, Winkler I, Kreuzer J, Saher M, Näätänen R, Ritter W. Temporal integration: intentional sound discrimination does not modulate stimulus-driven processes in auditory event synthesis. *Clin Neurophysiol* 2002;113:1909–1920. [PubMed: 12464328]
12. Sussman E, Winkler I, Ritter W, Alho K, Näätänen R. Temporal integration of auditory stimulus deviance as reflected by the mismatch negativity. *Neurosci Lett* 1999;264:161–164. [PubMed: 10320039]
13. Winkler I, Czigler I. Mismatch negativity: deviance detection or the maintenance of the ‘standard’. *NeuroReport* 1998;9:3809–3813. [PubMed: 9875709]
14. Winkler I, Czigler I, Jaramillo M, Paavilainen P, Näätänen R. Temporal constraints of auditory event synthesis: evidence from ERPs. *NeuroReport* 1998;9:495–499. [PubMed: 9512396]
15. Winkler I, Reinikainen K, Näätänen R. Event related brain potentials reflect traces of the echoic memory in humans. *Percept Psychophys* 1993;53:443–449. [PubMed: 8483708]
16. Yabe H, Tervaniemi M, Reinikainen K, Näätänen R. Temporal window of integration revealed by MMN to sound omission. *NeuroReport* 1997;8:1971–1974. [PubMed: 9223087]
17. Yabe H, Tervaniemi M, Sinkkonen J, Huotilainen M, Ilmoniemi RJ, Näätänen R. Temporal window of integration of auditory information in the human brain. *Psychophysiology* 1998;35:615–619. [PubMed: 9715105]
18. Yabe H, Sato Y, Sutoh T, Hiruma T, Shinozaki N, Nashida T, Saito F, Kaneko S. The duration of the integrating window in auditory sensory memory. *Electroencephalogr Clin Neurophysiol* 1999;49 (Suppl):166–169.
19. Zwislocki JJ. Theory of temporal auditory summation. *J Acoust Soc Am* 1960;32:1046–1060.
20. Zwislocki JJ. Temporal summation of loudness: an analysis. *J Acoust Soc Am* 1969;46:431–440. [PubMed: 5804115]

Isochronous Conditions

Separate Deviations

Combined Deviations

Paired Conditions

Separate Deviations

Combined Deviations

Standard

Frequency deviant

Intensity deviant

Frequency + intensity deviant

Fig. 1.

Schematic illustration of the stimulus sequences. The four experimental conditions are presented in separate rows. Time flows from left to right. Tones are shown by squares with the fill pattern marking the acoustic parameters. Standards tones had 440 Hz frequency and 50 dB (AHL) intensity, frequency-only deviants 470 Hz and 50 dB, intensity-only deviants 440 Hz and 30 dB, and frequency + intensity deviants 470 Hz and 30 dB.

Fig. 2.

ERP responses. Group-averaged ($N = 10$) frontal (Fz; left column) and left mastoid (Lm; right column) ERP responses elicited by double deviants (thick line) and corresponding control stimuli. The corresponding difference waveforms are plotted in the central column. 0 ms is at the onset of the first tone of the double deviant.

Table 1

MMN peak latencies and amplitudes in the four experimental conditions

Condition	Early MMN		Late MMN	
	Peak latency	Amplitude	Peak latency	Amplitude
ISD	159 (25)	-1.20 (1.62) ^a	236 (23)	-0.53 (0.58) ^a
ICD	156 (22)	-1.74 (1.44) ^b	257 (36)	-0.88 (1.23) ^a
PSD	175 (20)	-1.35 (1.72) ^a	278 (34) ^c	0.10 (0.93)
PCD	183 (14)	-1.82 (1.75) ^b	282 (32) ^c	0.25 (1.02)

Group averaged ($N = 10$) MMN peak latencies (ms) and frontal (Fz) MMN amplitudes (μV) measured from 20-ms long windows centered on the peaks found in the grand average difference waveforms in the four experimental conditions. Standard deviations are given in parentheses.

^aDenote $p < .05$ shown by Student's one-sample t tests.

^bDenote $p < .01$ shown by Student's one-sample t tests.

^cLatency taken from the corresponding single-deviant responses.