# Preattentive auditory context effects

ISTVÁN WINKLER
*Hungarian Academy of Sciences, Budapest, Hungary*
*and University of Helsinki, Helsinki, Finland*

ELYSE SUSSMAN
*Albert Einstein College of Medicine, New York, New York*

MARI TERVANIEMI
*University of Helsinki, Helsinki, Finland*

JÁNOS HORVÁTH
*Hungarian Academy of Sciences, Budapest, Hungary*

WALTER RITTER
*Nathan Kline Institute for Research in Schizophrenia, Orangeburg, New York*

and

RISTO NÄÄTÄNEN
*University of Helsinki, Helsinki, Finland*

The effects of auditory context on the preattentive and perceptual organization of tone sequences were investigated. Two sets of experiments were conducted in which the pitch of contextual tones was varied, bringing about two different contextual manipulations. Preattentive auditory organization was indexed by the mismatch negativity event-related potential, which is elicited by violations of auditory regularities even when participants ignore the sounds (e.g., by reading a book). The perceptual effects of the contextual manipulations on auditory grouping were assessed using target-detection and order-judgment tasks. The close correspondence found between the effects of auditory context on the perceptual and preattentive measures of auditory grouping suggests that a large part of contextual processing is preattentive.

The way sounds are perceived is largely determined by the auditory context in which they occur. The perceptual effects produced by the presentation of sounds simultaneously with or in close temporal proximity to a given auditory stimulus are usually interpreted as being caused by interference and/or integration during the automatic phase of auditory processing (see, e.g., Cowan, 1984; Massaro, 1975). In some cases, however, explanations based on selective attentional processes have also been suggested (such as for contralateral recognition masking; Hawkins & Presson, 1977; see, however, Kallman & Morris, 1984). Other forms of contextual influence on auditory perception, such as certain types of adaptation,

temporal distinctiveness, and auditory stream segregation (see, e.g., Bregman, 1990; Cowan, 1995; Glenberg, 1987), often require several seconds of exposure to the stimuli to be built up (in auditory streaming; see Bregman, 1978). These effects are mediated by the large-scale perceptual organization of the auditory input. That is, the global auditory context affects how the individual sounds in the sequence are grouped in perception—for example, whether they are perceived as belonging to a coherent series, the elements of which are typically attributed to a single sound source. The question addressed by the present study is whether this type of auditory contextual organization requires attention to be directed to the sounds or whether contextual grouping effects can occur independently of focused attention.

This question is related to the issue of whether or not the formation of object representations (including the feature conjunction) and, ultimately, their grouping into object files can occur without focused attention. Although the majority of the relevant research has been conducted on the visual modality, the principal ideas have been assumed to be applicable across different modalities. Treisman (1992; Kahneman, Treisman, &

Gibbs, 1992) suggested that setting up object files requires attention. Evidence suggesting that feature integration requires focused attention supports this notion (Treisman, 1982; for updates, see Treisman, 1993; for another theory maintaining that feature integration requires focused attention, see Wolfe, 1994; Wolfe, Cave, & Franzel, 1989). Results showing that participants have no explicit memory about the grouping of unattended display items (Mack, Tang, Tuma, Kahn, & Rock, 1992) are also compatible with Treisman's view.

However, proponents of the automatic grouping hypothesis point out that the lack of explicit memory does not necessarily rule out the possibility that elements of the unattended background were structured into groups (see Moore & Egeth, 1997). Furthermore, there exists a body of evidence indicating that grouping might occur without focused attention. For example, Driver and Mattingley (1998) showed that stimulus grouping affects awareness of visual material presented contralaterally to the damaged side in patients with unilateral neglect. Results showing that distractors forming a homogeneous group interfere less than do heterogeneous distractors with the detection of targets in visual search tasks (e.g., Duncan & Humphreys, 1989) also support the notion of preattentive grouping. Furthermore, distractors grouped together with the target impose a greater cost on discrimination performance than do distractors that belong to a different perceptual group (Baylis & Driver, 1992; Driver & Baylis, 1989).

The notion of preattentive grouping includes the assumption that all items of the visual field are automatically processed to a high degree even before grouping occurs (Driver, 1996; Duncan, 1984; Duncan & Humphreys, 1989; Kubovy, Cohen, & Hollier, 1999). Arguing against the notion of automatic grouping, Treisman (1993) suggested an alternative explanation of the reduced interfering effect of homogeneous distractors:

> Suppose that initially a single random element in a homogeneous conjunction display is identified with focused attention, setting up an object file for its parts and their spatial relations. This token is then used as a template to suppress matching objects across the whole display . . . . Because attention is needed to maintain an object file, spatially parallel suppression is possible only for one template at a time. If the distractors are not identical . . . a unique item no longer pops out. (p. 27)

Thus, according to Treisman (1993), Duncan and Humphreys's (1989) results do not force one to assume the existence of automatic grouping processes.

When applying these notions to auditory phenomena, one should be very cautious, because for concepts such as *object*, there is no one-to-one correspondence between the auditory and visual modalities. Nevertheless, the end product of perceptual analysis is comparable between the two modalities in that we experience the sensory world in terms of objects and events irrespective of the modality of the input. However, the route leading to coherent perception of the auditory world may differ from that corresponding to the visual world. This is because, unlike in vision, there is no single acoustic feature that separates sounds emitted from different sources. In short, there is no auditory equivalent of visual spatial maps. In fact, there may not even be sufficient information present in the acoustic input of the ears for calculating a unique solution of the auditory source-decomposition problem (see Stoffgren & Brady, 2001). The auditory system uses various and, in most situations, valid assumptions (such as those represented by the Gestalt principles of proximity, continuity, and similarity) to tease apart those parts of the acoustic input that belong to separate sources or patterns and to bind together those that belong to the same acoustic entity. This means that in the presence of multiple active sound sources (as is the case in most natural situations), the auditory system must organize its input by finding coherent threads of sounds within the composite stimulation. This process has been termed *auditory scene analysis*, within which the coherent sound threads have been called *auditory streams* (Bregman, 1990).

The question of whether or not the segregation or integration of sound streams requires attention has been debated in the behavioral literature. Bregman (1990; see also Bregman & Rudnicky, 1975) suggested that some forms of auditory stream segregation, such as the auditory streaming effect, are preattentive (see also Anstis & Saida, 1985; Beauvois & Meddis, 1991). An indirect effect of stream segregation provides support for this notion. Investigating the *irrelevant speech effect*,[1] D. M. Jones and his colleagues (D. M. Jones, Alford, Bridges, Tremblay, & Macken, 1999; D. M. Jones & Macken, 1995) found that the effect was weaker when the unattended irrelevant sounds were segregated from the standard than when they were grouped together with it. The authors reasoned that, since the auditory stimuli were irrelevant to the task and participants were instructed to ignore them, grouping (stream segregation) of the auditory material did not require focused attention.

However, the assumption of preattentive auditory stream segregation is not universally accepted. For example, M. R. Jones and her colleagues argued that auditory streaming represents a limitation of the ability to shift one's attention rapidly between two sounds with widely different auditory features (M. R. Jones, 1976; M. R. Jones, Kidd, & Wetzel, 1981; M. R. Jones, Maser, & Kidd, 1978). On a different basis, Carlyon, Cusack, Foxton, and Robertson (2001) also proposed that auditory stream segregation requires attention. These authors showed that a demanding primary task can interfere with the buildup of auditory streaming. Carlyon et al. also demonstrated that patients with unilateral neglect showed deficits in stream segregation when the sounds were presented to the affected side. From these results, the authors concluded that the formation of auditory streams probably requires focused attention. However, since most previous research on the structuring of auditory sequences used methods that required the partici-

pants to attend the sounds, the preattentive stream segregation hypothesis could not be tested directly.

The present study addressed the question of whether the preattentive organization of a sequence of sounds depends on what other sounds surround it. That is, if certain perceptual organizational processes were preattentive, would it then be possible that the auditory context within which a given sequence of sounds occurs could influence the perceptual organization of the sequence of sounds independently of the direction of focused attention?

Recent studies have provided electrophysiological evidence supporting the notion of preattentive auditory grouping (Ritter, Sussman, & Molholm, 2000; Sussman, Ceponiene, Shestakova, Näätänen, & Winkler, 2001; Sussman, Ritter, & Vaughan, 1998b, 1999; Winkler, Schröger, & Cowan, 2001; Yabe et al., 2001). These studies showed that an event-related brain potential (ERP) component termed the *mismatch negativity* (MMN) can be used to index grouping relations between auditory stimuli while participants ignore the test sounds. Therefore, in conjunction with traditional behavioral measures of auditory stream segregation, the research to be reported used MMN as a dependent variable to test the hypothesis of preattentive context effects.

MMN (Näätänen, Gaillard, & Mäntysalo, 1978; for recent reviews, see Näätänen & Winkler, 1999; Picton, Alain, Otten, & Ritter, 2000; Schröger, 1997) is elicited by sounds violating some characteristic regularity of the preceding auditory stimulus sequence. MMN appears with maximal (negative-polarity) amplitude over the frontocentral scalp 100–200 msec after the violation of the regularity commenced (which is usually at stimulus onset). In the auditory oddball paradigm, MMN is elicited by infrequent *deviant* sounds discriminably differing from the frequent *standard* stimulus. Deviations in simple as well as complex perceived auditory features (such as pitch, loudness, duration, virtual pitch, phonetic features, etc.) result in MMN elicitation (for reviews, see Näätänen & Alho, 1997; Näätänen & Winkler, 1999). There are also other paradigms in which MMN can be observed (for reviews of MMN elicited in non-oddball designs, see Näätänen, Tervaniemi, Sussman, Paavilainen, & Winkler, 2001; Näätänen & Winkler, 1999). In contrast, no MMN is elicited by frequent repetitive sounds or infrequent sounds presented alone (i.e., without the frequent repetitive stimulus; Näätänen, Paavilainen, Alho, Reinikainen, & Sams, 1989) or by a stimulus change at the beginning of an auditory sequence (Cowan, Winkler, Teder, & Näätänen, 1993).

Näätänen and his coworkers established that the process generating the MMN component is based on an auditory sensory-memory representation of the standard (for a discussion, see Näätänen, 1992; Näätänen & Alho, 1997; Näätänen & Winkler, 1999) and not on differences in the refractory state of the neural elements activated by the standard and the deviant stimuli (for a discussion of the *memory-trace* vs. the *release-from-refractoriness* explanation of the MMN component, see Näätänen, 1990,

1992). Winkler, Karmos, and Näätänen (1996) suggested that, in addition to sensory information, the memory system involved in the MMN-generating process contains records of the regularities extracted from the preceding auditory stimulation (see also Cowan et al., 1993; Ritter, Gomes, Cowan, Sussman, & Vaughan, 1998; Schröger, 1997; Winkler et al., 2001). MMN reflects an information-filtering process that can initiate the modification of the records that represent the regularities of the auditory environment (Sinkkonen, 1999; Winkler & Czigler, 1998; Winkler, Karmos, & Näätänen, 1996) and call for further processing of those sounds that carry new information (Näätänen, 1990, 1992). The observable MMN component seems to index the outcome of mismatch detection, rather than being an on-line correlate of the detection process (Ritter, Deacon, Gomes, Javitt, & Vaughan, 1995).

Two characteristics of the MMN-generating process make the MMN especially useful for the purposes of the present study. First, since MMN is elicited by violations of detected auditory regularities, it can be used to determine what regularities are currently tracked by the auditory system (Näätänen & Winkler, 1999; Ritter et al., 1995; Winkler, Karmos, & Näätänen, 1996). Second, MMN elicitation does not require participants to actively detect the deviant sounds. Focused attention in and of itself does not affect MMN elicitation and is not needed for it (Sussman, Winkler, & Wang, in press), and active discrimination of the deviant sounds or top-down predictive information about the occurrence of the deviants does not influence the MMN component (Rinne, Antila, & Winkler, 2001; Ritter, Sussman, Deacon, Cowan, & Vaughan, 1999; Sams, Paavilainen, Alho, & Näätänen, 1985; Sussman, Winkler, & Schröger, in press).[2] Therefore, MMN can be elicited when participants are instructed to ignore the sounds and to perform a visual primary task (such as reading a book or doing visual discrimination or tracking tasks; see, e.g., Alho, Woods, Algazi, & Näätänen, 1992; Winkler, Cowan, Csépe, Czigler, & Näätänen, 1996). The MMN elicited under these circumstances shows excellent correspondence with performance in active task situations, including its sensitivity to the amount of change, individual differences, arousal state, and learning effects (for reviews, see Näätänen & Alho, 1997; Näätänen & Winkler, 1999). Therefore, measuring MMN allows the assessment of the formation of regularity records without contamination from task-related top-down processes.

These properties of the MMN-generating process can be utilized in investigating preattentive sound organization. Several studies demonstrated that the perceptual organization of the auditory stimulus sequence may determine what regularities can be discerned from the auditory input (for a review, see Bregman, 1990). For example, in one demonstration of auditory stream segregation, two melodies (one familiar and one unfamiliar) were interleaved within an auditory stimulus sequence. When the two melodies were presented in similar frequency

ranges, participants could not discern the familiar tune. However, when the pitch of the familiar melody was transposed into a range that was sufficiently separate from that of the unfamiliar melody, participants could easily identify the familiar tune (Dowling, 1973). Thus, the organization of the auditory input determines the detectable auditory regularities and, hence, the elicitation of the MMN component.[3] The authors of the MMN studies of auditory grouping (Ritter et al., 2000; Shinozaki et al., 2000; Sussman et al., 2001; Sussman, Ritter, & Vaughan, 1998a, 1998b, 1999; Winkler et al., 2001; Yabe et al., 2001) based their conclusions on this logic.

Sussman et al. (1999) argued that the sorting of high-pitched and low-pitched sounds associated with the auditory streaming effect occurs preattentively. The perceptual organization of alternating high and low tones was controlled by presentation of the same stimulus sequence at two different delivery rates (a fast- and a slow-paced condition). This manipulation was based on results showing that the rate of stimulus delivery can control the emergence of auditory streaming: The faster the pace of stimulus presentation, the smaller the frequency separation that invokes auditory streaming (Bregman, 1990; Dannenbring & Bregman, 1976; van Noorden, 1975). Therefore, with a fixed frequency separation between the high and low tones, slow presentation rates result in the perception of a single stream of alternating high and low tones, whereas fast stimulus delivery brings about the perception of separate high-pitched and low-pitched sound streams. Sussman et al.'s (1999) participants were instructed to read a book and ignore the auditory stimuli. The test sequences were arranged so that a repetitive three-tone (standard) pattern occasionally broken by a deviant pattern (created by reversing the order of the tones in the standard pattern) appeared separately within the high-pitched and low-pitched sequences. These patterns emerged in perception only when the high and low tones were segregated (fast-paced condition). In consistency with this, MMN was elicited only by the deviant tone-patterns in the fast-paced condition, showing that segregation of the high and low tones occurred prior to the regularity detection underlying the MMN-generating process. Furthermore, the results suggest that auditory streaming does not require attention to be focused on the sounds. Preattentive auditory stream segregation based on multiple acoustic cues was demonstrated by Ritter et al. (2000).

Furthermore, the results of Sussman et al. (1998b) suggest that temporal grouping can also occur preattentively. These authors presented a repeating five-tone sequence consisting of four identical (standard) tones followed by a higher frequency (deviant) tone. Participants were instructed to read a book and to ignore the sounds. At the slow presentation rate (stimulus onset asynchrony [SOA] = 1.3 sec), the deviants elicited an MMN (see also Scherg, Vajsar, & Picton, 1989). However, when the same sequence was presented at a fast pace (SOA = 0.1 sec), no MMN was elicited even though the same deviant tones (presented with the same probability and SOA) elicited an MMN when these tones appeared randomly among the standard tones. The results can be explained by assuming that at the fast presentation rate, the periodically presented five-tone segment was grouped into a single pattern, and the repetition of this pattern became the regularity for which the MMN system was set. In this case, the high (deviant) tone became part of the regularity governing the elicitation of MMN and, therefore, it did not elicit an MMN. In contrast, at the slow presentation rate, tones were not grouped together, either because of their large temporal separation (>1 sec; see Handel, 1993) or because the periodicity of the presentation could not be preattentively detected (the time required for delivery of two full cycles [13 sec] probably exceeded the duration of auditory sensory memory). Therefore, the regularity for which the MMN system was set became the repetition of the standard tone and, in violation of this rule, the relatively rare higher frequency tones elicited an MMN. Another preattentive temporal grouping effect has been observed by Winkler et al. (2001).

The results of the experiments described above strongly support the notion that at least some forms of auditory grouping (such as streaming and temporal grouping) occur even when attention is directed away from the sounds. The present study takes the investigation of auditory grouping one step further. In two sets of experiments, we tested whether the auditory context surrounding a given set of sounds affects the organization of this sound set when attention is not focused on it. Alternatively, it is possible that the auditory organization of sound sequences is fully determined by the sequences themselves (i.e., independent of the context) in the absence of attention focused on the sounds. Therefore, the crucial feature that distinguishes the present experiments from those previously mentioned is that we kept the test sequence constant while manipulating the auditory context within which this sequence appeared. Whereas in previous studies the parameters affecting sound organization were investigated, in the present experiments we investigated whether grouping processes can adapt to different auditory environments in the absence of focused attention.

In addition, in the present experiments we tested the correspondence between preattentive and perceptual sound organization by comparing behavioral and ERP indices of the grouping of attended and ignored sounds. If a change of context had similar effects on the organization of the test sequences irrespective of the direction of focused attention, then these contextual effects on auditory grouping probably occurred preattentively. Conversely, if the results varied with the direction of focused attention, then one should conclude that the auditory context affected the organization of the test sequences after the stage of the MMN-generating process.

Experiment 1 was based on the notion that interference is stronger between sounds that are part of the same stream than between sounds that belong to separate

streams (Idson & Massaro, 1976; D. M. Jones, Macken, & Harries, 1997; for a compatible finding for the visual modality, see Baylis & Driver, 1992). The addition of random tones to a predominantly regular sequence of sounds can be expected to interfere with the detection of regularity in the resulting composite sound sequence. However, if the random and the regular-sequence tones are segregated into separate sound streams, the regularity may become detectable again.

These assumptions were based on results obtained in two-tone comparison tasks that interpolated irrelevant sounds between the standard and comparison tones (Deutsch, 1970; for a discussion, see Deutsch, 1978, 1984). In several studies, it was found that the decrease in recognition performance was a function of the similarity between the test and the intervening tones (Deutsch, 1970; Massaro, 1975; Pechmann & Mohr, 1992). These effects were explained in terms of retroactive memory interference, suggesting that the formation of memory traces for the sounds following the test tone may have damaged or eliminated the memory record of the test tone. D. M. Jones et al. (1997) showed that the organization of the sound sequence is a critical factor in the disruption of short-term recognition. When the interpolated tones are grouped together with the standard tone, they interfere with recognition, but when they are segregated from the standard tone, the effect on recognition performance is minimal. One possible explanation for this is that retroactive interference is considerably stronger within than across tonal groups. D. M. Jones et al. (1997), however, suggested that the memory traces of sounds belonging to separate auditory streams were maintained independently of each other, as has been demonstrated in

a number of behavioral studies (e.g., Bregman, 1990). Therefore, segregation of the standard and the interpolated tones could have eliminated interference between them. Comparing the two alternative explanations, D. M. Jones et al. (1997) showed that when similarity between sounds was equalized, intervening sounds gave rise to significantly less interference with pitch recognition when they were segregated from the standard tone than when they were grouped together with it.

In Experiment 1A, the effects of the pitch of the intervening tones on the detection of rare target tones in an auditory oddball paradigm was assessed. In Experiment 1B, the MMN method was used to investigate the effects of the same manipulation while participants ignored the auditory stimuli. Finally, Experiment 1C was conducted to test whether participants could have been aware of the frequent and rare tones during Experiment 1B.

## EXPERIMENT 1A

The purpose of Experiment 1A was to test the effects of random contextual tones on the detection of infrequent deviants in an auditory oddball paradigm. The *oddball-alone condition* was a simple auditory oddball design in which a repetitive (standard) tone was presented at a fixed time interval. The standard was infrequently replaced by a shorter (deviant) tone (Figure 1, top panel). The participant's task was to press a response key when a deviant tone was detected.

In the test conditions, the oddball sequence was interleaved with a sequence of random tones, with two of said tones being introduced between consecutive elements of
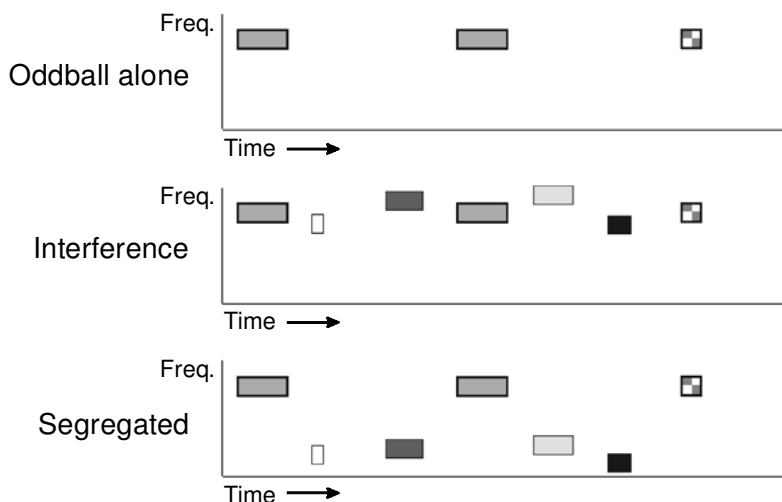


Figure 1. Schematic illustration of the stimulus paradigm for Experiment 1. The three experimental conditions are shown in separate panels; time increases along the *x*-axis. The position of the stimulus rectangles on the *y*-axis represents tone frequency, rectangle width represents tone duration, and darkness of shading, intensity. The outlines of the test-sequence tone rectangles are thicker than those of the contextual tones. Target (deviant) tones are marked by the checkered pattern. Note that the test sequence and the standard and deviant tones are identical in the three conditions.

the oddball sequence. The participants were required to press a response key to the same short oddball-deviant tones as were present in the oddball-alone condition. It was expected that the interfering effect of the intervening tones on detection performance would depend on the frequency separation between the oddball-sequence tones and the random intervening tones: Detection performance should be close to chance level when the intervening tone frequency is varied in a range that includes the oddball-sequence frequency (*interference condition*, Figure 1, middle panel), whereas the performance level should be close to that achieved in the oddball-alone condition when the frequency of the intervening tones varies in a range that widely differs from the frequency of the oddball tones (*segregated condition*, Figure 1, bottom panel).

## Method

**Participants**. Ten healthy participants (5 females, mean age 27.6 years) with normal hearing took part in Experiment 1A. The experiment was conducted in an acoustically dampened room at the Institute for Psychology, Budapest.

**Stimuli**. The stimuli were pure sinusoidal tones presented binaurally through headphones (NeuroStim stimulation system). The standard tone, which occurred with an 85% probability, was 1813 Hz in frequency, 75 dB (sound pressure level [SPL]) in intensity, and 275 msec in duration (2.5-msec rise and 2.5-msec fall times included). The deviant tone, which occurred randomly with a 15% probability, was 100 msec long (all other stimulus parameters being equal to those of the standard tone). The constant SOA was 1,200 msec in the oddball-alone condition (Figure 1, top panel).

Through presentation of two intervening tones between consecutive tones of the oddball-alone condition, the constant SOA of the test blocks became 400 msec for the interference and segregated conditions (Figure 1, middle and bottom panels). The intervening tones equiprobably assumed any of 12 different tone durations (75, 87.5, 112.5, 125, 137.5, 150, 162.5, 175, 187.5, 200, 212.5, and 225 msec; 12.5-msec steps, 2 below and 10 above the deviant duration) and four intensity levels (70, 72.5, 77.5, and 80 dB [SPL]; 2.5-dB steps, 2 below and 2 above the common intensity of the standard and deviant tones). Four different frequencies were used for the intervening tones in the interference and segregated conditions, separately, again with equal probability. In the interference condition (Figure 1, middle panel), the intervening tone frequencies were 1655, 1732.2, 1897.5, and 1986 Hz, centered on the common frequency of the standard and deviant tones and in equal log frequency steps. In the segregated condition (Figure 1, bottom panel), the intervening tones were 250, 261.7, 286.6, and 300 Hz in frequency (again, with equal log frequency steps, with the central level omitted). None of the intervening tones had frequency, intensity, or duration values equal to those of the standard or deviant tones in either condition. The ratio between the highest and lowest frequencies was equal between the two test conditions ($1986 \div 1655$ Hz = 1.2 in the interference condition and $300 \div 250$ Hz = 1.2 in the segregated condition). The order of the intervening tones was randomized, with the restriction that none of them appeared two or more times in a row.

**Procedure**. Stimulus blocks—two for each condition—contained 287 oddball-sequence tones (244 standard and 43 deviant tones). The first block of the experimental session was an oddball-alone stimulus block, which served to introduce the task to the participant. The participants were informed that the sequence would contain two types of tone, differing only in duration. They were instructed to press the response key as quickly as they could when they heard the shorter, infrequent tone. For half of the participants, the first oddball-alone block was followed by the two segregated blocks. After a short break, another oddball-alone block was presented, and it was followed by the two interference blocks. The other participants received the two halves of the experiment in the reversed order. Before the segregated and interference blocks, separately, the participants were informed about the structure of these sequences with the use of an illustration similar to that presented in Figure 1, including the type of intervening tones that occurred in them. They were instructed to press the response key to the same deviant tones as in the preceding oddball-alone block, but not to other short tones.

The median of the reaction times (RTs) for correctly detected deviant tones and the $d'$ measure of detection performance were calculated for each participant separately, in each condition. Null responses were not included in the analysis of RTs or $d'$ values. Differences between the conditions were tested by one-way analyses of variance (ANOVAs) with dependent measures (Greenhouse–Geisser correction applied) and Scheffé-type pairwise post hoc comparisons.

## Results

The participants found the task easy in the oddball-alone and segregated conditions. In the latter case, they reported hearing the low intervening tones as a separate channel that they could easily ignore to focus on the regular oddball sequence. In contrast, in the interference condition, with one exception,[4] the participants could not separate the intervening tones from the oddball tones and, consequently, they could not detect the target tones. Reportedly, all of the tones were part of the same seemingly chaotic sequence, and the fast presentation rate (which resulted from hearing all tones in a single stream) prevented the participants from keeping track of the oddball tones by counting. Some of the participants resorted to guessing, whereas others did not press the response key at all.

These subjective reports match the pattern of results that emerged from the responses (Table 1). The hit rate (HR) was equally high (above 90%), the false alarm (FA) rate equally low (below 0.2%), and $d'$ values were above 3.8 in the oddball-alone and segregated conditions, whereas the HR was low (28%), the FA rate relatively high (about 3%), and $d'$ values were below 2.7, except in one participant (see note 4) in the interference condition. RTs were shortest in the oddball-alone and longest in the interference condition, with that of the segregated condition falling closer to (and not being significantly different from) that of the oddball-alone condition. ANOVA and post hoc tests confirmed that detection performance was lower and RTs longer in the interference condition than in either the oddball-alone or the segregated condition, with no significant difference between the latter two (see Table 1).

## EXPERIMENT 1B

In Experiment 1B, we tested whether the regularity-detection processes underlying MMN elicitation would be affected by the contextual manipulation applied in

**Table 1**
**Grand-Average $d'$ Values and Reaction Times (± Standard Deviation [SD]) Obtained in the Control, Segregated, and Interference Conditions of Experiment 1A**

| Condition | $d'$ | | | | Reaction Time | | | |
|---|---|---|---|---|---|---|---|---|
| | Mean | SD | Scheffé $p$ values | | Mean (msec) | SD | Scheffé $p$ values | |
| | | | Segregated | Interference | | | Segregated | Interference |
| Oddball alone | 4.68 | 0.09 | .241 | .0001 | 401 | 64 | .5126 | .0056 |
| Segregated | 4.26 | 0.39 | | .0002 | 442 | 43 | | .0564 |
| Interference | 1.10 | 0.55 | | | 521 | 85 | | |
| ANOVA | $F(1.51,6.04) = 150.52, p < .0001$ | | | | $F(1.15,9.2) = 7.61, p < .02$ | | | |

Note—Results of the one-way ANOVAs with dependent measures (Greenhouse–Geisser correction applied; see degrees-of-freedom values) for the $d'$ and RT values are given in the bottom row. Results of the post hoc Scheffé-type pairwise comparisons are shown for each variable and pair of conditions.

Experiment 1A when participants ignore the auditory stimuli.

Deviant tones in the auditory oddball sequence of the oddball-alone condition (Figure 1, top panel) were expected to elicit an MMN peaking 200–300 msec from stimulus onset. The late peaking of this MMN would be in line with the known properties of MMN, which usually peaks 100–200 msec after onset of the deviation. In the present paradigm, the onset of the deviation is at the offset of the shorter, 100-msec-long (deviant) tone, because it is at this point that the shorter tones can be distinguished from the longer tones.

**Method**

**Participants**. Ten healthy adult participants (3 females, mean age 26.4 years) with normal hearing, none of whom had taken part in Experiment 1A, participated in Experiment 1B. Informed consent was obtained from all the participants after the testing procedure was explained to them.

**Stimuli and Procedure**. The electroencephalogram (EEG) was recorded in an electrically shielded, acoustically dampened room at the Institute for Psychology, Budapest. The participants sat in a comfortable chair and were instructed to read a book during the whole experiment and to ignore the auditory stimuli.

The stimulation was identical to that of Experiment 1A, except that the stimulus blocks were longer, consisting of 680 standard, 120 deviant, and—in the interference and segregated conditions—1,600 intervening tones. For each condition (oddball alone, interference, segregated), two stimulus blocks were delivered. In addition, three short (320 oddball-sequence tones) reversed-stimulus blocks, one for each condition, were also presented. In these reversed blocks, the standard and deviant tones were exchanged (i.e., the short tone was presented in 85% of the oddball sequences and the long tone in 15%, all other parameters and stimuli being identical to those of the corresponding test block). ERP responses recorded in the reversed-stimulus blocks provided a comparison for those obtained in the corresponding experimental blocks (see EEG recording and data analysis). The order of the blocks was separately randomized for each participant. Breaks (>5 min) were introduced after the third and sixth stimulus blocks.

The EEG was recorded with Ag/AgCl electrodes attached to the participants' scalps at the Fz, Cz, Pz, and Oz midline (10–20 system) locations, at the left and right mastoids (LM and RM, respectively), and at the one-third and two-thirds points on each side of the coronal line connecting the mastoids via Fz (starting from the midline, L1 and L2 on the left side and R1 and R2 on the right side of the head). Horizontal eye movements were measured by recording the electro-oculogram (EOG) between the outer canthi of the two eyes,

and vertical eye movements were measured by recording the vertical EOG between Fpz (10–20 system) and the common reference electrode, which was attached to the tip of the nose. The EEG was digitized (using SynAmps amplifiers) at a 250-Hz rate (100 Hz low pass) and then off-line filtered between 1 and 30 Hz. Epochs of 500-msec duration, starting 100 msec before and ending 400 msec after the onset of the standard and deviant tones, were collected. Epochs with a change exceeding 100 $\mu$V at any recording channel were rejected from subsequent processing. This procedure resulted in removal of the majority of trials contaminated by eye movement or other artifacts of noncortical origin.

For each participant, the artifact-free epochs were averaged separately for the three conditions and two stimulus types (standard and deviant). The mean amplitude in the 100-msec prestimulus period was subtracted from each point of the averaged ERP responses. This prestimulus period served as the reference ("biological zero"), relative to which the stimulus-elicited electrical responses were measured.

The MMN can be quantified by subtracting from the ERP elicited by the infrequent deviant tone, the response to the same tone recorded in a block in which this tone is the frequent stimulus. This subtraction eliminates those components that are independent of the role of the given stimulus in the sequence while leaving the MMN, which is only elicited by a deviant stimulus, unchanged (for a detailed description, see, e.g., Näätänen, 1992). Therefore, for each condition separately, we subtracted from the ERP elicited by the deviant tone, the ERP elicited by the corresponding *control tone* (which was the same as the deviant tone, but appearing as the frequent standard stimulus of the corresponding reversed sequence). The MMN amplitude was measured as the mean frontal (Fz) amplitude in the 230- to 270-msec interval of the deviant − control difference curves. This interval includes the grand-mean MMN peak latencies in those conditions in which MMN was elicited (see the Results section). One-sample $t$ tests were used to verify the elicitation of the MMN component.

**Results**

Figure 2 shows the ERP responses obtained in the three conditions. Sizable negative waves appear in the frontal (Fz) deviant − control difference curves (Figure 2, right column) between 200 and 300 msec in the oddball-alone (peak latency 240 msec) and segregated (peak latency 256 msec) conditions. In contrast, no MMN is observed in the interference condition. Accordingly, in the oddball-alone and segregated conditions, the deviant − control difference was significantly different from zero $[-1.21 \pm 0.34 \mu V$ (mean ± SEM), $t(9) = 3.51, p < .01$, and $-1.37 \pm 0.46 \mu V$, $t(9) = 3.0, p < .02$; one-sample
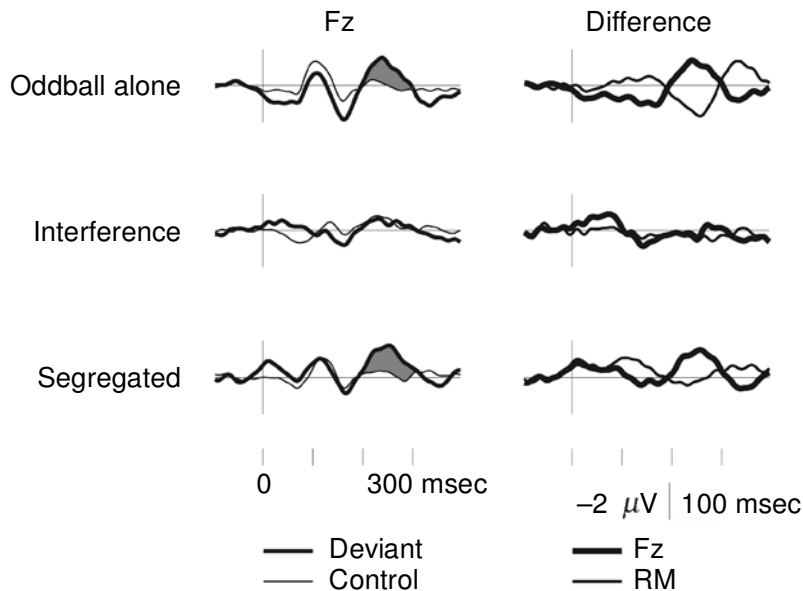
**Figure 2. Experiment 1B: The left panels show the grand-average frontal (Fz, left) responses to deviant (thick line) and control (reversed-block standard, thin line) tones in the three experimental conditions. MMN responses are marked by the shaded area between the deviant- and control-tone ERPs. In the right panels, the corresponding deviant − control difference waves at Fz (thick line) and RM (thin line) are shown.**

*t* tests of the average frontal difference amplitudes in the 230- to 270-msec interval for the oddball-alone and segregated conditions, respectively]. The same measure did not differ significantly from zero in the interference condition [$0.15 \pm 0.19 \ \mu V$, $t(9) = 0.76$].

The negative deviant − control difference waves observed in the oddball-alone and segregated conditions had a frontocentral maximum, appeared with reversed polarity at the mastoid leads, and were slightly larger over the right than over the left hemisphere. These are characteristic features of the MMN (see Alho, 1995; Näätänen, 1990, 1992). Therefore, we can conclude that MMN was elicited in the oddball-alone and segregated conditions, but not in the interference condition.

## EXPERIMENT 1C

The participants in Experiment 1B were instructed to read a book and to ignore the sounds. Although reading was checked on line by observation of the horizontal EOG record on the computer screen, the instructions did not prevent the participants from covertly attending the sounds either intermittently or by continuously dividing their attention between reading and listening to the tones. To check whether this was the case, we determined whether the participants were aware of which sound they heard more frequently in the oddball sequence. This test was done only for the segregated condition, which was the crucial condition in Experiment 1B

because it showed that the tone-repetition regularity was detected by the processes underlying MMN when the intervening tones were separated in frequency from the oddball tones. A test of memory for to-be-ignored stimulation can be accomplished only once per participant; otherwise, he/she will be prepared for subsequent tests. Therefore, this experiment was conducted once per participant, after all passive conditions and before any active ones included in a variety of other experiments, which were conducted at the Institute for Psychology, Budapest. For comparison, the same test was administered to another group of participants, who were instructed to attend the tone sequence.

### Method

**Participants**. One group of 29 healthy participants (19 females, mean age 21.2 years) and another group of 13 healthy participants (10 females, mean age 25.3 years), all with normal hearing, took part in Experiment 1C. None of these individuals had participated in Experiment 1A or 1B.

**Stimuli and Procedure**. The participants of the first (*ignore*) group sat in a comfortable chair and were instructed to read a book of their choice and to ignore the auditory stimuli delivered to them through headphones. The participants of the second (*attend*) group were told to attend all sounds, which were identical to those presented to the ignore group.

The participants were presented with a single block of sounds, which was a slightly modified version of the segregated-condition stimulus block of Experiment 1A. The modifications were aimed at preventing the participants from being able to answer the test question solely on the basis of their auditory sensory memory record of

the last part of the stimulus block or the memory they may have formed by attending the sounds at the beginning of the sequence. Therefore, during the first 150 and the last 14 high tones (the oddball sequence in Experiments 1A and 1B), the short and long tones appeared with equal probability. The probabilities of the short and long tones during the middle part of the block (150 tones) were identical to those in Experiments 1A and 1B (15:85). Altogether, 942 (3 × 314) tones were delivered. All other parameters were identical to those described for the segregated condition of Experiment 1A. When the stimulus block ended, the experimenter informed the participants that the tone sequence comprised two different high and many different low tones. The participants were then asked to listen to the two high tones again (one short and one long) and to tell whether they appeared with approximately equal probability within the whole stimulus block, or, if not, which of them appeared more frequently. If a participant claimed not to have a clear memory of the tone sequence, he/she was instructed to guess. The presentation order of the two tones following the stimulus block was balanced across participants.

### Results

Most of the participants of the ignore group reported that they were guessing the answer, since they had no memory of what tone sequence they had heard while they were reading. These subjective reports match the quantifiable results. Of the 29 participants, 10 reported that the short tones appeared more frequently, 11 that the long ones did, and 8 thought that the two tones appeared with equal probability. This distribution of the responses is close to chance level, suggesting that the participants did not notice the relationships between the sounds they had encountered while reading. The fact that there are fewer "equal" responses than those naming either of the tones showed that most of the participants relied neither on an impression they might have had if they had attended the beginning of the sequence nor on the sensory-memory record of the last part of the sequence, which would have encouraged "equal" responses.

In sharp contrast, 12 of the 13 participants of the attend group correctly identified the long high tone as appearing more frequently than the short high tone. The probability of this result's being due to chance is $p = .003$. Therefore, we can conclude that when the participants attended the tone sequence, they detected which of the high tones appeared more frequently and retained this information until they had answered the experimenter's question. It should be noted that before the tone sequences were presented, the participants were not told which tones or what aspect of the tones to focus on, nor were they told that their memory of the tones would be tested later.

### EXPERIMENT 1D

It is, however, also possible that attention to the tones had a graded effect on stream segregation. That is, the two streams would be separated only part of the time in the segregated condition of Experiment 1B, the ratio of segregated and integrated organization across time depending on the load of the primary task. Experiment 1D was designed to test this possibility by introducing a primary task with scalable memory load—the n-*back task*. In the *n*-back task, a series of stimulus items is presented to the participants, who are instructed to decide whether or not each item is identical to the one presented 1, 2, . . . (*n*) items before it. The *n*-back task requires continuous task performance, because each item is a target as well as a to-be-remembered item. In one sense, it is like memorizing a serial list, which changes with every new item, the oldest item being dropped off the list and the most recent one appended to it. Increasing *n* increases the number of past items to be held in memory at any given time. If attention has a graded effect on auditory streaming, one should expect that the MMN measure of streaming used in Experiment 1B would change as a function of task load and the *n* of the concurrent primary task, higher *n*s resulting in lower MMN amplitudes. No change in the MMN amplitude would mean that maintaining the segregation of two sound streams is not dependent on the amount of free attentional capacity.

### Method

**Participants**. Eleven healthy participants (7 females, mean age 19.4 years) with normal hearing volunteered for this experiment. None of the volunteers had taken part in Experiment 1A, 1B, or 1C.

**Stimuli and Procedure**. The participants performed a visual *n*-back task, similar to the one described by Watter, Geffen, and Geffen (2001). Visual stimuli were presented on a 14-in. computer monitor placed 1.2 m in front of the participant. The black fixation cross appearing at the center of the gray (61.0 cd/m²) screen was seen at 0.43º. Red (29.50 cd/m², red channel only on the RGB screen) circles at a 1.01º viewing angle were randomly presented in 1 of 12 possible positions. The target circles appeared on the perimeter of a 3.67º viewing-angle circle centered on the fixation cross. Target positions were placed equidistantly (30º apart) on the (invisible) central circle, shifted by 15º from the positions of the hour marks on a normal clock face. Targets appeared for 200 msec with an SOA of 2,195 msec; 75 visual trials were presented, equally distributed through the auditory stimulus block. The timing of the visual targets and that of the auditory stimuli were independent of each other.

In separate stimulus blocks, the participants were instructed to press a reaction key when (1) the target appeared in the same position as the previous one (*1-back task*) or (2) the target appeared at the same position as the item preceding it by three (*3-back task*). Matches occurred on 33% of the trials. Half of the participants started with the 1-back task and the other half with the 3-back task. The participants received training before each task and were motivated by bonus payments to perform the tasks as best they could. Eleven stimulus blocks were delivered for each condition, of which the 4th and the 8th were control (reversed) auditory stimulation blocks (see the Method section of Experiment 1B). The participants' task performance level was characterized by the $d'$ calculus (1 participant's data had to be omitted because of no misses or FAs in the 1-back task), response accuracy (for comparison with previous, similar experiments), and median RT for correct responses. Comparisons between conditions were performed by Student's *t* test.

Auditory stimulation was identical to that described for the segregated condition of Experiment 1B, except that the stimulus blocks consisted of 460 stimuli in this experiment—fewer than those in the passive condition of Experiment 1B, because of the demanding primary task. We recorded only the auditory ERP responses, which do

not show visual ERP components. This is because averaging smears out the visual ERP components in the auditory ERP waveforms, since the visual targets were not systematically time-locked to the auditory stimuli. The EEG was filtered with a pass-band of 1–20 Hz, and the artifact rejection level was set at 75 $\mu$V. Amplitude measurements were referred to the mean voltage in the 90-msec prestimulus period. MMN amplitudes were measured in the 232- to 272-msec interval. All other parameters of the ERP analysis were identical to those described for Experiment 1B.

## Results

**Task performance**. The participants performed the task significantly less accurately in the 3- than in the 1-back condition [$d'$ values, 1.78 ± 0.13 (mean ± SEM) vs. 4.39 ± 0.31 for the 3-back and 1-back tasks, respectively, [$t(9) = 9.37$, $p < .001$; accuracy .797 ± .016 vs. .975 ± .009, $t(10) = 12.85$, $p < .001$]. Also, the median RT for correct reactions was significantly longer in the 3-back than in the 1-back blocks [566.7 ± 61.1 msec vs. 460.5 ± 38.3 msec, $t(10) = 2.72$, $p < .03$]. These results closely follow those obtained by Watter et al. (2001) for their matching-position trials.

**ERP responses**. Figure 3 (left column) shows the ERP responses elicited by the standard and deviant tones in Experiment 1D. A frontally negative (positive at LM) difference wave peaking between 200 and 300 msec was elicited in both conditions (Figure 3, right column). These MMN responses were very similar to those obtained in the segregated condition of Experiment 1B (see Figure 2). The MMN amplitudes did not differ significantly between the two task conditions ($-0.79 ± 0.34$ at Fz and $0.55 ± 0.16$ at LM for the 1-back primary task,

and $-1.22 ± 0.22$ at Fz and $0.52 ± 0.16$ at LM for the 3-back primary task).

## Discussion of Experiments 1A, 1B, 1C, and 1D

The patterns of results obtained in Experiments 1A and 1B, showing the effects of contextual sounds on the voluntary (Experiment 1A) and unattended (Experiment 1B) detection of deviants in an auditory oddball paradigm, closely corresponded to each other. In the oddball-alone condition of Experiment 1A, in which no contextual tones were delivered, the participants performed at a high level of accuracy in detecting the infrequent short target tones within the series of a repetitive long tone. These short deviant tones elicited the MMN while the participants were reading a book and had no task related to the auditory stimuli (Experiment 1B). When random tones with frequencies close to that of the oddball-sequence (standard/deviant) tones were interpolated between successive tones of the oddball sequence in the interference condition, detection performance dramatically decreased; HR dropped from 94% to 28% and $d'$ from 3.8 to 2.7 (Experiment 1A), and no MMN was elicited (Experiment 1B). However, when the frequency separation between the oddball and intervening tones was large (the segregated condition), the accuracy of detection of the oddball deviant tones was high, close to that in the oddball-alone condition (Experiment 1A). Correspondingly, a sizable MMN component was elicited by these deviants in the corresponding passive situation (Experiment 1B). Moreover, the participants of Experiment 1A reported hearing a single stream of sounds in
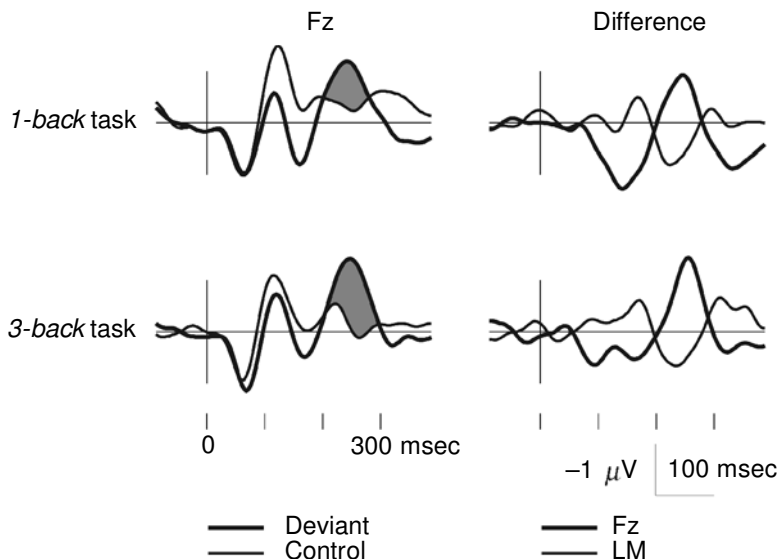


Figure 3. Experiment 1D: The left panels show the grand-average frontal (Fz, left) responses to deviant (thick line) and control (reversed-block standard, thin line) tones in the two experimental conditions. MMN responses are marked by the shaded area between the deviant- and control-tone ERPs. In the right panels, the corresponding deviant − control difference waves at Fz (thick line) and LM (thin line) are shown.

the interference condition, whereas in the segregated condition they perceived two separate streams: one consisting of the (low) intervening tones, and the other of the (high) oddball-sequence tones.

These subjective reports suggest that the present results reflect a grouping effect (auditory streaming) and can be interpreted similarly to D. M. Jones et al.'s (1997) explanation of the results obtained in Deutsch's (1970) paradigm. Because deviant tones appeared only infrequently in the oddball sequence, in Experiment 1A the participants likely used the strategy of maintaining a trace of the frequently presented standard stimulus (an *attentional template*) and comparing incoming stimuli with this template (Näätänen, 1985, 1992). The tones intervening between successive stimuli of the oddball sequence could interfere with the formation, preservation, or retrieval of this template. Although the attentional template strategy is somewhat different from the mode of operation of the MMN-generating process (which is based on auditory regularities), the participants' reports about the interference condition reveal that the template strategy also required regularity detection. The participants of Experiment 1 reported that they did not perceive anything regular in the interference condition. Specifically, they commented that, despite what was shown to them in the explanatory figure, all the tones seemed to belong to a single sequence and to vary randomly. That is, with one exception, the participants could not structure the sequence in a way that would have enabled them to identify the regularly repeating standard tones. Therefore, they could not detect the target deviant tones.[5] As a consequence, all the participants except one were unable to perform the task with any degree of success. In contrast, in the segregated condition, the intervening tones could be rejected as members of a group that contained only distractors (for an analogous explanation of the results of some visual search studies, see Duncan, 1993), which allowed the participants to detect the regularity of the oddball sequence and, hence, to form and maintain the attentional template necessary for performing the task.

Thus, it appears that the close correspondence between MMN elicitation in the participants who ignored the auditory stimuli and their performance in voluntary detection of targets in the same auditory stimulus sequences stemmed from a common source: The contextual tones could interfere with the analysis of the oddball sequence by affecting the organization of the auditory scene.

Alternative explanations should still be considered. First, as frequency separation affects tonal similarity and stream segregation in parallel, the results of Experiment 1A alone cannot distinguish between the retroactive-interference and grouping explanations. However, as has already been mentioned, D. M. Jones et al. (1997) showed that the amount of interference in the irrelevant sound paradigm more closely followed the formation of separate streams for the standard and the intervening sounds than

the similarity between these sounds. Therefore, the present results probably reflect a stream segregation effect.

Second, the frequency separation between the oddball and the interpolated tones could directly affect MMN elicitation. That is, frequency separation would separately determine perceptual stream segregation and MMN. However, Sussman et al. (1998a) has shown that MMN elicitation and perceptual auditory stream segregation go together even when all stimulation parameters (including frequency separation) are kept constant. This suggests that the present effects of frequency separation are not independent of each other but stem from a common cause: the organization of the auditory input. Because previous studies have also shown that temporal parameters can determine MMN by affecting the grouping of sounds when spectral parameters are kept constant (Sussman et al., 1998b, 1999; Winkler et al., 2001), it is not likely that the contextual effects observed in the present Experiment 1A and the MMN effects observed in Experiment 1B would reflect two separate, independent effects of the same experimental manipulation (cf. Experiments 2A–2C, which provide a strong test of this issue).

The present MMN results suggest that contextual effects on auditory grouping can occur even when attention is not focused on the auditory stimulation. Although reading (the primary task in Experiment 1B) did not prevent the participants from continuously or intermittently allocating a portion of their attention to the sounds, the results of Experiments 1C and 1D strongly argue against this possibility. Performance of the attend group of Experiment 1C in the memory test showed that when the participants attentively monitored the auditory stimuli throughout the stimulus block, they noticed the gross difference between the rates of recurrence with which the short and long tones appeared in the longest part of the sequence (the middle). The fact that they had no memory of this or of the structure of the sound sequence in general (according to the subjective reports) when they were instructed to ignore the sounds suggests that the participants of Experiment 1A and the ignore group of Experiment 1C did not attend the auditory stimulation. Lack of memory about the structure of the sequence, however, does not rule out the possibility that the auditory stimulus sequence was grouped preattentively (see Moore & Egeth, 1997).

Another possible attention-based explanation of the results of Experiments 1A and 1B was ruled out in Experiment 1D. According to this alternative, attention would have a graded effect on auditory grouping. Some grouping may occur even with relatively little attention directed to the sounds, although this amount of attention may not be sufficient for remembering the composition of the sound sequences. However, in Experiment 1D, the load of the concurrent primary task did not significantly affect the MMN amplitude, when MMN elicitation indexed the segregation of high and low tones in the test

sequences. These test sequences were identical to those presented in the segregated condition of Experiments 1A and 1B. If attention was necessary for maintaining the separation of the two streams, then increasing the task load of the primary task should have resulted in segregation in a smaller percentage of the total time. As a consequence, fewer deviant tones should have elicited MMN, thus reducing the mean MMN amplitude. The results of Experiment 1D, however, showed no indication of a task-load effect on the MMN amplitude, even though the participants' performance was significantly affected by the amount of task load. Therefore, the results obtained in Experiment 1 suggest that certain types of auditory grouping processes, such as those based on a large frequency separation between two sets of tones, may occur independently of the direction of focused attention. Furthermore, the results showed that this type of preattentive grouping affects perception and the deviance detection process reflected by the MMN component in similar ways.

This conclusion confirms assumptions made in previous studies about the correspondence between preattentive and perceptual auditory grouping (Ritter et al., 2000; Sussman et al., 1998b, 1999; Winkler et al., 2001; Yabe et al., 2001). It may also be noted that the results of Experiments 1B and 1D, showing (for the participants who ignored the auditory stimuli) that a large *frequency* separation between the intervening and oddball-sequence tones eliminated interference in the *duration* feature, indicate that integration of auditory features occurred preattentively, before the stage of the MMN-generating process. Finally, the insensitivity of the MMN to the load of the concurrent primary task supports the conclusion of Sussman, Winkler, and Wang (in press) that attention in and of itself does not affect the elicitation or the amplitude of the MMN response (see also Näätänen, 1990, 1992).

In Experiment 1, we investigated whether contextual manipulation could preattentively affect the detection of a given auditory regularity. To achieve this, the oddball sequence used in Experiment 1 had a simple fixed regularity (repetition of the standard tone), which could be obscured by the intervening tones. In Experiment 2, we investigated a different contextual manipulation: the effect of contextual sounds on the relationship (segregation vs. grouping together) of two sets of test sounds. Furthermore, instead of varying the detectability of a single regularity (as was done in Experiment 1), we designed Experiment 2 to manipulate the contextual tones (whereas the two sets of test tones remained the same), promoting a switch between two different perceptual organizations and, as a consequence, between two different preattentively detectable regularities.

In Experiment 2A, we investigated the perceptual effects of manipulating the contextual tones by asking listeners to judge the order of successive tones. Using the MMN response, in Experiment 2B we tested whether this manipulation affected the organization of the same tone sequences even when the participants ignored the auditory stimuli.

## EXPERIMENT 2A

The purpose of Experiment 2A was to assess the perceptual effects of contextual tones on the grouping of the two sets of test tones. Bregman (1978; see also van Noorden, 1975) showed that two test tones with a fixed frequency separation can become integrated into a common stream or segregated into two different streams, depending on the context provided by surrounding flanking tones. Bregman (1978) explained this phenomenon in terms of *competition among frequency proximities*. The relative frequency levels of a given set of tones, not their absolute levels, contribute to their grouping.

To study the effects of context on auditory grouping, in Experiment 2 we presented a modified version of Bregman's (1978) paradigm. In the design, two sets of test tones (A and B) and two additional sets of tones, called contextual (flanking) tones, were used. Members of each tone set varied in frequency, but not so much as to approach the frequencies of any other set (see below). In the *A-and-B-separate condition* (Figure 4, top panel), the frequencies of the X and Y tones closely surrounded those of the lower frequency A and higher frequency B tones; the X-tone frequencies were slightly lower than those of the A tones, and the Y-tone frequencies were slightly higher than those of the B tones. The frequency separation between the A and B tone sets was greater than that between the A and X or the B and Y set. On the basis of Bregman's (1978) results, we expected that presenting tones from the four sets with equal probability in a random order would promote the grouping of the A and X tones separately from the group formed of the B and Y tones. Increasing the X- and Y-tone frequencies far beyond those of the A and B tones should result in the A and B tones' being grouped together in perception (*A-and-B-together condition*), segregated from the group formed of the X and Y tones (Figure 4, bottom panel). Note that the A and B tones are identical in the two conditions— only the frequencies of the X and Y tones vary. The stimulus parameters were chosen on the basis of the results of previous behavioral studies (for a description of the factors that influence sound grouping in this paradigm, see Bregman, 1990).

In several previous studies, the accuracy of tone-order judgments was used to determine whether stream segregation occurred, since the order of tones is more accurately judged when the tones belong to the same stream than when they do not (Bregman, 1978, 1990; Bregman & Campbell, 1971; Bregman & Rudnicky, 1975; Broadbent, 1958). The present experiment used the order-judgment method to test the influence of the flanking tones on how the test tones were grouped into streams. In addition, the participants were asked to report how confident they were of their judgments, as it was expected that they would be more confident in judging the order of two
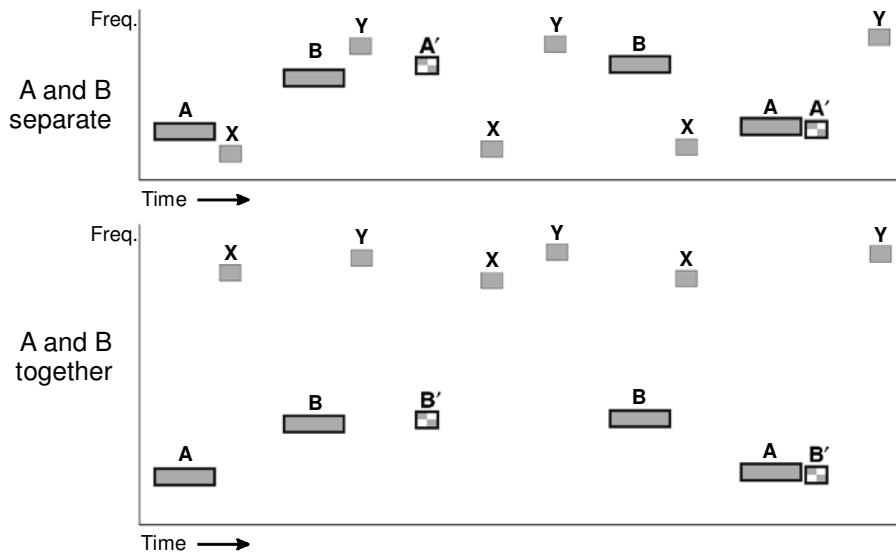
Figure 4. Schematic illustration of the stimulus paradigm for Experiment 2. The two experimental conditions are shown in separate panels; time increases along the *x*-axis. The set from which the tone was taken (A, B, X, and Y) is marked over each stimulus rectangle. The *y*-axis position of the stimulus rectangles represents tone frequency, and rectangle width represents tone duration. Probe tones (A′ and B′, used only in Experiments 2B and 2C) are marked by the checkered pattern. The outlines of the test-sequence tone rectangles (A, A′, B, and B′) are thicker than those of the contextual tones.

tones from the same stream than that of two tones falling across different streams.

## Method

**Participants**. Ten healthy adult participants (7 females, mean age 30.7 years) took part in Experiment 2A. All reported having normal hearing, and none of them had participated in Experiment 1. The experiment was conducted at the Albert Einstein College of Medicine, New York.

**Stimuli and Procedure**. Sequences of pure tones equiprobably selected from four tone sets (test tones: A and B; flanking tones: X and Y) were binaurally presented through headphones with a constant 275-msec SOA. Each tone set consisted of five different frequencies separated by equal 11-Hz steps (but see Experiment 1C) and delivered equiprobably within the test sequences. In Tone Set A, the frequencies ranged from 370 to 414 Hz and in Tone Set B, from 637 to 681 Hz. The A and B tones were 250 msec long (2.5-msec rise and 2.5-msec fall times included). As for the flanking tones, in the A-and-B-separate condition (Figure 4, top panel), Tone Set X (*X-separate*) frequencies ranged from 315 to 359 Hz (ending one step below the range of the A set), and Tone Set Y (*Y-separate*) frequencies ranged from 718 to 762 Hz (starting one step above the range of the B set). The X-separate and A tones formed a continuous frequency range, as did the B and Y-separate tones, separately. In the A-and-B-together condition (Figure 4, bottom panel), the flanking tones had much higher frequencies than did the test tones: Tone Set X (*X-together*) ranged from 1842 to 1886 Hz, and Tone Set Y (*Y-together*) ranged from 2415 to 2459 Hz. In both conditions, the flanking tones were 100 msec long (including 2.5-msec rise and 2.5-msec fall times), a feature required to ensure compatibility with Experiment 2B.

On each trial, a 15-sec tone sequence was presented, followed by 1 sec of silence and two comparison tones separated by a 275-msec SOA. The comparison tones repeated the last two tones of the test sequence, either in the same or in reversed order. The participant's task was to write down on a prepared form whether the two comparison tones occurred in the same order or in a different order in comparison with the last two tones of the test sequence (forced-choice task). In addition, the participants were instructed to rate, for each trial, how confident they were of their judgments, using a scale of 1–5, on which 1 meant *absolutely sure* and 5 meant *guessing*. The participants had as much time as they needed to make their responses. When they were ready for the next trial, they started it by pressing a response key.

Forty trials were presented in total, 20 for each condition, separately. The order of the tones within each sequence was randomized (and was the same for all the participants) up to the last two tones. Half of the test sequences (10 for each condition) ended in two tones that were expected to be in different perceptual streams (*expected across-stream* judgments): In the A-and-B-separate condition, an A appeared with a B tone; in the A-and-B-together condition, an A or a B tone was paired with an X-together or a Y-together tone. The other half of the test sequences (10 for each condition) ended in two tones that were expected to be in the same perceptual stream (*expected within-stream* judgments): In the A-and-B-separate condition, an A tone was presented with an X-separate tone or a B tone with a Y-separate tone; in the A-and-B-together condition, an A tone appeared with a B tone. On half of the trials (5 for each of the 4 cases described above), the two comparison tones were presented in the same order as the last two tones of the test sequence; on the other half of the trials, the two comparison tones were presented in reversed order in comparison with the last two tones of the test sequence. The order of the 40 trials was randomized, and they were presented in two blocks of 20 trials, with a resting period in between.

For each condition separately, the accuracy and confidence of expected within- and across-stream order judgments were compared (e.g., A–X and B–Y pairs vs. A–B pairs in the A-and-B-separate condition). In addition, the judgments and confidence ratings for the A–B pairs were compared between the two conditions, because

judgments of the same A–B pairs were expected to be across-stream judgments in the A-and-B-separate condition, but within-stream judgments in the A-and-B-together condition. Statistical testing was done by means of paired $t$ tests.

## Results

In both experimental conditions, the participants were more accurate in their order judgment in the expected within- than in the across-stream trial. In the A-and-B-separate condition, the mean number of correct judgments (10 judgments per case) was $9.5 \pm 0.7$ ($\pm SD$) for the expected within-stream tone pairs versus $7.9 \pm 1.5$ for the expected across-stream tone pairs [$t(9) = 3.21$, $p < .01$]. In the A-and-B-together condition, the analogous mean number of correct judgments was $9.9 \pm 0.3$ versus $8.0 \pm 1.16$ [$t(9) = 5.46$, $p < .001$]. In close correspondence, the participants expressed more confidence in their judgments in the expected within- than in the across-stream trials: The mean confidence rating values were $1.4 \pm 0.4$ versus $2.0 \pm 0.6$ [$t(9) = 4.38$, $p < .005$] and $1.2 \pm 0.3$ versus $2.1 \pm 0.5$ [$t(9) = 5.37$, $p < .001$] for expected within- versus across-stream judgments in the A-and-B-separate and A-and-B-together conditions, respectively (the rating expressing the highest confidence being 1).

Regarding the two sets of test tones (A and B), the participants judged the order between an A and a B tone more accurately and with more confidence in the A-and-B-together than in the A-and-B-separate condition: The mean numbers of correct judgments were $9.9 \pm 0.3$ versus $7.9 \pm 1.5$ [$t(9) = 4.05$, $p < .01$], and the mean confidence rating values were $1.2 \pm 0.3$ versus $2.0 \pm 0.6$ [$t(9) = 6.78$, $p < .001$] in the A-and-B-together versus the A-and-B-separate conditions, respectively.

## EXPERIMENT 2B

In Experiment 2B, MMN was used to test whether or not the contextual manipulation used in Experiment 2A promotes the same organization of the test sequences when participants perform a task that does not involve the auditory stimuli.

To this end, the sequences were set up to yield different regularities depending on whether the A and B tones were grouped together, separately from the X and Y tones (A-and-B-together condition), or segregated into separate streams, the A tones grouped with the X tones and the B tones grouped with the Y tones (A-and-B-separate condition). Different regularities were achieved by setting the duration of the A and B tones at 250 msec and the duration of the X and Y tones at 100 msec, thereby providing regularities for duration. In the A-and-B-together condition, 250 msec became the common tone duration in the A+B group, and 100 msec, that for the X+Y group. In contrast, in the A-and-B-separate condition, both tone groups (A+X and B+Y) had two equally frequent tone durations (250 and 100 msec), thereby providing a different set of regularities. Occasionally, A′ and B′ tones (probes) of 100-msec duration (randomly assuming any

of the five A-tone or five B-tone frequencies, respectively) were presented within the test sequences (Figure 4; the probes are marked by the checkered pattern). If the frequency relationships between the tones resulted in different groupings in the two experimental conditions while the participants were not focusing their attention on the auditory stimuli and this sound organization preceded the stage of the regularity detection processes underlying MMN generation, then the probe tones could be expected to elicit an MMN in the A-and-B-together, but not in the A-and-B-separate condition.

Our expectations about the MMN elicitation in Experiment 2B were based on the following considerations:

1. The A′ and B′ tones will always be grouped together with the A and B tones, respectively, because of their common frequencies. This is because frequency has been shown to be an effective cue for stream segregation (Bregman, 1990).

2. The probe tones will elicit the MMN when they violate some regularity of the group to which they belong irrespective of what regularities apply to other groups or to the full auditory sequence; no MMN is elicited by tones violating regularities of groups of which they are not a part (Ritter et al., 2000; Winkler et al., 2001).

3. Four tone sets (A, B, X, and Y) with five different tones in each were used instead of four distinct tones (as in Bregman, 1978). The reason for this change from the original paradigm was to prevent the probe tones from eliciting MMNs on the basis of their infrequent combination of frequency and duration within the sequence. Previous studies have shown that tones with infrequent combinations of two stimulus features can elicit MMN when they appear in sequences in which other combinations of the same feature levels occur frequently (Gomes, Bernstein, Ritter, Vaughan, & Miller, 1997; Sussman, Gomes, Nousak, Ritter, & Vaughan, 1998; Takegata, Paavilainen, Näätänen, & Winkler, 1999). In our design, each of the tone sets consisted of five tones of different frequencies. Using two different tone durations, tones could have any of the 10 different combinations of frequency and duration. These frequency–duration combinations appeared with approximately equal probabilities in any stream formed of two tone sets (e.g., A+X or B+Y). Thus, no frequency–duration combination could become frequent within any possible stream, thereby preventing the emergence of one or two feature-combination standards. Therefore, probe tones will not elicit MMNs by deviating from a standard feature combination.

4. A duration-regularity-violation MMN will be elicited by probe tones despite the variance in tone frequency, because MMN can be elicited by sounds deviating from the common level of an auditory feature present in all stimuli even when other features vary within the stimulus sequence (Gomes, Ritter, & Vaughan, 1995; Huotilainen et al., 1993; Nousak, Deacon, Ritter, & Vaughan, 1996; Winkler et al., 1990). This, in combination with point 2 above, implies that one can expect an MMN to be elicited by sounds deviating from a given level of an auditory feature common to members of the perceptual

group to which they belong despite variation in other stimulus features within this group.

On the basis of these considerations, the pattern of MMN responses emerging in Experiment 2B can be used to make inferences about the effects of the contextual tones on the preperceptual organization of the test sequences.

## Method

**Participants**. Ten healthy young adult participants (7 females, mean age 23.5 years), who reported having normal hearing, took part in Experiment 2B. None of the participants had taken part in Experiment 1 or 2A. Informed consent was obtained after the testing procedure was explained to them. Recordings were conducted in an electrically shielded, acoustically dampened room at the Cognitive Brain Research Unit of the University of Helsinki. The participants sat in a comfortable chair and were instructed to read a book and to ignore the auditory stimuli.

**Stimuli**. The stimuli were the same as those described in Experiment 2A, with the addition of the A′ and B′ (probe) tones, whose frequencies were the same as those of the A and B tones, respectively, but which had a duration of 100 msec. Stimulus sequences were composed as were those described for Experiment 2A, with two differences: (1) 10% of the A tones were exchanged for A′ tones, and 10% of the B tones for B′ tones; and (2) the stimulus blocks consisted of 4,000 stimuli. The tones were binaurally presented through headphones at 50 dB above the participant's hearing threshold.

**Procedure**. The participants were presented with two stimulus blocks for each experimental condition (A-and-B-separate and A-and-B-together). In addition, two control blocks (880 stimuli in each), one for each condition, were also included in the experi-

mental session. The control blocks were used to record ERP responses to the tones that served as probes in the experimental conditions in a situation in which these tones would not elicit MMN. Thus, MMN could be delineated by comparing between responses elicited by identical tones. In the control blocks, the stimulus durations were exchanged between the test and flanking tones (i.e., the A and B tones were 100 msec in duration, flanking tones 250 msec), and no rare-duration tones were presented. The order of stimulus blocks was counterbalanced across participants.

The EEG was recorded with Ag/AgCl electrodes attached to the scalp at Fz, Cz (10–20 system), LM, and RM. The common reference electrode was placed on the tip of the nose. Horizontal eye movements were recorded with a bipolar electrode pair from the outer canthi of the two eyes. Vertical ocular potentials were recorded between electrodes placed above and below the left eye. The EEG was digitized (using SynAmps amplifiers) at a 250-Hz sampling rate (0–40 Hz filtering limits) and off-line band-pass filtered between 1.5 and 30 Hz. Epochs of 450-msec duration starting 50 msec before and ending 400 msec after the onset of each stimulus were collected. Those containing a change exceeding 100 $\mu$V at any recording channel were rejected from subsequent processing.

Epochs were averaged separately for each condition and stimulus type (A and B, A′ and B′, and X and Y, separately collapsed). MMN elicitation was tested from the difference waveforms calculated by subtracting the collapsed averaged ERPs elicited by the A and B tones in the control blocks from those elicited by the physically identical probe tones in the corresponding experimental condition. (In the control blocks, the duration of the A and B tones was 100 msec, equal to the duration of the A′ and B′ tones in the experimental conditions.) The MMN amplitude was estimated by calculating the mean probe − control difference amplitude in the 240- to 280-msec interval. The measurement interval was determined by
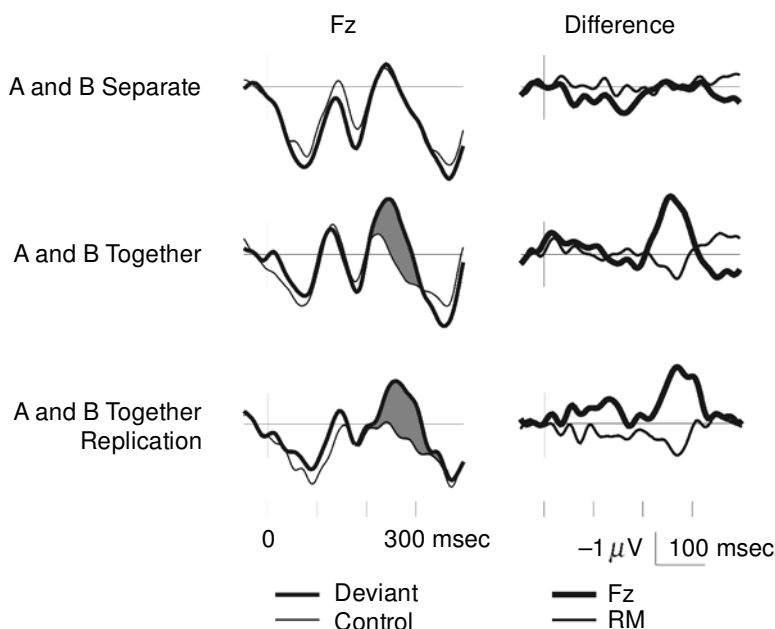


**Figure 5. Experiments 2B and 2C: The left panels show the grand-average frontal (Fz, left) responses to probe (A′ and B′ collapsed, thick line) and control tones (reversed-blocks A and B collapsed, thin line) in the A-and-B-separate and A-and-B-together conditions of Experiment 2B and of Experiment 2C (A-and-B-together condition only). MMN responses are marked by the shaded area between the probe- and control-tone ERPs. The corresponding probe − control difference waves at Fz (thick line) and RM (thin line) are shown in the right panels.**

finding the peak of the negative-polarity wave in the grand-average probe − control difference response of the A-and-B-together condition. The presence of the MMN component was tested using one-sample $t$ tests for the frontal (Fz) and RM responses.

## Results

Figure 5 (top two panels) presents the ERP responses elicited by the control and probe tones in the two experimental conditions. In the A-and-B-separate condition, no difference between the probe and control ERPs can be observed in the 200- to 300-msec latency range of the difference waveforms [values of $t(9) < 1$ were obtained for both Fz and RM]. Thus, it appears that the probe tones elicited no MMN in this condition.

In the A-and-B-together condition, a negative probe − control difference wave was elicited between 200 and 300 msec from stimulus onset. This difference was significantly different from zero at Fz [$-1.59 \pm 0.27$ $\mu$V, mean $\pm$ $SEM$; $t(9) = -5.89$, $p < .001$] as well as at RM [$0.56 \pm 0.17$ $\mu$V; $t(9) = 3.37$, $p < .01$], where it appeared with a reversed polarity. Therefore, one can conclude that the probe tones elicited the MMN in this condition.

## EXPERIMENT 2C

In Experiments 2A and 2B, the frequency step separating members of the four tonal groups was uniformly 11 Hz. Because of this, the X and Y ranges, as well as the separation between the individual tones within these ranges, were perceptually different between the conditions. We replicated Experiments 2A and the A-and-B-together condition of Experiment 2B, exchanging the original X and Y tones of this condition for ones whose frequency steps were proportionally equal to those of the X and Y tones presented in the A-and-B-separate conditions of Experiments 2A and 2B. The A-and-B-separate condition of Experiment 2B was not repeated, because the results of this condition obtained in Experiment 2B can be compared with the results of Experiment 1C. In summary, the difference between the previous two experiments and the present one is that, whereas in the previous experiments, the frequency steps between members of the X and Y tone sets were numerically equal in the two test conditions, in Experiment 2C the log frequencies of the steps were equalized between conditions.

## Method

**Participants**. Eight healthy young adult participants (5 females, mean age 28.3 years) with reportedly normal hearing took part in the replication of Experiment 2A. Six of them (4 females, mean age 27.2 years) also took part in the replication of the A-and-B-together condition of Experiment 2B, which was carried out before the active behavioral experiment. The experiment was conducted in the laboratories of the Cognitive Brain Research Unit of the University of Helsinki.

**Procedure**. The experimental procedure, data recording, and analysis were identical to those described for Experiments 2A and 2B, except that the participants were not asked to rate the confidence of their judgments (as they had been in Experiment 2A). The main change from the original experiments was that the frequencies of the X- and Y-together tones were calculated by multiplying the

frequency value of each of the original X- and Y-separate tones with a uniform 3.36 factor. Thus, the range of the X-together tones became 1059.5–1207.5 Hz, and that of the Y-together tones, 2415–2563 Hz.

## Results

As in Experiment 1A, the participants performed more accurately in the within-stream than in the across-stream trials. The mean ($\pm SD$) numbers of correct judgments were $9.3 \pm 1.39$ versus $7.5 \pm 1.6$ [$t(7) = 2.97$, $p < .02$] in the A-and-B-separate condition and $9.5 \pm 1.4$ versus $7.5 \pm 1.51$ [$t(7) = 2.94$, $p < .02$] in the A-and-B-together condition, in the within-stream trials and the across-stream trials, respectively. Also, the order between the A and B tones was judged more accurately [$t(7) = 3.35$, $p < .01$] in the A-and-B-together ($9.5 \pm 1.4$) than in the A-and-B-separate ($7.5 \pm 1.6$) condition.

Figure 5 (bottom panel) shows the ERP responses elicited by the control and probe tones in the replication of the A-and-B-together condition of Experiment 2B. A highly significant MMN response was elicited by the occasional probe tones [Fz, $-1.48 \pm 0.32$ ($\pm SEM$), $t(5) = 6.41$, $p < .01$; RM, $0.79 \pm 0.16$, $t(5) = 4.86$, $p < .01$].

## Discussion of Experiments 2A, 2B, and 2C

The accuracy of order judgments and the participants' confidence in their judgments in Experiments 2A and 2C provided evidence that in the A-and-B-separate condition, the A and B tones joined different streams: The A and X-separate tones formed one stream, whereas the B and Y-separate tones formed another. In the A-and-B-together condition, the participants' confidence and accuracy of order judgments suggest that the A and B tones joined together to form one stream that was separate from the stream created by the X-together and Y-together tones. That is, it appears that the test sequences were organized in two different ways, depending on the frequency proximity of the flanking tones.

The most striking evidence for this contextual grouping effect is that the accuracy and confidence of order judgments of two successive test tones—one taken from the A set and the other from the B set—was significantly influenced by the frequency of the flanking tones. That is, the judgment of the order between the same two tones (the A and B tones not having changed between the two conditions) was made easier or more difficult according to the frequency proximity of the contextual tones. Similarly, the frequency relationship between the test and flanking tones determined the elicitation of MMN in Experiment 2B (see also the results of Experiment 2C). MMN was elicited in the A-and-B-together but not in the A-and-B-separate condition. The elicitation of MMN was governed by the groupings resulting from the different frequency relationships between the test and the flanking tones in the two experimental conditions. In the A-and-B-together condition, the probe tones violated the duration-constancy regularity within the A+B stream and, therefore, elicited an MMN. In contrast, in the A-and-B-separate condition, the duration of the probe

tones appeared frequently within the streams to which they belonged (A+X or B+Y). Therefore, in this case, the duration of the probe tones was compatible with the regularities of their stream and, consequently, no MMN was elicited by these tones. This pattern of results supports our hypothesis that context effects on grouping can occur even when attention is not focused on the auditory stimuli.

The congruity between the perceptual grouping of the test sequences in Experiments 2A and 2C and the grouping of the auditory sequences inferred from the MMN results in Experiments 2B and 2C suggests that the contextual sounds had a similar effect on the organization of the sound sequences whether or not attention was focused on the sounds. Our results also corroborate previous studies that found a close correspondence between MMN and sound perception (Amenedo & Escera, 2000; Tiitinen, May, Reinikainen, & Näätänen, 1994; Winkler, Reinikainen, & Näätänen, 1993; Winker, Tervaniemi, & Näätänen, 1997; for reviews, see Näätänen & Alho, 1997; Näätänen & Winkler, 1999). Furthermore, it should be noted that because the order of tones was randomized in Experiment 2 (except for the last two tones of the order-judgment sequences), no regular rhythm could emerge in the segregated streams. Therefore, the present stream-segregation results contrasts the rhythm-based explanation of auditory stream segregation (M. R. Jones et al., 1981).

## GENERAL DISCUSSION

The present experiments were conducted to determine whether the context that influences the perceptual organization of sounds would similarly affect the organization of auditory stimulus sequences when attention was directed away from the sounds. We reported results from two sets of experiments, which tested the effects of the contextual tones on the organization of auditory sequences. Experiment 1A showed a differential effect of random tones on detecting targets in an auditory oddball paradigm, which was determined by whether the random tones were grouped together or separately from the oddball sequence (for compatible results and conclusions, see D. M. Jones et al., 1997). Experiment 1B showed corresponding effects on MMN elicitation. That is, in a passive situation, deviant tones elicited the MMN when the oddball sequence was segregated from the random intervening tones, but not when these tones were grouped together. Experiments 2A and 2C demonstrated that the same two sets of tones can be grouped together or separately on the basis of the frequency relationship between these and the additional contextual tones that are presented with them (for compatible results, see Bregman, 1978). Experiments 2B and 2C yielded matching MMN results when the participants ignored the auditory stimuli: The contextual manipulation that was shown to affect the perceptual organization of the tone sequences in Experiments 2A and 2C also determined the elicitation of the MMN by probe tones presented among the test tones.

In two different paradigms, we found good correspondence between contextual effects in active and passive situations (i.e., when the participants performed a task with the test sounds and when they were instructed to read a book and ignore the sounds, respectively). One explanation of this correspondence is that the perceptual effects tested in the present experiments are based on preattentive sound organization processes. This interpretation is compatible with Bregman's view that "the phenomenon of stream segregation [is] largely the result of the grouping performed by a pre-attentive mechanism" (Bregman, 1990, p. 206). Furthermore, a number of studies suggested that the processes providing the basis of MMN elicitation also determine voluntary task performance and conscious perception (see Näätänen, 1992; Näätänen & Alho, 1997; Näätänen & Winkler, 1999; Novak, Ritter, & Vaughan, 1992).

One might argue against the explanation presented above by questioning whether the effects obtained with MMN in the present passive conditions indeed index a preattentive stage of auditory information processing. One argument is that the MMN amplitude itself is not fully independent of attention (see, e.g., Näätänen, Paavilainen, Tiitinen, Jiang, & Alho, 1993; Woldorff, Hackley, & Hillyard, 1991). However, the intercondition differences in MMN elicitation could not result from an attentional effect on the MMN itself, unless one assumes that, despite the unchanging instructions, our participants independently decided to attend the sounds in some conditions (the segregated condition in Experiment 1B and the A-and-B-together condition in Experiments 2B and 2C) but not in the others. In addition, the results of Experiment 1D as well as those of Sussman, Winkler, and Wang (in press) strongly argue against a direct attentional effect on MMN.

Another argument against the interpretation of the MMN results might be that, despite the instruction to ignore the sounds, the participants divided their attention between reading and monitoring the sounds, either constantly or intermittently, by selecting one and then the other modality throughout all stimulus blocks. Thus, MMN could have shown the consequence of sound organization processes that required focused attention themselves. Support for this argument is provided by the results of Sussman et al. (1998a), who demonstrated that the voluntary organization of a sound sequence can affect which regularities the MMN-generating processes is set for. Sussman et al. (1998a) presented a sequence of high and low tones alternating at a relatively slow pace. Both high and low sequences, separately, consisted of repetitive three-tone patterns, which were occasionally broken by deviant tone patterns (reversing the tone order of the repetitive pattern). When the participants ignored the auditory stimuli (by reading a book), no MMN was elicited by the deviant patterns. However, in the attend condition, when the participants were instructed to actively segregate the sounds, selecting the high tones and ignoring the low ones, the deviant patterns elicited MMNs. Thus, MMN elicitation in Sussman et al.'s (1998a)

experiment was affected by attentive organization of the sound sequence.

However, the results of Experiments 1C and 1D makes the alternative explanation based on divided attention highly unlikely. In Experiment 1C, the participants in the ignore group did not remember which of two sounds they had heard more often in a stimulus sequence (similar to those used in Experiments 1A and 1B) that was presented to them while they were reading a book. If they were constantly or intermittently attending the sounds throughout the stimulus blocks, they should have been able to make the simple distinction that one tone occurred more frequently than another in the sequence, as was shown by the high-level performance found for the participants who were instructed to attend the sounds (i.e., the attend group in Experiment 1C). The possibility of a graded effect of attention on auditory stream segregation was ruled out in Experiment 1D. The results of that experiment showed that the MMN amplitude was not affected by differences in the load of a visual primary task, despite the large effect of memory load on task performance. The tone sequences in Experiment 1D were identical to the ones in the segregated condtion of Experiment 1B.

An additional argument against the divided-attention alternative was provided by the present ERP results. If the tones were attended in Experiments 1B, 1D, 2B, and 2C, the probe tones would have elicited the N2b component, an ERP correlate of the detection of attended deviant tones, which is a centrally maximal negative potential closely following the MMN component in the deviant ERP response. Unlike the MMN, however, N2b shows no polarity reversal at the mastoid leads. No such waveform can be observed in either experiment or condition (see the frontal deviant ERPs in Figures 2, 3, and 5). In conclusion, the absence of N2b to deviant sounds, together with the results of Experiment 1C, rules out the explanation that the participants attentively monitored the auditory stimulation throughout the stimulus blocks. Consequently, we conclude that the auditory context has effects that do not require attention to be constantly or intermittently focused on the sounds.

These conclusions are fully compatible with the hypothesis that object formation and the grouping of objects on the basis of similarity can occur without focused attention (Duncan & Humphreys, 1989; Kubovy et al., 1999; Näätänen, 1992; Näätänen & Winkler, 1999). However, attentional processes may still be required to initiate the grouping processes underlying the present contextual effects, but not, as we have shown, to maintain the grouped organization. It is possible to assume that the participants attended the auditory stimuli for at least a few seconds at the beginning of each stimulus block, which allowed them to create object files for the frequent sounds and form groups (streams) on the basis of these object files. This account (i.e., that participants engaged in a nonauditory primary task do not actively monitor the sounds throughout the stimulus blocks) would not con-

tradict the results of Experiments 1C and 1D or those of Carlyon et al. (2001) and is compatible with Treisman's (1993) views.

Nevertheless, the present results demonstrated that a substantial part of contextual processing in audition does not require attention to be focused on the sounds. Although the processes involved in the analysis and maintenance of the auditory scene are probably not automatic in the strict sense, neither do they require continuous attention. In some cases, sound organization can be altered by top-down effects (see Sussman et al., 1998a). However, sound organization processes are probably always in operation. One might term as *default processes* such processes that proceed even in the absence of attention, but which, under certain circumstances, can be reached by top-down effects. The analogy is taken from those computer functions that are executed even when the user gives no explicit instructions about them, but which can be altered by user interaction (e.g., the boot-up process of most operating systems will load the software with the most common mode unless the user explicitly selects an alternative mode).

If a large part of contextual processing occurs pre-attentively, then contextual information must be accessible by at least some of the processes operating within the large-capacity system. This information is required to evaluate incoming sounds in relation to the history of the auditory stimulation, parse the composite input into its constituents, or determine the source of a new auditory event. Therefore, a representation of the auditory context should include characteristics of the currently active sources in the auditory environment (spectral, spatial, and temporal features) as well as their relations to each other. This set of interrelated information, which subserves sound organization, can be regarded as a neural model of the auditory environment.

Winkler, Karmos, and Näätänen (1996) suggested that the MMN component is the product of a process whose primary function is to adjust the model when the previously detected regularities are violated (see also Sinkkonen, 1999; Winkler & Czigler, 1998). Since the auditory environment is constantly changing, the model must be adaptable. The results of a large number of MMN studies have demonstrated that changes in the regularities can affect the model along several different times scales. Even a few sounds, presented within an interval of a few hundred milliseconds, can set up a new regularity (Cowan et al., 1993; Schröger, 1997) or change an existing one (Winkler, Karmos, & Näätänen, 1996). Other changes, such as the broader temporal context effects discussed in the present study, require several seconds to be encoded in the structure of the model (see also Cowan et al., 1993; Winkler, Cowan, et al., 1996). Changes of longer term that take place within a few hours have been observed in studies investigating the effects of learning a difficult discrimination on MMN elicitation (Kraus et al., 1996; Näätänen, Schröger, Karakas, Tervaniemi, & Paavilainen, 1993). Finally, true long-term changes, such as the effects

of learning to speak a language (Cheour et al., 1998; Winkler et al., 1999) or acquiring musical expertise (Koelsch, Schröger, & Tervaniemi, 1999; Tervaniemi, Rytkönen, Schröger, Ilmoniemi, & Näätänen, 2001), were also found to influence the detection of violations of auditory regularities.

MMN studies demonstrated the presence of an auditory model when the participants did not pay attention to the sounds (for a review, see Näätänen & Winkler, 1999). However, the contents of this model can be modified by top-down processes (Sussman, Winkler, Huotilainen, Ritter, & Näätänen, 2002). Therefore, it can be suggested that the model may serve as a meeting point between stimulus-driven and top-down processes, integrating the information collected about the current acoustic environment.

The outcome of early auditory processing appears to be an interpretation of the ever-changing composite input. The ease with which listeners are able to select distinct sources of information supports the hypothesis that a large part of the processing that underlies the sorting of the spectrally and temporally overlapping patterns of the acoustic signal does not depend on limited attentional resources. In order to provide the information required for selection of auditory sources, the preattentive auditory system constantly monitors the full auditory environment, independently of the focus of one's attention. Ignoring the information delivered by a given source should not prevent some monitoring of the physical characteristics of that source, or else other sources could not be identified and selected. In conclusion, correct identification of even a single sound within a complex acoustic environment requires preattentive contextual processing.

The present results provided support for the notion of preattentive contextual auditory processing. Using the MMN, we have shown that the auditory context affects the preattentively maintained organization of sound sequences. Our behavioral results showed corresponding contextual effects on perception.

## REFERENCES

ALAIN, C., ACHIM, A., & RICHER, F. (1993). Perceptual context and the selective attention effect on auditory event-related brain potentials. *Psychophysiology*, **30**, 572-580.

ALAIN, C., & WOODS, D. L. (1994). Signal clustering modulates auditory cortical activity in humans. *Perception & Psychophysics*, **56**, 501-516.

ALHO, K. (1995). Cerebral generators of mismatch negativity (MMN) and its magnetic counterpart (MMNm) elicited by sound changes. *Ear & Hearing*, **16**, 38-51.

ALHO, K., WOODS, D. L., ALGAZI, A., & NÄÄTÄNEN, R. (1992). Intermodal selective attention II: Effects of attentional load on processing of auditory and visual stimuli in central space. *Electroencephalography & Clinical Neurophysiology*, **82**, 356-368.

AMENEDO, E., & ESCERA, C. (2000). The accuracy of sound duration representation in the human brain determines the accuracy of behavioural perception. *European Journal of Neuroscience*, **12**, 2570-2574.

ANSTIS, S., & SAIDA, S. (1985). Adaptation to auditory streaming of frequency-modulated tones. *Journal of Experimental Psychology: Human Perception & Performance*, **11**, 257-272.

BAYLIS, G. C., & DRIVER, J. (1992). Visual parsing and response competition: The effect of grouping factors. *Perception & Psychophysics*, **51**, 145-162.

BEAUVOIS, M. W., & MEDDIS, R. (1991). A computer model of auditory stream segregation. *Quarterly Journal of Experimental Psychology*, **43A**, 517-542.

BREGMAN, A. S. (1978). Auditory streaming: Competition among alternative organizations. *Perception & Psychophysics*, **23**, 391-398.

BREGMAN, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound.* Cambridge, MA: MIT Press.

BREGMAN, A. S., & CAMPBELL, J. (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of Experimental Psychology*, **89**, 244-249.

BREGMAN, A. S., & RUDNICKY, A. (1975). Auditory segregation: Stream or streams? *Journal of Experimental Psychology: Human Perception & Performance*, **1**, 263-267.

BROADBENT, D. E. (1958). *Perception and communication.* London: Pergamon.

CARLYON, R. P., CUSACK, R., FOXTON, J. M., & ROBERTSON, I. H. (2001). Effects of attention and unilateral neglect on auditory stream segregation. *Journal of Experimental Psychology: Human Perception & Performance*, **27**, 115-117.

CHEOUR, M., CEPONIENE, R., LEHTOKOSKI, A., LUUK, A., ALLIK, J., ALHO, K., & NÄÄTÄNEN, R. (1998). Development of language-specific phoneme representations in the infant brain. *Nature Neuroscience*, **1**, 351-353.

COWAN, N. (1984). On short and long auditory stores. *Psychological Bulletin*, **96**, 341-370.

COWAN, N. (1995). *Attention and memory: An integrated framework.* Oxford: Oxford University Press.

COWAN, N., WINKLER, I., TEDER, W., & NÄÄTÄNEN, R. (1993). Memory prerequisites of the mismatch negativity in the auditory event-related potential (ERP). *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 909-921.

DANNENBRING, G. L., & BREGMAN, A. S. (1976). The effect of silence on auditory stream segregation. *Journal of the Acoustical Society of America*, **59**, 987-989.

DEUTSCH, D. (1970). Tones and numbers: Specificity of interference in short-term memory. *Science*, **168**, 1604-1605.

DEUTSCH, D. (1978). Interference in pitch memory as a function of ear of input. *Quarterly Journal of Experimental Psychology*, **30A**, 283-287.

DEUTSCH, D. (1984). Memory for nonverbal auditory information: A link between behavioral and physiological studies. In L. R. Squire & N. Butters (Eds.), *Neuropsychology of memory* (pp. 45-54). New York: Guilford.

DOWLING, W. J. (1973). Rhythmic groups and subjective chunks in memory for melodies. *Perception & Psychophysics*, **14**, 37-40.

DRIVER, J. (1996). Attention and segmentation. *Psychologist*, **9**, 119-123.

DRIVER, J., & BAYLIS, G. C. (1989). Movement and visual attention: The spotlight metaphor breaks down. *Journal of Experimental Psychology: Human Perception & Performance*, **15**, 448-456.

DRIVER, J., & MATTINGLEY, J. B. (1998). Parietal neglect and visual awareness. *Nature Neuroscience*, **1**, 17-22.

DUNCAN, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General*, **113**, 501-517.

DUNCAN, J. (1993). Selection of input and goal in the control of behaviour. In A. Baddeley & L. Weiskrantz (Eds.), *Attention: Selection, awareness, & control. A Tribute to Donald Broadbent* (pp. 53-71). Oxford: Oxford University Press, Clarendon Press.

DUNCAN, J., & HUMPHREYS, G. (1989). Visual search and stimulus similarity. *Psychological Review*, **96**, 433-458.

GLENBERG, A. M. (1987). Temporal context and memory. In D. S. Gorfein & R. R. Hoffman (Eds.), *Memory and learning: The Ebbinghaus Centennial Conference* (pp. 173-190). Hillsdale, NJ: Erlbaum.

GOMES, H., BERNSTEIN, R., RITTER, W., VAUGHAN, H. G., JR., & MILLER, J. (1997). Storage of feature conjunctions in transient auditory memory. *Psychophysiology*, **34**, 712-716.

GOMES, H., RITTER, W., & VAUGHAN, H. G., JR. (1995). The nature of preattentive storage in the auditory system. *Journal of Cognitive Neuroscience*, **7**, 81-94.

HANDEL, S. (1993). The effect of tempo and tone duration on rhythm discrimination. *Perception & Psychophysics*, **54**, 370-382.

HANSEN, J. C., & HILLYARD, S. A. (1980). Endogenous brain potentials associated with selective auditory attention. *Electroencephalography & Clinical Neurophysiology*, **49**, 277-290.

HAWKINS, H. L., & PRESSON, J. C. (1977). Masking and preperceptual selectivity in auditory recognition. In S. Dornic (Ed.), *Attention and performance VI* (pp. 195-211). Hillsdale, NJ: Erlbaum.

HUOTILAINEN, M., ILMONIEMI, R. J., LAVIKAINEN, J., TIITINEN, H., ALHO, K., SINKKONEN, J., KNUUTILA, J., & NÄÄTÄNEN, R. (1993). Interaction between representations of different features of auditory sensory memory. *NeuroReport*, **4**, 1279-1281.

IDSON, W. L., & MASSARO, D. W. (1976). Cross-octave masking of single tones and musical sequences: The effects of structure on auditory recognition. *Perception & Psychophysics*, **19**, 155-175.

JONES, D. [M.], ALFORD, A., BRIDGES, A., TREMBLAY, S., & MACKEN, B. (1999). Organizational factors in selective attention: The interplay of acoustic distinctiveness and auditory streaming in the irrelevant sound effect. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **25**, 464-473.

JONES, D. M., & MACKEN, W. J. (1995). Organizational factors in the effect of irrelevant speech: The role of spatial location and timing. *Memory & Cognition*, **23**, 192-200.

JONES, D. M., MACKEN, W. J., & HARRIES, C. (1997). Disruption of short-term recognition memory for tones: Streaming or interference? *Quarterly Journal of Experimental Psychology*, **50A**, 337-357.

JONES, M. R. (1976). Time, our lost dimension: Toward a new theory of perception, attention, and memory. *Psychological Review*, **83**, 323-355.

JONES, M. R., KIDD, G., & WETZEL, R. (1981). Evidence for rhythmic attention. *Journal of Experimental Psychology: Human Perception & Performance*, **7**, 1059-1073.

JONES, M. R., MASER, D. J., & KIDD, G. R. (1978). Rate and structure in memory for auditory patterns. *Memory & Cognition*, **6**, 246-258.

KAHNEMAN, D., TREISMAN, A., & GIBBS, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, **24**, 175-219.

KALLMAN, H. J., & MORRIS, M. D. (1984). Backward recognition masking as a function of ear of mask presentation. *Perception & Psychophysics*, **35**, 379-384.

KOELSCH, S., SCHRÖGER, E., & TERVANIEMI, M. (1999). Superior preattentive auditory processing in musicians. *NeuroReport*, **10**, 1309-1313.

KRAUS, N., MCGEE, T. J., CARRELL, T. D., KING, C., TREMBLAY, K., & NICOL, T. (1996). Central auditory system plasticity associated with speech discrimination training. *Journal of Cognitive Neuroscience*, **7**, 25-32.

KUBOVY, M., COHEN, D. J., & HOLLIER, J. (1999). Feature integration that routinely occurs without focal attention. *Psychonomic Bulletin & Review*, **6**, 183-203.

MACK, A., TANG, B., TUMA, R., KAHN, S., & ROCK, I. (1992). Perceptual organization and attention. *Cognitive Psychology*, **24**, 475-501.

MASSARO, D. W. (1975). *Experimental psychology and information processing*. Chicago: Rand McNally.

MOORE, C. M., & EGETH, H. (1997). Perception without attention: Evidence of grouping under conditions of inattention. *Journal of Experimental Psychology: Human Perception & Performance*, **23**, 339-352.

NÄÄTÄNEN, R. (1985). Stimulus processing: Reflections in event-related potentials, magnetoencephalogram and regional cerebral blood flow. In M. I. Posner & O. S. M. Marin (Eds.), *Attention and performance XI* (pp. 355-373). Hillsdale, NJ: Erlbaum.

NÄÄTÄNEN, R. (1990). The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function. *Behavioral & Brain Sciences*, **13**, 201-288.

NÄÄTÄNEN, R. (1992). *Attention and brain function*. Hillsdale, NJ: Erlbaum.

NÄÄTÄNEN, R., & ALHO, K. (1997). Mismatch negativity (MMN)—the measure for central sound representation accuracy. *Audiology & Neuro-Otology*, **2**, 341-353.

NÄÄTÄNEN, R., GAILLARD, A. W. K., & MÄNTYSALO, S. (1978). Early

selective attention effect on evoked potential reinterpreted. *Acta Psychologica*, **42**, 313-329.

NÄÄTÄNEN, R., PAAVILAINEN, P., ALHO, K., REINIKAINEN, K., & SAMS, M. (1989). Do event-related potentials reveal the mechanism of the auditory sensory memory in the human brain? *Neuroscience Letters*, **98**, 217-221.

NÄÄTÄNEN, R., PAAVILAINEN, P., TIITINEN, H., JIANG, D., & ALHO, K. (1993). Attention and mismatch negativity. *Psychophysiology*, **30**, 436-450.

NÄÄTÄNEN, R., SCHRÖGER, E., KARAKAS, S., TERVANIEMI, M., & PAAVILAINEN, P. (1993). Development of a memory trace for a complex sound in the human brain. *NeuroReport*, **4**, 503-506.

NÄÄTÄNEN, R., TERVANIEMI, M., SUSSMAN, E., PAAVILAINEN, P., & WINKLER, I. (2001). Pre-attentive cognitive processing ("primitive intelligence") in the auditory cortex as revealed by the mismatch negativity (MMN). *Trends in Neurosciences*, **24**, 283-288.

NÄÄTÄNEN, R., & WINKLER, I. (1999). The concept of auditory stimulus representation in cognitive neuroscience. *Psychological Bulletin*, **125**, 826-859.

NOUSAK, J. M. K., DEACON, D., RITTER, W., & VAUGHAN, H. G., JR. (1996). Storage of information in transient auditory memory. *Cognitive Brain Research*, **4**, 305-317.

NOVAK, G. P., RITTER, W., & VAUGHAN, H. G., JR. (1992). The chronometry of attention-modulated processing and automatic mismatch detection. *Psychophysiology*, **29**, 412-430.

PECHMANN, T., & MOHR, G. (1992). Interference in memory for tonal pitch: Implications for a working-memory model. *Memory & Cognition*, **20**, 314-320.

PICTON, D. W., ALAIN, C., OTTEN, L., & RITTER, W. (2000). Mismatch negativity: Different water in the same river. *Audiology & Neuro-Otology*, **5**, 111-139.

RINNE, T., ANTILA, S., & WINKLER, I. (2001). MMN is unaffected by top-down predictive information. *NeuroReport*, **12**, 2209-2213.

RITTER, W., DEACON, D., GOMES, H., JAVITT, D. C., & VAUGHAN, H. G., JR. (1995). The mismatch negativity of event-related potentials as a probe of transient auditory memory: A review. *Ear & Hearing*, **16**, 52-67.

RITTER, W., GOMES, H., COWAN, N., SUSSMAN, E., & VAUGHAN, H. G., JR. (1998). Reactivation of a dormant representation of an auditory stimulus feature. *Journal of Cognitive Neuroscience*, **10**, 605-614.

RITTER, W., SUSSMAN, E., DEACON, D., COWAN, N., & VAUGHAN, H. G., JR. (1999). Two cognitive systems simultaneously prepared for opposite events. *Psychophysiology*, **36**, 835-838.

RITTER, W., SUSSMAN, E., & MOLHOLM, S. (2000). Evidence that the mismatch negativity system works on the basis of objects. *NeuroReport*, **11**, 61-63.

SAMS, M., PAAVILAINEN, P., ALHO, K., & NÄÄTÄNEN, R. (1985). Auditory frequency discrimination and event-related potentials. *Electroencephalography & Clinical Neurophysiology*, **62**, 437-448.

SCHERG, M., VAJSAR, J., & PICTON, T. W. (1989). A source analysis of the late human auditory evoked potentials. *Journal of Cognitive Neuroscience*, **1**, 336-355.

SCHRÖGER, E. (1997). On the detection of auditory deviants: A preattentive activation model. *Psychophysiology*, **34**, 245-257.

SHINOZAKI, N., YABE, H., SATO, Y., SUTOH, T., HIRUMA, T., NASHIDA, T., & KANEKO, S. (2000). Mismatch negativity (MMN) reveals sound grouping in the human brain. *NeuroReport*, **11**, 1597-1601.

SINKKONEN, J. (1999). Information and resource allocation. In R. Baddeley, P. Hancock, & P. Földiák (Eds.), *Information theory and the brain* (pp. 241-254). Cambridge: Cambridge University Press.

STOFFGREN, T. A., & BRADY, B. G. (2001). On specification and the senses. *Behavioral & Brain Sciences*, **24**, 195-222.

SUSSMAN, E., CEPONIENE, R., SHESTAKOVA, A., NÄÄTÄNEN, R., & WINKLER, I. (2001). Auditory stream segregation processes operate similarly in school aged children as adults. *Hearing Research*, **153**, 108-114.

SUSSMAN, E., GOMES, H., NOUSAK, J. M., RITTER, W., & VAUGHAN, H. G., JR. (1998). Feature conjunctions and auditory sensory memory. *Brain Research*, **793**, 95-102.

SUSSMAN, E., RITTER, W., & VAUGHAN, H. G., JR. (1998a). Attention affects the organization of auditory input associated with the mismatch negativity system. *Brain Research*, **789**, 130-138.

SUSSMAN, E., RITTER, W., & VAUGHAN, H. G., JR. (1998b). Predictability of stimulus deviance and the mismatch negativity. *NeuroReport*, **9**, 4167-4170.

SUSSMAN, E., RITTER, W., & VAUGHAN, H. G., JR. (1999). An investigation of the auditory streaming effect using event-related brain potentials. *Psychophysiology, 36*, 22-34.

SUSSMAN, E., WINKLER, I., HUOTILAINEN, M., RITTER, W., & NÄÄTÄNEN, R. (2002). Top-down effects on stimulus-driven auditory organization. *Cognitive Brain Research*, **13**, 393-405.

SUSSMAN, E., WINKLER, I., & SCHRÖGER, E. (in press). Top-down control over involuntary attention-switching in the auditory modality. *Psychonomic Bulletin & Review*.

SUSSMAN, E., WINKLER, I., & WANG, W. J. (in press). MMN and attention: Competition for deviance detection. *Psychophysiology*.

TAKEGATA, R., PAAVILAINEN, P., NÄÄTÄNEN, R., & WINKLER, I. (1999). Independent processing of changes in auditory single features and feature conjunctions in humans as indexed by the mismatch negativity. *Neuroscience Letters*, **226**, 109-112.

TERVANIEMI, M., RYTKÖNEN, M., SCHRÖGER, E., ILMONIEMI, R., & NÄÄTÄNEN, R. (2001). Superior formation of cortical memory traces for melodic patterns in musicians. *Learning & Memory*, **8**, 295-300.

TIITINEN, H., MAY, P., REINIKAINEN, K., & NÄÄTÄNEN, R. (1994). Attentive novelty detection in humans is governed by pre-attentive sensory memory. *Nature*, **372**, 90-92.

TREIJO, L. J., RYAN-JONES, D. L., & KRAMER, A. F. (1995). Attentional modulation of the mismatch negativity elicited by frequency differences between binaurally presented tone bursts. *Psychophysiology*, **32**, 319-328.

TREISMAN, A. (1982). Perceptual grouping and attention in visual search for features and for objects. *Journal of Experimental Psychology: Human Perception & Performance*, **8**, 194-214.

TREISMAN, A. (1992). Representing visual objects. In D. Meyer & S. Kornblum (Eds.), *Attention and performance XIV* (pp. 163-175). Hillsdale, NJ: Erlbaum.

TREISMAN, A. (1993). The perception of features and objects. In A. Baddeley & L. Weiskrantz (Eds.), *Attention: Selection, awareness, & control. A tribute to Donald Broadbent* (pp. 5-35). Oxford: Oxford University Press, Clarendon Press.

VAN NOORDEN, L. P. A. S. (1975). *Temporal coherence in the perception of tone sequences*. Unpublished doctoral dissertation, Eindhoven University of Technology.

WATTER, S., GEFFEN, G. M., & GEFFEN, L. B. (2001). The *n*-back as a dual-task: P300 morphology under divided attention. *Psychophysiology*, **38**, 998-1003.

WINKLER, I., COWAN, N., CSÉPE, V., CZIGLER, I., & NÄÄTÄNEN, R. (1996). Interactions between transient and long-term auditory memory as reflected by the mismatch negativity. *Journal of Cognitive Neuroscience*, **8**, 403-415.

WINKLER, I., & CZIGLER, I. (1998). Mismatch negativity: Deviance detection or the maintenance of the "standard." *NeuroReport*, **9**, 3809-3813.

WINKLER, I., KARMOS, G., & NÄÄTÄNEN, R. (1996). Adaptive modeling of the unattended acoustic environment reflected in the mismatch negativity event-related potential. *Brain Research*, **742**, 239-252.

WINKLER, I., KUJALA, T., TIITINEN, H., SIVONEN, P., ALKU, P., LEHTOKOSKI, A., CZIGLER, I., CSÉPE, V., ILMONIEMI, R. J., & NÄÄTÄNEN, R. (1999). Brain responses reveal learning of foreign language phonemes. *Psychophysiology*, **36**, 638-642.

WINKLER, I., PAAVILAINEN, P., ALHO, K., REINIKAINEN, K., SAMS, M., & NÄÄTÄNEN, R. (1990). The effect of small variation of the frequent auditory stimulus on the event-related brain potential to the infrequent stimulus. *Psychophysiology*, **27**, 228-235.

WINKLER, I., REINIKAINEN, K., & NÄÄTÄNEN, R. (1993). Event-related brain potentials reflect traces of echoic memory in humans. *Perception & Psychophysics*, **53**, 443-449.

WINKLER, I., SCHRÖGER, E., & COWAN, N. (2001). The role of large-scale perceptual organization in the mismatch negativity event-related brain potential. *Journal of Cognitive Neuroscience*, **13**, 1-13.

WINKLER, I., TERVANIEMI, M., & NÄÄTÄNEN, R. (1997). Two separate codes for missing fundamental pitch in the auditory cortex. *Journal of the Acoustical Society of America*, **102**, 1072-1082.

WOLDORFF, M. G., HACKLEY, S. A., & HILLYARD, S. A. (1991). The effects of channel-selective attention on the mismatch negativity wave elicited by deviant tones. *Psychophysiology, 28*, 30-42.

WOLDORFF, M. G., HILLYARD, S. A., GALLEN, C. C., HAMPSON, S. R., & BLOOM, F. E. (1998). Magnetoencephalographic recordings demonstrate attentional modulation of mismatch-related neural activity in human auditory cortex. *Psychophysiology*, **35**, 283-292.

WOLFE, J. M. (1994). Guided search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review*, **1**, 202-238.

WOLFE, J. M., CAVE, K. R., & FRANZEL, S. L. (1989). Guided search: An alternative to the feature integration model of visual search. *Journal of Experimental Psychology: Human Perception & Performance*, **15**, 419-433.

YABE, H., WINKLER, I., CZIGLER, I., KOYAMA, S., KAKIGI, R., SUTOH, T., HIRUMA, T., & KANEKO, S. (2001). Organizing sound sequences in the human brain: The interplay of auditory streaming and temporal integration. *Brain Research*, **897**, 222-227.

**NOTES**

1. The term *irrelevant speech effect* (the disruption of the serial recall for visually presented lists caused by irrelevant sounds) is derived from the original observations showing that the cross-modal interference caused by speech was far greater than that produced by other types of auditory stimulus material. However, the cited studies (D. M. Jones et al., 1999; D. M. Jones & Macken, 1995) proved that the effect is not specific to speech stimuli. Under the right circumstances, other types of sounds can also cause substantial disruption of the serial recall task, whereas the effect of speech sounds can be reduced. Thus, the effect is now termed the *irrelevant sound effect*.

2. Focusing one's attention on sounds presented to one ear and detecting infrequent deviants among them attenuates the MMN amplitude to similar deviants in the other ear (Näätänen, Paavilainen, Tiitinen, Jiang, & Alho, 1993; Treijo, Ryan-Jones, & Kramer, 1995; Woldorff, Hackley, & Hillyard, 1991; Woldorff, Hillyard, Gallen, Hampson, & Bloom, 1998). However, in the absence of such competition (e.g., when the attended and ignored deviations are different), MMN is not affected by attention manipulations (Sussman, Winkler, & Wang, in press).

3. It should be noted that the MMN component per se is not sensitive to grouping. In investigating the effects of auditory grouping on selection-related ERP components (processing negativity or Nd; see Näätänen et al., 1978, and Hansen & Hillyard, 1980, respectively), Alain, Achim, and Richer (1993) and Alain and Woods (1994) found no direct effect of large-scale auditory organization on the MMN.

4. The participant who succeeded at the task could hold on to the general structure of the sequence, regarding the oddball tones as beginning the bars of a regular 3/4 measure rhythm, thus achieving over 60% correct performance. The key to her success was structuring the sequence in a way that allowed detection of the regularity of the oddball sequence. Only this participant had extensive musical training.

5. Should the duration of the intervening tones have been equal to that of the standard tone, the participants would have been able to detect the short deviant tones despite the variation introduced by the intervening tones, even when the intervening tones were grouped together with the oddball tones. Such a situation would also have resulted in MMN being elicited by the deviant tones in the participants who ignored the auditory stimuli, since, as previous studies have demonstrated, MMN is elicited by violating the constancy of one feature even when other features of the tones are varied randomly in the auditory stimulus sequence (Gomes et al., 1995; Huotilainen et al., 1993; Nousak et al., 1996; Winkler et al., 1990).