

# Object representation in the human auditory system

István Winkler,<sup>1</sup> Titia L. van Zuijen,<sup>2</sup> Elyse Sussman,<sup>3</sup> János Horváth<sup>1</sup> and Risto Näätänen<sup>2</sup>

<sup>1</sup>Institute for Psychology, Hungarian Academy of Sciences, H-1394 Budapest, P.O.Box 398, Hungary

<sup>2</sup>Cognitive Brain Research Unit, Department of Psychology, University of Helsinki and Helsinki Brain Research Centre, P.O.Box 9, FIN-00014 University of Helsinki, Finland

<sup>3</sup>Department of Neuroscience and Department of Otorhinolaryngology-Head and Neck Surgery, Albert Einstein College of Medicine, 1410 Pelham Parkway S, Bronx, NY 10461, USA

**Keywords:** auditory sensory memory, auditory stream segregation, event-related potentials, implicit memory, spectro-temporal processing

## Abstract

One important principle of object processing is exclusive allocation. Any part of the sensory input, including the border between two objects, can only belong to one object at a time. We tested whether tones forming a spectro-temporal border between two sound patterns can belong to both patterns at the same time. Sequences were composed of low-, intermediate- and high-pitched tones. Tones were delivered with short onset-to-onset intervals causing the high and low tones to automatically form separate low and high sound streams. The intermediate-pitch tones could be perceived as part of either one or the other stream, but not both streams at the same time. Thus these tones formed a pitch 'border' between the two streams. The tones were presented in a fixed, cyclically repeating order. Linking the intermediate-pitch tones with the high or the low tones resulted in the perception of two different repeating tonal patterns. Participants were instructed to maintain perception of one of the two tone patterns throughout the stimulus sequences. Occasional changes violated either the selected or the alternative tone pattern, but not both at the same time. We found that only violations of the selected pattern elicited the mismatch negativity event-related potential, indicating that only this pattern was represented in the auditory system. This result suggests that individual sounds are processed as part of only one auditory pattern at a time. Thus tones forming a spectro-temporal border are exclusively assigned to one sound object at any given time, as are spatio-temporal borders in vision.

## Introduction

The visual input is rich in information about spatial and invariant surface characteristics of physical objects. These dominate our perception and play a crucial role in determining what is commonly regarded as an object (Lakoff & Johnson, 1999). In contrast, the dominant part of acoustic information can be better described in terms of events (such as a bird trill or a footstep) rather than static objects. Thus the notion of an auditory perceptual 'object' is not clear (Bregman, 1990; Blauert, 1997). Observations about the role of spectral information in selecting parts of an auditory scene led Kubovy (1981) to suggest that auditory objects are separated by spectro-temporal, rather than spatio-temporal, borders (see also Shamma, 2001). Sound patterns (spectro-temporal regions of the acoustic input) appear to be valid units of perception and they are represented both as perceptual entities and as abstract ones (Poeppe, 2003).

Modern theories define 'objects' in terms of processing principles applicable across different modalities (Kubovy, 1988; Griffiths & Warren, 2004; Handel, 1988a,b). A cross-modal notion of object can be based on the separability of objects (Kubovy, 1981; Kubovy & Valkenburg, 2001). Exclusive allocation is an important processing principle that governs the allocation of the sensory input into perceptual units and thus guides the separation of objects and the

distinction between foreground and background (Köhler, 1947). Exclusive allocation means that any given part of the sensory input (including borders separating two objects) can only belong to one object at a time. If the border separating two parts of a display can be assigned to either one of them, the result is ambiguous perception (see Rubin's famous face–vase illusion: Rubin, 1915; also Fig. 1).

To test whether the principle of exclusive allocation applies in audition, we constructed an auditory model of the ambiguous 'border' situation. We utilized the auditory streaming phenomenon to construct tone sequences with two distinct sound streams (one low, the other high), while intermediate-pitched tones could join either one of the streams in perception. Depending on the assignment of the intermediate-pitch 'border' sounds, different temporal patterns emerged in perception. We instructed participants to link the border sounds to one of the streams and investigated whether the brain constructs a neural representation only for the selected pattern or, simultaneously, also for the other possible sound pattern.

The question was studied using the mismatch negativity (MMN) event-related brain potential, which is elicited by sounds violating an acoustic regularity of the preceding sound sequence (Näätänen & Winkler, 1999; Picton *et al.*, 2000) whether or not the sounds are attended (Näätänen, 1990; Sussman *et al.*, 2003). It has been shown that MMN can be used to index the representation of sound patterns (Winkler & Schröger, 1995; Sussman *et al.*, 1998, 2002). Occasional changes were introduced into the tone sequences, which violated either the tone pattern selected by the subject or the alternative pattern but not both at the same time. This way, MMN elicitation indicates the

*Correspondence:* Dr István Winkler, Institute for Psychology, Hungarian Academy of Sciences, P.O.Box 398, Szondi u 83–85, H-1394 Budapest  
E-mail: iwinkler@cogpsyphy.hu

Received 22 March 2006, revised 2 May 2006, accepted 5 May 2006

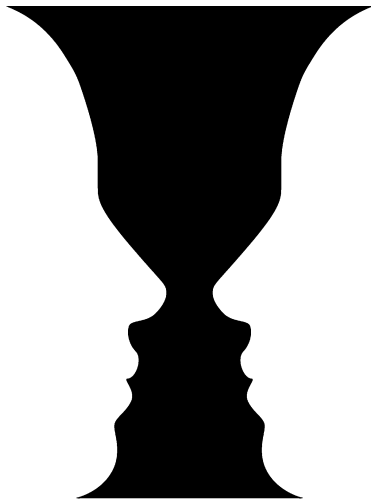


FIG. 1. Rubin's reversible face–vase illusion. This picture can be perceived either as a black vase in the centre over a white background or as two white profiles facing each other in front of a black background. The borders between the black and the white areas of the display belong to either one or the other area, but not to both of them at the same time (the exclusive allocation principle applied to the borders). The area 'receiving' the border is perceived as being in the foreground, whereas the other area is seen as being partially obscured by the former (i.e., it lies in the background).

presence of an auditory representation for the selected and/or the alternative tone pattern.

## Materials and methods

### Experimental subjects

Twenty-four young healthy volunteers participated in the experiment (8 male and 16 female, mean age 23.2 years). They were paid for taking part in the experiment. Subjects signed informed consent after the nature and procedures of the experiment were explained to them. The experiment was approved by the ethical committee of the Department of Psychology, Helsinki University. Data from four subjects were rejected during data analysis due to extensive electrical artifacts.

### Stimuli

Figure 2 shows a schematic diagram of the main test sequence. Three sets of tones, differing only in pitch, were presented in the sequences:

low tones [548 Hz, 50 dB above hearing threshold (AHT) of the individual], intermediate-pitch tones (750 Hz; 48 dB AHT) and high tones (1155 Hz; 45 dB AHT). Tone intensities were set to provide equal loudness across the three frequencies (Lindsay & Norman, 1977). Tone frequencies and the interstimulus intervals were chosen to ensure equal probability of the intermediate-pitched tones to be grouped with the high or the low tones (Baker *et al.*, 2000), but not both (Divenyi & Hirsh, 1978; Bregman *et al.*, 2000). Automatic segregation of the high and low tones as well as the perceptions resulting from joining the intermediate-pitch tones with either the high or the low tones was checked in an informal pilot study conducted with colleagues at the Cognitive Brain Research Unit in Helsinki using the same tone sequences as in the main experiment (subjects in the pilot study were 'naïve' with regards to the experiment). All subjects of the pilot study were able to hear both alternative groupings of the intermediate-pitch tone and none of them could join the high and low tones into a single pattern.

The common stimulus duration was 30 ms, including 2.5 ms linear rise and 2.5 ms fall times. The order between the tones was constant throughout the sequences, cyclically repeating the five tones in the following order: A, B, C, D, and E (Fig. 2). The stimulus onset asynchrony (SOA; onset-to-onset interval) between consecutive tones was randomly varied between predefined limits (see Table 1); the average duration of a cycle was 732 ms. Sequences could be perceived in two alternative ways. Grouping the intermediate tone with the high tones resulted in a repeating tone triplet that started with the intermediate tone followed by two high tones (E-A-C; marked by thin continuous frames on Fig. 2) with the two low tones occurring independently of the triplet (i.e. in a separate sound stream). Perception of the repeating E-A-C triplet was encouraged by the timing of the tones, which separated consecutive triplets with longer silent intervals than the intervals appearing within the triplets: The intervals separating the onsets of the E and A tones (median SOA 160 ms) and those separating the A and C tones (median SOA 252 ms) were substantially shorter than the interval between the C and E tones (median SOA 320 ms; see Table 1). Grouping the intermediate tone with the low ones resulted in the perception of a different repeating tone triplet, which started with the two low tones followed by the intermediate tone (B-D-E; marked with thin dashed frames on Fig. 2), whereas the two high tones were perceived as belonging to a different sound stream. Again, grouping occurred, because the SOAs between B and D (median 252 ms), and D and E (median 160 ms) were substantially shorter than the SOA between E and B (median 320 ms; see Table 1).

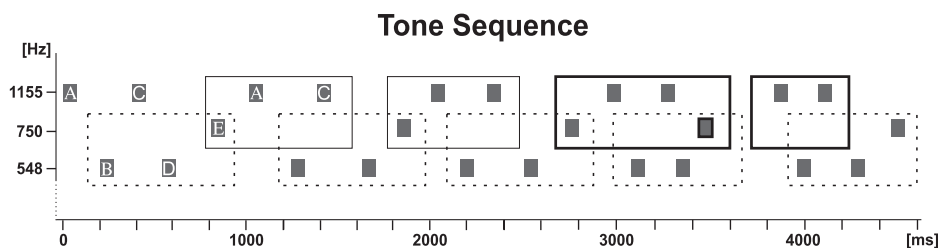


FIG. 2. Schematic illustration of the tone sequence. The x axis represents time, the y axis frequency. Tones are marked by grey rectangles and their positions within the repeating five-tone cycle are denoted by the letters A, B, C, D, and E. 'High-group' participants were asked to maintain perception of a repeating pattern formed from the high and intermediate tones (intermediate–high–high: E-A-C, marked by solid-line rectangles). 'Low-group' participants were asked to maintain perception of the repeating pattern formed from the low and intermediate tones (low–low–intermediate: B-D-E, marked by dashed-line rectangles). Occasionally, the silent interval preceding an intermediate-pitch (E) tone was shortened (for timing data, see Table 1). This change resulted in the formation of deviant patterns in the intermediate–high–high pitch pattern, but not in the low–low–intermediate pattern. The deviant E tone is marked by a thick black line surrounding the grey rectangle and the resulting deviant patterns are marked by thick-line frames. Note that early delivery of the intermediate-pitch tone only breaks the repetition of the intermediate–high–high pattern (E-A-C), whereas the low–low–intermediate pattern (B-D-E) continues unbroken.

TABLE 1. Distribution of the stimulus onset asynchronies (onset-to-onset intervals) between successive tones, separately for standard and deviant cycles

|                |       | Stimulus onset asynchrony (s) |         |        |                |                |
|----------------|-------|-------------------------------|---------|--------|----------------|----------------|
|                |       | Minimum                       | Maximum | Median | Lower quartile | Upper quartile |
| Standard cycle |       |                               |         |        |                |                |
| A-to-B         | 0.128 | 0.280                         | 0.160   | 0.144  | 0.176          |                |
| B-to-C         | 0.060 | 0.220                         | 0.092   | 0.076  | 0.108          |                |
| C-to-D         | 0.128 | 0.280                         | 0.160   | 0.144  | 0.180          |                |
| D-to-E         | 0.128 | 0.256                         | 0.160   | 0.144  | 0.180          |                |
| E-to-A         | 0.128 | 0.268                         | 0.160   | 0.148  | 0.180          |                |
| Deviant cycle  |       |                               |         |        |                |                |
| A-to-B         | 0.132 | 0.248                         | 0.160   | 0.148  | 0.176          |                |
| B-to-C         | 0.060 | 0.220                         | 0.080   | 0.068  | 0.096          |                |
| C-to-D         | 0.060 | 0.172                         | 0.080   | 0.072  | 0.092          |                |
| D-to-E         | 0.096 | 0.216                         | 0.130   | 0.112  | 0.148          |                |
| E-to-A         | 0.232 | 0.356                         | 0.264   | 0.248  | 0.280          |                |

Cycles consisted of five tones: A, B, C, D and E (see Fig. 1). Note that deviant cycles differed from standard cycles by shorter intervals between C and D and between D and E, and longer intervals between E and A.

Occasionally (in 8% of the cycles), the SOA between the C and E tones was shortened from 320 ms (median in regular cycles) to 210 ms (median). For participants who selected the repeating E-A-C pattern, this deviation resulted in the next E tone joining the pattern (E-A-C-E; see Fig. 2), because shortening the SOA between C and E brought the C-E interval into the range of the preceding E-A and A-C intervals (medians, 160 and 240 ms, respectively). Depending on the actual timing of the following A and C tones (which varied as described in Table 1), these tones could also join the preceding pattern (E-A-C-E-A-C) or form a separate tone pair (A-C; illustrated on Fig. 2). After that the regular cycle returned. The intermediate tone, which was delivered too early (termed 'deviant tone') and the resulting deviant pattern are marked by thick frames on Fig. 2. Importantly, for participants who selected the repeating B-D-E pattern, early delivery of the E tone did not result in a different grouping or in the temporal violation of the repeating tone pattern. This is because within the cycle of the 'deviant' E tone, the SOAs between the B and D and the D and E (medians 160 and 130 ms, respectively) remained substantially shorter than the SOA between the following E and B tones (median 424 ms), which was within the range of variation in the regular ('standard') pattern (standard E-B minimum-maximum SOA range, 256-548 ms; see Fig. 2 and Table 1). Thus the temporal deviation of the medium-pitch 'E' tone caused a large-scale change in one of the two alternative perceptions but no detectable change in the other alternative perception. Deviations occurred randomly within the sequence with the constraint that two deviant cycles were separated by at least two full standard cycles.

Half of the subjects (10 volunteers) were instructed to group the intermediate tone together with the high tones ('high group') and maintain this perception throughout all the stimulus blocks in the experiment. The other half of the subjects was instructed to group the intermediate tone with the low tones ('low group'). Deviations in the tone sequence illustrated in Fig. 2 violated the selected pattern for the high group of subjects and the alternative (not selected) pattern for the low group. For each group, the role of the high and low tones was reversed in half of the stimulus blocks: that is, the position of the high and low tones within the five-tone cycle was exchanged. Thus the same two alternative patterns emerged, but with the opposite grouping between the intermediate and high or low tones. In these reversed tone sequences, the same deviation (as described above) violated the

selected pattern for the low group and the alternative pattern for the high group.

In separate stimulus blocks, three experimental conditions were administered to each group of subjects. One experimental condition tested whether violations of the selected pattern elicited the MMN event-related brain potential. This condition is termed the 'Selected-pattern-deviant' condition. The second experimental condition tested whether violations of the alternative pattern elicited the MMN (termed 'Alternative-pattern-deviant' condition). For control purposes a third condition, termed 'Unambiguous-pattern' condition, was also administered. This condition tested the effects of the pattern violation without interference from those tones that were not included in the selected pattern. That is, in the Unambiguous-pattern condition, high-group subjects received the tone sequence shown on Fig. 2 but without the low tones. Low-group subjects received the reversed sequence, but without the high tones. Stimulus timing in the Unambiguous-pattern condition was identical to that in the corresponding Selected-pattern-deviant condition.

Subjects were presented altogether with 30 stimulus blocks of 205 cycles each (1025 tones; ~2.5 min duration per stimulus block). Each condition received 10 stimulus blocks, which together contained 160 deviant cycles. The order of the stimulus blocks of the Selected-pattern-deviant and the Alternative-pattern-deviant condition was balanced separately within each subject, whereas the Unambiguous-pattern condition was administered at the end of the session. Stimulus blocks for the Selected-pattern-deviant and the Alternative-pattern-deviant conditions started with five cycles during which those tones which were not part of the to-be-selected pattern were omitted (as in the Unambiguous-pattern condition). This induction presequence helped subjects to find the pattern they were to maintain throughout the stimulus block. Stimulus blocks were presented with short breaks between them and longer breaks (and the possibility to move about) after the 10th and the 20th stimulus block.

### Task

We monitored whether subjects were able to maintain perception of the designated tone pattern throughout the experiment by increasing the intensity (by 12 dB) of one tone in 5% of the selected pattern. That is, for the high group the intensity of either the intermediate-pitch or one of the two high tones was occasionally increased whereas for the low group either the intermediate-pitch or one of the two low tones changed. Altogether, 100 targets were presented in each condition. When detecting an intensity deviant, the subject was required to depress the response key whose number corresponded to the position of the intensity-deviant tone within the selected tonal pattern (i.e. 1, 2, or 3). Targets only appeared within the selected pattern (intermediate or high tone for high-group subjects, intermediate or low tone for low-groups subjects). That is, in the sequence shown in Fig. 2, high-group subjects were to press key 1 if they heard a louder E tone, key 2 for a louder A tone and key 3 for a louder C tone. Low-group subjects were to press 1 for louder B tones, 2 for louder D tones and 3 for louder E tones. Targets occurred randomly within the sequence with the constraint that they were separated from each other and from the temporal-deviant cycles by at least one full standard nontarget cycle. Participants could only perform this task successfully if they perceived the designated tone pattern. Detecting an intensity deviant in and of itself did not lead to a correct response because the task also included the requirement to indicate the position of the target within the selected pattern by pressing the appropriate response key.

### Procedures

Before training for the task started, the hearing threshold of the subject was determined for the intermediate tone using the staircase method so that tone intensities could be set accordingly. After the hearing-threshold measurement, the structure of the tone sequences was explained using a visual diagram similar to Fig. 2. Subjects then practiced maintaining perception of the designated tone pattern in Unambiguous-pattern condition sequences (i.e. with only the tones that formed the designated pattern). Once the subject could comfortably maintain perception of the designated pattern, sequences with all tones were presented, the task again being to maintain perception of the designated pattern. This phase lasted until the subject reported that he/she could maintain perception of the designated pattern. At this point the task was explained. The subject first practiced detecting target sounds in slower-paced sequences containing only the designated pattern and, subsequently, in sequences delivered at the pace used in the experiment. Finally, the subject practiced the task on the actual experimental stimulus sequences. Practicing usually lasted for ~30 min.

### EEG recording and data analysis

The electroencephalogram (EEG) was recorded with Ag/AgCl electrodes from eight scalp locations (F3, F4, C3, C4, P3 and P4 of the international 10–20 system and from the left and right mastoids, Lm and Rm, respectively) with the common reference electrode placed on the tip of the nose. The horizontal electrooculogram was monitored with a bipolar montage between electrodes placed lateral to the outer canthi on each side. The vertical electrooculogram was monitored between an electrode placed above and another below the right eye. Signals were digitized with a sampling frequency of 250 Hz and off-line-filtered between 2.5 and 16.0 Hz. Epochs within which the voltage difference between temporally adjacent sampling points exceeded 8  $\mu$ V on any channel were rejected from further analysis (Junghöfer *et al.*, 2000).

Due to the fast SOAs used, the short-latency event-related potentials (ERP) elicited by a given sound were expected to overlap the longer-latency potentials elicited by the previous sound. Because the deviant intermediate-pitch (E) tones were systematically delivered with a shorter SOA than the corresponding regular (standard) E tones, the ERP overlap effects from the preceding tone would be different on the ERPs recorded to standard and deviant E tones. To reduce this difference, which would confound the genuine ERP effects of regularity violation, we employed the ADJAR level 1 procedure (Woldorff, 1993), which aims to remove the ERP waveform elicited by the preceding tone from an ERP response. The ADJAR procedure was specifically developed for stimulus sequences delivered with fast and random SOAs. First, the average ERP elicited by the preceding tone was calculated separately for the standard and deviant E tones. These ERPs were then convolved with the corresponding normalized SOA distribution between the standard or deviant E tone and the preceding tone. Finally, the resulting waveforms were subtracted from the average ERP response elicited by the standard or deviant E tone. Statistical analysis and figures are based on these corrected waveforms for deviants and standards. Although the ADJAR procedure substantially reduced the overlap effect, residual differences can still be observed during the first ~100 ms of the standard vs. deviant E-tone responses. This is because the ERP waves elicited by the preceding tone are still relatively large in this time range, which falls between ~130 and 230 ms from the onset of the preceding tone. However, the overlap effect is minimal in later latency ranges. Therefore, we only

analysed ERPs from 100 ms onwards (starting with the auditory N1) for testing the questions of the current study. Because target tones were not shifted in time (compared with nontarget tones), no ADJAR correction was necessary, as the overlap effects from the ERP elicited by the preceding tone was the same for target and nontarget tones and thus they could be directly compared with each other. ERP responses elicited by the two tones (A and C) that follow the deviant E tone within the attended stream were not analysed. This is because perception of these tones was not uniform throughout the sequence. Sometimes these tones could be perceived as a separate pair while at other times they joined the preceding pattern (see the Stimuli section above). Averaging over the two cases would not yield meaningful results.

The corrected 600-ms-long ERP epochs elicited by the intermediate-pitch tones (including 200 ms prestimulus period) were separately averaged for the three different types of five-tone cycles (standard, deviant or target), condition (Selected-pattern-deviant, Alternative-pattern-deviant, Unambiguous-pattern), and participant group (high or low). Amplitude measurements were referred to the mean voltage of the prestimulus period.

The Unambiguous-pattern condition was used as a control testing the ERP effects of the pattern deviation used in the main test conditions (see Fig. 3, right column). ERPs in this condition showed a frontally negative wave (elicited by both standard and deviant E tones) in the 100–150 ms latency range from stimulus onset (the N1 wave; see Näätänen & Picton, 1987), a deviant-minus-standard negative difference, which showed slight polarity inversion at the mastoid leads in the 150–200 ms latency range (MMN; Näätänen *et al.*, 1978), a subsequent negative deviant-minus-standard difference with a clear same-polarity response at the mastoid leads in the 200–250 ms latency range (N2b; Ritter & Ruchkin, 1992), and two frontally positive differences in the 250–300 and 300–350 ms latency ranges (early and late P3a, respectively; see Escera *et al.*, 1998).

Based on the sequence of components found in the Unambiguous-pattern condition, responses in the two main experimental conditions were identified and statistically analysed. The analyses were conducted for the mean amplitudes in four latency ranges: a frontally negative wave in the 152–176 ms interval (N1), a frontally negative wave in the 212–236 ms interval (MMN), a frontally positive wave in the 276–300 ms interval (early P3a), and a frontally positive wave in the 332–356 ms interval (late P3a). No equivalent of the N2b component seen in the Unambiguous-pattern condition could be discerned in the main experimental conditions. In the Unambiguous-pattern condition, the N2b amplitude was measured in the 216–236 ms latency range, whereas MMN in this condition could be assessed from the 148–172 ms latency range. N1 and MMN appeared earlier in the Unambiguous-pattern than in the Selected-pattern-deviant condition, which was probably due to the overall higher density and increased variability of tones in the two main experimental conditions compared with the Unambiguous-pattern condition, as was shown in previous MMN studies (e.g. Winkler *et al.*, 1990; Wang *et al.*, 2005). ERP components are marked separately for the Selected-pattern-deviant and Unambiguous-pattern conditions on the frontal and mastoid ERP responses shown in Fig. 3.

Statistical comparisons of the ERP measurements for intermediate tones were conducted by ANOVA [Group (high vs. low)  $\times$  Condition (Selected-pattern-deviant vs. Alternative-pattern-deviant)  $\times$  Stimulus Type (standard vs. deviant)  $\times$  Electrode (F3 vs. F4)]. The Unambiguous-pattern condition was not included in these comparisons, as the larger differences between this and the two primary experimental conditions could obscure the answer to the main question. All factors except 'Group' were regarded as dependent. In addition, elicitation of

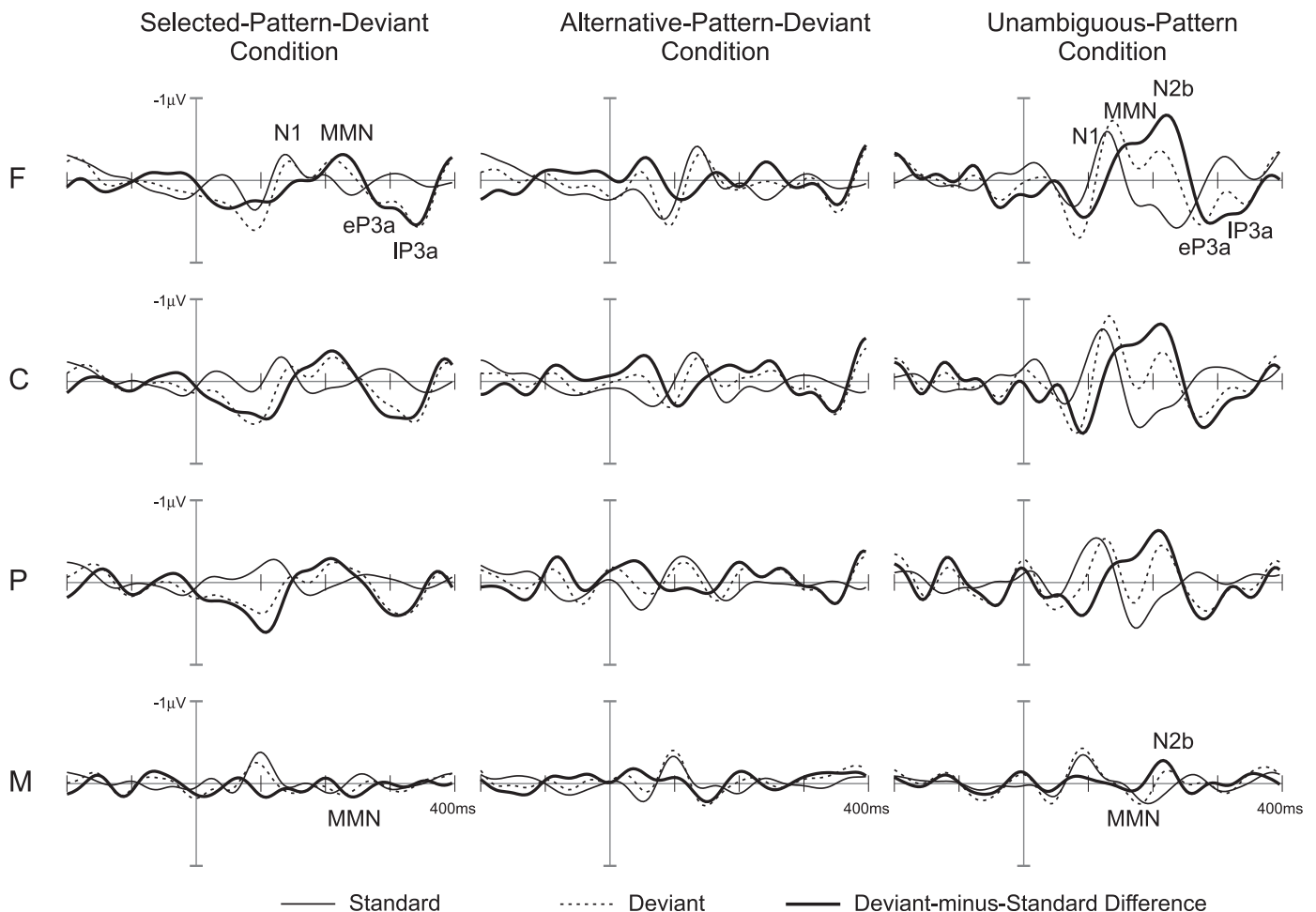


FIG. 3. Grand-average ( $n = 20$ ) responses elicited by standard (thin continuous line) and deviant (dashed line) intermediate (E) tones, together with the corresponding difference waveforms (thick continuous line), separately overlaid for the Selected-pattern-deviant (left column), Alternative-pattern-deviant (middle column) and Unambiguous-pattern (right column) conditions. Rows correspond to responses averaged between the signals recorded at F3 and F4 (marked F), C3 and C4 (C), P3 and P4 (P), and the left and right mastoids (M). Stimulus onset is at the crossing of the two axes.

the second negative and the first positive difference wave (MMN and P3a, respectively) were tested separately for all three experimental conditions by comparing the deviant-minus-standard differences against zero with Student's *t*-test. These tests were conducted on data pooled from the two groups of participants and the F3 and F4 electrodes, because these factors showed no effect in the ANOVA tests (see Results).

Responses to the target tones were accepted within the 150–3000 ms poststimulus period. Responses outside this interval were treated as false alarms, whereas incorrect responses to targets (e.g. pressing key 2 or 3 to the first tone of the target pattern) were marked as errors. In order to check whether subjects did maintain perception of the designated pattern, error rates were tested against a model based on the error rate expected if subjects reacted to the intensity deviants without perceiving the designated pattern. Responding to intensity deviants by randomly pressing one of the three response keys would lead to 33% hit and 67% error rate. Therefore, using Student's *t*-test, the number of errors divided by the sum of hits and errors (these together represent all detected intensity deviants) was compared with the number 0.67, separately for each condition. In addition to the above analysis, the pattern of reaction times (only correct responses) and hit rates were analysed by ANOVA [Group (high vs. low)  $\times$  Condition (Selected-pattern-deviant vs. Alternative-pattern-deviant)  $\times$

Position of the Target within the Pattern (1 vs. 2 vs. 3)]. False alarms were analysed by ANOVAs with the structure Group (high vs. low)  $\times$  Condition (Selected-pattern-deviant vs. Alternative-pattern-deviant).

ERP responses to target tones (N2b and P3b, shown by the responses in the Unambiguous-pattern condition; see Fig. 4, right column) were measured from 24-ms wide windows centred on the peaks found in the group-averaged target responses (high and low groups, separately) and analysed by ANOVA [Group (high vs. low)  $\times$  Condition (Selected-pattern-deviant vs. Alternative-pattern-deviant)  $\times$  Position of the Target within the Pattern (1 vs. 2 vs. 3)], similarly to the hit rate and reaction time measures. N2b was measured from the average of the signals recorded from C3 and C4, whereas P3b was from the average of the P3 and P4 signals, in line with the well-known scalp distribution of these components. The elicitation of these components was tested using Student's *t*-tests comparing the amplitudes against zero. For these analyses, data from the two groups of participants were pooled, as no group differences were found.

In all statistical analyses, Greenhouse–Geisser adjustment of the degrees of freedom was used where applicable ( $\epsilon$ -values and the uncorrected degrees of freedom are given in Results). Significant effects were further specified by Tukey's HSD *post hoc* tests.

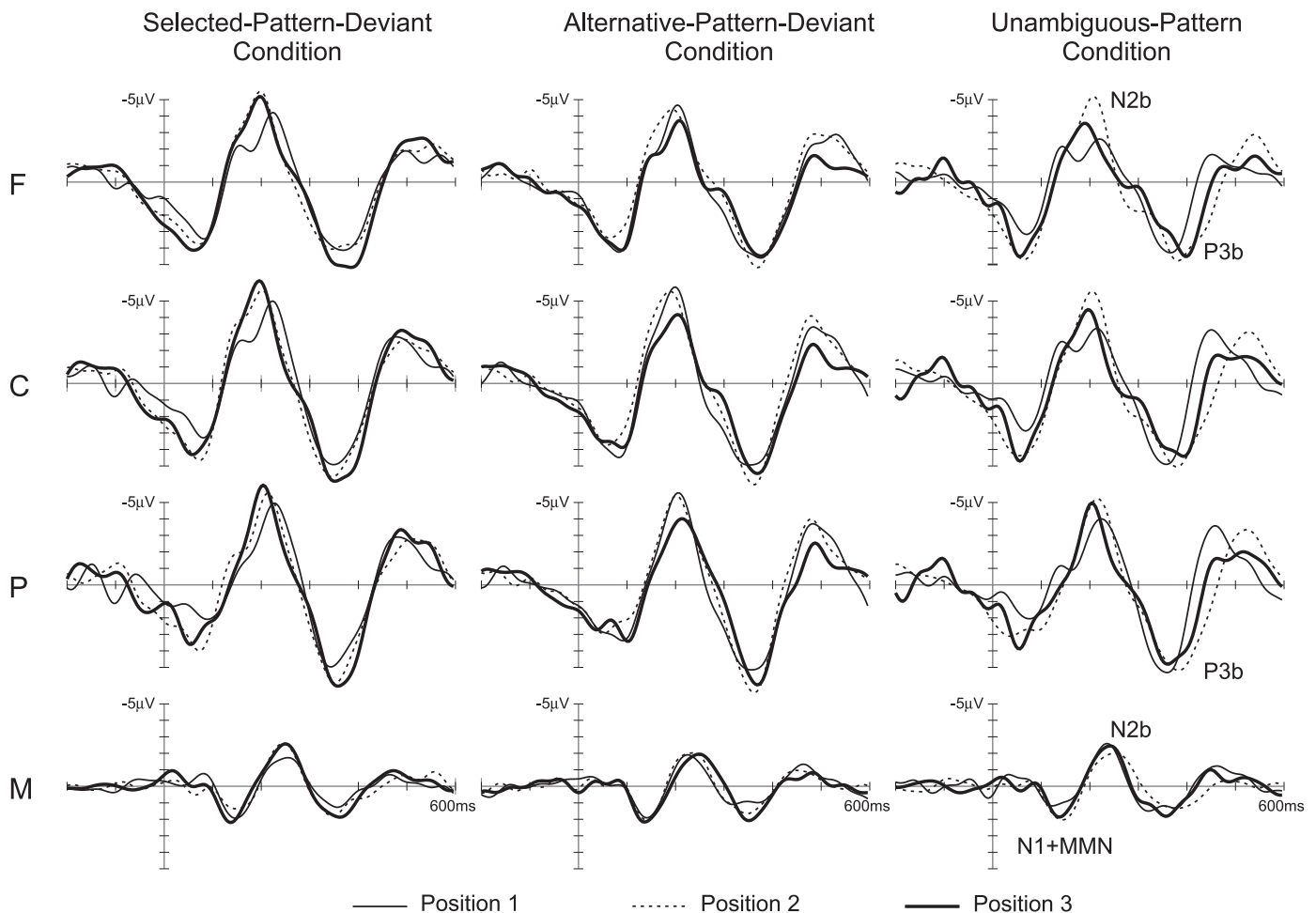


FIG. 4. Grand-average ( $n = 20$ ) ERP responses elicited by target tones in the Selected-pattern-deviant (left column), Alternative-pattern-deviant (middle column) and Unambiguous-pattern (right column) conditions. The three possible target positions have been overlaid (Position 1: thin continuous line; Position 2: dashed line; Position 3: thick continuous line). Rows correspond to responses averaged between the signals recorded at F3 and F4 (marked F), C3 and C4 (C), P3 and P4 (P), and the left and right mastoids (M). Stimulus onset is at the crossing of the  $x$  and  $y$  axes.

## Results

### Behavioural measures of target detection

Behavioural measures were analysed for two reasons: (i) to test whether subjects maintained perception of the designated pattern in the Selected-pattern-deviant and the Alternative-pattern-deviant condition; and (ii) to assess whether there were significant differences between the two groups of subjects in maintaining perception of the designated pattern.

Table 2 gives the grand-average hit, false-alarm and error rates (incorrect key depressed in response to a target tone), as well as the reaction times measured for the different conditions and target positions. Error rates were significantly below the 0.67 level predicted if participants detected the intensity deviants but were unable to tell their position within the designated pattern ( $t_{19} = -7.32, -13.08$  and  $-9.77, P < 0.00001$  each, for the Selected-pattern-deviant, Alternative-pattern-deviant and Unambiguous-pattern conditions, respectively). Because hits and errors together constitute those cases in which the subject correctly detected an intensity deviant (this is not sufficient for a correct response because the task was to respond according to the position of the intensity-deviant within the selected pattern), the above result also means that hit rates were significantly higher than what could be expected if participants were not able to maintain perception of the designated pattern. Thus, although the task

was not easy, participants maintained perception of the designated pattern during the EEG recordings.

For hit rates, the ANOVA (Group  $\times$  Condition  $\times$  Position of the Target within the Pattern) showed significant interaction between Condition and Position ( $F_{2,36} = 5.28, \epsilon = 0.87, P < 0.05$ ) and also a significant Condition effect ( $F_{1,18} = 12.50, P < 0.01$ ). Both effects were caused by the third-position target in the Alternative-pattern-deviant condition being detected more often than any other target in either condition (all  $P < 0.01$  in the Tukey HSD test). Furthermore, first-position targets were detected significantly faster than second-position ones ( $F_{2,36} = 4.79, \epsilon = 0.84, P < 0.05$  and Tukey HSD showing  $P < 0.05$ ; for mean reaction times, see Table 3). False alarm rates were unaffected by either Group or Condition. None of the analyses showed any difference between the two groups of subjects.

In summary, participants were mostly able to maintain perception of the designated pattern. The lack of task-performance differences between the two groups of subjects allows collapsing the ERP results across the two groups.

### ERP responses to standard and deviant tones

Figure 3 shows the grand-averaged ERP responses elicited by the deviant and standard intermediate (E) tones together with the

TABLE 2. Grand-average hit, false-alarm and error rates, and reaction times, measured in the different conditions

| Condition                   | Measured probabilities* |              |             | Reaction time (s) |
|-----------------------------|-------------------------|--------------|-------------|-------------------|
|                             | Hits                    | False alarms | Errors      |                   |
| Selected-pattern-deviant    | –                       | 0.02 ± 0.03  | 0.33 ± 0.21 | –                 |
| Position 1                  | 0.46 ± 0.25             | –            | –           | 1.05 ± 0.41       |
| Position 2                  | 0.50 ± 0.25             | –            | –           | 1.23 ± 0.42       |
| Position 3                  | 0.48 ± 0.28             | –            | –           | 1.25 ± 0.51       |
| Alternative-pattern-deviant | –                       | 0.03 ± 0.03  | 0.23 ± 0.15 | –                 |
| Position 1                  | 0.49 ± 0.24             | –            | –           | 1.11 ± 0.36       |
| Position 2                  | 0.51 ± 0.22             | –            | –           | 1.22 ± 0.38       |
| Position 3                  | 0.65 ± 0.27             | –            | –           | 1.13 ± 0.43       |
| Unambiguous-pattern         | –                       | 0.03 ± 0.03  | 0.27 ± 0.18 | –                 |
| Position 1                  | 0.60 ± 0.27             | –            | –           | 0.93 ± 0.22       |
| Position 2                  | 0.52 ± 0.23             | –            | –           | 1.18 ± 0.36       |
| Position 3                  | 0.53 ± 0.24             | –            | –           | 1.22 ± 0.54       |

Positions denote the sequential order of the tone within the pattern. Values are mean ± SD. \*Expected probabilities of hits and errors with random pressing of keys are 0.33 and 0.67, respectively.

TABLE 3. Grand-averaged MMN and P3a amplitudes by conditions, with N2b amplitudes for the condition under which it was elicited

| Condition                   | EEG amplitudes (µV) from |               |               |              |
|-----------------------------|--------------------------|---------------|---------------|--------------|
|                             | F3 and F4 (F)            | C3 and C4 (C) | P3 and P4 (P) | Mastoids (M) |
| Selected-pattern-deviant    |                          |               |               |              |
| MMN component               | -0.29 ± 0.47             | -0.30 ± 0.53  | -0.23 ± 0.57  | 0.07 ± 0.30  |
| P3a component               | 0.31 ± 0.46              | 0.38 ± 0.59   | 0.30 ± 0.65   | -0.01 ± 0.35 |
| Alternative-pattern-deviant |                          |               |               |              |
| MMN component               | -0.10 ± 0.50             | -0.15 ± 0.43  | -0.12 ± 0.45  | -0.06 ± 0.31 |
| P3a component               | 0.07 ± 0.38              | 0.13 ± 0.43   | -0.03 ± 0.62  | -0.03 ± 0.28 |
| Unambiguous-pattern         |                          |               |               |              |
| MMN component               | -0.42 ± 0.78             | -0.44 ± 0.71  | -0.27 ± 0.65  | 0.07 ± 0.34  |
| N2b component               | -0.75 ± 0.56             | -0.57 ± 0.70  | -0.47 ± 0.74  | -0.22 ± 0.35 |
| P3a component               | 0.49 ± 0.60              | 0.50 ± 0.79   | 0.38 ± 0.90   | 0.01 ± 0.45  |

Values are mean ± SD.

corresponding difference waveforms. A frontally negative wave was elicited by both standard and deviant tones. It peaked in the 140–180 ms latency range and was identified as the N1. The four-way ANOVA of the N1 amplitude showed only a significant two-way interaction between Stimulus Type (standard vs. deviant) and Electrode (F3 vs. F4:  $F_{1,18} = 5.33$ ,  $P < 0.05$ ). This interaction was explained by higher right frontal N1 amplitudes elicited by deviants than standards ( $P < 0.05$  in the Tukey HSD *post hoc* tests).

Another frontally negative wave was elicited only by deviant stimuli peaking in the 200–250 ms latency range in the Selected-pattern-deviant condition (see Fig. 3). The polarity of the deviant-minus-standard difference waveforms slightly reversed at the mastoid leads. This component was identified as the MMN, because the N2b obtained in the same time interval in the Unambiguous-pattern condition showed a clear same-polarity (negative) signal at the mastoid leads. The difference at the mastoid leads was significant between N2b (measured in the Unambiguous-pattern condition) and MMN (measured from the Selected-pattern-deviant and the Unambiguous-pattern condition):  $F_{2,38} = 5.87$ ,  $P < 0.01$  with Tukey HSD *post hoc* tests showing that the N2b signal at the mastoid was

significantly ( $P < 0.02$ , separately for each pair-wise comparison) more positive than the corresponding MMN signal.

The mean frontal MMN amplitudes showed only a main effect of Stimulus Type ( $F_{1,18} = 6.49$ ,  $P < 0.05$ ). Thus, overall, deviants elicited a more negative response than standards. Results did not differ between the two subject groups or the two electrodes included in the test. Thus the two groups were pooled together and the amplitudes were averaged between the F3 and F4 electrodes for testing the elicitation of MMN by Student's *t*-tests, separately for the three conditions (see Table 3 for the mean MMN amplitudes). Deviants elicited MMN in the Selected-pattern-deviant and Unambiguous-pattern, but not in the Alternative-pattern-deviant condition ( $t_{19} = -2.77$  and  $-2.43$ , both  $P < 0.05$ , for the Selected-pattern-deviant and Unambiguous-pattern conditions, respectively;  $t_{19} = -0.94$  for the Alternative-pattern-deviant condition).

MMN was followed by a fronto-centrally positive response with two peaks, the P3a component. P3a was elicited only by deviant stimuli. The first peak was in the 250–300 and the second in the 320–360 ms latency range. The early P3a amplitude showed an interaction between Condition and Stimulus Type ( $F_{1,18} = 5.72$ ,  $P < 0.05$ ) and a main effect of Stimulus Type ( $F_{1,18} = 6.37$ ,  $P < 0.05$ ). Both effects are explained by the significantly higher-amplitude response elicited by deviants than standards in the Selected-pattern-deviant but not in the Alternative-pattern-deviant condition (Tukey HSD: the amplitude elicited by the Selected-pattern deviant was more positive than that to either standards or the Alternative-pattern deviant by at least  $P < 0.05$ ). Again, no difference was found between the two groups of subjects or between F3 and F4. Deviants elicited the early P3a in the Selected-pattern-deviant and Unambiguous-pattern, but not in the Alternative-pattern-deviant, condition ( $t_{19} = 2.96$  and  $3.67$ ,  $P < 0.05$  and  $0.01$  for the Selected-pattern-deviant and Unambiguous-pattern conditions, respectively;  $t_{19} = 0.86$  for the Alternative-pattern-deviant condition). In the late P3a latency range, only a main effect of Stimulus Type was found on the ERP amplitudes ( $F_{1,18} = 15.88$ ,  $P < 0.01$ ), showing that, on average, deviants elicited a more positive response than standards.

### ERP responses to target tones

Figure 4 shows the responses to target tones, separately for the three possible target positions within the selected pattern. Target tones elicited a large negative response with a central maximum and no polarity inversion at the mastoid leads, which peaked in the 180–230 ms latency range. This response can be identified as N2b. The ANOVA test showed interaction between Condition and Position of the Target and Group ( $F_{2,36} = 4.78$ ,  $\epsilon = 0.979$ ,  $P < 0.05$ ). None of the Tukey HSD comparisons showed significant difference. Student's *t*-tests showed that N2b was elicited in all conditions and positions with at least  $P < 0.01$  (for mean amplitudes, see Table 4).

N2b was followed by a centro-parietal positive wave peaking in the 340–400 ms latency range, the target P3b response. The target P3b amplitude was significantly affected by the Position of the target ( $F_{2,36} = 4.53$ ,  $\epsilon = 0.864$ ,  $P < 0.05$ ), first-position targets eliciting slightly lower P3b responses than the second-position ones (Tukey HSD,  $P < 0.05$ ). Student's *t*-tests showed that P3b was elicited in all conditions and positions with at least  $P < 0.001$  (for mean amplitudes, see Table 4).

### Discussion

We investigated whether auditory spectro-temporal borders are treated similarly to spatio-temporal object borders in vision: they can only

TABLE 4. Grand-average N2b and P3b amplitudes for the different conditions

| Condition                          | EEG amplitudes ( $\mu\text{V}$ ) |                  |                  |                  |                        |                 |                 |                 |
|------------------------------------|----------------------------------|------------------|------------------|------------------|------------------------|-----------------|-----------------|-----------------|
|                                    | N2b component averages from      |                  |                  |                  | P3b component averages |                 |                 |                 |
|                                    | F3 and F4 (F)                    | C3 and C4 (C)    | P3 and P4 (P)    | Mastoids (M)     | F3 and F4 (F)          | C3 and C4 (C)   | P3 and P4 (P)   | Mastoids (M)    |
| <b>Selected-pattern-deviant</b>    |                                  |                  |                  |                  |                        |                 |                 |                 |
| Position 1                         | $-4.04 \pm 2.91$                 | $-4.80 \pm 2.88$ | $-4.76 \pm 3.05$ | $-1.34 \pm 1.57$ | $3.98 \pm 3.36$        | $4.87 \pm 3.65$ | $4.81 \pm 3.83$ | $1.21 \pm 1.55$ |
| Position 2                         | $-5.29 \pm 3.05$                 | $-5.47 \pm 3.80$ | $-4.78 \pm 4.02$ | $-0.42 \pm 1.86$ | $4.02 \pm 3.11$        | $5.58 \pm 3.25$ | $5.82 \pm 3.59$ | $1.75 \pm 1.47$ |
| Position 3                         | $-4.97 \pm 2.81$                 | $-5.95 \pm 3.30$ | $-5.73 \pm 3.54$ | $-0.74 \pm 1.50$ | $4.92 \pm 3.06$        | $5.86 \pm 3.25$ | $5.98 \pm 3.42$ | $1.79 \pm 1.66$ |
| <b>Alternative-pattern-deviant</b> |                                  |                  |                  |                  |                        |                 |                 |                 |
| Position 1                         | $-4.43 \pm 3.10$                 | $-5.65 \pm 3.86$ | $-5.32 \pm 3.69$ | $-1.09 \pm 3.55$ | $4.50 \pm 2.46$        | $4.51 \pm 3.36$ | $4.73 \pm 3.48$ | $0.88 \pm 1.70$ |
| Position 2                         | $-4.09 \pm 3.28$                 | $-4.94 \pm 3.25$ | $-5.23 \pm 3.08$ | $-1.53 \pm 1.29$ | $5.03 \pm 2.63$        | $5.98 \pm 2.85$ | $6.39 \pm 2.97$ | $1.43 \pm 1.19$ |
| Position 3                         | $-3.54 \pm 3.51$                 | $-4.00 \pm 3.26$ | $-3.86 \pm 2.65$ | $-0.94 \pm 1.46$ | $4.37 \pm 3.15$        | $5.37 \pm 3.72$ | $5.92 \pm 3.91$ | $1.76 \pm 1.90$ |
| <b>Unambiguous-pattern</b>         |                                  |                  |                  |                  |                        |                 |                 |                 |
| Position 1                         | $-2.50 \pm 4.29$                 | $-3.05 \pm 4.33$ | $-3.90 \pm 4.19$ | $-2.31 \pm 2.10$ | $4.20 \pm 3.51$        | $4.82 \pm 3.36$ | $5.24 \pm 3.64$ | $1.24 \pm 1.44$ |
| Position 2                         | $-5.00 \pm 3.06$                 | $-5.39 \pm 2.85$ | $-4.96 \pm 2.74$ | $-1.12 \pm 1.35$ | $4.66 \pm 3.03$        | $4.93 \pm 3.02$ | $5.12 \pm 3.17$ | $1.32 \pm 1.44$ |
| Position 3                         | $-3.44 \pm 2.71$                 | $-4.13 \pm 3.22$ | $-3.80 \pm 3.35$ | $-0.34 \pm 1.26$ | $4.36 \pm 3.11$        | $4.27 \pm 3.49$ | $3.96 \pm 3.36$ | $0.80 \pm 1.72$ |

Values are mean  $\pm$  SD. Positions denote the sequential order of the tone within the pattern.

belong to one sound pattern at a time. A repeating cycle of five tones were presented to subjects, who could perceive them in two mutually exclusive ways: grouping the intermediate-frequency tone either with the two high or with the two low tones, but not both at the same time. Thus the intermediate-frequency tones took the role of a border, whose allocation decided between two alternative perceptions of this ambiguous sequence (as is the case in Rubin's face-*vase* illusion; see Fig. 1). Participants were instructed to maintain one of the two alternative groupings. Infrequent deviants violated the temporal structure of either the selected tone pattern or the alternative one, but not both at the same time. If the border tones are exclusively allocated to the selected pattern, as is the rule for visual objects, then only violations of the selected pattern should elicit the MMN; those of the alternative pattern should not. If, however, both alternative patterns were processed in parallel, then MMN should be elicited by deviations of either pattern.

We found that participants were able to maintain perception of the designated repeating tonal pattern most of the time during the stimulus blocks, although short switches to the alternative grouping may have occurred, as is well-known for bi-stable perceptual configurations (cf. Leopold & Logothetis, 1999). This was shown by the low false-alarm and error rates, the latter being significantly lower than the level expected if participants could discriminate the target tones by their higher intensity but did not perceive the designated pattern. Furthermore, targets elicited the well-known target-related ERP components (N2b and P3b) in all three conditions.

MMN and the subsequent P3a component (its early part; see Escera *et al.*, 1998) were elicited by occasional violations of the structure of the repeating sound pattern in the Selected-pattern-deviant but not in the Alternative-pattern-deviant Condition. These results suggest that, at the stage of processing reflected by MMN, only the voluntarily selected sound organization was maintained in the brain. The differences found between the Alternative-pattern-deviant and the Selected-pattern-deviant condition could not have been caused by differences in the maintenance of the designated tone pattern because no significant performance or target ERP differences were found between these conditions.

Thus it appears that the principle of exclusive allocation applies also in the auditory modality. Each sound is assigned to one and only one auditory pattern, similarly to borders of objects in the visual modality. This suggests that the memory representation of pitch patterns may be

similar to that of visual objects, confirming the suggestion of Kubovy & Van Valkenburg (2001). The current results also support the suggestion of Kubovy (1981) that auditory objects are separated from each other by spectro-temporal borders and that, similarly to the allocation of spatio-temporal borders in vision, the allocation of spectro-temporal sound borders plays an important role in separating foreground and background objects in the auditory modality. For example, the siren sound of an ambulance car can be easily separated from the general street noise and from the sounds of an on-going conversation by its sharp spectro-temporal contours. One may then focus on the siren sound and look for the ambulance car. Alternatively, one can also let the siren sound be part of the background noise and follow the conversation, instead. The notion of temporal sound patterns acting as auditory objects is further supported by results showing that changes in a repeating sound elicits MMN only with respect to the regularities of the auditory stream to which the sound belongs (Ritter, Sussman & Molholm, 2000). This result suggests that individual sounds and their relationships (temporal, spectral, etc.) are represented as part of the description of the auditory 'object' to which they belong (cf. Winkler & Cowan, 2005).

The current as well as previous results also argue for object-based processing of sound. For example, when a tone sequence having the structure LLLLHLLLH... (where 'L' and 'H' represent a lower and a higher tone, respectively) was presented to participants at a slow pace [1.3 s SOA (onset-to-onset interval)] MMN was elicited by the relatively infrequent 'H' tones (Scherg, Vajsar & Picton, 1989). However, when the same sequence was presented at a fast pace (100 ms SOA), the 'H' tones did not elicit the MMN even though MMN was elicited by the same tones at the same delivery rate when the order of the tones was randomized (Sussman *et al.*, 1998; Sussman & Gumenyuk, 2005). These results suggest that the auditory regularity representations stored in the brain depend on the detection of the higher-order structure of the auditory input. The 'H' tones ceased to be 'deviants' when the repeating pattern was detected and thus they became part of the regularly repeating pattern, the object. Confirming this notion, the same sequence delivered at an intermediate pace (750 ms SOA) evoked MMN when participants were not aware of the higher-order structure of the tone sequence, but no MMN was elicited when participants were informed about the repeating pattern appearing in the sequence (Sussman *et al.*, 2002). Thus both stimulus-driven (rate of sound delivery) and top-down effects (knowledge of the



structure of the sequence) on pattern (object) formation can determine what is considered as change within the auditory input. These and similar results (e.g. Winkler, Sussman *et al.*, 2003), including the current ones, show similarity with the same-object advantage found in the visual modality, which show that searching for a combination of two target features is faster and easier when they appear on the same as opposed to two separate objects (Duncan, 1984; Valdes-Sosa *et al.*, 1998). Indeed, recent results strongly argue for preattentive binding of auditory stimulus features, which is an essential prerequisite of object formation (Takegata *et al.*, 2005; Winkler *et al.*, 2005a). Thus the view emerging from these investigations is that sound is processed in terms of sound patterns. Our current results argue that the auditory spectro-temporal patterns may be regarded as the true units of auditory processing, the auditory 'objects'. On the other hand, whereas the current results are compatible with the notion of preattentive formation of auditory objects they do not provide decisive evidence regarding this issue. Although a recent study showed that auditory streams can be formed even in the absence of focused attention (Sussman *et al.*, 2006), other results suggest that, when attention is strongly focused on a sound sequence, further streams may not be segregated (Brochard *et al.*, 1999; Sussman *et al.*, 2005; see, however, Winkler *et al.*, 2003). Future research using the current paradigm will ask the question of how the tones are grouped in the absence of focused attention.

The set of perceptual phenomena termed duplex perception (Rand, 1974; Lieberman, 1982) contradicts the principle of exclusive allocation. In its most widely studied case, one of the formant transitions of the syllable 'da' or 'ga' is separated from the rest of the acoustic signal forming the syllable. The two parts are then delivered to opposite ears of participants, who simultaneously hear both the original syllable and a separate chirp sound corresponding to the separated formant transition. Initially, this phenomenon was interpreted as demonstrating the existence of separate brain mechanisms processing speech sounds (Lieberman, 1982; Mathiak *et al.*, 2001). However, examples of duplex perception exclusively involving nonspeech stimuli have since been discovered (e.g. Fowler & Rosenblum, 1990). Bregman (1987) showed that multiple sound allocation can also occur in vision, when transparency allows elements of two objects to mingle in an ambiguous way. As sounds are transparent by nature, duplex perception could occur more often in the auditory modality. In fact, Bregman (1987, 1990) argued that under everyday circumstances, when two separate sounds share a common frequency, assigning the common frequency component to both sounds helps veridical perception in some auditory scenes. However, Bregman (1990) also pointed out that multiple allocation only occurred when two sound organizations received strong support from the primitive processes of auditory scene analysis and the two solutions were not contradictory. [Note that Bregman's description does not contradict the 'separate speech mechanism' interpretation of the language-related cases of duplex perception.] The same applies to duplex perception in vision. Therefore, in both modalities, the principle of exclusive allocation applies strictly only to stimulus configurations giving rise to contradictory alternative perceptual organizations. This is the case for Rubin's reversible face-vase illusion, the model of the auditory stimulation employed in the current study. Thus the current results showing exclusive allocation are compatible with the corresponding visual perceptual phenomenon and do not contradict the notion of duplex perception.

Two additional aspects of the current results may require discussion. First, the N1 peak latency was slightly longer than what is typical for the N1 wave. This was probably due to the fast and variable stimulus delivery, as has been found in previous experiments (Wang *et al.*, 2005). Second, MMN peaked earlier and N2b was elicited by deviants

in the Unambiguous-pattern, whereas MMN peaked somewhat later and no N2b was elicited in the Selected-pattern-deviant condition. The two components were clearly distinguished at the mastoid leads. The difference in the ERP results probably stems from differences in the complexity of the stimulation and in task difficulty. The overall stimulus presentation rate was much slower in the Unambiguous-pattern than in the other two conditions, because two sounds were omitted from each cycle (the sounds that did not belong to the selected pattern). This may have affected the component latencies. Moreover, maintaining perception of the designated pattern was much easier when no other sounds were present (in the Unambiguous-pattern condition). This might explain why deviations from the regular schedule were more distinct, thus eliciting the N2b component. If maintaining the same organization was effortful this suggests, that without the voluntary effort, perception would spontaneously flip between the two alternative perceptions, as is also the case for Rubin's reversible face-vase illusion. The current results suggest that the MMN measure will allow us to study the spontaneous fluctuation of perception, similarly to our previous study of an ambiguous case of auditory stream segregation (Winkler *et al.*, 2005b).

In summary, we found evidence suggesting that the principle of exclusive allocation applies to spectro-temporal sound patterns. This result supports the notion that sound patterns with pitch-time borders may fill the role of objects in sound processing. Our results are compatible with object-based theories of perception (Duncan & Humphreys, 1989).

## Acknowledgements

This research was supported by the National Institutes of Health grants TW005886 and DC04263, the Hungarian National Research Fund (OTKA T048383), the Academy of Finland and the Finnish Graduate School in Psychology.

## Abbreviations

AHT, above hearing threshold; ERP, event-related potentials; MMN, mismatch negativity; SOA, stimulus onset asynchrony.

## References

- Baker, K.L., Williams, S.M. & Nicolson, R.I. (2000) Evaluating frequency proximity in stream segregation. *Percept. Psychophys.*, **62**, 81–88.
- Blauert, J. (1997) *Spatial Hearing: the Psychophysics of Human Sound Localization*. MIT Press, Cambridge, MA.
- Bregman, A.S. (1987) The meaning of duplex perception: Sounds as transparent objects. In M.E.H.Schouten (ed.), *The Psychophysics of Speech Perception*. Martinus-Nijhoff NATO-ASI Series, Dordrecht, pp. 95–111.
- Bregman, A.S. (1990) *Auditory Scene Analysis: the Perceptual Organization of Sound*. MIT Press, Cambridge, MA.
- Bregman, A.S., Ahad, P.A., Crum, P.A.C. & O'Reilly, J. (2000) Effects of time intervals and tone durations on auditory stream segregation. *Percept. Psychophys.*, **62**, 626–636.
- Brochard, R., Drake, C., Botte, M.-C. & McAdams, S. (1999) Perceptual organization of complex auditory sequences: effect of number of simultaneous subsequences and frequency separation. *J. Exp. Psychol. Hum. Percept. Perform.*, **25**, 1742–1759.
- Divenyi, P.L. & Hirsh, I.J. (1978) Some figural properties of auditory patterns. *J. Acoust. Soc. Am.*, **64**, 1369–1385.
- Duncan, J. (1984) Selective attention and the organization of visual information. *J. Exp. Psychol. Gen.*, **113**, 501–517.
- Duncan, J. & Humphreys, G. (1989) Visual search and stimulus similarity. *Psychol. Rev.*, **96**, 458.
- Escera, C., Alho, K., Winkler, I. & Näätänen, R. (1998) Neural mechanisms of involuntary attention switching to novelty and change in the acoustic environment. *J. Cogn. Neurosci.*, **10**, 590–604.

- Fowler, C.A. & Rosenblum, L.D. (1990) Duplex perception: a comparison of monosyllables and slamming doors. *J. Exp. Psychol. Hum. Percept. Perform.*, **16**, 742–754.
- Griffiths, T.D. & Warren, J.D. (2004) What is an auditory object? *Nature Rev. Neurosci.*, **5**, 887–892.
- Handel, S. (1988a) Space is to time as vision is to audition: seductive but misleading. *J. Exp. Psychol. Hum. Percept. Perform.*, **14**, 315–317.
- Handel, S. (1988b) No one analogy is sufficient: rejoinder to Kubovy. *J. Exp. Psychol. Hum. Percept. Perform.*, **14**, 321.
- Junghöfer, M., Elbert, T., Tucker, D.M. & Rockstroh, B. (2000) Statistical control of artifacts in dense array EEG/MEG studies. *Psychophysiology*, **37**, 523–532.
- Köhler, W. (1947) *Gestalt Psychology*. Liveright, New York.
- Kubovy, M. (1981) Concurrent-pitch segregation and the theory of indispensable attributes. In: Kubovy, M. & Pomerantz, J. (eds), *Perceptual Organization*. Lawrence Erlbaum, Hillsdale, NJ, pp. 55–99.
- Kubovy, M. (1988) Should we resist the seductiveness of the space: time: vision: audition analogy? *J. Exp. Psychol. Hum. Percept. Perform.*, **14**, 318–320.
- Kubovy, M. & Van Valkenburg, D. (2001) Auditory and visual objects. *Cognition*, **80**, 97–126.
- Lakoff, G. & Johnson, M. (1999) *Philosophy in the Flesh: the Embodied Mind and its Challenge to Western Thought*. Basic Books, New York.
- Leopold, D.A. & Logothetis, N.K. (1999) Multistable phenomena: changing views in perception. *Trends Cogn. Sci.*, **3**, 254–264.
- Lieberman, A.M. (1982) On finding that speech is special. *Am. Psychol.*, **37**, 148–167.
- Lindsay, P.H. & Norman, D.A. (1977) *Human Information Processing*. Academic Press, New York.
- Mathiak, K., Hertrich, I., Lutzenberger, W. & Ackermann, H. (2001) Neural correlates of duplex perception: a whole-head magnetencephalography study. *Neuroreport*, **12**, 501–506.
- Näätänen, R. (1990) The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function. *Behav. Brain Sci.*, **13**, 201–288.
- Näätänen, R., Gaillard, A.W.K. & Mäntysalo, S. (1978) Early selective attention effect on evoked potential reinterpreted. *Acta Psychol.*, **42**, 313–329.
- Näätänen, R. & Picton, T.W. (1987) The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology*, **24**, 375–425.
- Näätänen, R. & Winkler, I. (1999) The concept of auditory stimulus representation in cognitive neuroscience. *Psychol. Bull.*, **125**, 826–859.
- Picton, T.W., Alain, C., Otten, L. & Ritter, W. (2000) Mismatch negativity: different water in the same river. *Audiol. Neuro-Otol.*, **5**, 111–139.
- Poeppl, D. (2003) The analysis of speech in different temporal integration windows: cerebral lateralization as ‘asymmetric sampling in time’. *Speech Comm.*, **41**, 245–255.
- Rand, T.C. (1974) Dichotic release from masking for speech. *J. Acoust. Soc. Am.*, **55**, 678–680.
- Ritter, W. & Ruchkin, D.S. (1992) A review of event-related potential components discovered in the context of studying P3. [In Friedman, D. & Bruder, G. (Eds), *Psychophysiology and experimental psychopathology – a tribute to Samuel Sutton*.] Ann. NY Acad. Sci., 658, pp. 1–32.
- Ritter, W., Sussman, E. & Molholm, S. (2000) Evidence that the mismatch negativity system works on the basis of objects. *Neuroreport*, **11**, 61–63.
- Rubin, E. (1915) *Synoplevede Figurer*. Gyldendalske, Copenhagen.
- Scherg, M., Vajsar, J. & Picton, T.W. (1989) A source analysis of the late human auditory evoked potentials. *J. Cogn. Neurosci.*, **1**, 336–355.
- Shamma, S. (2001) On the role of space and time in auditory processing. *Trends Cogn. Sci.*, **5**, 340–348.
- Sussman, E.S., Bregman, A.S., Wang, W.J. & Khan, F.J. (2005) Attentional modulation of electrophysiological activity in auditory cortex for unattended sounds within multistream auditory environments. *Cogn. Affect. Behav. Neurosci.*, **5**, 93–110.
- Sussman, E. & Gumenyuk, V. (2005) Organization of sequential sounds in auditory memory. *Neuroreport*, **16**, 1519–1523.
- Sussman, E., Horváth, J., Winkler, I. & Orr, M. (2006) The role of attention in the formation of auditory streams. *Percept. Psychophys.*, in press.
- Sussman, E., Ritter, W. & Vaughan, H.G. Jr (1998) Stimulus predictability and the mismatch negativity system. *Neuroreport*, **9**, 4167–4170.
- Sussman, E., Winkler, I., Huottilainen, M., Ritter, W. & Näätänen, R. (2002) Top-down effects on stimulus-driven auditory organization. *Cogn. Brain Res.*, **13**, 393–405.
- Sussman, E., Winkler, I. & Wang, W.J. (2003) MMN and attention: Competition for deviance detection. *Psychophysiol.*, **40**, 430–435.
- Takegata, R., Brattico, E., Tervaniemi, M., Varyiagina, O., Näätänen, R. & Winkler, I. (2005) Pre attentive representation of feature conjunctions for simultaneous, spatially distributed auditory objects. *Cogn. Brain Res.*, **25**, 169–179.
- Valdes-Sosa, M., Cobo, A. & Pinilla, T. (1998) Transparent motion and object-based attention. *Cognition*, **66**, B13–B23.
- Wang, W., Datta, H. & Sussman, E. (2005) The development of the length of the temporal window of integration for rapidly presented auditory information in 5–11-year-old children: Evidence from event-related brain potentials. *Clin. Neurophysiol.*, **116**, 1695–1706.
- Winkler, I. & Cowan, N. (2005) From sensory memory to long-term memory: Evidence from auditory memory reactivation studies. *Exp. Psychol.*, **52**, 3–20.
- Winkler, I., Czigler, I., Sussman, E., Horváth, J. & Balázs, L. (2005a) Preattentive binding of auditory and visual stimulus features. *J. Cogn. Neurosci.*, **17**, 320–339.
- Winkler, I., Horváth, J., Teder-Sälejärvi, W.A., Näätänen, R. & Sussman, E. (2003) Human auditory cortex tracks task-irrelevant sound sources. *Neuroreport*, **14**, 2053–2056.
- 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. & Schröger, E. (1995) Neural representation for the temporal structure of sound patterns. *Neuroreport*, **6**, 690–694.
- Winkler, I., Sussman, E., Tervaniemi, M., Ritter, W., Horváth, J. & Näätänen, R. (2003) Pre-attentive auditory context effects. *Cogn. Affect. Behav. Neurosci.*, **3**, 57–77.
- Winkler, I., Takegata, R. & Sussman, E. (2005b) Event-related brain potentials reveal multiple stages in the perceptual organization of sound. *Cogn. Brain Res.*, **25**, 291–299.
- Woldorff, M.G. (1993) Distortion of ERP averages due to overlap from temporally adjacent ERPs: Analysis and correction. *Psychophysiology*, **30**, 98–119.