Regular patterns stabilize auditory streams

Alexandra Bendixen[a)]
Department of General Psychology, Institute for Psychology, Hungarian Academy of Sciences, P.O. Box 398, H-1394 Budapest, Hungary

Susan L. Denham
School of Psychology, University of Plymouth, Drake Circus, Plymouth, Devon, PL4 8AA, United Kingdom

Kinga Gyimesi and István Winkler
Department of General Psychology, Institute for Psychology, Hungarian Academy of Sciences, P.O. Box 398, H-1394 Budapest, Hungary

(Received 14 January 2010; revised 20 September 2010; accepted 21 September 2010)

The auditory system continuously parses the acoustic environment into auditory objects, usually representing separate sound sources. Sound sources typically show characteristic emission patterns. These regular temporal sound patterns are possible cues for distinguishing sound sources. The present study was designed to test whether regular patterns are used as cues for source distinction and to specify the role that detecting these regularities may play in the process of auditory stream segregation. Participants were presented with tone sequences, and they were asked to continuously indicate whether they perceived the tones in terms of a single coherent sequence of sounds (integrated) or as two concurrent sound streams (segregated). Unknown to the participant, in some stimulus conditions, regular patterns were present in one or both putative streams. In all stimulus conditions, participants' perception switched back and forth between the two sound organizations. Importantly, regular patterns occurring in either one or both streams prolonged the mean duration of two-stream percepts, whereas the duration of one-stream percepts was unaffected. These results suggest that temporal regularities are utilized in auditory scene analysis. It appears that the role of this cue lies in stabilizing streams once they have been formed on the basis of simpler acoustic cues.
 [DOI: 10.1121/1.3500695]

PACS number(s): 43.66.Mk, 43.66.Ba, 43.66.Fe, 43.66.Qp [BCM] Pages: 3658–3666

I. INTRODUCTION

In a natural environment, the auditory system is constantly confronted with a mixture of sounds originating from several different sources. Relevant acoustic information can only be retrieved if the system succeeds in decomposing this mixture into meaningful auditory objects (Kubovy and Van Valkenburg, 2001; Griffiths and Warren, 2004; Winkler *et al*., 2009), also called streams (Bregman, 1990). In his influential monograph, Bregman (1990) introduced the term *auditory scene analysis* for the sound object decomposition problem. Almost 20 years later, understanding scene analysis still proves to be a major challenge for auditory research (for recent reviews, see Carlyon, 2004; Haykin and Chen, 2005; Denham and Winkler, 2006; Snyder and Alain, 2007). Two main types of cues for grouping sounds have been put forward (Bregman, 1990): Rapidly available cues, such as harmonicity and common onset (Alain *et al*., 2002; Hautus and Johnson, 2005), and sequential cues, such as sharing acoustic features over time (Moore and Gockel, 2002). The present study focuses on a sub-type of sequential cues based on the common experience that sound sources in everyday environments show characteristic temporal emission patterns (such as the sound of waves crashing on the shore). The aim of the study is to determine the role of such regular sound emission patterns in auditory stream segregation.

Static sequential cues for stream segregation, such as spectral separation and perceived location (van Noorden, 1975; Bregman, 1990; Moore and Gockel, 2002; Carlyon, 2004), are bound to fail in certain situations because most sources produce sounds covering a wide range of frequencies, and the location of the sound source relative to the listener is seldom constant. It is thus beneficial for the human auditory system to take into account the development of acoustic features over time (Jones, 1976; Darwin *et al*., 1995; Darwin, 1997). This enables the system to track and eventually predict changes in the source's behavior (e.g., due to movement) and, at the same time, to identify sound emission patterns that characterize a given source (e.g., due to constraints of the vocal tract).

Indeed, there is abundant evidence for the auditory system analyzing its input in a sequential manner in order to extract regular patterns. For example, studies measuring event-related brain potentials (ERPs) and behavioral responses have shown that violations of an established regularity are detected by the auditory system (e.g., Winkler and Schröger, 1995; Näätänen *et al*., 2001) and associated with processing costs (e.g., Schröger and Wolff, 1998), while processing benefits are observed for regularity-conforming tones (e.g., Jongsma *et al*., 2006). The extracted regularities are used for extrapolations about

[a)]Author to whom correspondence should be addressed. Current address: Institute for Psychology, University of Leipzig, Seeburgstr. 14-20, 04103 Leipzig, Germany. Electronic mail: alexandra.bendixen@uni-leipzig.de

 0001-4966/2010/128(6)/3658/9/$25.00

upcoming auditory events. This is shown by differences in auditory ERP responses elicited by predictable as compared to unpredictable tones (Winkler *et al*., 1996; Baldeweg, 2006; Bendixen *et al*., 2009). Such predictions have been discussed in terms of preparing for an event and thus saving processing resources (Schröger and Wolff, 1998; Sinkkonen, 1999; Bendixen *et al*., 2007). Jones and her colleagues (Jones, 1976; Jones *et al*., 1982; Jones and Boltz, 1989) have long argued for a more general role of temporal regularities in auditory perception. In particular, they suggested that the benefit of regularity extraction lies in providing a cue for auditory stream segregation (e.g., Jones *et al*., 1981). This suggestion has been taken up recently (Denham and Winkler, 2006; Winkler, 2007; Winkler *et al*., 2009; for a similar proposition, see Hung *et al*., 2001) and is tested in the present study.

A functional role of regularity extraction for auditory stream segregation can be claimed in a strong or in a weak sense. The strong hypothesis (Denham and Winkler, 2006; Winkler, 2007) states that regularities contribute to the initial grouping of the auditory input, during which representations for possible sound organizations are developed [the first stage of sound organization in Bregman's (1990) theory]. According to this hypothesis, sounds are more likely to be initially grouped together if they can be fitted into a common regular pattern. The weak hypothesis suggests that the representations of the initial sound organizations are developed on the basis of primary acoustic cues (such as separation in frequency and location), whereas regularities come into play only in the second stage of sound organization in which the auditory system decides between established representations of the alternative organizations (Bregman, 1990). The present study allows for a distinction between these variants of the general hypothesis on the function of regularities in stream segregation.

The study is based on a manipulation of the presence or absence of regularities in the *auditory streaming paradigm* (van Noorden, 1975). In this paradigm, participants are presented with a tone sequence that can be organized in two different ways, and they are then asked to indicate which organization they have heard (i.e., press one button for the one-stream percept and another button for the two-stream percept). The tone sequence is composed of a repetitive "ABA–" cycle, where "A" and "B" denote tones of different frequencies and "–" denotes a silent interval delivered with the same temporal parameters as the corresponding "B" tone. The perception of one vs two streams in the auditory streaming paradigm has been assumed to be largely determined by the spectral and temporal separation between the A and B tones (van Noorden, 1975). Following demonstrations that streams can be segregated on the basis of a variety of other cues as well (e.g., Vliegen and Oxenham, 1999; Grimault *et al*., 2002; Roberts *et al*., 2002; see also Akeroyd *et al*., 2005), the role of overall perceptual similarity between the A and B tones has been emphasized (Moore and Gockel, 2002).

Recent studies have presented long sequences of stimuli in the auditory streaming paradigm and asked participants about their current percept in a continuous manner (Roberts *et al*., 2002; Denham and Winkler, 2006; Pressnitzer and Hupé, 2006; Denham *et al*., 2010; see also Anstis and Saida, 1985). These studies revealed that, for a wide range of the acoustical parameters, the perception of such sequences fluctuates between alternative percepts, and that the characteristics of the alternations are very similar to bistable phenomena in vision, such as Rubin's vase-faces illusion (for a systematic comparison, see Pressnitzer and Hupé, 2006). The present study exploits this bistability and investigates how it can be biased by the introduction of regularities. First, spectral regularities are removed from the "ABA–" sequence by introducing jitter in frequency and intensity, separately for the "A" and "B" tones. As an important prerequisite to this approach, Denham *et al*. (2010) showed that perception remains bistable (i.e., oscillates between one- and two-stream percepts) with a moderate amount of random frequency jitter. In the present study, regularities are then re-established, separately in one or both features, and for one or both streams. These regularities are of higher order than the simple repetition used in the classical "ABA–" sequence. They preserve the range of variation in frequency and intensity, but replace the random occurrence of feature values with temporal patterns such as "low-middle-high-low-middle-high-…." The goal of the present study is to determine whether and, if so, how the presence of such regularities affects participants' reports of the perceived sound organization. Additionally, the effects of different numbers and types of regularities are assessed.

If, in line with the general hypothesis, auditory temporal regularities can serve as cues for stream segregation, the proportion of two-stream percepts should increase for stimulus configurations including separate regularities in one or both streams relative to fully irregular sequences. The weak variant of the general hypothesis predicts that this increase should be due to a prolongation of the mean duration of those perceptual phases in which participants hear two streams of sound. This is because, according to the weak hypothesis, the system has a tendency to maintain the segregated perceptual organization when it discovers that it is internally coherent (i.e., regular). The strong variant additionally predicts a shortening of the mean duration of the perceptual phases in which participants hear a single stream. This is because, in this variant of the main hypothesis, the regularities detected for the two-stream organization will influence grouping even when the integrated organization is currently dominant. The strong variant also predicts that the two-stream organization will emerge as the dominant organization earlier when the streams include separate regular patterns.

II. MATERIALS AND METHODS

A. Participants

Thirty healthy volunteers participated in the experiment. Data from four participants had to be excluded from the analysis due to difficulties in understanding or complying with the instructions (one responded before the stimulus was presented, another probably confused the two response buttons, and two appeared initially unsure of the task, then showed distinct change in their response pattern after reinstruction). The mean age of the remaining 26 participants (three left-handed, 16 male) was 21.7 yr. All participants had absolute thresholds not higher than 20 dB hearing level in the range 250–4000 Hz and no threshold difference exceeding 10 dB between the two ears (assessed with a Mediroll, SA-5 audiometer). None of the participants were taking any medication affecting the central nervous system. Prior to the beginning of the experiment, written informed consent was obtained from each participant according to the Declaration of Helsinki, after the experimental procedures and aims were explained to them. The study was approved by the Ethical Committee of the Institute for Psychology, Hungarian Academy of Sciences.

B. Apparatus and stimuli

Participants were seated in an acoustically shielded chamber. Sinusoidal tones with a mean level of 40 dB sensation level (above hearing threshold, adjusted individually for each participant) were presented binaurally via headphones in a continuous ABA– cycle. Participants were provided with two response buttons (one in each hand).

The ABA– cycle was delivered at a stimulus-onset asynchrony (SOA) of 150 ms for the individual constituents, resulting in a duration of 600 ms for the whole cycle. Consecutive "A" tones were separated by a 300-ms SOA with a "B" tone inserted midway between every second pair of consecutive "A" tones, thus giving a 600-ms SOA between consecutive "B" tones. The duration of each tone was 75 ms (including 10-ms rise and 10-ms fall times). The mean frequency of the "A" tones was 400 Hz, and the mean frequency of the "B" tones was seven semitones higher, i.e., 599 Hz.

The individual tones in the "A" and "B" sets varied both in frequency and level. The discrete frequency and intensity values in each set were chosen to preserve similarity within each set while promoting a clear differentiation between the two sets. The "A" tones took one of two frequency values (A1 = 384.1 Hz; A2 = 416.5 Hz) with equal probability. The "B" tones took one of three frequency values (B1 = 575.6 Hz; B2 = 599.3 Hz; B3 = 624.1 Hz) with equal probability. Both "A" and "B" tones were either *stressed* or *unstressed*, stressed tones being 6 dB higher in level than unstressed tones. Stress occurred on 25% of the "A" tones and 33% of the "B" tones in order to match the cycles of the frequency regularities (cf. next paragraph).

In ten stimulus conditions, the order of the different frequency and intensity values was either chosen randomly or followed predefined regular patterns separately for the "A" and "B" tones. Random sequences were created separately for each participant. Regularities were the same for all participants; they occurred in one or both sets of tones and in one or both features (see Fig. 1 for an overview and schematic illustration of the different conditions). Separate regularities for the "A" and "B" tones were chosen to ensure that no condition was fully regular in the integrated ("ABA") interpretation. The regularly repeating frequency pattern was "A2A2A1A1" for the "A" tones and "B1B2B3" for the "B" tones. The regular intensity pattern was a stress on every fourth "A" tone or on every third "B" tone. In the conditions in which frequency and intensity regularities were combined, stress coincided with the first tone of the repeating frequency pattern to facilitate pattern detection (Jones *et al.*, 1982).

C. Procedure

Participants were asked to listen to the tone sequences and to continuously indicate their percept by depressing one or both response buttons. Participants were instructed to choose between four response alternatives: *Integrated* (depress one button) if they heard the low and high tones within one coherent stream, *Segregated* (depress the other button) if they heard a low and a high stream in parallel, *Both* (depress both buttons) if they heard a pattern consisting of low and high tones and an additional separate stream consisting of only low or only high tones, and *Neither* (release both buttons) if their current percept did not fall into any of these categories. The assignment of *Integrated* and *Segregated* responses to the left- and right-hand buttons was counterbalanced across participants. Participants were encouraged to employ a neutral listening set, refraining from attempting to hear the sounds according to one or another perceptual organization. The experimenter made sure that participants understood the types of percepts they were required to report using both auditory and visual illustrations.

Each of the ten conditions was administered in two stimulus blocks of 2 min duration each. The order of the conditions was separately randomized for each participant with the restriction that each condition was presented once in the first half and once in the second half of the experiment. The net time of the experiment amounted to 40 min (10 conditions × 2 stimulus blocks × 2 min). There was at least a 30-s break between successive stimulus blocks, with additional relaxation time given to the participant as needed. All but one participant completed an ERP experiment prior to the current experiment. The ERP recording was unrelated to the present study.

D. Data recording and analysis

The state of the two response buttons was continuously recorded with a sampling rate of 250 Hz. Before analyzing the button presses, all states with duration shorter than 300 ms were discarded because these were assumed to represent transitions between two percepts. The underlying assumption was that participants may have been slightly inaccurate in timing their button presses and releases. After this correction (which on average led to the exclusion of 0.6% of the responses), perceptual *phases* were extracted from the participants' button presses. A perceptual phase is thus defined as the perception of the same sound organization for more than 300 ms.

Nine measures were derived separately for each participant and condition. The *proportion of Integrated* denotes the percentage of time in which an "Integrated" percept was reported. The *mean duration of Integrated* indicates the average duration of "Integrated" perceptual phases.[1] Proportions and mean durations for *Segregated*, *Both*, and *Neither* were defined in an analogous manner.[2] In addition, the *latency of the first Segregated percept* (the time from the start of the sound sequence to the onset of the first segregated perceptual phase) was determined.[3]

To test the hypotheses specified in Sec. I, in the first step of the analysis, the fully irregular condition 1 was compared with the fully regular condition 10 by conducting paired, two-tailed Student *t* tests on each of the nine dependent measures (see previous paragraph) with Bonferroni correction of the confidence levels. Based on the results of this analysis, in the second step, the relevant measures (those that were

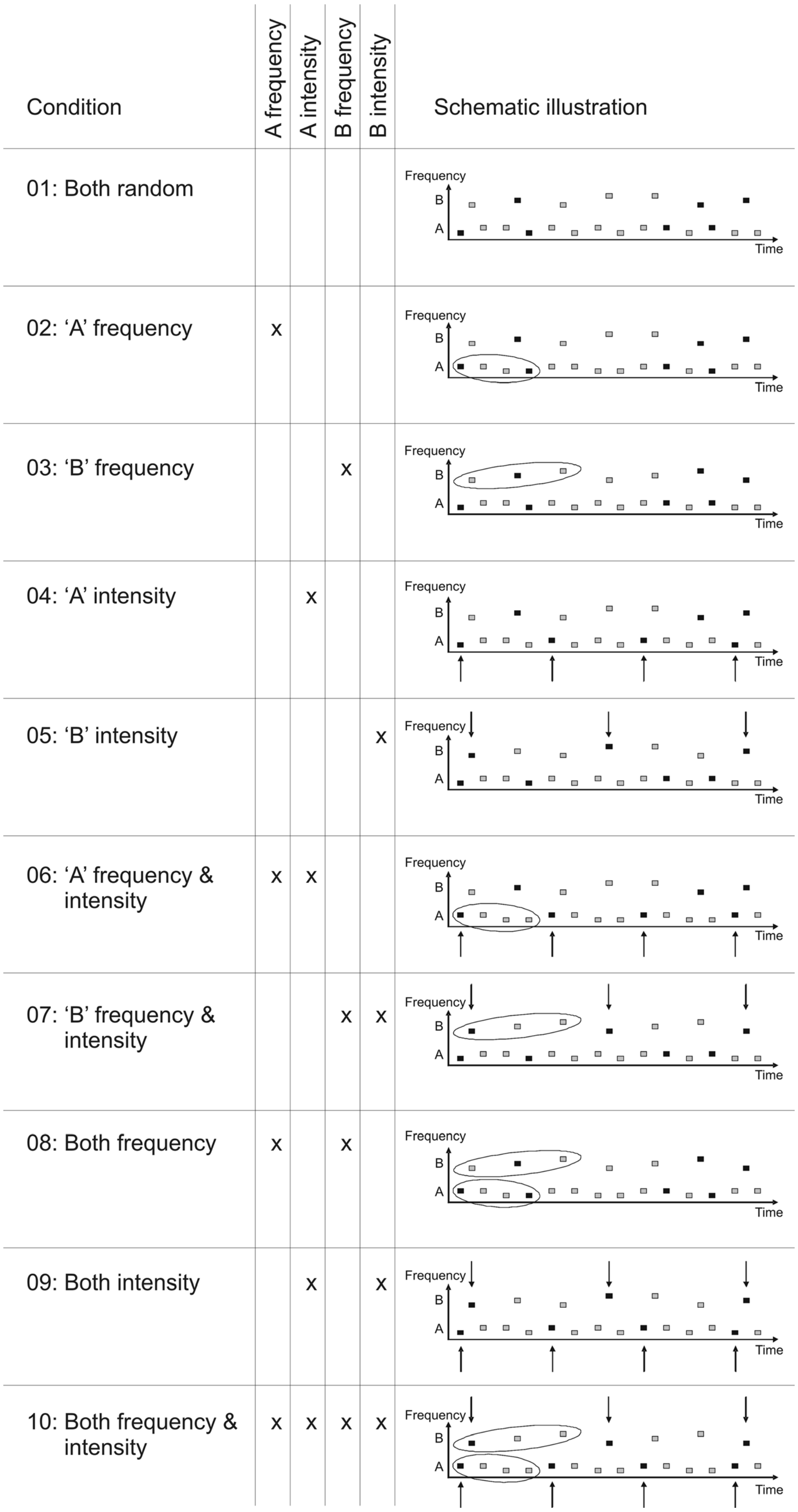

FIG. 1. Experimental design. Regularities (marked by "x" in the appropriate column) were manipulated independently for frequency and intensity and for the "A" and "B" tones. Condition names denote the regular feature (frequency, intensity, or frequency and intensity) as well as whether the regularity applies to one ("A" or "B") or both sets of tones. A stimulus example for each condition is schematically depicted in the right column. Filled and shaded squares indicate stressed and unstressed tones. Arrows indicate the cycle of intensity regularities in the schematic depictions; ellipses indicate the cycle of frequency regularities. Note that in order to reduce the length of the experimental session, the design was not fully crossed. Conditions with regularity in one feature for the "A" tones and regularity in the other feature for the "B" tones were not implemented; likewise, conditions with regularity in one feature for one of the tone sets and regularity in both features for the other tone set were not implemented.

affected by the overall presence of regularities) were selected and analyzed for conditions 2–9 to evaluate the impact of different types of regularities on stream segregation. First, the efficiency of the employed feature cues was determined by comparing the effects of frequency and intensity regularities in repeated-measures analyses of variance (ANOVAs) with the factors *Type of feature* (two levels: frequency, intensity) and *Stream* (three levels: A, B, both). Based on the results of these ANOVAs, data were collapsed across the *Type of feature* factor, and the final repeated-measures ANOVAs were calculated for the factors *Number of regular features* (two levels: one, two) and *Stream* (three levels: A, B, both). All

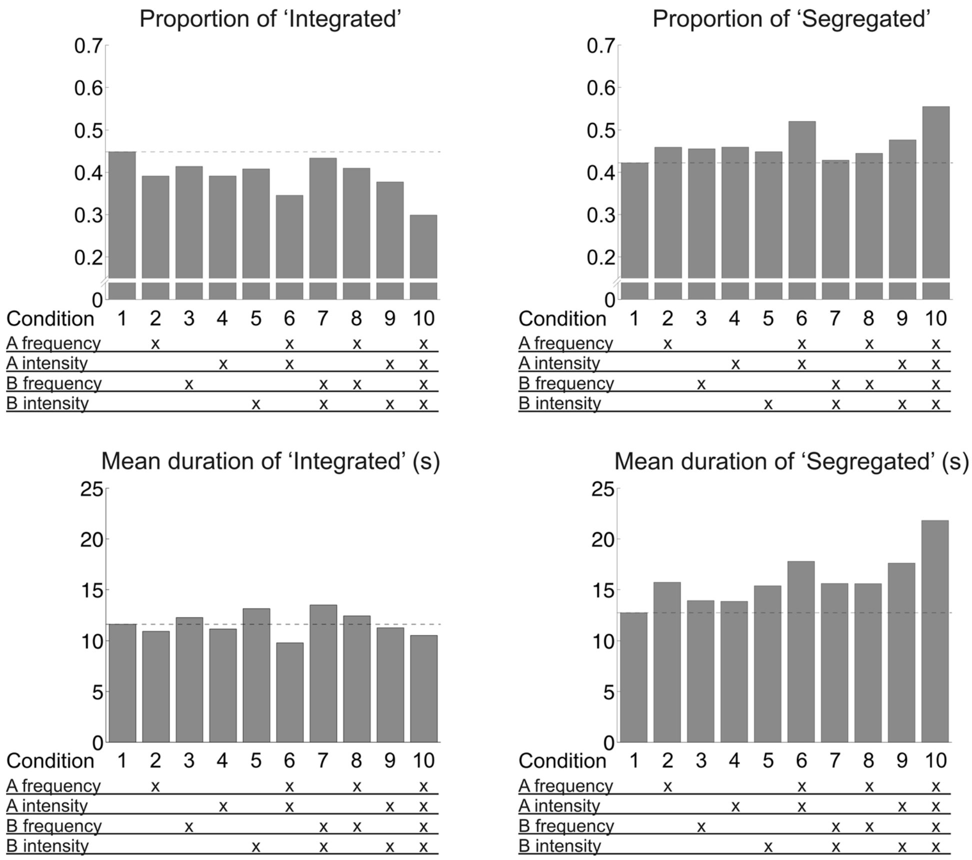

FIG. 2. Effects of the presence or absence of regularities on the proportion of the "Integrated" (top left) and "Segregated" (top right) percepts and on the average duration of the perceptual phases, separately for the "Integrated" (bottom left) and the "Segregated" (bottom right) percepts in each of the ten stimulus conditions. For each condition, the presence of frequency and intensity regularities in the "A" and "B" tones is marked by "x".

significant ANOVA effects and interactions are reported with the partial η^2 effect size measure. The Greenhouse–Geisser (1959) correction was applied where appropriate with the ϵ correction factors reported.

III. RESULTS

A. Comparison of the fully irregular and fully regular conditions

The proportions and mean durations of the "Integrated" and "Segregated" percepts in each condition are depicted in Fig. 2. The proportion of "Integrated" percepts was lower in the fully regular than in the fully irregular condition [$t(25) = 4.433$, $p_{\text{corrected}} = 0.0015$]. This was paralleled by an increase in the proportion of "Segregated" percepts [$t(25) = -3.527$, $p_{\text{corrected}} = 0.0149$]. The mean duration of "Integrated" percepts was not affected by the presence of the regularities [$t(25) = 0.760$, $p_{\text{corrected}} > 0.99$]. In contrast, the mean duration of "Segregated" percepts was significantly longer in the fully regular than in the fully irregular condition [$t(25) = -4.089$, $p_{\text{corrected}} = 0.0035$]. The latency of emergence of the first "Segregated" percept did not differ between the fully regular (15.23 s) and the fully irregular (16.76 s) conditions [$t(25) = 0.4521$, $p_{\text{corrected}} > 0.99$].

"Neither" percepts occurred very rarely (less than 0.6% in each condition), and their proportion and mean duration were unaffected by the presence of regularities (t values < 1.42, $p_{\text{corrected}} > 0.99$). "Both" percepts occurred with a frequency of 12%–17% in the various conditions. However, their proportion and mean duration were again unaffected by the presence of regularities (t values < 2.09, $p_{\text{corrected}} > 0.43$). Consequently, "Both" and "Neither" percepts were discarded from further analysis.

B. Detailed evaluation of the impact of different numbers and types of regularities

In a first step, frequency and intensity regularities were compared across the streams based on that subset of conditions in which only one feature was regular (2, 3, 4, 5, 8, and 9; see Fig. 1 for the assignment of condition numbers). In the ANOVA including the factors of *Type of feature* (*frequency*, *intensity*) and *Stream* (*A*, *B*, *both*), the *Type of feature* factor had no impact on the proportions and mean durations of "Integrated" and "Segregated" percepts (F values < 0.536, $p > 0.471$) nor was there a significant interaction between the *Type of feature* and *Stream* factors for any of these measures (F values < 0.666, $p > 0.518$). Thus, the effects of the two employed feature regularities did not significantly differ. Consequently, data were collapsed across the two features, which led to separate pooling of conditions 2 and 4 (one regular feature in the A stream), conditions 3 and 5 (one regular feature in the B stream), and conditions 8 and 9 (the same feature being regular in both streams).

In a second step of the analysis, these pooled conditions were contrasted with the corresponding conditions in which both features were regular (condition 6 for the A stream, 7 for the B stream, and 10 for both streams) in an ANOVA including the factors *Number of regular features* (two levels: one, two) and *Stream* (three levels: A, B, both). The results of this analysis are depicted in Fig. 3. The proportion of "Integrated" percepts was affected by the *Number of regular features* [$F(1,25) = 7.572$, $p = 0.011$, $\eta^2 = 0.232$], by the

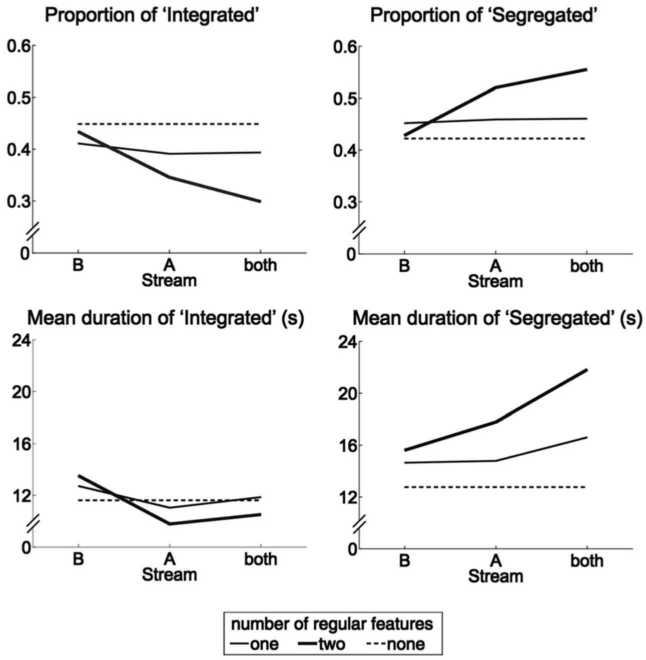

FIG. 3. Effects of the presence or absence of regularities in one or both sets of tones. Proportions (top) and average durations of the perceptual phases (bottom), separately for the "Integrated" (left) and "Segregated" (right) percepts, are compared between when the sequence of one ("A" or "B") or both sets of tones comprised a repeating pattern. The original experimental conditions have been pooled for this comparison (see main text for details). The baseline (i.e., fully irregular) condition is included for comparison (dashed line).

Stream factor [$F(2,50) = 4.449$, $p = 0.031$, $\epsilon = 1.46$, $\eta^2 = 0.151$], and by their interaction [$F(2,50) = 9.061$, $p = 0.0004$, $\eta^2 = 0.266$]. The interaction was due to the decrease in the proportion of the "Integrated" percept when both features were regular (as opposed to only one feature being regular) in the "A" stream and in both streams, but not in the "B" stream. A mirror-symmetric pattern was observed for the proportion of "Segregated" percepts, with significant effects of the *Number of regular features* [$F(1,25) = 6.294$, $p = 0.019$, $\eta^2 = 0.201$], of the *Stream* factor [$F(2,50) = 3.415$, $p = 0.041$, $\eta^2 = 0.120$], and of the interaction of the two factors [$F(2,50) = 6.362$, $p = 0.004$, $\eta^2 = 0.203$]. Again, two regular features had a larger effect than one regular feature when present in the "A" stream and in both streams, but not when present in the "B" stream.

The mean duration of "Integrated" percepts was not affected by either of the factors or by their interaction [*Number of regular features*: $F(1,25) = 0.520$, $p = 0.477$; *Stream*: $F(2,50) = 1.420$, $\varepsilon = 1.72$, $p = 0.248$; interaction: $F(2,50) = 2.477$, $p = 0.094$]. The mean duration of "Segregated" percepts was influenced by the *Number of regular features* [$F(1,25) = 9.908$, $p = 0.004$, $\eta^2 = 0.284$] with longer durations for two regular features than for one. No significant effect was observed for *Stream* [$F(2,50) = 1.584$, $\varepsilon = 1.47$, $p = 0.221$] or for the interaction of the two factors [$F(2,50) = 1.004$, $p = 0.374$].

IV. DISCUSSION

The present study was designed to investigate whether temporal regularities can serve as cues for auditory scene analysis and to specify the level at which the influence of such regularities might take place. The data support the view that temporal regularities are indeed used in auditory stream segregation, and that their impact occurs in the putative second stage of auditory scene analysis (Bregman, 1990), where they can stabilize the perception of separate streams, rather than in the first stage, in which the discovery of alternative sound organizations takes place.

As predicted by the general hypothesis, the presence of temporal regularities increased the proportion of the "Segregated" percept compared with the fully irregular stimulation. This effect can be attributed to the manipulation of the regularities, because all other acoustic cues (e.g., spectral separation, temporal structure) were held constant across the conditions. As our stimulus paradigm modeled natural source configurations in that different regularities were employed for the two sets of tones, the result can be interpreted as indicating that the auditory system recognizes the characteristic pattern of one or both sources and uses this information as an additional cue for maintaining their separation. Note that previous studies employing van Noorden's (1975) auditory streaming paradigm were ambiguous in this respect because the sequences were regular both within and across the streams. Predictability was manipulated in one previous study (French-St. George and Bregman, 1989), but the regularities in that study were simultaneously present in the two streams and in the whole sound sequence (i.e., for the segregated and the integrated organizations). The present results show that stimulus configurations containing different regular patterns for each putative stream are more likely to be interpreted as originating from multiple sound sources. This finding provides support for the proposition that the presence of regularities affects the decomposition of the auditory scene (Jones, 1976; Jones *et al.*, 1981; Hung *et al.*, 2001; Denham and Winkler, 2006; Winkler, 2007; Winkler *et al.*, 2009). Sensitivity to within-stream regularities enables the auditory system to detect characteristic sound emission patterns of concurrently active sources and to track changes in each source's behavior. Both of these functions are highly relevant in real environments.

It is important to consider how the increase in the proportion of "Segregated" percepts (and the parallel decrease in the proportion of "Integrated" percepts) emerged in the current experiment. These changes followed the pattern predicted by the weak variant of the general hypothesis. In particular, the mean phase duration of "Segregated" percepts was prolonged by the presence of the regularities, but the mean phase duration of "Integrated" percepts was not significantly affected. Both the comparison between the "extreme" conditions (fully irregular vs fully regular) and the detailed analysis of the intermediate conditions showed this pattern. Moreover, the presence of regularities did not decrease the latency of the first "Segregated" percept (for corroborating evidence, see French-St. George and Bregman, 1989). As specified in the weak hypothesis, this pattern of results suggests that once the "Segregated" organization became dominant, participants experienced this percept for longer durations in the presence of stream-specific temporal regularities than with the irregular stimulus configuration. However, the presence of temporal regularities did not promote switching to the "Segregated" organization when the currently dominant organization was "Integrated."

One way to conceptualize this effect is to assume that the "strength" of the representation of the dominant sound organization can be characterized by a dynamically changing threshold (Denham and Winkler, 2006; Pressnitzer and Hupé, 2006), which may decrease with time, e.g., by adaptation (Anstis and Saida, 1985). Switching occurs when the strength of the representation of the non-dominant sound organization exceeds the momentary threshold of the dominant one. Within this framework, the present results suggest that detection of temporal regularities within a given sound organization increases the strength of the dominant organization. However, temporal regularities are only discovered within the currently dominant sound organization. This notion is consistent with Bregman's (1978) view of pattern recognition processes operating upon sub-streams formed by preprocessing mechanisms. Further support for this assumption comes from results showing that the violation of a regular temporal pattern is only detected when the stimulus configuration or the attentional set strongly promote the corresponding sound organization (Sussman *et al.*, 1998, 1999; Shinozaki *et al.*, 2000; Winkler *et al.*, 2003; Sussman *et al.*, 2005). Thus temporal regularities may affect the competition between alternative sound organizations in an asymmetric way.

Unlike features such as spectral separation, the presence of a regular pattern does not constitute a "primitive" cue on which the initial grouping of the auditory scene is based (Bregman, 1990; see also French-St. George and Bregman, 1989), but a secondary factor influencing the stability of the perceptual organization with which it is associated. This interpretation is consistent with Bregman's (1990) view of the level at which regularities contribute to auditory scene analysis. Knowledge about the presence of regularities might influence the choice between alternative organizations, which have been formed in a first stage of grouping the auditory input (Bregman, 1990). The notion of two stages in auditory scene analysis has also received support from some recent ERP studies (Sussman, 2005; Winkler *et al.*, 2005; Snyder *et al.*, 2006).

The frequency and intensity regularities had effects of approximately equal size, but their presence in one or the other stream resulted in effects of different magnitude. There were larger benefits for the "Segregated" organization when regularities were introduced in the "A" than in the "B" stream, and this advantage was most pronounced when both features were regular. The asymmetry can be explained by assuming that the "A" stream was perceived as the foreground most of the time when participants reported hearing separate streams (the "Segregated" percept). Previous evidence suggests that when two or more concurrent streams are made up of qualitatively similar sounds, voluntarily selecting one stream as the foreground results in the rest of the auditory scene not being analyzed in a detailed manner (Alain and Woods, 1994; Brochard *et al.*, 1999; Cusack *et al.*, 2004; Sussman *et al.*, 2005). Assuming that (1) regularities are discovered only for the dominant sound organization and (2) concurrent streams made up of qualitatively similar sounds compete with each other (Cusack *et al.*, 2004; Denham and Winkler, 2006), selection of the "A" stream as the foreground most of the time during "Segregated" perceptual phases would explain the reduced impact of introducing regularities in the "B" stream. Please note, however, that in the present experiment, participants had no task-related reason to select one set of tones over the other as the foreground. Little is known about the stimulus-driven cues promoting the selection of foreground within an auditory scene. In the current stimulus sequences, the "A" and "B" tone sets differed in presentation rate, in mean frequency, and in the specific regularities employed for each set. Either one of these factors, or a combination of them, could have resulted in the "A" tones emerging predominantly as the foreground stream during the "Segregated" perceptual phases. For example, due to the higher presentation rate, regular patterns spanned less time in the "A" than in the "B" tones and could thus have become more salient. In any case, the asymmetry between the "A" and "B" streams suggests that regularities in the two streams are not processed together (e.g., looking for rhythmic differences). Rather, they are separately detected and evaluated within each stream (for corroborating evidence, see Ritter *et al.*, 2000). The independence of processing regularities within concurrent auditory streams further supports the notion that some initial grouping must occur before temporal regularities can affect perceptual organization.

Finally, the present data show that two features following a fully correlated regular sequential schedule stabilize the perception of two streams more than a single regular feature. This effect could be due to the fact that regularities are easier to discover when there is no uncorrelated variation in other sound features (Winkler *et al.*, 1990; Huotilainen *et al.*, 1993; Gomes *et al.*, 1995). Furthermore, regularities imply predictability of upcoming events, and fully predictable stimuli are known to be processed differently from partially predictable ones (Sussman and Winkler, 2001; Baess *et al.*, 2008; Bendixen *et al.*, 2009). Current theories suggest that predictive processing is an essential aspect of the auditory system (Winkler *et al.*, 1996; Baldeweg, 2006; Zanto *et al.*, 2006; Grimm and Schröger, 2007; Schröger, 2007; Bendixen *et al.*, 2009; Winkler *et al.*, 2009). A link between regularity extraction and stream segregation has been suggested on the basis of the similarity of the underlying neural mechanisms (Micheyl *et al.*, 2005) and of the common information requirements (Winkler, 2007; Winkler *et al.*, 2009). Moreover, some models of auditory scene analysis have successfully employed predictive principles (Ellis, 1999; Goto and Muraoka, 1999; Masuda-Katsuse and Kawahara, 1999; Dubnov, 2008). Improved predictability (due to two fully correlated temporal feature patterns) can thus result in the formation of a stronger or more stable representation of the corresponding auditory stream (cf. Denham and Winkler, 2006; Winkler *et al.*, 2009).

In conclusion, the present results support a functional role of regularities in auditory scene analysis. Regular temporal sound patterns appear to affect the second stage of auditory scene analysis, when the alternative sound organizations formed in the first stage compete for dominance (i.e., for conscious perception). In contrast, no evidence was obtained indicating that regularities are used as cues for initial grouping during the first stage of auditory scene analysis.

ACKNOWLEDGMENTS

This work was supported by the German Research Foundation (DFG, Grant No. BE 4284/1-1 to A.B.) and by the European Commission's Seventh Framework Programme (Grant No. ICT-FP7-231168 to I.W. and S.L.D.). The experiment was realized using Cogent 2000 developed by the Cogent 2000 team at the FIL and the ICN. The authors are grateful to Brian C.J. Moore and two anonymous reviewers for their constructive comments on an earlier version of the manuscript.

[1]All duration analyses were repeated with log-transformed duration values to allow for skewed duration distributions. Statistical results were similar to those obtained with the raw duration values.

[2]Proportions and mean durations were initially analyzed separately for all phases including the first phase (termed *all phases*) and with the first phase excluded (termed *subsequent phases*) because previous studies (Denham *et al*., 2010) revealed characteristic differences between the first and subsequently reported perceptual phases regarding the effects of primary cues for stream segregation. As the results were virtually identical regardless of whether the first phase was included or excluded (*all* vs *subsequent* reported perceptual phases), analyses are reported for *all phases* only. Furthermore, measures were initially calculated separately for the first and second presentation of each condition. The *Time of presentation* (first vs second run) had a significant effect on the proportion of "Both" percepts [$F(1,25) = 7.673$, $p = 0.010$] due to a higher proportion of "Both" percepts reported in the first than in the second run, and on the mean duration of "Both" percepts [$F(1,25) = 5.267$, $p = 0.030$] due to a longer duration of "Both" percepts in the first than in the second run. These effects might result from the adjustment of an adaptation level or response criterion as participants gained more experience regarding the variability present in the stimulus set. Importantly, no interaction was obtained between *Time of presentation* and *Condition* for any of the dependent variables (all p values > 0.05). The time of presentation thus affected participants' responses independently of the main experimental manipulation. Consequently, data were pooled for the two presentations. All subsequent reports and analyses are based on the pooled measures.

[3]The *latency of the first Segregated percept* was chosen to reflect early perceptual phases elicited in the various conditions. Please note that it is not possible to analyze the duration of the first phase separately for the four different types of percepts because there is only one type of percept per participant and condition (i.e., it is either Integrated, Segregated, or Both; Neither is indistinguishable from the delay to the first response). Therefore, each subject in each condition only contributes to the data of one of these possible percepts.

Akeroyd, M. A., Carlyon, R. P., and Deeks, J. M. (**2005**). "Can dichotic pitches form two streams?," J. Acoust. Soc. Am. **118**, 977–981.

Alain, C., Schuler, B. M., and McDonald, K. L. (**2002**). "Neural activity associated with distinguishing concurrent auditory objects," J. Acoust. Soc. Am. **111**, 990–995.

Alain, C., and Woods, D. L. (**1994**). "Signal clustering modulates auditory cortical activity in humans," Percept. Psychophys. **56**, 501–516.

Anstis, S., and Saida, S. (**1985**). "Adaptation to auditory streaming of frequency-modulated tones," J. Exp. Psychol. Hum. Percept. Perform. **11**, 257–271.

Baess, P., Jacobsen, T., and Schröger, E. (**2008**). "Suppression of the auditory N1 event-related potential component with unpredictable self-initiated tones: Evidence for internal forward models with dynamic stimulation," Int. J. Psychophysiol. **70**, 137–143.

Baldeweg, T. (**2006**). "Repetition effects to sounds: Evidence for predictive coding in the auditory system," Trends Cogn. Sci. **10**, 93–94.

Bendixen, A., Roeber, U., and Schröger, E. (**2007**). "Regularity extraction and application in dynamic auditory stimulus sequences," J. Cogn. Neurosci. **19**, 1664–1677.

Bendixen, A., Schröger, E., and Winkler, I. (**2009**). "I heard that coming: Event-related potential evidence for stimulus-driven prediction in the auditory system," J. Neurosci. **29**, 8447–8451.

Bregman, A. S. (**1978**). "Auditory streaming is cumulative," J. Exp. Psychol. Hum. Percept. Perform. **4**, 380–387.

Bregman, A. S. (**1990**). *Auditory Scene Analysis. The Perceptual Organization of Sound* (MIT Press, Cambridge, MA), pp. 47–184, 411–453.

Brochard, R., Drake, C., Botte, M.-C., and 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.

Carlyon, R. P. (**2004**). "How the brain separates sounds," Trends Cogn. Sci. **8**, 465–471.

Cusack, R., Deeks, J., Aikman, G., and Carlyon, R. P. (**2004**). "Effects of location, frequency region, and time course of selective attention on auditory scene analysis," J. Exp. Psychol. Hum. Percept. Perform. **30**, 643–656.

Darwin, C. J. (**1997**). "Auditory grouping," Trends Cogn. Sci. **1**, 327–333.

Darwin, C. J., Hukin, R. W., and Alkhatib, B. Y. (**1995**). "Grouping in pitch perception: Evidence for sequential constraints," J. Acoust. Soc. Am. **98**, 880–885.

Denham, S. L., Gyimesi, K., Stefanics, G., and Winkler, I. (**2010**). "Stability of perceptual organisation in auditory streaming," in *The Neurophysiological Bases of Auditory Perception*, edited by E. A. Lopez-Poveda, A. R. Palmer, and R. Meddis (Springer, New York), pp. 477–488.

Denham, S. L., and Winkler, I. (**2006**). "The role of predictive models in the formation of auditory streams," J. Physiol. Paris **100**, 154–170.

Dubnov, S. (**2008**). "Unified view of prediction and repetition structure in audio signals with application to interest point detection," IEEE Trans. Audio Speech Lang. Proc. **16**, 327–337.

Ellis, D. P. W. (**1999**). "Using knowledge to organize sound: The prediction-driven approach to computational auditory scene analysis and its application to speech/nonspeech mixtures," Speech Commun. **27**, 281–298.

French-St. George, M., and Bregman, A. S. (**1989**). "Role of predictability of sequence in auditory stream segregation," Percept. Psychophys. **46**, 384–386.

Gomes, H., Ritter, W., and Vaughan, H. G., Jr. (**1995**). "The nature of preattentive storage in the auditory system," J. Cogn. Neurosci. **7**, 81–94.

Goto, M., and Muraoka, Y. (**1999**). "Real-time beat tracking for drumless audio signals: Chord change detection for musical decisions," Speech Commun. **27**, 311–335.

Greenhouse, S. W., and Geisser, S. (**1959**). "On methods in the analysis of profile data," Psychometrika **24**, 95–112.

Griffiths, T. D., and Warren, J. D. (**2004**). "What is an auditory object?," Nat. Rev. Neurosci. **5**, 887–892.

Grimault, N., Bacon, S. P., and Micheyl, C. (**2002**). "Auditory stream segregation on the basis of amplitude-modulation rate," J. Acoust. Soc. Am. **111**, 1340–1348.

Grimm, S., and Schröger, E. (**2007**). "The processing of frequency deviations within sounds: Evidence for the predictive nature of the mismatch negativity (MMN) system," Restor. Neurol. Neurosci. **25**, 241–249.

Hautus, M. J., and Johnson, B. W. (**2005**). "Object-related brain potentials associated with the perceptual segregation of a dichotically embedded pitch," J. Acoust. Soc. Am. **117**, 275–280.

Haykin, S., and Chen, Z. (**2005**). "The cocktail party problem," Neural Comput. **17**, 1875–1902.

Hung, J., Jones, S. J., and Vaz Pato, M. (**2001**). "Scalp potentials to pitch change in rapid tone sequences. A correlate of sequential stream segregation," Exp. Brain Res. **140**, 56–65.

Huotilainen, M., Ilmoniemi, R. J., Lavikainen, J., Tiitinen, H., Alho, K., Sinkkonen, J., Knuutila, J., and Näätänen, R. (**1993**). "Interaction between representations of different features of auditory sensory memory," Neuroreport **4**, 1279–1281.

Jones, M. R. (**1976**). "Time, our lost dimension: Toward a new theory of perception, attention, and memory," Psychol. Rev. **83**, 323–355.

Jones, M. R., and Boltz, M. (**1989**). "Dynamic attending and responses to time," Psychol. Rev. **96**, 459–491.

Jones, M. R., Boltz, M., and Kidd, G. (**1982**). "Controlled attending as a function of melodic and temporal context," Percept. Psychophys. **32**, 211–218.

Jones, M. R., Kidd, G., and Wetzel, R. (**1981**). "Evidence for rhythmic attention," J. Exp. Psychol. Hum. Percept. Perform. **7**, 1059–1073.

Jongsma, M. L. A., Eichele, T., Van Rijn, C. M., Coenen, A. M. L., Hugdahl, K., Nordby, H., and Quian Quiroga, R. (**2006**). "Tracking pattern learning with single-trial event-related potentials," Clin. Neurophysiol. **117**, 1957–1973.

Kubovy, M., and Van Valkenburg, D. (**2001**). "Auditory and visual objects," Cognition **80**, 97–126.

Masuda-Katsuse, I., and Kawahara, H. (**1999**). "Dynamic sound stream formation based on continuity of spectral change," Speech Commun. **27**, 235–259.

Micheyl, C., Tian, B., Carlyon, R. P., and Rauschecker, J. P. (**2005**). "Perceptual organization of tone sequences in the auditory cortex of awake macaques," Neuron **48**, 139–148.

Moore, B. C. J., and Gockel, H. (**2002**). "Factors influencing sequential stream segregation," Acta Acust. United Acust. **88**, 320–333.

Näätänen, R., Tervaniemi, M., Sussman, E., Paavilainen, P., and Winkler, I. (**2001**). "Primitive intelligence in the auditory cortex," Trends Neurosci. **24**, 283–288.

Pressnitzer, D., and Hupé, J. M. (**2006**). "Temporal dynamics of auditory and visual bistability reveal common principles of perceptual organization," Curr. Biol. **16**, 1351–1357.

Ritter, W., Sussman, E., and Molholm, S. (**2000**). "Evidence that the mismatch negativity system works on the basis of objects," Neuroreport **11**, 61–63.

Roberts, B., Glasberg, B. R., and Moore, B. C. J. (**2002**). "Primitive stream segregation of tone sequences without differences in fundamental frequency or passband," J. Acoust. Soc. Am. **112**, 2074–2085.

Schröger, E. (**2007**). "Mismatch negativity: A microphone into auditory memory," J. Psychophysiol. **21**, 138–146.

Schröger, E., and Wolff, C. (**1998**). "Behavioral and electrophysiological effects of task-irrelevant sound change: A new distraction paradigm," Cogn. Brain Res. **7**, 71–87.

Shinozaki, N., Yabe, H., Sato, Y., Sutoh, T., Hiruma, T., Nashida, T., and 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 *Information Theory and the Brain*, edited by R. Baddeley, P. Hancock, and P. Foldiak (Cambridge University Press, Cambridge), pp. 241–254.

Snyder, J. S., and Alain, C. (**2007**). "Toward a neurophysiological theory of auditory stream segregation," Psychol. Bull. **133**, 780–799.

Snyder, J. S., Alain, C., and Picton, T. W. (**2006**). "Effects of attention on neuroelectric correlates of auditory stream segregation," J. Cogn. Neurosci. **18**, 1–13.

Sussman, E. (**2005**). "Integration and segregation in auditory scene analysis," J. Acoust. Soc. Am. **117**, 1285–1298.

Sussman, E., Bregman, A. S., Wang, W. J., and 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., Ritter, W., and Vaughan, H. G., Jr. (**1998**). "Attention affects the organization of auditory input associated with the mismatch negativity system," Brain Res. **789**, 130–138.

Sussman, E., Ritter, W., and Vaughan, H. G., Jr. (**1999**). "An investigation of the auditory streaming effect using event-related brain potentials," Psychophysiology **36**, 22–34.

Sussman, E., and Winkler, I. (**2001**). "Dynamic sensory updating in the auditory system," Cogn. Brain Res. **12**, 431–439.

van Noorden, L. P. A. S. (**1975**). "Temporal coherence in the perception of tone sequences," Doctoral dissertation, Technical University Eindhoven, Eindhoven.

Vliegen, J., and Oxenham, A. J. (**1999**). "Sequential stream segregation in the absence of spectral cues," J. Acoust. Soc. Am. **105**, 339–346.

Winkler, I. (**2007**). "Interpreting the mismatch negativity," J. Psychophysiol. **21**, 147–163.

Winkler, I., Denham, S. L., and Nelken, I. (**2009**). "Modeling the auditory scene: Predictive regularity representations and perceptual objects," Trends Cogn. Sci. **13**, 532–540.

Winkler, I., Karmos, G., and Näätänen, R. (**1996**). "Adaptive modeling of the unattended acoustic environment reflected in the mismatch negativity event-related potential," Brain Res. **742**, 239–252.

Winkler, I., Paavilainen, P., Alho, K., Reinikainen, K., Sams, M., and 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., and Schröger, E. (**1995**). "Neural representation for the temporal structure of sound patterns," Neuroreport **6**, 690–694.

Winkler, I., Sussman, E., Tervaniemi, M., Horváth, J., Ritter, W., and Näätänen, R. (**2003**). "Preattentive auditory context effects," Cogn. Affect. Behav. Neurosci. **3**, 57–77.

Winkler, I., Takegata, R., and Sussman, E. (**2005**). "Event-related brain potentials reveal multiple stages in the perceptual organization of sound," Cogn. Brain Res. **25**, 291–299.

Zanto, T. P., Snyder, J. S., and Large, E. W. (**2006**). "Neural correlates of rhythmic expectancy," Adv. Cogn. Psychol. **2**, 221–231.