DIFFERENT ROLES OF SIMILARITY AND PREDICTABILITY IN AUDITORY STREAM SEGREGATION

ALEXANDRA BENDIXEN[1,2], TAMÁS M. BŐHM[3,4], ORSOLYA SZALÁRDY[3], ROBERT MILL[5], SUSAN L. DENHAM[5] and ISTVÁN WINKLER[3,6]

[1]Institute of Psychology, University of Leipzig, Seeburgstr. 14–20, D-04103 Leipzig, Germany

[2]Institute of Psychology, Carl von Ossietzky University of Oldenburg, Ammerländer Heerstr. 114–118, D-26129 Oldenburg, Germany

[3]Institute of Cognitive Neuroscience and Psychology, Research Centre for Natural Sciences, Hungarian Academy of Sciences, H-1394 Budapest, P.O. Box 398, Hungary

[4]Department of Telecommunications and Media Informatics, Budapest University of Technology and Economics, H-1117 Budapest, Magyar tudósok krt. 2, Hungary

[5]Cognition Institute and School of Psychology, University of Plymouth, Drake Circus, Plymouth PL4 8AA, UK

[6]Institute of Psychology, University of Szeged, H-6722 Szeged, Petőfi S. sgt. 30–34, Hungary

Abstract

Sound sources often emit trains of discrete sounds, such as a series of footsteps. Previously, two different principles have been suggested for how the human auditory system binds discrete sounds together into perceptual units. The feature similarity principle is based on linking sounds with similar characteristics over time. The predictability principle is based on linking sounds that follow each other in a predictable manner. The present study compared the effects of these two principles. Participants were presented with tone sequences and instructed to continuously indicate whether they perceived a single coherent sequence or two concurrent streams of sound. We investigated the influence of separate manipulations of similarity and predictability on these perceptual reports. Both grouping principles affected perception of the tone sequences, albeit with different characteristics. In particular, results suggest that whereas predictability is only analyzed for the currently perceived sound organization, feature similarity is also analyzed for alternative groupings of sound. Moreover, changing similarity or predictability within an ongoing sound sequence led to markedly different dynamic effects. Taken together, these results provide evidence for different roles of similarity and predictability in auditory scene analysis, suggesting that forming auditory stream representations and competition between alternatives rely on partly different processes.

Keywords**:** sound perception, auditory scene analysis, streaming, auditory object formation, perceptual bi-stability, perceptual switching, primary sound features, sound patterns, higher-order sound feature, feature proximity

INTRODUCTION

Many auditory sources emit signals in a discontinuous manner over time. For instance, a human speaker pauses between words (at intonational phrase boundaries) or even within words (as part of articulating a stop or an affricate consonant). As a consequence, the auditory system is often confronted with discrete sound events that need to be bound together in order to retrieve relevant information. At the same time, binding events together that were actually emitted by two different sources (e.g., two speakers talking in parallel) usually needs to be avoided. This problem is described as *horizontal* or *sequential* grouping within the framework of *auditory scene analysis* (Bregman 1990). Auditory scene analysis describes the decomposition of a mixture of sounds into meaningful perceptual units (Bregman 1990; Kubovy and Van Valkenburg 2001; Griffiths and Warren 2004; Winkler et al. 2009), termed auditory streams. Several cues for sequential grouping are based on sounds sharing acoustic features over time (Moore and Gockel 2002). The notion of grouping based on feature similarity originates from the *Gestalt* law of proximity (Wertheimer 1923): If two sound events are similar in all of their features, they are likely to have been emitted by the same source (Bregman 1990). Natural sound sources seldom change the features of their emissions in an abrupt manner. Similarity is further qualified by the passage of time: Even small feature differences may lead to experiencing contrast when the sounds are delivered in short succession; in contrast, sounds with moderate amounts of feature difference may be regarded as similar when delivered with somewhat longer temporal separation (van Noorden 1975; Jones 1976; Bregman 1993). Therefore, we regard, and within this paper refer to, feature similarity as the inverse of the temporal rate of feature change. In other words, two sounds within a sequence are similar when the rate of change between them is low.

The feature similarity view of auditory grouping has been contrasted with the suggestion of feature predictability (e.g., Jones 1976; Jones et al. 1981), expressing the view that sound events are grouped together if their development over time is regular and thus predictable. Predictability is present in everyday environments when a sound source's parameters either stay constant or change in a regular manner over time (usually smoothly, such as when the changes have been caused by the relative movement of the source and the listener). It is beneficial for the auditory system to use such predictive information for tracking changes in the source's behavior (Winkler 2007; Winkler et al. 2009; Winkler et al. 2012). Predictability can also be found in sound emission patterns that characterize a given source (e.g., the pattern emitted by a train moving on the rails). The present study was conducted to characterize the relationship between the effects of feature similarity and predictability in auditory scene analysis.

Auditory stream segregation has been most widely studied 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. They are then asked to indicate which organization they 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 tones. The perception of one vs. two streams in the auditory streaming paradigm is largely determined by the feature separation between the A and B tones (Moore and Gockel 2002) evaluated against their temporal separation (van Noorden

1975). Perceptual similarity is based on a variety of features including spectral separation, differences in perceived location, and amplitude modulation differences (e.g., van Noorden 1975; Vliegen and Oxenham 1999; Grimault et al. 2002; Roberts et al. 2002; Akeroyd et al. 2005; Szalárdy et al. in press).

Such findings are consistent with the similarity view of sequential auditory grouping (Bregman 1990) and with our current qualification of similarity in terms of the inverse rate of feature change. Typically, the feature values within the 'A' and 'B' sets of tones are chosen to be identical; they are thus perfectly similar as well as entirely predictable. In this situation, it is difficult to demonstrate a specific benefit of predictability over and above that of similarity. When regarding predictability as a more specific version of the similarity principle, then either 1. predictability should take over the role of similarity in auditory grouping (i.e., groups can only be formed from a predictable sequence of sounds; similarity alone is not a sufficient prerequisite of grouping) or 2. binding by predictability can only occur for groups that have been initially formed on the basis of the similarity principle (with predictability providing further advantages to sound grouping). A study by Denham and colleagues (2010) showed that reducing predictability within the A and B set of sounds by introducing a moderate amount of random frequency jitter did not modify perceptual reports on sound organization. Moreover, streaming can be experienced when the temporal order of events from the A and B set of sounds is entirely random (e.g., Müller et al. 2005). These results rule out the first (strong) alternative of the role of predictability in auditory scene analysis. They were previously interpreted as suggesting that predictability is not an important cue in auditory scene analysis (see also French-St.George and Bregman 1989).

However, recent results of Bendixen and colleagues (2010) suggest otherwise. In this study, random frequency jitter was contrasted with regular patterns inserted separately into the A or B set of sounds. There were, for instance, three slightly different versions of the 'B' sound that were either arranged randomly or in a repeating regular order (B1–B2–B3). The introduction of separate regularities within each set of sounds ensured that no condition was fully regular in the integrated ('ABA') interpretation. Results of Bendixen and colleagues (2010) reveal that when regular patterns are present within one or both set of sounds, participants are more likely to hear the sequence as splitting into two separate streams. These results support the second (weak) alternative of the role of predictability in auditory scene analysis. Consistent evidence in favor of an effect of predictability on auditory scene analysis was obtained by Andreou and colleagues (2011).

Alternatively, similarity- and predictability-based grouping might operate in parallel in the human brain, neither being bound by the other. If this was the case, the effects of predictability should be independent of those of similarity. For instance, one should fi nd perceptual groups formed of predictable, but dissimilar sound events; as well as streams consisting of a predictable subset of similar sounds. The results of Winkler and colleagues (2003a; 2003b) argue against the latter possibility. These authors found that inserting random sounds between successive sounds of a regular sequence resulted in the auditory system losing track of the regularity when all sounds were similar, instead of segregating the regular and the random sounds. Furthermore, in Bendixen et al.'s (2010) study, the effect of within-stream predictability was limited to perceptual stabilization of the organization containing the regular pattern. The authors suggested that detecting predictability stabilizes perceptual coherence in an already-formed auditory stream, while the initial process of stream formation is driven by simpler analyses considering only feature similarity.

In Bendixen et al.'s (2010) study, the distinction between stream formation and stabilization was made possible by presenting long sequences of stimuli in the auditory streaming paradigm and asking participants about their current percept in a continuous manner. This procedure was introduced by Anstis and Saida (1985) and subsequently employed in a number of studies (Roberts et al. 2002; Denham and Winkler 2006; Pressnitzer and Hupé 2006; Kondo and Kashino 2009; Denham et al. 2010, in press; Szalárdy et al. in press). These studies revealed that, for a wide range of the acoustical parameters, the perception of such sequences fluctuated between alternative percepts, and that the characteristics of the perceptual switches were very similar to bi-stable phenomena in vision, such as the Necker cube (for a systematic comparison, see Pressnitzer and Hupé 2006). By separately analyzing the overall distribution of perceptual phases and the average phase duration for each percept, it is possible to separate cues that initiate perceptual switches from cues that stabilize a given percept but do not cause switching towards it (Bendixen et al. 2010; Szalárdy et al. in press).

The argument that similarity triggers stream formation and predictability stabilizes the resulting streams was based on comparing the effects of similarity (Denham and Winkler 2006; Denham et al. 2010) and predictability (Bendixen et al. 2010) across different experiments using the bi-stable version of the auditory streaming paradigm. A direct comparison of the two types of cues has not yet been provided. The present study was thus designed to manipulate similarity and predictability within the same experiment in order to confirm previous observations and test some further predictions derived from them. Specifically, similarity was manipulated by changing the frequency separation between the A and B tones, and predictability was manipulated by introducing jitter in frequency and intensity within each set of tones, which was implemented either in a random manner or in a predictable manner by means of stream-specific regularities (e.g., 'low-middle-high-low-middle-high-...'). Furthermore, in some conditions all cues were retained throughout the block whereas in others either similarity or predictability was changed in the middle of the block in order to investigate the temporal dynamics of how changes in perceptual organization follow the parameter changes.

Based on Bendixen et al.'s (2010) suggestion that auditory streams are initially formed on the basis of feature separation (the lack of similarity), whereas predictability can stabilize them, the following pattern of results is expected. 1. The proportion of two-stream percepts is higher in conditions with high frequency separation (low similarity) than in conditions with low frequency separation (high similarity). 2. The increased proportion of two-stream percepts is caused by a prolongation of those perceptual phases in which participants report hearing two sound streams and a parallel shortening of those ones in which participants report hearing a single stream. 3. The proportion of two-stream percepts is higher in conditions with stream-specific regular patterns (high predictability) than in conditions without stream-specific regular patterns (low predictability). 4. In this case, the increased proportion of two-stream percepts is caused by a prolongation of those perceptual phases in which participants report hearing two streams of sound, with no shortening of those perceptual phases in which participants report hearing a single stream. 5. The two-stream organization emerges as the dominant organization earlier in conditions with high frequency separation (low similarity) than in conditions with low frequency separation (high similarity). 6. The two-stream organization does not emerge as the dominant organization earlier in conditions with stream-specific regular patterns (high predictability) than in conditions without stream-specific regular patterns (low predictability).

MATERIALS AND METHODS

Participants

Thirty-three healthy volunteers with self-reported normal hearing participated in the experiment. Data from three participants had to be excluded from the analysis due to difficulties in fulfilling the task (all three appeared unsure of the task; two completed the experiment but used the option to not report any percept most of the time, one asked to cancel the experiment). The mean age of the remaining 30 participants (all right-handed, 8 male) was 22.4 years. 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.

Apparatus and stimuli

Participants were seated in an acoustically shielded chamber. Sinusoidal tones with a mean level of 70 dB sound pressure level were presented binaurally via headphones in a continuous 'ABA–' cycle. Participants were provided with a response keypad containing four buttons to be pressed with the middle and index fingers of their left and right hands.

The 'ABA–' cycle was delivered at a stimulus-onset asynchrony (SOA) of 150 ms between successive elements, thus giving 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 100 ms (including 10 ms rise and 10 ms fall times).

Stimuli were arranged in 12 conditions[1] (see *Figure 1* for an overview and schematic illustration) defined by the factors frequency separation (high vs. low), regularity (present vs. absent), and change (no change vs. change in frequency separation vs. change in regularity). Each condition was administered as a 4-minute block during the experimental session.

In the *high frequency separation* conditions, the mean frequency of the 'A' tones was 400 Hz, and the mean frequency of the 'B' tones was 7 semitones higher, i.e., 599 Hz. In the *low frequency separation* conditions, the mean frequency of the 'A' tones was increased by one semitone to 424 Hz, and the mean frequency of the 'B' tones was decreased by one semitone to 566 Hz, leaving a frequency separation of 5 semitones between the 'A' and 'B' tones. The rationale for using such a small range of frequency separation values (5 vs. 7 semitones) was to achieve comparable effect sizes for the similarity and predictability manipulations (based on the predictability effect obtained in Bendixen et al., 2010).

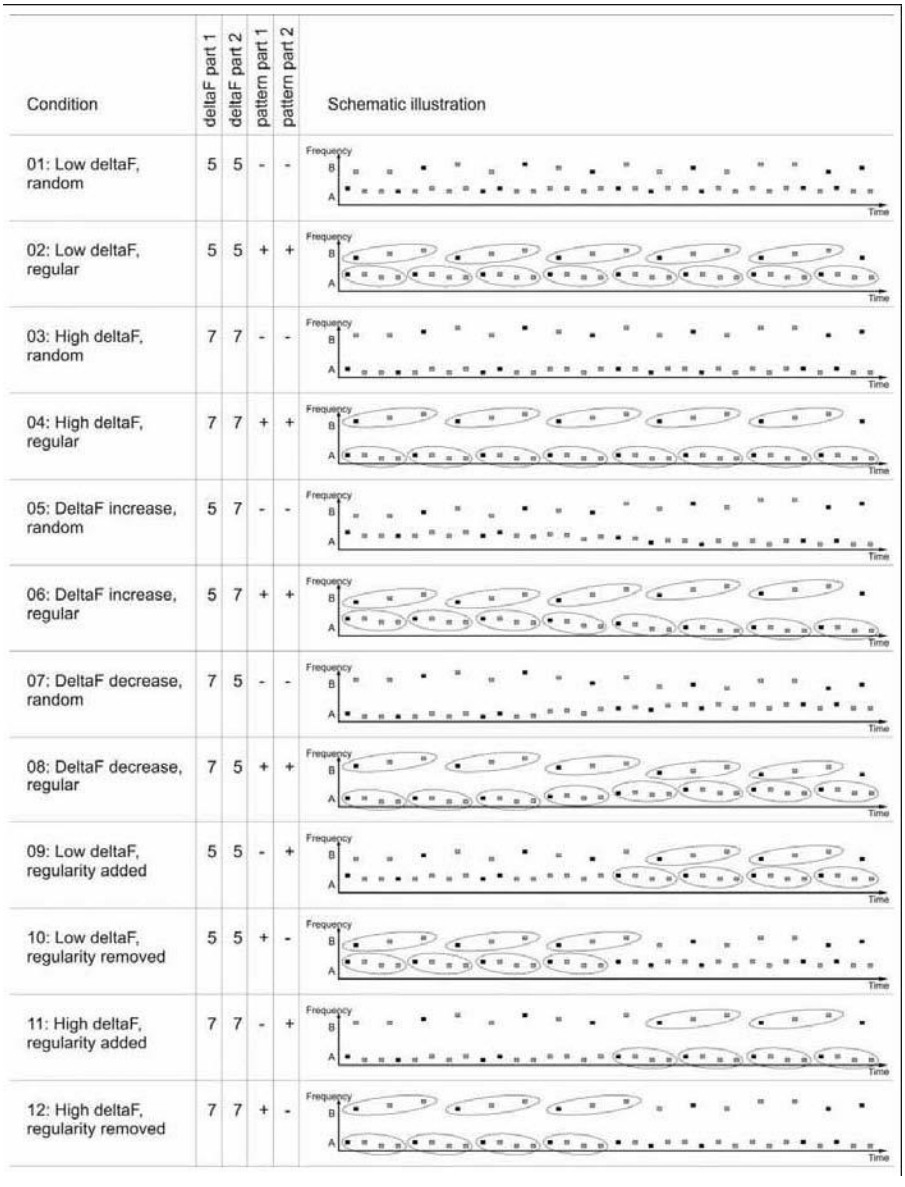

Figure 1. Experimental conditions. Frequency separation (5 semitones – low vs. 7 semitones – high) and the presence vs. absence of regularities (random vs. regular pattern) were manipulated independently for the first and second halves of each block. All manipulations were applied to the 'A' and 'B' tones in parallel. Condition names denote the feature values as well as the presence and direction of changes. A stimulus example for each condition is schematically depicted in the right column. Filled and shaded squares indicate stressed and unstressed tones. Ellipses indicate the cycle of the frequency–intensity regularities. Note that in order to reduce the length of the experimental session, the design was not fully crossed: Conditions with changes in both frequency separation and regularities were not implemented.

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 with equal probability (A1: below the mean 'A' frequency by 10% of the current frequency separation between the 'A' and 'B' sets, A2: above the mean 'A' frequency by 10% of the current frequency separation). The 'B' tones took one of three frequency values with equal probability (B1: below the mean 'B' frequency by 10% of the current frequency separation, B2: identical to the mean 'B' frequency, B3: above the mean 'B' frequency by 10% of the current frequency separation). Both 'A' and 'B' tones were either *stressed* or *unstressed.* Stressed tones had a 6 dB higher 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 (see below).

The order of the different frequency and intensity values was either chosen randomly and independently for the two features *(regularity-absent* conditions) or followed predefined regular patterns, separately for the 'A' and 'B' tones *(regularity-present* conditions) (matching conditions 1 and 10 of Bendixen et al. 2010). Random sequences were separately randomized for each participant. Regular sequences were the same for all participants. The regularly repeating frequency pattern was 'A2A2A1A1' for the 'A' tones and 'B1B2B3' for the 'B' tones. Stress coincided with the first tone of these repeating frequency patterns to facilitate pattern detection (Jones et al. 1982). The different cycle lengths of the regularities for the 'A' and 'B' tones made the regular overall ('ABA') cycle too long for being detected.

In the conditions with a change in frequency separation, the frequency values of the 'A' and 'B' tones converged or diverged gradually, following a linear function over a period of 30 seconds starting at 1.45 min from the beginning of the block and reaching the final frequency separation at 2.15 min. In the conditions with a change in regularity, the regular patterns were either introduced or removed from the sequence at the middle of the block (i.e., at 2.00 min). The rationale for introducing sudden rather than gradual changes in regularity was that patterns would not be extracted instantaneously (Schröger 2007; Winkler 2007; Bendixen et al. 2008). Therefore, the resulting effect would be gradual in any case.

Procedure

Participants were asked to listen to the tone sequences and to continuously indicate their percept by depressing one of three specified buttons on the response pad (lower left, lower right, or upper right). The fourth (upper left) button only served to initiate the block. Participants were instructed to choose between four response alternatives: *Integrated* (depress one button) when they heard the low and high tones within one coherent stream, *Segregated* (depress another button) when they heard a low and a high stream in parallel, *Both* (depress a third button) when they heard a stream consisting of low and high tones and an additional separate stream consisting of only low or only high tones, and *Neither* (release all buttons) when their current percept did not fall into any of these categories. The assignment of *Integrated, Segregated* and *Both* responses to the three 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.

The order of the twelve conditions (blocks) was separately randomized for each participant. A break of at least 30 seconds separated successive stimulus blocks, with additional time given to the participant as needed. Before the first experimental block, 1-min practice blocks with an intermediate frequency separation (6 semitones) containing no regular pattern were presented as long as needed to clarify the instructions and the button-response assignments.

Data recording and analysis

The state of the three 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 the upper limit of the response delay from a change in the participant's percept (Moreno-Bote et al. 2010). After this correction (which on average led to the exclusion of 0.64% 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.

Data were analyzed separately for the first and second halves of each block to study the effects of parameter change. The first half was taken from the beginning of the sound sequence to 1.45 min, the second half from 2.15 min to the end of the sound sequence
(4.00 min). Nine perceptual measures were derived separately for each participant, condition, and part of block (first vs. second half). 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. Proportions and mean durations for *Segregated, Both* and *Neither* percepts were defined in an analogous manner. 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[2]. All duration analyses were carried out with log-transformed duration values to accommodate skewed duration distributions. The log scale was converted back to seconds for reporting mean durations and for display purposes.

To test hypotheses 1 to 6 specified in the Introduction without the possibly confounding influence of parameter changes, only data from the first half of each condition were taken into consideration. Each of the nine dependent measures (see previous paragraph) was analyzed in a repeated-measures analysis of variance (ANOVA) with the factors *frequency separation* (2 levels: low, high) and *regularity* (2 levels: absent, present). Conditions were pooled for this analysis according to the cue combination in the first half of the block (i.e., conditions 1, 5, and 9 for low frequency separation × regularity absent; conditions 2, 6, and 10 for low frequency separation × regularity present; conditions 3, 7, and 11 for high frequency separation × regularity absent; conditions 4, 8, and 12 for high frequency separation × regularity present).

For testing the effects of the parameter changes in the middle of the sequence, data from the second half of each condition were analyzed, and parameter values of the first half were used as a factor in order to determine carry-over effects. Only the proportions of each percept (i.e., four dependent measures) were studied. The latency of the first 'Segregated' percept is not meaningful for the second half of a stimulus block; the computation of average phase durations was not reliable for the second half of the blocks because too many subjects reported only one percept throughout the second half of some of the conditions. The effects of change in

frequency separation were studied in repeated-measures ANOVAs including the factors *frequency separation in 1st half* (2 levels: low, high), *frequency separation in 2nd half* (2 levels: low, high) and *regularity* (2 levels: absent, present). Conditions 1–8 were used for this analysis. The effects of change in regularity were investigated in repeated-measures ANOVAs including the factors *regularity in 1st half* (2 levels: absent, present), *regularity in 2nd half* (2 levels: absent, present) and *frequency separation* (2 levels: low, high). Conditions 1–4 and 9–12 were used for this analysis.

All significant ANOVA effects and interactions are reported with the partial η^2 effect size measure. The Bonferroni correction of confidence level was applied to accommodate for testing multiple dependent variables.

RESULTS

The proportions and average phase durations of the four percepts during the first half of each block are depicted in *Figure 2*. The proportion of 'Integrated' percepts was affected by *frequency separation* [$F(1,29) = 55.296$, $p < 0.001$, $\eta^2 = 0.656$], due to a higher proportion of 'Integrated' percepts with the small (41.0%) than with the large frequency difference (25.4%) between the A and B tones. The proportion of 'Integrated' percepts was also influenced by *regularity* [$F(1,29) = 19.311$, $p < 0.001$, $\eta^2 = 0.400$]. This was due to a higher proportion of 'Integrated' percepts with random (37.4%) than with regular arrangement (29.0%) separately within each set of tones. There was no significant interaction between *frequency separation* and *regularity* [$F(1,29) = 0.010$, $p > 0.99$].

The average phase duration of 'Integrated' percepts was affected by *frequency separation* [$F(1,29) = 28.664$, $p < 0.001$, $\eta^2 = 0.497$], due to longer 'Integrated' phase durations with the small (7.1 s) than with the large frequency difference (4.84 s). The average phase duration of 'Integrated' percepts was not influenced by *regularity* [$F(1,29) = 2.232$, $p > 0.99$], nor was there a significant interaction between *frequency separation* and *regularity* [$F(1,29) = 0.014$, $p > 0.99$].

The proportion of 'Segregated' percepts was affected by *frequency separation* [$F(1,29) = 44.553$, $p < 0.001$, $\eta^2 = 0.606$], due to a lower proportion of 'Segregated' percepts with the small (48.8%) than with the large frequency difference (63.5%). The proportion of 'Segregated' percepts was also influenced by *regularity* [$F(1,29) = 10.195$, $p < 0.05$, $\eta^2 = 0.260$]. This was due to a lower proportion of 'Segregated' percepts with random (52.4%) than with regular arrangement (59.8%). There was no significant interaction between *frequency separation* and *regularity* [$F(1,29) = 0.181$, $p > 0.99$].

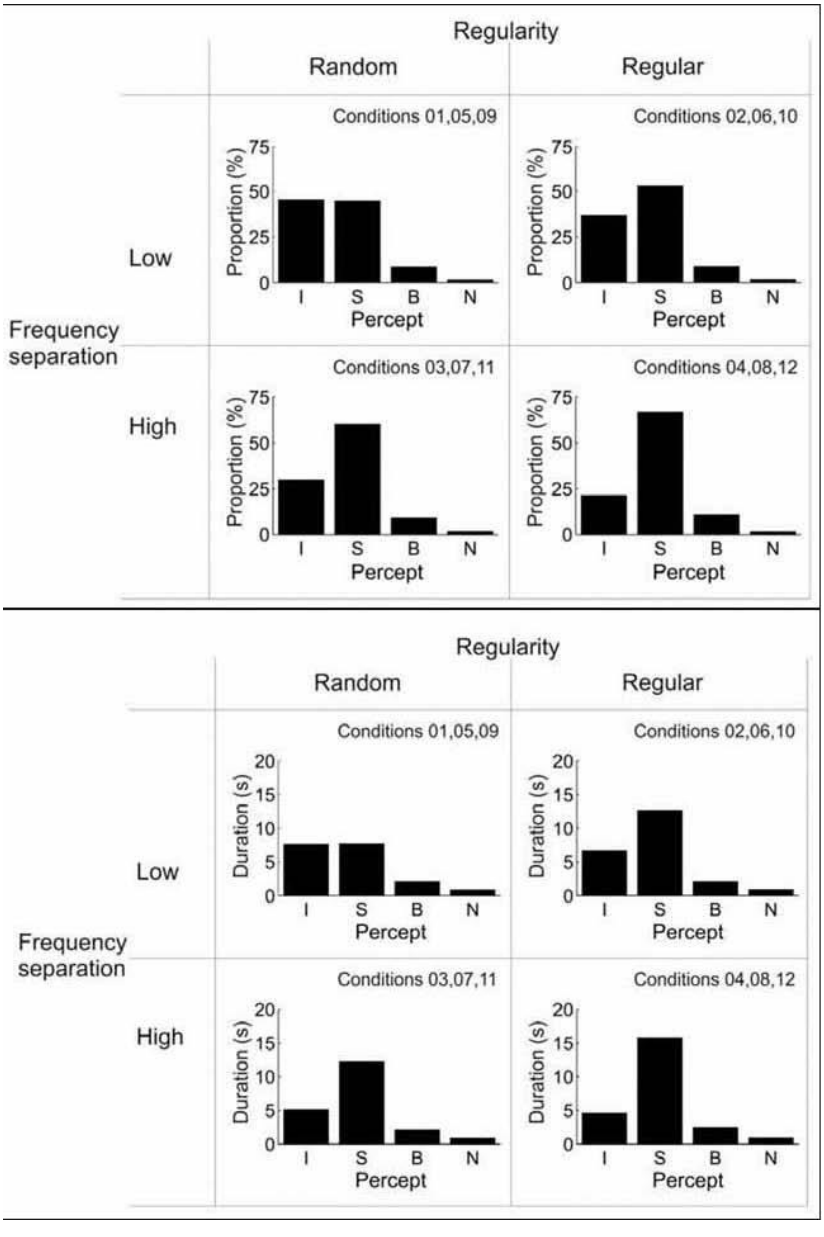

Figure 2. Effects of frequency separation and of the presence or absence of regularities on the proportion of 'Integrated' (I), 'Segregated' (S), 'Both' (B) and 'Neither' (N) percepts (top panel) and on the average duration of the perceptual phases for each percept (bottom panel). Only data from the first half of each block are shown. Experimental conditions with identical first parts were pooled for this comparison (see the text for details). Duration data were log-transformed for the analyses. They are shown here after being converted back to seconds for display purposes.

The average phase duration of 'Segregated' percepts was affected by *frequency separation* [$F(1,29) = 23.209$, $p < 0.001$, $\eta^2 = 0.445$], due to shorter 'Segregated' phase durations with the small (9.82 s) than with the large frequency difference (13.87 s). The average phase duration of 'Segregated' percepts was also influenced by *regularity* [$F(1,29) = 12.285$, $p < 0.05$, $\eta^2 = 0.298$]. This was due to shorter 'Segregated' phase durations with random (9.66 s) than with regular arrangement (14.13 s). There was no significant interaction between *frequency separation* and *regularity* [$F(1,29) = 2.418$, $p > 0.99$].

'Neither' percepts occurred very rarely (1.5% on average), and their proportion and mean duration were unaffected by *frequency separation*, by *regularity*, as well as by their interaction (all F values < 1.66, $p > 0.99$). 'Both' percepts occurred with a frequency of 9% on average.

However, their proportion and mean duration were again unaffected by *frequency separation,* by *regularity,* as well as by their interaction (all *F* values < 2.25, *p* > 0.99).

The latency of the first 'Segregated' percept was influenced by *frequency separation* [*F*(1,29) = 18.375, *p* < 0.01, η^2 = 0.388]. The two-stream organization emerged as the dominant organization earlier with high frequency separation (after 7.11 s on average) than with low frequency separation (11.43 s). The latency of the first 'Segregated' percept was not significantly influenced by *regularity* [*F*(1,29) = 0.437, *p* > 0.99] nor by the interaction of *frequency separation* and *regularity* [*F*(1,29) = 0.620, *p* > 0.99].

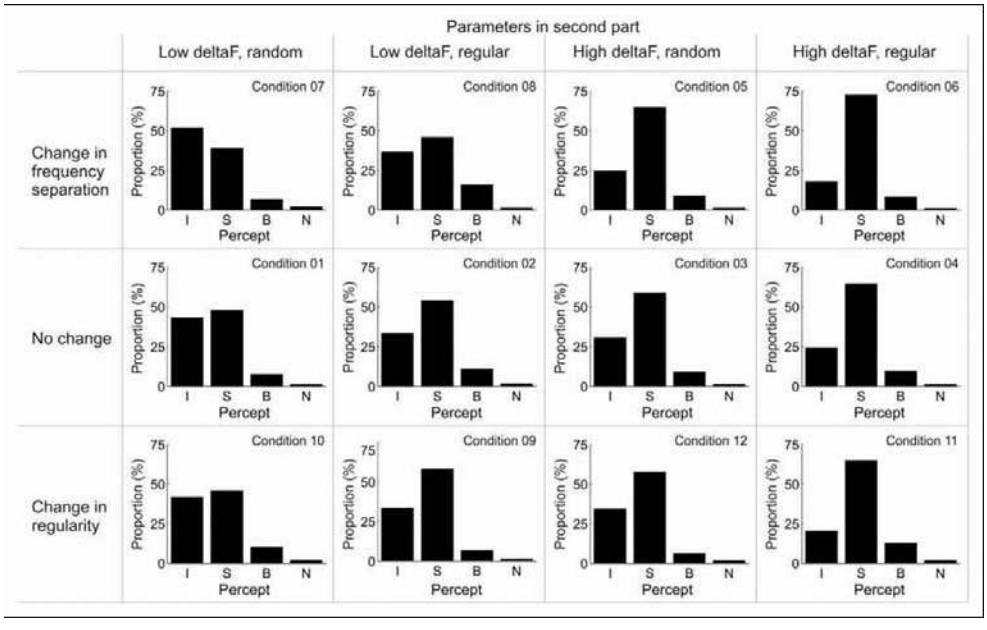

Figure 3. Effects of frequency separation, of the presence or absence of regularities, and of a change in one of these cues on the proportion of 'Integrated' (I), 'Segregated' (S), 'Both' (B) and 'Neither' (N) percepts. Only data from the second half of each block are shown. Note that the conditions in which the regularity has been changed closely resemble the conditions without any change, whereas the conditions in which the frequency separation has been changed yield a more pronounced data pattern suggesting a contrast effect.

The proportions of the four percepts during the second half of each block are depicted in *Figure 3*. In the ANOVA for the frequency-change conditions, the proportion of 'Integrated' percepts in the second half of the blocks was affected by *frequency separation in 2nd half* [*F*(1,29) = 47.836, *p* < 0.001, η^2 = 0.623] and by *regularity* [*F*(1,29) = 29.291, *p* < 0.001, η^2 = 0.502] in the same way as for the first half (see above). Likewise, the proportion of 'Segregated' percepts in the second half was affected by *frequency separation in 2nd half [F*(1,29) = 51.585, *p* < 0.001, η^2 = 0.640] and by *regularity* [*F*(1,29) = 12.594, *p* < 0.01, η^2 = 0.303] in the same way as for the first half. In addition, both measures were influenced by the *frequency separation in 1st half* of the block ['Integrated', *F*(1,29) = 7.275, *p* < 0.05, η^2 = 0.201, 'Segregated', *F*(1,29) = 9.396, *p* < 0.05, η^2 = 0.245]. The underlying pattern was that of a contrast effect: A lower proportion of 'Integrated' percepts was reported in the second half when the first half had included the small frequency difference (29.8%) than when the first half had included the large frequency difference (36.0%), and likewise a higher proportion of 'Segregated' percepts was reported in the second half when the first half had included the small (59.8%) than the large frequency difference (52.2%). No significant two-way or three-way interactions between *frequency*

separation in 1st half, *frequency separation in 2nd half,* and *regularity* were observed (all *F* values < 0.137, $p > 0.99$). Proportions of 'Both' and 'Neither' responses were again unaffected by all experimental factors (all *F* values < 4.54, $p > 0.167$).

In the ANOVA for the regularity-change conditions, the proportion of 'Integrated' percepts in the second half was affected by *regularity in 2nd half* [$F(1,29) = 24.052$, $p < 0.001$, $\eta^2 = 0.453$] and by *frequency separation* [$F(1,29) = 20.624$, $p < 0.001$, $\eta^2 = 0.416$] in the same way as for the first half (see above). Likewise, the proportion of 'Segregated' percepts in the second half was affected by *regularity in 2nd half* [$F(1,29) = 15.873$, $p < 0.01$, $\eta^2 = 0.354$] and by *frequency separation* [$F(1,29) = 18.498$, $p < 0.001$, $\eta^2 = 0.389$] in the same way as for the first half. Both measures were unaffected by *regularity in 1st half* (both *F* values < 0.62, $p > 0.99$). No significant two-way or three-way interactions between *regularity in 1st half*, *regularity in 2nd half*, and *frequency separation* were observed (all *F* values < 1.26, $p > 0.99$). Proportions of 'Both' and 'Neither' responses were unaffected by all experimental factors (all *F* values < 6.642, $p > 0.06$).

DISCUSSION

The present study was designed to compare the effects of feature similarity and feature predictability in auditory scene analysis. The data support the view that similarity and predictability are both used as cues in stream segregation, but that they exert their influence at different stages within auditory scene analysis. Only feature similarity seems to contribute to the putative first stage of auditory scene analysis (Bregman 1990), in which alternative sound organizations are discovered. Both feature similarity and predictability appear to contribute to the second stage of auditory scene analysis, where competition between alternative sound organizations takes place.

As predicted on the basis of previous studies (Denham and Winkler 2006; Denham et al. 2010, in press; Szalárdy et al. in press), low frequency separation (high similarity) increased the proportion of 'Integrated' percepts by prolonging the duration of 'Integrated' percepts and by shortening the duration of 'Segregated' percepts. The prolongation of 'Integrated' percepts is consistent with the notion of stabilizing the currently dominant organization. The shortening of 'Segregated' percepts shows that feature similarity can trigger switching to the 'Integrated' organization when the currently dominant organization is 'Segregated'. This pattern of results suggests that alternative organizations (i.e., those that are incompatible with the currently dominant organization) are continuously explore d and evaluated in terms of feature similarity. One further piece of evidence supporting this interpretation is the faster emergence of the 'Segregated' percept in conditions with high frequency separation (low similarity).

The results regarding the presence or absence of regularities agree with those obtained in our previous study (Bendixen et al. 2010). The presence of stream-specific regularities (high predictability) increased the proportion of 'Segregated' percepts by prolonging their phase duration, but it did not affect the duration of the 'Integrated' percepts. The prolongation of 'Segregated' percepts is consistent with the notion of stabilizing the currently dominant organization. The fact that the duration of 'Integrated' percepts remained unaffected by the presence or absence of regularities suggests that alternative organizations are not evaluated in terms of feature predictability – at least not to the extent of being able to trigger perceptual switches. The fact that the two-stream percept did not emerge earlier in conditions with the

predictability cue, as compared to those without this cue, lends further support for this conclusion. Although we cannot exclude the possibility that the lack of predictability effects during 'Integrated' percepts were due to floor effects (i.e., generally short duration of 'Integrated' percepts), the fact that changing the frequency separation had a significant effect on the duration of 'Integrated' percepts and on the latency of the first 'Segregated' percept suggests that there was enough room above the floor for these effects to be detected.

Our interpretation is consistent with that of various authors arguing that predictability in a currently non-dominant sound organization is not analyzed at all (Bregman 1978; Sussman et al. 1998, 1999; Shinozaki et al. 2000; Winkler et al. 2003a; Sussman et al. 2005; Winkler et al. 2006). One may, however, object that only one form of predictability (defined separately within each of the putative streams) was investigated in the present study and that a different pattern might be obtained when setting up predictable relations between streams (just like similarity is used as a between-stream cue in the present study). This possibility is currently being explored in another experiment.

The finding that predictability within one organization has no effect during the perception of an alternative organization has further implications regarding the processing of predictability cues. As dominance switches back and forth between the alternative organizations, memory regarding predictability within the currently non-dominant organization could be preserved even if it is not actively explored. However, it appears that either no memory of the predictability cue is maintained or it is not used in determining when to switch between alternative organizations. Otherwise, one would expect an effect of changing predictability within the sequence. Yet after the sudden introduction (or removal) of stream-specific regularities, perceptual reports were practically identical to the control conditions in which the regularities had been present (or absent) all along. It appears that the history of stream- specific regularities is not taken into account. There may be good reason for not maintaining or using a memory of predictability. Unless continuously monitored, this information can get outdated during the non-dominant period. Thus, there is no advantage in using such a memory when deciding whether to switch to the currently non-dominant organization.

In contrast, when changing the frequency similarity of the two streams, the similarity experienced before the change continued to have an effect on perceptual reports after the change. This carry-over effect was additive to the effect of the new parameters (i.e., frequency similarity after the change). It exhibited the pattern of a contrast effect: that is, an increase in similarity led to a higher probability of perceiving 'Integration' than when the same level of similarity had been experienced all along; and likewise, a decrease in similarity led to a higher probability of perceiving 'Segregation'. This contrast effect might be surprising in view of priming studies that use the system's tendency to retain the previously dominant percept in spite of parameter changes (Sussman and Steinschneider 2006; Rahne and Sussman 2009). It is, however, consistent with recent observations by Snyder and colleagues (2009a; 2009b), who showed that changes in frequency similarity can lead to either priming or contrast effects depending on the time-scales over which the change and subsequent perceptual evaluations take place (see also Rogers and Bregman 1993).

It is worth noting that none of the parameter changes employed in the present study led to a *reset* of stream segregation; that is, perceptual reports did not re-start with the 'Integrated' percept after any of the changes. This is in contrast to previous observations in which even the introduction of some small and short-lived change was associated with perceptual reset (Rogers and Bregman 1998; Cusack et al. 2004; Roberts et al. 2008; Haywood and Roberts 2010). In

these studies, however, the change was introduced shortly after the beginning of the auditory sequence, during a period in which, probably, not all the alternative organizations had yet been fully formed or experienced. One possible explanation for these contrasting results is that no reset is needed when the auditory system has already developed a full description of all alternatives, including statistics on their reliability in explaining the incoming sequence of sounds. If the currently dominant organization is then challenged by an input change, there may be alternatives that still apply. Keeping them provides continuity for perception. Other alternatives may be weakened by the change and possibly eliminated in favor of new ones. In contrast, when the change occurs at a time when few descriptions are available and possibly the reliability of even those descriptions has not yet been established, the existing description are more vulnerable and may be eliminated quickly. Thus, the analysis practically starts all over again. A detailed discussion of this issue is provided in Winkler et al. (2012).

Similarity and predictability did not significantly interact with each other for any of the manipulations and on any of the variables measured in the current experiment. This result strongly suggests that the two types of cues act independently of each other, possibly because they partly act upon different stages in auditory scene analysis. Feature similarity constitutes a primary cue on which the initial grouping of the auditory scene is based (the first stage of Bregman's 1990 description), whereas, as we argued above, feature predictability does not appear to act as a cue for forming sound groups. Although both factors influence the stability of the perceptual organization with which they are associated (Bendixen et al. 2010; Szalárdy et al. in press), they appear to do so in an additive manner. The present findings are thus fully compatible with Bregman's (1990) notion of two stages in auditory scene analysis, which has also received support from ERP studies (Sussman 2005; Winkler et al. 2005; Snyder et al. 2006).

It is reasonable to assume that the detection of *similar* events is based on at least partly different mechanisms than that of *sequentially predictable* events. Based on the differences in how changes in one and the other affect perception, it is likely that the memory mechanisms supporting them are also different. This view is compatible with previous conclusions about the importance of regularities and predictions in auditory scene analysis (Jones 1976; Jones et al. 1981; Jones et al. 1982; Jones and Boltz 1989; Hung et al. 2001; Denham and Winkler 2006; Winkler et al. 2009). However, here we regard them as principles complementing each other in achieving stable, yet flexible perception of complex auditory scenes.

In conclusion, the present study confirms the important roles of both similarity and predictability in auditory stream segregation. It demonstrates that the two types of cues exhibit different patterns with respect to various aspects of auditory scene analysis. Based on these differences, it is suggested that similarity and predictability involve at least partially separate mechanisms and act upon different stages in auditory scene analysis.

ACKNOWLEDGEMENTS

This work was supported by the European Commission's Seventh Framework Programme (ICT-FP7-231168), by the German Research Foundation [Deutsche Forschungsgemeinschaft, DFG, SCH 375/20-1]; by the German Academic Exchange Service (Deutscher Aka demischer Austauschdienst, DAAD, Project 50345549), by the Hungarian Scholarship Board (MagyarÖsztöndíj Bizottság, MÖB, Project P-MÖB/853), and by the Lendület project awarded to IW by the Hungarian Academy of Sciences (contract number LP2012-36/2012). The experiment was realized using Cogent 2000 developed by the Cogent 2000 team at the FIL and the ICN. The authors thank Susann Duwe and Marie-Luise Schmidt for collecting the data.

NOTES

[1] The original design contained one more control condition that is not reported here for brevity. This 13th condition contained a random variation of frequency and intensity values throughout the block, it started with a low frequency separation, and changed abruptly to a high frequency separation at the middle of the block (i.e., at 2.00 minutes). The condition was always administered last within the experimental session to avoid participants noticing the abrupt change and searching for similar events in subsequent blocks. The condition was included for comparability with previous experiments involving parameter changes within the auditory streaming paradigm, because these changes were typically administered abruptly (but see Rogers and Bregman 1998). The results of condition 13 were found to be essentially identical to those of condition 5 (same parameters but with gradual change). The results are now reported in Winkler et al. (2012).

[2] The *latency of the first Segregated percept* reflects the time it takes for the organization based on two separate streams to become the dominant percept. An 'Integrated' phase may or may not be contained within the *latency of the first Segregated percept.* Because the 'Both' response very rarely appears as the first reported percept in a stimulus block and the 'Neither' response is indistinguishable from the delay to the first reported response, this measure fairly well captures the dynamics of perception at the beginning of the stimulus blocks. Please note that because each subject in each condition only records a single first percept, it is not possible to statistically analyze first percepts separately for the four possible percepts.

REFERENCES

Akeroyd, M. A., Carlyon, R. P., Deeks, J. M. (2005): Can dichotic pitches form two streams? *Journal of the Acoustical Society of America,* 118, 977–981.

Andreou, L.-V., Kashino, M., Chait, M. (2011): The role of temporal regularity in auditory segregation. *Hearing Research,* 280, 228–235.

Anstis, S., Saida, S. (1985): Adaptation to auditory streaming of frequency-modulated tones. *Journal of Experimental Psychology: Human Perception and Performance,* 11, 257–271.

Bendixen, A., Denham, S. L., Gyimesi, K., Winkler, I. (2010): Regular patterns stabilize auditory streams. *Journal of the Acoustical Society of America,* 128, 3658–3666.

Bendixen, A., Prinz, W., Horváth, J., Trujillo-Barreto, N. J., Schröger, E. (2008): Rapid extraction of auditory feature contingencies. *Neuroimage,* 41, 1111–1119.

Bregman, A. S. (1978): Auditory streaming is cumulative. *Journal of Experimental Psychology: Human Perception and Performance,* 4, 380–387.

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

Bregman, A. S. (1993): Auditory scene analysis: Hearing in complex environments. In: McAdams, S., Bigand, E. (eds.): *Thinking in Sound. The Cognitive Psychology of Human Audition.* Clarendon Press, Oxford, pp. 10–36.

Cusack, R., Deeks, J., Aikman, G., Carlyon, R. P. (2004): Effects of location, frequency region, and time course of selective attention on auditory scene analysis. *Journal of Experimental Psychology: Human Perception and Performance,* 30, 643–656.

Denham, S. L., Gyimesi, K., Stefanics, G., Winkler, I. (2010): Stability of perceptual organisation in auditory streaming. In: Lopez-Poveda, E. A., Palmer, A. R., Meddis, R. (eds.): *The neurophysiological bases of auditory perception.* Springer, New York, pp. 477 –488.

Denham, S. L., Gyimesi, K., Stefanics, G., Winkler, I. (2013): Perceptual bistability in auditory streaming: How much do stimulus features matter? *Learning and Perception, 5*(Suppl. 2), 73–100. (this issue)

Denham, S. L., Winkler, I. (2006): The role of predictive models in the formation of auditory streams. *Journal of Physiology, Paris,* 100, 154–170.

French-St.George, M., Bregman, A. S. (1989): Role of predictability of sequence in auditory stream segregation. *Perception & Psychophysics,* 46, 384–386.

Griffiths, T. D., Warren, J. D. (2004): What is an auditory object? *Nature Reviews Neuroscience,* 5, 887–892.

Grimault, N., Bacon, S. P., Micheyl, C. (2002): Auditory stream segregation on the basis of amplitude-modulation rate. *Journal of the Acoustical Society of America,* 111, 1340 –1348.

Haywood, N. R., Roberts, B. (2010): Build-up of the tendency to segregate auditory streams: Resetting effects evoked by a single deviant tone. *Journal of the Acoustical Society of America,* 128, 3019–3031.

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

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

Jones, M. R., Boltz, M. (1989): Dynamic attending and responses to time. *Psychological Review,* 96, 459–491.

Jones, M. R., Boltz, M., Kidd, G. (1982): Controlled attending as a function of melodic and temporal context. *Perception & Psychophysics,* 32, 211–218.

Jones, M. R., Kidd, G., Wetzel, R. (1981): Evidence for rhythmic attention. *Journal of Experimental Psychology: Human Perception and Performance,* 7, 1059–1073.

Kondo, H. M., Kashino, M. (2009): Involvement of the thalamocortical loop in the spontaneous switching of percepts in auditory streaming. *Journal of Neuroscience,* 29, 12695–12701.

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

Moore, B. C. J., Gockel, H. (2002): Factors influencing sequential stream segregation. *Acta Acustica United with Acustica,* 88, 320–333.

Moreno-Bote, R., Shpiro, A., Rinzel, J., Rubin, N. (2010): Alternation rate in perceptual bistability is maximal at and symmetric around equi-dominance. *Journal of Vision,* 10, 1.1–18.

Müller, D., Widmann, A., Schröger, E. (2005): Auditory streaming affects the processing of successive deviant and standard sounds. *Psychophysiology,* 42, 668–676.

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

Rahne, T., Sussman, E. (2009): Neural representations of auditory input accommodate to the context in a dynamically changing acoustic environment. *European Journal of Neuroscience,* 29, 205–211.

Roberts, B., Glasberg, B. R., Moore, B. C. J. (2002): Primitive stream segregation of tone sequences without differences in fundamental frequency or passband. *Journal of the Acoustical Society of America,* 112, 2074–2085.

Roberts, B., Glasberg, B. R., Moore, B. C. J. (2008): Effects of the build-up and resetting of auditory stream segregation on temporal discrimination. *Journal of Experimental Psychology – Human Perception and Performance,* 34, 992–1006.

Rogers, W. L., Bregman, A. S. (1993): An experimental evaluation of three theories of auditory stream segregation. *Perception & Psychophysics,* 53, 179–189.

Rogers, W. L., Bregman, A. S. (1998): Cumulation of the tendency to segregate auditory streams: Resetting by changes in location and loudness. *Perception & Psychophysics,* 60, 1216–1227.

Schröger, E. (2007): Mismatch negativity: A microphone into auditory memory. *Journal of Psycho-physiology,* 21, 138–146.

Shinozaki, N., Yabe, H., Sato, Y., Sutoh, T., Hiruma, T., Nashida, T., Kaneko, S. (2000): Mismatch negativity (MMN) reveals sound grouping in the human brain. *Neuroreport,* 11, 1597–1601.

Snyder, J. S., Alain, C., Picton, T. W. (2006): Effects of attention on neuroelectric correlates of auditory stream segregation. *Journal of Cognitive Neuroscience,* 18, 1–13.

Snyder, J. S., Carter, O. L., Hannon, E. E., Alain, C. (2009a): Adaptation reveals multiple levels of representation in auditory stream segregation. *Journal of Experimental Psychology: Human Perception and Performance,* 35, 1232–1244.

Snyder, J. S., Holder, W. T., Weintraub, D. M., Carter, O. L., Alain, C. (2009b): Effects of prior stimulus and prior perception on neural correlates of auditory stream segregation. *Psychophysiology,* 46, 1208–1215.

Sussman, E. (2005): Integration and segregation in auditory scene analysis. *Journal of the Acoustical Society of America,* 117, 1285–1298.

Sussman, E., 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. *Cognitive, Affective and Behavioral Neuroscience,* 5, 93–110.

Sussman, E., Ritter, W., Vaughan, H. G., Jr. (1998): Attention affects the organization of auditory

input associated with the mismatch negativity system. *Brain Research,* 789, 130–138.
Sussman, E., Ritter, W., Vaughan, H. G., Jr. (1999): An investigation of the auditory streaming effect using event-related brain potentials. *Psychophysiology,* 36, 22–34.
Sussman, E., Steinschneider, M. (2006): Neurophysiological evidence for context-dependent encoding of sensory input in human auditory cortex. *Brain Research,* 1075, 165–174.
Szalárdy, O., Bendixen, A., Tóth, D., Denham, S. L., Winkler, I. (2013): Modulation frequency acts as a primary cue for auditory stream segregation. *Learning and Perception, 5*(Suppl. 2), 149–161. (this issue)
Van Noorden, L. P. A. S. (1975): *Temporal coherence in the perception of tone sequences*. Doctoral dissertation, Technical University Eindhoven.
Vliegen, J., Oxenham, A. J. (1999): Sequential stream segregation in the absence of spectral cues. *Journal of the Acoustical Society of America,* 105, 339–346.
Wertheimer, M. (1923): Untersuchungen zur Lehre von der Gestalt II [Laws of organization in perceptual forms II]. *Psychologische Forschung,* 4, 301–350.
Winkler, I. (2007): Interpreting the mismatch negativity. *Journal of Psychophysiology,* 21, 147–163.
Winkler, I., Denham, S. L., Nelken, I. (2009): Modeling the auditory scene: Predictive regularity representations and perceptual objects. *Trends in Cognitive Sciences,* 13, 532–540.
Winkler, I., Denham, S. L., Mill, R., Bőhm, T. M., Bendixen, A. (2012): Multistability in auditory stream segregation: A predictive coding view. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 367,* 1001–1012.
Winkler, I., Kushnerenko, E., Horváth, J., Čeponienė, R., Fellman, V., Huotilainen, M., Näätänen, R., Sussman, E. (2003a): Newborn infants can organize the auditory world. *Proceedings of the National Academy of Sciences of the United States of America,* 100, 11812–11815.
Winkler, I., Sussman, E., Tervaniemi, M., Horváth, J., Ritter, W., Näätänen, R. (2003b): Preattentive auditory context effects. *Cognitive, Affective, and Behavioral Neuroscience,* 3, 57–77.
Winkler, I., Takegata, R., Sussman, E. (2005): Event-related brain potentials reveal multiple stages in the perceptual organization of sound. *Cognitive Brain Research,* 25, 291 –299.
Winkler, I., Van Zuijen, T. L., Sussman, E., Horváth, J., Näätänen, R. (2006): Object representation in the human auditory system. *European Journal of Neuroscience,* 24, 625–634.