PERCEPTUAL BISTABILITY IN AUDITORY STREAMING: HOW MUCH DO STIMULUS FEATURES MATTER?

SUSAN L. DENHAM[1], KINGA GYIMESI[2,3], GÁBOR STEFANICS[2,4] and ISTVÁN WINKLER[2,5]

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

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

[3] Department of Cognitive Science, Budapest University of Technology and Economics, H-1111 Budapest, Stoczek u. 2, Hungary

[4] Laboratory for Social and Neural Systems Research, University of Zurich, P.O. Box 146, CH-8091 Zurich, Switzerland

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

Abstract

The auditory two-tone streaming paradigm has been used extensively to study the mechanisms that underlie the decomposition of the auditory input into coherent sound sequences. Using longer tone sequences than usual in the literature, we show that listeners hold their first percept of the sound sequence for a relatively long period, after which perception switches between two or more alternative sound organizations, each held on average for a much shorter duration. The first percept also differs from subsequent ones in that stimulus parameters influence its quality and duration to a far greater degree than the subsequent ones. We propose an account of auditory streaming in terms of rivalry between competing temporal associations based on two sets of processes. The formation of associations (discovery of alternative interpretations) mainly affects the first percept by determining which sound group is discovered first and how long it takes for alternative groups to be established. In contrast, subsequent percepts arise from stochastic switching between the alternatives, the dynamics of which are determined by competitive interactions between the set of coexisting interpretations.

INTRODUCTION

In order to make sense of real world environments it is necessary to extract and organize relevant information from the wealth of incoming sensory data. The potential amount of information far exceeds the processing capacity of any living system. Hence, biological organisms are not idle perceivers (Brunswik 1956); rather they seek out information about the world and the objects in it (Neisser 1967). The challenge is to create and maintain appropriate object representations on the fly (Hochberg 1981). Thus the process of perceptual organization is fundamental to effective perception.

An important problem for auditory perception is that sound sources may emit discontinuous sequences of sounds, so some means for forming associations between discrete events is required. The formation of 'sequential' associations has been extensively studied with the help of the auditory streaming paradigm (for recent reviews see, Moore and Gockel 2002; Cusack, Deeks et al. 2004; Micheyl, Carlyon et al. 2007; Snyder and Alain 2007; Bee and Micheyl 2008; Winkler, Denham et al. 2009). In a typical streaming experiment (van Noorden 1975), a tone sequence of the structure ABA-ABA-ABA-... is presented. 'A' and 'B' denote tones differing from each other in some acoustic feature, such as frequency, and '-' stands for a silent interval equal to the time interval between the onsets of successive tones, the stimulus onset asynchrony (SOA); see *Figure 1A*. When all sounds are grouped together into a single coherent sequence (termed *auditory stream),* a galloping rhythm is typically heard. By increasing the frequency separation (Δf) between the A and B tones and/or by shortening the time interval between the tones, perception of the sound sequence changes to that of two homogeneous isochronous streams; a faster paced one consisting of A tones and a slower paced one consisting of B's (van Noorden 1975).

In general, there is a trade-off between Δf and SOA in determining the dominant perceptual organization. In his experiments, van Noorden (1975) identified three separate regions of the Δf–SOA space with different characteristic perceptual organizations (see *Figure 1B*). With very low Δf's, participants always heard the galloping rhythm, showing that they organized all tones into a single sound stream. With slightly larger Δf's, and moderate SOA's, participants were able to hear either two separate sound streams or a single integrated stream and could influence their perception by altering their organizational bias at will; termed the *ambiguous region.* Further increasing Δf and decreasing SOA resulted in participants not being able to hear the integrated pattern, which suggests that perception of two streams became the dominant sound organization.

In classical accounts of auditory streaming (for a review, see Bregman 1990), the process of stream segregation was described in terms of an initial build-up period during which evidence in favour of one or other interpretation of the sound sequence is considered, after which a decision is made in favour of segregation or integration. However, recently the stability of auditory perceptual organization has been more closely examined (Winkler, Takegata et al. 2005; Denham and Winkler 2006; Pressnitzer and Hupe 2006; Kondo and Kashino 2009; Bendixen, Denham et al. 2010; Schadwinkel and Gutschalk 2011) and it has been shown that rather than reaching a stable decision, auditory perception switches back and forth between alternative perceptual organizations; a phenomenon termed *bi-stability* or *multi-stability*.
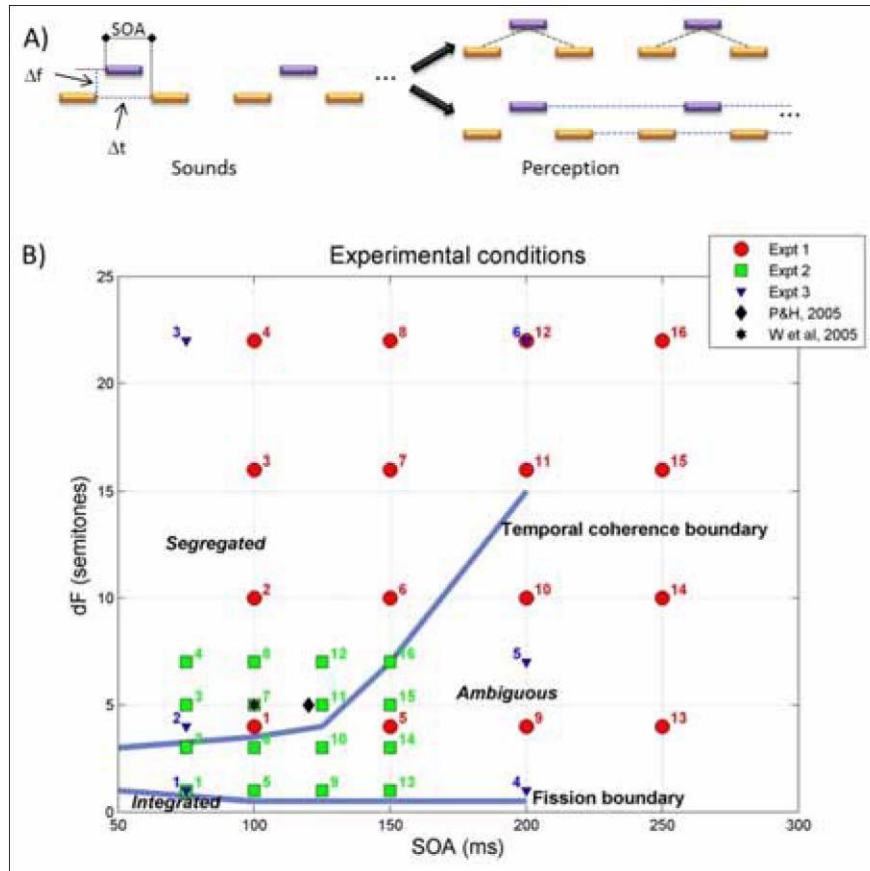
Figure 1. The auditory streaming paradigm. A) A sequence of low (A) and high (B) tones presented repeatedly in ABA- groups can be perceived as a single stream with a galloping rhythm (upper right), or as two segregated streams (lower right), each with an isochronous rhythm. B) Experimental conditions for Experiments 1 (red circles), 2 (green squares) and 3 (blue triangles) reported here, marked on parameter space of stimulus onset asynchrony (*x* axis) and frequency separation (*y* axis). The diagram also shows the perceptual regions found in human perceptual experiments using sound sequences of the same structure (van Noorden 1975; Beauvois and Meddis 1996). Stimulus sequences in the region of the parameter space above the 'temporal coherence boundary' have been found to be generally perceived as two segregated streams, and those with parameters in the region below the 'fission boundary' as a single stream. Those falling in the ambiguous region were perceived in either way, and perception could be influenced by top-down processes (van Noorden 1975). Conditions are numbered separately for Experiments 1 and 2 in the following way. Number 1 denotes the shortest SOA and smallest Δf (100 ms / 4 ST, 75 ms / 1 ST, and 75 ms / 1 ST for Experiments 1, 2, and 3, respectively). Numbers increase faster through the four different Δf's (e.g. 2 marks 100 ms / 10 ST, 75 ms / 3 ST, and 75 ms / 4 ST for Experiments 1, 2, and 3, respectively) and slower for the four different SOAs (e.g. 5 marks 150 ms / 4 ST and 100 ms / 1 ST for Experiments 1 and 2, respectively). Stimulus parameters used in previous experiments investigating perceptual switching are indicated as follows: black diamond (Pressnitzer and Hupe 2006), black star (Winkler, Takegata et al. 2005).

Bistability in visual perception (for a review, see Blake and Logothetis 2002) has been studied extensively since it offers the possibility of identifying the neural correlates of visual awareness, and ultimately consciousness, by allowing perceptual changes to be dissociated from changes in the stimulus (Rees, Kreiman et al. 2002). The finding of perceptual bistability in auditory streaming is important in that it raises questions about similarities between the mechanisms underlying perceptual organization in different sensory modalities. One concept that pervades theories and models of visual bistability is that of competition or rivalry (Blake and Logothetis 2002). Here we investigate whether auditory bistability can be similarly understood in terms of competing interpretations of the auditory scene (Denham and Winkler 2006; Winkler, Denham et al. 2009), i.e. competition between different sequential associations

(see *Figure 2A*) where:

First-order associations (Bregman 1990; Horváth, Czigler et al. 2001), refer to links between temporally adjacent events. These associations are supported by temporal proximity between sequential sound events, thus linking them is relatively easy, at least at small and intermediate Δt's; the links are stronger for small Δf and weaker for large Δf, i.e. similar sounds form better groups than dissimilar ones (Köhler 1947).

Higher order associations (Bregman 1990; Horváth, Czigler et al. 2001), refer to links between non-adjacent events. In our simple streaming sequence these sounds are identical, i.e. A-A… or B---B…; therefore, temporal separation alone governs the strength of this type of grouping, which is stronger for small Δt and weaker for large Δt.
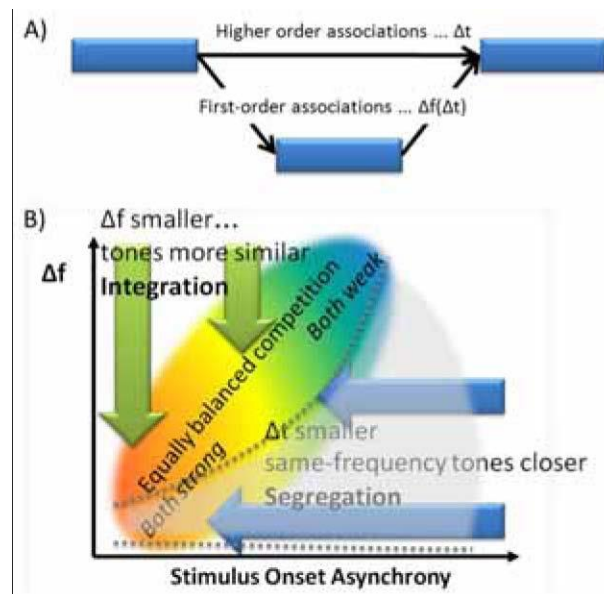

Figure 2. Alternative sequential associations and their possible influence on competitive strength. A) Cartoon indicating the influence of the parameters Δf and Δt on the most prominent sequential associations that can be made in the auditory streaming paradigm. Δt is placed in parenthesis with respect to the first-order association, because with short-medium Δt's, the association–competition account of auditory streaming suggests that the effect of changes in Δt on the formation and representation of the local rule is relatively small. This is because 1) the sounds to be connected are adjacent and 2) with relatively short Δt's separating the sounds, the neural after-effects of the first sound are still strong when the second sound arrives. B) Diagram showing how changes in stimulus parameters would increase (decrease) the competitive strength of alternative interpretations of the sequence according to the association–competition account of auditory stream segregation. The figure shows that balanced region of competition where perceptual switching is predicted to be maximal, with a hypothesized gradient determined by the strength of the competing interpretations from higher (orange) to lower (blue). Superimposed is on the diagram is the 'ambiguous' region identified by van Noorden (1975), shown in grey.

The notion of continuous competition between possible alternative sound organizations is well suited to describing demands on perception in everyday listening environments in which perceptual decisions as to the appropriate way to interpret incoming sound events have to be made as they occur. A strategy that allows the ongoing formation and maintenance of possible alternatives and switching between them mediates an effective trade-off between perceptual flexibility and perceptual stability.

In vision, it has been demonstrated that perceptual switching is greatest in conditions where the competition is evenly balanced (Moreno-Bote, Shpiro et al. 2010). Based on this observation, the amount of switching in auditory stream segregation should be high in the region where competition between integration and segregation is evenly balanced, i.e. along a

roughly diagonal region (see *Figure 2B*). The shape of this region is explained by the effects of Δf and Δt on the strength of the competing associations (see above): decreasing Δt increases the strength of higher-order associations, whereas decreasing Δf increases the strength of first-order associations. In addition, one may assume that competition and, therefore, switching will reach its maximum in the region where Δf and SOA are both small and thus both associations are quite strong, i.e. at the lower left area within the balanced diagonal region *(Figure 2B)*. In contrast, the traditional interpretation of auditory streaming provides no specific predictions regarding the distribution of the amount of switching beyond that it should occur where both alternative perceptions are possible, that is, in the ambiguous region found by van Noorden (1975), which occupies a roughly triangular region in the parameter space, narrowing quite sharply at small Δf and SOA values and broadening out with larger SOA's.

Here we report a detailed study of the influence of Δf and SOA on perceptual bistability in auditory streaming. After specifying the experimental conditions, we present the raw perceptual switching data and then examine the distribution of perceptual switching. Next we test differences in the distribution of first and later percept choices and durations and the time course of auditory perceptual organization. We also describe our finding that contrary to the common assumption that integration and segregation are mutually exclusive, participants sometimes report simultaneously perceiving both organizations. Finally, we discuss the implications of our findings for models and theories of auditory perceptual organization.

METHODS

Participants

The experiments were conducted in the sound-attenuated experimental chamber of the Institute for Psychology, Hungarian Academy of Sciences. They were approved by the Ethical Committee of the Institute for Psychology. After the aims and procedures of the study were explained to them, participants signed an informed consent form before starting the experiment. Participants received modest financial compensation for their participation. Fifteen young healthy volunteers (7 male, 18–25 years of age, average 20.9 years) participated in Experiment 1 and a different fifteen (7 male, 21–25 years of age, average 22.2 years) in Experiment 2. In Experiment 3, a supplementary experiment designed to explore whether perceptual switching would decline with time, a total of 48 participants (19 male, 18–29 years of age, average 22.01 years) took part, but due to the increased length of the stimulus trains, each participated only in a subset of the conditions. Thus, there were 30 participants in conditions 1 and 6 of Experiment 3 *(Figure 1B)*, whereas there were 24 participants in conditions 2–5; the overlap between participants was balanced across the conditions. Half of the participants in Experiment 3 also took part in either Experiments 1 or 2. Participants in each experiment were pre-selected on the basis of the results of clinical audiometry with the criteria that the hearing threshold between 250 and 6000 Hz should not exceed 25 dB, and the difference between the two ears should not exceed 15 dB in the same frequency range.

Stimulus paradigm

Experiment 1 was designed to investigate the distribution of perceptual switching across a relatively large range of the parameter space. Based on the results of Experiment 1 and the theoretical prediction suggesting that the highest amount of switching should occur with parameters strongly supporting both alternative sound organizations, Experiment 2 was then focussed on the region of small to medium Δf (promoting integration) and short to medium SOA (promoting segregation). The experimental conditions used are illustrated in *Figure 1B*.

In Experiments 1 and 2 participants were presented with 4-minute long trains of the ABA-structure, and in Experiment 3 with 10-minute long trains, where A and B were pure tones of 75 ms duration, including 5 ms linear onset and offset ramps. The frequency of the lower-pitched, more frequent tones ("A" in *Figure 1A*) was kept constant at 400 Hz across the different stimulus conditions. In separate trains, Δf was 4, 10, 16, or 22 semitones (ST) in Experiment 1, and 1, 3, 5, or 7 ST in Experiment 2; SOA was 100, 150, 200, or 250 ms and 75, 100, 125, or 150 ms for Experiments 1 and 2, respectively. Altogether, 4 × 4 = 16 different types of trains were tested in Experiments 1 and 2, separately. Experiment 3 was intended to explore how switching continued beyond 4 minutes; parameters were as follows: Δf was 1, 4, and 22 ST at 75 ms SOA (conditions 1–3; *Figure 1B*) and 1, 7, and 22 ST at 200 ms SOA (conditions 4–6). These parameters were chosen to include some of the previous conditions tested and also to probe stability in even more extreme parts of the parameter space (Δf = 1 ST, SOA = 200 ms; Δf = 22 ST, SOA = 75 ms).

Sounds were generated on an IBM PC computer (MEL 2.0 stimulus presentation software – Psychology Software Tools Inc.), amplified using a custom-made sound mixer and amplifier, and delivered through Sennheiser HD 430 headphones at a comfortable 70-dB (SPL) intensity level. The order of the experimental conditions was randomized separately for each participant.

Procedure

In order to eliminate possible confusion caused by the perception of rhythms other than the 'galloping' rhythm, the notion of an integrated percept was generalized and defined for participants as hearing a repeating pattern, which contained both low and high tones. The notion of a segregated percept was similarly generalized and defined for participants as hearing some repeating pattern(s) formed either exclusively of high or exclusively of low tones, with the possibility that multiple repeating segregated patterns (i.e., A---A---A... and B-B-B...) may be perceived concurrently. Participants were asked to depress one response key so long as they experienced an integrated percept and the other key when they experienced a segregated percept. The role of the two keys was randomly assigned across participants. When participants heard no repeating tone pattern, they were instructed to release both keys. Participants were asked to mark their perception throughout the duration of the stimulus train and not to attempt hearing the sound 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.

In the light of results from a pilot study, we also informed participants that it was possible that they may sometimes hear both types of patterns at the same time; i.e., the instructions were "...you may hear a repeating integrated, or some repeating segregated tone patterns,

possibly even both at the same time, or no repeating pattern at all…". In the case that they heard both types of patterns at the same time, they were instructed to keep both buttons depressed. However, they were also cautioned to be sure to release a button when they stopped hearing the corresponding pattern.

Participants sat in a comfortable reclining chair in the experimental chamber, holding a response button in each hand. Short 1–3 minute breaks were inserted between consecutive stimulus trains with longer breaks, when the participant could move about, scheduled just before the start of the experiment (i.e. after the initial explanation of the task) and at the half-time of the session. Further longer breaks were inserted into the session if and when required. Each experiment took about two hours, altogether (instruction time and breaks included). The state of the two response keys was sampled at 10 Hz (100 ms sampling time) by a NeuroScan Synamps EEG recording system. Response key states were then extracted from the signals and encoded separately for each participant and condition for further analysis.

Data analysis

We analysed the data in terms of "perceptual phases". A perceptual phase is the continuous interval within which the participant depressed the same response button combination, marking that he/she perceived the sound sequence throughout this interval as either integrated, segregated, both at the same time, or neither (see the definition of the percepts to be discriminated by the participants above). Most perceptual phases are delimited by two perceptual switches, except for the first perceptual phase, which starts with the first button press after the onset of the stimulus train and the last perceptual phase, which ends with the end of the stimulus train. Theoretical considerations suggest that the first perceptual phase may be qualitatively different from the rest of the perceptual phases (termed "subsequent perceptual phases") (Mamassian and Goutcher 2005; Denham, Gyimesi et al. 2010). Therefore, we separately analysed the first and the subsequent perceptual phases and compared them where appropriate.

When analysing the data, we discarded all those responses that we assumed to represent inaccurate coordination of key presses and releases during the transition between two percepts; i.e. all perceptual phases with durations shorter than 300 ms, our estimate of the upper limit of the response delay with respect to a change in the participant's percept, in accord with (Moreno-Bote, Shpiro et al. 2010).

Group-average probability of perceiving the segregated percept was calculated for the first perceptual phase for illustration purposes only *(Figures 5)*. This is because each stimulus condition was presented only once and therefore, the probability of the different types of first percepts cannot be estimated separately for each participant. Similarly, for the time periods used in the statistical analyses (*Figures 8* and *9*), the number of participants reporting segregation at each time step was divided by the total number of participants to calculate the probability that any participant would report segregation at that time step. For the subsequent perceptual phases the probability of the segregated percept was estimated separately for each participant by dividing the time spent perceiving the segregated sound organization within the period by the duration of the whole period. It should be noted that for better visualisation of the distributions we discuss in *Figure 5* (and also *Figures 4, 6, 7* and *11*) we interpolate the colour scale, between the measured data points using the standard Matlab bicubic spline interpolation method.

We characterized the type (i.e. integrated or segregated) and stability of perceptual phases

by assigning arithmetic signs to phase durations, with the duration of integrated phases treated as positive, and segregated phases as negative. The phase durations of the other two percepts ('both' and 'neither'), which occurred much less frequently, were discarded from this analysis. This allowed us to treat the duration of the first percept as a continuous variable in the [–4, 4] minutes interval and to test the effects of the parameters on the type and stability of the first percept as well as comparing their distribution with those of the subsequent perceptual phases.

Switching rate, the number of perceptual switches per second, is the inverse of phase duration. The time course of the mean switching rate was calculated by taking the mean across participants of the inverse of their current phase duration at each time step. Both the time course of switching rate and probability of segregation were smoothed using a moving average window of 2 seconds.

In order to assess ambiguity, we calculated the ratio between the proportion of perceiving the integrated and the segregated percepts, always dividing the shorter by the longer of the two. This measure lies in the range [0,1] with a maximum of 1 (50–50% integration and segregation) and decreases with diverging balance between the percepts in either direction.

Finally, because Experiment 3 was designed to answer a specific question (i.e., how switching occurs when listening to sequences longer than 4 minutes), the results of this experiment are only reported at the appropriate section. In the rest of the sections, we describe the data obtained in Experiments 1 and 2.

RESULTS

Initial reaction time and perceptual switching

There is an effect of stimulus parameters on the initial reaction time (the delay of the first response from the onset of the stimulus block). In Experiment 1, there was a significant main effect on initial reaction time of Δf and SOA; initial reaction times are shorter for smaller than for larger Δf's ($F[3,42] = 4.44$, $p < 0.02$, $\varepsilon = 0.73$, effect size $\eta^2 = 0.24$), and for longer than for shorter SOA's ($F[3,42] = 7.86$, $p < 0.005$, $\varepsilon = 0.74$, effect size $\eta^2 = 0.36$). There was no interaction between Δf and SOA. In Experiment 2, there was a significant main effect of SOA; initial reaction times are longer for shorter than for longer SOA's ($F[3,42] = 7.14$, $p < 0.01$, $\varepsilon = 0.45$, effect size $\eta^2 = 0.34$). The Δf effect was not significant, and there was no interaction between Δf and SOA. The initial reaction time reflects the conscious discovery of the first perceptual alternative. The SOA effect *(longer* initial reaction times with *shorter* SOA's) rules out the suggestion that this discovery is tied to experiencing a certain number of cycles of the stimulation. Thus these results suggest that we do not start from a default sound organization, which is immediately available. Rather, it suggests that perceptual groups are formed on-line.

We found perceptual bistability over a very wide range of the parameter space typically used in auditory streaming experiments, including conditions for which stable organization was expected, i.e. conditions with very large frequency differences and fast presentation rates, and small frequency differences and slow presentation rates. We found no condition, of all those tested, that was stable across all participants, and no participant who experienced stable perceptual organization for all conditions. The switching results for each participant, condition, and position of the condition within the experimental session are illustrated in *Figure 3*. On average, there were 16.72 switches per condition in Experiment 1, and 36.61 in Experiment 2.

This corresponds, to one switch every 14.35 and 6.06 seconds, respectively; showing that perceptual switching occurs quite often when listening to these tone sequences.

We found no significant effect of the position of the condition within the experimental session for either experiment (F[15,210] = 1.90, $p > 0.1$, Greenhouse–Geisser ε = 0.27 and F[15,210] = 0.81, $p > 0.1$, ε = 0.26, for Experiments 1 and 2, respectively; one-way dependent ANOVA of the number of perceptual switches with the factor position-number [1…16]). This suggests that the observed perceptual switching does not result from learning or fatigue within the experimental session.

The number of switches per train appears to be higher in Experiment 2 than in Experiment 1. This may be an effect of the parameters and/or a difference between the participant groups. Note, the amount of switching varies considerably across participants – see the middle column of Figure 3. The following sections examine perceptual switching in more detail.

Distribution of perceptual switching

The next question we address is to what extent perceptual switching is parameter dependent. In Experiment 1, switching is highest for the smallest two Δf's combined with the shortest two SOA's (see Figure 4, left panel). The observation of the maximum of switching is supported by the results of Experiment 1 using an ANOVA test with the structure SOA (4 levels) × Δf (4 levels) (see *Figure 1B*), which showed a significant effect of SOA (F[3,42] = 4.62, $p < 0.02$, 8 = 0.68, effect size η^2 = 0.79). Although the SOA × Δf interaction did not reach significance, the amount of switching for SOA's of 100 or 150 ms combined with Δf's of 4 or 10 ST was significantly higher than that outside this region (Tukey HSD post-hoc test with $df = 126$, $p < 0.05$ at least for comparisons between within and outside this region).

The observation of the number of switches peaking (although a non-monotonic function of SOA and Δf) in the short-SOA/moderate-Δf region was also confirmed by the statistical analysis of the number of switches observed in Experiment 2. An ANOVA with structure SOA (4 levels, see Figure 1B) × Δf (4 levels) yielded a significant effect of SOA (F[3,42] = 4.72, p < 0.02, 8 = 0.70, η2 = 0.25), Δf (F[3,42] = 10.01, p < 0.001, 8 = 0.68, η2 = 0.42), and an interaction between the two factors (F[9,126] = 2.89, p < 0.02, 8 = 0.57, η2 = 0.17). Tukey HSD post-hoc tests showed that the amount of switching at the SOA of 125 ms combined with Δf's of 3 or 5 ST was higher than that for parameter combinations at the edge of the tested parameter space (df = 126, p < 0.02 or less).

In summary, the region of maximum switching does not coincide with the ambiguous region defined by van Noorden (1975). For example, switching is quite infrequent with medium to long ΔSOA's and medium Δf's, a position well within the classically defined region of ambiguity.
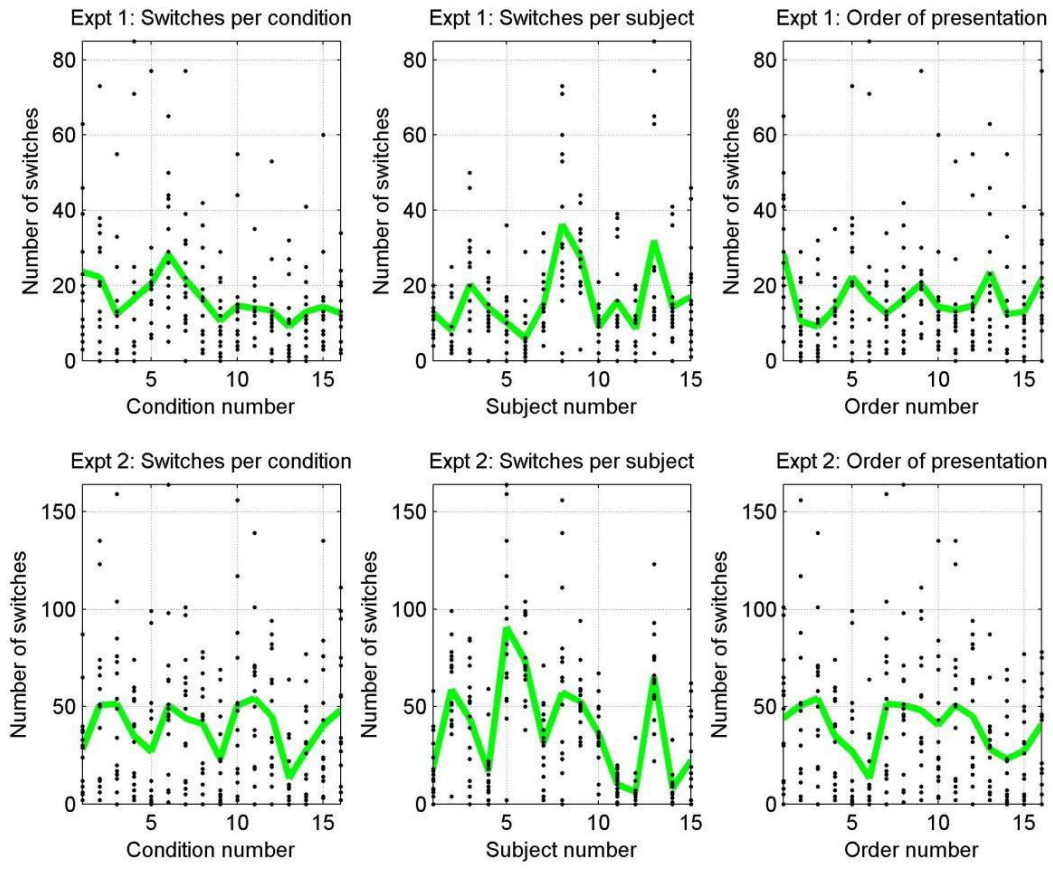
Figure 3. Perceptual switching. Average total number of perceptual switches (green lines) and individual participant data (black dots) plotted against a) condition, b) participant, and c) order of presentation of the condition within the experimental session. Results from Experiment 1 are plotted in the top row; those from Experiment 2 in the bottom row. Condition numbers are as defined in *Figure 1B*. Note that due to the randomized order of the stimulus conditions, stimulus blocks with any set of parameters could occur in any position within the experimental session.
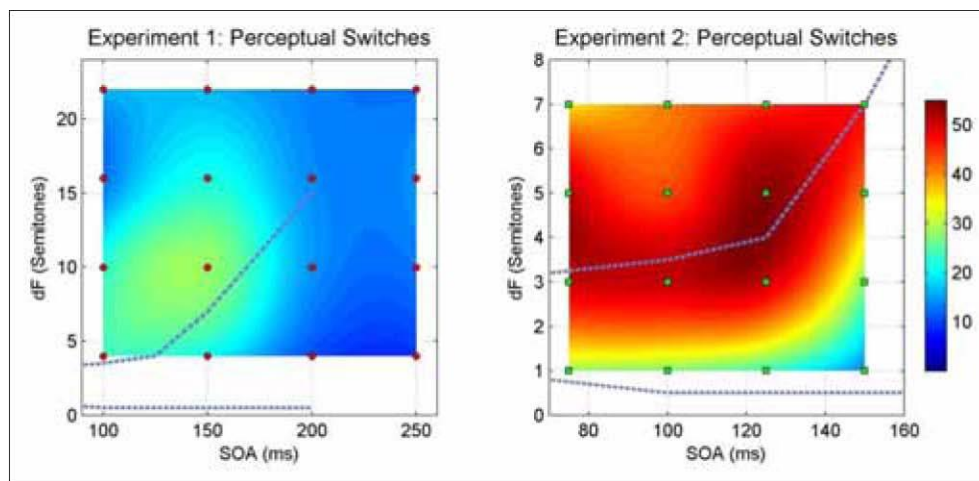

Figure 4. Distribution of perceptual switching. Group-mean distribution of switching in Experiment 1 (left panel) and Experiment 2 (right panel). The colour scale indicates number of switches.

Differences between first and subsequently perceived organizations

Most previous studies of auditory streaming have shown a strong initial bias towards integration. This has led to the suggestion that auditory streaming can be interpreted as a process whereby the auditory system 'accumulates evidence' before making a final organizational decision (Bregman 1990), and segregation only emerges after a gradual *build-up* process Anstis and Saida 1985). We found that for some combinations of stimulus parameters participants are highly likely to report segregation first, as illustrated in Figure 5. In addition, the distribution of the probability of reporting segregation with respect to stimulus parameters changes substantially after the first perceptual phase distribution, and becomes far more evenly spread, and less dependent on stimulus parameters during subsequent perceptual phases.

For testing the effects of the stimulation parameters on the percept type and stability of the first and subsequent perceptual phases, ANOVA's of the signed phase durations (Figure 6) were conducted with the structure Phase (first vs. subsequent) × SOA (4 levels, see Figure 2) × Δf (4 levels) (see Figure 1B). For Experiment 1, this test yielded significant main effects of all three factors (Table 1). Note that the Phase main effect is meaningless in the analysis of signed phase durations, because the value for subsequent phases is the product of averaging perceptual phases of different arithmetic sign, whereas that for the first perceptual phase it is from a single perceptual phase. The interactions between Phase and SOA/Δf describe the *differential effects* of the parameters on the first vs. subsequent phases. Differences in the (absolute) duration of the first and subsequent phases are reported in the next section.
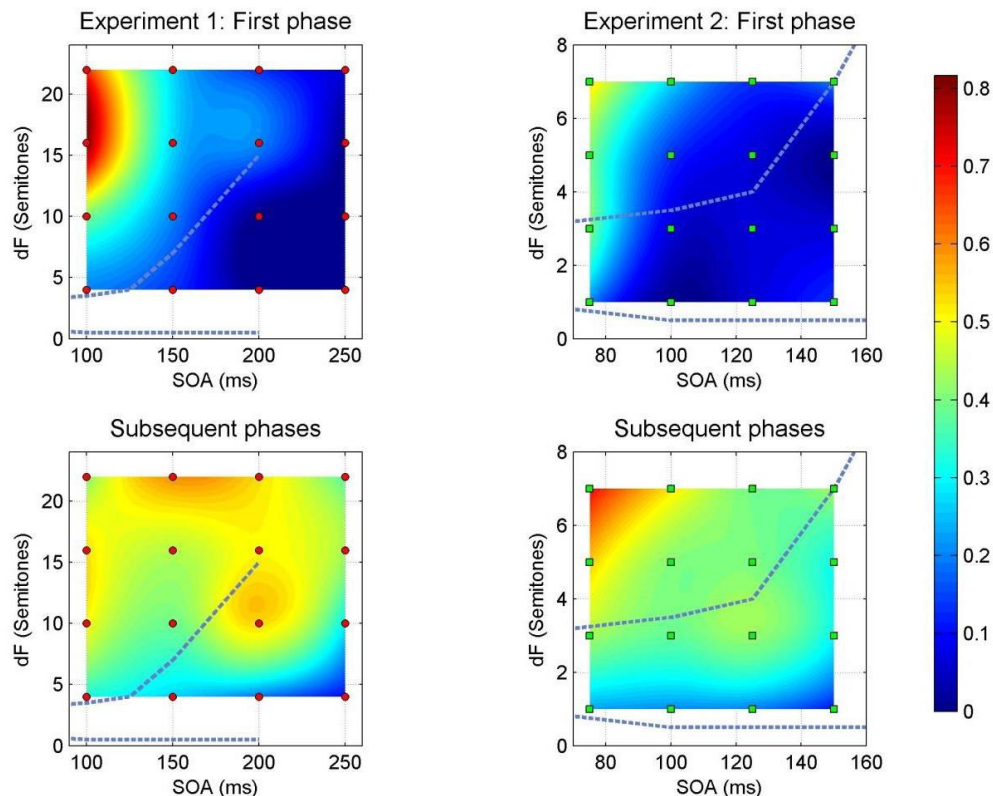

Figure 5. Distribution of the probability of segregation. Group-mean distribution of the probability of segregation in Experiment 1 (left column) and Experiment 2 (right column) are shown for the first (upper row) and the subsequent perceptual phases (lower row). The colour scale, which is the same for all plots, indicates the probability of segregation.
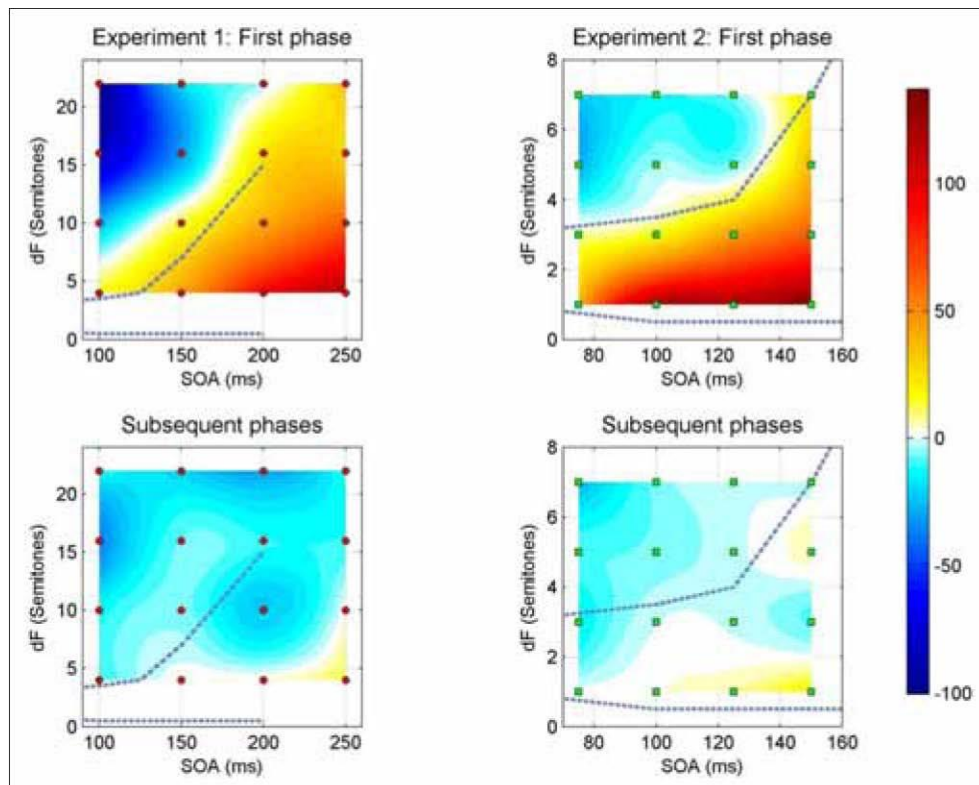
Figure 6. Distribution of signed phase durations. Group-mean distribution of signed phase durations in Experiment 1 (left column) and Experiment 2 (right column) are shown for the first (upper row) and the subsequent perceptual phases (lower row). The colour scale, which is the same for all plots, indicates duration in seconds.

Figure 6 clearly shows a region where first-phase segregation is most common and also that segregation dominates over a wider range of stimulus parameters in subsequent phases.

Both SOA and Δ*f* significantly interacted with Phase *(Table 1)*. The interactions between Phase and SOA and Phase and Δ*f* were caused by the SOA effect (more segregation with shorter SOA's) and the Δ*f* effect (more integration with smaller Δ*f*'s) only characterizing the first perceptual phase. In contrast, neither SOA nor Δ*f* had a significant effect on the balance between integration and segregation in subsequent phases (for the Phase × SOA interaction: Tukey HSD post-hoc test with *df* = 42, $p < 0.05$ at least for comparisons between firstphase/200- and 250-ms SOA and all other cells; for the Phase × Δ*f* interaction: Tukey HSD post-hoc test with *df* = 42, $p < 0.05$ at least for comparisons between first-phase/5- and 10-ST Δ*f* and all other cells). No other interactions yielded significant results.

Table 1. Significant results of the ANOVA [Phase (first vs. subsequent) × SOA (4 levels) × Δf (4 levels)] of the signed perceptual phase durations for Experiments 1 and 2

Factor	df	F	p<	η^2	ε
Experiment 1					
Phase	1,14	15.04	.002	0.52	
SOA	3,42	30.11	.0001	0.73	0.68
Δf	3,42	22.35	.0001	0.61	0.60
Phase × SOA	3,42	9.44	.0001	0.40	0.57
Phase × Δf	3,42	9.09	.0001	0.39	0.90
Experiment 2					
Phase	1,14	22.48	.001	0.62	
SOA	3,42	12.04	.001	0.46	0.69
Δf	3,42	34.01	.0001	0.71	0.50
Phase × SOA	3,42	5.26	.01	0.27	0.47
Phase × Δf	3,42	13.99	.0001	0.50	0.66

The ANOVA of the signed phase durations for Experiment 2 produced a very similar set of results. All three main effects were significant *(Table 1)*. Integration was generally more common in the first perceptual phase, whereas the subsequent phases showed a more balanced picture. Both SOA and Δf significantly interacted with Phase *(Table 1)*. The interactions between Phase and SOA and Phase and Δf were again caused by both SOA and Δf only affecting signed phase durations in the first perceptual phase (for the Phase × SOA interaction: Tukey HSD post-hoc test with $df = 42$, $p < 0.05$ at least for comparisons between first-phase/100–150-ms SOA and all other cells; for the Phase × Δf interaction: Tukey HSD post-hoc test with $df = 42$, $p < 0.05$ at least for comparisons between first-phase/1- and 3-ST Δf and all other cells). No other interactions yielded significant results.

In summary, whereas SOA and Δf had the expected effects on the first phase percept, their influence was weaker during subsequent perceptual phases, which were more balanced. Very importantly, segregation was often reported first in the region of small SOA and large Δf.

Differences between phase durations in the first and subsequent perceptual phases

In addition to differences in the distribution of sound organization between first and subsequent perceptual phases we also found a difference in mean phase durations, as illustrated in *Figure 7*.

Absolute (unsigned) perceptual phase durations were entered into the ANOVA's with the same factors as in the previous section: Phase (first vs subsequent) × SOA (4 levels) × Δf (4 levels). The results of the analysis are displayed in *Table 2* below. For Experiment 1, this analysis yielded main effects of Phase and Δf. The Phase main effect was caused by longer first than subsequent phase durations. The Δf main effect was caused by phase durations being longer for 4-ST than for 10-ST Δf. Significant interactions were found between Phase and Δf and between SOA and Δf. The latter stemmed from the opposite tendency of the SOA effect at low and high Δf's: at low Δf's, phase durations increased with increasing SOA's, whereas at high Δf's, they decreased with increasing SOA's (Tukey HSD post-hoc test with $df = 126$, $p < 0.05$ at least, for comparisons between 10-ST Δf combined with 150-ms SOA and 16- or 22-ST Δf combined with

100-ms SOA; as well as 4-ST Δf differing from 16- and 22-ST Δf at 250-ms SOA). Finally, the significant triple interaction revealed that the above described SOA × Δf interaction only characterized the first perceptual phase, whereas subsequent perceptual phases showed a largely uniform distribution of phase durations. No other main effect or interaction reached significance.
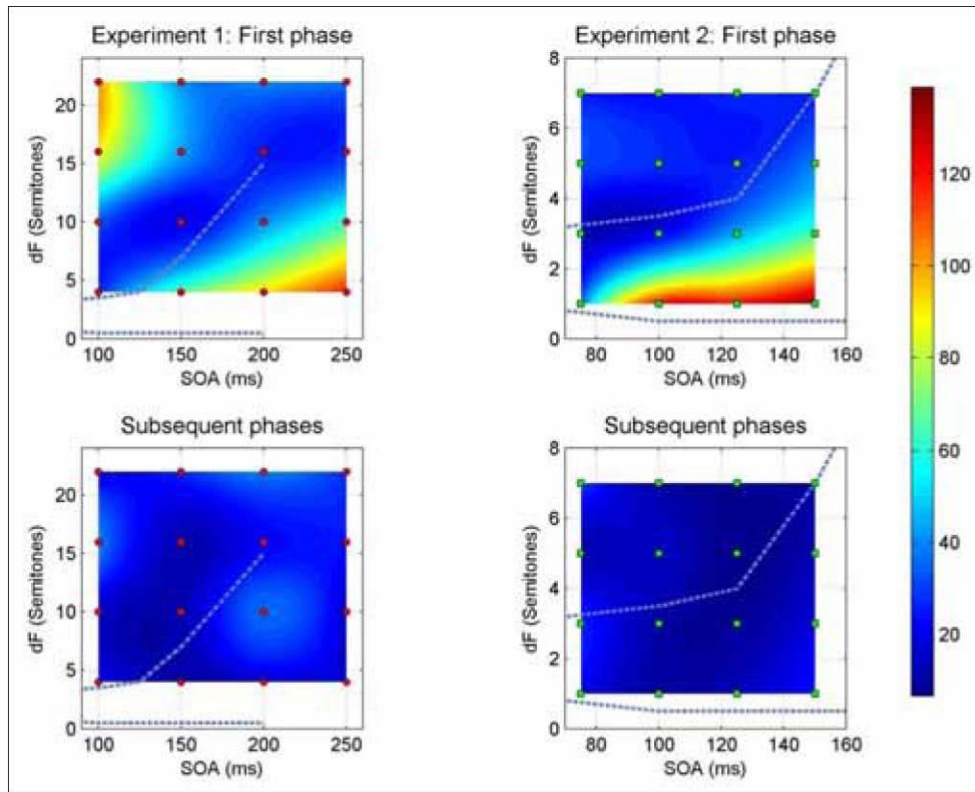

Figure 7. Distribution of phase durations (absolute). Group-mean durations of the first (top) and the subsequent phases (bottom) are shown for Experiment 1 (left) and Experiment 2 (right). The colour scale, which is the same for all plots, indicates duration in seconds.

In Experiment 2, an ANOVA of the same structure as above yielded significant main effects for Phase and Δf. Similarly to Experiment 1, the first perceptual phase was significantly longer than the subsequent phases. The Δf main effect was caused by phase durations monotonically decreasing with increasing Δf. The interaction between the Phase and SOA factors, was the product of increasing phase durations with increasing SOA's in the first phase, only (Tukey HSD with $df = 42$, $p < 0.05$ at least, between any pair of first- and second-phase cells at 100–150-ms SOA). The interaction between the Phase and Δf factors stemmed from the duration of the first perceptual phase at the smallest Δf (1-ST) being significantly longer than all other perceptual phase durations, including all other first phase durations (Tukey HSD with $df = 42$, $p < 0.001$ in all cases). This result revealed that longer first phase durations mainly occur at very low Δf's (qualifying the Phase main effect). Finally, there was a significant interaction between the SOA and Δf factors. No other main effect or interaction reached significance.

Table 2. Significant results of the ANOVA [Phase (first vs. subsequent) × SOA (4 levels) × Δ*f* (4 levels)] of the absolute (unsigned) perceptual phase durations for Experiments 1 and 2

Factor	df	F	p<	η^2	ε
Experiment 1					
Phase	1,14	46.11	.0001	0.77	
Δf	3,42	3.141	.05	0.18	0.83
Phase × Δf	3,42	4.25	.05	0.23	0.86
SOA × Δf	9,126	5.45	.001	0.28	0.59
Phase × SOA × Δf	9,126	3.39	.01	0.20	0.57
Experiment 2					
Phase	1,14	21.34	.001	0.60	
Δf	3,42	18.55	.0001	0.57	0.71
Phase × SOA	3,42	5.58	.01	0.28	0.81
Phase × Δf	3,42	9.88	.001	0.41	0.86
SOA × Δf	9,126	5.07	.01	0.27	0.47

In summary, the first perceptual phase was usually longer than subsequent ones. This result is consistent with previous findings in auditory (Denham, Gyimesi et al. 2010) as well as in visual bistability (Mamassian and Goutcher 2005; Carter and Cavanagh 2007). Furthermore, stimulus parameters primarily affect the durations of the first perceptual phase. Finally, the duration of the first perceptual phase is shortest with short SOA's and low-moderate (but not very low) Δ*f*'s.

Time course of perceptual organization

Classical findings of a build-up of segregation (Anstis and Saida 1985) suggest that segregation becomes more likely later during the stimulus trains. We quantified this observation by comparing the probabilities of segregation across three time ranges, selected from early (20–50 s), middle (120–150 s) and late (200–230 s) periods of the stimulus trains, using ANOVAs of the structure: Time-range (early, middle, late) × SOA (4 levels) × Δ*f* (4 levels).

In Experiment 1 (Figure 8), increasing SOA induced a monotonically decreasing probability of perceiving the segregated organization (F[3,42] = 29.92, p < 0.0001, 8 = 0.89, η^2 = 0.68), whereas increasing Δf induced a monotonically increasing probability for perceiving the segregated organization (F[3,42] = 45.18, p < 0.0001, ε = 0.78, η2 = 0.76). These results are fully compatible with the known effects of SOA and Δf on the perception of these typical two tone streaming sequences. The significant interaction between Time-range and SOA (F[6,84] = 3.37, p < 0.001, ε = 0.74, η2 = 0.19) reflects a gradual decrease in the steepness of the SOA effect (decreasing segregation with increasing SOA's) at later time ranges. No other main effect or interaction reached significance.
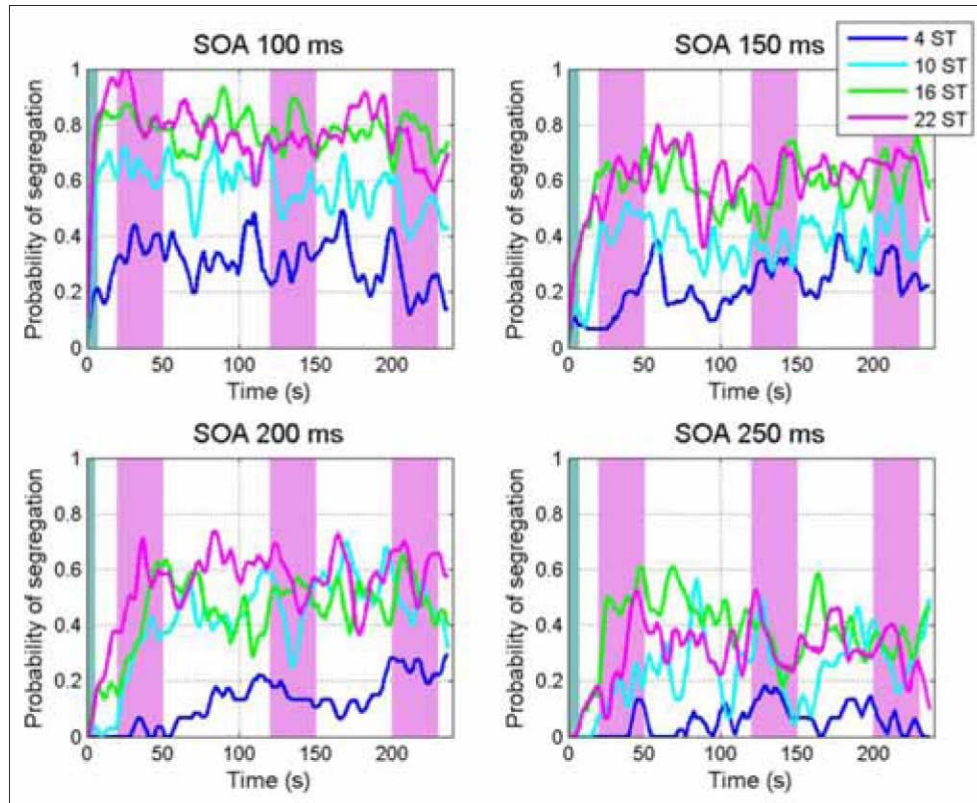
Figure 8. The time course of the probability of segregation in Experiment 1. The group-average time course of the probability of segregation is shown as a function of SOA (panels) and Δf (line colours). Note that the blue-grey bars indicate the initial period during which not all participants had yet made their first choice; the data during this time is, therefore averaged over only those participants who had already indicated some percept by the given time. The violet bars mark the periods used for the statistical analysis.

Similarly, in Experiment 2 *(Figure 9)*, increasing SOA induced a monotonically decreasing probability for perceiving the segregated organization ($F[3,42] = 17.90$, $p < 0.0001$, $\varepsilon = 0.69$, $\eta^2 = 0.56$), whereas increasing Δf induced a monotonically increasing probability for perceiving the segregated organization ($F[3,42] = 74.24$, $p < 0.0001$, $\varepsilon = 0.95$, $\eta^2 = 0.84$). The Time-range × SOA ($F[6,84] = 3.40$, $p < 0.05$, $\varepsilon = 0.59$, $\eta^2 = 0.20$) and Time-range × Δf ($F[6,84] = 3.55$, $p < 0.05$, $\varepsilon = 0.68$, $\eta^2 = 0.20$) interactions reflect a gradual decrease in the steepness of the related main effects at later time ranges. No other main effect or interaction reached significance.

These results suggest that the well-known effects of SOA and Δf on the perceptual organization of the stimulus trains, while prominent at the beginning of the stimulus trains, diminish with time. This observation shows a further difference between the first and subsequent perceptual phases.
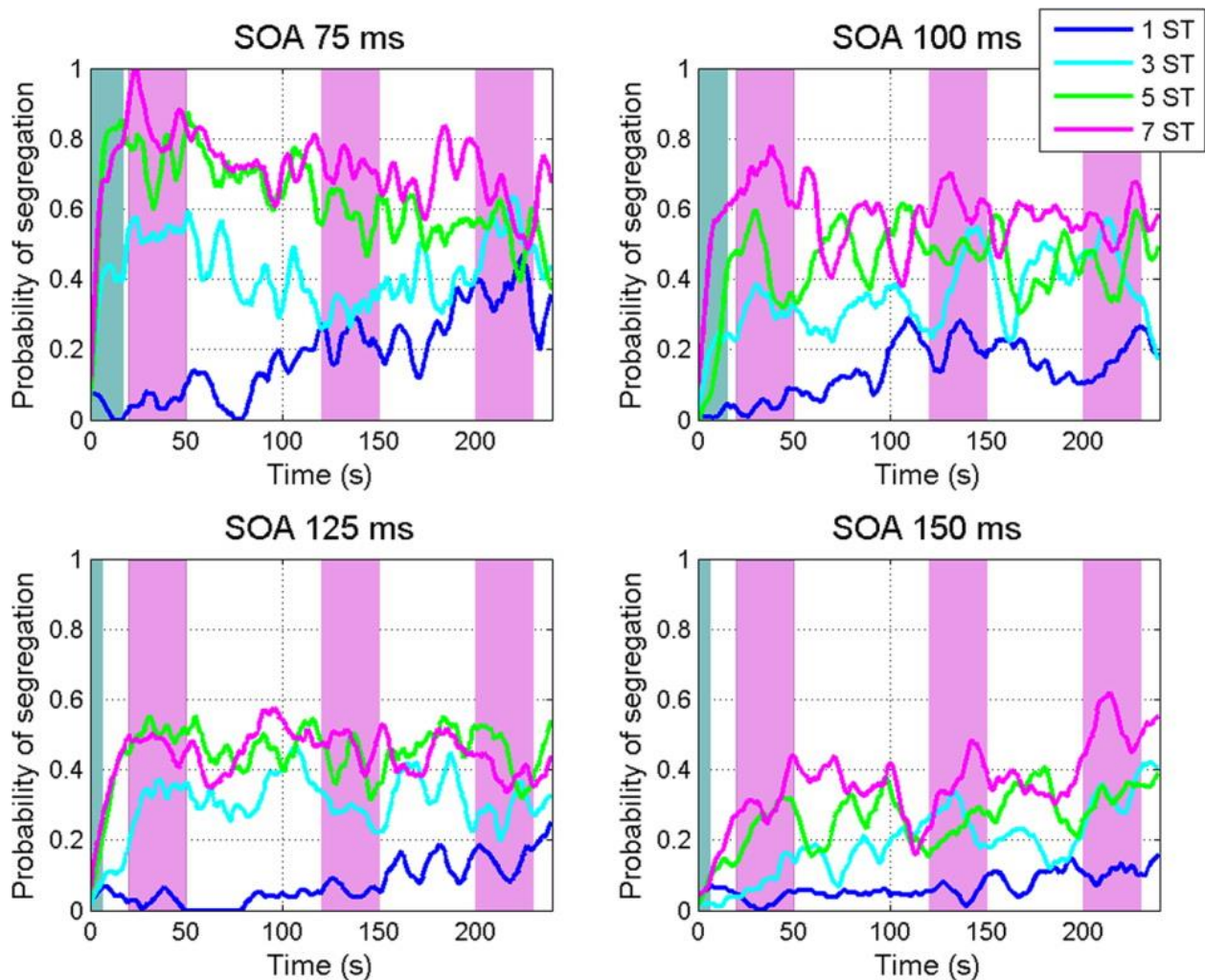
Figure 9. The time course of the probability of segregation in Experiment 2. The group-average time course of the probability of segregation is shown as a function of SOA (panels) and Δf (line colours). See description in the caption of Figure 8.

Does the rate of perceptual switching decline with time?

In order to explore the time course of auditory perceptual switching further we conducted a third experiment in which each stimulus train lasted for 10 minutes. We found no sign that perceptual organization eventually stabilises. This can best be seen by considering the time course of the rate of perceptual switching (see definition in the "Data analysis" section). In *Figure 10* (top panel), switching rate averaged across all participants is plotted as a function of time, and although there are both fast and slow fluctuations, clearly the rate never drops towards zero. We notice here the two rather stable conditions (1-ST 200-ms; 22-ST 75-ms) have a lower switching rate than the other conditions, especially at the start, but even for these conditions switching rate gradually increases through the duration of the stimulus trains.
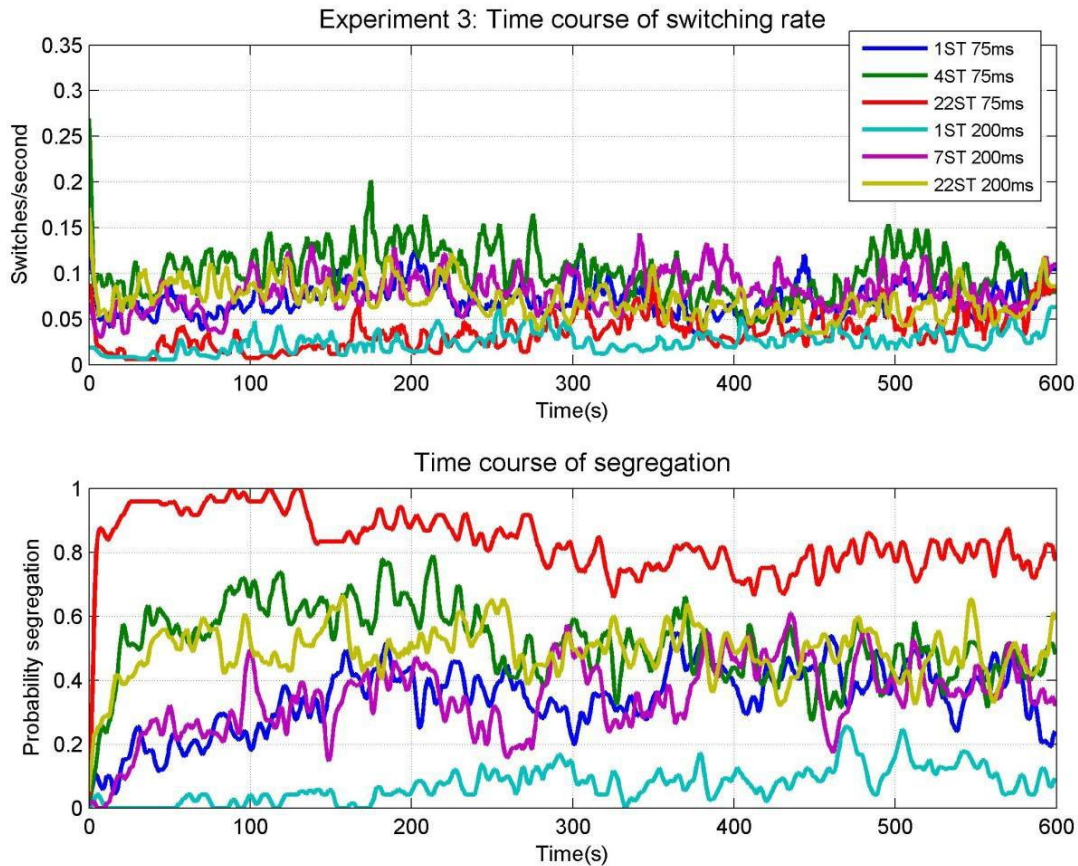
Figure 10. Time course of the rate of perceptual switching and the probability of segregation in Experiment 3. Top panel: Switching rate averaged across all participants as a function of time within the stimulus train. Bottom panel: Probability of segregation for each condition (SOA/Δf combinations marked by colour code) as a function of time within the stimulus train. The legend indicates the colour of the line corresponding to each set of stimulus parameters, which is the same for both panels.

Two observations can be made when considering the time course of the probability of segregation in this experiment (Figure 10, bottom panel). Firstly, as for Experiments 1 and 2, stimulus parameters have much larger effects on perceptual measures near the beginning of the stimulus trains than later in the stimulus blocks. Secondly, towards the middle of the stimulus trains, at around 5 minutes, the probability of segregation approaches ca. 0.5 for all except the most extreme conditions. Thereafter, these intermediate conditions diverge little despite many fluctuations.

Transition Phases: Distribution of both *responses*

The distribution of cases in which participants reported perceiving integrated and segregated percepts simultaneously *(both* responses; *Figure 11,* top panels) is not uniform and, with the exception of one case, *both* never occurred as the first percept. In Experiment 1, an ANOVA of the probability of *both* responses during subsequent phases with the structure of SOA (4 levels) × Δf (4 levels) yielded only a significant interaction between the two factors ($F[9,126] = 3.06$, $p < 0.05$, $\varepsilon = 0.59$, $\eta^2 = 0.18$), which was caused by opposite effects of Δf at different SOAs. At the shortest SOA, increasing Δf decreased the proportion of *both* responses, whereas at the longest SOA, increasing Δf increased the proportion of *both* responses (Tukey HSD post-hoc test with *df*

= 126, *p* < 0.05 between the proportion of *both* responses at shortest-SOA (100-ms) and largest Δ*f* (22ST) and that with the two shortest SOA's (100 and 150 ms) and the smallest Δ*f* (4ST)).
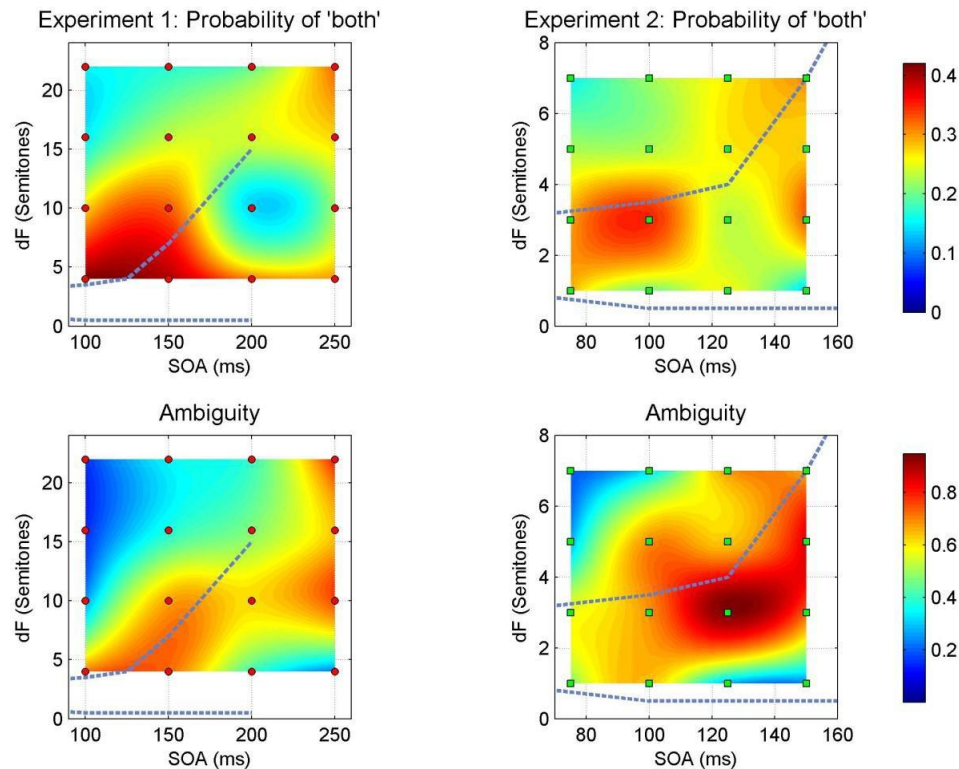

Figure 11. Distribution of probability of simultaneously experiencing both integration and segregation and ambiguity in Experiments 1 and 2. Upper panels: Group-mean probability of both responses during the subsequent phases in Experiment 1 (left) and Experiment 2 (right). Lower panels: Group-mean distribution of ambiguity during subsequent phases in Experiment 1 (left) and Experiment 2 (right). Colour scale indicates probability/ambiguity; separately for each row.

Figure 11 (bottom panels) present the distribution of ambiguity (see the definition in the "Data analysis" section). The balanced region of high ambiguity falls also along a diagonal in the parameter space. However, while there is some similarity between them, the two distributions are not the exactly the same.

In summary, segregated and integrated sound organizations can be perceived simultaneously and there is a tendency for participants to report perceiving both organizations simultaneously most often along a diagonal ridge in the parameter space, apparently in correspondence with our measure of ambiguity.

GENERAL DISCUSSION

Data from the experiments reported here allow some important observations:

1. Switching between integrated and segregated percepts continues throughout the stimulus sequences with any combination of Δf and SOA and for all participants Switching is not a product of learning or fatigue. Rather, there appears to be no "final" stable percept in the auditory streaming paradigm.

2. The first perceptual phase differs from subsequent phases; first phase durations are on average longer than the durations of the subsequent phases, and the influence of stimulus parameters, Δf and SOA, on phase duration is stronger during the first phase.

3. Integration is not always the first percept; for large Δf and short SOA the first reported percept is often segregation.

4. The mean number of perceptual switches is highest in the region of moderate Δf and short SOA, and the first perceptual phase in this region is typically shorter than for other combinations of the two parameters.

5. With time, the proportion of integration and segregation tends to become more balanced irrespective of the combination of Δf and SOA, although in very extreme regions of the parameter space, the tendency to strongly favour one or other organisation remains.

6. Participants sometimes report simultaneous perception of integrated and segregated tone patterns *(both* responses); i.e. segregated and integrated percepts are not mutually exclusive.

7. Some of these observations pose difficulties for traditional descriptions of auditory streaming.

Switching throughout

Perceptual switching is pervasive, even for combinations of stimulus parameters that strongly promote one or other percept and are thus expected to be rather stable. The distribution of perceptual switching in relation to the stimulus parameters is not well described by the *ambiguous* region established in the classical streaming experiments (van Noorden 1975), rather it appears to be more consistent with an account in terms of the relative strength of competing perceptual organizations. Furthermore, perceptual switching continues throughout the duration of the stimulus trains. Even in Experiment 3, which used 10-minute long trains, there was no sign of the rate of perceptual switching lessening *(Figure 10)*. The current results do not support the notion that after gathering evidence during the initial build-up period, the system settles on a stable sound organization. Thus the notion of a final perceptual decision must be abandoned.

One possible explanation is that switching results from attentional fluctuations. Cusack et al. (2004) found that short breaks in the sound sequences or briefly switching attention away and then back to the sounds resulted in a reset of stream segregation, marked by a re-emergence of the integrated percept. This suggests that the switching observed in the current and many other studies could have resulted from listeners' attention switching away and then back to the sounds. However, in a previous study (Denham, Gyimesi et al. 2010) we found no sign of perceptual reset with short breaks included within the sound sequences; for corroborating evidence, see (Bendixen, Bőhm et al. 2011) in this issue. The difference between our and Cusack et al.'s paradigm was that whereas in our study, the short breaks occurred during the

period of subsequent phases, Cusack et al. (2004) introduced them with much less delay from the onset of the stimulus trains, presumably within the first perceptual phase. Thus we explain the difference in the results as yet another difference between the processes occurring during the first and subsequent perceptual phases. Since switching occurs during the subsequent perceptual phases, attentional switches here probably do not reset stream segregation. Furthermore, if perceptual switches were driven by attentional fluctuations, one would expect to see a fatigue effect on perceptual switches (i.e., more attentional fluctuation and thus more perceptual switches during stimulus blocks occurring later in the experimental session) and no effect of the stimulus parameters on the distribution of switches. However, we found no effect of the position of the stimulus block within the experimental session (see *Figure 3*), whereas stimulus parameters had an effect on the number of switches (see *Figure 4;* cf. the next section).

Phase durations and the distribution of switching

First-phase durations were found to be shortest with short SOA's and moderate Δf's. The more uniformly distributed phase durations in subsequent perceptual phases produce more overall switching with shorter first phase durations, because a short first perceptual phase leaves more time for switching in subsequent phases within the uniformly long (4-minute) stimulus blocks. This partly explains how characteristics of the first perceptual phase can affect the distribution of switches through their effect on the duration of the first perceptual phase. Thus the highest switching region occurs where both first- and higher-order associations are strong. This allows both sound organizations to be discovered quickly and thus switching between them starts early. However, comparing the distribution of the duration of the first perceptual phase (*Figure 5,* top panels) with that of the number of switches *(Figure 4)* reveals that stimulus parameters also have a direct effect on the amount of switching. For example, in Experiment 1, short first phases have been observed for long SOA's (200 and 250 ms) combined with small to intermediate Δf's (4 and 10 ST; see *Figure 5,* top left panel). However, the number of switches is rather low in this region of the parameter space (*Figure 4*, right panel). Thus the strength of the different associations also affects switching during the subsequent phases.

Although an increase in switching with increasingly strong sequential associations is intuitively easy to understand, the non-monotonic relationship between association strength and switching (see *Figure 4*) requires further consideration. Throughout these experiments we used uniform tone durations of 75 ms; thus for SOA's less than 100 ms, there was no (or almost no) silent gap between successive A and B tones. Therefore, it is possible that for small Δf's, triplets of three successive tones (ABA) may have formed unitary events. In this case, for segregation to occur, the system would first have to extract the components from the composite before other sequential associations could be established. This notion is supported by the literature on temporal integration, showing that auditory input within 150–200 ms is integrated into a single unit and processed in many ways differently from successive sounds exceeding this period (Cowan 1984; Czigler and Winkler 1996; Yabe, Tervaniemi et al. 1997). However, temporal integration can only win over segregation with small Δf's (Yabe, Winkler et al. 2001; Shinozaki, Yabe et al. 2003; Sussman 2005). Therefore, we suggest that when all three tones fall within the temporal integration window and Δf is not too large, establishing higher-order associations is more difficult and, as a consequence, the amount of switching decreases compared with the more balanced cases.

First phase segregation and the time course of perceptual organization

Our results contradict the notion that segregation can only occur after sufficient evidence supporting the segregated organization is gathered by the auditory system (Bregman, Ahad et al. 2000).

One possible explanation for the immediate dominance of segregation at small SOA and large Δf is that large Δf's give rise to a relatively large spatial distance between stimulus-driven neural activity in the tonotopically organized part of the afferent auditory system, thus weakening and delaying interactions between the neural activity evoked by the two different tones delivered in the auditory streaming paradigm. At the same time, short SOA's allow the after-effect of the more frequent tones to stay relatively strong at the time of the arrival of the next identical tone. This may enable direct linking of successive identical tones. In accordance with this account, dominance of segregation over integration at large Δf's has been demonstrated within the temporal integration period of ca. 200 ms (Yabe, Winkler et al. 2001; Shinozaki, Yabe et al. 2003; Sussman 2005).

Many previous studies presented short (typically <30 s) trains and asked participants about their percept at the end of the trains. Looking at the cross-section of *Figures 8* and *9* at ca. 15 s, and also the early part of the bottom panel of *Figure 10,* we find that our results closely match those of e.g., van Noorden (1975). Indeed, the initial 30 seconds of the curves shown in these figures indeed give the impression of a fast but gradual build-up of streaming. However, whereas this "build-up" has been usually interpreted as "more and more participants reaching the final percept", our data shows that the group-average probability of perceiving the sounds in terms of one or another percept is a product of averaging between perceptual states, which continue to switch back and forth all the time.

Our results further show that there are some changes in the probabilities of the perceptual organizations that become evident over longer time scales. In *Figures 8, 9* and *10* three distinct trends can be discerned. With parameter combinations that strongly promote segregation (short SOA, large Δf), following a fast overshoot of the probability of segregation, the ratio between segregation and integration declines slowly (e.g. *Figure 10,* SOA = 75 ms, Δf = 22-ST). With parameter combinations regarded to promote integration, the probability of segregation appears to increase slowly throughout the whole duration of the stimulus (e.g. *Figure 8,* SOA = 200 ms, Δf = 4-ST). Finally, for most less extreme parameter combinations the initial increase of segregation is followed by a period, in which the probability of segregation converges into a narrow region between 0.4 and 0.6, as it most clearly discernible over the longer durations used in Experiment 3 *(Figure 10)*.

It thus appears that distinguishing between the first and subsequent perceptual phases, as is often the case in the analysis of visual bistability (Carter and Cavanagh 2007), may prove a more fruitful description of the temporal behaviour of auditory perceptual organization than the assumption of a build-up of segregation. Based on the notion of competing sequential associations, this difference can be explained by assuming that the organization discovered first has a "grace period", during which no competition occurs (as the alternatives have not yet been discovered). This initial phase may also be prolonged by the currently dominant (perceived) sound organization being strengthened by correctly describing the stimulus configuration (Pastukhov and Braun 2008). However, after a while, the neural associations underlying alternative sound organizations become stronger thus enabling them to vie for dominance. Since at this stage the competition is presumably between patterns of neural activity very far

removed from the physical stimulus parameters and much more dependent on the intrinsic brain circuitry and neural mechanisms involved, the discovered organisations tend to compete on more equal terms. Thus the probabilities of perceiving different organizations become more balanced with time.

Mutual exclusivity?

The perception of both integrated and segregated patterns at the same time contradicts our intuitive assumption regarding the exclusivity of two competing perceptual organizations, as well as the findings of Pressnitzer and Hupé (2006). However, it has been shown in vision that, contrary to the usual assumptions of exclusivity (Leopold and Logothetis 1999), periods of transition during which neither eye is clearly dominant can be of rather long duration; comparable with eye dominance durations (Brascamp, van Ee et al. 2006; Lee, Blake et al. 2007). This is consistent with the findings in our experiments, where the mean proportion of *both* responses overall is ca. 25% (see *Figure 11*).

It is not clear at this stage what gives rise to the perception of both types of pattern simultaneously. One possibility is that there is a very rapid switching between the two alternatives of integration and segregation but that conscious perception is more sluggish, and unable to follow this rapid switching. Hence there is a sort of stroboscopic effect in which both perceptual organizations are perceived as being present although there is actually switching between them. Alternatively it is possible that "sluggishness" arises because of attempts by the system to update the currently dominant percept by exploring other interpretations in response to on-going prediction errors (Hohwy, Roepstorff et al. 2008); although precisely what such errors might be when listening to an unchanging stimulus sequence is unclear.

We suggest another possible explanation for the emergence of the *both* response. Most studies of the auditory streaming paradigm consider three possible sequential groups: ABA-ABA-, A-A-, and B---B---. However, results of a pilot study in which participants were asked to verbally describe their perceptions highlighted other possible groups, such as AB--AB--, A----A----, BA--BA--, etc. Although it appears that these groups emerge only seldom in perception, some of them, such as the AB-- group would satisfy our description of an integrated pattern, while the A--- on its own would satisfy the description of a segregated pattern. Hence, experiencing these groupings together would be correctly reported as a *both* response. Furthermore, although such groupings break the mutual exclusivity of integrated versus segregated percepts, they do not break the exclusivity of allocating individual sounds to groups (i.e., each sound is part of exactly one group). However, we cannot exclude the possibility that some of the *both* responses reflect groupings that involve duplex perception (Liberman 1982; Fowler and Rosenblum 1990; Ramnani 2006), that is, the same sound being part of two separate groups. The latter would require further consideration for explanations of auditory streaming (cf. Bregman, 1990). Future experiments should attempt to separate these different types of *both* responses, and probe further the basis upon which participants make the *both* perceptual decision.

Theoretical insights

In summary, interesting theoretical insights into the processes underlying streaming can be

obtained from the experiments reported here. At the onset of a new stimulus train, i.e. at the beginning of the first perceptual phase, the system is essentially concerned with the *formation* of perceptual organizations. The organization perceived first is determined largely by the stimulus parameters, which allow some organizations to be discovered faster than others. Consistent with findings in vision that local competition is a necessary prerequisite of perceptual bistability (Anourova, Rama et al. 1999), results of the current experiments suggest that the discovery of feature-sensitive associations is a necessary step for triggering changes in global perceptual organization. Consideration of the typical mean first phase durations suggests that previous streaming experiments, which used relatively short stimulus sequences, have largely characterised the initial phase of perceptual organization. In this respect the theoretical proposals of Micheyl and colleagues (2007), and the recent model of Elhilali and Shamma (2008; Elhilali, Ma et al. 2009) are also primarily concerned with the choice of first phase percept, thus they do not address questions relating to the stability of perceptual choice or the basis for perceptual switching. We regard the model proposed by Elhilali and colleagues (2009) as a possible solution to the initial segregation of overlapping events; as such it addresses an issue we do not address here and may in the future be usefully integrated within the current proposal.

Once the various regularities have been discovered, and if the stimulus sequence continues as before, generic mechanisms of switching between alternative perceptual organizations come into play, and alternative interpretations of the sensory environment coexist. Our observations show that the initial parameter bias becomes weaker and a balance emerges between the alternatives, with the dominant organisations being segregation and integration. The initially rather surprising finding of a significant proportion of *both* responses provides further support for the notion that alternative perceptual organizations reported by listeners for a given sound sequence are simultaneously represented within the brain, even if we are not always aware of them (Bregman 1990).

CONCLUSIONS

When the auditory system is exposed to an unchanging sequence of sounds, which can be organized in more than one way, perceptual bistability is pervasive. There is no combination of features that we have tested for which perception remains stable for even a few minutes. Analysis of the experimental data revealed that the perceptual organization resulting from listening to such sound sequences can be characterized by two distinct set of processes: *formation* of sequential associations and *coexistence* between alternative interpretations. The first percept is co-determined by these two types of processes and its duration and sensitivity to stimulus features differs from those in subsequent perceptual phases, which are determined by the processes of coexistence only (i.e., the on-going competition between alternative organizations). A possible advantage of such processing is that the perceptual flexibility necessary for effective operation in real world environments may depend upon having a system which can balance on the verge of instability (Noest, van Ee et al. 2007). This would allow top-down processes to influence stimulus driven activity and rapidly select the perceptual organization best suited to the current behavioural goals.

ACKNOWLEDGEMENTS

REFERENCES

Anourova, I., Rama, P., et al. (1999): Selective interference reveals dissociation between auditory memory for location and pitch. *Neuroreport, 10*(17), 3543–3547.

Anstis, S., Saida, S. (1985): Adaptation to auditory streaming of frequency-modulated tones. *J Exp Psychol Hum Percept Perform, 11*(3), 257–271.

Beauvois, M. W., Meddis, R. (1996): Computer simulation of auditory stream segregation in alternating-tone sequences. *J Acoust Soc Am, 99*(4 Pt 1), 2270–2280.

Bee, M. A., Micheyl, C. (2008): The cocktail party problem: what is it? How can it be solved? And why should animal behaviorists study it? *J Comp Psychol, 122*(3), 235–251.

Bendixen, A., Bőhm, T. M., et al. (2011): Different roles of proximity and predictability in auditory stream segregation, <u>submitted</u>.

Bendixen, A., Denham, S. L., et al. (2010): Regular patterns stabilize auditory streams. *J Acoust Soc Am, 128*(6), 3658–3666.

Blake, R., Logothetis, N. K. (2002): Visual competition. *Nat Rev Neurosci, 3*(1), 13–21.

Brascamp, J. W., van Ee, R., et al. (2006): The time course of binocular rivalry reveals a fundamental role of noise. *J Vis, 6*(11), 1244–1256.

Bregman, A. S. (1990): *Auditory Scene Analysis,* MIT Press.

Bregman, A. S., Ahad, P. A., et al. (2000): Effects of time intervals and tone durations on auditory stream segregation. *Percept Psychophys, 62*(3), 626–636.

Brunswik, E. (1956): *Perception and the representative design of psychological experiments*. University of California Press, Berkeley, CA.

Carter, O., Cavanagh, P. (2007): Onset rivalry: brief presentation isolates an early independent phase of perceptual competition. *PLoS ONE, 2*(4), e343.

Cowan, N. (1984): On short and long auditory stores. *Psychol Bull, 96,* 351–370.

Cusack, R., Deeks, J., et al. (2004): Effects of location, frequency region, and time course of selective attention on auditory scene analysis. *J Exp Psychol Hum Percept Perform, 30*(4), 643–656.

Czigler, I., Winkler, I. (1996): Preattentive auditory change detection relies on unitary sensory memory representations. *Neuroreport, 7,* 2413–2417.

Denham, S. L., Gyimesi, K., et al. (2010): Stability of perceptual organisation in auditory streaming. In: *The Neurophysiological Bases of Auditory Perception*. Lopez-Poveda, E. A., Palmer, A. R., Meddis, R. (eds.) Springer, pp. 477–488.

Denham, S. L., Winkler, I. (2006): The role of predictive models in the formation of auditory streams. *J Physiol Paris, 100*(1–3), 154–170.

Elhilali, M., Ma, L., et al. (2009): Temporal coherence in the perceptual organization and cortical representation of auditory scenes. *Neuron, 61*(2), 317–329.

Elhilali, M., Shamma, S. A. (2008): A cocktail party with a cortical twist: how cortical mechanisms contribute to sound segregation. *J Acoust Soc Am, 124*(6), 3751–3771.

Fowler, C. A., Rosenblum, L. D. (1990): Duplex perception – a comparison of monosyllables and slamming doors. *J Exp Psychol Hum Percept Perform, 16*(4), 742–754.

Hochberg, J. (1981): Levels of perceptual organization. In: *Perceptual organization*. Pomerantz, M. K. J. (ed.), Erlbaum, Hillsdale, NJ, pp. 255–278.

Hohwy, J., Roepstorff, A., et al. (2008): Predictive coding explains binocular rivalry: an epistemological review. *Cognition, 108*(3), 687–701.

Horváth, J., Czigler, I., et al. (2001): Simultaneously active pre-attentive representations of local and global rules for sound sequences in the human brain. *Brain Res Cogn Brain Res, 12*(1), 131–144.

Köhler, W. (1947): *Gestalt Psychology*. Liveright, New York.

Kondo, H. M., Kashino, M. (2009): Involvement of the thalamocortical loop in the spontaneous switching of percepts in auditory streaming. *J Neurosci, 29*(40), 12695–12701.

Lee, S. H., Blake, R., et al. (2007): Hierarchy of cortical responses underlying binocular rivalry. *Nat Neurosci, 10*(8), 1048–1054.

Leopold, D. A., Logothetis, N. K. (1999): Multistable phenomena: changing views in perception. *Trends Cogn Sci, 3*(7), 254–264.

Liberman, A. M. (1982): On finding that speech is special. *American Psychologist, 37*(2), 148–167.

Mamassian, P., Goutcher, R. (2005): Temporal dynamics in bistable perception. *J Vis, 5*(4), 361–375.

Micheyl, C., Carlyon, R. P., et al. (2007): The role of auditory cortex in the formation of auditory streams. *Hear Res, 229*(1–2), 116–131.

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

Moreno-Bote, R., Shpiro, A., et al. (2010): Alternation rate in perceptual bistability is maximal at and symmetric around equi-dominance. *J Vis, 10*(11), 1.

Neisser, U. (1967): *Cognitive Psychology*. Appleton–Century–Crofts, New York.

Noest, A. J., van Ee, R., et al. (2007): Percept-choice sequences driven by interrupted ambiguous stimuli: a low-level neural model. *J Vis, 7*(8), 10.

Pastukhov, A., Braun, J. (2008): A short-term memory of multi-stable perception. *J Vis, 8*(13), 7 1–14.

Pressnitzer, D., Hupé, J. M. (2006): Temporal dynamics of auditory and visual bistability reveal common principles of perceptual organization. *Curr Biol, 16*(13), 1351–1357.

Ramnani, N. (2006): The primate cortico-cerebellar system: anatomy and function. *Nat Rev Neurosci, 7*(7), 511–522.

Rees, G., Kreiman, G., et al. (2002): Neural correlates of consciousness in humans. *Nat Rev Neurosci, 3*(4), 261–270.

Schadwinkel, S., Gutschalk, A. (2011): Transient bold activity locked to perceptual reversals of auditory streaming in human auditory cortex and inferior colliculus. *J Neurophysiol, 105*(5), 1977–1983.

Shinozaki, N., Yabe, H., et al. (2003): Spectrotemporal window of integration of auditory information in the human brain. *Brain Res Cogn Brain Res, 17*(3), 563–571.

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

Sussman, E. (2005): Integration and segregation in auditory scene analysis. *J. Acoust. Soc. Am., 117*(3), 1285–1298.

van Noorden, L. P. A. S. (1975): *Temporal coherence in the perception of tone sequences*, PhD Thesis. Eindhoven University of Technology, Leiden, The Netherlands.

Winkler, I., Denham, S. L., Nelken, I. (2009): Modeling the auditory scene: predictive regularity representations and perceptual objects. *Trends Cogn Sci, 13*(12), 532–540.

Winkler, I., Takegata, R., et al. (2005): Event-related brain potentials reveal multiple stages in the perceptual organization of sound. *Brain Res Cogn Brain Res, 25*(1), 291–299.

Yabe, H., Tervaniemi, M., et al. (1997): Temporal window of integration revealed by MMN to sound omission. *Neuroreport, 8,* 1971–1974.

Yabe, H., Winkler, I., et al. (2001): Organizing sound sequences in the human brain: the interplay of auditory streaming and temporal integration. *Brain Res, 897*(1–2), 222–227.