Stability of perceptual organisation in auditory streaming

Susan L. Denham[1], Kinga Gyimesi[2,3], Gábor Stefanics[2,4], and István Winkler[2,5]

sdenham@plymouth.ac.uk, kinga@cogpsyphy.hu, gstefan@cogpsyphy.hu, iwinkler@cogpsyphy.hu

[1]Centre for Theoretical and Computational Neuroscience, University of Plymouth, Drake Circus, Plymouth PL4 8AA, UK

[2]Department of General Psychology, Institute for Psychology, Hungarian Academy of Sciences, 1394 Budapest, P.O. Box 398, Hungary

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

[4]Department of Experimental Zoology and Neurobiology, University of Pécs, 7624 Pécs, Ifjúság st. 6, Hungary

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

Corresponding author; Susan Denham tel: +44 1752584913 sdenham@plymouth.ac.uk

Abstract

In everyday situations, we perceive sounds organized according to their source, and can follow someone's speech or a musical piece in the presence of other sounds without apparent effort. Thus it is surprising that recent evidence obtained in the most widely used experimental test-bed of auditory scene analysis, the two-tone streaming paradigm, demonstrated extensive bistability even in regions of the parameter space previously thought to be strongly biased towards a particular organisation. This raises the question of what aspects of the rich natural input allow the auditory system to form stable representations of concurrently active sound sources. Here we report the results of perceptual studies aimed at testing this issue. It is possible that the extreme repetitiveness of the alternating two-tone sequence, i.e., lack of change, causes perceptual instability. Our first experiment addressed this hypothesis by introducing random changes in the stimulation. It is also possible that under natural conditions, multiple redundant cues stabilize perception. The second experiment tested this hypothesis by adding a second cue which favoured one organisation. Much to our surprise, neither one of these manipulations stabilized the perception of two-tone streaming sequences. We discuss these experimental results in the light of our previous theoretical proposals and findings of significant differences between the first and later perceptual phases. We argue that multi-stability is inherent in perception. However, it is normally hidden by switches of attention which allow the return of the dominant perceptual organisation resulting in the subjective experience of perceptual stability. In our third experiment, we explored this possibility by inserting short gaps into the sequences, since gaps have been shown to reset auditory streaming similarly to switches in attention.

Keywords

auditory streaming, bi-stability, perceptual switching, auditory scene analysis

Introduction

The phenomenon of bistability in auditory perceptual organisation has been highlighted in a number of recent studies (Denham et al., 2006; Denham et al., 2008a; Pressnitzer et al., 2005; Pressnitzer et al., 2006; Winkler et al., 2005). These investigations have shown, using the auditory two tone streaming paradigm, that bistability is found extensively even in regions of the parameter space previously thought to be strongly biased towards a particular organisation (Denham et al., 2008a). Very similar characteristics of perceptual switching are observed in the visual and auditory modalities (Denham et al., 2008a; Pressnitzer et al., 2006), suggesting that at some level of perceptual processing, generic mechanisms may be involved. Furthermore, a notable difference was identified between the first percept evoked by the auditory streaming sequence (first perceptual phase) and subsequent percepts (subsequent perceptual phases), with the duration of the first phase being stimulus-parameter dependent and an order of magnitude longer in duration than the parameter-independent subsequent phases (Denham et al., 2008a).

These rather surprising results raised a number of further questions. Firstly, given that subjective impressions of the world tend to be rather stable, what factors are responsible for stabilizing perceptual organisation? Is it the case that the extreme repetitiveness of alternating two tone sequences causes instability (e.g., by exhausting the neural networks specifically responding to the given stimuli), or is it that that these stimuli are inherently ambiguous, and that in such cases perception simply entertains the alternatives? To address these questions, in the first experiment we introduced random changes in the frequency and timing of the tones, and in the second experiment a cue which favoured streaming was included. Finally, in a third experiment, we explored the possibility that switches in attention could trigger the return of the dominant perceptual organisation, and hence give rise to perceptual stability. In this experiment, we inserted short gaps into the two tone sequences, since gaps have been shown to reset auditory streaming similarly attentional switches (Cusack et al., 2004).

Experiment 1

The first experiment was designed to explore whether the extreme repetitiveness of the alternating two tone sequence was responsible for the observed bistability in perceptual organisation. In order to do this we introduced random changes by jittering both the timing and the frequencies of the alternating tones.

Participants

Eleven young healthy volunteers (6 female, 19-24 years of age, average: 21.0 years) participated in the first experiment. For all the experiment reported here, participants received modest financial compensation for their participation. The study was conducted in the sound-attenuated experimental chamber of the Institute for Psychology, Hungarian Academy of Sciences. It was 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 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 not exceed 15 dB in the same frequency range.

Stimulus paradigm

Participants were presented with 4-minute long trains of the ABA- structure, where A and B were pure tones of 75 ms duration, including 5 ms linear onset and offset ramps. In separate trains, Δf was 1, 4, 7 or 10 semitones (ST) and stimulus onset asynchrony (SOA) was 75, 100, 150, or 200 ms. Each combination of the parameters was presented in one stimulus block (altogether, 4 × 4 = 16 stimulus conditions), delivered in an order randomized separately for each participant. The mean frequency of the lower-pitched, more frequent A tones was a nominal 400 Hz, and that of the B tones was higher by the Δf for the corresponding condition. The actual frequency of each tone in the sequence was chosen randomly from a uniform distribution centred on the nominal frequency +/- 10% of the Δf. The onset of each tone was jittered in the range +/-20% of the SOA.

Procedure

The procedure for all experiments was as described here. While listening to the test sound sequences, participants sat in a comfortable chair in the experimental chamber. They were instructed 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. 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, and the notion of a segregated percept was defined 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 mark their perception throughout the duration of the stimulus sequence and not to attempt hearing the sound according to one or another perceptual organisation. The experimenter made sure that participants understood the types of percepts they were required to report, using both auditory and visual illustrations. To avoid possible implications of exclusivity between the two potential organisations, subjects were explicitly told that it was possible that they may sometimes hear both types of patterns at the same time. However, they were also cautioned to be sure to release the button when they stopped hearing the corresponding pattern. When analyzing the responses, we discarded all phases with duration shorter than 300 ms in order to avoid cases in which participants may simply have been slightly inaccurate in synchronising their button presses and releases.

Results

The principal finding was that the introduction of jitter in the frequencies and timing of the tones did not stabilize perceptual organisation. Once again perceptual switching was found in all conditions and for all subjects; the distribution of switching across conditions and subjects is shown in Figure 1.

Figure 1 here …………………………………………………………………………

We have previously found a profound difference between the first phase duration, and that of subsequent phases, with first phases typically an order of magnitude longer than the mean of subsequent phases (Denham et al., 2008b). An ANOVA test comparing the durations of the first and the mean of all subsequent phases (First vs. Subsequent Phases × Δf × SOA) showed that the first phase was significantly longer than the mean of subsequent phases (63.40 vs. 35.69 s; $F(1,10)=6.48$, $p<0.05$ with the η^2 effect size being 0.73). This effect was stronger for small than large Δf's, as was shown by the significant interaction between the First vs. Subsequent Phases and the Δf factors ($F(3,30)=4.25$, $p<0.05$, $\eta^2=0.30$ with the ε Greenhouse-Geisser correction of the degrees of freedom being 0.61). The mean duration of subsequent phases was longer than those we found in our previous study. So, perhaps, random jittering of the stimulus parameters results in a less distinct representation of the regularities, which is reflected in increased duration of phases following the first percept.

The results of this experiment suggest that it is not change *per se* which stabilises perceptual organisation. There may be some tendency for less switching in the presence of varying stimulus parameters compared to the case when stimulus parameters are kept constant throughout the stimulus, but it is not clear whether this is significant as the reduction in switching is well below the high inter-subject variability. Thus we can conclude that the bistability of the auditory streaming sequence is not an artefact produced by the constancy of stimulus parameters.

Experiment 2

In the second experiment, we investigated whether perceptual bistability in auditory streaming occurred because of the inherent ambiguity of the tone sequences. In order to do so we introduced a sound location cue supporting the segregated perceptual organisation. Our hypothesis was that this may stabilise perception since the integrated organisation would include tones coming from opposite directions.

Participants

Twelve young healthy volunteers participated in the experiment (5 female, 19-24 years of age, average: 21.4 years).

Stimulus paradigm

As in experiment 1, participants were presented with 4-minute long trains of the ABA-structure, where the A and B were pure tones of 75 ms duration, including 5 ms linear onset and offset ramps. In separate trains, Δf was 0, 0.5, 2, 4 or 6 semitones (ST) and SOA was 125 or 150 ms. Each of the combinations except for the 0 ST condition was presented with and without the location cue. Location was simulated by increasing the level of the signal to one ear by 6dB for all high tones and decreasing it by 5dB for the other ear, while doing the opposite for all low tones. Altogether, 4 × 4 +2 = 18 conditions were tested in separate blocks, the order of which was randomized separately for each participant. All other parameters and procedures were identical to Experiment 1.

Results

The principal finding was that the introduction of an additional unambiguous cue, which favoured segregation, did not stabilise perceptual organisation. Once again we found perceptual switching in all conditions and for all subjects; as illustrated in Figure 2.

Figure 2 here ………………………………………………………………………………

As was expected, adding the location cue significantly increased the probability of perceiving segregation (means: 0.32 and 0.53 without and with the location cue). The ANOVA (No vs. Location Cue × Δf × SOA) showed significant effects of all three factors

($F(1,11)=52.20$, $p<0.0001$, $\eta^2=0.83$; $F(3,33)=29.92$, $p<0.00001$, $\varepsilon=0.68$, $\eta^2=0.73$; $F(1,11)=5.98$, $p<0.05$, $\eta^2=0.35$ for the three factors, respectively) with no significant interaction between them. Although a previous similar study using short tone sequences showed that Δf and location difference interacted in determining the probability of hearing the sound sequence as two segregated sound streams (Farkas et al., 2006), the current results did not show significant interaction between the two cues. The probability of segregation calculated by assuming independence of the two cues ($P_{\Delta f+loc}(\Delta f, SOA)= P_{\Delta f}(\Delta f, SOA)+ P_{loc}(SOA)- P_{\Delta f}(\Delta f, SOA)\times P_{loc}(SOA)$) did not significantly differ from the probability of segregation measured when both cues were present (ANOVA: Independence-assuming model vs. Actual measurement × Δf × SOA; $F(1,11)=1.74$ for the comparison between the model and the measurement). This result can be explained by our previous finding showing that stimulus parameters mainly affect the first perceptual phase of auditory stream segregation, but not so much the subsequent perceptual phases (Denham et al., 2008a). However, independence of the frequency and locations cues can also be investigated by plotting the correlation of the number of switches between conditions with and without the location cue as a function of Δf and SOA. Figure 2C shows that the correlations between corresponding conditions with and without the location cue monotonically and dramatically increase at both SOA's with increasing Δf's; i.e. as Δf increases it begins to dominate perceptual organisation irrespective of the location cue. Thus the two cues do not act independently of each other.

Experiment 3

In the final experiment we tested the effects of short gaps on auditory streaming. The effects of gaps were assumed to model the effect of switching attention back to the sequence, as it has been shown that switching attention away and then back to the sound sequence, as well as the insertion of gaps as short as 500 ms in duration, restarted the build-up of auditory streaming (Cusack et al., 2004). In addition to testing the effects of gaps on the stability of auditory streams, we also tested 1) whether the effect was due to the introduction of a silent interval or whether any temporal violation would result in the reset of streaming, and 2) if the effect was indeed due to the introduction of silent gaps, what was the minimal duration of the gap needed to trigger a reset.

Participants

18 young healthy volunteers participated in the experiment (6 female, 18-25 years of age, average: 22.7 years).

Stimulus paradigm

As in experiment 1, participants were presented with 4-minute long trains of the ABA-structure, where the A and B were pure tones of 75 ms duration, including 5 ms linear onset and offset ramps. In separate trains, Δf was 4, 16 or 22 semitones (ST) and SOA was 150, 200 or 250 ms. In each stimulus block, two gaps were introduced. The first gap was inserted 90 s, and the second 180 s into the stimulus sequence. Each combination of Δf and SOA was delivered twice within the experimental session: once with the first gap being 500 ms and the second 150 ms long and once with the first gap being 200 ms and the second -200 ms long. The -200 ms value means that the onset of the ABA tone triplet following the 'gap' was moved closer to the previous tone triplet by 200 ms. The -200 ms gap was used to test whether breaking the uniform rhythm of the sequence without introducing a silent interval would have an effect similar to that of the silent gaps. The other gap durations were selected on the basis of pilot studies suggesting that only gaps exceeding the circa 170 ms long temporal window of integration (for a review, see (Cowan, 1984)) had a significant effect on auditory streaming.

Results

We found that 1) breaking the rhythm did not by itself affect auditory streaming, and 2) only silent gaps exceeding 150 ms had a significant effect on streaming. A comparison of the probability of integration in the 10 s long intervals preceding and following a gap, ANOVA (Before vs. After the gap × Gap duration × SOA × Δf), showed significant main effects of all four factors ($F(1,17)=20.72$, $p<0.001$, $\eta^2=0.55$, $F(3,51)=7.59$, $p<0.001$, $\varepsilon=0.84$, $\eta^2=0.90$, $F(2,34)=14.04$, $p<0.001$, $\varepsilon=0,77$, $\eta^2=0.45$, and $F(2,34)=48,28$, $p<0,001$, $\varepsilon=0,70$, $\eta^2=0.74$ for Before vs. After the gap, Gap duration, SOA, and Δf, separately). The probability of perceiving integration significantly increased after the gap compared to before the gap and with increasing gap duration. However, there was a significant interaction between Before vs. After the gap and Gap duration ($F(3,51)=8.53$, $p<0.0002$, $\varepsilon=0.80$, $\eta^2=0.33$). Figure 3 shows the probability of integration as a function of gap duration, separately before and after the gap. Whereas the duration of the gap obviously had no effect on the probability of integration during the 10-s interval preceding the gap, the probability of perceiving the integrated organization after the gap increased with increasing gap duration. With the negative 'gap' and the 150 ms gap, the difference between the pre- and post-gap probability of integration did not reach significance. In contrast, with the 200 ms long gap, there was a tendency for a difference ($p<0.08$, post-hoc Sheffé pair-wise comparisons with df=51 and MSE=0.98), which became significant by the 500-ms long gap ($p<0.0001$). Figure 3 also shows that the probability of perceiving integration after the gap increases rapidly between gap durations of 150 and 200 ms, suggesting that temporal integration plays an important role in this effect.

The influence of gaps on the stability of perception was studied by comparing the duration of the perceptual phase preceding and following the gaps. We found no significant increase in the duration of the perceptual phase immediately following the gap at any gap duration (ANOVA: Before vs. After the gap × Gap duration × SOA × Δf; $p<0.09$, F(1,17)=3.36=0.09 for the Before vs. After the gap factor). This means that gaps do not have a stabilizing effect on the perception of the auditory streaming sequence. Thus either attentional switches do not help in stabilizing auditory perception or, despite the similarities found between the effects of gaps and attention switching on the reset of streaming (Cusack et al., 2004), the two manipulations affect different processes.

Discussion

The experiments reported here were designed to investigate what aspects of natural sounds allow the auditory system to form stable representations of concurrently active sound sources, given the demonstration that auditory perceptual organisation can be bi- or even multi-stable. We tested 1) whether the extreme repetitiveness of the alternating two-tone sequence causes perceptual instability; 2) whether under natural conditions, multiple redundant cues stabilize perception; and 3) whether attentional switches lead to a reset of perceptual organisation thereby leading to apparently stable perception.

Much to our surprise we found that randomly jittering the frequencies and timing of the tones did not stabilize perceptual organisation; i.e. neither constancy nor precise predictability of stimulus parameters are necessary for bi-stable perception. As long as a predictable *distribution* of stimulus features exists, perceptual switching can occur. There was a tendency for phase durations to increase with a decrease in the precision with which the regularity can be represented, but this cannot explain the perceptual stability experienced in real-life situations. Even so, this aspect of the results requires further investigation.

The introduction of an additional cue, which unambiguously favoured segregation rather than integration, similarly failed to stabilise perceptual organisation. Subjects continued to report hearing the integrated organisation even in the face of interaural level differences which gave the impression that the alternating tones came from different directions.

However, the biasing cue did increase the probability of perceiving segregation, even though it did not prevent perceptual switching. The influence of this cue was strongly modulated by the strength of the primary cue to segregation, namely Δf. This is consistent with other auditory streaming experiments showing the dominance of the spectral composition of the alternating tones over other features such as pitch (Vliegen et al., 1999).

The final hypothesis we explored was whether switches in attention may lead to perceptual stability. We investigated this by introducing silent gaps into the tone sequence, as Cusack and colleagues have previously argued that gaps have a similar effect on the build up of streaming to switching attention (Cusack et al., 2004). The main conclusion that we can draw from this experiment is that the introduction of even very brief gaps can increase the probability of reporting integration. However, this does not amount to a complete return to the starting state, as evidenced by the lack of a long duration perceptual phase (a 'first phase'), following the gap.

Violation of predictability is not sufficient for a reset to occur. The -200ms 'gap' condition violated the expected timing of the following triplet significantly, but did not increase the probability of a return to integration. A further requirement is that the gap should exceed the window of temporal integration. This finding suggests that temporal integration may play an important role in perceptual sound organization. It is, as if gaps of shorter durations are disregarded by the system, possibly regarded as natural variation in timing. A more formal description of the effect could be based on extending Zwislocki's model of temporal integration (Zwislocki, 1969) to expected sounds.

The results of the gap experiment can be compared to the effect of gaps in binocular rivalry experiments. It has been shown that when subjects view a bi-stable visual stimulus which is presented intermittently, then they often resume the perceptual organisation that they experienced before the interruption (Leopold et al., 2002). Frequently interrupting the stimulus can lead to prolonged periods of perceptual stability. Further experiments will be needed to investigate whether a similar effect holds in auditory streaming. Nevertheless, the increased tendency to report integration after a gap suggests that intermittent presentation may lead to increased perceptual stability.

Acknowledgements

This work was supported by the European Research Area Specific Targeted Projects EmCAP (IST-FP6-013123), and SCANDLE (IST-FP7-231168).

Figure Legends

Figure 1. The number of perceptual switches; left: for each condition with the Δf and SOA indicated; right: for each subject. Black dots indicate individual subjects (left) or conditions (right); solid lines connect the mean for each condition/subject.

Figure 2. The number of perceptual switches for each condition; A: the 125ms SOA conditions, and B: the 150ms SOA conditions. ‘L’ indicates the presence of the location cue, no ‘L’ indicates that the location cue was absent. Black dots are for individual subjects, and solid lines connect the mean of each condition. C. Interactions between Δf and the location cue, showing the correlation coefficients between corresponding Δf conditions with and without the location cue.

Figure 3. Group-averaged (n=18) probability of perceiving integration in the 10-s long periods before (dashed line) and after (solid line) the gap as a function of the gap duration.

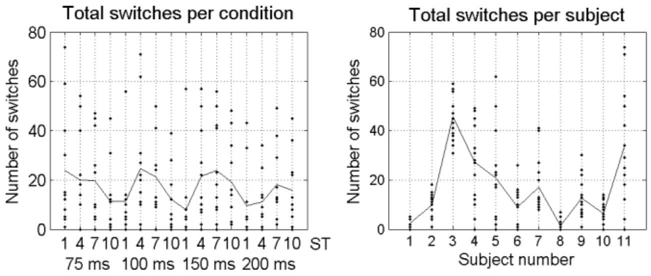

Figure 1

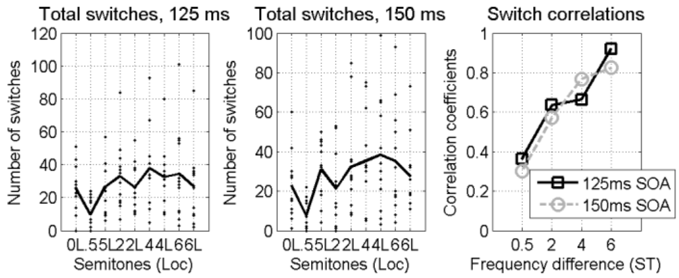

Figure 2

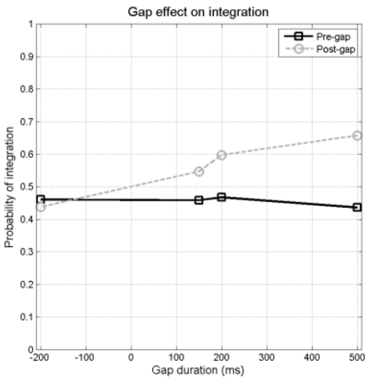

Figure 3

References

Cowan, N. 1984. On short and long auditory stores. Psychol Bull 96, 351-370.

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. J Exp Psychol Hum Percept Perform 30, 643-56.

Denham, S.L., Winkler, I. 2006. The role of predictive models in the formation of auditory streams. J Physiol Paris 100, 154-70.

Denham, S.L., Gyimesi, K., Stefanics, G., Winkler, I. 2008a. Perceptual bi-stability in auditory streaming: How much do stimulus features matter? Hearing Research under revision.

Denham, S.L., Montbrio, E., M, C., Smith, L.M., Honing, H., Ladinig, O., E, B.-B., Purwins, H. 2008b. Theoretical insights into music cognition and proposals for a generic model. University of Plymouth, Plymouth.

Farkas, A., Dudás, T., Horváth, J., Sussman, E., Winkler, I. 2006. Jelzőmozzanatok integrációja hangláncok elválasztásában [Cue integration in separating auditory streams]. A Magyar Pszichológiai Társaság XVII., Országos Tudományos Nagygyűlése, Budapest.

Leopold, D.A., Wilke, M., Maier, A., Logothetis, N.K. 2002. Stable perception of visually ambiguous patterns. Nat Neurosci 5, 605-9.

Noest, A.J., van Ee, R., Nijs, M.M., van Wezel, R.J. 2007. Percept-choice sequences driven by interrupted ambiguous stimuli: a low-level neural model. J Vis 7, 10.

Pressnitzer, D., Hupé, J.M. 2005. Is auditory streaming a bistable percept?, Forum Acusticum, Budapest. pp. 1557-1561.

Pressnitzer, D., Hupe, J.M. 2006. Temporal dynamics of auditory and visual bistability reveal common principles of perceptual organization. Curr Biol 16, 1351-7.

Vliegen, J., Moore, B.C., Oxenham, A.J. 1999. The role of spectral and periodicity cues in auditory stream segregation, measured using a temporal discrimination task. J Acoust Soc Am 106, 938-45.

Winkler, I., Takegata, R., Sussman, E. 2005. Event-related brain potentials reveal multiple stages in the perceptual organization of sound. Brain Res Cogn Brain Res 25, 291-9.

Zwislocki, J.J. 1969. Temporal summation of loudness: An analysis. . J Acoust Soc Am 46, 431 -440.