

MODULATION-FREQUENCY ACTS AS A PRIMARY CUE FOR AUDITORY STREAM SEGREGATION

ORSOLYA SZALÁRDY^{1,2,}, ALEXANDRA BENDIXEN^{3,4,} DÉNES TÓTH^{1,} SUSAN L. DENHAM⁵ and
ISTVÁN WINKLER^{1,6}

¹ Institute of Psychology and Cognitive Neuroscience, Research Centre for Natural Sciences, Hungarian Academy of Sciences, H-1394 Budapest, P.O. Box 398, Hungary

² Department of Cognitive Science, Faculty of Natural Sciences, Budapest University of Technology and Economics, H-1111 Budapest, Stoczek u. 2, Hungary

³ Institute of Psychology, University of Leipzig, Seeburgstr. 14–20, D-04103 Leipzig, Germany

⁴ Institute of Psychology, Carl von Ossietzky University of Oldenburg, Ammerländer Heerstr. 114–118, D-26129 Oldenburg, Germany

⁵ Cognition Institute and Centre for Theoretical and Computational Neuroscience, University of Plymouth, Drake Circus, Plymouth PL4 8AA, UK

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

ABSTRACT

In our surrounding acoustic world sounds are produced by different sources and interfere with each other before arriving to the ears. A key function of the auditory system is to provide consistent and robust descriptions of the coherent sound groupings and sequences (auditory objects), which likely correspond to the various sound sources in the environment. This function has been termed auditory stream segregation. In the current study we tested the effects of separation in the frequency of amplitude modulation on the segregation of concurrent sound sequences in the auditory stream-segregation paradigm (van Noorden 1975). The aim of the study was to assess 1) whether differential amplitude modulation would help in separating concurrent sound sequences and 2) whether this cue would interact with previously studied static cues (carrier frequency and location difference) in segregating concurrent streams of sound. We found that amplitude modulation difference is utilized as a primary cue for the stream segregation and it interacts with other primary cues such as frequency and location difference.

Keywords: auditory streaming, perceptual bi-stability, amplitude modulation

INTRODUCTION

In everyday auditory environments several sound sources may produce discontinuous sequences of sounds which arrive to the ears as a mixture of overlapping events. In order to allow us to make sense of our surroundings, the auditory system decomposes the mixture of sounds and forms coherent sequences from them, termed *auditory streams*. Auditory streams often match the concurrently active sound sources in the environment (Bregman 1990, Griffiths and Warren 2004). Since Bregman's (1990) seminal book, the organization of the acoustic input into auditory streams has been studied within the framework of "auditory scene analysis" (for recent reviews, see Snyder and Alain 2007, Ciocca 2008, Winkler et al. 2009, Shamma et al. 2011).

The formation of streams is often tested by mixing together two sets of sounds that are separated by a difference in some acoustic feature. Such sequences are often heard as two separate streams of sound (termed "auditory streaming"; van Noorden 1975, Bregman 1990). Although most of the cues underlying auditory stream segregation are based on spectral differences between the two sets of sounds (Hartmann and Johnson 1991), Grimault and colleagues (Grimault et al. 2000, 2002) showed that auditory streaming can also be based on time-varying cues alone. For example, Grimault et al. (2002) evoked auditory streaming by differences in amplitude modulation (AM) frequency (see also Dollezal et al. 2012).

Recently, Bendixen and colleagues (Bendixen et al. 2010) showed that auditory stream segregation is also supported by higher-order temporal regularities. However, in contrast to spectral cues, which can initiate the formation of auditory streams and are thus considered primary cues (see Bendixen et al., and Denham et al. in this issue), higher-order temporal regularities only stabilized streams that had been formed on the basis of spectral cues. Following up on these findings, we set up the present experiment to investigate how differences in AM frequency affect auditory stream segregation. Can one form streams on the basis of such cues alone (i.e., do they act as primary cues) or do they only stabilize streams (similar to higher-order temporal patterns)? Furthermore, we wished to determine the relationship between cues based on the fluctuation of the temporal envelope, spectral and spatial cues of auditory stream segregation. Do they increase the likelihood of streaming independently of each other, or do their effects interact with each other as was found for spectral and spatial cues (Denham et al. 2010, see, however, Du et al. 2011), suggesting that they rely at least partly on common mechanisms? In the current study, we tested the effects of AM frequency differences on stream segregation, using this cue alone, as well as combining it with spectral and spatial cues.

In the traditional "auditory streaming paradigm" (van Noorden 1975), participants are presented with a repeating three-tone (ABA-) sound sequence, where 'A' and 'B' denote two different sounds and '-' stands for a silent interval equaling the common duration of the two tones. When participants are asked about their perception after listening to the sequence for a few seconds, they may report hearing a single stream with a galloping rhythm (termed the "integrated percept"), or two separate homogeneous streams with different but even presentation rates, one for each sound (the "segregated percept"). These studies showed that by increasing the feature separation (e.g., frequency difference: Δf) between the 'A' and 'B' sounds (typically tones), and decreasing the silent interval separating the sounds in time (Δt), the segregated percept became more and more probable; whereas decreasing the feature

difference and/or increasing the Δt increased the probability of the integrated percept (Moore and Gockel 2002). Van Noorden (1975) found a trade-off between Δf and the stimulus onset asynchrony (SOA; onset-to-onset interval, which also characterizes Δt with constant sound durations).

Whereas most early studies using the auditory streaming paradigm presented short (< 20 s) sound sequences and asked listeners about their perception at the termination of the sequence (but see Anstis and Saida 1985), some recent studies (Roberts et al. 2002, Denham and Winkler 2006, Pressnitzer and Hupe 2006, Denham et al. 2010, Bendixen et al. 2010, see also Bendixen et al., and Denham et al. in this issue) delivered longer sequences (e.g., of 4 minutes duration) and asked participants to report their perception continuously throughout the sound sequences. These studies found that perception switches back and forth between the integrated and the segregated percept throughout the whole stimulus train, and likened this phenomenon to visual observations of perceptual bistability (Pressnitzer and Hupe 2006, for reviews of visual bistability, see Leopold and Logothetis 1999, Blake and Logothetis 2002).

In one of these studies, Denham and colleagues (Denham et al. 2010) found that although the combination of two primary static cues (separation in tone frequency and source location) increased the probability of the segregated percept, these cues did not substantially reduce switching between the two sound organizations; thus they did not stabilize perception. This is in contrast to everyday environments in which auditory percepts do not seem to fluctuate. One may argue that static cues do not fully characterize natural sound streams. Veridical decomposition of complex natural auditory scenes may additionally benefit from monitoring the time-varying cues of stream segregation. In the current experiment we tested whether the effects of separation in the frequency of AM (Grimault et al. 2002, Dollezal et al. 2012), combined with spectral and spatial cues might lead to perceptual stabilization in the auditory streaming paradigm. Both Grimault et al. (2002) and Dollezal et al. (2012) presented the AM frequencies above 20 Hz. Previous studies suggested that AM frequencies below 20 Hz are associated with the sensation of rhythm, whereas above 20 Hz they influence pitch (Schulze and Langner 1997). In this case, we should expect that a difference between AM's of >20 Hz frequencies will be utilized as a primitive cue of stream segregation similarly to pitch difference. However, Bregman et al. (1985) reported that when two sinusoidal tones, set in the region of the first and second formants of the human voice, respectively, were modulated by the same ca. 100 Hz frequency, they tended to be fused more strongly than when the two tones were modulated by different frequencies. Bregman and colleagues (1985) concluded that AM information may be utilized independently from frequency difference in separating two sounds. Thus it is possible that AM frequency difference will have different effects from pitch or location difference, perhaps similar to higher-order dynamic cues.

Continuous recording of the participants' perceptual report allows one to measure the average duration of each percept. This allows one to separately assess the effects of a cue on the alternative percepts. Primary static cues for sound source segregation, such as frequency separation, both increase the duration of the segregated percepts and decrease the duration of the integrated percept. Based on the assumption that bistability of auditory streaming reflects continuous competition between alternative sounds organizations (Denham and Winkler 2006, Winkler et al. 2009), cues increasing the duration of the segregated percepts are regarded as having a stabilizing effect on the segregated organization. On the other hand, cues decreasing the duration of the integrated percepts increase the likelihood of a switch towards the

segregated organization (competitive effect). Primary static cues exert both effects on auditory streaming. Bendixen et al. (2010) found that introducing regularly repeating sequential patterns separately into the high and low tones of the ABA sequence influenced only the duration of the segregated percept, but not that of the integrated one. This result indicated that higher-order sequential cues can only stabilize a sound organization already formed on the basis of primary cues, but they cannot help in the initial discovery of or switching towards a sound organization (see also Bendixen et al. in this issue). In the current study, we tested the effect of AM frequency difference on the proportion and duration of the segregated and integrated percepts in the auditory streaming paradigm. If the auditory system utilizes the AM frequency difference as a primary cue, then with increasing AM frequency difference, we should see an increase in the proportion and average duration of the segregated percept as well as a decrease in the proportion and average duration of the integrated percept. In contrast, if the effect of the AM cue is similar to higher-order dynamic cues, then with increasing AM frequency difference, we should see an increase in the average duration of the segregated percept, but no decrease in the average duration of integrated phases.

We also tested the interaction of the AM frequency difference cue with two types of static cues: carrier-frequency difference and perceived sound source location difference (henceforth termed 'location'). If the AM frequency difference cue is processed independently from these static cues, then their effects on the proportion of the different percepts should be additive.

METHODS

Participants

Twenty-eight young healthy volunteers (18–27 years of age, average: 21.5, 15 female and 13 male) participated in the experiment. Participants gave written informed consent after the experimental procedures were explained to them. The study was approved by the Ethical Committee of the Institute for Psychology of the Hungarian Academy of Sciences. Participants were pre-selected by a clinical audiometric test, with the criteria that their hearing threshold would not exceed 25 dB hearing level and 10 dB difference between the ears in the 250 to 8000 Hz frequency range. The audiometric test was performed using an SA-5 audiometer (Mediroll, Hungary). The study was conducted in a sound-attenuated experimental chamber at the Institute for Psychology of the Hungarian Academy of Sciences. Participants received modest financial compensation.

Stimulus paradigm

Participants were presented with four-minute trains of the 'ABA-' structure. 'A' and 'B' were tones of 75 ms duration including 10 ms raised-cosine onset and offset ramps. Each tone was amplitude-modulated with a sinusoidal envelope; the depth of the modulation was 100%. The 'A' tone frequency was always 1000 Hz. In the stimulus conditions employing a carrier-frequency difference, the 'B' tone frequency was 4 semitones [ST] higher. The 'A' tone was always modulated at 100 Hz. In the stimulus conditions employing an AM frequency difference, the AM frequency of the 'B' tone was 0.5, 1, or 1.5 octaves higher than that of the 'A' tones (these values were selected based on the results of Dollezal et al. (2012)). In conditions without

location difference, the 'A' and 'B' tones were delivered binaurally with Sennheiser HD 600 headphones at 40 dB intensity above the participant's individual hearing level established for the 'A' tones. In the stimulus conditions employing a location difference, the 'A' tone was 6 dB louder on the left channel, and 5 dB softer on the right channel, whereas 'B' tones were 5 dB softer on the left and 6 dB louder on the right channel. The stimulus onset asynchrony (SOA) was uniformly 150 ms, including that before the 'silent' ('-') part of the 'ABA-' pattern (i.e., the SOA between consecutive 'A' tones was always 300 ms, whereas the SOA between consecutive 'B' tones was always 600 ms). Each stimulus condition was delivered separately in a single stimulus block; the order of the stimulus blocks was separately randomized for each participant. Stimulus conditions, altogether 16, differed in the cues that separated the 'A' and 'B' tones. Each level of the AM frequency difference (0, 0.5, 1, and 1.5 octaves) appeared either a) alone (4 conditions), b) together with carrier-frequency difference (4 conditions), c) together with location difference (4 conditions), or d) together with both carrier-frequency and location difference (4 conditions).

The stimulus block with no difference between the 'A' and 'B' tones served as a catch trial. For the statistical analyses, participants were grouped according to whether they reported the same percept throughout the catch trial stimulus block ("faithfully reporting" group) or reported switching more than once or after the first 60 seconds of this stimulus block ("doubtfully reporting" group). Although it is possible that some of the "doubtfully reporting" participants indeed experienced illusory perceptual changes during the presentation of the identical sound triplets, it is also possible that they responded to the implicit pressure of reporting perceptual switches suggested by the presence of two response keys and the careful instructions regarding how to mark the possible changes of perception (see below). In order to account for this possible bias, we pooled the responses separately for the two participants groups and compared them in our statistical analyses (see below). Overall, 14 participants (19–24 years of age, average: 22.1, 8 female and 6 male) fell into the "faithfully reporting" group, and 14 participants (18–27 years of age, average: 21.0, 7 female and 7 male) fell into the "doubtfully reporting" group.

Procedure

Participants were instructed to report their perception continuously during the stimulus blocks by depressing and releasing two response keys, one held in each hand. The assignment of the two response keys was counterbalanced across participants to eliminate the effects of handedness. Participants were instructed not to press either of the response keys if they did not hear any regularly repeating sound pattern (termed the "Neither" percept). When they heard regularly repeating identical tones of only one type (either 'A' or 'B'), they were asked to press one of the keys and keep it depressed as long as they heard the same repeating pattern. The same response key was to be depressed also when the participant heard two different tones repeating at the same time, as long as the different tones were not joined together in a repeating pattern (the "Segregated" percept). Participants were to press the other response key when they heard some regularly repeating pattern that included both 'A' and 'B' tones (the "Integrated" percept; e.g., ABA-ABA-...). Finally, participants were instructed to depress both keys when they heard a repeating pattern which included both types of tones (e.g. AB--AB...) as well as a repeating pattern made up of identical tones (e.g., --A---A...): the "Both" percept. They

were asked to follow their perception closely with the key combinations, so that they always marked their actual perception.

The experimenter made sure that each participant understood the types of percepts they were required to report, using both auditory and visual illustrations. The instruction phase lasted for 15–25 minutes, depending on the participant. The 16 stimulus blocks were delivered in a randomized order. During the experiment, participants were seated in a comfortable chair. They were allowed to relax for about a minute between consecutive stimulus blocks with longer breaks inserted whenever the participant asked for it. On average, the experimental session lasted for 120 minutes.

Measurement and data analysis

The continuous response signal was broken down into “perceptual phases”. A perceptual phase is the continuous interval within which the participant depressed the same response key combination, marking that he/she perceived the sound sequence throughout this interval as either integrated, segregated, both, or neither (see the definition of the percepts to be discriminated by the participants above). Perceptual phases shorter than 300 ms were not analyzed, because they may represent inaccurate coordination of key presses and releases (Moreno-Bote et al. 2010). The type and duration of each perceptual phase was extracted for each subject and condition. From this information, the proportion of each type of percept and the average phase duration were calculated. The proportion and phase duration of a percept was taken to be 0 for participants/conditions where the given percept did not appear within the corresponding stimulus block.

The conditions with no AM frequency difference were introduced for comparison with our previously reported experiment using the same stimulation parameters (Denham et al. 2010), for the catch trial (see above), and for testing the independence of cue effects. Results in these conditions were not included in testing the effects of AM frequency difference because the no-difference between ‘A’ and ‘B’ conditions cannot be compared with the conditions including an AM frequency difference, only.

Because the “both” and “neither” percepts appeared relatively rarely (both <15% and neither <1%) during the stimulus blocks and not at all in many combination of participant & condition, we only conducted statistical analyses for the “integrated” and the “segregated” percepts, which were perceived by most participants in most stimulus conditions. Percept proportion and average phase duration values were separately entered into mixed-mode analyses of variance (ANOVA¹) of the structure: Participant group (faithfully reporting vs. doubtfully reporting; grouping factor) × AM frequency difference (0.5 vs. 1 vs. 1.5 octave) × Carrier-frequency difference (present vs. absent) × Location difference (present vs. absent). Where appropriate, Greenhouse–Geisser correction of sphericity violations was applied and the ϵ correction factor is reported together with the η^2 effect size. Post-hoc analyses were conducted using the Tukey HSD test. In addition, the independence of the effects of the different cues on the proportion of the segregated and integrated percepts was tested by comparing with dependent Student’s *t*-tests the proportion of the integrated/segregated percept when both or all three cues were present with that modeled by assuming independent effects of the individual cues (e.g., P_{ab} was compared with $P_{model} = P_a + P_b - P_a \times P_b$, where P_a and P_b denote the observed proportions of the given percept with only cue ‘a’ or ‘b’ present and P_{ab} the

observed proportion with both cues ‘a’ and ‘b’ present; a similar model can be set up for three independent cue effects). Statistical analyses were conducted with the STATISTICA software. All statistically significant effects are reported.

RESULTS

Table 1 summarizes all significant effects obtained in the ANOVA tests.

Table 1. Significant effects obtained in the ANOVA tests. “Measure” indicates the measure compared: intProp = proportion of integrated phases; segProp = proportion of segregated phases; intDur = average duration of all integrated phases; segDur = average duration of all segregated phases. “Factors” are the ANOVA factors: AM = AM frequency difference; CF = carrier-frequency difference, LOC = location difference; Group = “Faithfully reporting” vs. “Doubtfully reporting” participants. Degrees of freedom (*df*), *F* values, significance levels (*p*), Greenhouse–Geisser correction factors (where applicable), and η^2 effect sizes are shown for each significant ANOVA effect.

Measure	Factor	df	F	p	G-G ϵ	η^2
intProp	AM	2, 52	51.071	.001	0.958	0.663
	CF	1, 26	82.782	.001		0.761
	LOC	1, 26	124.416	.001		0.827
	LOC×Group	1, 26	11.045	.01		0.298
	AM×CF	1, 52	36.733	.001	0.826	0.586
	AM×LOC	2, 52	15.019	.001	0.823	0.366
segProp	AM	2, 52	14.073	.001	0.875	0.351
	CF	1, 26	50.121	.001		0.658
	LOC	1, 26	79.960	.001		0.755
	AM×CF	2, 52	12.455	.001	0.941	0.324
	AM×LOC	2, 52	9.028	.001	0.861	0.258
intDur	AM	2, 52	15.164	.001	0.734	0.368
	CF	1, 26	32.037	.001		0.552
	LOC	1, 26	26.518	.001		0.505
	AM×CF	2, 52	20.784	.001	0.656	0.444
	AM×LOC	2, 52	14.685	.001		0.361
segDur	AM	2, 52	13.533	.001	0.785	0.342
	LOC	1, 26	17.362	.001		0.400

Proportions of the integrated and segregated percepts

With increasing AM frequency difference the proportion of the integrated percept decreased and the proportion of the segregated percept increased. Similar effects were observed for carrier-frequency and for location difference. When present, these cues decreased the proportion of the integrated percept and increased the proportion of the segregated percept within the stimulus block.

AM frequency difference interacted with carrier-frequency difference as well as with location difference in decreasing the proportion of the integrated percept (*Fig. 1*). In both cases, the slope of the AM frequency effect was steeper without than with the additional (carrier-frequency/location) difference. This was shown by the lack of significant differences between the proportions recorded for the 1 and 1.5 octaves AM frequency difference in the

presence of either one of the additional cues (Tukey HSD with $df = 52$: $p = 0.451$ and $p = 0.977$ for the carrier-frequency and location difference, respectively), whereas the same differences were significant in the absence of additional cues (Tukey HSD with $df = 52$: $p < 0.001$, both).

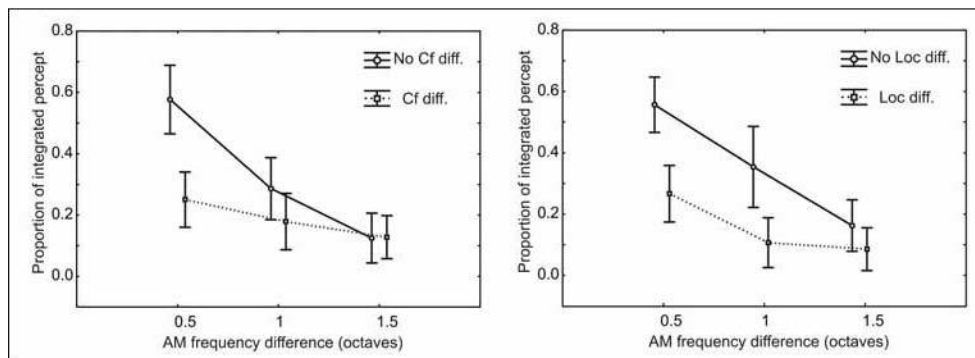


Figure 1. Proportion of the integrated percept: Interactions between AM frequency difference and carrier-frequency difference (left) and between AM frequency difference and location difference (right). The x axes show the levels of the AM frequency difference in octaves; the y axes show the proportion of the integrated percept within the stimulus blocks. Solid lines indicate the absence of carrier-frequency/ location difference, dotted lines indicate their presence.

The tests of the independence of cue effects revealed that the smallest AM frequency difference (0.5 octaves) and carrier-frequency difference did not interact with each other in determining the proportion of the integrated percept: The observed proportion was not significantly different from the proportion calculated by the model for independence. However, with higher AM frequency differences (1 and 1.5 octaves) these two cues significantly interacted with each other ($p < 0.01$). Furthermore, when adding the location cue on top of the AM frequency difference and carrier-frequency difference, significant interactions were found for each level of the AM frequency difference (all p values < 0.05) both in the proportion of the integrated and the segregated percept. In all of these cases the theoretical proportion, which was calculated by assuming independent cue effects, was lower for the integrated, and higher for the segregated percept than the proportion actually measured. This means that these cues are sub-additive in determining the segregation of auditory streams.

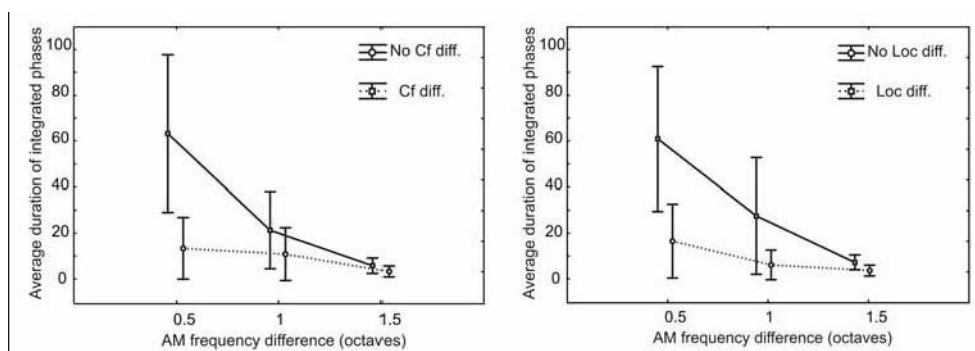


Figure 2. Phase duration of the integrated percept: Interactions between AM frequency difference and carrier-frequency difference (left) and between AM frequency difference and location difference (right). The x axes show the levels of the AM frequency difference in octaves; the y axes show the average phase duration of the integrated percept in seconds. Solid lines indicate the absence of carrier-frequency/ location difference, dotted lines indicate their presence.

Average perceptual phase durations

All three cues significantly influenced the durations of the integrated phases, and AM frequency difference and location difference had a significant effect on the duration of segregated phases: Increasing separation between the 'A' and 'B' tones decreased the average duration of the integrated phases and increased that of the segregated phases (*Fig. 2*). None of the interactions between the effects of the different cues on average phase durations were significant for the segregated percept. For the integrated percept, AM frequency difference interacted with the carrier-frequency difference as well as with location difference. Similarly to the proportion variable, the interaction was caused by differences in the slope of the AM frequency effect in the presence vs. absence of an additional cue (carrier-frequency/location difference). This was shown by the lack of significant differences between different AM frequency differences in the presence of another cue ($df = 52, p > 0.42$) whereas the same differences were significant in the absence of either of the other cues ($df = 52, p < 0.001$). The interaction between the carrier-frequency difference and the location difference was also significant. Either one of these two cues alone or both together significantly decreased the duration of the integrated percept ($df = 26, p < 0.001$) but the effect caused by a single cue alone did not differ from the effect caused by the presence of both cues together ($df = 26, p > 0.39$).

Possible biasing effects of the instructions

The only significant effect of sorting participants into "Faithfully reporting" and "Doubtfully reporting" groups on the measured variables was an interaction between the Participant group factor and the effect of presence vs. absence of the location difference cue on the proportion of the integrated percept: A higher proportion of the integrated percept was reported by "Faithfully reporting" than "Doubtfully reporting" participants in the absence of the location difference cue. No other significant main effect or interaction included the Participant group factor. Therefore, we conclude that the pattern of results was largely similar in the two groups of participants.

DISCUSSION

We found that with increasing separation in AM frequency, the proportion of perceiving two segregated streams increased and the proportion of perceiving a single integrated stream decreased. These results are consistent with those of Grimault and colleagues (Grimault et al. 2002) as well as Dollezel et al. (2012). Therefore, we conclude that separation in the frequency of amplitude modulation is utilized as a cue in auditory stream segregation. Importantly, we found that increasing the separation in AM frequency increased the duration of segregated perceptual phases and decreased the duration of integrated phases. Thus, similarly to primary static cues, separation in AM frequency is a cue that has both competitive and stabilizing effects on auditory streaming. That is, when integration is the currently dominant percept, the AM frequency cue increases the chances of a switch towards the segregated percept. And, further, when segregation is the currently dominant percept, the AM frequency cue helps to maintain the dominance of this percept. Similar effects were found for the carrier-frequency and the

location difference tested, replicating previous findings (Denham et al. 2010, and in press, Bendixen et al. in this issue). These results suggest that AM frequency difference similarly affects the segregation of concurrent sound streams as primary static cues. Therefore we conclude that AM frequency difference is a primary cue of auditory stream segregation.

We found strong interactions between the effects of the AM frequency cue and those of the two static cues on most measures. The nature of these interactions appears to be a ceiling type of effect. That is, with the exception of the smallest tested amount of AM frequency difference, adding either or both of these other cues causes little further increase in segregation and/or reduction in integration. The test of the independence model suggests that AM frequency and carrier-frequency difference have partly overlapping effects on the segregation of auditory streams. Except for the smallest AM frequency difference, these cues interacted with each other in their effects on the proportion of integrated percept. At the smallest (0.5 octave) AM frequency difference, perhaps there was enough room for each cue to reach its full effect. These results suggest that AM frequency and carrier-frequency difference may affect auditory stream segregation through partly overlapping mechanisms. This result is compatible with the observation that an AM of >20 Hz frequency modulation rate (such as 100 Hz) may influence the perceived pitch, which would explain the interaction between the effects of AM frequency difference and pitch difference on auditory streaming. Adding the location cue extended the interactions also to the proportion of the segregated percept. In all cases, the values measured for these combinations of cues showed less segregation and more integration than those predicted by the model based on the independence of these effects. These observations as well as the lack of interaction at lower levels of the AM frequency difference suggest that the interactions could have resulted from a ceiling effect. That is, it is possible that overall effect of these types of cues is limited. Full stability of the percept can only be achieved by qualitatively different cues. This suggestion is compatible with the results of Denham et al. (2010), who also found bistability with a number of different manipulations of the auditory streaming paradigm. If this was the case, then the interactions between the cues tested in the current study may have resulted from a bottleneck separating the stage of processing at which these cues are taken into account and another stage, which imposes bi/multi-stability on the outcome of this stage. Further research should be directed towards discovering the prerequisites of unambiguous perception of multi-stream stimulus configurations.

Some of the participants reported switching between integrated and segregated sound organizations even during those stimulus blocks in which the 'A' and 'B' sounds were identical. This may be an auditory illusion. However, it is also possible that these participants implicitly tried to "please" the experimenter by assuming that they were expected to report perceptual switching in this experiment. However, despite this difference between the "faithfully reporting" and the "doubtfully reporting" participants, the two groups showed similar patterns of results. Only one interaction involving the Group variable reached significance. It remains to be clarified in further experiments whether "doubtfully reporting" participants are indeed influenced by implicit expectations. The current results suggest that, on the one hand, this issue should be kept in mind when designing future experiments, but on the other hand, mixing together "doubtfully reporting" and "faithfully reporting" participants does not qualitatively affect the results; it probably only increases the inter-subject variance.

CONCLUSIONS

We found that separation in the frequency of amplitude modulation acts as a primary cue in auditory streaming. It increases the strength of the segregated organization in the competition as well as the stability of this percept. This cue affects auditory streaming through partly overlapping mechanisms as separation in pitch and, possibly location.

ACKNOWLEDGEMENTS

This work was funded by EU FP7-ICT-231168-SCANDLE project: acoustic SCene ANalysis for Detecting Living Entities, and the Lendület project awarded to IW by the Hungarian Academy of Sciences (contract number LP2012-36/2012). The experiment was realized using Cogent 2000 developed by the Cogent 2000 team at the FIL and the ICN.

NOTE

¹ We are aware that the current data are not ideally suited for ANOVA type of statistical testing of continuous variables. This is mainly due to the limited observation period. When a given percept does not appear within the 4 minutes stimulus block, assuming that it would never be observed (as is done by filling empty observations with zeros) is not warranted. Similarly, the last perceptual phase of each stimulus block was cut short by the end of the sound sequence. Thus we cannot assess, how long it would have been, had the stimulus block continued. Finally, the phase-duration variable does not conform to the normal distribution. To ascertain that the ANOVA's did not yield substantially distorted results, we re-analyzed the data using more advanced statistical methods implemented in the "gamlss" and survival R-packages (Rigby and Stasinopoulos 2005, Therneau and Lumley 2011, respectively) and by which 1. there was no need for averaging, that is, all individual phases could be analyzed, 2. fixed and random (subject) effects were separated, 3. censoring was explicitly modeled, 4. various distributions besides the normal distribution could be fitted and 5. instead of p values the Bayesian information criterion (BIC) was applied for model selection. However, because these new statistical methods are not yet fully documented and they yielded consistent results with the ANOVA's, we did not wish to burden the ms with detailed descriptions of statistical methods and decided to present the results obtained with the traditional ANOVA method. As a further caution against spurious effects, we set the α level to 0.01 for the ANOVA tests.

REFERENCES

- Anstis, S., Saida, S. (1985): Adaptation to Auditory Streaming of Frequency-Modulated Tones. *Journal of Experimental Psychology – Human Perception and Performance*, 11, 257–271.
- Bendixen, A., Böhm, T. M., Szalárdy, O., Mill, R., Denham, S. L., Winkler, I. (2012): Different roles of similarity and predictability in auditory stream segregation. *Learning and Perception*, 4, 37–54.
- Bendixen, A., Denham, S. L., Gyimesi, K., Winkler, I. (2010): Regular patterns stabilize auditory streams. *Journal of the Acoustical Society of America*, 128, 3658–3666.
- Blake, R., Logothetis, N. K. (2002): Visual competition. *Nature Reviews Neuroscience*, 3, 13–23.
- Bregman, A. S. (1990): *Auditory Scene Analysis: The Perceptual Organization of Sound*. MIT Press, Cambridge, MA.
- Bregman, A. S., Abramson, J., Doehring, P., Darwin, C. J. (1985): Spectral integration based on common amplitude-modulation. *Perception & Psychophysics*, 37, 483–493.
- Ciocca, V. (2008): The auditory organization of complex sounds. *Frontiers in Bioscience*, 13, 148–169.
- Denham, S. L., Gyimesi, K., Stefanics, G., Winkler, I. (2012): Perceptual bistability in auditory streaming: How much do stimulus features matter? *Learning and Perception*, 4, 73–96.
- Denham, S. L., Gyimesi, K., Stefanics, G., Winkler, I. (2010): Stability of perceptual organisation in auditory streaming. In: Lopez-Poveda, E. A., Palmer, A. R., Meddis, R. (eds.): *The Neurophysiological Bases of Auditory Perception*. Springer, New York, pp. 477–487.
- Denham, S. L., Winkler, I. (2006): The role of predictive models in the formation of auditory streams. *Journal of Physiology, Paris*, 100, 154–70.
- Dollezal, L. V., Beutelmann, R., Klump, G. M. (2012): Stream segregation in the perception of sinusoidally amplitude-modulated tones. *Plos One*, 7, e43615.
- Du, Y., He, Y., Ross, B., Bardouille, T., Wu, X. H., Li, L. A., Alain, C. (2011): Human auditory cortex activity shows additive effects of spectral and spatial cues during speech segregation. *Cerebral Cortex*, 21, 698–707.
- Griffiths, T. D., Warren, J. D. (2004): What is an auditory object? *Nature Reviews Neuroscience*, 5, 887–892.
- Grimault, N., Bacon, S. P., Micheyl, C. (2002): Auditory stream segregation on the basis of amplitude-modulation rate. *Journal of the Acoustical Society of America*, 111, 1340–1348.
- Grimault, N., Micheyl, C., Carlyon, R. P., Arthaud, P., Collet, L. (2000): Influence of peripheral resolvability on the perceptual segregation of harmonic complex tones differing in fundamental frequency. *Journal of the Acoustical Society of America*, 108, 263–271.
- Hartmann, W. M., Johnson, D. (1991): Stream segregation and peripheral channeling. *Music Perception*, 9, 155–184.
- Leopold, D. A., Logothetis, N. K. (1999): Multistable phenomena: Changing views in perception. *Trends in Cognitive Sciences*, 3, 254–264.
- Moore, B. C. J., Gockel, H. (2002): Factors influencing sequential stream segregation. *Acta Acustica United with Acustica*, 88, 320–333.
- Moreno-Bote, R., Shpiro, A., Rinzel, J., Rubin, N. (2010): Alternation rate in perceptual bistability is maximal at and symmetric around equi-dominance. *Journal of Vision*, 10, 1–18.
- Pressnitzer, D., Hupe, J. M. (2006): Temporal dynamics of auditory and visual bistability reveal common principles of perceptual organization. *Current Biology*, 16, 1351–1357.

- Rigby, R. A., Stasinopoulos, D. M. (2005): Generalized additive models for location, scale and shape. *Applied Statistics*, 54, 507–554.
- Roberts, B., Glasberg, B. R., Moore, B. C. J. (2002): Primitive stream segregation of tone sequences without differences in fundamental frequency or passband. *Journal of the Acoustical Society of America*, 112, 2074–2085.
- Schulze, H., Langner, G. (1997): Periodicity coding in the primary auditory cortex of the Mongolian gerbil (*Meriones unguiculatus*): two different coding strategies for pitch and rhythm? *Journal of Comparative Physiology a-Neuroethology Sensory Neural and Behavioral Physiology*, 181, 651–663.
- Shamma, S. A., Elhilali, M., Micheyl, C. (2011): Temporal coherence and attention in auditory scene analysis. *Trends in Neurosciences*, 34, 114–123.
- Snyder, J. S., Alain, C. (2007): Toward a neurophysiological theory of auditory stream segregation. *Psychological Bulletin*, 133, 780–799.
- Therneau, T., Lumley, T. (2011): survival: Survival Analysis Including Penalised Likelihood. *R package version 2.36-3*, url: <http://CRAN.R-project.org/package=survival>.
- van Noorden, L. P. A. S. (1975): *Temporal coherence in the perception of tone sequences*. Doctoral dissertation, Technical University Eindhoven.
- Winkler, I., Denham, S. L., Nelken, I. (2009): Modeling the auditory scene: predictive regularity representations and perceptual objects. *Trends in Cognitive Sciences*, 13, 532–540.