Sound predictability as a higher-order cue in auditory scene analysis

Alexandra Bendixen[1], Susan L. Denham[2], István Winkler[3]

[1] *University of Leipzig, 04105 Leipzig, Germany, Email: alexandra.bendixen@uni-leipzig.de*
[2] *University of Plymouth, PL4 8AA Plymouth, UK, Email: sdenham@plymouth.ac.uk*
[3] *Hungarian Academy of Sciences, 1394 Budapest, Hungary, Email: iwinkler@cogpsyphy.hu*

Introduction

A major challenge for the auditory system is to disentangle signals emitted by two or more sound sources that are active in a temporally interleaved manner (*sequential stream segregation* [1]). Besides distinct characteristics of the individual signals (e.g., their timbre, location, and pitch), one important cue for distinguishing the sound sources is how their emitted signals unfold over time. It seems intuitively plausible that signals that unfold predictably with respect to their acoustic features and time-points of occurrence, such as the repetitive signature of a train moving on the rails, can be more readily identified as originating from one sound source. Based on this rationale, predictive elements have successfully been incorporated into computational models of auditory scene analysis for many years [2].

In contrast, empirical evidence for contributions of signal predictability to the decomposition of sound mixtures in human listeners has remained quite elusive. Some early studies have concluded that predictability does not affect sound source formation [3]; others have suggested that the effect of a sound source's predictability is confined to a late stage of auditory scene analysis and contingent upon the listener attending the predictable sound source [4]. Recently, driven by significant advances in knowledge on the auditory system's predictive processing capacities [5,6], experimental psychology has started to re-examine the role of predictability in auditory scene analysis.

Here we present two studies [7,8] demonstrating that a predictable arrangement of the emitted signals of a putative sound source automatically increases the tendency to isolate this source from a sound mixture. This predictability effect is not dependent on the listener attempting to hear out the predictable sound source(s). We applied a subjective-report procedure in which participants were asked to continuously indicate their momentary perception of a sound sequence. They were encouraged to maintain a neutral listening set, not attempting to hear the sequence in any particular way. The general configuration of the sequences followed a classical auditory streaming paradigm [9]. Unknown to the participants, predictable frequency and intensity patterns were hidden in some of the tone sequences, which were then contrasted with random arrangements of these tone features in other sequences.

Prolonged exposure to ambiguous tone sequences leads to perception switching back and forth between various alternatives of sound grouping [10]. Participants' perceptual reports can then be analysed in terms of the switching dynamics between the 'Integrated' (one-source) and 'Segregated' (two-sources) interpretations. This, in turn, allows for a specification of the mechanisms underlying the effects of a given cue (here: predictability) within auditory scene analysis (Figure 1).

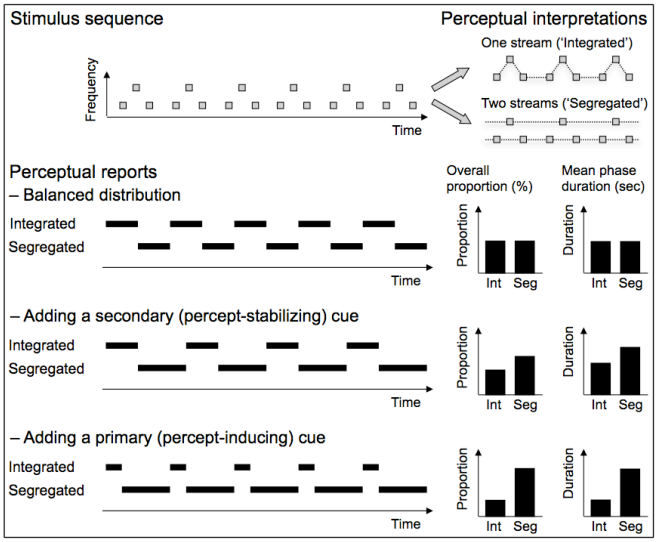

Figure 1: Top panel: Stimulus configuration (left) and typical perceptual interpretations (right) in the bi-stable auditory streaming paradigm [9]. Bottom panel: The effects of different types of cues on subjective perceptual reports. Simplified time-courses of perceptual switching were generated for the purpose of illustration. See the main text for distinguishing the effects of 'percept-inducing' and 'percept-stabilizing' cues.

As shown in Figure 1, a 'percept-stabilizing' cue prolongs the mean experienced duration of one of the perceptual alternatives (in this case, the 'Segregated' percept) but does not affect the mean duration of the other alternative. In contrast, a 'percept-inducing' (primary) cue prolongs the mean duration of one perceptual alternative and also shortens the duration of the other alternative (by causing perceptual switching back to the compatible percept). This implies that a percept-inducing cue contributes to the actual grouping of the auditory input and thus affects an early stage of auditory scene analysis. In contrast, a percept-stabilizing cue acts upon a later stage of auditory scene analysis, by providing differential support to the currently dominant grouping [1,10]. Note that the distinction between 'percept-inducing' and 'percept-stabilizing' cues is possible only via the analysis of the average perceptual phase durations, whereas the proportions of the different percepts are affected in a qualitatively similar way by both types of cues.

Experiment 1

Methods

Healthy adult participants listened to 'ABA_' sequences [9] with small amounts of variation in the frequency and

intensity values of the 'A' and 'B' sets of tones. The variation in the stimulus features was implemented randomly in some experimental conditions, and arranged in predictable patterns in some other conditions (Figure 2). Note that the range and the average amount of jitter was the same for both types of sequences, excluding simple effects of physical (dis)similarity to account for any observed condition differences.

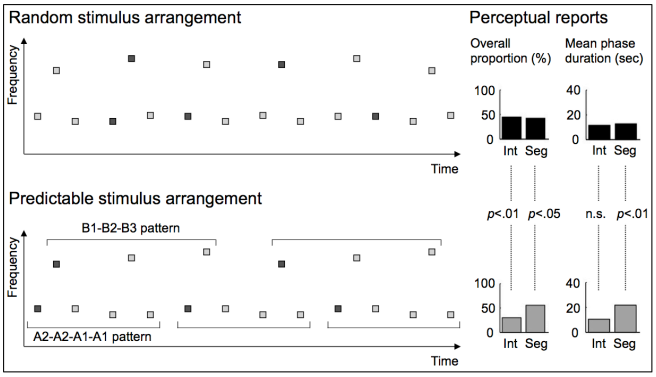

Figure 2: Example tone sequences used in Experiment 1 (left panel) and their effects on the perceptual reports of 'Integration' (*Int*) and 'Segregation' (*Seg*) (right panel) for unpredictable (upper row) and predictable (lower row) tone configurations. Vertical position of the rectangles indicates tone frequency; shading indicates intensity of the tones. For the predictable configuration, the cyclically repeating patterns in the 'A' and 'B' tone sets are marked.

The predictability manipulation was designed to selectively affect the 'Segregated' organization of the auditory input. This was achieved by inserting different predictable patterns into the two putative streams (i.e., separately into the 'A' and 'B' sets of tones). By necessity, this manipulation also changed the formal (i.e., mathematical) predictability of the 'Integrated' perceptual organization; however, one full cycle of the pattern in the 'Integrated' organization spanned 18 tones, which is beyond the capacity of auditory pattern extraction [11]. Therefore, the 'Integrated' organization can be considered as unpredictable from the auditory system's point of view in all stimulus conditions, and only the predictability of the 'Segregated' perceptual organization varies. Note that in addition to the conditions depicted in Figure 2, several intermediate conditions were employed in which only one of the tone sets or only one of the features was predictable; these are not reported here (for details, see [7]).

If predictability supports auditory scene analysis, this should be indicated by an increase in the proportion of perceptual reports of 'Segregation' and a corresponding decrease in the proportion of perceptual reports of 'Integration'. An analysis of the average phase durations will allow us to determine whether predictability acts as a 'percept-inducing' or a 'percept-stabilizing' cue.

Results

All statistical results are reported with *p* values corrected for multiple comparisons to account for statistical testing of several dependent variables. The effects of the predictability manipulation corresponded to the result pattern expected for a percept-stabilizing cue (cf. Figures 1, 2): The proportion of 'Segregated' percepts was higher for predictable than for unpredictable sequences [$t(25) = 3.527$, $p_{\text{corrected}} < .05$]; the proportion of 'Integrated' percepts was correspondingly lower [$t(25) = 4.433$, $p_{\text{corrected}} < .01$]. The change in proportions by predictability was brought about by a selective prolongation of the average phase duration of 'Segregated' percepts [$t(25) = 4.089$, $p_{\text{corrected}} < .01$], whereas the mean phase duration of 'Integrated' percepts was not affected [$t(25) = 0.760$, $p_{\text{corrected}} > .99$].

Discussion

The present data provide evidence that predictability affects auditory stream segregation: Inserting separate predictable patterns into two sets of tones increases the tendency to perceive the two sets as originating from two different sources. The demonstration that predictability is effective as a cue for auditory scene analysis even with neutral listening instructions (i.e., without voluntary effort to 'hear out' the predictable source) qualitatively differs from the findings of previous studies [3]. This was achieved by selectively manipulating the predictability of the 'Segregated' perceptual organization without the confounding effect of changing the predictability of the 'Integrated' perceptual organization in parallel.

The effect of predictability was brought about by selectively prolonging the duration of experiencing 'Segregation'. In contrast, predictability had no effect while participants were experiencing 'Integration'. Predictability thus shows the characteristics of a 'percept-stabilizing' cue. Within a two-stage model of auditory scene analysis [1], these results suggest that predictability does not affect the first stage during which the auditory input is decomposed into groups of sounds, but only the second stage during which the sound groups are evaluated. Predictability can be conceptualized as giving more or less support to the sound configurations provided by the first stage depending on how successfully they predict incoming sounds [10].

If this view holds, the initial grouping (i.e., the first stage of auditory scene analysis) would be affected mainly by the so-called primary cues [1] such as spectral separation and other acoustic differences between the tone sets. Predictability would exert its influence only after the primary cues had been considered. In order to substantiate this conclusion on the time-course of the different auditory cues, a further study was designed to manipulate predictability and a primary grouping cue (spectral separation, i.e., dissimilarity of the two tone sets) independently within the same experiment.

Experiment 2

Methods

Healthy adult participants listened to tone sequences that were presented in a two-factorial design with the factors predictability (2 levels: predictable vs. unpredictable) and spectral separation (2 levels: low vs. high). The predictability manipulation was the same as applied in Experiment 1 (cf. Figure 2). The two levels of the spectral separation were 5 and 7 semitones mean difference between the 'A' and 'B' sets of sounds. In some experimental

conditions, predictability or spectral separation were altered after the first half of a stimulus block to investigate how dynamic changes in the two cue types affect perceptual reports; only data prior to these changes are reported here (for details, see [8]).

As in Experiment 1, participants were asked to continuously report their current perception of the sequence with neutral listening instructions, refraining from the attempt to hear the sequence in any particular manner. Their responses were recorded and analysed in terms of the proportion and mean duration of 'Integrated' and 'Segregated' perceptual phases.

Results

Again, statistical results are reported with p values corrected for multiple comparisons to account for statistical testing of multiple dependent variables. The effect of the predictability manipulation (cf. Figure 3) replicates the pattern observed in Experiment 1: The proportion of 'Segregated' percepts was higher for predictable than for unpredictable sequences [$F(1,29) = 10.195$, $p_{corrected} < .05$]; the proportion of 'Integrated' percepts was correspondingly lower [$F(1,29) = 19.311$, $p_{corrected} < .001$]. The effect of predictability on the proportions was due to prolonging the average phase duration of 'Segregated' percepts [$F(1,29) = 12.285$, $p_{corrected} < .05$], while the average phase duration of 'Integrated' percepts was not affected [$F(1,29) = 2.232$, $p_{corrected} > .99$].

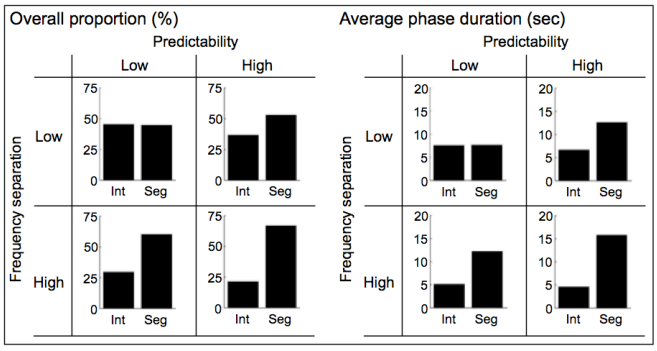

Figure 3: Results of Experiment 2: Effects of manipulating predictability (predictable vs. unpredictable arrangement of the tones within each set) and spectral separation (high vs. low frequency difference between the two tone sets) on the perceptual reports of 'Integration' (*Int*) and 'Segregation' (*Seg*).

Increasing the spectral separation between the tone sets likewise increased the proportion of 'Segregated' perceptual reports [$F(1,29) = 44.553$, $p_{corrected} < .001$] and decreased the proportion of 'Integrated' perceptual reports [$F(1,29) = 55.296$, $p_{corrected} < .001$]. Unlike for the predictability manipulation, the effect of increased spectral separation on the overall proportions was brought about not only by prolonging the average duration of 'Segregated' phases [$F(1,29) = 23.209$, $p_{corrected} < .001$], but also by shortening the average duration of 'Integrated' phases [$F(1,29) = 28.664$, $p_{corrected} < .001$].

The effects of the two cues on the proportions and average phase durations of 'Segregated' and 'Integrated' percepts did not significantly interact with each other [all F values < 2.5, all corrected p values > .999].

Discussion

As observed in Experiment 1, predictability acted to stabilize source segregation whenever a 'Segregated' percept emerged. In contrast, spectral separation not only stabilized the 'Segregated' perceptual organization, but also caused switching towards that organization, thereby cutting short the 'Integrated' perceptual phases. Thus spectral separation shows the specified properties of a cue that not only stabilizes but also induces grouping. This constitutes not just a quantitative but a qualitative difference in the way the two cue types exert their influence on auditory scene analysis.

Additional support for independent processing of the cues is provided by the absence of statistical interactions: The effects of predictability and spectral separation were fully additive for each of the dependent measures. Altogether, the present data support the conclusion that the two types of cues act upon temporally and functionally different stages within auditory scene analysis. Whereas spectral separation acts as an 'early' (primary [1]) cue contributing to the initial decomposition of the auditory input, predictability can be considered a 'late' (higher-order) cue involved in evaluating the decompositions derived during the first stage. This evaluation might be conceptualized as a feedback loop, determining whether the way the sounds were grouped at an earlier stage should be maintained or not.

General Discussion

Taken together, results of Experiments 1 and 2 suggest that predictability is automatically (i.e., without conscious effort) taken into account as a cue in auditory scene analysis by human listeners. Moreover, predictability exerts its influence only after primary acoustic grouping cues have been considered. This temporal dissociation can inform computational models of auditory scene analysis aimed at mimicking how human perception solves the source separation problem.

Consistent evidence for a role of predictability in auditory stream segregation has recently been obtained in objective-listening paradigms [12,13]. It has also been suggested that the beneficial effects of predictability for solving the source segregation problem may show an age-related decline [13]. This calls for further investigations into the underlying mechanisms and into the possible reasons for their age-related impairment. Such investigations should include the attempt to establish links with the physiological processes known to underlie predictive processing [5,6,10] in order to eventually shape predictability-based computational models of auditory scene analysis in a biologically plausible way.

References

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

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

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

[4] Jones, M.R., & Boltz, M.: Dynamic attending and responses to time. Psychological Review **96** (1989), 459-491

[5] Baldeweg, T.: Repetition effects to sounds: Evidence for predictive coding in the auditory system. Trends in Cognitive Sciences **10** (2006), 93-94

[6] Friston, K.: A theory of cortical responses. Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences **360** (2005), 815-836

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

[8] Bendixen, A., Bőhm, T.M., Szalárdy, O., Mill, R., Denham, S.L., & Winkler, I.: Different roles of similarity and predictability in auditory stream segregation. Learning and Perception (in press)

[9] van Noorden, L.P.A.S.: Temporal coherence in the perception of sound sequences. Technical University Eindhoven, The Netherlands, 1975

[10] Winkler, I., Denham, S.L., Mill, R., Bőhm, T.M., & Bendixen, A.: Multistability in auditory stream segregation: A predictive coding view. Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences **367** (2012), 1001-1012

[11] Boh, B., Herholz, S.,C., & Pantev, C.: Processing of complex auditory patterns in musicians and nonmusicians. PLoS One **6** (2011), e21458

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

[13] Rimmele, J.M., Schröger, E., & Bendixen, A.: Age-related changes in the use of regular patterns for auditory scene analysis. Hearing Research **289** (2012), 98-107