Budapest University of Technology and Economics

PhD School in Psychology – Cognitive Science

Orsolya Kisné Szalárdy

PERCEPT-INDUCING AND PERCEPT-STABILIZING CUES OF AUDITORY STREAM SEGREGATION

PhD thesis booklet

Supervisor:
Prof. István Winkler

Budapest, 2016

Introduction

In everyday life, the auditory system is usually confronted with several acoustic events produced by different sound sources operating concurrently. Imagine a noisy street you are walking on. The street noise consists of many different sounds produced by several sources at the same time, such as the noise of the cars, the footsteps of the people, the ringing of cellphones, human speech, the sound of the wind blowing, etc. Separating the sound of an approaching car from the other sound sources can be crucial for avoiding an accident. In other words, in order to interact with our environment, we need to organize the acoustic events into meaningful streams, which typically represent the available sound sources (Bregman, 1990). The formation of sound streams (termed *auditory streams*) has been investigated within the framework of *auditory scene analysis* (Bregman, 1990; for recent reviews see Snyder & Alain, 2007; Winkler et al., 2009; Gutschalk & Dykstra, 2014).

The formation of auditory streams is often investigated in the classical *auditory streaming paradigm*, a stimulus configuration that consist of a repeating 'ABA_' pattern of sounds (van Noorden, 1975), where 'A' and 'B' denote two sounds differing in some feature(s) and '_' stands for a silent interval equaling the common duration of the two sounds. Depending on the stimulus parameters, this stimulus configuration is usually perceived in one of two different ways: Either all tones form a single sound stream termed the *integrated percept* or two streams are heard concurrently, one consisting only of the 'A' and the other only of the 'B' sounds (the *segregated percept*). Several studies have shown that the auditory system utilizes many different cues for segregating sound streams (Moore & Gockel, 2002; Moore & Gockel, 2012). According to Bregman's theory (1990), there are two stages of auditory stream segregation. Initially, alternative sound groupings (proto-objects: groupings of sounds that may appear in perception) are formed by linking together sounds based on similarity in various features (such as pitch, temporal proximity, source location, etc.; also

termed as similarity-based cues). In the second stage, the alternative sound groupings compete with each other and the strongest one becomes the dominant percept, while compatible other groupings form the background. Foreground and background together provide a possible full description of the auditory scene, a *sound organization*. Whereas the classical view was that perception settles on one of the alternative sound organizations after a few seconds after the beginning of the sequence (the "buildup" of streams), some studies showed that for longer sound sequences (> 1 minute) perception switches back and forth between the alternative interpretations (Anstis & Saida, 1985; Denham & Winkler, 2006; Pressnitzer & Hupé, 2006). This phenomenon is termed *perceptual bi-/multistability* (Blake & Logothetis, 2002; Leopold & Logothetis, 1999; for a review, see Schwartz et al., 2012). In order to explain the multistability discovered for auditory stream segregation, Denham and Winkler (2006) suggested that perceptual multistability reflects continuous competition between the alternative sound organizations. Further, Bendixen and colleagues (2010; Bendixen et al., 2013) showed that beyond similarity-based cues, auditory stream segregation can also be supported by temporal regularities. However, similarity and regularity affect stream segregation in different ways. Directly comparing the effects of feature similarity and temporal regularities Bendixen and colleagues (2013) found that for multistable stimulus configurations, cues which can induce a percept (termed *percept-inducing cues*) facilitated switching from another percept toward the particular percept, thus reducing the length of the time intervals during which other percepts are heard. In contrast, cues which can only stabilize a percept (termed *percept-stabilizing cues*) extend the time interval during which this percept is heard, but they do not affect the intervals during which some other percept is heard. So far, all percept-inducing cues found were also percept-stabilizing cues, whereas the percept-stabilizing cues are not necessarily also percept-inducing ones.

More than 20 years after Bregman's (1990) influential work, understanding auditory scene analysis still poses several questions for research (for reviews, see Carlyon, 2004; Denham & Winkler, 2006; Haykin & Chen, 2005; Snyder & Alain, 2007; Winkler et al., 2012). Perceptual bi-/multistability provides an important tool for investigating stream segregation as the mental representation of the acoustic scene can be tested without changing the physical parameters of the stimuli. The main objectives of the thesis are to shed light on two topics using the phenomena of auditory multistability: The first goal is to test different cues of auditory stream segregation regarding their ability to induce the formation of and/or to stabilize sound streams. The second goal is to extend our knowledge about electrophysiological correlates of the perceived sound organization.

Within the first topic, we investigated the effects of amplitude modulation and its interactions between multiple percept-inducing cues delivering multistable stimulus configurations. Grimault et al. (2002) found that when they increased the difference in the rate of amplitude modulation (AM) between two sets of interleaved tones, participants were more likely to perceive the sequence as two streams. Using the auditory streaming paradigm we tested the effects of AM on the proportion and phase duration of the alternative percepts. (The term *phase* is used for the continuous time interval during which the same percept is heard, phase duration then refers to the length of this interval.) As AM is a time-varying cue, its effect may be independent from static cues such as frequency or perceived location difference. However AM influences the perceived pitch (Meddis & Hewitt, 1991) therefore its effect may be overlapping with carrier frequency difference. Thesis I presents the conclusions from our study in which we have investigated the effects and interactions between separations in amplitude modulation, pitch, and perceived location.

The second question we asked was whether initial stream segregation is driven solely by similarity-based cues or whether higher-order cues can also induce stream segregation. For

instance, Devergie et al. (2010) demonstrated that participants separated streams in the absence of similarity-based cues when familiar sound patterns were present in the streams to be separated. Recent studies demonstrated that predictable sound patterns have an important role in stream segregation (Andreou et al., 2011; Rimmele et al., 2012; Snyder & Weintraub, 2011; Bendixen et al., 2013; Bendixen et al., 2010). However, these cues were more likely to stabilize sound streams rather than initiating their formation (Bendixen et al., 2013; Bendixen et al., 2010). Thesis II summarizes our results investigating the effects of melody, rhythm, and familiarity with the melody on auditory stream segregation.

By applying electrophysiological methods, sources of information inaccessible to behavioral methods can be tapped into. Mismatch negativity (MMN) component of the event-related brain potentials (ERP) is regarded as the electrophysiological correlate of detecting the violation of a regularity extracted from the preceding sound sequence by the auditory system (Winkler, 2007). MMN paradigms have been successfully used to investigate auditory stream segregation (Rahne et al., 2007; Sussman et al., 1998, 1999; Winkler et al., 2003a; Winkler et al., 2003b; Winkler et al., 2003c; Winkler et al., 2005). However, the relationship between auditory stream segregation and the deviance-detection process reflected by MMN is indirect (Schröger et al., 2014; Winkler et al., 2009) and even the relationship between MMN elicitation and perception is not entirely straightforward (Ross et al., 1996; Paavilainen et al., 2007; van Zuijen et al., 2006). Percept-dependent differences related to auditory stream segregation have also been found for the magnetic P1 and N1 responses. Thesis III summarizes our findings regarding the percept-dependent processing of regular and deviant sounds in the classical auditory streaming paradigm, while Thesis IV introduces a new multistable auditory stimulus paradigm permitting the investigation of auditory foreground/background decomposition by means of ERPs and the investigation of percept-dependent MMN elicitation.

New scientific results

Thesis I. The effects of separation in amplitude modulation frequency on auditory stream segregation

The aim of the study was to test the effect of the AM difference on auditory stream segregation in the auditory streaming paradigm and its interactions with carrier frequency difference and perceived location. We continuous recorded the listener's perception during the stimulus blocks. The results showed that larger separation in any of the tested stream-segregation cues have cut the integrated phases shorter and extended the duration of the segregated phases. Furthermore, we found interactions between the AM rate difference and carrier frequency difference, AM rate difference and location difference, and carrier frequency difference and location difference, suggesting a ceiling effect in stream segregation. These results suggest that similarity/separation in any of these features is utilized as a percept-inducing cue.

Thesis II. The effects of sequence-structure cues on auditory stream segregation

In this study, we aimed at testing the effects of familiar and unfamiliar rhythms and melodies on stream segregation. Specifically, we investigated whether or not the presence of tunes separately in two interleaved tone sequences helps to segregate them into separate streams. The tunes were pre-selected on the basis of their pitch range and mean frequency similarity to avoid stream segregation on the basis of pitch difference. Two Hungarian and two German songs were used. A group of Hungarian and another of English volunteers participated in the experiments: Hungarian participants were familiar with the Hungarian but not with the German songs whereas English participants were not familiar with any of the songs. None of the participants had received formal musical training. The songs were

presented either with their original rhythm, or with an even rhythm (i.e., each tone and the following rest had constant durations). We found that the presence of melody helped segregating the sound sequences irrespective of whether participants were familiar with the tunes or not, but more so for familiar tunes. Rhythm also strongly promoted stream segregation regardless of the familiarity. However, the latter was caused by overlaps between the tones of the two interleaved sequences, rather than the temporal structure, *per se*. These results confirm that higher-order cues can induce stream segregation in some particular cases.

Thesis III. Short- and longer-latency ERP correlates of perceived sound organization in the auditory streaming paradigm

The main objective of the study was to test whether the currently perceived sound organization affects the processing of regular and pitch-deviant sounds. For testing this question the auditory streaming paradigm was used with continuous recording of both the listener's perception and EEG. Occasionally, standard sounds were exchanged for deviants that were slightly or moderately different from the standard sounds. We found that the ERPs elicited by regular and irregular (deviant) sounds were affected as early as ca. 75 ms from stimulus onset (the P1 component), at the time of feature extraction. Further modulations were found for deviant tones: deviant sound processing was affected by the current percept, as was reflected by the N2 and P3a ERP components. This suggests that sound evaluation is affected by the current interpretation of the auditory scene. Taken together, these results indicate that there are multiple interactions between the dominant sound organization and the processing of incoming sounds.

Thesis IV. A new paradigm for measuring the ERP correlates of auditory multistable perception

The aim of the study was twofold: using a new auditory multistable stimulus paradigm, we investigated the foreground-background discrimination by ERPs and tested whether or not MMN elicitation dynamically follows the perceptual organization of sounds. We have designed our paradigm based on Wessel's (1979) repeating three-tone patterns of rising pitch combined with alternation in timbre. Such sequences can be perceived as rising (based on temporal proximity) or as falling (based on timbre similarity). We continuously recorded the participants' percept during the experiment. Thus the current perceptual foreground was determined behaviorally. For testing whether MMN elicitation follows the perceived sound organization, two additional feature regularities were introduced, one of which was expected to be discovered only when the percept was that of the rising pitch, and the other only when the percept was that of the falling pitch. If the perceptual groupings of these sound sequences affected the representations underlying MMN, then each feature-regularity violation should elicit MMN only when the listener experiences the corresponding percept. Our results showed that in this paradigm, MMN elicitation did not follow the perceptual organization experienced by the listener. Rather, it followed the temporal arrangement of the tones: One MMN was elicited with both percepts, the other with neither percept. We suggest that in this case, the feature regularities had been extracted before perceptual grouping occurred. However, we found that the P1 amplitude correlated with one aspect of the perceptual organization: whether the sound was part of the foreground or the background. This result confirmed that obtained in Thesis III: the perceived sound organization influences the processing of sounds as early as 75 ms from stimulus onset.

Conclusions

In conclusion, we found that both spectral and temporal-structure-based cues can induce auditory stream segregation. The fact that unfamiliar melodies induced stream segregation brings up the possibility that tonality analysis may influence auditory stream segregation. Determining the nature of these processes requires further studies. Our results showed multiple interactions between the processing of incoming sounds and the currently dominant representation of the sound sequence. Early effects of percept-dependent modulation of the ERP components have been found on the P1 and N1 elicited by regular sounds. The processing of deviance was affected in the later, evaluation stage, as reflected by modulations of the N2 and P3a ERP components. We have also tested a new auditory multistability paradigm for investigating the foreground-background decomposition of auditory scenes. With this paradigm, percept-dependent modulation of the P1 was obtained, yet MMN elicitation was independent of the percept. Our results for the P1 suggest that enhancement of this component may reflect inhibition of the background tones. Future studies should be conducted to clarify the relationship between the representations underlying the MMN ERP component and those involved in conscious perception.

List of publications related to the thesis

I. Szalárdy, O., Bendixen, A., Tóth, D., Denham, S. L., Winkler, I. (**2013**). Modulation-frequency acts as a primary cue for auditory stream segregation. *Learning & Perception*, 5(2): 149-161.

II. Szalárdy, O., Bendixen, A., Bőhm, T. M., Davies, L. A., Denham, S. L., Winkler, I. (**2014**). The effects of rhythm and melody on auditory stream segregation. *Journal of the Acoustical Society of America*, 135: 1392-1405.

III. Szalárdy, O., Bőhm, T. M., Bendixen, A., Winkler, I. (**2013**). Event-related potential correlates of sound organization: early sensory and late cognitive effects. *Biological Psychology*, 93: 97-104.

IV. Szalárdy, O., Winkler, I., Schröger, E., Widmann, A., Bendixen, A. (**2013**). Foreground-background discrimination indicated by event-related brain potentials in a new auditory multistability paradigm. *Psychophysiology*, 50(12): 1239-1250.

References

Andreou, L. V., Kashino, M., & Chait, M. (2011). The role of temporal regularity in auditory segregation. *Hear. Res., 280*(1-2), 228-235.

Anstis, S., & Saida, S. (1985). Adaptation to auditory streaming of frequency-modulated tones. *Journal of Experimental Psychology: Human Perception and Performance, 11*(3), 257-271.

Bendixen, A., Bőhm, T., Szalárdy, O., Mill, R., Denham, S. L., & Winkler, I. (2013). Different roles of similarity and predictability in auditory stream segregation. *Learning and Perception, 5(2)*, 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*(6), 3658-3666.

Blake, R., & Logothetis, N. K. (2002). Visual competition. *Nature Reviews Neuroscience, 3*(1), 13-23.

Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge, MA: MIT Press.

Carlyon, R. P. (2004). How the brain separates sounds. *Trends in Cognitive Sciences, 8*(10), 465-471.

Denham, S. L., & Winkler, I. (2006). The role of predictive models in the formation of auditory streams. *Journal of Physiology, Paris, 100*(1-3), 154-170.

Devergie, A., Grimault, N., Tillmann, B., & Berthommier, F. (2010). Effect of rhythmic attention on the segregation of interleaved melodies. *Journal of the Acoustical Society of America, 128*(1), 1-7.

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*(3), 1340-1348.

Gutschalk, A., & Dykstra, A. R. (2014). Functional imaging of auditory scene analysis. *Hearing Research, 307*, 98-110.

Haykin, S., & Chen, Z. (2005). The cocktail party problem. *Neural Comput, 17*(9), 1875-1902.

Klimesch, W. (2011). Evoked alpha and early access to the knowledge system: the P1 inhibition timing hypothesis. *Brain Research, 1408*, 52-71.

Leopold, D. A., & Logothetis, N. K. (1999). Multistable phenomena: Changing views in perception. *Trends in Cognitive Sciences, 3*(7), 254-264.

Meddis, R., & Hewitt, M. J. (1991). Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: Pitch identification. *Journal of the Acoustical Society of America, 89*(6), 2866-2882.

Moore, B. C. J., & Gockel, H. (2002). Factors influencing sequential stream segregation. *Acta Acustica United with Acustica, 88*(3), 320-333.

Moore, B. C. J., & Gockel, H. E. (2012). Properties of auditory stream formation. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 367*(1591), 919-931.

Paavilainen, P., Arajärvi, P., & Takegata, R. (2007). Preattentive detection of nonsalient contingencies between auditory features. *Neuroreport, 18*(2), 159-163.

Pressnitzer, D., & Hupé, J.-M. (2006). Temporal dynamics of auditory and visual bistability reveal common principles of perceptual organization. *Current Biology, 16*(13), 1351-1357.

Rahne, T., Böckmann, M., von Specht, H., & Sussman, E. S. (2007). Visual cues can modulate integration and segregation of objects in auditory scene analysis. *Brain Research, 1144*, 127-135.

Rimmele, J., Schröger, E., & Bendixen, A. (2012). Age-related changes in the use of regular patterns for auditory scene analysis. *Hearing Research, 289*(1-2), 98-107.

Ross, J., Tervaniemi, M., & Näätänen, R. (1996). Neural mechanisms of the octave illusion: electrophysiological evidence for central origin. *Neuroreport, 8*(1), 303-306.

Schröger, E., Bendixen, A., Denham, S. L., Mill, R. W., Bőhm, T. M., & Winkler, I. (2014). Predictive regularity representations in violation detection and auditory stream segregation: from conceptual to computational models. *Brain Topography, 27*(4), 565-577.

Schwartz, J. L., Grimault, N., Hupé, J. M., Moore, B. C., & Pressnitzer, D. (2012). Multistability in perception: binding sensory modalities, an overview. *Philosophical Transactions of the Royal Society London B: Biological Sciences, 367*(1591), 896-905.

Snyder, J. S., & Alain, C. (2007). Toward a neurophysiological theory of auditory stream segregation. *Psychological Bulletin, 133*(5), 780-799.

Snyder, J. S., & Weintraub, D. M. (2011). Pattern specificity in the effect of prior Δf on auditory stream segregation. *Journal of Experimental Psychology- Human Perception and Performance, 37*(5), 1649-1656.

Sussman, E., Ritter, W., & Vaughan, H. G. (1999). An investigation of the auditory streaming effect using event-related brain potentials. *Psychophysiology, 36*(1), 22-34.

Sussman, E., Ritter, W., & Vaughan, H. G., Jr. (1998). Attention affects the organization of auditory input associated with the mismatch negativity system. *Brain Research, 789*(1), 130-138.

van Noorden, L. P. A. S. (1975). *Temporal coherence in the perception of tone sequences.* Unpublished Doctoral dissertation, Technical University Eindhoven.

van Zuijen, T. L., Simoens, V. L., Paavilainen, P., Näätänen, R., & Tervaniemi, M. (2006). Implicit, intuitive, and explicit knowledge of abstract regularities in a sound sequence: an event-related brain potential study. *Journal of Cognitive Neuroscience, 18*(8), 1292-1303.

Wessel, D. L. (1979). Timbre space as a musical control structure. *Computer Music Journal, 3*, 45-52.

Winkler, I. (2007). Interpreting the mismatch negativity. *Journal of Psychophysiology, 21*(3-4), 147-163.

Winkler, I., Denham, S., Mill, R., Bőhm, T. M., & Bendixen, A. (2012). Multistability in auditory stream segregation: a predictive coding view. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 367*(1591), 1001-1012.

Winkler, I., Denham, S. L., & Nelken, I. (2009). Modeling the auditory scene: predictive regularity representations and perceptual objects. *Trends in Cognitive Sciences, 13*(12), 532-540.

Winkler, I., Kushnerenko, E., Horváth, J., Čeponienė, R., Fellman, V., Huotilainen, M., et al. (2003a). Newborn infants can organize the auditory world. *Proceedings of the National Academy of Sciences of the United States of America, 100*(20), 11812-11815.

Winkler, I., Sussman, E., Tervaniemi, M., Horváth, J., Ritter, W., & Näätänen, R. (2003b). Preattentive auditory context effects. *Cognitive, Affective, and Behavioral Neuroscience, 3*(1), 57-77.

Winkler, I., Takegata, R., & Sussman, E. (2005). Event-related brain potentials reveal multiple stages in the perceptual organization of sound. *Cognitive Brain Research, 25*(1), 291-299.

Winkler, I., Teder-Sälejärvi, W. A., Horvath, J., Näätänen, R., & Sussman, E. (2003c). Human auditory cortex tracks task-irrelevant sound sources. *Neuroreport, 14*(16), 2053-2056.