Budapest University of Technology and Economics

PhD School in Psychology – Cognitive Science

MŰEGYETEM 1782

Zsuzsanna Kocsis

EEG signatures of separating concurrent sounds

PhD thesis booklet

Supervisor:

Dr. István Winkler

Budapest, 2017

General background and aim of the thesis

The main aim of this thesis was to investigate how the human brain separates concurrent sounds or a figure from the background and what brain correlates can be found for this process (Bregman, 1990; Ciocca, 2008). Concurrent sound segregation is a basic grouping process of auditory scene analysis, yet many features of it are not well understood. Two ERP components were found to reflect processes related to concurrent sound segregation: the object-related negativity (ORN; Alain et al., 2001) is elicited by the presence of more than one concurrent sound; when the sound manipulations that elicit the ORN are task-relevant, a later positive component, the P400 is also elicited. To gain better understanding of the concurrent sound segregation, a more ecologically valid paradigm was examined for separating auditory objects and the combinations of simple cues of concurrent sound segregation were tested. Specific goals of the thesis were to shed light on the following questions: First, what are the even-related potential (ERP) correlates of separating a figure from the background when segregation requires integration over both time and spectral components? Is the ORN elicited in this case? Thus *Thesis I* focused on the generality of the ORN response. This thesis point summarizes the electrophysiological effects of relevant parameters of stochastic figure-ground segregation. Second, how does the auditory system utilize different concurrent grouping cues? How do these cues interact with each other in sound grouping? What does the ERP correlate of concurrent sound segregation (the ORN) reflect in terms of cue evaluation: separate assessment of the cues or an integrated process? *Thesis II* presents the conclusions from our study regarding these questions. The rest of the studies utilized the ORN for investigating specific questions regarding concurrent sound segregation using the ORN response and brain oscillatory responses accompanying the ORN. Third, what types of EEG oscillations can be observed during concurrent sound segregation?

Thesis III summarizes our results regarding the neural oscillations accompanying concurrent sound segregation. Finally, we asked about the EEG correlates of concurrent sound segregation when multiple convergent or divergent cues were employed. *Thesis IV* introduces a paradigm in which the perception of two or three concurrent sounds was promoted by manipulating concurrent grouping cues either convergently or divergently. The results of the ERP study designed to test the paradigm are summarized in this thesis point.

Overall we sought to answer whether the ORN reflects a global readout of the available cues in the auditory system, or it shows the contributions to concurrent sound segregation from each cue.

New scientific results

Thesis I: Auditory figure-ground segregation

In everyday acoustic scenes, the acoustic signals comprising different objects often overlap in time as well as in the spectrum. Therefore to segregate these auditory objects and select one to be the "figure" over the background summing the signals of other auditory objects, one is typically required to group together sound elements over both time and frequency. Event-related brain potential (ERP) correlates of simultaneous and sequential grouping have been mostly studied separately until now. Our aim was to investigate the ERP responses emerging in relatively natural acoustic contexts in which a combination of both types of grouping processes is required for veridical perception. The salience of the figure was varied systematically by independently manipulating sequential and simultaneous cues supporting figure detection. This design allowed us to determine the electrophysiological correlates of the emergence of an auditory object from a stochastic background.

We found that correct identification of the figure led to the elicitation of an ORN and a P400 component providing evidence that these components are also observed when concurrent sound segregation requires the integration of spectral cues over time. Furthermore, we found that P400 correlated with detection performance. Moreover, the source localization results showed that both the ORN and the P400 had generators in the temporal cortex, which was also in line with previous results (c.f., Alain and McDonald, 2007).

Thesis II: Effects of multiple congruent cues on concurrent sound segregation

Concurrent sounds segregation is based on instantaneously available cues, such as difference in pitch, sound onset, source location, or some combination of the above. Most cues have been studied alone, or combined with one other cue, but no systematic study of these cues has been conducted in both active and passive listening conditions.

Here we investigated the effects of combining different cues on the ORN and P400 ERP components. Participants were presented with complex tone sequences with half of the tones manipulated: one or two harmonic partials were mistuned, delayed or presented from a different source location than the rest. In separate stimulus blocks, one, two or three of these manipulations were combined. Participants either watched a silent, subtitled movie, or were instructed to respond to each tone by pressing down one of two response buttons indicating whether they perceived one or two concurrent sounds. We tested whether the salience of the harmonicity-based cue is increased by mistuning two partials in a congruent manner. Further, we tested whether the effects of combined cues are additive, sub- or superadditive compared to the single-cue effects. If each cue elicits a separate ORN response, the ORN elicited by multiple congruent cues will be as large as the summed amplitudes of the ORN components elicited by the contributing cues. This would suggest that ORN reflects processes that are closely related to cue evaluation and so the processes underlying ORN are farther upstream from what appears in perception. Alternatively, ORN may reflect the system's overall assessment of the likelihood that the auditory input consists of two concurrent sounds. That is, ORN could reflect the readout of a process combining the evidence from the available cues. In this case, we should find sub- or superadditivity between the contributing cues' ORN components.

We found that different cues are not equally effective in eliciting the ORN (the location cue appeared to be weaker than the others tested in the study) and that the ORN responses are not additive when two or more concurrent sound segregation cues are combined. The latter result indicates that the ORN component does not reflect the evaluation of individual sensory cues. The P400 component was also observed in most but not all conditions during active listening, and it was less sensitive to sound segregation cues and their combinations than the ORN.

Thesis III: Theta oscillations accompanying concurrent sound segregation

This study was aimed at assessing the large scale brain oscillations associated with concurrent sound segregation as there have been few previous reports of oscillatory networks during concurrent sound segregation. The complex tones could be perceived as a single sound or as two concurrent sounds based on differences in the harmonic template, sound onset, or sound source location. In separate task conditions, participants performed a visual change detection task (visual control), watched a silent movie (passive listening) or reported for each tone whether they perceived one or two concurrent sounds (active listening). Our goal was to determine the neural oscillatory correlates of concurrent sound segregation by comparing the oscillatory activity between concurrent sound segregation supported by different cues and between different attentional conditions.

We found increased theta activity in two intervals in response to simultaneous sound segregation cues – in an early and in a late time window. These time windows corresponded to the time windows of the ORN and P400. Event-related spectral perturbations were larger in active than in passive listening. Our results suggest that the neural networks underlying the generation of the ERP components ORN and P400 may communicate via theta rhythms.

Thesis IV: Two and three concurrent sound objects

No previous study tested the segregation of three concurrent sounds. Complex sounds containing sound segregation cues and cue combinations were set up to promote different separations of the tonal elements into one, two or three concurrent sounds.

Thus here we examined the effects of divergent cues on perceptual and neural (ORN) indicators of concurrent sound segregation. If ORN elicited by the divergent manipulations (three-objects conditions) is as large as the summed amplitudes of the ORNs elicited separately by the individual cues, this would be consistent with the interpretation that ORN reflects the independent evaluation of the divergent cues of concurrent sound segregation (i.e., more closely related to the evaluation of cues). Alternatively, if ORN generally shows sub-additivity relative to the ORN amplitudes elicited separately by the individual cues, this would support the interpretation that ORN reflects the auditory system's overall readout of the presence of multiple auditory objects, regardless of the congruency of the manipulations or the number of objects and different cues. A further question was whether the ORN amplitude reflects the actual number of the auditory objects present in the auditory scene or it only indicates the presence of more than one simultaneous objects.

In this study, we found that it is possible to promote the perception of three concurrent sounds. However, listeners cannot reliably tell apart two- vs. three concurrent sound objects. Nevertheless, we found that ORN was elicited in almost all conditions in the ERP study. Sounds with AM or FM difference between the manipulated and the base harmonics yielded the smallest ORN amplitudes. When two partials were divergently manipulated, numerically larger ORN amplitudes were elicited as compared to the convergently manipulated tones. Based on the results, the ORN reflects the overall readout of the auditory system regarding the presence of multiple concurrent objects, and not the actual number of the object present at any moment.

Conclusions and further directions

The results show that the ORN event-related potential component reflects the emergence of an auditory object from the background in a stochastic figure-ground segregation paradigm. Segregation of simultaneous sound objects is also indexed by the ORN when single or multiple cues are used on complex tones. ORN is only followed by the P400 component in situations in which the sounds are task-relevant. ORN peaks around 150-250 ms after stimulus onset reflecting an earlier, attention-independent processing stage, whereas P400 peaks after 400 ms indexing a later, top-down controlled processing stage.

The picture emerging from studies II and IV is that ORN reflects the overall readout of the auditory system regarding the presence of multiple concurrent objects. P400 could then represent the outcome of the perceptual decision. Thus, whereas ORN reflects a bottom-up primitive grouping mechanism, P400 likely reflects a process which incorporates top-down effects on perceptual decisions.

The properties of ORN and P400 are compatible with Bregman’s (1990) two-stage theoretical framework. In the first stage of processing, in which the acoustic input is decomposed into putative perceptual groupings (i.e., proto-objects are formed) ORN reflects an on/off flag signalling whether one or more sound objects are likely to be present. If no cue or cue combination reaches a certain threshold, the incoming signal is represented as a single sound object. Once a threshold is reached (regardless of the number of congruent cues, using the horse-race principle) the presence of more than one object is signalled, and competition between the alternatives commences. This competition can be biased by attention, thus choosing which alternative is consciously perceived.

The processes of concurrent sound segregation probably precede those of sequential sound segregation (see Winkler and Schröger's (2015) model of auditory event formation). This allows sequential grouping processes to integrate the outcome of concurrent sound segregation thus forming proto-objects from a subset of the simultaneously encountered sounds.

However, the current results are also compatible with the temporal coherence model of auditory scene analysis (Shamma and Micheyl, 2010; Shamma et al., 2011). Within this framework, the ORN would appear as a temporal coherence detector. Teki and colleagues (2011; 2013) suggest that in the first stage of processing, feature analysis is done, whereas in the second stage, the output of the previous stage is grouped according to temporal modulation. Since ORN reflects the aggregate readout of cues, it can be seen as temporal coherence detector. Teki and colleagues (2013) also suggest that in addition to the first two stages, attention plays a key role in the formation of the streams as attention biases the auditory system toward a particular grouping of sound source attributes that depend on the listener's current behavioural and perceptual goals. In this light, P400 reflects the outcome of the biasing process: the detection of the target features. It is important to note that there are neurophysiological signals sensitive to temporal coherence (O'Sullivan et al., 2015). Thus it is likely that temporal coherence is indeed computed in the human brain. However, this does not decide between Bregman's theoretical framework and temporal coherence models, because temporal coherence based grouping could occur during the first phase of Bregman's auditory scene analysis (i.e., it would be regarded as the heuristic implementation of the Gestalt principle of "common fate").

Deciding between Bregman's framework and the temporal coherence model requires further research. Insights could be gained from stream segregation experiments using stimuli that can be grouped by mechanisms outside temporal coherence, such as, inserting predictable patterns separately into the putative streams (e.g., a familiar melody within an unfamiliar background; c.f., Szalárdy and colleagues, 2014), salient or unexpected stimuli that are inconsistent with the listening situations (a dog bark in the bathroom), or presenting speech streams in which the semantic congruity can also help segregation. Another possible piece of evidence could come from finding representations of alternative (non-dominant) sound organizations, as these are not computed by temporal coherence models. The existence of neuronal populations encoding sound objects that are currently not perceived would support Bregman's model of auditory scene analysis.

Other open issues are listed below. First, not much is known about the neural substrates of concurrent sound segregation. fMRI measures could shed further light on the brain regions involved in these processes. Second, the stochastic figure-ground stimuli bring the research closer to more realistic sound environments, but there is still space to improve: concurrent sound segregation could be studied by the use of streams of natural sounds (not random noise background); for example, environmental sounds containing a meaningful target. Third, so far the paradigms used have been employing mostly short sounds, future studies should be conducted to investigate the differences if continuous sounds were used instead, and the mistuning happened during listening and not only in a very short time window. Finally, integrating concurrent sound segregation into segregating speech streams would result in better understanding of how our auditory system copes with the cocktail party problem.

List of publications related to theses

I. Tóth, B., Kocsis, Z., Háden, G.P., Szerafin, Á., Shinn-Cunningham, B., & Winkler, I. (2016). EEG signatures accompanying auditory figure-ground segregation. *Neuroimage*, 141, 108-119. DOI: 10.1016/j.neuroimage.2016.07.028.

II. Kocsis, Z., Winkler, I., Szalárdy, O., & Bendixen, A. (2014). Effects of multiple congruent cues on concurrent sound segregation during passive and active listening: An event-related potential (ERP) study. *Biological Psychology*, 100, 20-33. DOI: 10.1016/j.biopsycho.2014.04.005.

III. Tóth, B., Kocsis, Z., Urbán, G., & Winkler, I. (2016). Theta oscillations accompanying concurrent auditory stream segregation. *International Journal of Psychophysiology*, 106, 141-151. DOI: 10.1016/j.ijpsycho.2016.05.002.

IV. Kocsis, Z., Winkler, I., Bendixen, A., & Alain, C. (2016). Promoting the perception of two and three concurrent sound objects: an event-related potential study. *International Journal of Psychophysiology*, 107, 16-28. DOI: 10.1016/j.ijpsycho.2016.06.016.

References

Alain, C., Arnott, S. R., & Picton, T. W. (2001). Bottom-up and top-down influences on auditory scene analysis: Evidence from brain potentials. *Journal of Experimental Psychology,* 27, 1072-1089.

Alain, C., & McDonald, K. L. (2007). Age-related differences in neuromagnetic brain activity underlying concurrent sound perception. *Journal of Neuroscience*, 27, 1308- 1314.

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

Ciocca, V. (2008). The auditory organization of complex sounds. *Frontiers in Bioscience*, 13, 148-169.

O'Sullivan, J. A., Shamma, S. A., & Lalor, E. C. (2015). Evidence for neural computations of temporal coherence in an auditory scene and their enhancement during active listening. *Journal of Neuroscience,* 35(18), 7256-7263.

Shamma, S. A., & Micheyl, C. (2010). Behind the scenes of auditory perception. *Current Opinion in Neurobiology,* 20, 361-366.

Shamma, S. A., Elhiali, M., & Micheyl, C. (2011). Temporal coherence and attention in auditory scene analysis. *Trends in Neuroscience,* 34(3), 114–123.

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(3), 1392-1405.

Teki, S., Chait, M., Kumar, S., von Kriegstein, K., & Griffiths, T. D. (2011). Brain bases for auditory stimulus-driven figure-ground segregation. *The Journal of Neuroscience,* 31, 164-171.

Teki, S., Chait, M., Kumar, S., Shamma, S., & Griffiths, T. D. (2013). Segregation of complex acoustic scenes based on temporal coherence. *eLife,* 2: e00699.

Winkler, I., & Schröger, E. (2015). Auditory perceptual objects as generative models: Setting the stage for communication by sound. *Brain and Language*, 148, 1-22.