EEG signatures accompanying auditory figure-ground segregation

Brigitta Tóth1,2 Zsuzsanna Kocsis1,3, Gábor P. Háden1, Ágnes Szerafin1,3, Barbara Shinn-Cunningham2, István Winkler1,4

1) Institute of Cognitive Neuroscience and Psychology, Research Centre for Natural Sciences, Hungarian Academy of Sciences, Budapest, Hungary

2) Center for Computational Neuroscience and Neural Technology, Boston University, Boston, USA

3) Department of Cognitive Science, Faculty of Natural Sciences, Budapest University of Technology and Economics, Budapest, Hungary

4) Department of Cognitive and Neuropsychology, Institute of Psychology, University of Szeged, Szeged, Hungary

Corresponding author:

Brigitta Tóth

Research Centre for Natural Sciences, Hungarian Academy of Sciences

P.O.Box 1519

H-1519

Budapest, Hungary

Tel.: +3613826809

E-mail address: toth.brigitta@ttk.mta.hu

Abstract

In everyday acoustic scenes, figure-ground segregation typically requires one to group together sound elements over both time and frequency. Electroencephalogram was recorded while listeners detected repeating tonal complexes composed of a random set of pure tones within stimuli consisting of randomly varying tonal elements. The repeating pattern was perceived as a figure over the randomly changing background. It was found that detection performance improved both as the number of pure tones making up each repeated complex (figure coherence) increased, and as the number of repeated complexes (duration) increased – i.e., detection was easier when either the spectral or temporal structure of the figure was enhanced. Figure detection was accompanied by the elicitation of the object related negativity (ORN) and the P400 event-related potentials (ERPs), which have been previously shown to be evoked by the presence of two concurrent sounds. Both ERP components had generators within and outside of auditory cortex. The amplitudes of the ORN and the P400 increased with both figure coherence and figure duration. However, only the P400 amplitude correlated with detection performance. These results suggest that 1) the ORN and P400 reflect processes involved in detecting the emergence of a new auditory object in the presence of other concurrent auditory objects; 2) the ORN corresponds to the likelihood of the presence of two or more concurrent sound objects, whereas the P400 reflects the perceptual recognition of the presence of multiple auditory objects and/or preparation for reporting the detection of a target object.

Keywords: perceptual object, auditory scene analysis, figure-ground segregation, event-related brain potentials (ERP), object-related negativity (ORN), ERP source localization

1. Introduction

Selectively hearing out a sound from the background of competing sounds (referred to as auditory figure–ground segregation) is one of the main challenges that the auditory system faces in everyday situations. In ordinary acoustic scenes, figure and ground signals often overlap in time as well as in frequency content. In such cases, auditory objects are extracted by integrating sound components both over time and frequency. Auditory figure–ground segregation thus involves most of the processes of auditory scene analysis (Bregman, 1990): 1) grouping simultaneous components from disparate spectral regions and 2) across time into perceptual objects or sound streams, while 3) separating them from the rest of the acoustic scene. Event-related brain potential (ERP) correlates of simultaneous and temporal/sequential grouping have been studied, but they have generally been treated separately. As a result, little is known about the responses emerging in more natural situations where both grouping processes are required for veridical perception. The aim of the present study was to investigate electrophysiological correlates of figure–ground segregation by using auditory stimuli with high spectro-temporal complexity. The salience of the figure was varied systematically by independently manipulating sequential and simultaneous cues supporting figure detection. This design allowed us to investigate the electrophysiological correlates of the emergence of an auditory object from a stochastic background.

Auditory objects are formed by grouping incoming sound components over frequency and time (Kubovy and van Valkenburg, 2001; Griffiths & Warren, 2004; Shinn-Cunningham, 2008; Winkler, et al., 2009; Bizley & Cohen, 2013) on the basis of various grouping heuristics (Bregman, 1990; Denham and Winkler, 2014). Simultaneous grouping is driven by various sound features such as common onset/offset (Lipp, Kitterick, Summerfield, Bailey, & Paul-Jordanov, 2010; Weise, Bendixen, Müller, & Schröger, 2012), location, loudness (Bregman, 1990; Darwin, 1997), as well as harmonic structure, or, more generally, spectral templates (Lin and Hartman, 1998; Alain, Schuler, & McDonald, 2002; for a review, see Ciocca, 2008). Feature similarity promotes sequential grouping (van Noorden, 1975; Moore and Gockel, 2002; for reviews see Bregman, 1990; Carlyon et al, 2001). It interacts with the temporal separation of successive sounds, such that longer gaps between sounds reduce the likelihood of grouping even similar sounds into the same perceptual stream (Winkler et al., 2012; Mill et al., 2013). Temporal structure has been suggested to guide attentive grouping processes through rhythmic processing (Jones and Kidd et al. 1981) and/or temporal coherence between elements of the auditory input (Shamma et al., 2011, 2013). For example, within a stochastic background, the spectrotemporal regularity of a repeating cluster of synchronous tones causes them to stream together into a perceptual object distinct from the acoustic background (Elhilali, Xiang, Shamma, & Simon, 2009; Elhilali, Ma, Micheyl, Oxenham, & A, 2010). Indeed, temporal regularity also aids temporal/sequential segregation by allowing listeners to predict upcoming sounds (Dowling et al., 1987; Bendixen et al., 2010; Devergie et al., 2010; Szalárdy et al., 2014).

Few past studies addressed interactions between simultaneous and temporal grouping cues. Differences in amplitude modulation, a cue that helps simultaneous grouping through the gestalt “common

fate" principle, has been also found effective for temporal grouping (Grimault et al., 2002; Szalárdy et al., 2013; Dolležal et al., 2012). Testing temporal coherence and harmonicity separately and together, Micheyl and colleagues (2013) found that the two cues separately facilitated auditory stream segregation. Teki and colleagues (2011, 2013) designed a new stimulus for testing both simultaneous and sequential grouping in auditory figure-ground segregation. The stimuli consist of a sequence of chords that are made up of pure tones with random frequency values and no harmonic relation to each other. When a subset of these tonal components is repeated several times, they form an auditory object (figure) which pops out from the rest of the stimulus (ground). The *coherence* of the figure is controlled by the number of frequencies in the subset making up the repeating chords, while the number of repetitions sets the *duration* of the figure. The separation of the figure from the ground requires integrating across both frequency and time. Specifically, there are no low-level feature differences between the figure and the ground; the subset of repeated components making up the figure chord is randomly chosen for each trial and each frequency can serve as part of the figure or of the ground, depending on the trial. Listeners are sensitive to the appearance of the spectro-temporally coherent figure in such stimuli, and figure salience systematically increases with increasing figure *coherence* and increasing figure *duration* (Teki et al., 2011; Teki et al., 2013, O'Sullivan et al., 2015).

Neural correlates of auditory stream segregation originate from a distributed network including the primary and non-primary auditory cortices and the superior temporal and intraparietal sulci (Teki et. al. 2011; Alain, 2007; Alain & McDonald, 2007; Alain et al., 2002; O'Sullivan et al., 2015). Electrophysiological correlates of figure ground segregation have been investigated by using linear regression for extracting a signature of the neural processing of different temporal coherence defining a foreground object over a stochastic background (O'Sullivan et al., 2015). The results showed fronto-central activity suggesting early pre-attentive neural computation of temporal coherence between 100 and 200 ms post-stimulus, which was extended beyond 250 ms when listeners were instructed to detect the figure. Further, a frontocentrally negative event-related potential (ERP) component of sound segregation, which typically peaks between 150 and 300 ms from cue onset, is elicited by auditory objects segregated by simultaneous cues (Alain et al., 2003, 2001, Alain & McDonald, 2007, 2005). The object-related negativity (ORN) appears to reflect the outcome of the simultaneous segregation process (i.e., the perceptual decision that the acoustic input carries two or more concurrent sounds) rather than the processes leading to the perceptual decision (Kocsis, Winkler, Szalárdy, & Bendixen, 2014). Sound segregation by simultaneous cues interacts with the temporal/sequential probability of the presence of these cues within the sound sequence, thus providing some evidence for joint processing of simultaneous and sequential cues of auditory stream segregation (Bendixen et al., 2010a; Bendixen et al., 2010b). When listeners are instructed to report whether they heard one or two sounds, ORN is followed by the centro-parietal P400 component peaking at about 450 ms from cue onset (Alain et al., 2001, 2002). P400 amplitude correlates with the likelihood of consciously perceiving two concurrent sound objects (Alain et al., 2001, 2002; Johnson, Hautus, & Clapp, 2003). As for the ERP correlates of sequential sound

segregation, the auditory P1 and N1 have been shown to be modulated by whether the same sound sequence is perceived in terms of a single (integrated) or two separate (segregated) streams (Gutschalk, 2005; Micheyl et al., 2007; Snyder & Alain, 2007; Szalárdy, Bőhm et al. 2013). The mismatch negativity (MMN) ERP can also be used as an index of sequential auditory stream segregation when the auditory regularities that can be detected from the stimulus sequences differ between the alternative sound organizations (Sussman et al., 1999; for reviews, see Winkler et al., 2009; Spielmann et al., 2014). However, MMN does not reflect auditory stream segregation *per se*; it can only be used as an indirect index of segregation in certain paradigms where the way in which the auditory scene is organized determines whether or not a particular sound will be perceived as a predicted or an unexpected event.

In two experiments, we employed the figure-ground stimuli adapted from Teki and colleagues' study (Teki et al., 2011) to analyze figure-ground segregation-related ERPs as a function of figure coherence and duration. Experiment 1 used behavioral methods a) to assess the optimal parameter ranges for figure coherence and duration to be used in the electrophysiological experiment (Experiment 2) and b) to test whether location difference between the frequency components assigned to the figure and the ground enhanced their separation. For Experiment 2, we hypothesized that concurrent sound segregation will lead to the elicitation of ORN and P400 (as listeners were instructed to detect the emergence of the figure) and further that the P400 and possibly the ORN amplitude will increase together with figure coherence, whereas figure duration may gate the emergence of these components. We further hypothesized that interactions between the effects of these parameters on the ERP components would arise, supporting the view that simultaneous (figure coherence) and temporal/sequential (figure duration) grouping cues interact when listeners parse complex acoustic scenes.

2. Experiment 1

2.1. Methods

2.1.1. Participants

20 young adults (10 female; mean age: 22.4 years) participated in the experiment. They received modest financial compensation for participation. All participants had normal hearing and reported no history of neurological disorders. The United Ethical Review Committee for Research in Psychology (EPKEB; the institutional ethics board) approved the study. At the beginning of the experimental session, written informed consent was obtained from participants after the aims and methods of the study were explained to them.

2.1.2. Stimuli

The auditory stimuli (see a schematic example in Figure 1) were adapted from Teki and colleagues' study (Teki et al., 2011). Each sound consisted of a sequence of 40 random chords of 50 ms duration with no inter-chord interval (total sound duration: 2000 ms). Chords consisted of 9- 21 pure tone components.

Component frequencies were drawn with equal probability from a set of 129 frequency values equally spaced on a logarithmic scale between 179 and 7246 Hz. The onset and offset of the chords were shaped by 10 ms raised-cosine ramps. In half of the stimuli, the same chord (containing 4 or 6 tonal components) was repeated 2, 3, or 4 times in a row (resulting in 3, 4, or 5 identical chords, respectively), thus forming a "figure" over the background of random chords. In the other half of the stimuli, random chords of 4 or 6 tonal components ("control") were added to 3, 4, or 5 consecutive chords (control chords). Past work showed that listeners could segregate repeating chords (but not additional random chords) from the other concurrent chords ("ground"), resulting in the perception of a foreground auditory object and a variable background (Teki et al., 2011). Each figure/control chord had a unique spectral composition with their frequencies randomly chosen from the set. The figure/control chords appeared at a random time between 200–1800 ms from stimulus onset (between the 5^{th} and the 35^{th} position within the sequence of 40 chords).

The figure chord sequences differed across trials on three dimensions: duration (the number of chords: 3, 4, or 5), coherence (the number of tonal components comprising the chord: 4 or 6), and perceived difference in lateral direction relative to the background (no difference, roughly 45˚ difference, or roughly 90˚ difference). The tones forming the background were always presented dichotically (perceived as originating from a midline location). In contrast, the interaural time and level differences (ITDs and ILDs, respectively) of the figure/control chords were manipulated to change their perceived laterality, either set to zero (heard at the same midline location as the background), heard at a lateral angle of roughly ±45˚ (ITD=±395 μs and ILD=±5.7 dB), or heard at a lateral angle of roughly ±90˚ (ITD=±680 μs and ILD=±9.08 dB). Thus, the figure and the ground overlapped spectrally; they could only be separated based on the figure's coherence and, when different from the background, the differences in perceived location.

……..FIGURE 1……...

Consecutive trials were separated by an inter-trial interval of 2000 ms. Listeners were presented with 20 trials of each stimulus type (figure vs. control × 2 coherence levels × 3 duration levels × 3 perceived location difference levels = 72 stimulus types, each appearing with equal probability) in a randomized order.

Stimuli were created using MATLAB 11b software (The MathWorks) at a sampling rate of 44.1 kHz and 16-bit resolution. Sounds were delivered to the listeners via Sennheiser HD600 headphones (Sennheiser electronic GmbH & Co. KG) at a comfortable listening level of 60–70 dB SPL (self-adjusted by each listener). Presentation of the stimuli was controlled by Cogent software (developed by the Cogent 2000

team at the FIL and the ICN and Cogent Graphics developed by John Romaya at the LON) under MATLAB.

2.1.3. Procedure

Listeners were tested in an acoustically attenuated room of the Research Centre for Natural Sciences, MTA, Budapest, Hungary. Each trial consisted of the presentation of the 2000-ms long sound, during which they were asked to focus their eyes on a fixation cross that appeared simultaneously at the center of a 19" computer screen (directly in front of the listener at a distance of 125 cm). After the stimulus ended, a black screen was presented for 2000 ms. Listeners were instructed to press one of two response keys either during the stimulus or the subsequent inter-trial interval to indicate whether or not they detected the presence of a "figure" (repeating chord). The instruction emphasized the importance of responding correctly over response speed. The response key assignment (left or right hand) remained the same throughout the experiment and was counterbalanced across participants.

Prior to conducting the main experiment, listeners performed a 15 min practice session with feedback. The practice session consisted of two parts. In the first part, six stimulus sequences were presented. Each sequence consisted of 5 examples of the figure and 5 of the control condition, delivered in a randomized order (60 trials, altogether). In the practice session, the duration and coherence values used covered a larger range than in the main experiment, but all components were presented dichotically (no spatial location difference was employed). The figure stimuli were categorized into easy-to-detect (duration=5, coherence=6 and duration=3, coherence=8), moderately-difficult-to-detect (duration=4, coherence=4 and duration=3, coherence=6), and difficult-to-detect (duration=3, coherence=4 and duration=2, coherence=3) groups. In order to help listeners to learn the task, practice trials were organized into sequences consisting of sounds with the same difficulty level; these sequences were presented in descending order of detectability, from easy-to-detect to difficult-to-detect. All other parameters were identical to those described for the main experiment. To accustom listeners to the perceived location manipulation, 6 additional practice blocks were presented, one for each of the six levels of perceived location difference presented (0, 15, 30, 45, 60, and 90°). In these practice sequences, the figure duration was always 5 and the coherence level 6. Each level of the perceived location difference was presented for 12 trials (6 with a figure and another 6 with the control; 72 overall). These were presented in a fixed order (90 60, 0, 45 30, and 15°). All other stimulus parameters were identical to those described for the main experiment.

No feedback was provided to listeners in the main experiment, which lasted for about 1.5 hours. The main experiment was divided into 20 blocks, each consisting of 72 trials. The order of the different types of trials was randomized separately for each listener. Listeners were allowed a short rest between stimulus blocks.

2.1.4. Data analysis

Reaction times were not analyzed, because listeners were instructed to respond accurately rather than as fast as they could. For the d‘ values (the standard measure for discrimination sensitivity; see, for example, Green and Swets, 1988) a repeated-measures ANOVA was performed with the factors of Coherence (2 levels: 4 vs. 6 tonal components) × Duration (3 levels: 3 vs. 4 vs. 5 chords) × Location difference (3 levels: 0 vs. 45 vs. 90˚). Statistical analyses were performed with the Statistica software (version 11.0). When the assumption of sphericity was violated, degrees of freedom values were adjusted using the Greenhouse-Geisser correction. Bonferroni's post hoc test was used to qualify significant effects. All significant results are described. The ε correction values for the degree of freedom (where applicable) and the partial η^2 values representing the proportion of explained variance are shown.

……..FIGURE 2……...

2.2. Results and Discussion

The results of Experiment 1 are presented in Figure 2. The fact that the d' values exceeded 2 for several parameter combinations demonstrates that listeners were sensitive to the appearance of figure in the stimuli, confirming that the auditory system possesses mechanisms that process cross-frequency/time correlations (Teki et al., 2011). The main effect of Coherence ($F(1,19) = 97,05$, $p < 0.001$; $\eta^2 = 0.83$) demonstrates that listeners were better at detecting figures containing six tonal components than those comprising four components. The main effect of Duration was also significant ($F(2,38) = 114.98$, $p < 0.001$; $\eta^2 = 0.85$). Pairwise post-hoc comparisons showed that the d' values were significantly higher for figure duration of 5 than for durations of 3 or 4 chords ($p < 0.001$, both), and that the d' for figure duration of 4 chords was significantly higher than for duration of 3 chords ($p < 0.001$). Location difference also yielded a significant main effect ($F(2,38)=9,96$, $p < 0.01$; $\eta^2 = 0.34$). Post hoc pairwise comparisons showed that the d' for figures with 90° difference from the ground was significantly lower than that for figures with 0 or 45° location difference ($p<0.01$, both). There were no significant interactions between the three factors.

Similarly to previous results (Teki et al., 2011), we found that increasing figure coherence and duration helped listeners to separate the figure from the ground in the expected way and without interactions between these factors. We expected that increasing location difference between the figure and the ground would help figure-ground segregation, helping the detection of the figure. Instead we found that a large separation between the figure and ground interfered with detection of the figure. We ascribe this difference to an effect of top-down attention: the figure could appear at any lateral angle, from roughly -90˚ to +90˚; listeners may have adopted a strategy of listening for the figure near midline (at the center of the range). If the actual figure was too far from this attended direction (e.g., at the extreme locations of ±90˚), it may have fallen outside the focus of attention. Given that our focus was on bottom-up, automatic processes involved in segregating figure and group, we excluded the location manipulation from Experiment 2.

3. Experiment 2

3.1. Methods

3.1.1. Participants

27 young adults (17 female; mean age 21.9 years) with normal hearing and no reported history of neurological disorders participated in the experiment. None of the participants were taking medications affecting the nervous system and none of them participated in Experiment 1. The study was approved by the institutional ethics board (EPKEB). At the beginning of the experimental session, written informed consent was obtained from participants after the aims and methods of the study were explained to them. Participants were university students who received course credit for their participation. Data of one participant was excluded from the analysis due to a technical problem in the data recording.

3.1.2. Stimuli

The stimuli were identical to those delivered in the "no location difference" condition of Experiment 1 except that the test sounds were composed of 41 tonal segments. The stimulus set in the EEG experiment therefore comprised six stimulus conditions: 2 coherence levels (4, 6 tonal components) × 3 duration levels (3, 4, 5 chords). Fifty percent of the sounds carried a figure, which appeared between 200 and 1800 ms (5^{th}–35^{th} chord) from onset.

3.1.3. Procedure

Participants were tested in an acoustically attenuated and electrically shielded room of the Research Centre for Natural Sciences, MTA, Budapest, Hungary. Each trial started with the delivery of the sound with a concurrent presentation of the letter "S" at the center of a 19" computer screen placed directly in front of the participant (distance: 125 cm). Following the stimulus presentation, the letter "S" was replaced by a question mark on the screen denoting the response period which lasted until a response was made. After the response was recorded, the screen was blanked for a random inter-trial interval of 500-800 ms (uniform distribution) before the next trial began. Listeners were instructed to press one of two response keys during the response period to mark whether or not they detected the presence of a "figure" (repeating chord). The instruction emphasized the importance of confidence in the response over speed. The response key assignment (left or right hand) remained the same during the experiment and was counterbalanced across participants.

Before the main experiment, participants completed a short practice session (10 minutes) during which they received feedback. The practice session was identical to the first part of the practice session of Experiment 1. (The second part, training for the perceived location manipulation, was skipped.)

The main experiment lasted about 90 minutes. Overall, listeners received 130 repetitions of each stimulus type (2 coherence levels × 3 duration levels × figure present vs. absent), divided into 10 stimulus blocks of 156 trials each. The order of the different types of trials was separately randomized for each listener. Participants were allowed a short rest between stimulus blocks.

3.1.4. Data analysis

3.1.4.1. Behavioral responses

Figure detection was assessed by means of the sensitivity index (d' value), separately for each figure type, with the control trials serving as distractors. For the d' data, a repeated-measures ANOVA was performed with the factors of Coherence (2 levels: 4 vs. 6 tonal components) × Duration (3 levels: 3 vs. 4 vs. 5 chords).

3.1.4.2. EEG recording and preprocessing

EEG was recorded from 64 locations of the scalp with Ag/AgCl electrodes placed according to the international 10-20 system with Synamps amplifiers (Neuroscan Inc.) at 1 kHz sampling rate. Vertical and horizontal eye movements were recorded by electrodes attached above and below the left eye (VEOG) and lateral to the left and right outer canthi (HEOG). The tip of the nose was used as reference and an electrode placed between Cz and FCz was used as ground (AFz). The impedance of each electrode was kept below 15 kΩ. Signals were filtered on-line (70 Hz low pass, 24dB/octave roll off).

The analysis of EEG data was performed using Matlab 7.9.1 (Mathworks Inc.) The continuous EEG signal was filtered between 0.5-45 Hz by band-pass finite impulse response (FIR) filter (Kaiser windowed, Kaiser $\beta = 5.65$, filter length 4530 points). EEG signals were converted to average reference. In order to exclude EEG segments containing infrequent electrical artifacts (rare muscle and movement artifacts etc.), the data were visually screened and the affected segments were rejected. Next the Infomax algorithm of Independent Component Analysis (ICA) (as implemented in EEGlab; for detailed mathematical description and validation, see Delorme and Makeig, 2004) was performed on the continuous filtered dataset of each subject, separately. ICA components constituting blink artifacts were removed via visual inspection of their topographical distribution and frequency content.

3.1.4.3. ERP data analysis

For the ERP analysis, the EEG signals were down-sampled to 250 Hz and filtered between 0.5–30 Hz by a band-pass finite impulse response (FIR) filter (Kaiser windowed, Kaiser $\beta = 5.65$, filter length 4530 points). EEG epochs of 850 ms duration were extracted separately for each stimulus from 50 ms before the onset of the figure/control within each trial and baseline corrected by the average voltage in the pre-stimulus period. Epochs with an amplitude change exceeding 100 μV at any electrode were rejected from

further analysis. The data of one subject were excluded from further analysis due to low signal to noise ratio: we obtained fewer than 20 artifact free epochs for one of the stimulus types. Overall, 84.2% of the data was retained.

Difference waveforms were calculated between ERPs elicited by the figure- and the control-trial responses. Inspecting the group-averaged difference waveforms elicited by the figure trials in each condition, we observed an earlier negative and a later positive centroparietal response in most conditions. We tentatively identified them as ORN and P400, respectively. Using the typical latency windows for ORN (150-300 ms) and P400 (450-600 ms) we performed peak detection for ORN and P400 at their typical maximal scalp location (maximal negative value at Cz and maximal positive value Pz within the ORN and P400 time window, respectively) on the group-averaged waveforms, separately for each condition. Based on these peak latencies, ORN and P400 amplitudes were then averaged from 100 ms wide windows centered on the detected peaks (see Table 1 for descriptive statistics of the ERP amplitudes). Individual peak latencies were determined from the same latency windows and electrode location as was described above. For assessing whether ORN and/or P400 were elicited, ERP amplitude differences were tested against zero by one-sample t-tests, separately for each stimulus condition and time window. For testing the effects of coherence and duration on figure vs. control trials, central (Cz) ORN and parietal (Pz) P400 amplitudes and peak latencies were compared by repeated-measures ANOVA with the factors of Coherence (2 levels: 4 vs. 6 tonal components) x Duration (3 levels: 3 vs. 4 vs. 5 chords).

For testing the effects of coherence and duration on hit and miss trials, difference waveforms were calculated between ERPs elicited by hit (correct response to figure trials) and miss trials (no response to figure trials). Peak latency and subsequent amplitude measurements were performed by the same procedure as those described for figure vs. control trial analyses. Measurement windows and descriptive statistics are shown in Table 2. Because both this and the following analyses were based on the figure trials alone, only half of the trials were used. In the Coherence-4/Duration-3 and in the Coherence-6/Duration-5 conditions, very few hit or miss trials were obtained because of the very low and very high detection rates (respectively). Therefore, these stimulus conditions were excluded from further analysis. Paired-samples t-tests were performed separately for the remaining four stimulus types to compare the trial types (hits vs. misses). In order to determine whether the processes indexed by ORN and P400 are related to the inter-individual variability in figure detection sensitivity, the amplitude differences between hit and miss trials in the ORN (Cz) and P400 (Pz) time windows were correlated with d' (Pearson correlation), separately for each stimulus condition.

Statistical analyses were performed with the Statistica software (version 11.0). When the sphericity assumption was violated, the degrees of freedom were adjusted using the Greenhouse-Geisser correction. Bonferroni's post hoc test was used to qualify significant effects. All significant results are described. The ε correction values for the degree of freedom (where applicable) and the partial η^2 values representing the proportion of variance explained are shown.

3.1.4.4. Source localization by sLORETA

The sLORETA software (standardized Low Resolution Brain Electromagnetic Tomography; Pascual-Marqui et al., 2002) allows the location of the neural generators of the scalp-recorded EEG to be estimated. The algorithm limited the solution to the cortical and hippocampal grey matter according to the probability template brain atlases based on template structural MRI data provided by the Montreal Neurological Institute (MNI). Electrode locations were calculated according to the 10-20 system without individual digitization. The solution space is divided into 6239 voxels (5x5x5 mm resolution). Source localization computations are based on a three-shell spherical head model registered to the Talairach human brain atlas. Because the highest-amplitude sound segregation related ERP responses were obtained for the Coherence-6 stimuli, current density maps were generated from the ORN (200-350 m) and P400 (460-600) measurement windows of the figure and control trials collapsing across durations 3-5, separately for each participant. For comparisons of the electrical source activity between the figure and the control trials, Student's t value maps were generated using the LORETA-Key software package's statistical nonparametric mapping voxel-wise comparison calculation tool.

3.2. Results

3.2.1. Behavioral responses

Group-averaged d' values are presented in Figure 3. There was a significant main effect of Coherence ($F(1,24) = 153.84$, $p < 0.001$, $\eta 2 = 0.865$), confirming that d' was greater for figures consisting of 6 compared to 4 tonal components. The main effect of Duration was also significant ($F(2,48) = 193.51$, $p < 0.001$, $\eta 2 = 0.89$, $\varepsilon = 0.89$). Pairwise post hoc comparisons showed that the d' values for figure duration of 5 chords were significantly higher than those for durations of 3 or 4 chords ($p < 0.001$, both), and the d' values for figure duration of 4 chords were significantly higher than those for duration of 3 chords ($p < 0.001$). There was also a significant interaction between Duration and Coherence ($F(2,48) = 18.52$, $p < 0.001$, $\eta 2 = 0.44$). All post hoc pairwise comparisons between different figure types yielded significant ($p < 0.001$) results, except that between Coherence-6/Duration-3 and Coherence-4/Duration-4. These results are compatible with those of Teki et al. (2011) and of Experiment 1.

……..FIGURE 3……...

3.2.2. ERP responses

3.2.2.1. Comparison between the Figure and Control responses

Mean ERP responses elicited by all figure and control sounds are shown in Figure 4. Figure-minus-control difference amplitudes measured from the ORN and P400 time windows (at Cz and Pz, respectively) significantly differed from zero for all stimulus types except for Coherence-4/Duration-3 (see Table 1). The ORN shows a lateral central maximum extending to central and parietal scalp locations with increasing Coherence and Duration. The P400 shows a midline parietal maximum extending towards lateral and central scalp locations with increasing Coherence and Duration. Table 2 shows all significant results for the ANOVAs of the ORN and P400 amplitudes.

The ANOVA comparing the central (Cz) ORN amplitudes showed a significant main effect of Coherence ($F(1,24) = 24.61$, $p < 0.001$, $\eta 2 = 0.506$), which was due to significantly larger amplitudes for Coherence-6 than for Coherence-4 stimuli ($p < 0.001$). The main effect of Duration was also significant ($F(2,48) = 8.288$, $p < 0.001$, $\eta 2 = 0.257$); post-hoc pairwise comparisons showed significantly larger amplitudes for Duration 5 than for the 3 or 4 conditions ($p < 0.001$ and $p = 0.047$, respectively). The ANOVA comparing the ORN peak latencies showed a significant main effect of Duration ($F(2, 48)=9.12$, $p < 0.001$, $\eta 2 = 0.275$) with post-hoc pairwise comparisons indicating significantly shorter ORN latencies in the 3 than the 4 or 5 chords conditions ($p < 0.02$ and $p < 0.001$, respectively). Note that the peak-latency effect was caused by the increased ORN duration and amplitude elicited at longer figure durations (see Figure 4).

The ANOVA comparing the parietal (Pz) P400 amplitudes showed significant main effects of Coherence ($F(1,24) = 37.856$, $p < 0.001$, $\eta 2 = 0.611$) due to significantly higher amplitudes for the 6 tonal components than for 4 tonal components ($p < 0.001$) and Duration ($F(2,48) = 51.944$, $p < 0.001$, $\eta 2 = 0.684$), post-hoc pairwise comparisons showed significantly higher amplitudes for 5 than for 3 or 4 chords and for 4 than for 3 chords; $p < 0.001$ in all comparisons. There was also a significant interaction between Coherence and Duration ($F(2,48) = 4.005$, $p = 0.025$, $\eta 2 = 0.143$). Post hoc ANOVAs were performed with the factors of Coherence (2 levels: 4 vs. 6 tonal components) separately for each level of Duration. These revealed significant Coherence main effects at each level of Duration ($F(1,24) = 9.32$, $p = 0.005$, $\eta 2 = 0.279$; $F(1,24) = 29.11$, $p < 0.001$, $\eta 2 = 0.548$; $F(1,24) = 21.91$, $p < 0.001$, $\eta 2 = 0.477$; for Durations levels 3, 4, and 5, respectively). The Coherence main effect size was lower for stimuli with Duration 3 than for stimulus with Duration 4 or 5. These results indicate that the source of interaction between Coherence and Duration is that the effect of Coherence is larger at the two longer than at the shortest duration. The ANOVA comparing the P400 peak latencies showed a significant main effect of Coherence ($F(1, 24)=11.49$, $p= 0.002$; $\eta 2 = 0.323$) due to significantly shorter ERP latency for Coherence-6 than for Coherence-4 stimuli.

……..FIGURE 4……..

……..TABLE 1……..

3.2.2.2. Comparison between the hit and miss figure trial responses

ERP responses from the hit and miss figure trials are shown in Figure 5. The central (Cz) hit and miss amplitudes measured in the ORN latency range significantly differed from each other for all but one of the tested stimulus condition: Coherence-4/Duration-3 (see Table 2).[1] The parietal (Pz) amplitudes measured from the P400 latency range significantly differed between hit and miss trials for each of the tested conditions (see Table 2).

……..FIGURE 5……...

……..TABLE 2…….....

3.2.2.3. ORN and P400 source localization

LORETA paired-sample t-tests revealed significantly higher current source density in response to figure than control trials corresponding to the sources of ERPs at the ORN and P400 time windows. LORETA t value maps superimposed on the MNI152 standard brain are shown in Figure 6, while the statistical results are shown in Tables 3 and 4 for the ORN and P400 ERPs, respectively. In both time windows, Brodmann area 41 (BA 41) on the right hemispheres, the anterior transverse temporal part of the primary auditory cortices, and the anterior cingulate cortex (ACC, BA 25, 33) were found to be more active during figure compared to control trials. At the ORN time window, activity was greater for figure than control trials also in the cortical regions of BA 39, including areas of the superior temporal gyrus and the inferior parietal sulcus (angular gyrus). In the time window of P400, several other brain regions were observed to be more active for figure than for control stimuli. These include frontal cortical areas such as the medial and superior frontal gyri (BA 6, 32, 31), the cingulate cortices (BA 23,24, 29, 30,31,32), and also areas in the visual cortices (BA 7,18, 19).

……..FIGURE 6……...

……..TABLE 3..……...

……..TABLE 4..……..

3.2.2.4. Correlation between behavioral and ERP measures

[1] Note that the number of trials averaged for the compared hit and miss responses differed from each other. However, the difference never exceeded the ~1:2 ratio, because the *t* tests were only conducted for those conditions in which the number of hit and miss trials separately exceeded 30% of the total number of trials. The Coherence-4/Duration-3 and Coherence-6/Duration-5 conditions were dropped from these analyses due to this reason.

Discrimination sensitivity (d') was correlated with the amplitude difference between hit and miss trials in the ORN and P400 time window. No significant correlation was found for the central (Cz) amplitude difference in the ORN time window. However, significant positive correlations were obtained between the parietal (Pz) hit-minus-miss amplitude difference measured from the P400 time window and d' for four of the six stimulus conditions (see Figure 7).

……..FIGURE 7……...

4. General Discussion

In accordance with the findings of Teki and colleagues (2011 and 2013), the results of both Experiment 1 and 2 showed that both the coherence of the figure and its duration promoted figure–ground segregation: Figure detection performance improved as the number of repeated tonal components increased and as the number of repetitions of the figure elements increased. In other words, the perceptual salience of the figure increased parametrically with increasing figure coherence and duration. This result confirms that the segregation of the figure from the concurrently presented stochastic background required the integration of acoustic elements over time and frequency. Teki and colleagues (2013) showed that the effects of figure coherence and duration on figure–ground segregation can be explained by the temporal coherence principle (Shamma et al., 2011 and 2013). In the temporal coherence model, auditory features (such as location, pitch, timbre, loudness, etc.) are first extracted in auditory cortex by distinct neuron populations. Correlations between the dynamic activity of these distinct cortical populations cause perceptual streams to emerge, as described by the resulting correlational matrix of activity patterns.

We found no evidence that spatial separation between the figure and the background led to an automatic enhancement of figure–ground segregation; instead, when the figure came from the most extreme lateral locations, detection of the figure was poorer than when it came from closer to midline. Taken together with the results of previous studies of simultaneous sound segregation (McDonald and Alain, 2005; Kocsis et al., 2014, Lee and Shinn-Chunningham 2008), this finding supports the idea that spectrotemporal cues contribute automatically to figure–ground segregation, while spatial cues are more influential in directing top-down, volitional attention. This conclusion is also compatible with that of Bregman (1990), who argued that source location is a weak cue of auditory stream segregation.

Correct identification of the figure resulted in the elicitation of a centrally maximal negative response between 200 and 300 ms from the figure onset and a parietally maximal positive response between 450 and 600 ms (Experiment 2). Based on the observed scalp distributions, their cortical source origin, and the latency range, these ERP responses could be identified as the ORN and P400 (Alain and McDonald, 2007; Lipp, Kitterick, Summerfield, Bailey, and Paul-Jordanov, 2010; Johnson, Hautus, Duff, &

Clapp, 2007, Bendixen et al. 2010), respectively, which are known to be elicited when two concurrent sounds are attentively segregated (Alain et al., 2001 and 2002). However, ORN (and P400) have been previously observed only in the context of one vs. two discrete concurrent complex tones, whereas the present figure stimuli formed a coherent stream that was separated from the randomly changing background. Thus, the current results demonstrate that ORN and P400 are elicited also in cases when concurrent sound segregation requires integrating spectral cues over time to form a new stream. In turn, the elicitation of these ERP components suggests that the brain mechanisms underlying figure–ground segregation by spectral coherence over time may reflect some common processes with those involved in simpler forms of simultaneous sound segregation, such as some common segregation mechanism or common consequence of detecting two concurrent sounds. If ORN is based on deviation from some template (Alain et al., 2002), then the current results suggest that the template does not have to be fixed, such as a template of harmonicity (Lin and Hartman, 1998). Rather, it can be built dynamically by extracting higher-order spectro-temporal statistics of the input stimulus. This conclusion is also supported by the results of O'Sullivan and colleagues (2015), who manipulated the coherence level of the figure under both active and passive listening conditions. These authors found that a neual response appearing in the same latency range as the present ORN was correlated with the coherence level of the figure stimuli. It is possible that this neural activity (extracted from the EEG by a linear regression method) corresponds to or at least overlaps with the ORN response obtained with the ERP method in the current study. It is then likely that the early negative response reported in the present and in O'Sullivan et al.'s (2015) study reflect at least partly the same underlying spectrotemporal computations. O'Sullivan et al., however found an effect of the coherence level on the onset latency (the first time point that significantly differed from zero) of their response: lower levels of coherence elicited responses with longer onset latencies. This effect held for stimuli with 6,8,or 10 coherence levels, but not for coherence levels of 2 or 4. In the current study stimuli with 4 vs. 6 coherence levels were tested and no coherence effect on the peak latency of the ORN response was found. One explanation is that the correlation between coherence level and the onset latency of the response only holds for more salient auditory objects. Another alternative is that the onset latency is more sensitive to coherence levels than the peak latency.

There are, however, other event-related brain responses that may also be related to the current early response. Most notable of them is the auditory evoked awareness related negativity (ARN, Gutschalk et al., 2008). ARN was described in an auditory detection task in which listeners were instructed to detect a repeating tone embedded in a stochastic multi-tone background (masker). This paradigm is similar to the current one. The main differences are that in Gutshalk et al.'s (2008) study, only a single tone was repeated and that it was separated in frequency from the tones of the background by a protected band surrounding the frequency of the target tone. Gutshalk and colleagues observed an auditory cortical magnetoencephalographic response in the latency range of 50–250 ms, which was elicited by detcted targets and also in a passive condition (with higher amplitudes for cued than uncued repeating tones). The authors did not discuss the relation of the response they termed ARN to the ORN. One possibility is that

the two components are similar and the current early response matches both. However, the ORN and the ARN may also be separate components. One possible difference between them is that whereas ORN was found rather insensitive to task load (Alain and Izenberg, 2003), no ARN was obtained when the ARN-eliciting stimulus was presented to one ear while attention was strongly focused on sounda presented to the opposite ear (Gutschalk et al., 2008). However, the two tests of attention are not compatible. Thus they do not definitively prove whether ORN and ARN are different responses or not. In the current study, the auditory stimuli were always task-relevant. Therefore, if the ORN and ARN components differ from each other, further experiments are needed to determine which if any matches the the observed early negative response.

Comment [IW1]: Alain, C., & Izenberg, A. (2003). Effects of attentional load on auditory scene analysis. Journal of Cognitive Neuroscience 15(7), 1063–1073.

The N2 ERP responses are also elicited in the same latency range. However, the current early negative ERP response cannot be analogous to either the N2b or the MMN component. Unlike to the N2b, the current early response was found to be generated in the temporo-parietal regions (see source localization results), and unlike to the MMN, the current early response was elicited even though the figure and control trials were delivered with equal probabilities.

The ORN and the P400 amplitude increased together with figure coherence and duration, both of which increase the salience of the figure, as shown by the behavioral results. Further the P400 peak latency decreased with increasing figure coherence. These findings suggest that both the ORN and P400 reflect processes affected by the integrated impact of the different cues of concurrent sound segregation rather than processes affected by individual cues (cf. Kocsis et al., 2014). This conclusion is also compatible with results of studies in the visual domain, which demonstrated that in a visual figure identification task neural responses emerging at about 200 ms reflect perceptual salience rather than physical cue contrast (Straube, Grimsen, & Fahle, 2010). The fact that the ORN peak latency increased together with figure duration increasing from 3 to 4 but not from 4 to 5 segments suggests that ORN reflects the outcome of temporal integration of the cues, at least until some threshold is reached (sufficient evidence is gathered for the presence of multiple concurrent sounds).

The P400 amplitude was significantly correlated with figure detection performance, at least when figure salience was sufficiently high so that detection performance was above chance level. Hence, the inverse relationship between P400 amplitude and task difficulty is clear for stimuli above the perceptual threshold. A similar relationship to behavioral sensitivity has been reported for the P300 component (see Polich & Kok, 1995). Convergent results were obtained in a visual figure identification task: Straube et al (2010) found that increasing the salience of the visual object resulted in increasing P300 amplitudes. An alternative explanation would suggest that P400 reflects attention capture by the presence of the figure. Although one cannot rule out this alternative based on the current results, P400 was found to be elicited by mistuning a partial of a complex tone even when tones with mistuned partials appeared with higher probability than fully harmonic ones within the sequences (Alain, Arnott, & Picton, 2001), making it unlikely that they would have captured attention. There is one more result dissociating ORN and P400 within the

current data: Whereas no significant interaction was observed between the effects of the two cues of figure–ground segregation on the ORN amplitude, the effects of the two cues interacted significantly for the P400 amplitude as well as for discrimination performance (in Experiment 2). Thus, the P400 amplitude is linked directly to behavioral performance in two different ways, whereas the ORN amplitude does not show a similar correspondence to behavior. Furthermore, while ORN is elicited in passive situations (similarly to the brain electric activity observed by O'Sullivan et al., 2015) and has been observed in newborns and 6-month-old infants (Bendixen et. al, 2015; Folland, Butler, Smith, & Trainor, 2012), P400 is only elicited when listeners are instructed to report whether they heard one or two concurrent objects (e.g., Alain et al.2001; McDonald and Alain 2005; Kocsis et al., 2014). These results suggest that ORN reflects the likelihood of the presence of two or more concurrent sounds (the outcome of cue evaluation), whereas P400 relates to the outcome of perceptual decisions (Alain, 2007; Synder and Alain, 2007). The lack of interaction between the effects of the spectral and the temporal figure–ground segregation cue on ORN suggests that these cues independently affect the auditory system's assessment of the likelihood that multiple concurrent sounds are present in an acoustic mixture. Moreover, the significant interaction found between the P400 amplitude and discrimination performance hints that perceptual decisions are non-linearly related to this likelihood, at least for high likelihoods.

Our source localization results suggest that in both the early (ORN) and the late (P400) time intervals, the temporal cortices are involved in the segregation of the figure from the rest of the acoustic scene. This result is in line with previous reports about the sources of concurrent sound segregation-related ERP components (Alain and McDonald, 2007; Snyder, Alain, & Picton, 2006; Wilson et al., 2007) and also with the location of the effects of concurrent sound segregation on transient and steady-state evoked responses, as well as induced gamma oscillations (Bidet-Caulet et al., 2007 and 2009). ERP studies showed that the source waveforms of ORN and P400 were located in bilateral regional dipoles of the primary auditory cortex, whereas direct electrophysiological recording from auditory cortex revealed the involvement of secondary auditory areas, such as the lateral superior temporal gyrus. Furthermore, in auditory cortex, attention to a foreground object leads to sustained steady state power and phase coherence (regular auditory targets) compared to attention to an irregular background (Elhilali et al., 2009). In Elhilali and colleagues' study, the enhancement varied with the salience of the target. For the same type of stimuli as the current study, a previous fMRI study showed that activity in the intraparietal and superior temporal sulci increased when the stimulus parameters promoted the perception of two streams as opposed to one (Teki et al., 2011). However, in contrast to our experimental design, the BOLD responses were recorded during a passive listening condition and analyzed over the whole duration of the stimuli. Thus it is possible that whereas the auditory cortical electrophysiological responses evoked or induced by the emergence of the figure reflect processes directly involved in detecting the emergence of auditory objects and making perceptual decisions, the full network of perceptual object representations extends also to higher auditory cortical and parietal areas. Consistent with this, we find that in the ORN time window, stimuli including a figure elicited higher activity than control trials in areas of the superior

temporal gyrus and the inferior parietal sulcus (angular gyrus), which are also linked with attention towards salient features (for review see Seghier, 2012). The scalp distributions of the figure–ground segregation related neural activity found by O'Sullivan et al (2015) are compatible with the current observations. The angular gyrus is known to receive connections from the parahippocampal gyrus (Rushworth, et al., 2006), which have been shown to have greater activity in response to figure than control stimuli at both the ORN and the P400 time windows. Further, the anterior cingulate cortex (ACC, BA 25, 33), which also showed higher activity for figure than for control stimuli in both time windows, has previously been associated with attentional control processes (Wang et al., 2009). Finally, further brain regions associated with attention control, such as the medial and superior frontal gyri (BA 6, 32, 31) showed higher activation during figure than control trials in the P400 time window. Although the current localization results are either compatible with those of previous studies localizing the neural generators responsible of figure–ground segregation or they can be interpreted in a consistent manner, nevertheless, the precision of our source localization is restricted by the relatively low number of electrodes (N=64), the lack of individual digitization of structural MRI scans and the general limitations of the solutions for EEG source localization (the accuracy with which a source can be located is affected by the factors such as head-modelling errors, source-modelling errors, and instrumental or biological EEG noise, for review see Grech et al., 2008).

5. Summary

Figures with multiple temporally coherent tonal components can be perceptually separated from a randomly varying acoustic ground. Two ERP responses, the ORN and the P400, were elicited when listeners detected the emergence of figures in this situation. Both of these components were at least partly generated in auditory cortex. The ORN and P400 amplitudes were correlated with the salience of the figure, but only the P400 amplitude was correlated with behavioral detection performance. The figures used in our study were defined by their spectro-temporal structure: their emergence depended jointly on integrating information over both time (duration) and frequency (coherence). Our results suggest that auditory cortex is involved in both the integration across time and frequency and the grouping of sound that leads to the emergence of such a figure. ORN probably reflects the likelihood of the presence of multiple concurrent sounds based on the evaluation of the available perceptual cues, whereas P400 appears to be related to the perceptual decision. These ERP components are reliably elicited even in stimulus configurations the complexity of which approaches that of real-life auditory scenes.

Acknowledgments

This work was funded by the Hungarian Academy of Sciences (Magyar Tudományos Akadémia [MTA], post-doctoral fellowship and internship of Erasmus Mundus Student Exchange Network in Auditory Cognitive Neuroscience to B.T. and the MTA Lendület project (LP2012-36/2012) to I.W. The authors are grateful to Tamás Kurics for programming assistance and Emese Várkonyi, Zsófia Zavecz, Csenge Török for collecting the EEG data.

Reference

Alain, C., Arnott, S.R., Picton, T.W., 2001. Bottom-up and top-down influences on auditory scene analysis: evidence from event-related brain potentials. J. Exp. Psychol. Hum. Percept. Perform. 27, 1072–1089.

Alain, C. (2007). Breaking the wave: Effects of attention and learning on concurrent sound perception. Hearing Research, 229(1-2), 225–236.

Alain, C., McDonald, K.L. (2007). Age-related differences in neuromagnetic brain activity underlying concurrent sound perception. The Journal of Neuroscience, 27(6), 1308–1314.

Alain, C., Schuler, B.M., McDonald, K.L. (2002). Neural activity associated with distinguishing concurrent auditory objects. The Journal of the Acoustical Society of America, 111(2), 990–995.

Bendixen, A., Denham, S.L., Gyimesi, K., Winkler, I. (2010). Regular patterns stabilize auditory streams. The Journal of the Acoustical Society of America, 128(6), 3658–3666.

Bendixen, A., Háden, G.P., Németh, R., Farkas, D., Török, M., Winkler, I. (2015). Newborn infants detect cues of concurrent sound segregation. Developmental Neuroscience, 37(2), 172-181.

Bendixen, A., Jones, S.J., Klump, G., Winkler, I. (2010). Probability dependence and functional separation of the object-related and mismatch negativity event-related potential components. Neuroimage, 50, 285-290.

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

Bidet-Caulet, A., Bertrand, O. (2009). Neurophysiological mechanisms involved in auditory perceptual organization. Frontiers in Neuroscience, 3(09), 182–191.

Bidet-Caulet, A., Fischer, C., Bauchet, F., Aguera, P.-E., Bertrand, O. (2008). Neural substrate of concurrent sound perception: direct electrophysiological recordings from human auditory cortex. Frontiers in Human Neuroscience, 1, 5.

Bidet-Caulet, A., Fischer, C., Besle, J., Aguera, P.-E., Giard, M.-H., Bertrand, O. (2007). Effects of selective attention on the electrophysiological representation of concurrent sounds in the human auditory cortex. The Journal of Neuroscience, 27, 9252–9261.

Bizley, J.K., Cohen, Y.E. (2013). The what, where and how of auditory-object perception. Nature Review Neuroscience, 14(10), 693–707.

Carlyon, R.P., Cusack, R., Foxton, J.M., Robertson, I.H. (2001). Effects of attention and unilateral neglect on auditory stream segregation. Journal of Experimental Psychology: Human Perception and Performance, 27, 115-127.

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

Darwin, C.J. (1997). Auditory grouping. Trends in Cognitive Sciences, 1(9), 327–333.

Delorme, A., Sejnowski, T., Makeig, S. (2007). Improved rejection of artifacts from EEG data using high-order statistics and independent component analysis. Neuroimage, 34, 1443-1449.

Denham, S.L., Winkler, I. (2014). Auditory perceptual organization. In J. Wagemans (Ed.), Oxford Handbook of Perceptual Organization , 601-620 Oxford, U.K.: Oxford University Press.

Devergie, A., Grimault, N., Tillmann, B., and Berthommier, F. (2010). Effect of rhythmic attention on the segregation of interleaved melodies. Journal of the Acoustical Society of America, 128, EL1-EL7.

Dolležal, L.V., Beutelmann, R., Klump G.M. (2012). Stream segregation in the perception of sinusoidally amplitude-modulated tones. PLoS One, 7(9):e43615.

Dowling, W.J., Lung, K.M., Herrbold, S. (1987). Aiming attention in pitch and time in the perception of interleaved melodies, Perceptual Psychophysiology, 41, 642-656.

Elhilali, M., Ma, L., Micheyl, C., Oxenham, A.J., Shamma, S.A. (2010). Representation of Auditory Scenes. Computer, 61(2), 317–329.

Elhilali, M., Xiang, J., Shamma, S.A., Simon, J.Z. (2009). Interaction between Attention and Bottom-Up Saliency Mediates the Representation of Foreground and Background in an Auditory Scene. PLoS Biology, 7(6), 1000129.

Folland, N., Butler, B.E., Smith, N., Trainor, L.J. (2012). Processing simultaneous auditory objects: Infants' ability to detect mistuning in harmonic complexes. The Journal of the Acoustical Society of America, 131(1), 993.

Grech, R., Cassar, T., Muscat, J., Camilleri, K.P., Fabri, S.G., Zervakis, M., Xanthopoulos, P., Sakkalis, V., Vanrumste, B. (2008) Review on solving the inverse problem in EEG source analysis, Journal of NeuroEngineering and Rehabilitation, 5:25.

Green, D.M., Swets, J.A. (1988). Signal Detection Theory and Psychophysics. Los Altos, CA: Peninsula Publishing.

Griffiths, T.D., Warren, J.D. (2004). What is an auditory object? Nature Reviews. Neuroscience, 5(11), 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.

Gutschalk, A. (2005). Neuromagnetic Correlates of Streaming in Human Auditory Cortex. Journal of Neuroscience, 25(22), 5382–5388.

Gutschalk A, Micheyl C, Oxenham AJ (2008) Neural Correlates of Auditory Perceptual Awareness under Informational Masking. PLoS Biol 6(6): e138.

Johnson, B.W., Hautus, M., Clapp, W.C. (2003). Neural activity associated with binaural processes for the perceptual segregation of pitch. Clinical Neurophysiology, 114(12), 2245–2250.

Johnson, B.W., Hautus, M.J., Duff, D.J., Clapp, W.C. (2007). Sequential processing of interaural timing differences for sound source segregation and spatial localization: Evidence from event-related cortical potentials. Psychophysiology, 44(4), 541–551.

Jones, M., Kidd, G., Wetzel, R. (1981). Evidence for rhythmic attention. Journal of Experimental Psychology: Human Perception and Performance, 7, 1059-1073.

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(1), 20–33.

Kubovy, M., van Valkenburg, D. (2001). Auditory and visual objects. Cognition, 80(1-2), 97-126.

Lee, A.K., Shinn-Cunningham, B.G. (2008). Effects of frequency disparities on trading of an ambiguous tone between two competing auditory objects. Journal of the Acoustical Society of America, 123, 4340-4351.

Lipp, R., Kitterick, P., Summerfield, Q., Bailey, P.J., Paul-Jordanov, I. (2010). Concurrent sound segregation based on inharmonicity and onset asynchrony. Neuropsychologia, 48(5), 1417–1425.

Micheyl, C., Carlyon, R.P., Gutschalk, A., Melcher, J.R., Oxenham, A. J., Rauschecker, J.P., Courtenay Wilson, E. (2007). The role of auditory cortex in the formation of auditory streams. Hearing Research, 229(1-2), 116–131.

Micheyl, C., Kreft, H., Shamma, S., Oxenham, A. J. (2013). Temporal coherence versus harmonicity in auditory stream formation. Journal of the Acoustical Society of America, 133(3), 188-194.

Mill, R.W., Bőhm, T.M., Bendixen, A., Winkler, I., Denham, S.L. (2013). Modelling the emergence and dynamics of perceptual organisation in auditory streaming. PLoS Computational Biology, 9(3), 1-21.

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

Pascual-Marqui, R.D., Esslen, M., Kochi. K, Lehmann, D. (2002). Functional imaging with low resolution brain electromagnetic tomography (LORETA): review, new comparisons, and new validation. Japanese Journal of Clinical Neurophysiology 30, 81-94.

Pelli, D.G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies, Spatial Vision 10:437-442.

Polich, J., Kok, A. (1995). Cognitive and biological determinants of P300: an integrative review. Biological Psychology, 41, 103–146.

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. Clinical Neurophysiology, 118(10), 2128–2148.

Rainer, G., Asaad, W.F., Miller, E.K. (1998). Memory fields of neurons in the primate prefrontal cortex. Proceedings of the National Academy of Sciences of the United States of America, 95(25), 15008–13.

Rushworth, M.F., Behrens, T.E., Johansen-Berg, H. (2006). Connection patterns distinguish 3 regions of human parietal cortex. Cerebral Cortex, 16, 1418–1430.

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

Shamma, S., Elhilali, M., Ma, L., Micheyl, C., Oxenham, A.J., Pressnitzer, D., Yin, P., Xu, Y. (2013). Temporal coherence and the streaming of complex sounds. Advance Experimental Medical Biology, 787, 535-43.

Seghier, M.L. (2012). The angular gyrus: multiple function ad multiple subdivisions. Neuroscientist, 19(1), 43–61.

Shinn-Cunningham, B.G. (2008). Object-based auditory and visual attention. Trends Cognitive Sciences, 12(5), 182-186.

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

Snyder, J.S., Alain, C., Picton, T.W. (2006). Effects of attention on neuroelectric correlates of auditory stream segregation. Journal of Cognitive Neuroscience, 18(1), 1–13.

Spielmann, M.I., Schröger, E., Kotz, S.A., Bendixen, A. (2014). Attention effects on auditory scene analysis: insights from event-related brain potentials. Psychological Research 78(3), 361-78.

Straube, S., Grimsen, C., Fahle, M. (2010). Electrophysiological correlates of figure-ground segregation directly reflect perceptual saliency. Vision Research, 50(5), 509–521.

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

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.

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

O'Sullivan, J.A., Shamma, A.S., Lalor, E.C. (2015). Evidence for Neural Computations of Temporal Coherence in an Auditory Scene and Their Enhancement during Active Listening. The Journal of Neuroscience, 35(18):7256-7263.

Teki, S., Chait, M., Kumar, S., von Kriegstein, K., and 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. E life. 2, 00699.

Van Noorden, L.P.A.S. (1975). Temporal coherence in the perception of tone sequences. Unpublished doctoral dissertation, Eindhoven University of Technology.

Wang, L., Liu, X., Guise, K.G., Knight, R.T., Ghajar, J., Fan, J. (2009). Effective Connectivity of the Fronto-parietal Network during Attentional Control. Journal of Cognitive Neuroscience 22:3, pp. 543–553.

Weise, A., Bendixen, A., Müller, D., Schröger, E. (2012). Which kind of transition is important for sound representation? An event-related potential study. Brain Research, 1464, 30–42.

Wilson, E. C., Melcher, J.R., Micheyl, C., Gutschalk, A., Oxenham, A.J. (2007). Cortical FMRI activation to sequences of tones alternating in frequency: relationship to perceived rate and streaming. Journal of Neurophysiology, 97, 2230–2238.

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., Denham, S.L., Mill, R., Bőhm, T.M., Bendixen, A. (2012). Multistability in auditory stream segregation: A predictive coding view. Philosophical Transactions of the Royal Society B, 367, 1001–1012.

Whittingstall K, Stroink G, Gates L, Connolly JF, Finley A. (2003). Effects of dipole position, orientation and noise on the accuracy of EEG source localization. Biomedical Engineering Online, 2:14.

Figure Captions

Figure 1. Schematic illustration of a stimulus including a "figure" component. Black dots depict random tonal components while red represent repeating components. The onsets of the chords are represented as vertical lines. The *x* axis shows both time and the serial position of the chord within the stimulus. Stimuli consisted of 40 chords, each of 50-ms duration, and each containing a random set of 9 to 21 pure tone components. In half of the stimuli, an additional set of 4 or 6 tonal components was repeated 2, 3, or 4 times (resulting in 3, 4, or 5 consecutive chords) to create a "figure" that could be perceptually segregated from the rest of the random chords ("ground"). In the other half of the stimuli, random chords with the same numbers of tonal components were added to the ground ("control"). The figure/control started between 200 –1800 ms from the stimulus onset.

Figure 2. In Experiment 1, detection improved with increasing figure coherence and increasing figure duration, but was worse when the figure and background were separated by a large spatial separation (see text). Group-averaged (N=20) d' values (standard error of mean represented by bars) are shown as a function of figure duration separately for the two coherence levels (marked by the different line types). The three levels of location difference between the figure and the ground are shown in the three separate panels.

Figure 3. In Experiment 2, detection improved with increasing figure coherence and increasing figure duration, consistent with Experiment 1. Group-averaged (N=25) d' values (standard error of mean represented by bars) are shown as a function of figure duration separately for the two coherence levels (marked by the different line types).

Figure 4. Group-average (N=25) ERPs elicited by figure (green lines) and control stimuli (blue lines) triggered from the figure/control segment onset (0 ms at the x axis) at Cz (top of each panel) and at Pz (bottom of each panel) for the 6 stimulus conditions (Coherence: 4 or 6; Duration: 3, 4, or 5). Boxes mark the measurement windows for ORN at Cz and P400 at Pz; a red box indicates that the figure-minus control difference significantly differed from zero ($p<0.05$) within the measurement window, while a grey box indicates no significant amplitude difference. The scalp distribution of the mean difference amplitude within the measurement window is shown to the right of each panel. Color calibration is at the right side of the figure.

Figure 5. Group-average (N=25) ERPs elicited for hit (green lines) and miss trials (blue lines) triggered from the figure segment onset (0 ms at the x axis) at Cz (top of each panel) and at Pz (bottom of each panel) for the 6 stimulus types (Coherence: 4 or 6; Duration: 3, 4, or 5. Boxes mark the measurement windows for ORN at Cz and P400 at Pz; a red box indicates significant amplitude difference ($p<0.05$) between hit and corresponding miss trials within the measurement window, a grey box indicates no significant amplitude difference. Note that due to the low number of hit or miss trials in the Coherence-4/Duration-3 and Coherence-6/Duration-5 conditions, no response amplitudes were measured. The scalp

distribution of the mean hit-minus-miss difference amplitudes within the measurement window is shown to the right of each panel. Color calibration is at the right side of the figure.

Figure 6. LORETA t-value maps from voxel-by-voxel paired t-tests contrasting current density values between figure and control stimuli for the ORN (left) and P400 (right) latency range. Red color corresponds to higher current source density magnitudes (indexed by positive t values) for the figure compared to control trials (color scales are at the bottom of the left and right panels). A) Maps are displayed on the 3D inflated cortex. The 3D inflated cortex plots present the right hemisphere on the top and left hemisphere below. B) Maps shown on the MNI152 standard brain template. Coordinates are scaled in cm; origin is at the anterior commissure; (X) = left (−) to right (+); (Y) = posterior (−) to anterior (+); (Z) = inferior (−) to superior (+). The maps corresponding to the ORN time window (200-350 ms) are shown at the x=-40 mm, y=-25 mm, z=0 mm MNI cordinates; the maps corresponding to the P400 time window (460-600 ms) are shown at the x=30 mm, y=-25 mm, z=15 mm MNI cordinates.

Figure 7. Across individual subjects, the change in the size of the P400 amplitude difference for hit-miss trials (measured at Pz) correlates with figure-detection peformance (d') for four of the six stimulus conditions. The dots represent the different listeners' data. Pearson correlation r values and R2 determination coefficients and p-values are shown on each panel. A regression line is shown on each panel representing the relationship between P400 amplitudes and d'.

8. Table1

Table 1. Group-average (N = 25) central (Cz) ORN (top) and parietal (Pz) P400 amplitudes and peak latencies (bottom) of the figure-minus-control difference waveforms, separately for the six stimulus conditions

		Coherence 4			Coherence 6	
ORN	Duration 3	Duration 4	Duration 5	Duration 3	Duration 4	Duration 5
Mean amplitude at Cz (μV)	-0.37	-0.86	-1.17	-1.30	-1.70	-2.85
SD	1.22	1.72	1.42	1.58	1.90	2.03
t(24)	-1.48	-2.44*	-4.04***	-4.03***	-4.38***	-6.87***
Amplitude measurement window (ms)	200-300	200-300	232-332	172-272	200-300	232-332
ORN peak latency	258.08	263.04	272.16	242.40	268.80	273.28
SD	5.61	6.37	6.02	4.76	5.05	6.34
P400						
Mean amplitude at Pz (μV)	0.33	1.60	4.08	1.58	4.35	6.79
SD	1.39	2.04	2.76	1.72	2.93	4.05
t(24)	1.16	3.84***	7.23***	4.48***	7.27***	8.23***
Amplitude measurement window (ms)	452-552	520-620	580-680	500-600	480-580	480-580
P400 peak latency	554.08	561.92	556.32	542.24	545.12	536.80
SD	8.14	6.09	7.46	6.61	7.48	6.99

Notes: Significant differences from zero are marked by asterisks (p<.05,***p<.001)*

8. Table2

Table 2. Group-average (N = 25) central (Cz) ORN (top) and parietal (Pz) P400 amplitudes and peak latencies (bottom) of the hit-minus-miss difference waveforms, separately for the four tested stimulus conditions

	Coherence 4		Coherence 6	
ORN	Duration 4	Duration 5	Duration 3	Duration 4
Mean amplitude at Cz (μV)	-0.84	-2.57	-0.02	-2.03
SD	1.89	2.24	1.71	1.71
t(24)	-2.17*	-5.62***	-0.04	-5.82***
Amplitude measurement window (ms)	200-300	240-340	200-300	200-300
P400				
Mean amplitude at Pz (μV)	4.10	4.67	3.51	5.40
SD	2.87	3.89	3.47	3.81
t(24)	6.99***	5.88***	4.96***	6.95***
Amplitude measurement window (ms)	552-652	500-600	472-572	500-600

Notes: Significant differences from zero are marked with asterisks (p<.05,***p<.001); due to the low number of Coherence-4/Duration-3 hit trials and Coherence-6/Duration-5 miss trials (<30% of all trials), the ERP measures are not reliable for these conditions.*

8. Table3

Table 3. Summary of significant differences of LORETA-based estimates of neural activity for figure versus control in the Coherence 6 conditions in the time ORN window (200-350 ms). The anatomical regions, MNI coordinates, and BAs of maximal t-values are listed.

Region	BA	MNI coordinates (mm)			voxels (N)	t value	p value
		x	y	z			
Transverse Temporal Gyrus	41	40	-25	10	3	1,33	<0.001
Superior Temporal Gyrus	39	45	-60	30	1	1,27	<0.001
Angular Gyrus	39	50	-60	30	1	1,26	<0.001
Anterior Cingulate	25	0	0	-5	3	1,55	<0.001
Parahippocampal Gyrus	25, 27, 28, 30, 34. 35	0	-35	0	27	1,41	<0.001

Note: Positive t-values indicate stronger current density for figure than for control trials. The numbers of voxels exceeding the statistical threshold ($p < 0.01$) are also reported. The origin of the MNI space coordinates is at the anterior commissure; (X) = left (−) to right (+); (Y) = posterior (−) to anterior (+); (Z) = inferior (−) to superior (+).

8. Table4

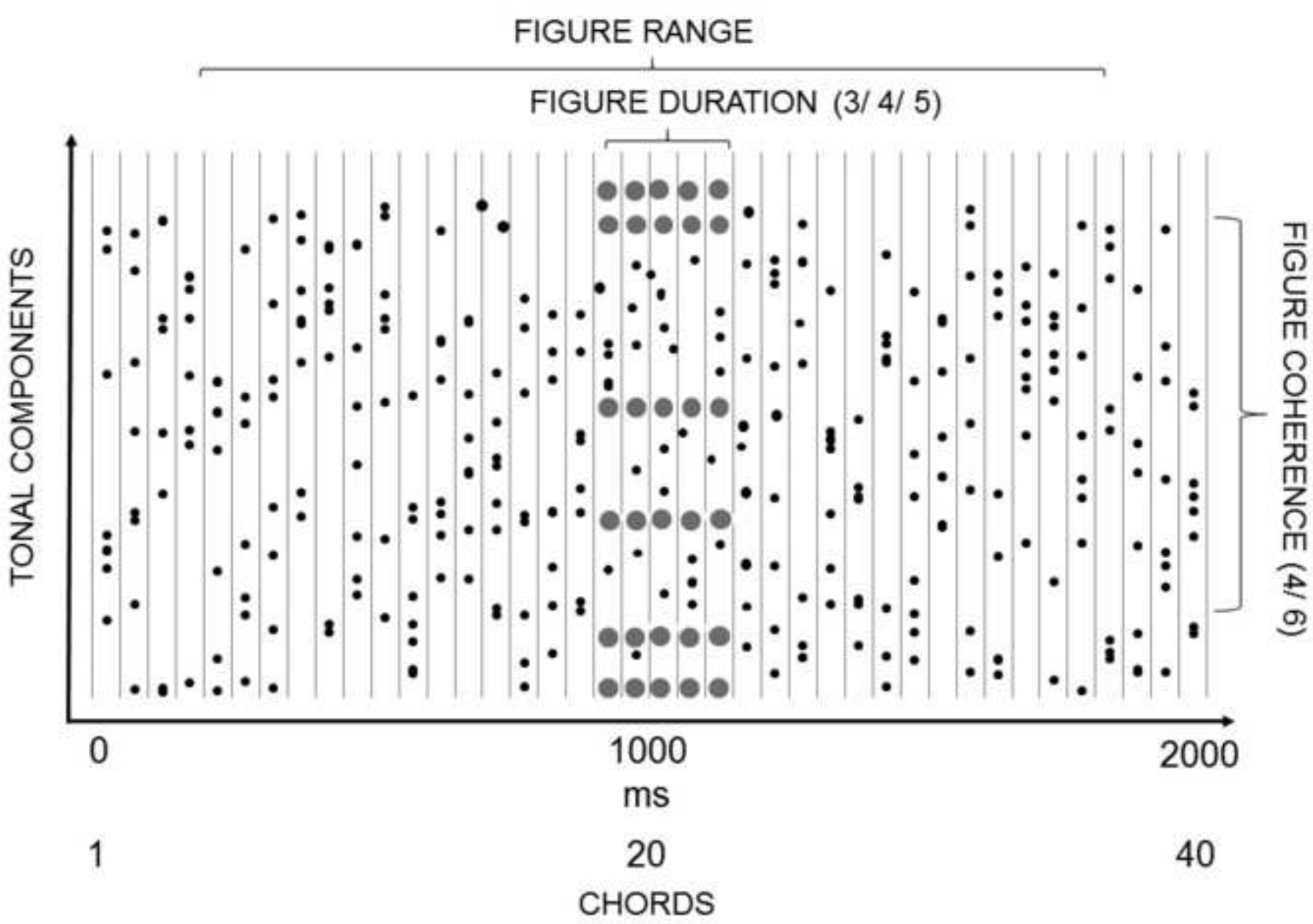
Table 4. Summary of significant differences of LORETA-based estimates of neural activity for figure versus control in the Coherence 6 conditions in the P400 time window (460-600 ms). The anatomical regions, MNI coordinates, and BAs of maximal t-values are listed

Region	*BA*	*MNI coordinates (mm)*			*voxels (N)*	*t value*	*p value*
		x	y	z			
Superior Temporal Gyrus	41	40	-40	10	1	1.86	<0.001
Medial Frontal Gyrus	6,32	0	5	50	20	2.01	<0.001
Paracentral Gyrus	5,31	-15	-40	50	7	1.91	<0.001
Superior Frontal Gyrus	6	0	5	55	21	1.99	<0.001
Cingulate Gyrus	23,24,31,32	0	-40	25	178	2.38	<0.001
Anterior Cingulate Gyrus	33	5	10	25	5	1.98	<0.001
Posterior Cingulate Gyrus	23, 29, 30,31	5	-40	25	50	2.38	<0.001
Parahippocampal Gyrus	27, 30	10	-35	0	13	2.06	<0.001
Cuneus	7,18,19	0	-75	20	138	2.21	<0.001
Precuneus	7,19,31	0	-50	30	152	2.23	<0.001
Middle Occipital Gyrus	18	-15	-90	15	10	1.94	<0.001

Note: Positive t-values indicate stronger current density for figure than for control trials. The numbers of voxels exceeding the statistical threshold ($p < 0.01$) are also reported. The origin of the MNI space coordinates is at the anterior commissure; (X) = left (−) to right (+); (Y) = posterior (−) to anterior (+); (Z) = inferior (−) to superior (+).

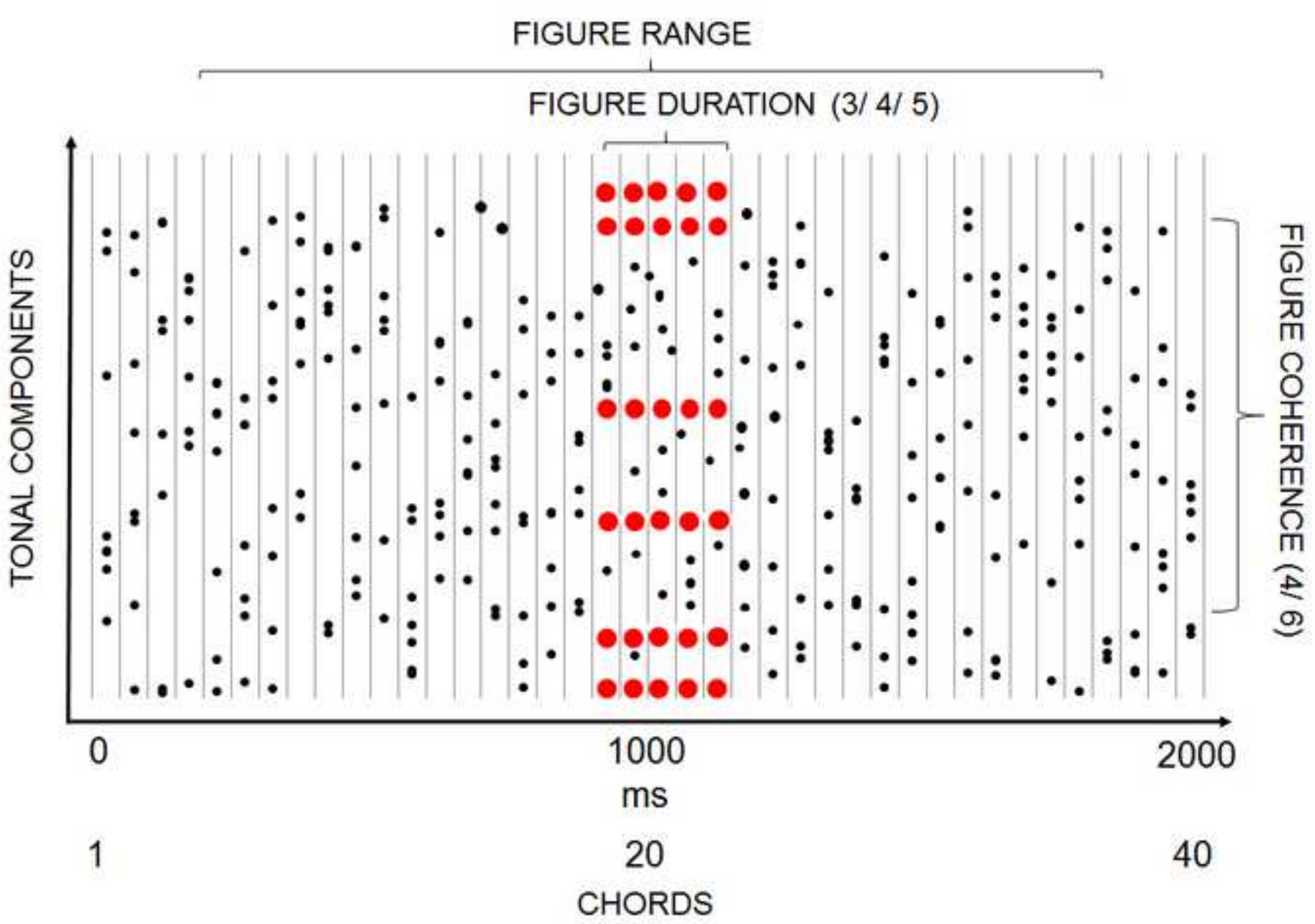
9. Figure1

Click here to download high resolution image



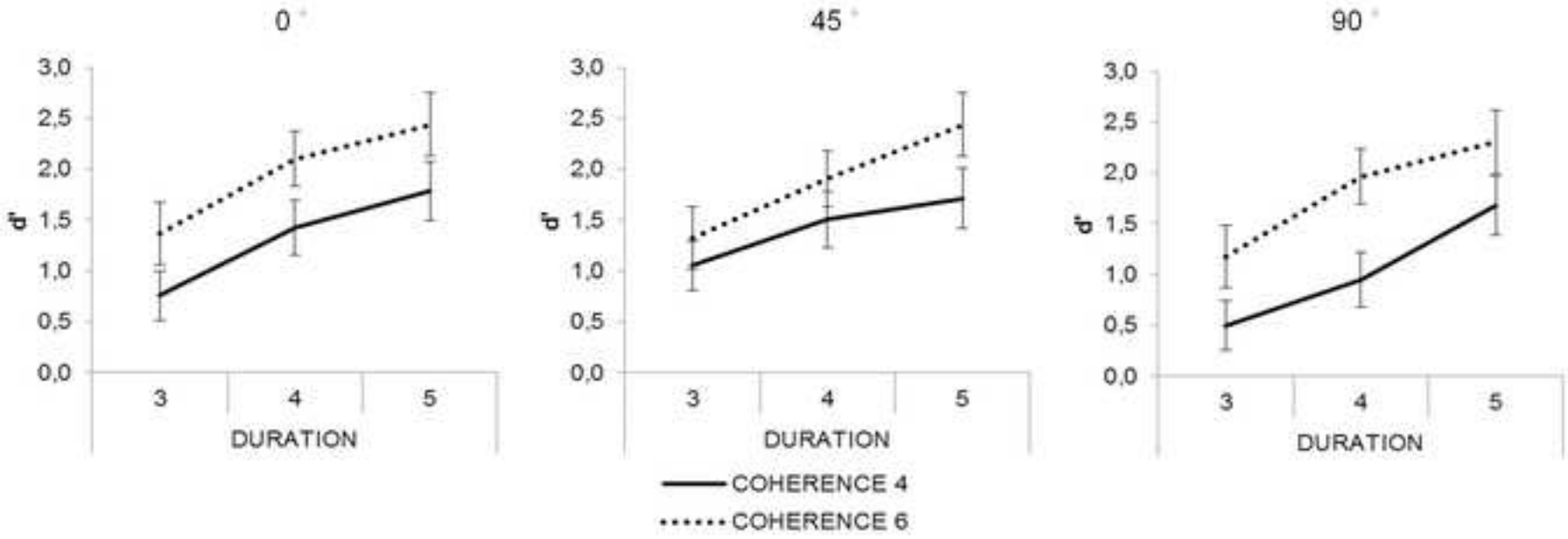
9. Figure1_RGB

Click here to download high resolution image



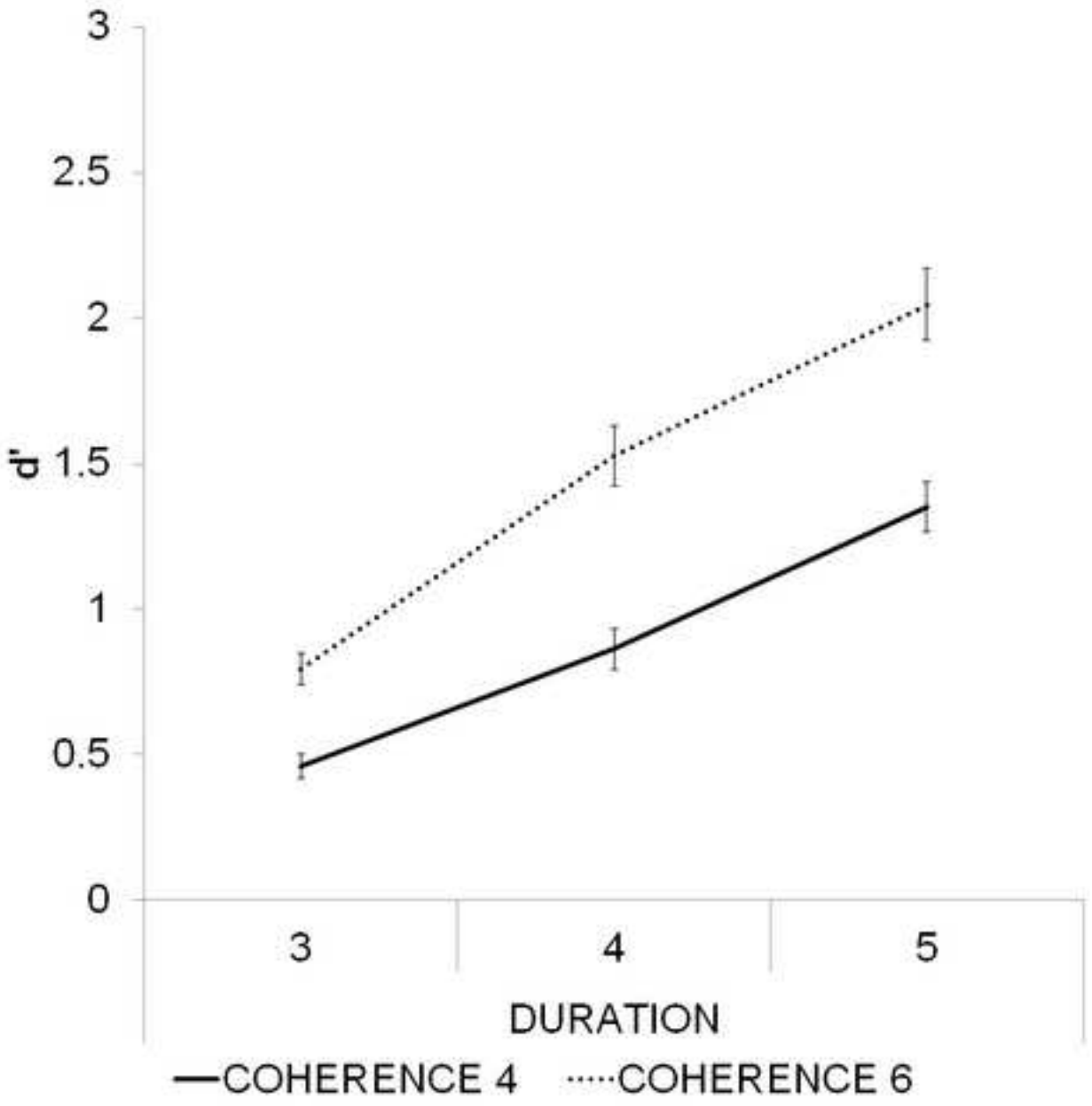
9. Figure 2

Click here to download high resolution image



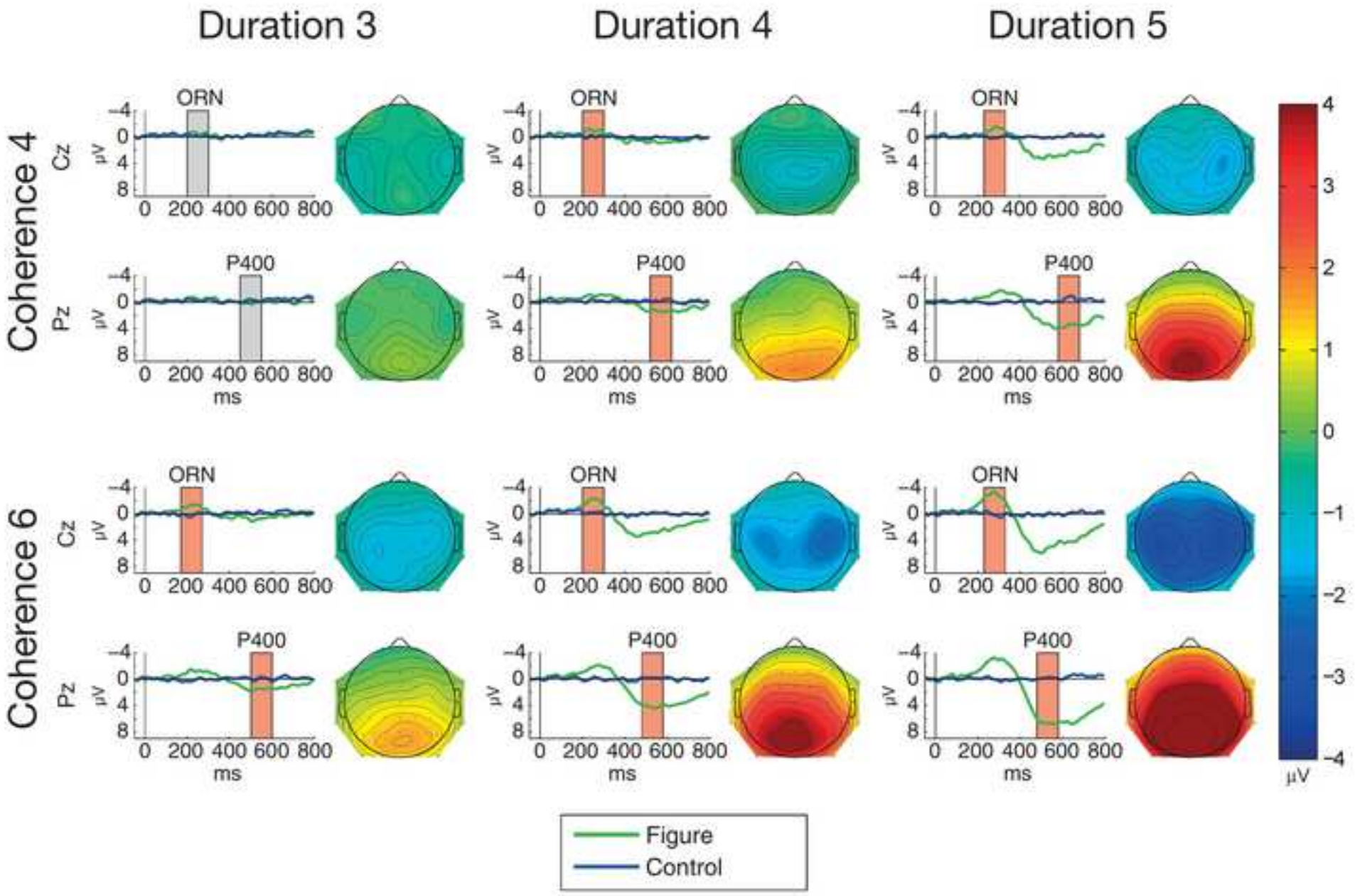
9. Figure 3

Click here to download high resolution image



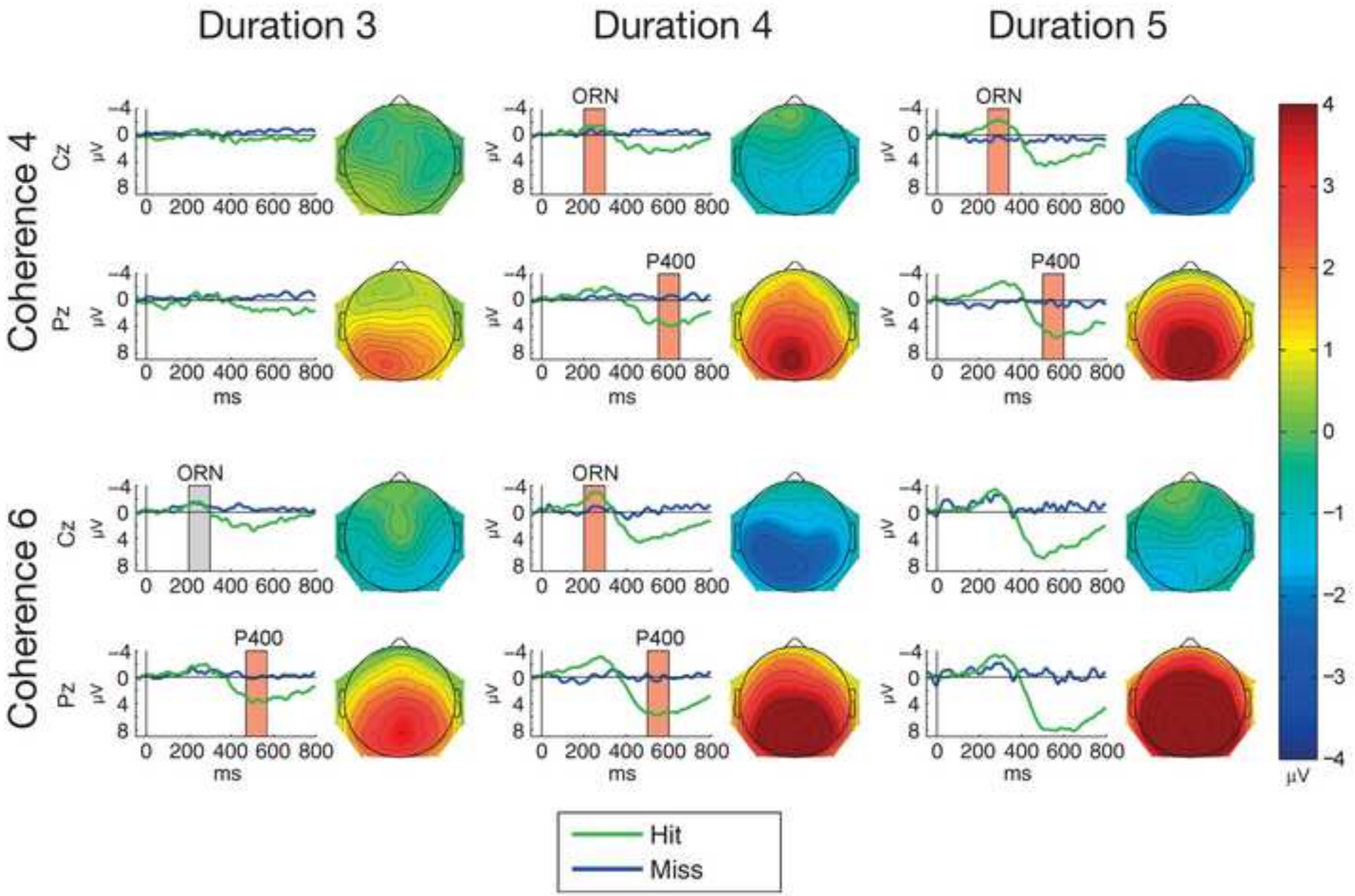
9. Figure4

Click here to download high resolution image



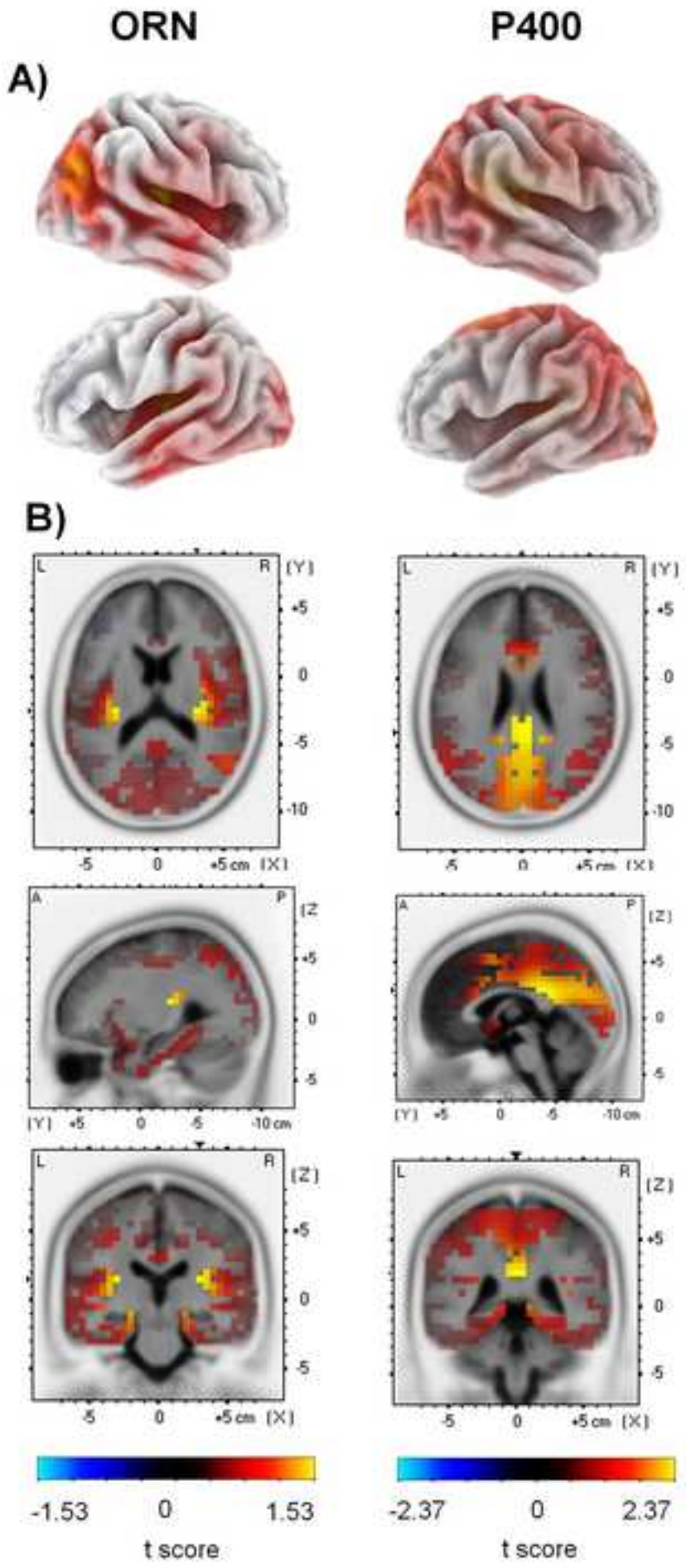
9. Figure5

Click here to download high resolution image



9. Figure6

Click here to download high resolution image



9. Figure 7

Click here to download high resolution image