Effects of multiple congruent cues on concurrent sound segregation during passive and active listening: An event-related potential (ERP) study

Kocsis, Z.[1,2], Winkler, I. [1,3], Szalárdy, O. [1], & Bendixen, A. [4,5]

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

[2]Budapest University of Technology and Economics, Budapest, Hungary

[3]Institute of Psychology, University of Szeged, Szeged, Hungary

[4]Department of Psychology, Cluster of Excellence "Hearing4all", European Medical School, Carl von Ossietzky University of Oldenburg, Oldenburg, Germany

[5]Department of Psychology, University of Leipzig, Leipzig, Germany

Corresponding author:

Zsuzsanna Kocsis

Institute of Cognitive Neuroscience and Psychology, Research Centre for Natural Sciences, Hungarian Academy of Sciences

Magyar tudósok körútja 2.

Budapest, Hungary

H-1117

Tel.: +3613826809

E-mail address: kocsis.zsuzsanna@ttk.mta.hu

Acknowledgments

This work was funded by the Hungarian Academy of Sciences (Magyar Tudományos Akadémia [MTA], Lendület project LP2012-36/2012 to I.W.), by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG Cluster of Excellence 1077 "Hearing4all"), by the German Academic Exchange Service (Deutscher Akademischer Austauschdienst [DAAD], Project 56265741), and by the Hungarian Scholarship Board (Magyar Ösztöndíj Bizottság [MÖB], Project 39589). The experiment was realized using Cogent 2000 developed by the Cogent 2000 team at the FIL and the ICN. EEG data were analyzed with EEGlab (Delorme & Makeig, 2004) and additional plugins written by Andreas Widmann, University of Leipzig. The authors are grateful to Zsuzsanna D'Albini for assistance in data acquisition.

Abstract

In two experiments, we assessed the effects of combining different cues of concurrent sound segregation on the object-related negativity (ORN) and the P400 event-related potential components. Participants were presented with sequences of complex tones, half of which contained some manipulation: One or two harmonic partials were mistuned, delayed, or presented from a different location than the rest. In separate conditions, one, two, or three of these manipulations were combined. Participants watched a silent movie (passive listening) or reported after each tone whether they perceived one or two concurrent sounds (active listening). ORN was found in almost all conditions except for location difference alone during passive listening. Combining several cues or manipulating more than one partial consistently led to sub-additive effects on the ORN amplitude. These results support the view that ORN reflects a combined, feature-unspecific assessment of the auditory system regarding the contribution of two sources to the incoming sound.

Keywords:

object-related negativity (ORN); P400; multiple congruent cues; concurrent sound segregation; active and passive listening

Introduction

In everyday situations, we are constantly confronted with mixtures of sounds emitted by concurrently active sources. The human auditory system needs to parse this mixture to allow us to perceive the world in terms of meaningful objects and events. Cues that support the parsing process are traditionally divided into two main categories (Bregman, 1990; Carlyon, 2004; Haykin and Chen, 2005; Snyder and Alain, 2007): those that group together sound elements along time (horizontal or sequential sound organization) and those that group them at one particular moment of time (vertical or concurrent sound organization). Concurrent segregation is based on instantaneously available cues, such as differences in pitch, sound onset, and source location. Whereas no direct event-related potential (ERP) correlate of sequential segregation has been discovered yet, concurrent segregation appears to have such an ERP correlate: The object-related negativity (ORN) component has been shown to follow the listener's perception of two concurrent sounds (Alain, Arnott, & Picton, 2001). The present study was designed to systematically investigate how combinations of the three most well-known cues of concurrent sound segregation (different source location, onset asynchrony, and inharmonic relation between the partials of complex tones) affect the ORN component. Specifically, we wished to assess whether ORN sums together the outputs of three independent detectors of concurrent sound segregation, or whether it is a read-out of the system's overall assessment of the likelihood that the sound input carries contributions from two sound sources[1].

The ORN peaks between 150 and 180 ms from cue onset, reaches its maximum at frontocentral electrode sites, and inverts polarity at the mastoids (Alain, Schuler, and McDonald, 2002; Alain and McDonald, 2007). Alain and colleagues (2001) found that ORN was larger at the mastoid electrodes during active listening (listeners were required to judge whether they heard one or two concurrent sounds) than passive listening situations (listeners had no task related to the sounds), indicating attentional modulation of the ORN amplitude.

The presence and amplitude of ORN is correlated with manipulations that typically lead to listeners reporting two sound sources compared to one (Alain, Theunissen, Chevalier, Batty, and Taylor, 2003; McDonald and Alain, 2005; Alain and McDonald, 2007). Previous studies

[1] Another possibility is that the amplitude of the ORN reflects the number of perceived auditory objects, although modulation of the ORN amplitude by the amount of mistuning (Alain et al., 2001) makes this alternative unlikely.

have shown that ORN can be elicited by different cues, such as inharmonicity (Alain et al., 2001 and 2002; Bendixen, Jones, Klump, and Winkler, 2010), onset asynchrony (Lipp, Kitterick, Summerfield, Bailey, and Paul-Jordanov, 2010; Weise, Schröger, and Bendixen, 2012); dichotic pitch (Johnson, Hautus, and Clapp, 2003; Hautus, Johnson, and Colling, 2009), separation in the fundamental frequency of speech sounds (Snyder and Alain, 2005; Alain, Reinke, He, Wang, and Lobaugh, 2005), and simulated echo (Sanders, Joh, Keen, and Freyman, 2008; Sanders, Zobel, Freyman, and Keen, 2008). There are also some reports of ORN emerging with a combination of some of the above cues, such as inharmonicity and location difference (McDonald and Alain, 2005) or inharmonicity and onset asynchrony (Weise et al., 2012).

ORN is elicited in both passive and active listening situations (Alain et al., 2001, 2002; Alain, 2007) and its amplitude is independent of the task demands (Alain and Izenberg, 2003). In active listening situations, ORN elicitation is accompanied by a late positive wave that peaks about 400 ms after stimulus onset, the P400 component. P400 amplitude also correlates with the likelihood of perceiving two concurrent sound objects compared to one (Alain et al., 2001, 2002; Hautus & Johnson, 2005), but P400 does not follow the ORN in an obligatory manner (Johnson, Hautus, Duff, and Clapp, 2007). Johnson et al. (2007) proposed that P400 is influenced by the task context. In their study, they used two different tasks: In the detection task, participants were to indicate whether they heard dichotic pitch or a control stimulus, whereas in the localization task, only dichotic pitch stimuli were presented, and participants were to decide where the sound was located. In the latter case, no P400 was elicited (Johnson et al., 2007).

Whereas ORN is assumed to reflect an automatic process of detecting the difference between the physical features (e.g., frequency) extracted from the incoming stimulus and a template of the complex sound (e.g., based on its fundamental frequency), P400 appears to reflect a controlled process that uses prior knowledge to extract meaning from the incoming auditory information (Alain et al., 2002; Hautus & Johnson, 2005; Johnson et al., 2007). Studies showing that ORN is not only elicited by harmonic cues suggest that the template underlying ORN also includes information about the timing and source location of the partials of complex sounds.

Previous studies suggested that the ORN amplitude is modulated by the strength or salience of the cues supporting the segregation of concurrent sounds. For example, Alain and colleagues

(2001) found larger ORN amplitudes with increasing amounts of inharmonicity (larger ORN amplitude for the 16% than 8%, or 4% mistuning). In this study, participants reported hearing two sounds more often with higher amounts of mistuning. In another paradigm, using dichotic pitch, Clapp, Johnson, and Hautus (2007) found that the largest ORN was elicited in response to the most salient dichotic pitch cue, and the ORN amplitude decreased with decreasing cue saliency. These authors also found a similar pattern for P400 amplitude.

Perception of concurrent sounds can be made more likely not only by strengthening one particular cue (e.g., increasing the amount of mistuning for inharmonicity-based segregation, cf. Alain et al., 2001), but also by combining two different cues (e.g., frequency and location). In this case, the multiple congruent cues may strengthen the impression of the presence of separate sound sources. Using MEG, such a combined effect was found in a speech segregation task (Du, He, Ross, Bardouille, Wu, Li, and Alain, 2011). Du and colleagues (2011) hypothesized that separation in both base frequency and source location contribute to speech segregation, and combining these cues would result in additivity or superadditivity between the ORN components elicited by the two cues, separately. They found that the ORN elicited by the combination of the base-frequency separation and the location cue equaled the sum of the responses elicited by the two cues alone. A similar effect of summing two different types of cues was obtained by Hautus and colleagues (2009), although these authors did not directly test whether the effect of the cue combination was strictly additive when compared to the sum of the effects of the single cues alone (see also McDonald and Alain, 2005; Weise et al., 2012).

Here we report the results of a study in which we systematically investigated combinations of inharmonicity, onset asynchrony, and location difference under passive (Experiment 1) and active (Experiment 2) listening conditions. Based on previous studies (Alain et al., 2001, 2002, 2003; McDonald & Alain, 2005), we expected that ORN will be present in both listening situations, whereas P400 will only be present in the active listening situation. First, we tested whether the salience of the harmonicity-based cue can be further increased by mistuning two partials in a congruent manner (as opposed to mistuning only one partial). We hypothesized that mistuning two partials would enhance the ORN amplitude by providing redundant information for harmonicity-based segregation. Second, we aimed to assess the effects of combining different cues of concurrent segregation on ORN and P400. As the cues are congruent in supporting the same decomposition of the input into two sounds in

perception, we regard them as redundant with respect to concurrent sound segregation. By investigating whether the effects of combined cues are additive, sub- or superadditive compared to the single-cue effects, our goal was to separate two possible interpretations of the ORN component. It is possible that each cue elicits a separate ORN response and the observed response sums together the individual ORN components. In this case, 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 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, depending on the way the cues are combined, we should find sub- or superadditivity between the contributing cues' ORN components. Subadditivity of the contributing cues' ORN amplitudes would occur for example if the cue-combination algorithm evaluated the likelihood of the presence of two concurrent sounds by passing on the signal resulting from the most salient cue. Superadditivity of the ORN amplitudes would occur if cue combination took into account partial cues, which alone would not be sufficient to support the presence of two concurrent sounds. The two methods are not mutually exclusive; thus one may find both sub- and superadditivity depending on the strength of the available contributing cues. Either one of these possibilities would mean that the process reflected by ORN is less directly related to cue evaluation; rather it is closer to what appears in perception.

EXPERIMENT 1

Methods

Participants

Twenty healthy volunteers (eight female, mean age 23.5 years, SD = 2.42) participated in the experiment. Participants received modest financial compensation. None of the participants were taking any medication affecting the central nervous system. Prior to the beginning of the experiment, written informed consent was obtained from each participant according to the

Declaration of Helsinki after the experimental procedures and aims of the study were explained to them. The study was approved by the Ethical Committee of the Institute of Cognitive Neuroscience and Psychology, Research Centre for Natural Sciences, Hungarian Academy of Sciences.

Apparatus, stimuli, and procedure

The study was conducted in a sound-attenuated experimental chamber at the Institute of Cognitive Neuroscience and Psychology, Research Centre for Natural Sciences, Hungarian Academy of Sciences.

Complex tones with an intensity of 40 dB sensation level (above hearing threshold, adjusted individually for each participant) were presented binaurally via headphones with a 1100 ms onset-to-onset interval. In each stimulus block, 2 types of tones were presented in random succession with equal probabilities: the "base" tone, a fully harmonic tone of 250 ms duration (including 10 ms rise and 10 ms fall times) comprising the 5 lowest partials (all having the same amplitude and starting in sine phase), and a manipulated version of this tone. The manipulated tones had the same base frequency (see below) and duration as their base versions. The manipulations were administered either to one (the 2^{nd}) or two (2^{nd} and 4^{th}) partials. Three simple manipulations and their combinations (altogether 11 different manipulations) were tested. The simple manipulations were: a) mistuning the 2^{nd} partial (or 2^{nd} and 4^{th} partials) by +8 %, or b) delaying the same partial(s) by 100 ms (but ending at the same time as the other partials), or c) delivering the same partial(s) with a different interaural time (ITD) and level difference (ILD) compared to those of the other harmonics (location difference). For the purpose of adding the location manipulation in some of the conditions without making these conditions stand out from the other conditions, the location of each individual tone could take one of two positions in each condition, regardless of whether a location manipulation was applied or not. Hence in each condition, half of the tones were presented with parameters promoting the listener to hear the tones as originating from ca. 45˚ right and the other half from 45˚ left from the midline (ITD of +/- 200 μs and ILD of +/-5 dB, applied congruently). Tones with the two perceived locations were delivered in a fully randomized order, which was independent from the manipulation (i.e., the probability of a manipulation was equal for the left and right tones). Thus in all conditions, base-left, base-

right, manipulated-left, and manipulated-right tones each made up 25% of the stimuli. The location difference cue for the manipulated partial was set up by employing the opposite parameter combination than for the rest of the partials, thus creating a ca. 90° location difference between the manipulated and the other partials. A summary of the experimental manipulations is given in Table 1.

----------Insert Table 1 about here ----------

Each stimulus condition was presented in a separate block consisting of 200 base and 200 manipulated tones. Stimulus blocks commenced with 10 base-version tones, which were excluded from the analyses. In each stimulus block, all tones had the same fundamental frequency, whilst the fundamental frequency changed from block to block. Eleven fundamental frequencies were used with the lowest frequency being 200 Hz, and the rest following in one-semitone steps (i.e., the highest fundamental frequency being 378 Hz). The order of the fundamental frequencies and the order of the different stimulus blocks (conditions) were randomized separately for each participant.

Participants watched a silent, subtitled movie of their own choice on a computer screen that was placed in front of them at a distance of 1.15 m. They were asked to ignore the sounds. Total duration of the experimental blocks amounted to 83 minutes. Short breaks were inserted between stimulus blocks with at least one longer break, set between the 6th and the 7th stimulus block, when the participant was allowed to leave the chamber. Further longer breaks were inserted if the participant asked for it. The total time of the session (including electrode mounting and removal) was ca. 3 hours.

Electrophysiological recording and data analysis

Electroencephalogram (EEG) was continuously recorded with Ag/AgCl electrodes. 63 electrodes were placed on the scalp according to the extended international 10–20 system (Jasper, 1958, Chatrian, Lettich and Nelson, 1985). An additional electrode was placed on the tip of the nose, which served as the reference. Eye movements were monitored by bipolar

recordings from two electrodes placed above and below the left eye (vertical electrooculogram, VEOG) and two placed lateral to the outer canthi of both eyes (horizontal electrooculogram, HEOG). EEG and EOG signals were amplified (0–40 Hz) by SynAmps amplifiers (Neuroscan Inc.), sampled at 500 Hz. Data were resampled to 250 Hz and filtered off-line using a 0.1–30 Hz band-pass finite impulse response (FIR) filter (Kaiser windowed, Kaiser $\beta = 5.65$, filter length 4530 points).

For each tone, an epoch of 400-ms duration including a 100 ms pre-stimulus baseline was extracted from the continuous EEG record. Epochs with an amplitude change exceeding 100 μV at any electrode were rejected from further analysis, which led to retaining 84.0% of the responses on average. Epochs for the two stimulus types (base version and manipulated) were separately averaged for each of the 11 conditions, collapsing over the two possible locations (left vs. right presentation).

Difference waveforms were calculated between ERPs elicited by the manipulated and the corresponding base tones for identifying and measuring the ORN component. Except for the scalp topography analyses, all measurements were taken from the recordings at the Cz electrode. Average ORN amplitudes were measured from 72-ms wide windows centered on the average peak latency for each condition. To account for the observed latency variation between conditions, peak latencies were determined separately for each condition by the jackknifing method (Kiesel, Miller, Jolicœur & Brisson, 2008; Miller, Ulrich & Schwarz, 2009). Epochs for the base versions and manipulated tones were averaged separately.

Following visual inspection of the responses, the N1 amplitude differences were also investigated. N1 difference amplitudes were measured from 40-ms wide windows centered on the average peak latency for each condition.

All ERP difference amplitudes were tested against zero using one-sample, two-tailed *t* tests. For testing possible differences in the ORN amplitudes and scalp distributions across the three single-cue manipulations, ORN amplitudes were averaged separately for the following six electrode clusters: left frontal (Fp1, AF7, AF3, F7, F5, F3, F1), left central (FT7, FC5, FC3, FC1, C5, C3, C1), left parietal (CP5, CP3, CP1, P7, P5, P3, P1), right frontal (Fp2, AF8, AF4, F8, F6, F4, F2), right central (FT8, FC6, FC4, FC2, C6, C4, C2), right parietal (CP6, CP4, CP2, P8, P6, P4, P2). ORN amplitudes and scalp topographies were then compared by a repeated-measures ANOVA with the factors *Manipulation* (3 levels: 2nd partial mistuned vs.

2^{nd} partial delayed vs. 2^{nd} partial with location difference) × *Frontality* (3 levels: frontal vs. central vs. parietal) × *Laterality* (2 levels: left vs. right).

The effects of providing multiple congruent cues were tested by a repeated-measures ANOVA of the ORN amplitudes at Cz with the factors *Number of mistuned partials* (2 levels: 2^{nd} partial vs. 2^{nd} and 4^{th} partials) × *Delay* (2 levels: delay present vs. absent) × *Location difference* (2 levels: location difference present vs. absent). Additivity between cue effects was tested with paired two-tailed *t* tests comparing the multiple-cues ORN amplitudes with the summed amplitudes of the ORN components elicited by the corresponding cues.

All significant statistical results are reported. ANOVA effects are reported together with the partial η^2 effect size measure. The Greenhouse–Geisser correction was applied when the assumption of sphericity was violated; the ε correction factor is reported in these cases. Post hoc tests for repeated-measures ANOVAs were carried out with the Bonferroni correction of the confidence level for multiple comparisons.

Results

ORN

ERP responses elicited by the base and the manipulated tones as well as the corresponding difference waveforms are shown in Figure 1 for all experimental conditions at Cz. In the conditions with delay, the delayed partials commenced 100 ms later, causing the resulting ORN to be delayed. ORN amplitudes were found to be significant in almost all conditions, except for the condition where the 2nd partial was presented with location difference alone (see Figure 1 and Table 2 for the full list of results).

----------Insert Figure 1 about here----------

The ANOVA comparing ORN amplitudes and topographies across the three single cues showed a significant main effect of *Frontality* [$F(2,38) = 12.840$, $p = 0.001$, $\eta^2 = 0.403$, $\varepsilon = 0.63$], which was due to significantly larger amplitudes at frontal ($p = 0.015$) and central ($p <$

0.001) than parietal electrodes. This verifies the typical topography pattern of ORN (e.g. Alain et al., 2002). Importantly, there was no significant interaction between *Manipulation* and either one of the topography factors (*Frontality* or *Laterality*), suggesting that ORN topography did not significantly differ between the 3 types of manipulation. The scalp topographies are shown in Figure 2, top row.

----------Insert Table 2 about here ----------

----------Insert Figure 2 about here----------

In the ANOVA assessing the effects of providing multiple congruent cues, no significant effects or interactions were observed for any of the experimental manipulations (*Number of mistuned partials, Delay* or *Location difference*), all p values > 0.07. This means that adding delay and/or location difference, and/or mistuning more than one partial, did not significantly change the ORN amplitude as compared to that elicited by mistuning only one partial, the most commonly used condition for studying ORN.

Multiple congruent cues always elicited numerically smaller ORN amplitudes than the sum of the ORN amplitudes elicited by the contributing cues, although the differences did not reach significance in each case. The sum of the contributing cues' ORN amplitude values, the corresponding multiple-cue ORN amplitude and the results of the additivity tests are given in Table 3 for each comparison.

----------Insert Table 3 about here ----------

N1

Significant differences between the base and the manipulated tones in the N1 latency range (i.e., preceding the ORN) were found for several conditions, mostly those where the

manipulated partials were delayed and/or had a different location than the rest: 2nd partial delayed, 2nd partial mistuned and delayed, 2nd and 4th partials mistuned and delayed, 2nd partial with location difference, 2nd partial delayed with location difference and 2nd and 4th partials mistuned, delayed and with location difference (see Table 2 for a full list of the mean difference amplitudes and the results of the corresponding *t* tests against zero).

Discussion

In Experiment 1, we studied the ORN components elicited by three different cues of concurrent sound segregation and their combination in a passive listening situation. All of these cues and cue combinations elicited significant ORN components, except for location difference alone, which appeared to be a weaker cue of ORN elicitation with the current parameters.

We found no significant increase of the ORN amplitude when congruently manipulating multiple partials, i.e., there was no significant difference between those conditions where only one partial was manipulated as compared to those conditions where two partials were manipulated. Similarly, adding delay and/or location difference on top of mistuning did not lead to a significant increase in the ORN amplitude. Furthermore, combining several cues always elicited numerically (and in most cases significantly) smaller ORN amplitudes than the sum of the contributing ORN amplitudes. In other words, multiple congruent cues were processed in a subadditive manner. We found no evidence pointing towards superadditivity for any of the combinations, nor did any of the combinations appear to follow a strictly additive model. Note that the amount of mistuning employed in the current study (+8%) did not force a ceiling effect on the ORN amplitude, as a previous study found an increase of the ORN amplitude by increasing the amount of mistuning from 8 to 16% (Alain et al., 2001). Taken together, these results suggest that ORN may reflect a combined assessment of the likelihood of the presence of two concurrent sounds, as opposed to summing the strength of sensory evidence for the presence of two concurrent sounds.

Some effects of the cues of concurrent sound segregation were observed in a latency range preceding that of the ORN. Specifically, significantly larger N1 components were elicited in conditions where one or two partials were delayed or presented with location difference and in some of the conditions where these cues appeared in combination. These results were

unexpected, and the current paradigm was not designed to separate whether the N1 increase was related to the presence of two concurrent sounds or to the specific acoustic manipulations. A test of this issue was therefore included in the follow-up Experiment 2. We introduced control blocks in which only one sound object was delivered at any time. To control for the delay, we tested tones with the two (2^{nd} and 4^{th}) partials omitted; thus the initial 100-ms segment was identical to the delay manipulation, but no additional partials commenced after 100 ms as that would promote concurrent sound segregation. To control for the location of the tones, we recorded responses separately for the two source locations, using only the base versions of the tones. If the acoustic manipulations accounted for the N1 effect, the difference would be apparent between the base versions presented in the two different locations.

Besides these control blocks, the main purpose of the follow-up Experiment 2 was to repeat the manipulations employed in Experiment 1 in an active listening situation. Previous studies (Alain et al., 2001; 2002) have shown that ORN is also elicited during active listening, and that it is followed by a late positive peak (the P400) when listeners are asked to give perceptual judgments as to the presence of one or two concurrent sounds. Thus in Experiment 2 we investigated whether a) a similar pattern for the processing of the cues and their combinations is observed when participants are asked to attend to the sounds, b) whether this pattern translates into perceptual judgments of the sounds as coming from one or two sources, c) how the different cues and combinations affect the P400 response. Finally, we also assessed d) whether attention affects the ORN amplitude with multiple concurrent cues.

EXPERIMENT 2

Methods

Participants

Twenty-three healthy volunteers (twelve female, mean age 22.1 years, SD = 1.62) participated in the experiment. None of the participants had taken part in Experiment 1.

Apparatus and stimuli

The stimulus paradigm employed in Experiment 2 was identical to that of Experiment 1 with the following exceptions.

In the beginning of the experimental session, two N1 control blocks were administered. In one stimulus block, two partials (2nd and 4th) were omitted from the complex tone taking the place of the manipulated sounds of the corresponding stimulus block of Experiment 1 (50%), whereas the base version of the sound stayed the same (50%). The order of the base and manipulated sounds was fully randomized. In the other control block, only the base versions of the left and right tones were delivered. As in Experiment 1 as well as all other stimulus blocks of Experiment 2, in the two control blocks, half of the tones were presented with parameters promoting the listener to perceive the tones as originating from ca. 45° right and the other half from 45° left from the midline; tones with the two perceived locations were delivered in a fully randomized order. All other stimulus parameters were identical to those of Experiment 1. During these control blocks, participants watched a subtitled, silent movie and were asked to disregard the sounds.

For the remainder of the session, participants were given two response keys (one in each hand), and were instructed to perform tasks as detailed below by pressing one or the other key with their left or right thumb.

The next (3rd) stimulus block served as control for a different analysis, which is not reported here. In this stimulus block, half of the sounds were base-version complex tones, whereas for the other half, the 2nd and 4th partials were mistuned, delayed, and with location difference as described for Experiment 1 (condition 11). 140 stimuli of the base version and 140 of the manipulated version were delivered with an onset-to-onset interval of 1400 ms. Participants were instructed to watch a fixation cross continuously present at the center of the computer screen placed at 1.15 m directly in front of them (visual angle of 0.4°) and to press either one of the responses buttons when the fixation cross changed to an "X" for 100 ms, after which it returned to the regular "+" sign. The change appeared at a random time point between 550 and 750 ms after each tone onset. Participants were asked to ignore the tones.

Participants then received two blocks of training (blocks 4 and 5) in the task they were asked to do during the rest of the stimulus blocks. In the first training block, 20 base-version tones

and 20 tones with both the 2nd and 4th partials mistuned, delayed, and with location difference (see Experiment 1, condition 11) were delivered in a randomized order. In the second training block, 40 tones were presented in a randomized order, 10 of which were of the base version, 10 with the 2nd partial being mistuned, 10 with the 2nd partial being delayed, and 10 with the 2nd partial with location difference. The participants' task was to mark for each tone whether he/she perceived one or two concurrent sounds by depressing one or the other pre-assigned response button. Button assignment remained the same for the rest of the experiment within one participant; it was counterbalanced across participants. Responses were scored as "corresponding" (participant responded 'one sound' for a base tone or 'two sounds' for a manipulated tone) or "non-corresponding" (the converse cases: participant responded 'two sounds' for a base tone or 'one sound' for a manipulated tone). The training blocks were repeated when the percentage of "corresponding" responses was below 65%. None of the subjects needed more than two training sessions.

From the remaining 12 stimulus blocks, 11 blocks (blocks 6-10 and 12-17) matched the stimuli and experimental conditions of Experiment 1, except that the onset-to-onset interval was increased to 1400 ms, and only 140 tones (instead of 200) of both the base and the manipulated tone versions were delivered in each of the 11 conditions. Participants were instructed to indicate whether they perceived one or two sound objects, but mark their answer only once the fixation cross changed to "X" on the screen, which occurred at a random time between 550 and 750 ms after the tone onset. Stimulus blocks commenced with 10 base version sounds, which were not included in either the behavioral or the electrophysiological data analysis.

Between the main stimulus blocks 10 and 12, participants received another control stimulus block (11), the data of which are not reported here. In this stimulus block, no sounds were presented. Participants were instructed to press either one of the response keys when the fixation cross changed to "X". The temporal schedule of delivering the cross-changes was the same as in the other control block (3).

The total net time of the experiment was 104 minutes. Short and long breaks were inserted as in Experiment 1. The session lasted for ca. 4 hours (including instructions, electrode mounting and removal).

Electrophysiological recording and data analysis

Parameters for the EEG recording were identical to Experiment 1, except that signals were sampled at 2000 Hz, and resampled offline to 250 Hz for data analysis.

For each tone, an epoch of 650 ms duration including a 100 ms pre-stimulus baseline was extracted from the continuous EEG record. Epochs with an amplitude change exceeding 100 μV at any electrode were rejected from further analysis, which led to retaining 84.6% of the epochs on average.

For evaluating the perceptual judgments, the percentage of correspondence between the presence or absence of a manipulation and the listeners' judgments (two vs. one sound) was calculated separately for each condition for the base and manipulated versions of the tones. The effects of stimulus condition on the perceptual judgments were assessed by a repeated-measures ANOVA with the factors *Type of tone* (2 levels: base version vs. manipulated version of tones) × *Condition* (11 levels).

Difference waveforms were calculated as described in Experiment 1. For N1 and ORN, the measurements are identical as Experiment 1 and P400 amplitudes were measured in 100-ms wide intervals centered on the average peak latency per condition.

ORN amplitudes underwent the same statistical analyses as employed in Experiment 1. For P400 amplitudes, a one-way repeated-measures ANOVA with the factor *Manipulation* (3 levels: 2^{nd} partial mistuned vs. 2^{nd} partial delayed vs. 2^{nd} partial with location difference) as well as two-tailed paired-sample *t* tests investigating additivity effects were administered.

ORN amplitudes and scalp topographies were also compared between the two experiments by two mixed-model ANOVAs. The ANOVA comparing ORN amplitudes was based on amplitude measures from Cz and had the factors *Listening condition* (2 levels, across groups: active listening vs. passive listening) × *Number of mistuned partials* (2 levels: 2^{nd} partial vs. 2^{nd} and 4^{th} partials) × *Delay* (2 levels: delay present vs. absent) × *Location difference* (2 levels: location difference present vs. absent). The ANOVA comparing the scalp distributions of the single-cue based ORN components had the factors *Listening condition* (2 levels, across groups: active listening vs. passive listening) × *Manipulation* (3 levels: 2^{nd} partial mistuned vs. 2^{nd} partial delayed vs. 2^{nd} partial with location difference) × *Frontality* (3 levels: frontal

vs. central vs. parietal) × *Laterality* (2 levels: left, right), where each electrode cluster (as defined for Experiment 1) was represented by the mean amplitude measured from the electrodes in the cluster.

For the N1 control blocks, difference waveforms were calculated between the responses elicited by the two types of tones. N1 amplitudes were measured in 40-ms wide intervals centered on the average peak latency per condition. N1 difference amplitudes were tested against zero using one-sample, two-tailed *t* tests.

In all other respects, methods were the same as in Experiment 1.

Results

Behavioral data

The percentage of the corresponding one-sound answers to the base versions was 95.43%, while the percentage of corresponding two-sound answers to the manipulated versions was 85.69% (averaged across the 11 conditions; see Table 4).

The repeated-measures ANOVA showed a significant main effect of *Type of tone* [$F(1,22) = 28.699$, $p < 0.001$, $\eta^2 = 0.566$], where the number of corresponding answers to base versions was significantly higher than the corresponding answers to the manipulated versions of tones. There was also a significant main effect of *Condition* [$F(10,220) = 74.298$, $p < 0.001$, $\eta^2 = 0.772$, $\varepsilon = 0.413$], which was due to a significantly smaller number of corresponding answers in the condition with the 2nd partial with location difference than in all the other conditions (all p values < 0.001). A significant interaction of *Type of tone* and *Condition* was also found [$F(10,220) = 63.953$, $p < 0.001$, $\eta^2 = 0.744$, $\varepsilon = 0.235$]. This was due to a significant main effect of *Condition* for corresponding answers to the manipulated versions [$F(10,220) = 95.622$, $p < 0.001$, $\eta^2 = 0.813$, $\varepsilon = 0.286$]. On the other hand, no significant main effect of *Condition* for corresponding answers to the base versions was found [$F(10,220) = 1.578$, $p = 0.2$, $\eta^2 = 0.067$, $\varepsilon = 0.317$].

----------Insert Table 4 about here ----------

Electrophysiological data

ORN

Figure 3 shows the ERP responses elicited by the base and the manipulated tones as well as the corresponding difference waveforms for all experimental conditions at Cz. In the conditions with delay, the delayed partials commenced 100 ms later, causing the resulting ORN to be delayed. Significant ORN responses were elicited in all conditions (see Table 5, middle).

----------Insert Figure 3 about here----------

The ANOVA comparing ORN amplitudes and scalp topographies (Figure 2, bottom row) across the three single-cue conditions showed a significant main effect of *Manipulation* [$F(2,44) = 3.840$, $p = 0.029$, $\eta^2 = 0.149$], but the post hoc test only showed a tendency towards significance for the 2nd partial delayed having larger amplitudes than the 2nd partial with location difference ($p = 0.081$). There was also a significant main effect of *Frontality* [$F(2,44) = 19.787$, $p < 0.001$, $\eta^2 = 0.474$, $\varepsilon = 0.635$], which resulted from significantly larger amplitudes in the frontal ($p = 0.005$) and central ($p < 0.001$) than the parietal clusters. No significant interaction of *Manipulation* and either one of the topography factors (*Frontality* or *Laterality*) was observed, suggesting that ORN topography did not differ between the 3 types of manipulation.

----------Insert Table 5 about here ----------

In the ANOVA assessing the effects of providing multiple congruent cues, no significant effects or interactions were observed for any of the experimental manipulations (*Number of mistuned partials, Delay* or *Location difference*), all p values > 0.12. This is consistent with the pattern of results found in Experiment 1.

In the cue additivity tests, we found that multiple congruent cues always elicited numerically smaller ORN amplitudes than the sum of the ORN amplitudes elicited by the contributing cues, although not all of these differences were significant (see the corresponding amplitude values and statistical test results in Table 6).

----------Insert Table 6 about here ----------

P400

P400 difference amplitudes were not significant in the following four stimulus conditions: 2nd and 4th partials mistuned, 2nd partial with location difference, 2nd and 4th partials mistuned with location difference and 2nd and 4th partials mistuned, delayed and with location difference (see Table 5, bottom for all results). In the ANOVA comparing the three single-cue conditions, we found a significant main effect of *Manipulation* [$F(2,44)$ = 6.145, p = 0.004, η^2 = 0.218,] due to significantly larger amplitudes in the 2nd partial delayed condition than in the 2nd partial with location difference condition (p = 0.015). P400 amplitude in the 2nd partial mistuned condition did not significantly differ from either of the other conditions (both p values > 0.069). The scalp topographies for the P400 components in the three single-cue conditions are shown in Figure 4.

----------Insert Figure 4 about here----------

The cue additivity tests for the P400 showed less homogeneous results than those for the ORN. About half of the comparisons numerically pointed towards sub-, the other half towards super-additivity, while only one comparison in either direction was significant (subadditivity for the combination of 2nd partial delayed plus location difference with 2nd partial mistuned; superadditivity for the combination of 2nd partial delayed with 2nd partial with location difference). No other significant results were obtained (see Table 7 for all results).

----------Insert Table 7 about here----------

Comparing ORN between the active and the passive listening conditions

In the mixed-model ANOVA comparing the ORN amplitudes between Experiments 1 and 2, only *Listening condition* (i.e., the difference between the two experiments) had a significant effect [$F(1,19)$ = 6.9535, p = 0.016, η^2 = 0.268]. ORN amplitudes were larger in the active than in the passive listening situation. No other main effects were observed, and no interactions between *Listening condition* and any of the other factors (*Number of mistuned partials*, *Delay*, and *Location difference*; all p values > 0.078).

When comparing the scalp topographies of the ORN components between the passive and active listening conditions, no main effect of or interaction involving the *Listening condition* was found. The significant main effects of *Manipulation* [$F(2,38)$ = 4.349, p = 0.019, η^2 = 0.186] and of *Frontality* [$F(2,38)$ = 35.206, p < 0.001, η^2 = 0.649, ε = 0.635] replicated the similar results obtained in the analyses that were conducted separately for the two experiments, with the post hoc tests also showing the same origin for these effects.

N1

No significant N1 differences were found between the two different tones in either of the N1 control conditions. In the condition controlling for the effects of delay, the mean amplitude difference at Cz was -0.4809 μV ($t(22)$ = -1.2619, p = 0.22), and in the condition controlling for the location difference the mean amplitude difference was 0.2798 μV ($t(22)$ = 0.9284, p = 0.36). The N1 effects are shown in Figure 5.

In contrast, manipulated tones elicited significantly larger N1 components than the base versions of the tones in most of the 11 stimulus conditions (see Table 5, top for all results).

----------Insert Figure 5 about here----------

Discussion

In Experiment 2, all investigated cues and cue combinations elicited significant ORN responses. Even the condition with location difference alone showed a significant ORN response, although participants' perceptual judgments showed that few of the manipulated tones evoked the perception of two concurrent sounds in this stimulus condition. This suggests that the cues of concurrent sound segregation employed in the current experiment were picked up by the processes underlying ORN generation. On the other hand, ORN cannot fully govern perceptual judgments. Many psychophysical studies have shown that perceptual judgments are codetermined by the criteria used by the observer in deciding between the behavioral alternatives (Green and Swets, 1966). The present instructions did not attempt to manipulate the decision level – listeners were free to choose their criteria. Given that in most stimulus blocks the cues (when present) strongly promoted perception of two concurrent sounds, it is plausible to assume that listeners accommodated to the high distinctiveness of these cues and set their criteria to be high. This should have resulted in dismissing most exemplars with the relatively weak location cue while the obligatory evaluation of the auditory system – as reflected by the elicitation of the ORN component – suggested the presence of two concurrent sound sources. The significantly higher percentage of corresponding responses for the base version tones is also compatible with this interpretation.

In full correspondence with the results obtained in Experiment 1, ORN amplitude did not significantly increase with congruently manipulating two partials (as compared to manipulating only one), or with adding delay and/or location difference on top of mistuning. Furthermore, combining several cues tended to elicit ORN components of smaller amplitude than the sum of the contributing ORN components. No superadditive effects were found for any cue combination. Further, no significant differences were found between the ORN effects or scalp topographies between the two experiments. The results of Experiment 2 thus replicated those obtained in Experiment 1 and support the interpretation given for them in the Discussion of Experiment 1.

The P400 results were not as clear-cut as those for the ORN component. In some cases, no significant P400 was observed, even though the ORN was elicited in those conditions as well. This contrasts the results of some previous studies (Alain et al., 2002). Note, however, that even in the conditions with non-significant P400, positive deflections in the P400 latency range were observed in the difference waveforms between the ERPs elicited by the

manipulated and the base tones (cf. Figure 3). It is possible that the lack of significant detection of the P400 component was caused by some of the preceding ORN components not yet having terminated at the onset of the P400.

In terms of cue redundancy, results for the P400 were equivocal: we found significant superadditivity in one case and significant subadditivity in another case. Altogether there was no clear tendency towards either pattern. Again, these results may be partly obscured by the preceding ORN components.

Significant N1 differences were obtained in most conditions. Unlike in the passive listening situation (Experiment 1), in the active listening situation, these N1 differences were not confined to conditions where delay and location difference manipulations were employed, but also extended to conditions with mistuning alone. The two control conditions suggest that these N1 effects were not solely due to the acoustic differences between the base and the manipulated tones, but rather they may reflect some aspect of processing concurrent sounds.

General discussion

In the present study, we systematically combined three cues of concurrent sound segregation (mistuning, onset asynchrony, and location difference) for testing how the ORN event-related potential component (and in Experiment 2, also the P400 component as well as perceptual judgments) reflects the joint evaluation of these cues. We employed two listening conditions in two separate experiments in which participants were instructed to either disregard the tones (passive listening) or to focus their attention on the tones and judge whether they heard one or two sounds (active listening). The pattern of ORN elicitation in response to the different cues and their combination was highly similar under the two listening conditions. This pattern is consistent with the notion that the ORN response reflects the auditory system's overall assessment of the likelihood that the sound input carries contributions from two sound sources, rather than summing together the outputs of independent detectors of concurrent sound segregation.

We found that ORN was elicited by all of the tested combinations of cues and also by each of the cues individually, with the exception of location difference alone during passive listening. Location difference was also the weakest cue during active listening; although it elicited a

small-amplitude ORN, it seldom (< 20%) led to the perception of two separate sounds. Our location-cue results are fully consistent with those of McDonald and Alain (2005) who also showed a small ORN for location difference alone during active listening, no significant ORN during passive listening, and little effect of location difference on perceptual segregation. Notably, these authors used a similar amount of location difference (90°) but with free field presentation, which suggests that the lack of an effect in the present study should not be attributed to the artificial manipulation of location via headphones. Instead, the present and previous results (McDonald and Alain, 2005) suggest that location difference alone is not a strong cue of concurrent sound segregation, or at least its effects are easily counteracted by other cues pointing towards integration (i.e., harmonicity and common onset). Alternatively, it is possible that the location cue was not sufficiently salient, although the locations used for the different harmonics could be clearly distinguished as determined by informal perceptual reports. The saliency of the location cue may have been reduced by the fact that for 25% of the tones, all harmonics were delivered at the same location where the location-manipulated harmonics of half of the manipulated tones appeared (see Methods). That is, unlike the mistuned and delayed harmonics, the harmonics separated in location from the other harmonics of the same tone appeared with equal probability as part of tones in which all harmonics were delivered at the same (perceived) location. Bendixen and colleagues (2010) have found an effect of the probability of mistuning on the ORN amplitude. The lack of significant ORN elicitation by the current location-separation cue may indicate a similar contextual effect on ORN. Finally, the lack of consistent perceptual judgments for the location-cue manipulated tones demonstrated that although the information provided by the processes underlying ORN may reflect the full assessment of the auditory system regarding the presence of two concurrently active sound sources, perceptual judgments are codetermined by other effects (cf. the Discussion of the results of Experiment 2).

Mistuning one of the partials of the complex sound or delaying its onset elicited a clear ORN component and led to a robust two-object percept, as was shown in previous studies (e.g. Alain et al., 2001, 2002, 2003; Lipp et al., 2010; Weise et al., 2012). Importantly, ORN amplitude remained unchanged when manipulating not only one (the 2nd) but two (the 2nd and 4th) partials in a congruent manner. We had hypothesized that involving two partials would increase the saliency of the manipulation and thereby boost effects on ORN and perception. Such result patterns have been previously reported for increasing the amount of mistuning (e.g. Alain et al., 2001) or increasing the strength of a dichotic pitch manipulation (Clapp et

al., 2007). The present results suggest that manipulating more than one partial of a complex sound does not lead to a similar increase in strength or saliency. Alternatively, it is possible that the 4th partial is not sufficiently influential in assessing the source of complex sounds (cf. Alain et al., 2001, who showed that mistuning the 4th partial alone causes weaker effects than mistuning the 2nd partial alone; therefore, the 2nd partial may have dominated the present results).

Besides manipulating more than one partial, we pursued a second approach for increasing the strength of the sensory evidence in favor of concurrent sound segregation. This approach was based on employing multiple cues in parallel (i.e., onset asynchrony and/or location difference in addition to mistuning). Previous studies suggested that ORN increases with such cue combinations (Hautus et al., 2009; McDonald and Alain, 2005, Weise et al., 2012), and that this increase may follow a fully additive pattern (Du et al., 2011). In contrast, here we mostly found subadditivity for the amplitude of ORN elicited by multiple cues. That is, the combinations of cues of concurrent segregation elicited lower-amplitude ORN components compared with the sum of the ORN amplitudes elicited by each contributing cue separately. In most cases, the increase of the ORN amplitude from single to multiple cues was non-significant, as was shown by the ANOVAs assessing the effects of multiple congruent cues.

One reason for these weak, subadditive effects of the cue combinations may be that the employed cues, at least as far mistuning and onset asynchrony are concerned, were clearly supra-threshold: Each of them alone was sufficient to elicit as much as 90% correspondence between the presence of the cue and the perceptual judgment. Hence the cues can be regarded as fully redundant with respect to each other. Similarly, adding the (weak) location cue on top of a strong cue of concurrent segregation probably also provided only redundant information (cf. McDonald and Alain, 2005, for high amounts of mistuning). Therefore, one may not be surprised that ORN amplitude does not increase further by adding more cues. Note, however, that this explanation implies that the ORN reflects the outcome of a process that combines the sensory evidence to provide an overall assessment of the likelihood that one or two sound sources were present in the environment. If, on the contrary, ORN were to reflect the strength of the sensory evidence that drives the decision between one and two sound sources, then each additional cue should increase ORN amplitude in an additive manner, even if it is redundant with the other cues in terms of the eventual perceptual decision. Our results are not consistent with this latter view; they support the interpretation that ORN underlies the actual

perceptual decision. Subadditivity between the ORN components elicited by the contributing cues suggests that the likely method of combining cues of concurrent stream segregation is the selection of the individually most salient cue. Several authors have already argued that the ORN probably reflects a perceptual grouping mechanism rather than a cue-related response; their arguments were based on the highly different nature of the stimuli eliciting ORN (e.g., Hautus and Johnson 2005; Hautus et al., 2009; Johnson et al. 2007; Lipp et al., 2010; McDonald & Alain, 2005). We add here the argument of subadditivity for multiple congruent cues. Another supporting piece of evidence in the present data is given by the highly similar ORN topographies across different types of cues and listening conditions (cf. Figure 2).

The P400 component was less tightly related to the perceptual reports. Unlike in previous studies (Alain et al., 2001, 2002; Hautus & Johnson, 2005), P400 failed to reach significance in some conditions despite clear perceptual distinction between one- and two-sound objects. One might speculate that a procedural difference between our and previous studies may account for this (namely, subjects were to withhold their response for several hundred milliseconds). However, a similar dissociation between P400 and perceptual decisions was observed by Hautus and colleagues (2009) with the instruction to respond as quickly as possible. Hence the results are more in line with the view that ORN and P400, as well as P400 and behavior, are not as tightly connected as previously assumed (Johnson et al, 2007). In terms of cue redundancy, there was no clear sub-, super- or fully additive pattern for the P400; the results are thus not informative regarding this question.

Unexpectedly, we observed effects related to the cues of concurrent segregation also in the N1 latency range. Some previous studies reported mistuning-related effects preceding the ORN latency range in MEG (Alain and McDonald, 2007; Lipp et al., 2010). These effects were, however, even earlier than in the N1 range; it remains unclear whether they relate to the present effects. Because our control conditions suggest that the N1 effects were not caused by the acoustic differences between the base and the manipulated stimuli, we tentatively suggest that the N1 differences are related to the automatic processing of the cues of concurrent sound segregation, but not necessarily to the resulting percept (following the interpretations of Alain and McDonald, 2007, as well as Lipp et al., 2010).

Finally, a significant effect of listening condition was found for the ORN amplitude, with larger amplitudes during active than passive listening. This is in accordance with some (e.g. Alain et al., 2001) but not all (e.g. Alain & Izenberg, 2003; Lipp et al., 2010) previous studies.

Alain and colleagues (2001) found that attention only affected the ORN amplitude when the same fundamental frequency and number of manipulated partial have been used throughout a stimulus block, but not when the fundamental frequency and/or the manipulated partial were randomly varied. The authors suggested that under constant stimulus conditions, participants start to actively search for the mistuned partial in order to perform the task, and that this search process caused the attention effect. In the present study, although a random variation in perceived location was present, neither the fundamental frequency nor the number of the manipulated partial varied within the stimulus blocks. It is thus possible that the modulation of ORN amplitude by listening condition reflects task-specific preparation during active listening rather than a genuine attention effect on ORN. If this was the case, then the search process assumed by Alain and colleagues (2001) appears to be insensitive to location variation – i.e., location information may not be part of the search template.

In conclusion, we provide evidence for the ORN component reflecting the auditory system's combined assessment as to whether one or more sources contributed to the incoming sound. Our results are not consistent with the view that ORN would directly reflect the processing of the sensory cues that underlie this perceptual decision. This further qualifies the ORN component as an indicator of concurrent sound segregation, and shows that the brain accomplishes this complex operation in a short time (<200 ms).

References

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

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., Izenberg, A. (2003). Effects of attentional load on auditory scene analysis. Journal of Cognitive Neuroscience, 15, 1063-1073.

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

Alain, C., Reinke, K. S., He, Y., Wang, C., Lobaugh, N. (2005). Hearing two things at once: Neurophysiological indices of speech segregation and identification. Journal of Cognitive Neuroscience, 17, 811-818.

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

Alain, C., Theunissen, E.L., Chevalier, H., Batty, M., Taylor, M.J. (2003). Developmental changes in distinguishing concurrent auditory objects. Cognitive Brain Research, 16, 210–218.

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, Massachusetts: The MIT Press.

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

Chatrian, G.E., Lettich, E., Nelson, P.L. (1985). Ten percent electrode system for topographic studies of spontaneous and evoked EEG activity. American Journal of EEG Technology, 25, 83-92.

Clapp, W.C., Johnson, B.W., Hautus, M.J. (2007). Graded cue information in dichotic pitch: effects on event-related potentials. Neuroreport, 18, 365-368.

Delorme, A., Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. Journal of Neuroscience Methods, 134, 9-21.

Du, Y., He, Y., Ross, B., Bardouille, T., Wu, X., Li, L., Alain, C. (2011). Human auditory cortex activity shows additive effects of spectral and spatial cues during speech segregation. Cerebral Cortex 21, 698-707.

Green D.M, Swets J.A. Signal detection theory and psychophysics. New York: Wiley; 1966.

Hautus, M.J., Johnson, B.W. (2005). Object-related brain potentials associated with the perceptual segregation of a dichotically embedded pitch. Journal of the Acoustical Society of America, 117, 275–280.

Hautus, M.J., Johnson, B.W., Colling, L.J. (2009). Event-related potentials for interaural time differences and spectral cues. Neuroreport, 20, 951-956.

Haykin, S., Chen, Z. (2005). The cocktail party problem. Neural Computation, 17, 1875-1902.

Jasper, H.H. (1958). The ten-twenty electrode system of the International Federation. Electroencephalography and Clinical Neurophysiology, 10, 370-375.

Johnson, B.W., Hautus, M., Clapp, W.C. (2003). Neural activity associated with binaural processes for the perceptual segregation of pitch. Clinical Neurophysiology, 114, 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, 541-551

Lipp, R., Kitterick, P., Summerfield, Q., Bailey, P.J., Paul-Jordanov, I. (2010). Concurrent sound segregation based on inharmonicity and onset asynchrony. Neuropsychologia, 48, 1417-25.

McDonald, K.L., Alain, C. (2005). Contribution of harmonicity and location to auditory object formation in free field: Evidence from event-related brain potentials. Journal of the Acoustical Society of America, 118, 1593-1604.

Sanders, L.D., Joh, A.S., Keen, R.E., Freyman, RL. (2008). One sound or two? Object-related negativity indexes echo perception. Perceptual Psychophysiology, 70, 1558-1570.

Sanders, L.D.. Zobel, B.H., Freyman, R.L., Keen, R. (2008). Manipulations of listeners' echo perception are reflected in event-related potentials. Journal of the Acoustical Society of America, 129, 301-309.

Snyder, J.S., Alain, C. (2005). Age-related changes in neural activity associated with concurrent vowel segregation. Cognitive Brain Research, 24, 492-499.

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

Weise, A., Schröger, E., Bendixen, A. (2012). The processing of concurrent sounds based on inharmonicity and asynchronous onsets: An object-related negativity (ORN) study. Brain Research, 1439, 73-81.

Table(s)

Table 1
Summary of experimental manipulations

Condition	Experimental manipulation(s)	1 or 2 partials	Mistuning	Delay	Location
1	2nd partial mistuned	1	+	-	-
2	2nd and 4th partials mistuned	2	+	-	-
3	2nd partial delayed	1	-	+	-
4	2nd partial delayed and mistuned	1	+	+	-
5	2nd and 4th delayed and mistuned	2	+	+	-
6	2nd partial with location difference	1	-	-	+
7	2nd partial mistuned with location difference	1	+	-	+
8	2nd and 4th partials mistuned with location difference	2	+	-	+
9	2nd partial delayed with location difference	1	-	+	+
10	2nd partial mistuned, delayed and with location difference	1	+	+	+
11	2nd and 4th partials mistuned, delayed and with location difference	2	+	+	+

Table(s)

Table 2

Grand-average (N=20) ERP amplitudes at Cz measured in the N1 (top) and the ORN (bottom) latency range of the manipulated-minus-base difference waveform for the 11 stimulus conditions of Experiment 1 (passive listening)

N1	2nd partial mistuned	2nd and 4th partials mistuned	2nd partial delayed	2nd partial mistuned and delayed	2nd and 4th mistuned and delayed	2nd partial with location difference	2nd partial mistuned with location difference	2nd and 4th partials mistuned with location difference	2nd partial delayed with location difference	2nd partial mistuned, delayed and with location difference	2nd and 4th partials mistuned, delayed and with location difference
Mean amplitude at Cz (μV)	-0.1504	-0.5522	-0.6983	-0.8475	-0.8587	-0.6698	-0.2249	-0.5698	-0.8764	-0.3931	-0.6877
t(19)	-0.53	-1.4025	-2.2885	-3.4962	-2.8153	-2.3047	-0.9353	-1.8026	-3.184	-1.0468	-2.3775
p	0.6023	0.1769	0.0337*	0.0024**	0.0110*	0.0326*	0.3614	0.0873	0.0049**	0.3084	0.0281*
Time window for measurement (ms)	80-120	84-125	80-120	80-120	80-120	60-100	80-120	80-120	80-120	80-120	80-120
ORN											
Mean amplitude at Cz (μV)	-1.0827	-1.5911	-2.1033	-1.7498	-1.9277	-0.5292	-1.0487	-1.2053	-1.6150	-1.2779	-1.7862
t(19)	-3.917	-4.6625	-5.0669	-5.0297	-5.2885	-1.6832	-4.4899	-3.3015	-4.5769	-3.3739	-6.3604
p	<0.001***	<0.001***	<0.001***	<0.001***	<0.001***	0.1087	<0.001***	0.004**	<0.001***	0.003**	<0.001***
Time window for measurement (ms)	124-196	124-196	180-252	180-252	180-252	152-224	160-232	116-188	180-252	192-264	188-260

Notes: Significant differences from zero are marked with asterisks (p<.05, **p<.01, ***p<.001)*

Table(s)

Table 3

Additivity tests for combining cues of concurrent sound segregation during passive listening (Experiment 1). The sum of the amplitudes of the ORN components elicited by the contributing cues and the corresponding multiple-cue ORN amplitudes (at Cz) are given together with the t and p values for the paired two-tailed t tests between them.

Conditions	Mean amplitude at Cz (μV)	t(19)	p
2nd partial mistuned + 2nd partial delayed (condition 1+3)	-3.186		
2nd partial mistuned and delayed (condition 4)	1.74979	-1.991	0.061
2nd partial mistuned+ 2nd partial with location difference (condition 1+6)	-1.61194		
2nd partial mistuned and with location difference (condition 7)	-1.04872	-1.306	0.207
2nd partial delayed + 2nd partial with location difference (condition 3+6)	-2.63249		
2nd partial delayed and with location difference (condition 9)	-1.61503	-1.59	0.128
2nd partial mistuned + 2nd partial delayed + 2nd partial with location difference (condition 1+3+6)	-3.71521		
2nd partial mistuned, delayed with location difference (condition 10)	-1.27789	-3.143	0.005**
2nd partial mistuned and delayed + 2nd partial with location difference (condition 4+6)	-2.27901		
2nd partial mistuned, delayed and with location difference (condition 10)	-1.27789	-1.871	0.077
2nd partial mistuned and with location difference + 2nd partial delayed (condition 7+3)	-3.152		
2nd partial mistuned, delayed and with location difference (condition 10)	-1.27789	-2.431	0.025*
2nd partial delayed and with location difference + 2nd partial mistuned (condition 9+1)	-2.69775		
2nd partial mistuned, delayed and with location difference (condition 10)	-1.27789	-2.225	0.038*

Notes: Significant differences are marked with asterisks (p<.05, **p<.01, ***p<.001)*

Table(s)

Table 4

Group mean (N=23) percentages of correspondence between the presence vs. absence of cue manipulation and the listener's judgment, as well as percentage of corresponding responses separately for the base and manipulated versions in the 11 experimental conditions

Conditions	Percentage of responses corresponding to the base version	Percentage of responses corresponding to the manipulated version
2nd partial mistuned	94.78%	90.78%
2nd and 4th partials mistuned	96.34%	90.06%
2nd partial delayed	94.88%	87.48%
2nd partial mistuned and delayed	96.21%	93.70%
2nd and 4th partials mistuned and delayed	97.30%	94.13%
2nd partial with location difference	90.75%	18.32%
2nd partial mistuned and with location difference	94.47%	93.91%
2nd and 4th partials mistuned and with location difference	97.05%	93.79%
2nd partial delayed and with location difference	95.96%	91.86%
2nd partial mistuned, delayed and with location difference	95.90%	93.79%
2nd and 4th partials mistuned, delayed and with location difference	96.09%	94.75%

Table(s)

Table 5

Grand-average (N=23) ERP amplitudes at Cz measured in the N1(top), ORN (middle), and P400 (bottom) latency range of the manipulated-minus-base difference waveforms for the 11 stimulus conditions of Experiment 2 (active listening)

N1	2nd partial mistuned	2nd and 4th partials mistuned	2nd partial delayed	2nd partial mistuned and delayed	2nd and 4th mistuned and delayed	2nd partial with location difference	2nd partial mistuned with location difference	2nd and 4th partials mistuned with location difference	2nd partial delayed with location difference	2nd partial mistuned, delayed and with location difference	2nd and 4th partials mistuned, delayed and with location difference
Mean amplitude at Cz (μV)	-0.43356	-0.84169	-1.1528	-0.84069	-1.0484	-0.70872	-0.65448	-1.0609	-0.64766	-0.63116	-1.3187
t(22)	-0.96744	-2.3124	-3.5627	-2.5776	-2.8308	-2.4446	-2.1502	-2.8216	-1.603	-1.9076	-5.3944
p	0.344	0.03*	0.002**	0.017*	0.01*	0.023*	0.043*	0.01*	0.123	0.07	<0.001***
Time window for measurement (ms)	80-120	80-120	80-120	80-120	80-120	72-112	80-120	80-120	80-120	80-120	80-120
ORN											
Mean amplitude at Cz (μV)	-1.1785	-1.9523	-2.5055	-2.2847	-1.9592	-0.9867	-1.8816	-2.4056	-2.0920	-2.0828	-2.9394
t(22)	-3.3805	-5.2447	-6.0678	-5.061	-3.8983	-2.6315	-4.9561	-5.4187	-5.8903	-3.5371	-11.027
p	0.003**	<0.001***	<0.001***	<0.001***	<0.001***	0.015*	<0.001***	<0.001***	<0.001***	0.002**	<0.001***
Time window for measurement (ms)	112-184	108-180	192-264	188-260	188-260	176-248	152-224	172-244	192-264	200-272	208-280
P400											
Mean amplitude at Cz (μV)	1.4768	0.8754	2.197	3.8848	2.832	-0.2733	1.9441	0.9587	3.429	3.1226	1.5044
t(22)	2.5041	1.479	3.3252	4.862	4.032	-0.7086	3.1062	1.7178	5.5549	4.4944	1.4776
p	0.02*	0.153	0.003**	<0.001***	<0.001***	0.486	0.005**	0.1	<0.001***	<0.001***	0.162
Time window for measurement (ms)	352-452	300-400	400-500	420-520	416-516	384-484	316-416	292-392	400-500	400-500	392-492

Notes: Significant differences from zero are marked with asterisks (p<.05, **p<.01, ***p<.001)*

Table(s)

Table 6

Additivity tests for combining cues of concurrent sound segregation during active listening (Experiment 2). The sum of the amplitudes of the ORN components elicited by the contributing cues and the corresponding multiple-cue ORN amplitudes (at Cz) are given together with the t and p values for the paired two-tailed t tests between them.

Conditions	Mean amplitude at Cz (μV)	t(19)	p
2^{nd} partial mistuned + 2^{nd} partial delayed (condition 1+3)	-3.683962		
2^{nd} partial mistuned and delayed (condition 4)	-2.284653	-2.474	0.022*
2^{nd} partial mistuned+ 2^{nd} partial with location difference (condition 1+6)	-2.165166		
2^{nd} partial mistuned and with location difference (condition 7)	-1.881584	-0.445	0.661
2^{nd} partial delayed + 2^{nd} partial with location difference (condition 3+6)	-3.49219		
2^{nd} partial delayed and with location difference (condition 9)	-2.092038	-2.359	0.028*
2^{nd} partial mistuned + 2^{nd} partial delayed + 2^{nd} partial with location difference (condition 1+3+6)	-4.670659		
2^{nd} partial mistuned, delayed with location difference (condition 10)	-2.082845	-3.073	0.006**
2^{nd} partial mistuned and delayed + 2^{nd} partial with location difference (condition 4+6)	-3.271350		
2^{nd} partial mistuned, delayed and with location difference (condition 10)	-2.082845	-1.856	0.077
2^{nd} partial mistuned and with location difference + 2^{nd} partial delayed (condition 7+3)	-3.270507		
2^{nd} partial mistuned, delayed and with location difference (condition 10)	-2.082845	-1.408	0.173
2^{nd} partial delayed and with location difference + 2^{nd} partial mistuned (condition 9+1)	-4.387077		
2^{nd} partial mistuned, delayed and with location difference (condition 10)	-2.082845	-3.414	0.002**

Notes: Significant differences are marked with asterisks (p<.05, **p<.01, ***p<.001)*

Table(s)

Table 7

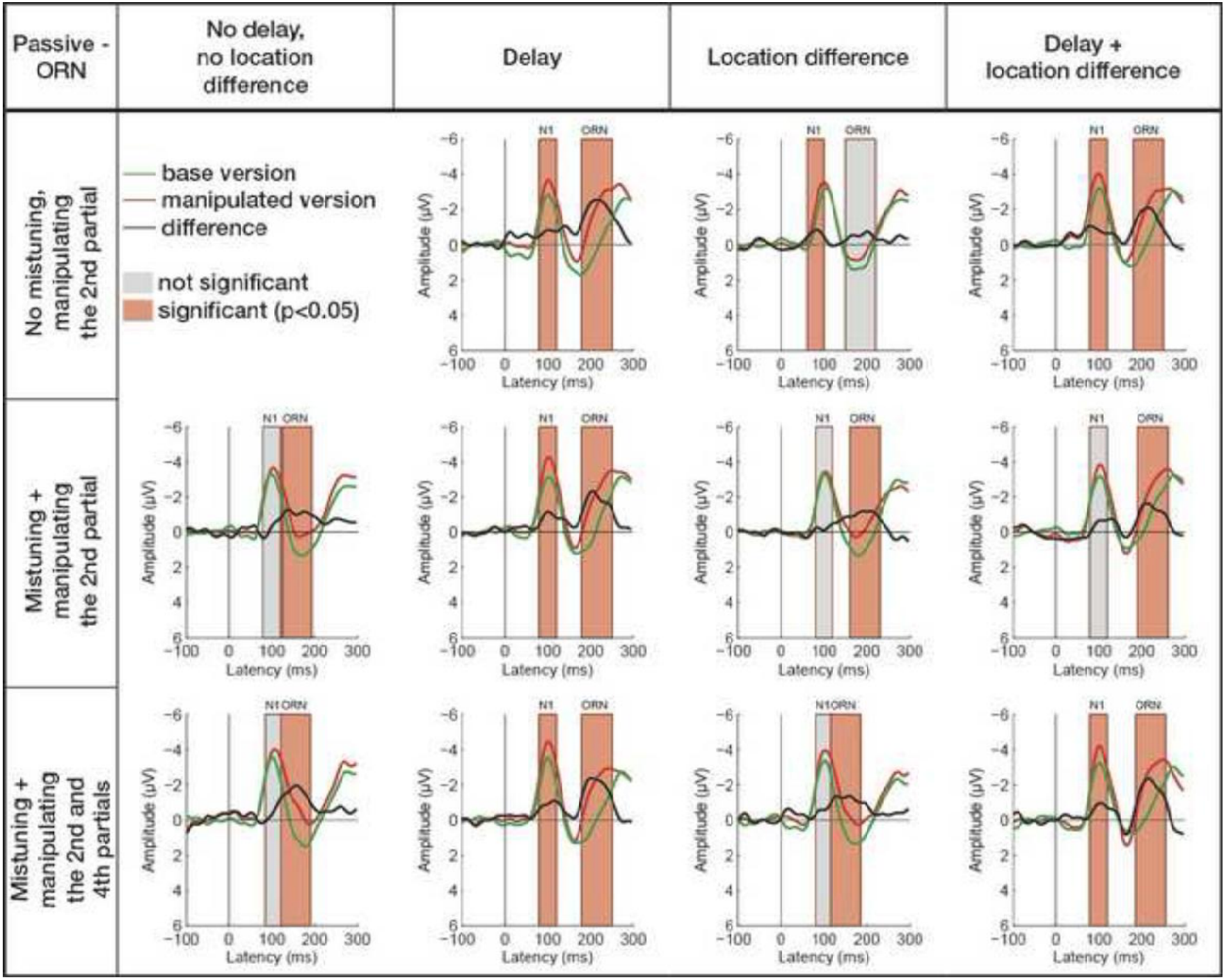
Additivity tests for combining cues of concurrent sound segregation during active listening (Experiment 2). The sum of the amplitudes of the P400 components elicited by the contributing cues and the corresponding multiple-cue P400 amplitudes (at Cz) are given together with the t and p values for the paired two-tailed t tests between them.

Conditions	Mean amplitude at Cz (μV)	t(22)	p
2nd partial mistuned + 2nd partial delayed (condition 1+3)	3.673826		
2nd partial mistuned and delayed (condition 4)	3.884773	-0.223	0.825
2nd partial mistuned+ 2nd partial with location difference (condition 1+6)	1.203539		
2nd partial mistuned and with location difference (condition 7)	1.944099	-1.147	0.264
2nd partial delayed + 2nd partial with location difference (condition 3+6)	1.923776		
2nd partial delayed and with location difference (condition 9)	3.428958	-2.398	0.025*
2nd partial mistuned + 2nd partial delayed + 2nd partial with location difference (condition 1+3+6)	3.40057		
2nd partial mistuned, delayed with location difference (condition 10)	3.122613	0.338	0.739
2nd partial mistuned and delayed + 2nd partial with location difference (condition 4+6)	3.611517		
2nd partial mistuned, delayed and with location difference (condition 10)	3.122613	0.781	0.443
2nd partial mistuned and with location difference + 2nd partial delayed (condition 7+3)	4.14113		
2nd partial mistuned, delayed and with location difference (condition 10)	3.122613	1.39	0.178
2nd partial delayed and with location difference + 2nd partial mistuned (condition 9+1)	4.905753		
2nd partial mistuned, delayed and with location difference (condition 10)	3.122613	2.147	0.043*

Notes: Significant differences are marked with asterisks (p<.05, **p<.01, ***p<.001)*

Figure(s)

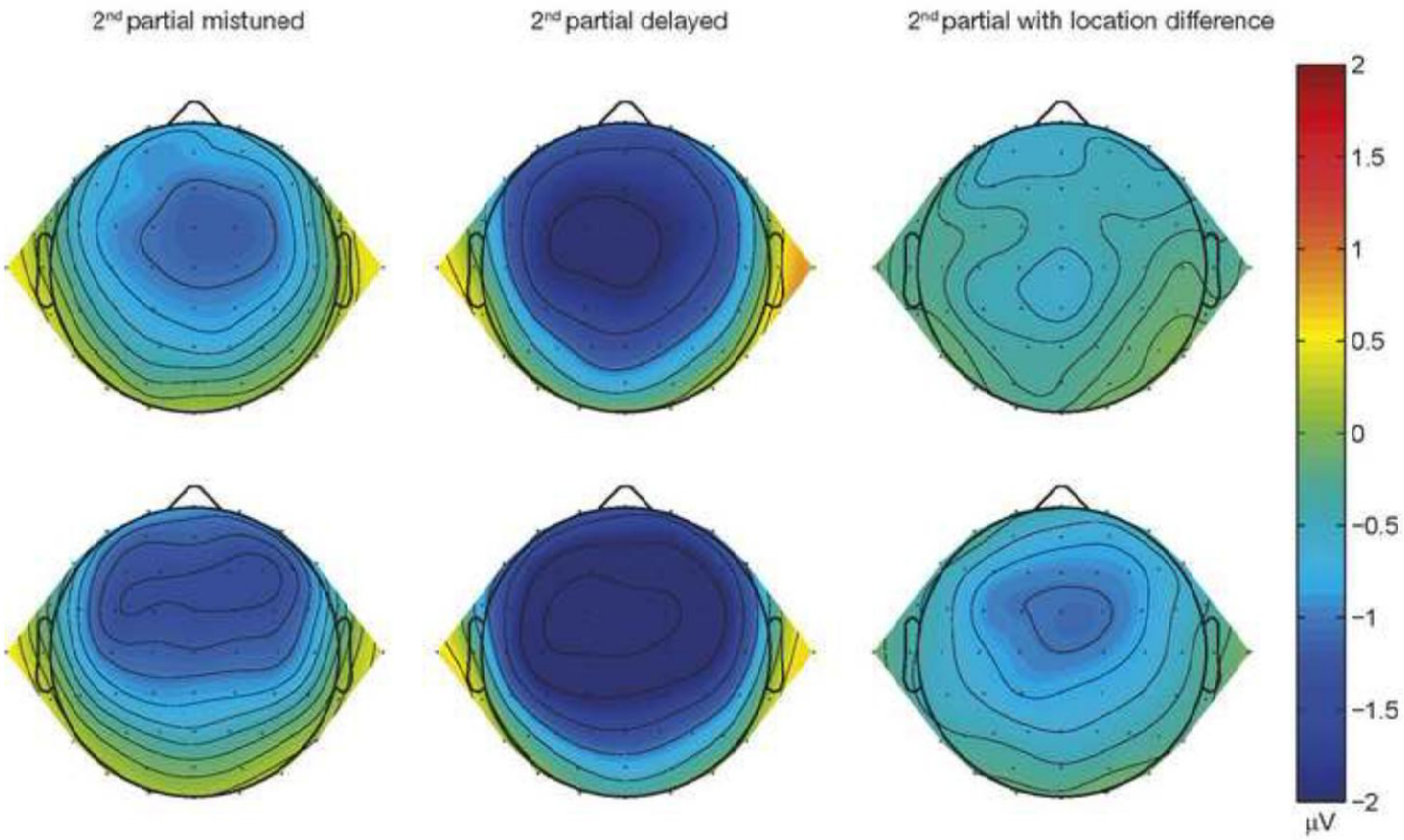
Click here to download high resolution image



Figure(s)

Figure 1. Grand-average (N=20) ERPs elicited at Cz in the 11 conditions of Experiment 1 (passive listening) by the manipulated (red lines) and base-version tones (green), together with their difference waveforms (black). Stimulus onset is at the crossing of the *x* and *y* axes. Note that in the conditions with delay, the delayed partials commenced 100 ms later and the resulting ORN was also delayed.

Figure(s)
Click here to download high resolution image

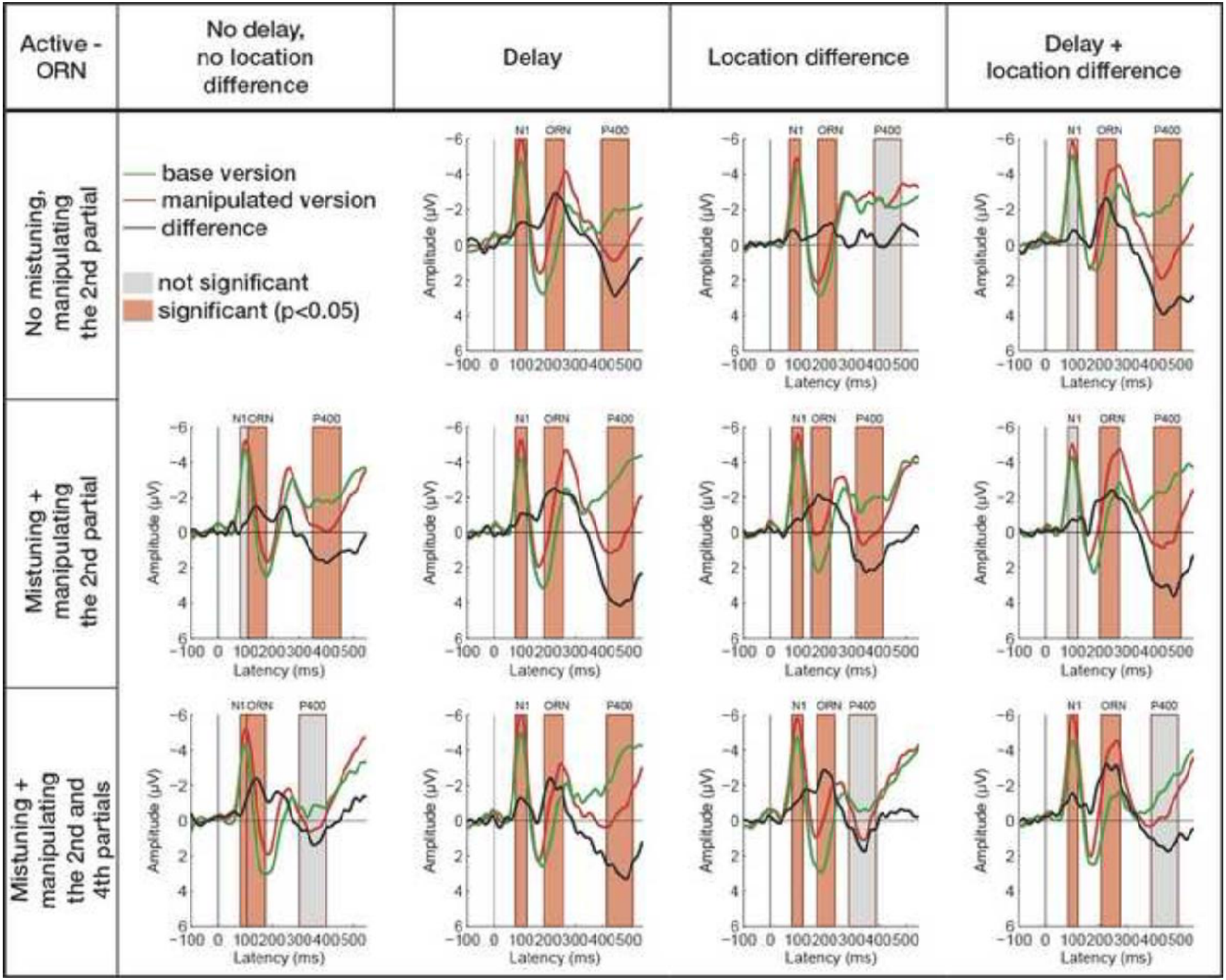


Figure(s)

Figure 2. Grand-average ORN scalp topographies under passive (Experiment 1: N=20; top row) and active (Experiment 2: N= 23; bottom row) listening conditions for the three single-cue conditions (mistuning: left column; delay: middle column; location difference: right column). The common voltage scale is placed at the right side of the figure.

Figure(s)

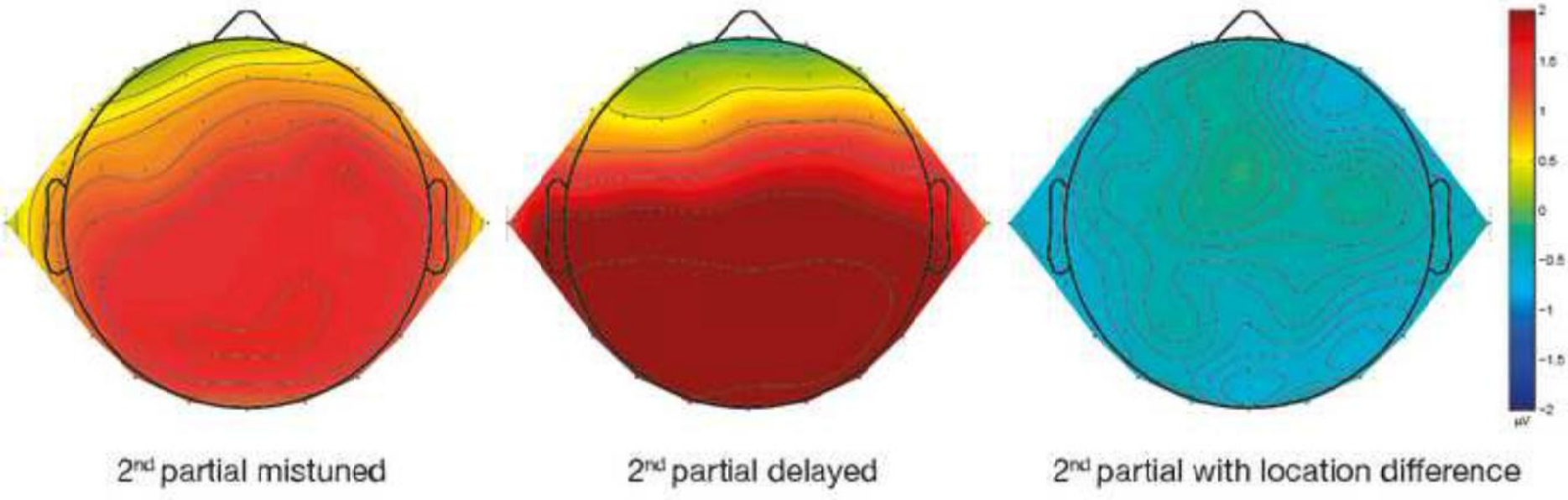
Click here to download high resolution image



Figure(s)

Figure 3. Grand-average (N=23) ERPs elicited at Cz in the 11 conditions of Experiment 2 (active listening) by the manipulated (red lines) and base-version tones (green), together with their difference waveforms (black). Stimulus onset is at the crossing of the *x* and *y* axes. Note that in the conditions with delay, the delayed partials commenced 100 ms later and the resulting ORN was also delayed.

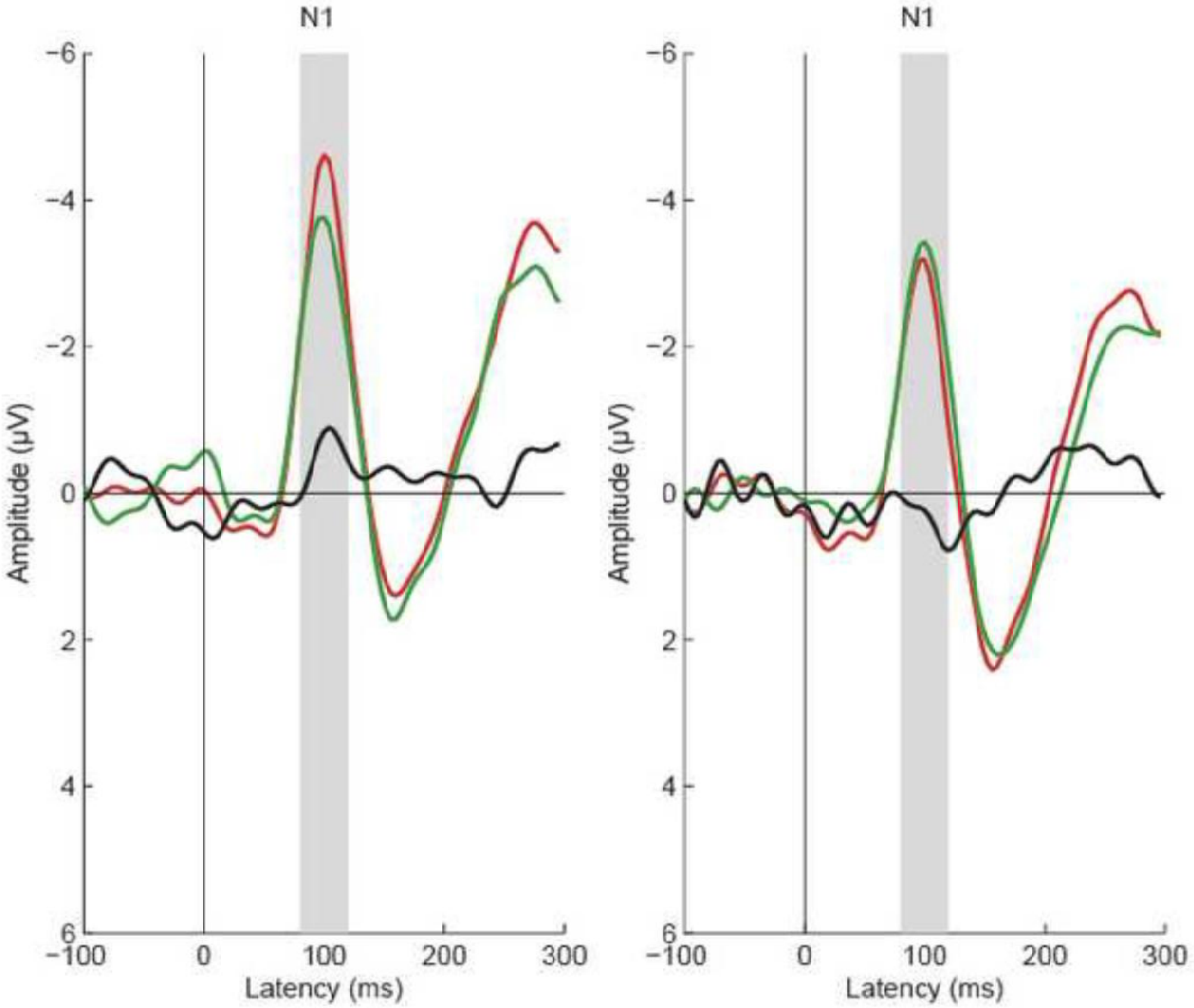
Figure(s)
Click here to download high resolution image



Figure(s)

Figure 4. Grand-average ORN scalp topographies active listening (Experiment 2: N= 23) for the three single-cue conditions (mistuning: left column; delay: middle column; location difference: right column) in the P400 time window. The common voltage scale is placed at the right side of the figure.

Figure(s)
Click here to download high resolution image



Figure(s)

Figure 5. Grand-average (N=23) ERPs elicited at Cz by the control tones together with their difference waveforms (left: testing the effects of partials being delayed; right: testing the effects of partials delivered from different locations).

***Highlights (for review)**

Highlights:

- We assessed the effects of multiple cues on the object-related negativity (ORN).
- Combining cues consistently led to sub-additive effects on the ORN amplitude.
- ORN reflects the integrated assessment of the presence of more than one sound source.
- We used two experimental situations to investigate the effects of attention.