

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

Kocsis, Z.<sup>1,2</sup>, Winkler, I.<sup>1,3</sup>, Szalárdy, O.<sup>1</sup>, & Bendixen, A.<sup>4,5</sup>

<sup>1</sup>Institute of Psychology and Cognitive Neuroscience, Research Centre for Natural Sciences, Hungarian Academy of Sciences, Budapest, Hungary

<sup>2</sup>Budapest University of Technology and Economics, Budapest, Hungary

<sup>3</sup>Institute of Psychology, University of Szeged, Szeged, Hungary

<sup>4</sup>Department of Psychology, Cluster of Excellence “Hearing4all”, European Medical School, Carl von Ossietzky University of Oldenburg, Oldenburg, Germany

<sup>5</sup>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](mailto:kocsis.zsuzsanna@ttk.mta.hu)

## Acknowledgments

1  
2  
3 This work was funded by the Hungarian Academy of Sciences (Magyar Tudományos  
4 Akadémia [MTA], Lendület project LP2012-36/2012 to I.W.), by the German Research  
5 Foundation (Deutsche Forschungsgemeinschaft, DFG Cluster of Excellence 1077  
6 “Hearing4all”), by the German Academic Exchange Service (Deutscher Akademischer  
7 Austauschdienst [DAAD], Project 56265741), and by the Hungarian Scholarship Board  
8 (Magyar Ösztöndíj Bizottság [MÖB], Project 39589). The experiment was realized using  
9 Cogent 2000 developed by the Cogent 2000 team at the FIL and the ICN. EEG data were  
10 analyzed with EEGLab (Delorme & Makeig, 2004) and additional plugins written by Andreas  
11 Widmann, University of Leipzig. The authors are grateful to Zsuzsanna D’Albini for  
12 assistance in data acquisition.  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

## **Abstract**

1  
2  
3 In two experiments, we assessed the effects of combining different cues of concurrent sound  
4 segregation on the object-related negativity (ORN) and the P400 event-related potential  
5 components. Participants were presented with sequences of complex tones, half of which  
6 contained some manipulation: One or two harmonic partials were mistuned, delayed, or  
7 presented from a different location than the rest. In separate conditions, one, two, or three of  
8 these manipulations were combined. Participants watched a silent movie (passive listening) or  
9 reported after each tone whether they perceived one or two concurrent sounds (active  
10 listening). ORN was found in almost all conditions except for location difference alone during  
11 passive listening. Combining several cues or manipulating more than one partial consistently  
12 led to sub-additive effects on the ORN amplitude. These results support the view that ORN  
13 reflects a combined, feature-unspecific assessment of the auditory system regarding the  
14 contribution of two sources to the incoming sound.  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27

## **Keywords:**

28  
29 object-related negativity (ORN); P400; multiple congruent cues; concurrent sound  
30 segregation; active and passive listening  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

## Introduction

1  
2  
3 In everyday situations, we are constantly confronted with mixtures of sounds emitted by  
4 concurrently active sources. The human auditory system needs to parse this mixture to allow  
5 us to perceive the world in terms of meaningful objects and events. Cues that support the  
6 parsing process are traditionally divided into two main categories (Bregman, 1990; Carlyon,  
7 2004; Haykin and Chen, 2005; Snyder and Alain, 2007): those that group together sound  
8 elements along time (horizontal or sequential sound organization) and those that group them  
9 at one particular moment of time (vertical or concurrent sound organization). Concurrent  
10 segregation is based on instantaneously available cues, such as differences in pitch, sound  
11 onset, and source location. Whereas no direct event-related potential (ERP) correlate of  
12 sequential segregation has been discovered yet, concurrent segregation appears to have such  
13 an ERP correlate: The object-related negativity (ORN) component has been shown to follow  
14 the listener's perception of two concurrent sounds (Alain, Arnott, & Picton, 2001). The  
15 present study was designed to systematically investigate how combinations of the three most  
16 well-known cues of concurrent sound segregation (different source location, onset  
17 asynchrony, and inharmonic relation between the partials of complex tones) affect the ORN  
18 component. Specifically, we wished to assess whether ORN sums together the outputs of  
19 three independent detectors of concurrent sound segregation, or whether it is a read-out of the  
20 system's overall assessment of the likelihood that the sound input carries contributions from  
21 two sound sources<sup>1</sup>.

22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39 The ORN peaks between 150 and 180 ms from cue onset, reaches its maximum at  
40 frontocentral electrode sites, and inverts polarity at the mastoids (Alain, Schuler, and  
41 McDonald, 2002; Alain and McDonald, 2007). Alain and colleagues (2001) found that ORN  
42 was larger at the mastoid electrodes during active listening (listeners were required to judge  
43 whether they heard one or two concurrent sounds) than passive listening situations (listeners  
44 had no task related to the sounds), indicating attentional modulation of the ORN amplitude.

45  
46  
47  
48  
49  
50  
51 The presence and amplitude of ORN is correlated with manipulations that typically lead to  
52 listeners reporting two sound sources compared to one (Alain, Theunissen, Chevalier, Batty,  
53 and Taylor, 2003; McDonald and Alain, 2005; Alain and McDonald, 2007). Previous studies

---

54  
55  
56  
57  
58 <sup>1</sup> Another possibility is that the amplitude of the ORN reflects the number of perceived auditory objects,  
59 although modulation of the ORN amplitude by the amount of mistuning (Alain et al., 2001) makes this  
60 alternative unlikely.  
61  
62  
63  
64  
65

1 have shown that ORN can be elicited by different cues, such as inharmonicity (Alain et al.,  
2 2001 and 2002; Bendixen, Jones, Klump, and Winkler, 2010), onset asynchrony (Lipp,  
3 Kitterick, Summerfield, Bailey, and Paul-Jordanov, 2010; Weise, Schröger, and Bendixen,  
4 2012); dichotic pitch (Johnson, Hautus, and Clapp, 2003; Hautus, Johnson, and Colling,  
5 2009), separation in the fundamental frequency of speech sounds (Snyder and Alain, 2005;  
6 Alain, Reinke, He, Wang, and Lobaugh, 2005), and simulated echo (Sanders, Joh, Keen, and  
7 Freyman, 2008; Sanders, Zobel, Freyman, and Keen, 2008). There are also some reports of  
8 ORN emerging with a combination of some of the above cues, such as inharmonicity and  
9 location difference (McDonald and Alain, 2005) or inharmonicity and onset asynchrony  
10 (Weise et al., 2012).

11 ORN is elicited in both passive and active listening situations (Alain et al., 2001, 2002; Alain,  
12 2007) and its amplitude is independent of the task demands (Alain and Izenberg, 2003). In  
13 active listening situations, ORN elicitation is accompanied by a late positive wave that peaks  
14 about 400 ms after stimulus onset, the P400 component. P400 amplitude also correlates with  
15 the likelihood of perceiving two concurrent sound objects compared to one (Alain et al., 2001,  
16 2002; Hautus & Johnson, 2005), but P400 does not follow the ORN in an obligatory manner  
17 (Johnson, Hautus, Duff, and Clapp, 2007). Johnson et al. (2007) proposed that P400 is  
18 influenced by the task context. In their study, they used two different tasks: In the detection  
19 task, participants were to indicate whether they heard dichotic pitch or a control stimulus,  
20 whereas in the localization task, only dichotic pitch stimuli were presented, and participants  
21 were to decide where the sound was located. In the latter case, no P400 was elicited (Johnson  
22 et al., 2007).

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

31 Previous studies suggested that the ORN amplitude is modulated by the strength or salience of  
32 the cues supporting the segregation of concurrent sounds. For example, Alain and colleagues  
33

1 (2001) found larger ORN amplitudes with increasing amounts of inharmonicity (larger ORN  
2 amplitude for the 16% than 8%, or 4% mistuning). In this study, participants reported hearing  
3 two sounds more often with higher amounts of mistuning. In another paradigm, using dichotic  
4 pitch, Clapp, Johnson, and Hautus (2007) found that the largest ORN was elicited in response  
5 to the most salient dichotic pitch cue, and the ORN amplitude decreased with decreasing cue  
6 saliency. These authors also found a similar pattern for P400 amplitude.  
7  
8  
9

10  
11 Perception of concurrent sounds can be made more likely not only by strengthening one  
12 particular cue (e.g., increasing the amount of mistuning for inharmonicity-based segregation,  
13 cf. Alain et al., 2001), but also by combining two different cues (e.g., frequency and location).  
14 In this case, the multiple congruent cues may strengthen the impression of the presence of  
15 separate sound sources. Using MEG, such a combined effect was found in a speech  
16 segregation task (Du, He, Ross, Bardouille, Wu, Li, and Alain, 2011). Du and colleagues  
17 (2011) hypothesized that separation in both base frequency and source location contribute to  
18 speech segregation, and combining these cues would result in additivity or superadditivity  
19 between the ORN components elicited by the two cues, separately. They found that the ORN  
20 elicited by the combination of the base-frequency separation and the location cue equaled the  
21 sum of the responses elicited by the two cues alone. A similar effect of summing two different  
22 types of cues was obtained by Hautus and colleagues (2009), although these authors did not  
23 directly test whether the effect of the cue combination was strictly additive when compared to  
24 the sum of the effects of the single cues alone (see also McDonald and Alain, 2005; Weise et  
25 al., 2012).  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39

40 Here we report the results of a study in which we systematically investigated combinations of  
41 inharmonicity, onset asynchrony, and location difference under passive (Experiment 1) and  
42 active (Experiment 2) listening conditions. Based on previous studies (Alain et al., 2001,  
43 2002, 2003; McDonald & Alain, 2005), we expected that ORN will be present in both  
44 listening situations, whereas P400 will only be present in the active listening situation. First,  
45 we tested whether the salience of the harmonicity-based cue can be further increased by  
46 mistuning two partials in a congruent manner (as opposed to mistuning only one partial). We  
47 hypothesized that mistuning two partials would enhance the ORN amplitude by providing  
48 redundant information for harmonicity-based segregation. Second, we aimed to assess the  
49 effects of combining different cues of concurrent segregation on ORN and P400. As the cues  
50 are congruent in supporting the same decomposition of the input into two sounds in  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1 perception, we regard them as redundant with respect to concurrent sound segregation. By  
2 investigating whether the effects of combined cues are additive, sub- or superadditive  
3 compared to the single-cue effects, our goal was to separate two possible interpretations of the  
4 ORN component. It is possible that each cue elicits a separate ORN response and the  
5 observed response sums together the individual ORN components. In this case, the ORN  
6 elicited by multiple congruent cues will be as large as the summed amplitudes of the ORN  
7 components elicited by the contributing cues. This would suggest that ORN reflects processes  
8 that are closely related to cue evaluation and farther upstream from what appears in  
9 perception. Alternatively, ORN may reflect the system's overall assessment of the likelihood  
10 that the auditory input consists of two concurrent sounds. That is, ORN could reflect the  
11 readout of a process combining the evidence from the available cues. In this case, depending  
12 on the way the cues are combined, we should find sub- or superadditivity between the  
13 contributing cues' ORN components. Subadditivity of the contributing cues' ORN amplitudes  
14 would occur for example if the cue-combination algorithm evaluated the likelihood of the  
15 presence of two concurrent sounds by passing on the signal resulting from the most salient  
16 cue. Superadditivity of the ORN amplitudes would occur if cue combination took into account  
17 partial cues, which alone would not be sufficient to support the presence of two concurrent  
18 sounds. The two methods are not mutually exclusive; thus one may find both sub- and  
19 superadditivity depending on the strength of the available contributing cues. Either one of  
20 these possibilities would mean that the process reflected by ORN is less directly related to cue  
21 evaluation; rather it is closer to what appears in perception.  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41

## 42 **EXPERIMENT 1**

### 43 **Methods**

#### 44 *Participants*

45  
46  
47  
48  
49  
50  
51  
52  
53 Twenty healthy volunteers (eight female, mean age 23.5 years, SD = 2.42) participated in the  
54 experiment. Participants received modest financial compensation. None of the participants  
55 were taking any medication affecting the central nervous system. Prior to the beginning of the  
56 experiment, written informed consent was obtained from each participant according to the  
57  
58  
59  
60  
61  
62  
63  
64  
65

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

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<sup>nd</sup>) or two (2<sup>nd</sup> and 4<sup>th</sup>) partials. Three simple manipulations and their combinations (altogether 11 different manipulations) were tested. The simple manipulations were: a) mistuning the 2<sup>nd</sup> partial (or 2<sup>nd</sup> and 4<sup>th</sup> 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  $\mu$ s 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-



1 right, manipulated-left, and manipulated-right tones each made up 25% of the stimuli. The  
2 location difference cue for the manipulated partial was set up by employing the opposite  
3 parameter combination than for the rest of the partials, thus creating a ca. 90° location  
4 difference between the manipulated and the other partials. A summary of the experimental  
5 manipulations is given in Table 1.  
6  
7  
8  
9

10  
11  
12  
13 -----Insert Table 1 about here -----  
14  
15  
16  
17  
18

19 Each stimulus condition was presented in a separate block consisting of 200 base and 200  
20 manipulated tones. Stimulus blocks commenced with 10 base-version tones, which were  
21 excluded from the analyses. In each stimulus block, all tones had the same fundamental  
22 frequency, whilst the fundamental frequency changed from block to block. Eleven  
23 fundamental frequencies were used with the lowest frequency being 200 Hz, and the rest  
24 following in one-semitone steps (i.e., the highest fundamental frequency being 378 Hz). The  
25 order of the fundamental frequencies and the order of the different stimulus blocks  
26 (conditions) were randomized separately for each participant.  
27  
28  
29  
30  
31  
32  
33  
34

35 Participants watched a silent, subtitled movie of their own choice on a computer screen that  
36 was placed in front of them at a distance of 1.15 m. They were asked to ignore the sounds.  
37 Total duration of the experimental blocks amounted to 83 minutes. Short breaks were inserted  
38 between stimulus blocks with at least one longer break, set between the 6<sup>th</sup> and the 7<sup>th</sup>  
39 stimulus block, when the participant was allowed to leave the chamber. Further longer breaks  
40 were inserted if the participant asked for it. The total time of the session (including electrode  
41 mounting and removal) was ca. 3 hours.  
42  
43  
44  
45  
46  
47  
48  
49  
50

### 51 *Electrophysiological recording and data analysis*

52  
53

54 Electroencephalogram (EEG) was continuously recorded with Ag/AgCl electrodes. 63  
55 electrodes were placed on the scalp according to the extended international 10–20 system  
56 (Jasper, 1958, Chatrian, Lettich and Nelson, 1985). An additional electrode was placed on the  
57 tip of the nose, which served as the reference. Eye movements were monitored by bipolar  
58  
59  
60  
61  
62  
63  
64  
65

1 recordings from two electrodes placed above and below the left eye (vertical  
2 electrooculogram, VEOG) and two placed lateral to the outer canthi of both eyes (horizontal  
3 electrooculogram, HEOG). EEG and EOG signals were amplified (0–40 Hz) by SynAmps  
4 amplifiers (Neuroscan Inc.), sampled at 500 Hz. Data were resampled to 250 Hz and filtered  
5 off-line using a 0.1–30 Hz band-pass finite impulse response (FIR) filter (Kaiser windowed,  
6 Kaiser  $\beta = 5.65$ , filter length 4530 points).  
7

8  
9  
10 For each tone, an epoch of 400-ms duration including a 100 ms pre-stimulus baseline was  
11 extracted from the continuous EEG record. Epochs with an amplitude change exceeding 100  
12  $\mu\text{V}$  at any electrode were rejected from further analysis, which led to retaining 84.0% of the  
13 responses on average. Epochs for the two stimulus types (base version and manipulated) were  
14 separately averaged for each of the 11 conditions, collapsing over the two possible locations  
15 (left vs. right presentation).  
16  
17

18  
19  
20  
21  
22 Difference waveforms were calculated between ERPs elicited by the manipulated and the  
23 corresponding base tones for identifying and measuring the ORN component. Except for the  
24 scalp topography analyses, all measurements were taken from the recordings at the Cz  
25 electrode. Average ORN amplitudes were measured from 72-ms wide windows centered on  
26 the average peak latency for each condition. To account for the observed latency variation  
27 between conditions, peak latencies were determined separately for each condition by the  
28 jackknifing method (Kiesel, Miller, Jolicœur & Brisson, 2008; Miller, Ulrich & Schwarz,  
29 2009). Epochs for the base versions and manipulated tones were averaged separately.  
30  
31  
32

33  
34  
35  
36  
37  
38 Following visual inspection of the responses, the N1 amplitude differences were also  
39 investigated. N1 difference amplitudes were measured from 40-ms wide windows centered on  
40 the average peak latency for each condition.  
41  
42

43  
44  
45 All ERP difference amplitudes were tested against zero using one-sample, two-tailed  $t$  tests.  
46 For testing possible differences in the ORN amplitudes and scalp distributions across the three  
47 single-cue manipulations, ORN amplitudes were averaged separately for the following six  
48 electrode clusters: left frontal (Fp1, AF7, AF3, F7, F5, F3, F1), left central (FT7, FC5, FC3,  
49 FC1, C5, C3, C1), left parietal (CP5, CP3, CP1, P7, P5, P3, P1), right frontal (Fp2, AF8, AF4,  
50 F8, F6, F4, F2), right central (FT8, FC6, FC4, FC2, C6, C4, C2), right parietal (CP6, CP4,  
51 CP2, P8, P6, P4, P2). ORN amplitudes and scalp topographies were then compared by a  
52 repeated-measures ANOVA with the factors *Manipulation* (3 levels: 2<sup>nd</sup> partial mistuned vs.  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

2<sup>nd</sup> partial delayed vs. 2<sup>nd</sup> 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<sup>nd</sup> partial vs. 2<sup>nd</sup> and 4<sup>th</sup> 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  $\eta^2$  effect size measure. The Greenhouse–Geisser correction was applied when the assumption of sphericity was violated; the  $\epsilon$  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* [ $F(2,38) = 12.840, p = 0.001, \eta^2 = 0.403, \epsilon = 0.63$ ], 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

1 manipulated partials were delayed and/or had a different location than the rest: 2<sup>nd</sup> partial  
2 delayed, 2<sup>nd</sup> partial mistuned and delayed, 2<sup>nd</sup> and 4<sup>th</sup> partials mistuned and delayed, 2<sup>nd</sup>  
3 partial with location difference, 2<sup>nd</sup> partial delayed with location difference and 2<sup>nd</sup> and 4<sup>th</sup>  
4 partials mistuned, delayed and with location difference (see Table 2 for a full list of the mean  
5 difference amplitudes and the results of the corresponding *t* tests against zero).  
6  
7  
8  
9

## 10 11 12 **Discussion**

13  
14  
15 In Experiment 1, we studied the ORN components elicited by three different cues of  
16 concurrent sound segregation and their combination in a passive listening situation. All of  
17 these cues and cue combinations elicited significant ORN components, except for location  
18 difference alone, which appeared to be a weaker cue of ORN elicitation with the current  
19 parameters.  
20  
21  
22  
23

24  
25 We found no significant increase of the ORN amplitude when congruently manipulating  
26 multiple partials, i.e., there was no significant difference between those conditions where only  
27 one partial was manipulated as compared to those conditions where two partials were  
28 manipulated. Similarly, adding delay and/or location difference on top of mistuning did not  
29 lead to a significant increase in the ORN amplitude. Furthermore, combining several cues  
30 always elicited numerically (and in most cases significantly) smaller ORN amplitudes than  
31 the sum of the contributing ORN amplitudes. In other words, multiple congruent cues were  
32 processed in a subadditive manner. We found no evidence pointing towards superadditivity  
33 for any of the combinations, nor did any of the combinations appear to follow a strictly  
34 additive model. Note that the amount of mistuning employed in the current study (+8%) did  
35 not force a ceiling effect on the ORN amplitude, as a previous study found an increase of the  
36 ORN amplitude by increasing the amount of mistuning from 8 to 16% (Alain et al., 2001).  
37 Taken together, these results suggest that ORN may reflect a combined assessment of the  
38 likelihood of the presence of two concurrent sounds, as opposed to summing the strength of  
39 sensory evidence for the presence of two concurrent sounds.  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52

53  
54 Some effects of the cues of concurrent sound segregation were observed in a latency range  
55 preceding that of the ORN. Specifically, significantly larger N1 components were elicited in  
56 conditions where one or two partials were delayed or presented with location difference and  
57 in some of the conditions where these cues appeared in combination. These results were  
58  
59  
60  
61  
62  
63  
64  
65

1 unexpected, and the current paradigm was not designed to separate whether the N1 increase  
2 was related to the presence of two concurrent sounds or to the specific acoustic manipulations.  
3 A test of this issue was therefore included in the follow-up Experiment 2. We introduced  
4 control blocks in which only one sound object was delivered at any time. To control for the  
5 delay, we tested tones with the two (2<sup>nd</sup> and 4<sup>th</sup>) partials omitted; thus the initial 100-ms  
6 segment was identical to the delay manipulation, but no additional partials commenced after  
7 100 ms as that would promote concurrent sound segregation. To control for the location of the  
8 tones, we recorded responses separately for the two source locations, using only the base  
9 versions of the tones. If the acoustic manipulations accounted for the N1 effect, the difference  
10 would be apparent between the base versions presented in the two different locations.  
11  
12  
13  
14  
15  
16  
17  
18

19 Besides these control blocks, the main purpose of the follow-up Experiment 2 was to repeat  
20 the manipulations employed in Experiment 1 in an active listening situation. Previous studies  
21 (Alain et al., 2001; 2002) have shown that ORN is also elicited during active listening, and  
22 that it is followed by a late positive peak (the P400) when listeners are asked to give  
23 perceptual judgments as to the presence of one or two concurrent sounds. Thus in Experiment  
24 2 we investigated whether a) a similar pattern for the processing of the cues and their  
25 combinations is observed when participants are asked to attend to the sounds, b) whether this  
26 pattern translates into perceptual judgments of the sounds as coming from one or two sources,  
27 c) how the different cues and combinations affect the P400 response. Finally, we also  
28 assessed d) whether attention affects the ORN amplitude with multiple concurrent cues.  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40

## 41 **EXPERIMENT 2**

### 42 43 44 45 **Methods**

#### 46 47 48 *Participants*

49  
50  
51  
52  
53 Twenty-three healthy volunteers (twelve female, mean age 22.1 years, SD = 1.62) participated  
54 in the experiment. None of the participants had taken part in Experiment 1.  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

## *Apparatus and stimuli*

1  
2  
3 The stimulus paradigm employed in Experiment 2 was identical to that of Experiment 1 with  
4 the following exceptions.  
5  
6

7  
8 In the beginning of the experimental session, two N1 control blocks were administered. In one  
9 stimulus block, two partials (2<sup>nd</sup> and 4<sup>th</sup>) were omitted from the complex tone taking the place  
10 of the manipulated sounds of the corresponding stimulus block of Experiment 1 (50%),  
11 whereas the base version of the sound stayed the same (50%). The order of the base and  
12 manipulated sounds was fully randomized. In the other control block, only the base versions  
13 of the left and right tones were delivered. As in Experiment 1 as well as all other stimulus  
14 blocks of Experiment 2, in the two control blocks, half of the tones were presented with  
15 parameters promoting the listener to perceive the tones as originating from ca. 45° right and  
16 the other half from 45° left from the midline; tones with the two perceived locations were  
17 delivered in a fully randomized order. All other stimulus parameters were identical to those of  
18 Experiment 1. During these control blocks, participants watched a subtitled, silent movie and  
19 were asked to disregard the sounds.  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29

30 For the remainder of the session, participants were given two response keys (one in each  
31 hand), and were instructed to perform tasks as detailed below by pressing one or the other key  
32 with their left or right thumb.  
33  
34  
35  
36

37 The next (3<sup>rd</sup>) stimulus block served as control for a different analysis, which is not reported  
38 here. In this stimulus block, half of the sounds were base-version complex tones, whereas for  
39 the other half, the 2<sup>nd</sup> and 4<sup>th</sup> partials were mistuned, delayed, and with location difference as  
40 described for Experiment 1 (condition 11). 140 stimuli of the base version and 140 of the  
41 manipulated version were delivered with an onset-to-onset interval of 1400 ms. Participants  
42 were instructed to watch a fixation cross continuously present at the center of the computer  
43 screen placed at 1.15 m directly in front of them (visual angle of 0.4°) and to press either one  
44 of the responses buttons when the fixation cross changed to an “X” for 100 ms, after which it  
45 returned to the regular “+” sign. The change appeared at a random time point between 550  
46 and 750 ms after each tone onset. Participants were asked to ignore the tones.  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56

57 Participants then received two blocks of training (blocks 4 and 5) in the task they were asked  
58 to do during the rest of the stimulus blocks. In the first training block, 20 base-version tones  
59  
60  
61  
62  
63  
64  
65

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

and 20 tones with both the 2<sup>nd</sup> and 4<sup>th</sup> 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 2<sup>nd</sup> partial being mistuned, 10 with the 2<sup>nd</sup> partial being delayed, and 10 with the 2<sup>nd</sup> 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  $\mu$ V 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)  $\times$  *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<sup>nd</sup> partial mistuned vs. 2<sup>nd</sup> partial delayed vs. 2<sup>nd</sup> 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)  $\times$  *Number of mistuned partials* (2 levels: 2<sup>nd</sup> partial vs. 2<sup>nd</sup> and 4<sup>th</sup> partials)  $\times$  *Delay* (2 levels: delay present vs. absent)  $\times$  *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)  $\times$  *Manipulation* (3 levels: 2<sup>nd</sup> partial mistuned vs. 2<sup>nd</sup> partial delayed vs. 2<sup>nd</sup> partial with location difference)  $\times$  *Frontality* (3 levels: frontal

1 vs. central vs. parietal)  $\times$  *Laterality* (2 levels: left, right), where each electrode cluster (as  
2 defined for Experiment 1) was represented by the mean amplitude measured from the  
3 electrodes in the cluster.  
4  
5

6 For the N1 control blocks, difference waveforms were calculated between the responses  
7 elicited by the two types of tones. N1 amplitudes were measured in 40-ms wide intervals  
8 centered on the average peak latency per condition. N1 difference amplitudes were tested  
9 against zero using one-sample, two-tailed *t* tests.  
10  
11  
12  
13

14 In all other respects, methods were the same as in Experiment 1.  
15  
16  
17  
18  
19

## 20 **Results**

### 21 *Behavioral data*

22  
23  
24  
25  
26 The percentage of the corresponding one-sound answers to the base versions was 95.43%,  
27 while the percentage of corresponding two-sound answers to the manipulated versions was  
28 85.69% (averaged across the 11 conditions; see Table 4).  
29  
30  
31  
32

33 The repeated-measures ANOVA showed a significant main effect of *Type of tone* [ $F(1,22) =$   
34  $28.699, p < 0.001, \eta^2 = 0.566$ ], where the number of corresponding answers to base versions  
35 was significantly higher than the corresponding answers to the manipulated versions of tones.  
36  
37 There was also a significant main effect of *Condition* [ $F(10,220) = 74.298, p < 0.001, \eta^2 =$   
38  $0.772, \varepsilon = 0.413$ ], which was due to a significantly smaller number of corresponding answers  
39 in the condition with the 2<sup>nd</sup> partial with location difference than in all the other conditions  
40 (all *p* values  $< 0.001$ ). A significant interaction of *Type of tone* and *Condition* was also found  
41 [ $F(10,220) = 63.953, p < 0.001, \eta^2 = 0.744, \varepsilon = 0.235$ ]. This was due to a significant main  
42 effect of *Condition* for corresponding answers to the manipulated versions [ $F(10,220) =$   
43  $95.622, p < 0.001, \eta^2 = 0.813, \varepsilon = 0.286$ ]. On the other hand, no significant main effect of  
44 *Condition* for corresponding answers to the base versions was found [ $F(10,220) = 1.578, p =$   
45  $0.2, \eta^2 = 0.067, \varepsilon = 0.317$ ].  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59

60 -----Insert Table 4 about here -----  
61  
62  
63  
64  
65

*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 2<sup>nd</sup> partial delayed having larger amplitudes than the 2<sup>nd</sup> partial with location difference ( $p = 0.081$ ). There was also a significant main effect of *Frontality* [ $F(2,44) = 19.787, p < 0.001, \eta^2 = 0.474, \epsilon = 0.635$ ], 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.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

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: 2<sup>nd</sup> and 4<sup>th</sup> partials mistuned, 2<sup>nd</sup> partial with location difference, 2<sup>nd</sup> and 4<sup>th</sup> partials mistuned with location difference and 2<sup>nd</sup> and 4<sup>th</sup> 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* [ $F(2,44) = 6.145, p = 0.004, \eta^2 = 0.218, ]$  due to significantly larger amplitudes in the 2<sup>nd</sup> partial delayed condition than in the 2<sup>nd</sup> partial with location difference condition ( $p = 0.015$ ). P400 amplitude in the 2<sup>nd</sup> 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 2<sup>nd</sup> partial delayed plus location difference with 2<sup>nd</sup> partial mistuned; superadditivity for the combination of 2<sup>nd</sup> partial delayed with 2<sup>nd</sup> partial with location difference). No other significant results were obtained (see Table 7 for all results).

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

1  
2  
3  
4  
5 Comparing ORN between the active and the passive listening conditions  
6

7  
8 In the mixed-model ANOVA comparing the ORN amplitudes between Experiments 1 and 2,  
9 only *Listening condition* (i.e., the difference between the two experiments) had a significant  
10 effect [ $F(1,19) = 6.9535, p = 0.016, \eta^2 = 0.268$ ]. ORN amplitudes were larger in the active  
11 than in the passive listening situation. No other main effects were observed, and no  
12 interactions between *Listening condition* and any of the other factors (*Number of mistuned*  
13 *partials, Delay, and Location difference*; all  $p$  values  $> 0.078$ ).  
14  
15  
16  
17  
18  
19

20 When comparing the scalp topographies of the ORN components between the passive and  
21 active listening conditions, no main effect of or interaction involving the *Listening condition*  
22 was found. The significant main effects of *Manipulation* [ $F(2,38) = 4.349, p = 0.019, \eta^2 =$   
23  $0.186$ ] and of *Frontality* [ $F(2,38) = 35.206, p < 0.001, \eta^2 = 0.649, \varepsilon = 0.635$ ] replicated the  
24 similar results obtained in the analyses that were conducted separately for the two  
25 experiments, with the post hoc tests also showing the same origin for these effects.  
26  
27  
28  
29  
30  
31

32  
33  
34  
35 *N1*  
36

37  
38 No significant N1 differences were found between the two different tones in either of the N1  
39 control conditions. In the condition controlling for the effects of delay, the mean amplitude  
40 difference at Cz was  $-0.4809 \mu\text{V}$  ( $t(22) = -1.2619, p = 0.22$ ), and in the condition controlling  
41 for the location difference the mean amplitude difference was  $0.2798 \mu\text{V}$  ( $t(22) = 0.9284, p =$   
42  $0.36$ ). The N1 effects are shown in Figure 5.  
43  
44  
45  
46  
47

48 In contrast, manipulated tones elicited significantly larger N1 components than the base  
49 versions of the tones in most of the 11 stimulus conditions (see Table 5, top for all results).  
50  
51  
52  
53  
54  
55

56 -----Insert Figure 5 about here-----  
57  
58  
59  
60  
61  
62  
63  
64  
65

## Discussion

1  
2  
3 In Experiment 2, all investigated cues and cue combinations elicited significant ORN  
4 responses. Even the condition with location difference alone showed a significant ORN  
5 response, although participants' perceptual judgments showed that few of the manipulated  
6 tones evoked the perception of two concurrent sounds in this stimulus condition. This  
7 suggests that the cues of concurrent sound segregation employed in the current experiment  
8 were picked up by the processes underlying ORN generation. On the other hand, ORN cannot  
9 fully govern perceptual judgments. Many psychophysical studies have shown that perceptual  
10 judgments are codetermined by the criteria used by the observer in deciding between the  
11 behavioral alternatives (Green and Swets, 1966). The present instructions did not attempt to  
12 manipulate the decision level – listeners were free to choose their criteria. Given that in most  
13 stimulus blocks the cues (when present) strongly promoted perception of two concurrent  
14 sounds, it is plausible to assume that listeners accommodated to the high distinctiveness of  
15 these cues and set their criteria to be high. This should have resulted in dismissing most  
16 exemplars with the relatively weak location cue while the obligatory evaluation of the  
17 auditory system – as reflected by the elicitation of the ORN component – suggested the  
18 presence of two concurrent sound sources. The significantly higher percentage of  
19 corresponding responses for the base version tones is also compatible with this interpretation.  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33

34  
35 In full correspondence with the results obtained in Experiment 1, ORN amplitude did not  
36 significantly increase with congruently manipulating two partials (as compared to  
37 manipulating only one), or with adding delay and/or location difference on top of mistuning.  
38 Furthermore, combining several cues tended to elicit ORN components of smaller amplitude  
39 than the sum of the contributing ORN components. No superadditive effects were found for  
40 any cue combination. Further, no significant differences were found between the ORN effects  
41 or scalp topographies between the two experiments. The results of Experiment 2 thus  
42 replicated those obtained in Experiment 1 and support the interpretation given for them in the  
43 Discussion of Experiment 1.  
44  
45  
46  
47  
48  
49  
50

51  
52 The P400 results were not as clear-cut as those for the ORN component. In some cases, no  
53 significant P400 was observed, even though the ORN was elicited in those conditions as well.  
54 This contrasts the results of some previous studies (Alain et al., 2002). Note, however, that  
55 even in the conditions with non-significant P400, positive deflections in the P400 latency  
56 range were observed in the difference waveforms between the ERPs elicited by the  
57  
58  
59  
60  
61  
62  
63  
64  
65

1 manipulated and the base tones (cf. Figure 3). It is possible that the lack of significant  
2 detection of the P400 component was caused by some of the preceding ORN components not  
3 yet having terminated at the onset of the P400.  
4

5  
6 In terms of cue redundancy, results for the P400 were equivocal: we found significant  
7 superadditivity in one case and significant subadditivity in another case. Altogether there was  
8 no clear tendency towards either pattern. Again, these results may be partly obscured by the  
9 preceding ORN components.  
10

11  
12 Significant N1 differences were obtained in most conditions. Unlike in the passive listening  
13 situation (Experiment 1), in the active listening situation, these N1 differences were not  
14 confined to conditions where delay and location difference manipulations were employed, but  
15 also extended to conditions with mistuning alone. The two control conditions suggest that  
16 these N1 effects were not solely due to the acoustic differences between the base and the  
17 manipulated tones, but rather they may reflect some aspect of processing concurrent sounds.  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29

## 30 **General discussion**

31  
32 In the present study, we systematically combined three cues of concurrent sound segregation  
33 (mistuning, onset asynchrony, and location difference) for testing how the ORN event-related  
34 potential component (and in Experiment 2, also the P400 component as well as perceptual  
35 judgments) reflects the joint evaluation of these cues. We employed two listening conditions  
36 in two separate experiments in which participants were instructed to either disregard the tones  
37 (passive listening) or to focus their attention on the tones and judge whether they heard one or  
38 two sounds (active listening). The pattern of ORN elicitation in response to the different cues  
39 and their combination was highly similar under the two listening conditions. This pattern is  
40 consistent with the notion that the ORN response reflects the auditory system's overall  
41 assessment of the likelihood that the sound input carries contributions from two sound  
42 sources, rather than summing together the outputs of independent detectors of concurrent  
43 sound segregation.  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53

54  
55 We found that ORN was elicited by all of the tested combinations of cues and also by each of  
56 the cues individually, with the exception of location difference alone during passive listening.  
57 Location difference was also the weakest cue during active listening; although it elicited a  
58  
59  
60  
61  
62  
63  
64  
65

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
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).

46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
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 2<sup>nd</sup>) but two (the 2<sup>nd</sup> and 4<sup>th</sup>) 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



1 al., 2007). The present results suggest that manipulating more than one partial of a complex  
2 sound does not lead to a similar increase in strength or saliency. Alternatively, it is possible  
3 that the 4<sup>th</sup> partial is not sufficiently influential in assessing the source of complex sounds (cf.  
4 Alain et al., 2001, who showed that mistuning the 4<sup>th</sup> partial alone causes weaker effects than  
5 mistuning the 2<sup>nd</sup> partial alone; therefore, the 2<sup>nd</sup> partial may have dominated the present  
6 results).  
7  
8  
9

10  
11 Besides manipulating more than one partial, we pursued a second approach for increasing the  
12 strength of the sensory evidence in favor of concurrent sound segregation. This approach was  
13 based on employing multiple cues in parallel (i.e., onset asynchrony and/or location difference  
14 in addition to mistuning). Previous studies suggested that ORN increases with such cue  
15 combinations (Hautus et al., 2009; McDonald and Alain, 2005, Weise et al., 2012), and that  
16 this increase may follow a fully additive pattern (Du et al., 2011). In contrast, here we mostly  
17 found subadditivity for the amplitude of ORN elicited by multiple cues. That is, the  
18 combinations of cues of concurrent segregation elicited lower-amplitude ORN components  
19 compared with the sum of the ORN amplitudes elicited by each contributing cue separately.  
20 In most cases, the increase of the ORN amplitude from single to multiple cues was non-  
21 significant, as was shown by the ANOVAs assessing the effects of multiple congruent cues.  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31

32  
33 One reason for these weak, subadditive effects of the cue combinations may be that the  
34 employed cues, at least as far mistuning and onset asynchrony are concerned, were clearly  
35 supra-threshold: Each of them alone was sufficient to elicit as much as 90% correspondence  
36 between the presence of the cue and the perceptual judgment. Hence the cues can be regarded  
37 as fully redundant with respect to each other. Similarly, adding the (weak) location cue on top  
38 of a strong cue of concurrent segregation probably also provided only redundant information  
39 (cf. McDonald and Alain, 2005, for high amounts of mistuning). Therefore, one may not be  
40 surprised that ORN amplitude does not increase further by adding more cues. Note, however,  
41 that this explanation implies that the ORN reflects the outcome of a process that combines the  
42 sensory evidence to provide an overall assessment of the likelihood that one or two sound  
43 sources were present in the environment. If, on the contrary, ORN were to reflect the strength  
44 of the sensory evidence that drives the decision between one and two sound sources, then  
45 each additional cue should increase ORN amplitude in an additive manner, even if it is  
46 redundant with the other cues in terms of the eventual perceptual decision. Our results are not  
47 consistent with this latter view; they support the interpretation that ORN underlies the actual  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1 perceptual decision. Subadditivity between the ORN components elicited by the contributing  
2 cues suggests that the likely method of combining cues of concurrent stream segregation is  
3 the selection of the individually most salient cue. Several authors have already argued that the  
4 ORN probably reflects a perceptual grouping mechanism rather than a cue-related response;  
5 their arguments were based on the highly different nature of the stimuli eliciting ORN (e.g.,  
6 Hautus and Johnson 2005; Hautus et al., 2009; Johnson et al. 2007; Lipp et al., 2010;  
7 McDonald & Alain, 2005). We add here the argument of subadditivity for multiple congruent  
8 cues. Another supporting piece of evidence in the present data is given by the highly similar  
9 ORN topographies across different types of cues and listening conditions (cf. Figure 2).  
10  
11  
12  
13  
14  
15

16  
17 The P400 component was less tightly related to the perceptual reports. Unlike in previous  
18 studies (Alain et al., 2001, 2002; Hautus & Johnson, 2005), P400 failed to reach significance  
19 in some conditions despite clear perceptual distinction between one- and two-sound objects.  
20 One might speculate that a procedural difference between our and previous studies may  
21 account for this (namely, subjects were to withhold their response for several hundred  
22 milliseconds). However, a similar dissociation between P400 and perceptual decisions was  
23 observed by Hautus and colleagues (2009) with the instruction to respond as quickly as  
24 possible. Hence the results are more in line with the view that ORN and P400, as well as P400  
25 and behavior, are not as tightly connected as previously assumed (Johnson et al, 2007). In  
26 terms of cue redundancy, there was no clear sub-, super- or fully additive pattern for the P400;  
27 the results are thus not informative regarding this question.  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37

38  
39 Unexpectedly, we observed effects related to the cues of concurrent segregation also in the N1  
40 latency range. Some previous studies reported mistuning-related effects preceding the ORN  
41 latency range in MEG (Alain and McDonald, 2007; Lipp et al., 2010). These effects were,  
42 however, even earlier than in the N1 range; it remains unclear whether they relate to the  
43 present effects. Because our control conditions suggest that the N1 effects were not caused by  
44 the acoustic differences between the base and the manipulated stimuli, we tentatively suggest  
45 that the N1 differences are related to the automatic processing of the cues of concurrent sound  
46 segregation, but not necessarily to the resulting percept (following the interpretations of Alain  
47 and McDonald, 2007, as well as Lipp et al., 2010).  
48  
49  
50  
51  
52  
53  
54  
55

56 Finally, a significant effect of listening condition was found for the ORN amplitude, with  
57 larger amplitudes during active than passive listening. This is in accordance with some (e.g.  
58 Alain et al., 2001) but not all (e.g. Alain & Izenberg, 2003; Lipp et al., 2010) previous studies.  
59  
60  
61  
62  
63  
64  
65

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
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.

23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
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

- 1  
2  
3  
4 Alain, C. (2007). Breaking the wave: Effects of attention and learning on concurrent sound  
5 perception. *Hearing Research*, 229, 225–236.  
6  
7  
8 Alain, C., Arnott, S.R., Picton, T.W. (2001). Bottom-up and top-down influences on  
9 auditory scene analysis: Evidence from brain potentials. *Journal of Experimental*  
10 *Psychology*, 27, 1072-1089.  
11  
12  
13 Alain, C., Izenberg, A. (2003). Effects of attentional load on auditory scene analysis.  
14 *Journal of Cognitive Neuroscience*, 15, 1063-1073.  
15  
16  
17 Alain, C., McDonald, K.L. (2007). Age-related differences in neuromagnetic brain activity  
18 underlying concurrent sound perception. *Journal of Neuroscience*, 27, 1308–1314.  
19  
20  
21 Alain, C., Reinke, K. S., He, Y., Wang, C., Lobaugh, N. (2005). Hearing two things at  
22 once: Neurophysiological indices of speech segregation and identification. *Journal*  
23 *of Cognitive Neuroscience*, 17, 811-818.  
24  
25  
26 Alain, C., Schuler, B.M., McDonald, K.L. (2002). Neural activity associated with  
27 distinguishing concurrent auditory objects. *Journal of the Acoustical Society of*  
28 *America*, 111, 990-995.  
29  
30  
31 Alain, C., Theunissen, E.L., Chevalier, H., Batty, M., Taylor, M.J. (2003). Developmental  
32 changes in distinguishing concurrent auditory objects. *Cognitive Brain Research*,  
33 16, 210–218.  
34  
35  
36 Bendixen, A., Jones, S.J., Klump, G., Winkler, I. (2010). Probability dependence and  
37 functional separation of the object-related and mismatch negativity event-related  
38 potential components. *Neuroimage*, 50, 285-290.  
39  
40  
41 Bregman, A.S. (1990). *Auditory Scene Analysis: The Perceptual Organization of*  
42 *Sound*. Cambridge, Massachusetts: The MIT Press.  
43  
44  
45 Carlyon, R.P. (2004). How the brain separates sounds. *Trends in Cognitive Science*, 8,  
46 465-471.  
47  
48  
49 Chatrjian, G.E., Lettich, E., Nelson, P.L. (1985). Ten percent electrode system for  
50 topographic studies of spontaneous and evoked EEG activity. *American Journal of*  
51 *EEG Technology*, 25, 83-92.  
52  
53  
54 Clapp, W.C., Johnson, B.W., Hautus, M.J. (2007). Graded cue information in dichotic  
55 pitch: effects on event-related potentials. *Neuroreport*, 18, 365-368.  
56  
57  
58 Delorme, A., Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-  
59 trial EEG dynamics including independent component analysis. *Journal of*  
60 *Neuroscience Methods*, 134, 9-21.  
61  
62  
63  
64  
65

- 1 Du, Y., He, Y., Ross, B., Bardouille, T., Wu, X., Li, L., Alain, C. (2011). Human auditory  
2 cortex activity shows additive effects of spectral and spatial cues during speech  
3 segregation. *Cerebral Cortex* 21, 698-707.  
4
- 5 Green D.M, Swets J.A. Signal detection theory and psychophysics. New York: Wiley;  
6 1966.  
7
- 8 Hautus, M.J., Johnson, B.W. (2005). Object-related brain potentials associated with the  
9 perceptual segregation of a dichotically embedded pitch. *Journal of the Acoustical*  
10 *Society of America*, 117, 275–280.  
11
- 12 Hautus, M.J., Johnson, B.W., Colling, L.J. (2009). Event-related potentials for interaural  
13 time differences and spectral cues. *Neuroreport*, 20, 951-956.  
14
- 15 Haykin, S., Chen, Z. (2005). The cocktail party problem. *Neural Computation*, 17, 1875-  
16 1902.  
17
- 18 Jasper, H.H. (1958). The ten-twenty electrode system of the International Federation.  
19 *Electroencephalography and Clinical Neurophysiology*, 10, 370-375.  
20
- 21 Johnson, B.W., Hautus, M., Clapp, W.C. (2003). Neural activity associated with binaural  
22 processes for the perceptual segregation of pitch. *Clinical Neurophysiology*, 114,  
23 2245-2250.  
24
- 25 Johnson, B. W., Hautus, M. J., Duff, D. J., Clapp, W. C. (2007). Sequential processing of  
26 interaural timing differences for sound source segregation and spatial localization:  
27 evidence from event-related cortical potentials. *Psychophysiology*, 44, 541-551  
28
- 29 Lipp, R., Kitterick, P., Summerfield, Q., Bailey, P.J., Paul-Jordanov, I. (2010). Concurrent  
30 sound segregation based on inharmonicity and onset asynchrony.  
31 *Neuropsychologia*, 48, 1417-25.  
32
- 33 McDonald, K.L., Alain, C. (2005). Contribution of harmonicity and location to auditory  
34 object formation in free field: Evidence from event-related brain potentials. *Journal*  
35 *of the Acoustical Society of America*, 118, 1593-1604.  
36
- 37 Sanders, L.D., Joh, A.S., Keen, R.E., Freyman, RL. (2008). One sound or two? Object-  
38 related negativity indexes echo perception. *Perceptual Psychophysiology*, 70, 1558-  
39 1570.  
40
- 41 Sanders, L.D., Zobel, B.H., Freyman, R.L., Keen, R. (2008). Manipulations of listeners'  
42 echo perception are reflected in event-related potentials. *Journal of the Acoustical*  
43 *Society of America*, 129, 301-309.  
44
- 45 Snyder, J.S., Alain, C. (2005). Age-related changes in neural activity associated with  
46 concurrent vowel segregation. *Cognitive Brain Research*, 24, 492-499.  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

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

3  
4 Weise, A., Schröger, E., Bendixen, A. (2012). The processing of concurrent sounds based  
5 on inharmonicity and asynchronous onsets: An object-related negativity (ORN)  
6 study. *Brain Research*, 1439, 73-81.  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

**Table 1***Summary of experimental manipulations*

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

**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)*

	2 <sup>nd</sup> partial mistuned	2 <sup>nd</sup> and 4 <sup>th</sup> partials mistuned	2 <sup>nd</sup> partial delayed	2 <sup>nd</sup> partial mistuned and delayed	2 <sup>nd</sup> and 4 <sup>th</sup> mistuned and delayed	2 <sup>nd</sup> partial with location difference	2 <sup>nd</sup> partial mistuned with location difference	2 <sup>nd</sup> and 4 <sup>th</sup> partials mistuned with location difference	2 <sup>nd</sup> partial delayed with location difference	2 <sup>nd</sup> partial mistuned, delayed and with location difference	2 <sup>nd</sup> and 4 <sup>th</sup> partials mistuned, delayed and with location difference
<b>N1</b>											
Mean amplitude at Cz ( $\mu$ V)	-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
<b>ORN</b>											
Mean amplitude at Cz ( $\mu$ V)	-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 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 ( $\mu$ V)	<i>t</i> (19)	<i>p</i>
2 <sup>nd</sup> partial mistuned + 2 <sup>nd</sup> partial delayed (condition 1+3)	-3.186		
2 <sup>nd</sup> partial mistuned and delayed (condition 4)	1.74979	-1.991	0.061
2 <sup>nd</sup> partial mistuned+ 2 <sup>nd</sup> partial with location difference (condition 1+6)	-1.61194		
2 <sup>nd</sup> partial mistuned and with location difference (condition 7)	-1.04872	-1.306	0.207
2 <sup>nd</sup> partial delayed + 2 <sup>nd</sup> partial with location difference (condition 3+6)	-2.63249		
2 <sup>nd</sup> partial delayed and with location difference (condition 9)	-1.61503	-1.59	0.128
2 <sup>nd</sup> partial mistuned + 2 <sup>nd</sup> partial delayed + 2 <sup>nd</sup> partial with location difference (condition 1+3+6)	-3.71521		
2 <sup>nd</sup> partial mistuned, delayed with location difference (condition 10)	-1.27789	-3.143	0.005**
2 <sup>nd</sup> partial mistuned and delayed + 2 <sup>nd</sup> partial with location difference (condition 4+6)	-2.27901		
2 <sup>nd</sup> partial mistuned, delayed and with location difference (condition 10)	-1.27789	-1.871	0.077
2 <sup>nd</sup> partial mistuned and with location difference + 2 <sup>nd</sup> partial delayed (condition 7+3)	-3.152		
2 <sup>nd</sup> partial mistuned, delayed and with location difference (condition 10)	-1.27789	-2.431	0.025*
2 <sup>nd</sup> partial delayed and with location difference + 2 <sup>nd</sup> partial mistuned (condition 9+1)	-2.69775		
2 <sup>nd</sup> 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 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 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)*

	2 <sup>nd</sup> partial mistuned	2 <sup>nd</sup> and 4 <sup>th</sup> partials mistuned	2 <sup>nd</sup> partial delayed	2 <sup>nd</sup> partial mistuned and delayed	2 <sup>nd</sup> and 4 <sup>th</sup> mistuned and delayed	2 <sup>nd</sup> partial with location difference	2 <sup>nd</sup> partial mistuned with location difference	2 <sup>nd</sup> and 4 <sup>th</sup> partials mistuned with location difference	2 <sup>nd</sup> partial delayed with location difference	2 <sup>nd</sup> partial mistuned, delayed and with location difference	2 <sup>nd</sup> and 4 <sup>th</sup> partials mistuned, delayed and with location difference
<b>N1</b>											
Mean amplitude at Cz ( $\mu$ V)	-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
<b>ORN</b>											
Mean amplitude at Cz ( $\mu$ V)	-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
<b>P400</b>											
Mean amplitude at Cz ( $\mu$ V)	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 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 ( $\mu$ V)	<i>t</i> (19)	<i>p</i>
2 <sup>nd</sup> partial mistuned + 2 <sup>nd</sup> partial delayed (condition 1+3)	-3.683962		
2 <sup>nd</sup> partial mistuned and delayed (condition 4)	-2.284653	-2.474	0.022*
2 <sup>nd</sup> partial mistuned+ 2 <sup>nd</sup> partial with location difference (condition 1+6)	-2.165166		
2 <sup>nd</sup> partial mistuned and with location difference (condition 7)	-1.881584	-0.445	0.661
2 <sup>nd</sup> partial delayed + 2 <sup>nd</sup> partial with location difference (condition 3+6)	-3.49219		
2 <sup>nd</sup> partial delayed and with location difference (condition 9)	-2.092038	-2.359	0.028*
2 <sup>nd</sup> partial mistuned + 2 <sup>nd</sup> partial delayed + 2 <sup>nd</sup> partial with location difference (condition 1+3+6)	-4.670659		
2 <sup>nd</sup> partial mistuned, delayed with location difference (condition 10)	-2.082845	-3.073	0.006**
2 <sup>nd</sup> partial mistuned and delayed + 2 <sup>nd</sup> partial with location difference (condition 4+6)	-3.271350		
2 <sup>nd</sup> partial mistuned, delayed and with location difference (condition 10)	-2.082845	-1.856	0.077
2 <sup>nd</sup> partial mistuned and with location difference + 2 <sup>nd</sup> partial delayed (condition 7+3)	-3.270507		
2 <sup>nd</sup> partial mistuned, delayed and with location difference (condition 10)	-2.082845	-1.408	0.173
2 <sup>nd</sup> partial delayed and with location difference + 2 <sup>nd</sup> partial mistuned (condition 9+1)	-4.387077		
2 <sup>nd</sup> partial mistuned, delayed and with location difference (condition 10)	-2.082845	-3.414	0.002**

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

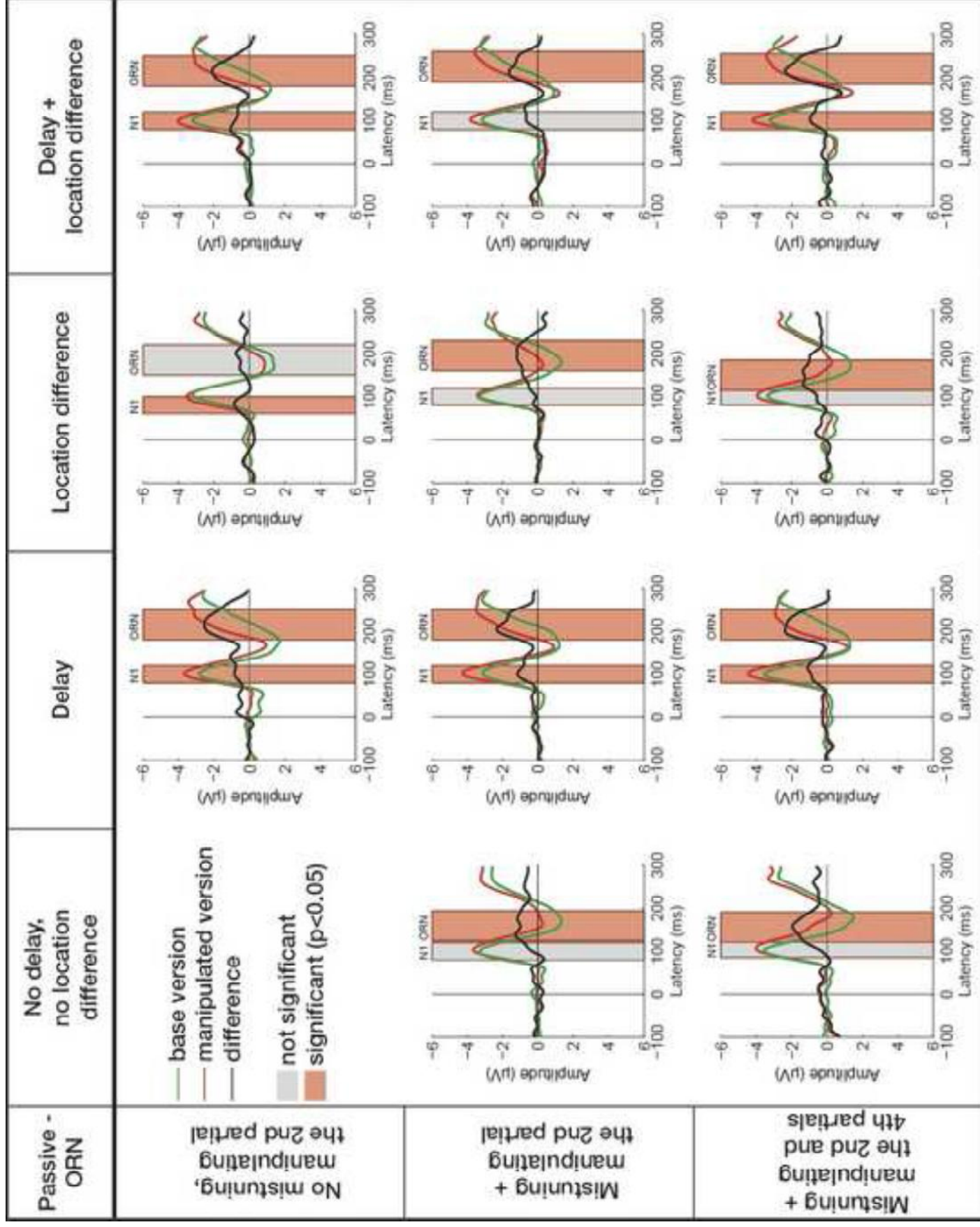
**Table 7**

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 ( $\mu$ V)	<i>t</i> (22)	<i>p</i>
2 <sup>nd</sup> partial mistuned + 2 <sup>nd</sup> partial delayed (condition 1+3)	3.673826		
2 <sup>nd</sup> partial mistuned and delayed (condition 4)	3.884773	-0.223	0.825
2 <sup>nd</sup> partial mistuned+ 2 <sup>nd</sup> partial with location difference (condition 1+6)	1.203539		
2 <sup>nd</sup> partial mistuned and with location difference (condition 7)	1.944099	-1.147	0.264
2 <sup>nd</sup> partial delayed + 2 <sup>nd</sup> partial with location difference (condition 3+6)	1.923776		
2 <sup>nd</sup> partial delayed and with location difference (condition 9)	3.428958	-2.398	0.025*
2 <sup>nd</sup> partial mistuned + 2 <sup>nd</sup> partial delayed + 2 <sup>nd</sup> partial with location difference (condition 1+3+6)	3.40057		
2 <sup>nd</sup> partial mistuned, delayed with location difference (condition 10)	3.122613	0.338	0.739
2 <sup>nd</sup> partial mistuned and delayed + 2 <sup>nd</sup> partial with location difference (condition 4+6)	3.611517		
2 <sup>nd</sup> partial mistuned, delayed and with location difference (condition 10)	3.122613	0.781	0.443
2 <sup>nd</sup> partial mistuned and with location difference + 2 <sup>nd</sup> partial delayed (condition 7+3)	4.14113		
2 <sup>nd</sup> partial mistuned, delayed and with location difference (condition 10)	3.122613	1.39	0.178
2 <sup>nd</sup> partial delayed and with location difference + 2 <sup>nd</sup> partial mistuned (condition 9+1)	4.905753		
2 <sup>nd</sup> partial mistuned, delayed and with location difference (condition 10)	3.122613	2.147	0.043*

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

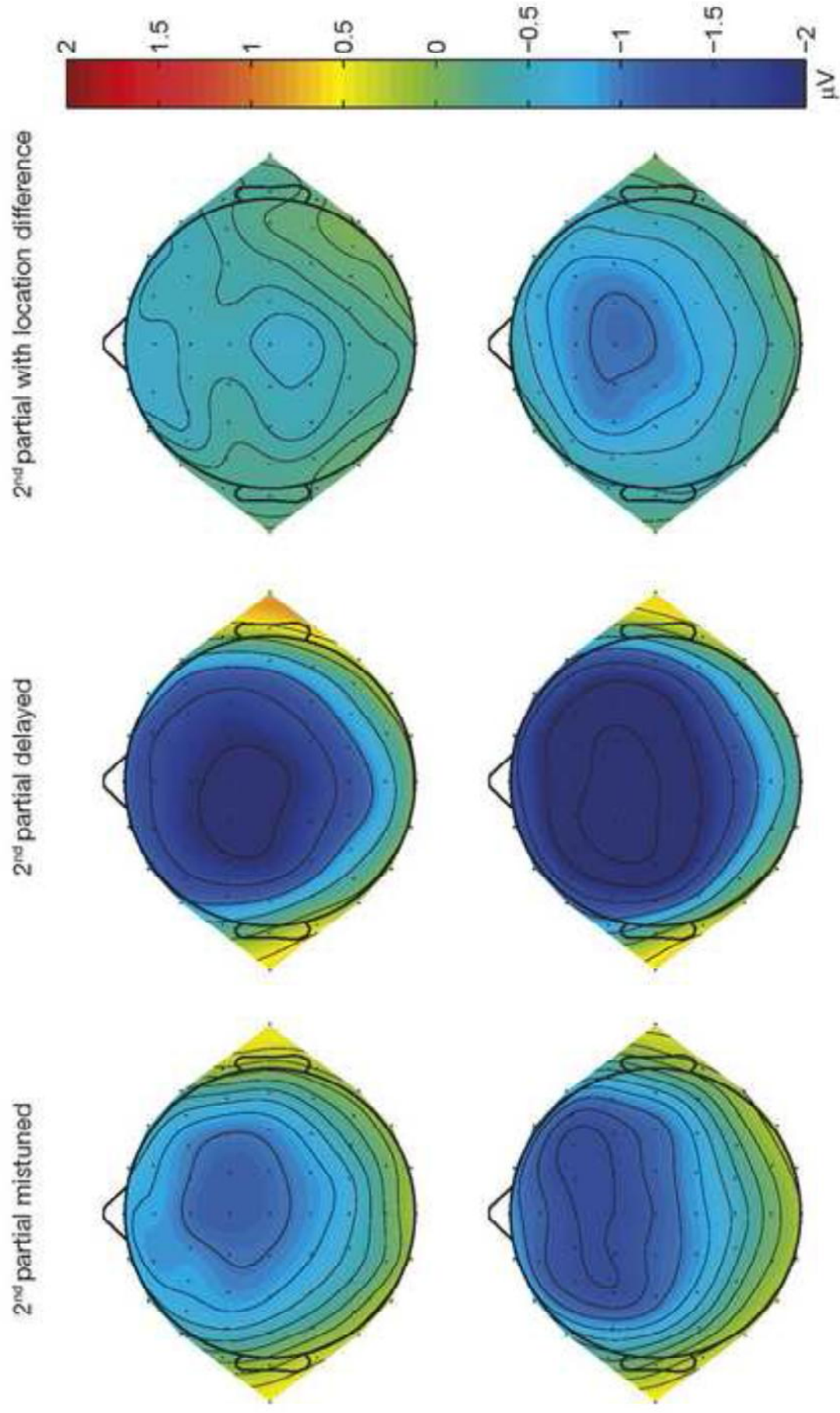
Figure(s)  
[Click here to download high resolution image](#)



## 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](#)

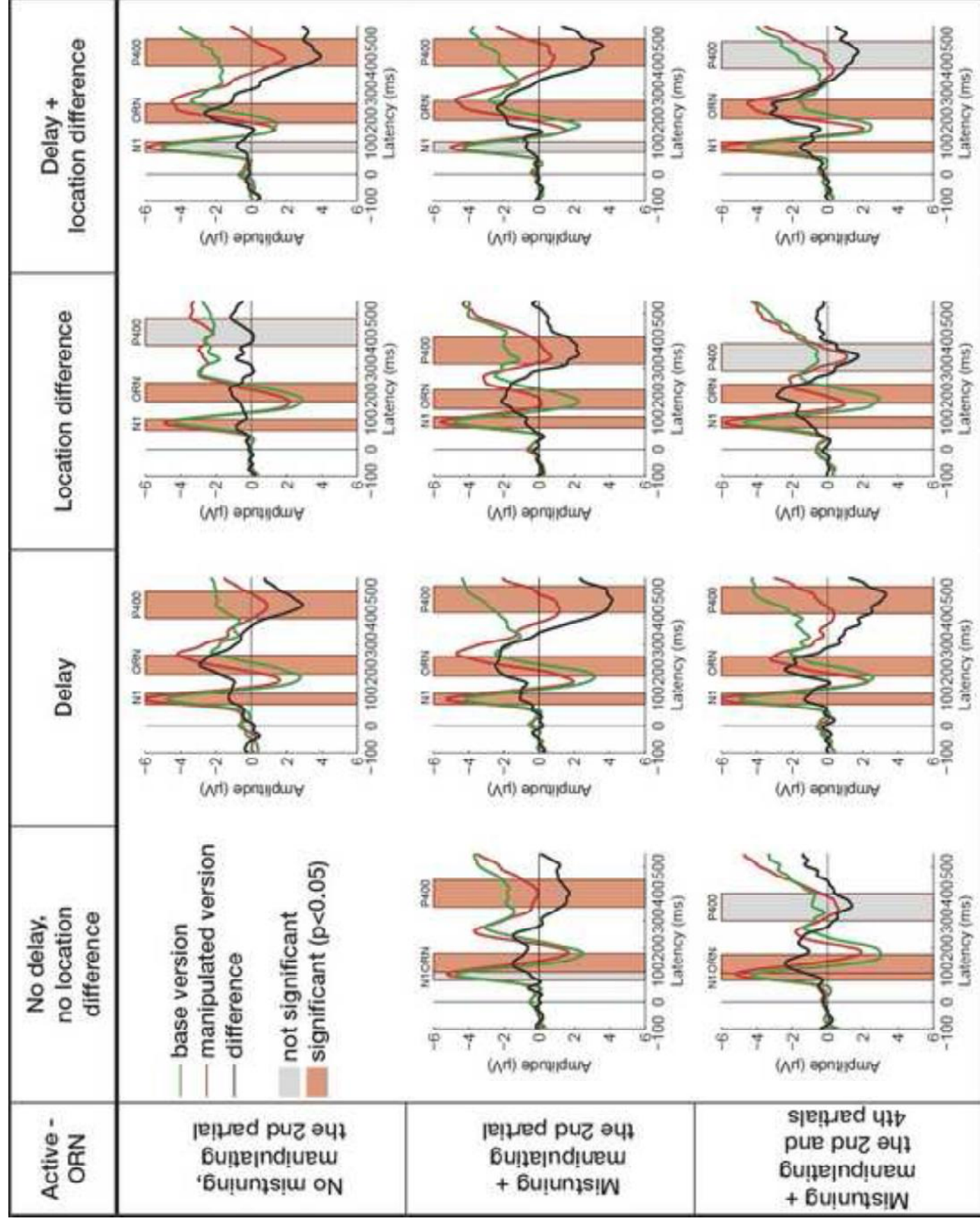




## Figure(s)

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.

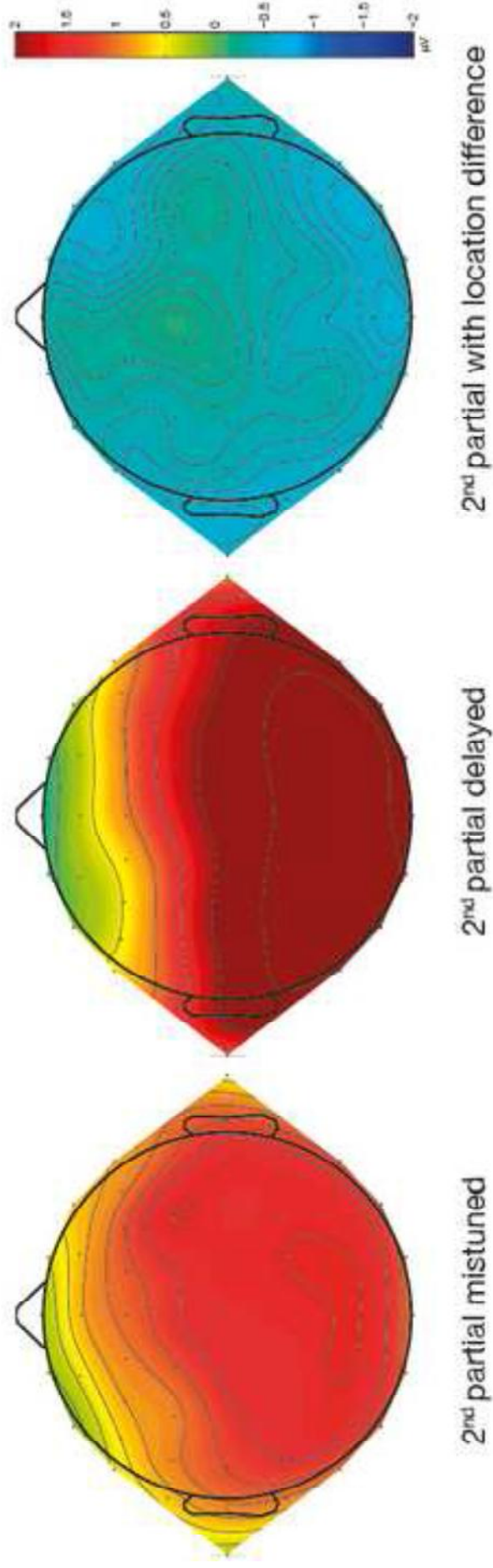
Figure(s)  
[Click here to download high resolution image](#)



## 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.

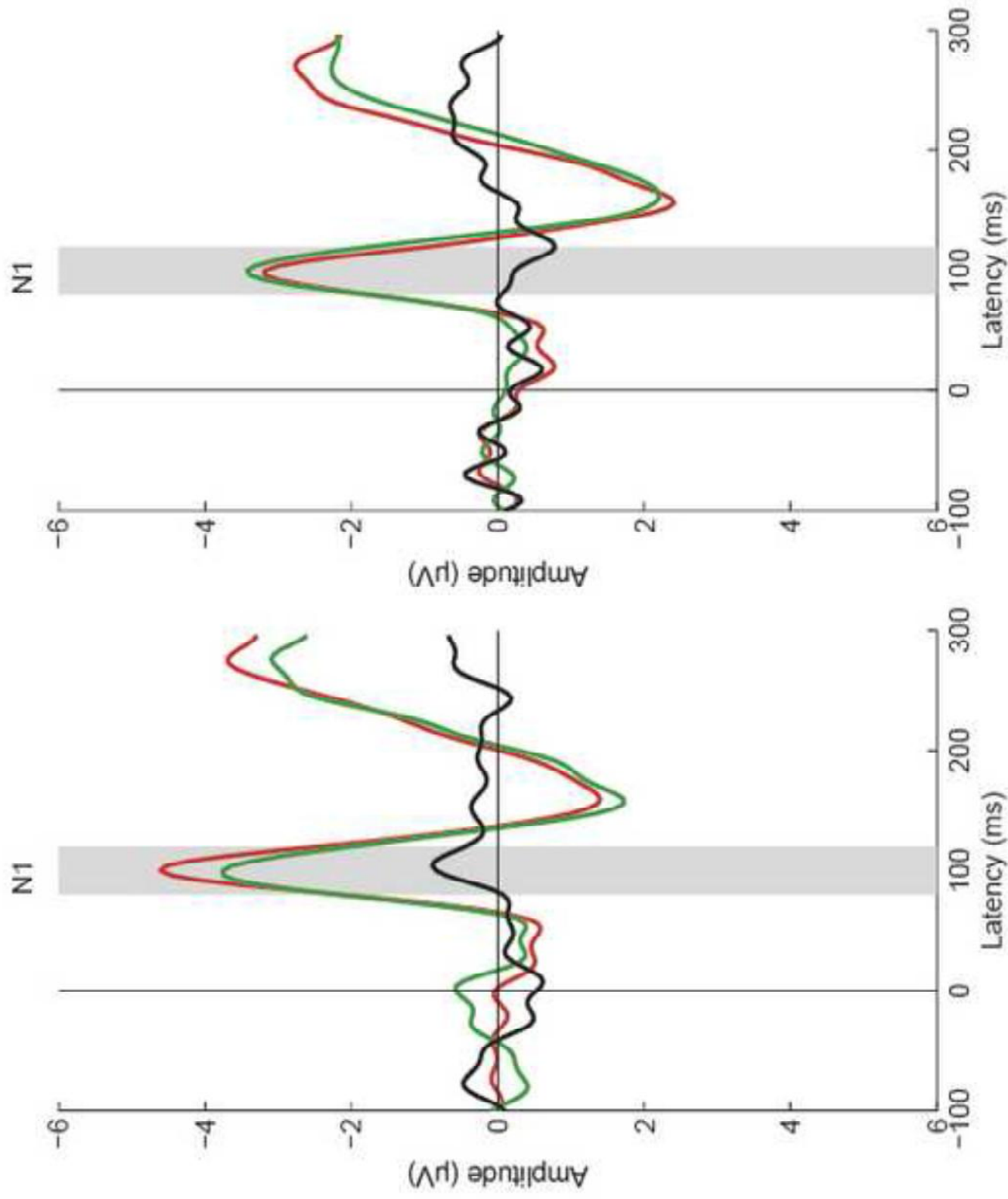
Figure(s)  
[Click here to download high resolution image](#)



## 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:

- 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.