

# Neural activity associated with distinguishing concurrent auditory objects

Claude Alain,<sup>a)</sup> Benjamin M. Schuler, and Kelly L. McDonald

*Rotman Research Institute, Baycrest Centre for Geriatric Care, 3560 Bathurst Street, Toronto, Ontario M6A 2E1, Canada and Department of Psychology, University of Toronto, Sidney Smith Hall, 100 St. George Street, Toronto, Ontario M5S 1A1, Canada*

(Received 25 September 2001; accepted for publication 20 November 2001)

The neural processes underlying concurrent sound segregation were examined by using event-related brain potentials. Participants were presented with complex sounds comprised of multiple harmonics, one of which could be mistuned so that it was no longer an integer multiple of the fundamental. In separate blocks of trials, short-, middle-, and long-duration sounds were presented and participants indicated whether they heard one sound (i.e., buzz) or two sounds (i.e., buzz plus another sound with a pure-tone quality). The auditory stimuli were also presented while participants watched a silent movie in order to evaluate the extent to which the mistuned harmonic could be automatically detected. The perception of the mistuned harmonic as a separate sound was associated with a biphasic negative–positive potential that peaked at about 150 and 350 ms after sound onset, respectively. Long duration sounds also elicited a sustained potential that was greater in amplitude when the mistuned harmonic was perceptually segregated from the complex sound. The early negative wave, referred to as the object-related negativity (ORN), was present during both active and passive listening, whereas the positive wave and the mistuning-related changes in sustained potentials were present only when participants attended to the stimuli. These results are consistent with a two-stage model of auditory scene analysis in which the acoustic wave is automatically decomposed into perceptual groups that can be identified by higher executive functions. The ORN and the positive waves were little affected by sound duration, indicating that concurrent sound segregation depends on transient neural responses elicited by the discrepancy between the mistuned harmonic and the harmonic frequency expected based on the fundamental frequency of the incoming stimulus. © 2002 Acoustical Society of America.

[DOI: 10.1121/1.1434942]

PACS numbers: 43.64.Qh, 43.64.Ri, 43.66.Lj [LHC]

## I. INTRODUCTION

In most everyday situations, there is often more than one audible sound source at any given moment. Given that the acoustic components from simultaneously active sources impinge upon the ear at the same time, how does the auditory system sort which elements of the mixture belong to a particular source and which originate from a different sound source?

Psychophysical research has identified several factors that can help listeners to segregate co-occurring events. For example, sound components that are harmonically related or that rise and fall in intensity together usually arise from a single physical source and tend to be grouped into one perceptual object. Conversely, sounds are more likely to be assigned to separate objects (i.e., sources) if they are not harmonically related and if they differ widely in frequency and intensity (for a review, see Bregman, 1990; Hartmann, 1988, 1996). The present study focuses on concurrent sound segregation based on harmonicity.

One way of investigating concurrent sound segregation based on harmonicity is by means of the mistuned harmonic experiment. Usually, the listener is presented with two stimuli successively, one of them with perfectly harmonic components, the other with a mistuned harmonic. The task of

the listener is to indicate which one of the two stimuli contains the mistuned harmonic. Several factors influence the perception of the mistuned harmonic as a separate tone, including degree of inharmonicity, harmonic number, and sound duration (Hartmann, McAdams, and Smith, 1990; Lin and Hartmann, 1998; Moore, Peters, and Glasberg, 1985).

This effect of mistuning on concurrent sound segregation is consistent with Bregman's account of auditory scene analysis (Bregman, 1990). Within this model, the acoustic wave is first decomposed into perceptual groups (i.e., objects) according to Gestalt principles. Partials that are harmonically related are grouped together into one entity, while the partial that is sufficiently mistuned stands out as a separate object. It has been proposed that the perception of the mistuned harmonic as a separate object depends on a pattern-matching process that attempts to adjust a harmonic template, defined by a fundamental frequency, to fit the spectral pattern (Goldstein, 1978; Hartmann, 1996; Lin and Hartmann, 1998). When a harmonic is mistuned by a sufficient amount, a discrepancy occurs between the perceived frequency and that expected on the basis of the template. The purpose of this pattern-matching process could be to signal to higher auditory centers that more than one auditory object might be simultaneously present in the environment.

One important question concerns the nature of the mismatch process that may underlie concurrent sound segrega-

<sup>a)</sup>Electronic mail: calain@rotman-baycrest.on.ca

tion. For instance, it is unclear whether the mismatch process is transient in nature or whether it remains present for the whole duration of the stimulus. Previous behavioral studies have shown that perception of the mistuned harmonic as a separate tone improved with increasing sound durations (e.g., Moore *et al.*, 1986). This suggests that perception of concurrent auditory objects may depend on a continuous analysis of the stimulus rather than on a transient detection of inharmonicity.

Event-related brain potentials (ERPs) provide a powerful tool for exploring the neural mechanisms underlying concurrent sound segregation. In a series of experiments, Alain, Arnott, and Picton (2001) measured ERPs to complex sounds that either had all harmonics in tune or included one mistuned harmonic so that it was no longer an integer multiple of the fundamental. When individuals reported perceiving two concurrent auditory objects (i.e., a buzz plus another sound with a pure-tone quality), a phasic negative deflection was observed in the ERP. This negative wave peaked around 180 ms after sound onset and was referred to as the object-related negativity (ORN) because its amplitude correlated with perceptual judgment, being greater when participants reported hearing two distinct perceptual objects. The ORN was also present even when participants were asked to ignore the stimuli and read a book of their choice. This suggests that this component indexes a relatively automatic process that occurs even when auditory stimuli are not task relevant.

Distinguishing concurrent auditory objects was also associated with a late positive wave that peaked at about 400 ms following stimulus onset (P400). Like the ORN, the P400 amplitude correlated with perceptual judgment, being larger when participants perceived the mistuned harmonic as a separate tone. However, in contrast with the ORN, this component was present only when participants were required to respond whether they heard one or two auditory stimuli.

The aim of the present study was to further investigate the nature of the neural processes underlying concurrent sound segregation using sounds of various durations. In Alain *et al.*'s study, it was unclear whether the ORN and P400 indexed a transient or a sustained process because the sound duration was always kept constant. Examining the ORN and P400 for sounds of various duration can give clues about the processes involved in concurrent sound segregation. If concurrent sound segregation depends on a transient process that detects a mismatch between the mistuned harmonic and the harmonic template, then these ERP components should be little affected by sound duration. However, if concurrent sound segregation depends on the ongoing analysis of the stimulus, then the effect of mistuning on ERPs should vary with sound duration. Because the stimuli in Alain *et al.*'s study were always 400 ms in duration, it was also difficult to determine the contributions of the offset responses and the response selection processes to the P400 component. In the present study, participants were presented with sounds of various durations and were asked to respond at the end of the sound presentation to reduce contamination by response processes. If the P400 component received contribution from the offset responses and/or from the response

processes, then the P400 amplitude should vary as a function of stimulus duration.

## II. METHOD

### A. Participants

Thirteen adults provided written informed consent to participate in the study. The data of three participants were excluded from further analysis because they showed extensive ocular contaminations or had extreme difficulty in distinguishing the different stimuli. Four women and six men form the final sample (aged between 22 and 37 years, mean age =  $25.7 \pm 4.67$  years). All participants were right-handed and had pure-tone thresholds within normal limits for frequencies ranging from 250 to 8000 Hz (both ears).

### B. Stimuli and task

All stimuli had a fundamental frequency of 200 Hz. The tuned stimuli consisted of a complex sound obtained by combining 12 pure tones with equal intensity. In the mistuned stimuli the third harmonic was shifted either up- or downwards by 16% of its original value (696 or 504 Hz instead of 600 Hz). The intensity level of each sound was 80 dB SPL. The durations of the sounds were short (100 ms), medium (400 ms), or long (1000 ms), including 5-ms rise/fall time. The sounds were generated digitally with a sampling rate of 50 kHz and presented binaurally through Sennheiser HD 265 headphones. Participants were presented with 18 blocks of trials. Each block consisted of 130 stimuli of short, medium, or long duration sounds. Half of the stimuli in each block were tuned while the other half were mistuned. Tuned and mistuned stimuli were presented in a random order. The short, medium, and long duration blocks were presented in a random order across participants.

Each participant took part in active and passive listening conditions (nine blocks of trials in each condition). In the active listening condition, participants indicated whether they perceived one tuned sound or two sounds (i.e., a buzz plus another sound with a pure-tone quality) by pressing one of two buttons on a response box using the right index and middle fingers. Participants were asked to withhold their response until the end of the sound to reduce motor-related potentials during sound presentation. The intertrial interval, i.e., the interval between the participant's response and the next trial, was 1000 ms. No feedback was provided after each response. In the passive condition, participants watched a silent movie with subtitles and were asked to ignore the auditory stimuli. In the passive listening condition, the interstimulus interval varied randomly between 800 and 1000 ms. The order of the active and passive conditions was counter-balanced across participants.

### C. Electrophysiological recording and analysis

The electroencephalogram (EEG) was digitized continuously (bandpass 0.05–50 Hz; 250-Hz sampling rate) from an array of 64 electrodes using NeuroScan SynAmps and stored for offline analysis. Eye movements were monitored with electrodes placed at the outer canthi and at the superior and

inferior orbit. During the recording, all electrodes were referenced to the midline central electrode (i.e., Cz); for data analysis they were re-referenced to an average reference and the electrode Cz was reinstated.

The analysis epoch included 200 ms of prestimulus activity and 800, 1000, or 1600 ms of poststimulus activity for the short, medium, and long duration sounds, respectively. Trials contaminated by excessive peak-to-peak deflection ( $\pm 200 \mu\text{V}$ ) at the channels not adjacent to the eyes were automatically rejected before averaging. ERPs were then averaged separately for each site, stimulus duration, stimulus type, and listening condition. ERPs were digitally low-pass filtered to attenuate frequencies above 15 Hz. For each individual average, the ocular artifacts (e.g., blinks, saccades, and lateral movements) were corrected by means of ocular source components using the Brain Electrical Source Analysis (BESA) software (Picton *et al.*, 2000).

The ERP waveforms were quantified by computing mean values in selected latency regions, relative to the mean amplitude of the 200-ms prestimulus activity. The intervals chosen for the ORN and P400 mean amplitude were 100–200 ms and 300–400 ms, respectively. To ease the comparison between active and passive listening, the ERPs for correct and incorrect trials in the active listening condition were lumped together. Trials with an early response (i.e., response during sound presentation) were excluded from the analysis.

The effects of sound duration on perceptual judgment were subjected to a repeated measures within-subject analysis of variance (ANOVA) with sound duration and stimulus type as factors. Accuracy was defined as hits minus false alarms. For the ERP data, the independent variables were participants' listening condition (active versus passive), sound duration (short, medium, long), stimulus type (tuned versus mistuned), and electrode (Fz, F1, F2, FCz, FC1, FC2, Cz, C1, and C2). Scalp topographies using the 61 electrodes (omitting the periocular electrodes) were statistically analyzed after scaling the amplitudes to eliminate amplitude differences between stimuli and conditions. For each participant and each condition, the mean voltage measurements were normalized by subtracting the minimum value from each data point and dividing by the difference between the maximum and minimum value from the electrode set (McCarthy and Wood, 1985). Whenever appropriate, the degrees of freedom were adjusted with the Greenhouse–Geisser epsilon. All reported probability estimates are based on these reduced degrees of freedom.

### III. RESULTS

#### A. Behavioral data

Overall, participants were more likely to report hearing two concurrent stimuli when the complex sound included a mistuned harmonic. Conversely, they were more likely to report perceiving one complex sound when the sound components were all harmonically related. The main effects of stimulus type and sound duration on perceptual judgment were not significant. However, there was a significant interaction between sound duration and stimulus type,  $F(2,18) = 5.75$ ,  $p < 0.02$  (Fig. 1). Analyses of simple main effects

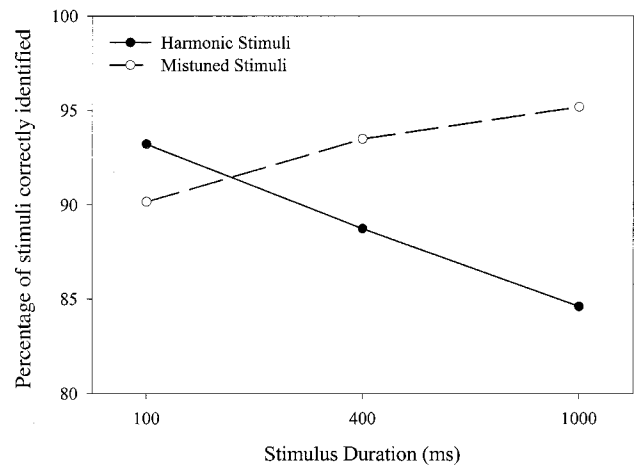


FIG. 1. Probability of reporting hearing one sound or two sounds as a function of stimulus duration.

revealed that participants were significantly less likely to report hearing one complex sound when the tuned stimuli increased in duration,  $F(2,36) = 5.63$ ,  $p < 0.01$ . In comparison, the perception of the mistuned harmonic as a separate tone was little affected by increasing sound duration,  $F(2,36) = 1.97$ .

#### B. Electrophysiological data

Figure 2 shows the group mean ERPs elicited by tuned and mistuned stimuli as a function of sound duration during passive and active listening. In both listening conditions, tuned and mistuned stimuli elicited a clear N1–P2 complex. At the midline frontocentral site (i.e., FCz), the N1 and P2 deflections peaked at about 125 and 195 ms after sound onset, respectively. Middle and long duration sounds generated a sustained potential and a small offset response. The N1 amplitude was larger during active than passive listening,  $F(1,9) = 21.48$ ,  $p < 0.001$ . The effect of sound duration on the N1 amplitude was not significant nor was the interaction

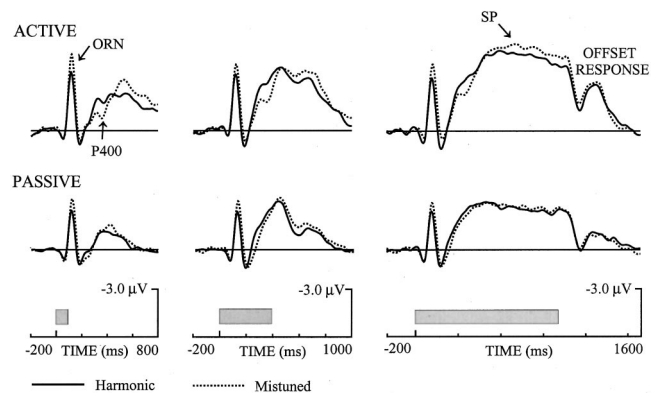


FIG. 2. Group mean event-related brain potentials (ERPs) from the midline frontocentral site (FCz) as a function of sound duration and harmonicity. Top: ERPs recorded when individuals were required to decide whether one sound or two sounds were present (active listening). Bottom: ERPs recorded when individuals were asked to watch a movie and to ignore the auditory stimuli (passive listening). The gray rectangle indicates the duration of the stimulus.

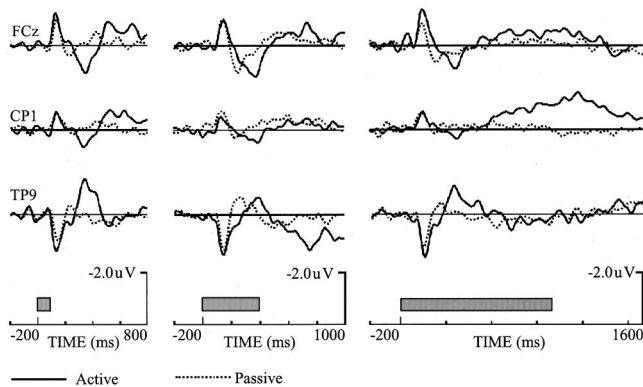


FIG. 3. Group mean difference waves between ERPs elicited by harmonic and inharmonic stimuli during passive and active listening at the midline frontocentral site (FCz), the left central parietal site (CP1), and the left inferior and posterior temporal site (TP9). The tick marks indicate 200 ms for the short and middle duration sounds, and 300 ms for the long duration sound.

between sound duration and listening condition. The P2 wave amplitude and latency were not significantly affected by the listening condition or sound duration.

The ERPs to mistuned stimuli showed a negative displacement compared to those elicited by tuned stimuli. The effects of mistuning on ERPs can best be illustrated by subtracting ERPs to tuned stimuli from ERPs elicited by mistuned stimuli (Fig. 3). In the active listening condition, the difference waves revealed a biphasic negative–positive potential that peaked at about 160 and 360 ms poststimulus. The negative wave, referred to as the “object-related negativity” (ORN), was maximum at frontocentral sites and inverted in polarity at inferior temporal sites. ANOVA with stimulus type, listening condition, stimulus duration, and electrode as factors yielded a main effect of stimulus type,  $F(1,9) = 20.54$ ,  $p < 0.001$ , and a main effect of listening condition,  $F(1,9) = 16.48$ ,  $p < 0.01$ . The interaction between listening condition and stimulus type was not significant,  $F(1,9) = 3.69$ ,  $p = 0.09$ . A separate ANOVA on ERP data recorded during passive listening yielded a main effect of stimulus type,  $F(1,9) = 17.16$ ,  $p < 0.01$ . This indicates that a significant ORN was present during passive listening. In both listening conditions, the ORN amplitude and latency was little affected by sound duration.

In the active listening condition, the ORN was followed by a positive wave peaking at 350 ms poststimulus referred to as the P400. Like the ORN, the P400 was biggest over frontocentral sites and was inverted in polarity at occipital and temporal sites (see Figs. 3 and 4). Complex sounds with the mistuned harmonic generated greater positivity than tuned stimuli,  $F(1,9) = 7.90$ ,  $p < 0.05$ . The interaction between stimulus type and listening condition was significant,  $F(1,9) = 7.32$ ,  $p < 0.05$ , reflecting greater P400 amplitude during active than passive listening. A separate ANOVA on the ERPs recorded during passive listening yielded no main effect of stimulus type,  $F(1,9) = 0.28$ . Like the ORN, there was no significant interaction between sound duration and stimulus type,  $F(2,18) = 1.69$ ,  $p = 0.214$ , indicating that P400 amplitude was not significantly affected by the duration of the mistuned stimulus.

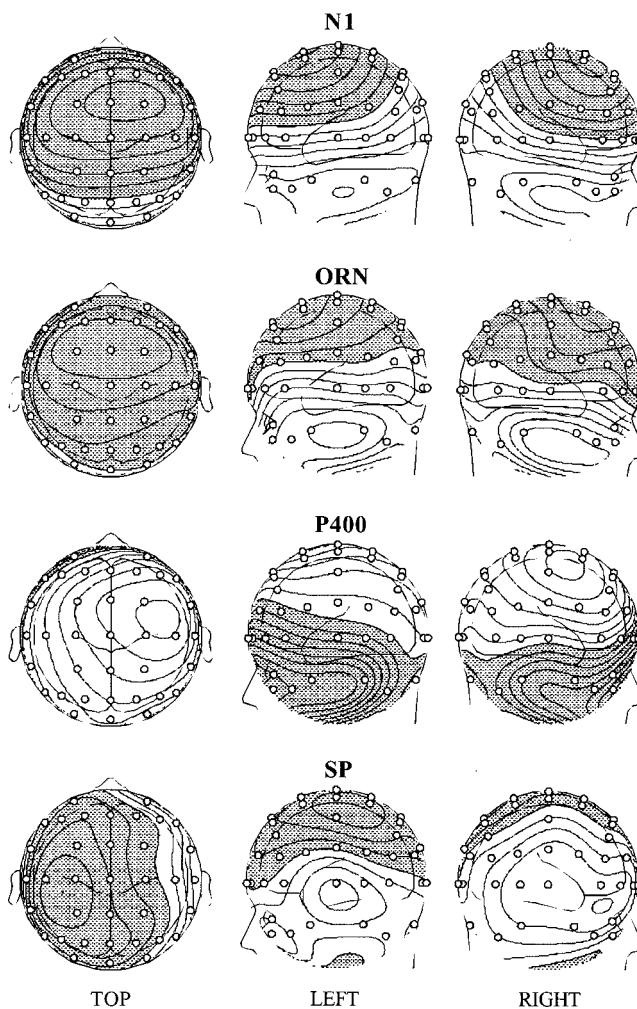


FIG. 4. Contour maps for the N1 (120 ms), ORN (160 ms), P400 (360 ms), and sustained potential (800 ms). The N1, ORN, and P400 topographies represent the peak amplitude measurement for the short duration signal (i.e., 100 ms). The sustained potential (SP) topography represents the amplitude measurement for the long duration signal (i.e., 1000 ms). Shade indicates negativity, whereas light indicates positivity. For the N1 wave the contour spacing was set at  $0.6 \mu\text{V}$ . For the ORN, P400, and sustained potential the contour spacing was set at  $0.2 \mu\text{V}$ . The negative polarity is illustrated by the shaded area. The open circle indicates electrode position.

A visual inspection of the data revealed a positive wave that peaked at 245 ms following sound onset that was present during the passive listening. This positive wave peaked earlier than the P400 and was more frontally distributed than the P400. The positive wave recorded during passive listening was affected by sound duration,  $F(2,18) = 8.99$ ,  $p < 0.01$ , being larger for middle than the short or the long duration sounds ( $p < 0.05$ , in both cases).

### 1. Sustained potentials

Long duration stimuli elicited a large and widespread sustained potential that was maximum at frontocentral sites. To take into account the widespread nature of the sustained response, the effects of mistuning and listening condition on the sustained potentials were quantified using a larger array of electrodes (i.e., F1, F2, F3, F4, F5, F6, FC1, FC2, FC5, FC6, C1, C2, C3, C4, C5, C6). ANOVA for the 600–1200-ms interval following sound onset yielded a main effect

of listening condition,  $F(1,9)=12.12$ ,  $p<0.01$ , reflecting greater amplitude during active than passive listening (Fig. 3). The main effect of mistuning was not significant nor was the interaction between listening condition and mistuning. However, there was a significant interaction between mistuning and hemisphere,  $F(1,9)=6.34$ ,  $p<0.05$ , and a three-way interaction including listening condition, mistuning, and hemisphere,  $F(1,9)=9.64$ ,  $p<0.02$ . Therefore, the effect of mistuning on the sustained potential was examined separately for the left and right hemispheres. The effect of mistuning on the sustained potential was significant only over the left hemisphere,  $F(1,9)=7.25$ ;  $p<0.05$  (Fig. 4). The interaction between listening condition and mistuning was not significant for the selected electrodes. However, it was highly significant for central electrodes near the midline [e.g., C1 and C3,  $F(1,9)=10.77$ ,  $p<0.01$ ].

## 2. Scalp distribution

Scalp distributions are an important criterion in identifying and distinguishing between ERP components. The assumption is that different scalp distributions indicate different spatial configurations of intracranial current sources. In the present study, we analyzed scalp distributions to examine whether the observed ERP component generation (i.e., N1, ORN, P400, sustained potentials) depends on distinct neural networks.

Figure 4 shows the amplitude distribution for the N1, ORN, P400, and mistuning-related changes in the sustained potential. The N1 was largest at frontocentral sites and inverted polarity at inferior temporal sites. The ORN amplitude distribution was not significantly different from that of the N1 wave. There was no significant difference in N1 and ORN amplitude distribution elicited by short, medium, and long duration sounds. In comparison with the N1 and the ORN, the P400 response was more lateralized over the right central areas. This difference in topography was present for short, medium, and long duration sounds,  $F(60,540)>9.50$ ,  $p<0.001$ , in all cases. The N1, ORN, and P400 scalp distributions were not significantly affected by sound durations. Last, the mistuning-related change in sustained potential was greater over the left central parietal area than the N1, ORN, and P400 responses,  $F(60,540)>5.00$ ,  $p<0.01$  in all cases.

## IV. DISCUSSION

Participants were more likely to report hearing two distinct stimuli when the complex sound contained a mistuned harmonic. This is consistent with previous research (e.g., Alain *et al.*, 2001; Hartmann *et al.*, 1990; Moore, Glasberg, and Peters, 1986), and shows that frequency periodicity provides an important cue in parsing co-occurring auditory objects.

The ability to perceive the mistuned harmonic as a separate tone was little affected by increasing sound duration. Given that the amount of mistuning was well above threshold, it is not surprising that sound duration had little impact on perceiving the mistuned harmonic as a separate tone. More surprising was the finding that for tuned stimuli participants were more likely to report hearing two auditory

objects when the complex sound was long rather than short. Because the third harmonic was the only harmonic that was mistuned in the present study, participants may have realized that the only changing component was always in the same frequency region and therefore listened more carefully for sounds at that particular frequency. It has been shown that individuals are able to identify a single harmonic in a complex sound if they have previously listened to that harmonic presented alone (for a review, see Bregman, 1990). A similar effect could have taken place in the present study. Participants could have heard the mistuned partial as a separate tone and this tone may have primed them to hear, in the tuned stimuli, the third harmonic which was the most similar in frequency with the mistuned harmonic. Hence, the relevant figure, which was identified by the attention processes, was not the whole Gestalt of the complex sound but the changing third harmonic over different trials.

Two ERP components were associated with the perception of the mistuned harmonic as a separate tone. The first one was the ORN, which was maximum at frontocentral sites and inverted in polarity at inferior parietal and occipital sites. This amplitude distribution is consistent with generators in auditory cortices along the Sylvian fissure. Like participants' perception of the mistuned harmonic as a separate tone, the ORN amplitude and latency were little affected by increasing sound duration. This suggests that concurrent sound segregation depends on a transient neural response triggered by the automatic detection of inharmonicity. As previously suggested by Alain *et al.*, the ORN may index an automatic mismatch detection process between the mistuned harmonic and the harmonic frequency expected based upon the harmonic template extrapolated from the incoming stimulus.

Mistuned stimuli generated a significant ORN even when participants were not actively attending to the stimuli. In addition, the ORN amplitude was similar in both active and passive listening conditions. These findings replicate those of Alain *et al.* (2001), and are consistent with the proposal that this component indexes a relatively automatic process. The results are also consistent with the proposal that the ORN indexes primarily bottom-up processes and that concurrent sound segregation may occur independently of listener's attention. However, the role of attention in detecting a mistuned harmonic will require further empirical research. In the present study, listeners' attention may have wandered to the auditory stimuli while they watched the subtitled movie, thereby contributing to the ORN recorded during passive listening.

The ORN presents some similarities in latency and amplitude distribution with another ERP component called the mismatch negativity, or MMN. The MMN is elicited by the occurrence of rare deviant sounds embedded in a sequence of homogeneous standard stimuli. Like the ORN, the MMN has a frontocentral distribution and its latency peaks at about 150 ms after the onset of deviation. Both ORN and MMN can be recorded while listeners are reading or watching a video and therefore are thought to index bottom-up processing of auditory scene analysis. A crucial difference between the two components is that while the MMN generation is highly sensitive to the perceptual context, the ORN generation is not.

That is, the MMN is elicited *only* by rare deviant stimuli, whereas the ORN is elicited by mistuned stimuli whether they are presented occasionally or frequently (Alain *et al.*, 2001). Thus, the MMN reflects a mismatch between the incoming auditory stimulus and what is expected based on the previously occurring stimuli, whereas the ORN indexes a discrepancy between the mistuned harmonic and the harmonic template that is presumably extrapolated from the incoming stimulus. As mentioned earlier, scalp distributions and dipole source modeling are important criteria in identifying and distinguishing between ERP components. Thus, further research comparing the scalp distributions of the ORN and MMN may provide evidence that these two ERP components index different processes and recruit distinct neural networks.

The second component associated with concurrent sound segregation was the P400, which was present only when participants were asked to make a response. The P400 has a more lateralized and widespread distribution than the N1 or the ORN and seems to be more related to perceptual decisions. Given that participants indicated their response after the sound was presented, the P400 generation cannot be easily accounted for by motor processes. The P400 may index the perception and recognition of the mistuned harmonic as a separate object, distinct from the complex sound. As with the ORN, the P400 amplitude was little affected by sound duration, although the P400 tended to be smaller for long than middle or short duration stimuli. This result suggests that for shorter and intermediate duration sounds, the P400 amplitude may be partly superimposed by the offset response elicited by the end of the stimulus.

Long duration sounds generated a sustained potential, which was larger during active than passive listening. This enhanced amplitude may reflect additional attentional resources dedicated to the analysis of the complex sounds. Within the active listening condition, the perception of the mistuned harmonic as a separate sound generated greater sustained potential amplitude than sounds that were perceived as a single object. This suggests that concurrent sound segregation can involve both transient and sustained neural events when individuals are required to pay attention to the auditory scene. The role of the transient neural event may be to signal to higher auditory centers that more than one sound source is present in the mixture. In comparison, the enhanced sustained potential for mistuned stimuli may reflect an ongoing analysis of both sound sources for an eventual response, context updating, or a second evaluation of the mistuned harmonic. Interestingly, the mistuning-related changes in the sustained potential were lateralized to the left hemisphere and could partly reflect motor-preparation processes because participants were required to indicate their response with their right hand. However, this cannot easily account for the differences between tuned and mistuned stimuli because both stimuli required a response from the right hand, unless the differences in sustained potentials between tuned and mistuned stimuli reflect the activation of different motor programs. It is also possible that the enhanced sustained potential to mistuned stimuli reflects enhanced processing allocated to the mistuned harmonic. Perhaps there is an ad-

ditional and ongoing analysis of the sound quality when one partial stands out from the complex as a separate object.

## V. CONCLUSION

In summary, the perception of concurrent auditory objects is associated with two neural events that peak, respectively, at about 160 and 360 ms poststimulus. The scalp distribution is consistent with generators in auditory cortices, reinforcing the role of primary and secondary auditory cortex in scene analysis. Although it cannot be excluded that concurrent sound segregation may have taken place at some stage along the auditory pathway before auditory cortices, the perception of the mistuned harmonic as a separate sound does involve primary and secondary auditory cortices.

The ORN was little affected by sound duration and was present even when participants were asked to ignore the stimuli. We propose that this component indexes a transient and automatic mismatch process between the harmonic template extrapolated from the incoming stimulus and the harmonic frequency expected based upon the fundamental of the complex sound. As with the ORN, the P400 was little affected by sound duration. However, the P400 is present only when individuals are required to discriminate between tuned and mistuned stimuli, suggesting that the P400 generation depends on controlled processes responsible for the identification of the stimuli and the generation of the appropriate response. Last, the perception of the mistuned harmonic generated larger sustained potentials than the perception of tuned stimuli. The effect of mistuning on the sustained potential was present only during active listening, suggesting that attention to complex auditory scenes recruits both transient and sustained processes but that scene analysis of sounds presented outside the focus of attention may depend primarily on transient neural events.

- Alain, C., Arnott, S. R., and Picton, T. W. (2001). "Bottom-up and top-down influences on auditory scene analysis: Evidence from event-related brain potentials," *J. Exp. Psychol. Hum. Percept. Perform.* **27**(5), 1072–1089.
- Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sounds* (The MIT Press, London).
- Goldstein, J. L. (1978). "Mechanisms of signal analysis and pattern perception in periodicity pitch," *Audiology* **17**(5), 421–445.
- Hartmann, W. M. (1988). "Pitch, perception and the segregation and integration of auditory entities," in *Auditory Function: Neurobiological Bases of Hearing*, edited by G. M. Edelman, W. E. Gall, and W. M. Cowan (Wiley, New York), pp. 623–645.
- Hartmann, W. M. (1996). "Pitch, periodicity, and auditory organization," *J. Acoust. Soc. Am.* **100**, 3491–3502.
- Hartmann, W. M., McAdams, S., and Smith, B. K. (1990). "Hearing a mistuned harmonic in an otherwise periodic complex tone," *J. Acoust. Soc. Am.* **88**, 1712–1724.
- Lin, J. Y., and Hartmann, W. M. (1998). "The pitch of a mistuned harmonic: Evidence for a template model," *J. Acoust. Soc. Am.* **103**, 2608–2617.
- McCarthy, G., and Wood, C. C. (1985). "Scalp distributions of event-related potentials: An ambiguity associated with analysis of variance models," *Electroencephalogr. Clin. Neurophysiol.* **62**, 203–208.
- Moore, B. C., Glasberg, B. R., and Peters, R. W. (1986). "Thresholds for hearing mistuned partials as separate tones in harmonic complexes," *J. Acoust. Soc. Am.* **80**, 479–483.
- Moore, B. C., Peters, R. W., and Glasberg, B. R. (1985). "Thresholds for the detection of inharmonicity in complex tones," *J. Acoust. Soc. Am.* **77**, 1861–1867.
- Picton, T. W., van Roon, P., Armiljo, M. L., Berg, P., Ille, N., and Scherg, M. (2000). "The correction of ocular artifacts: A topographic perspective," *Clin. Neurophysiol.* **111**(1), 53–65.