

Bottom-Up and Top-Down Influences on Auditory Scene Analysis: Evidence From Event-Related Brain Potentials

Claude Alain, Stephen R. Arnott, and Terence W. Picton

Rotman Research Institute, Baycrest Centre for Geriatric Care, and University of Toronto

The physiological processes underlying the segregation of concurrent sounds were investigated through the use of event-related brain potentials. The stimuli were complex sounds containing multiple harmonics, one of which could be mistuned so that it was no longer an integer multiple of the fundamental. Perception of concurrent auditory objects increased with degree of mistuning and was accompanied by negative and positive waves that peaked at 180 and 400 ms poststimulus, respectively. The negative wave, referred to as *object-related negativity*, was present during passive listening, but the positive wave was not. These findings indicate bottom-up and top-down influences during auditory scene analysis. Brain electrical source analyses showed that distinguishing simultaneous auditory objects involved a widely distributed neural network that included auditory cortices, the medial temporal lobe, and posterior association cortices.

Fundamental to human hearing is the ability to distinguish co-occurring auditory objects, such as the voice of a tenor against the background orchestra. Auditory scene analysis requires parsing the incoming acoustic signal to identify separate and potentially meaningful auditory objects, each attributable to a sound source in the auditory field. Psychophysical studies have shown that listeners can use multiple cues to distinguish simultaneous auditory objects. For example, sounds that have different stimulus onsets, different intensities, or different periodicities are more likely to be considered as coming from different sound sources than sounds that have the same stimulus onset, intensity, or periodicity (for reviews, see Bregman, 1990; Hartmann, 1988).

Bregman (1990) has proposed a model of auditory scene analysis based primarily on the Gestalt laws of organization (Koffka, 1935). In this model, auditory scene analysis involves an initial preattentive process that partitions the acoustic wave into distinct groups of sounds according to Gestalt principles such as grouping by physical similarity, temporal proximity, and good continuity. Sounds are more likely to be assigned to separate sources if they differ widely in frequency, intensity, and spatial location. This stage is independent of listeners' attention being driven by the incoming acoustic data (bottom-up). The outcome of the preattentive analysis may then be subjected to a more detailed analysis by controlled processes (top-down). Whereas the preattentive pro-

cesses group sounds on the basis of physical similarity, the controlled schema-driven processes use prior knowledge to extract meaning from the acoustic data. As such, the schema-driven process depends on representations of previous experiences that have been acquired through learning and a comparison of the incoming sounds with those representations. The use of contexts or knowledge, or both, is particularly evident in unfavorable signal-to-noise listening situations (e.g., the cocktail party example). In the laboratory, individuals correctly perceive more words when a sentence's final words are contextually predictable than when they are unpredictable (Pichora-Fuller, Schneider, & Daneman, 1995). Thus, schema-driven processes provide a way to resolve perceptual ambiguity in complex listening situations.

Understanding how the brain solves the problem of discriminating simultaneous sounds is a major goal for both psychological and physiological sciences. Yet, the notion of auditory grouping, which is central to auditory scene analysis (Bregman, 1990), has received little attention from a neurobiological perspective. How are groups of features combined and represented in the human brain? How long does it take the auditory system to parse the acoustic wave so that multiple auditory objects can be identified?

Evidence from single-cell recording (Cariani & Delgutte, 1996; Palmer, 1990) and computer modeling studies (Beauvieux & Meddis, 1991; Brown & Wang, 1997; McCabe & Denham, 1997; von der Malsburg & Schneider, 1986) suggests that concurrent sound segregation depends on both peripheral and central auditory processes. The present study investigated the neural processes associated with distinguishing concurrent auditory objects using human event-related brain potentials (ERPs). ERPs are particularly well suited to an examination of bottom-up and top-down influences in auditory scene analysis because they provide information-processing sequences in the brain with excellent temporal resolution, even for sounds that are outside the focus of attention. Through analysis of changes in ERP amplitude or latency as a function of perception, inferences can be made about the timing, level of processing, and anatomical location of auditory scene analysis.

Claude Alain, Stephen R. Arnott, and Terence W. Picton, Rotman Research Institute, Baycrest Centre for Geriatric Care, Toronto, Ontario, Canada, and Department of Psychology, University of Toronto, Toronto, Ontario, Canada.

This research was funded by grants from the Canadian Institute for Health Research. We gratefully thank Yu He and Kelly McDonald for technical assistance and Lori Bernstein for helpful comments on earlier versions of this article.

Correspondence concerning this article should be addressed to Claude Alain, Rotman Research Institute, Baycrest Centre for Geriatric Care, 3560 Bathurst Street, Toronto, Ontario M6A 2E1, Canada. Electronic mail may be sent to calain@rotman-baycrest.on.ca.

Auditory ERPs are characterized by obligatory negative and positive potentials peaking, respectively, at about 100 and 200 ms poststimulus (Näätänen & Picton, 1987; Woods, 1995). This N1-P2 complex is related to signal detection, being present when a transient auditory stimulus is audible. However, conscious identification of an auditory event is often associated with an additional late positive wave peaking between 250 and 600 ms poststimulus, referred to as the P300 or P3b (Hillyard, Squires, Bauer, & Lindsay, 1971; Martin, Sigal, Kurtzberg, & Stapells, 1997; Parasuraman & Beatty, 1980; Parasuraman, Richer, & Beatty, 1982). Many studies have examined the processing of auditory material presented sequentially either without noise or against background noise, but only a few studies have compared the N1-P2 complex under these conditions (e.g., Davis, Bowers, & Hirsh, 1967), and none have examined the pattern of neural activity associated with the perception of concurrent auditory objects.

Experiment 1

The first experiment was designed to examine the different patterns of brain activity that can occur in situations that promote the perception of one or two auditory objects. We used a paradigm similar to that of Moore, Glasberg, and Peters (1986), who used loudness-matched complex sounds that involved either all tuned harmonics (or partials) or one mistuned harmonic. The participants indicated whether they heard a single sound (i.e., a regular "buzz" at a pitch equivalent to the fundamental) or two sounds (i.e., the buzz plus a separate sound with a pure-tone quality at the frequency of the mistuned harmonic). When one harmonic in the complex sound was mistuned from its original value by more than 3%, listeners heard it as a separate tone (Moore et al., 1986). Several factors influence the ability to hear a mistuned harmonic as a separate tone, including degree of inharmonicity, harmonic number, and duration of the stimulus (Hartmann, McAdams, & Smith, 1990; Moore et al., 1986).

In the first experiment, we manipulated the degree of mistuning of the second harmonic within a complex sound and used both active and passive listening conditions. In the active listening condition, participants indicated whether they heard a single sound or two sounds while ERPs were recorded conjointly with behavioral measures. This permitted a direct comparison between the ERPs and participants' perceptual experience. In the passive listening condition, ERPs were recorded while participants read a book. This condition allowed us to examine whether concurrent sound segregation can occur outside the focus of attention. Our hypotheses were that an increased mistuning of the second harmonic would promote the perception of two concurrent auditory objects (e.g., Moore et al., 1986) and that a specific ERP signature would emerge when listeners perceived two distinct sounds. If auditory scene analysis involves an early preattentive stage of processing, then mistuning should also affect ERPs during the passive listening condition.

Method

Participants. Twelve young adults (between 24 and 38 years of age; 6 men) participated in the experiment. All participants reported normal hearing and provided informed consent. All but 2 were right-handed.

Stimuli and task. The stimuli consisted of complex sounds obtained by combining 12 pure tones. Each harmonic had an intensity level of 70 dB

SPL with a 400-ms duration including 10-ms rise-fall time. Stimuli were generated digitally at a sampling rate of 20 kHz with a 16-bit Tucker Davis System converter and presented binaurally through TDH-49 headphones. The fundamental frequency was either 200 or 400 Hz. The independent variable was the frequency of the second harmonic that was shifted upward by 1%, 2%, 4%, 8%, or 16% of its original value. For example, when the fundamental frequency was 200 Hz, the 400-Hz second harmonic was replaced with tones of 404, 408, 416, 432, or 464 Hz. Stimuli at the 200-Hz and 400-Hz fundamental frequencies were presented in separate blocks of trials. Within a block of trials, six different (one harmonic and five inharmonic) and equiprobable stimuli were presented in a random order. Order of presentation of 200- and 400-Hz blocks was counterbalanced between participants.

In the active listening condition, the participants were required, on each trial, to indicate whether they heard (a) a single complex sound with one pitch or (b) two sounds, a complex tone and a component with a pure-tone quality not belonging to the complex. They indicated their responses by pressing 1 or 2 on the number pad of the keyboard with their right hand. The intertrial interval (time between response and next stimulus) varied randomly from 800 to 1,200 ms. No feedback was provided on performance. In the passive listening condition, the participants read a book of their choice and ignored the stimuli. The interstimulus intervals during passive listening varied randomly from 800 to 1,200 ms. The order of the listening conditions was counterbalanced across participants.

Electrophysiological recording and analysis. Electroencephalograms (EEGs) were recorded from an array of 47 electrodes including those from the standard 10-20 placement. Vertical and horizontal eye movements were recorded with electrodes at the outer canthi and at the superior and inferior orbits. Electrophysiological signals were digitized continuously (band pass: 0.05-50 Hz; 250-Hz sampling rate) through NeuroScan Syn-Amps and stored for off-line analysis. During recording, all electrodes were referenced to 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 1,000 ms of poststimulus activity. 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 type, and listening condition. ERPs were digitally low-pass filtered to attenuate frequencies above 32 Hz. For each individual average, ocular artifacts (e.g., blinks and lateral movements) were removed by means of ocular source components with Brain Electrical Source Analysis (BESA) software (Berg & Scherg, 1994; Lins, Picton, Berg, & Scherg, 1993a, 1993b).

ERP waveforms were quantified by computing mean values in selected latency regions relative to the mean amplitude of the 200-ms prestimulus activity. Latency regions of 150-250 and 350-450 ms were chosen to encompass the N1, P2, and P3b deflections. Unless otherwise specified, the effects of mistuning on the ERPs were analyzed at a subset of frontocentral electrodes (Fz, FC1, FC2, and Cz). Scalp topographies using the 39 electrodes (omitting the periorcular electrodes) were statistically analyzed after scaling of the amplitudes to eliminate amplitude differences between conditions (McCarthy & Wood, 1985). All measurements were subjected to repeated measures within-subject analyses of variance (ANOVAs) with stimulus type (harmonic or inharmonic stimuli), fundamental frequency (200 or 400 Hz), and electrode as variables. The original degrees of freedom for all analyses are reported throughout the article. Type I errors associated with inhomogeneity of variance were controlled by decreasing the degrees of freedom through the Greenhouse-Geisser epsilon, and probability estimates are based on these reduced degrees of freedom.

Results

Behavioral data. The behavioral results are summarized in Figure 1. The probability of reporting two simultaneous stimuli increased with mistuning, $F(5, 55) = 69.67, p < .001, \epsilon = .312$.

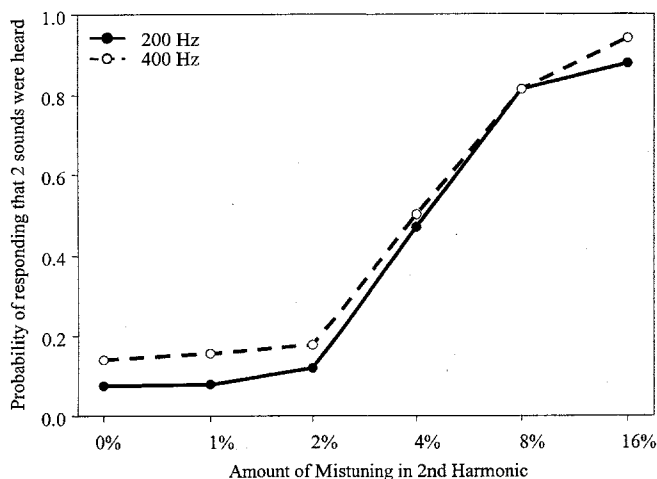


Figure 1. Probability of signaling two auditory objects as a function of mistuning in Experiment 1: Group mean participant subjective reports for the 200-Hz and 400-Hz stimuli.

Participants were more likely to report hearing two distinct auditory stimuli when the second harmonic was mistuned by 4% or more. Pairwise comparisons revealed no differences in perceptual judgments among mistunings of 0%, 1%, or 2%. There was no difference in performance between the 200- and 400-Hz sounds.

ERP data. Figure 2 shows the group mean ERPs recorded during the active listening condition, averaged over the 200- and 400-Hz stimuli, for harmonic and inharmonic stimuli with the second harmonic mistuned by 4% or 16%. All stimuli elicited a large negative-positive (N1-P2) complex peaking respectively at about 110 and 200 ms poststimulus over the midline central and frontal electrodes. The N1-P2 complex was followed by a clear sustained potential and a small offset response over the frontal regions. The ERPs also contained a late positive wave (P3b) maximum over the parietal regions.

An ANOVA on the N1 wave recorded over the midline frontal (Fz) and temporal (T7 and T8) sites did not yield significant differences in amplitude or latency elicited by harmonic and inharmonic stimuli. The first reliable effect of mistuning on the ERPs occurred over the frontocentral region during the P2 interval (i.e., between 150 and 250 ms poststimulus). An ANOVA on the mean voltage amplitude recorded in this latency range yielded a main effect of mistuning, $F(5, 55) = 13.89$, $p < .001$, $\epsilon = .493$. Planned comparisons revealed that complex sounds with the second harmonic mistuned by 16% generated more negative ERPs than any other stimuli ($p < .01$ in all cases). Complex sounds with the second harmonic mistuned by 8% also generated more negative ERP amplitudes than sounds with a 0%, 1%, or 2% mistuned harmonic ($p < .05$ in all cases). There was no significant difference between ERPs elicited by sounds with one harmonic mistuned by 0%, 1%, 2%, or 4%.

The effects of mistuning on ERP amplitude paralleled those observed for the perceptual judgment. Figure 3 shows group mean performance and group mean normalized ERP amplitudes for the 150–250-ms interval at the frontocentral sites. For each participant, a correlation coefficient was calculated between probability of reporting two concurrent stimuli and ERP mean amplitudes. The

significance of these correlations was then examined through a t test on the group mean correlation. A significant negative correlation was found between ERP mean amplitude and perceptual judgment, $r = -.72$, $t(11) = 11.93$, $p < .001$. That is, the larger the negative displacement, the more likely participants were to report perceiving two auditory objects.

Increased mistuning was also associated with a widely distributed late positivity peaking at the midline central site (Cz) between 350 and 450 ms poststimulus. This brain response preceded the P3b response recorded at the midline parietal site. An ANOVA with mistuning, electrode (Fz, Cz, or Pz), and fundamental frequency (200 or 400 Hz) as variables yielded a main effect of mistuning, $F(5, 55) = 10.92$, $p < .001$, $\epsilon = .519$. Complex sounds with the second harmonic mistuned by 8% or 16% generated more positive ERPs than stimuli with the harmonic mistuned by 4% or less ($p < .02$ for all pairwise comparisons). Complex sounds with the second harmonic mistuned by 0%, 1%, 2%, or 4% generated ERPs with similar amplitudes. The interaction between mistuning and electrode was not significant, $F(10, 110) = 2.25$, $p = .09$, $\epsilon = .361$.

The positive displacement at the frontocentral sites also paralleled the effects of mistuning on perceptual judgment (Figure 3). The correlation between ERP amplitude and perceptual judgment was significant, $r = .52$, $t(11) = 4.22$, $p < .002$, with positive displacement increasing with the probability of reporting two concurrent stimuli.

The enhanced positivity over frontocentral regions may partly reflect differences in P3b amplitude–latency. At the midline parietal electrode (i.e., Pz), the P3b showed an earlier onset and peaked earlier when participants reported hearing two distinct stimuli, $F(5, 55) = 4.64$, $p < .01$, $\epsilon = .782$ (Figure 2). Pairwise comparisons revealed a shorter P3b latency for complex sounds with a 16% mistuned harmonic than for any other sounds ($p < .01$ in all cases). P3b latencies were similar for complex sounds with the second harmonic mistuned by less than 8%. P3b peak amplitude also increased with mistuning, $F(5, 55) = 5.81$, $p < .02$, $\epsilon = .351$; the largest amplitude was elicited by stimuli with the 16% mistuned harmonic. There was no difference in P3b peak amplitude elicited by complex stimuli with the harmonic mistuned by 8% or 16%.

The neural activity associated with distinguishing concurrent auditory objects is best illustrated by subtracting ERPs to harmonic stimuli from those elicited by inharmonic stimuli (Figure 4). In the active listening condition, the difference wave revealed a biphasic negative-positive potential that peaked at about 180 and 400 ms after stimulus onset. This biphasic response reversed in polarity at inferior temporal and occipital sites. The negative component, which we refer to as *object-related negativity* (ORN), was present during both active and passive listening conditions, whereas the P400 was present only during the active listening condition.

The effect of mistuning on ERP amplitude recorded during the passive listening condition was examined in a separate ANOVA. For the 150–250-ms intervals, the effect of mistuning on ERP amplitude was significant, $F(5, 55) = 3.12$, $p < .05$, $\epsilon = .750$. Complex sounds that contained the second harmonic mistuned by 16% generated more negative ERPs than stimuli with a mistuned harmonic of 0%, 1%, or 2% ($p < .05$ in all cases). The ORN amplitude recorded during passive listening also correlated with perceptual judgment and the ORN recorded during active listening,

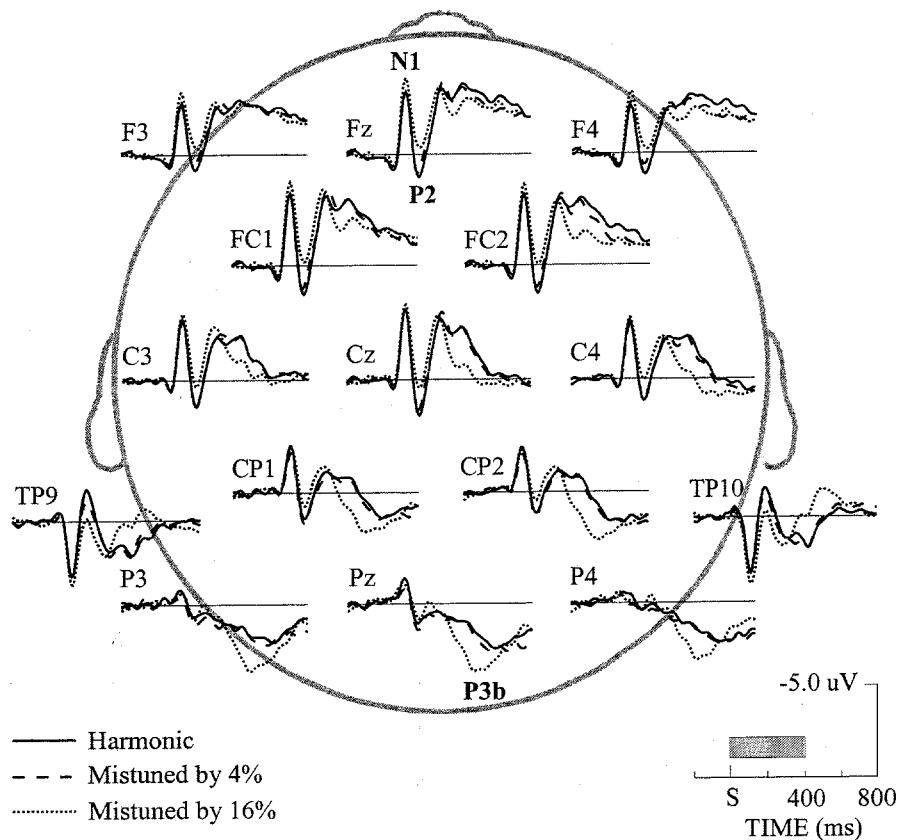


Figure 2. Experiment 1: Group mean event-related brain potentials for complex sounds with all partials at harmonic frequencies or with the second harmonic mistuned by 4% or 16% of the original value. The duration of the stimuli is illustrated by the gray rectangle. Negativity is plotted upward.

$r = -.42$ and $.33$, $t(11) = 3.71$ and 3.06 , respectively ($p < .02$ in both cases). For the 350–450-ms interval, the effect of mistuning on ERP amplitude was not significant.

The ORN and the P400 components were differentially affected by the listening conditions. Whereas the ORN was present in both passive and active listening, the P400 component was present only when participants were required to decide whether one or two stimuli were present. This difference in attentional sensitivity was quantified by comparing the ERP mean amplitude for the 150–250-ms and 350–450-ms intervals elicited by harmonic and inharmonic stimuli with the second harmonic mistuned by 16% in both active and passive listening conditions. Repeated measures multivariate analyses of variance (MANOVAs) with component (ORN or P400), listening condition (active or passive), stimulus type (harmonic or inharmonic), and electrode (Fz, FC1, FC2, or Cz) as variables yielded a significant Component \times Listening Condition \times Stimulus Type interaction, $F(1, 11) = 43.66$, $p < .001$, reflecting greater differences in P400 than ORN amplitude as a function of the listening conditions.

The ORN amplitude was larger when participants were attending to the stimuli than when they were reading: Stimulus Type \times Listening Condition interaction, $F(5, 55) = 3.07$, $p < .05$, $\epsilon = .768$. This enhanced amplitude may have been caused by a change in the gain of the neural generators of the ORN or a superimposition of an attention-related negative wave with a different scalp

topography. To distinguish between these two possibilities, we compared the amplitude distribution of ORNs elicited by stimuli with the second harmonic mistuned by 16% in the active and passive listening conditions. Mean values for the 150–250-ms interval were compared across 39 electrode sites. The interaction among listening condition, mistuning, and electrodes was not significant, $F(38, 418) = 1.45$, indicating that the ORNs recorded during active and passive listening had a similar distribution.

Discussion

When the second harmonic of a complex sound was mistuned from its harmonic value to a sufficient degree, participants reported hearing two distinct auditory objects: one complex sound and one sound with a pure tone quality. This finding is consistent with psychophysical studies showing that mistuning in complex sounds causes the perception of distinct auditory objects (e.g., Carlyon, 1992; Moore et al., 1986). Most of our participants required more than 4% mistuning to report hearing two distinct sounds reliably. This estimate is larger than the 2% reported in other studies (e.g., Moore et al., 1986) and may be related to procedural differences or differences in the participants' experience with the stimuli. The participants in previous studies were usually well-trained listeners.

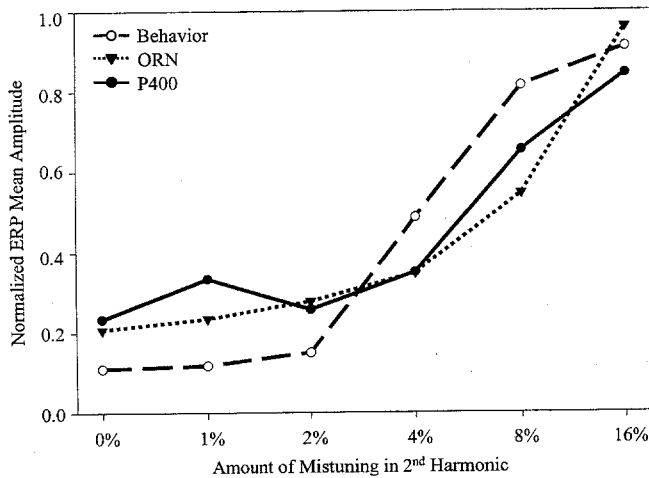


Figure 3. Effects of mistuning on event-related brain potential (ERP) amplitudes recorded during active listening: Experiment 1. As a means of easing the comparison between the ERP amplitudes and behavior, the ERP amplitudes were normalized for each participant, with 1.0 reflecting the largest displacement in negativity (object-related negativity [ORN]) or positivity (P400) for the 150–250-ms and 350–450-ms intervals, respectively. Perceptual judgments for the 200-Hz and 400-Hz stimuli were averaged together. For the behavioral data, the scale reflects the percentage of trials in which participants reported hearing two auditory objects.

Two ERP components were related to the discrimination of simultaneous auditory objects. The first component, the ORN, overlapped the N1 and P2 deflections and peaked at about 180 ms after stimulus onset. The ORN amplitude was maximum at frontocentral sites and inverted in polarity at the mastoid sites, consistent with generators located in the supratemporal plane within the Sylvian fissure. Its amplitude correlated with perceptual judgment being greater when participants were more likely to report that two perceptual objects were present in the mixture. The ORN was present in both active and passive listening conditions, suggesting that this component indexes a relatively automatic process that occurs even when auditory stimuli are not task relevant (e.g., when participants are reading a book). For example, ORN generation may be associated with automatic detection of the mistuned harmonic from the harmonic template extrapolated from the components of the incoming stimulus. This is consistent with 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 & Hartmann, 1998). When a harmonic is mistuned to a sufficient amount, a mismatch occurs between the perceived frequency and that expected on the basis of the template.¹

The second ERP component associated with distinguishing simultaneous auditory objects was a late positive wave (P400) with a widespread distribution. Like the ORN, its amplitude correlated with perceptual judgment, being larger when participants were more likely to perceive the mistuned harmonic as a separate tone. However, in contrast with the ORN, this component was present only when participants were required to indicate whether they heard one or two auditory stimuli. Moreover, P400 generation was more sensitive to attentional manipulation than ORN generation. Thus, whereas the ORN appears to be more associated with auto-

matic processing, the P400 seems to be more related to controlled processes. One possibility would be that the P400 reflects the conscious decision about the presence (or not) of two auditory objects. Participants may reach a decision earlier when the two stimuli can be clearly distinguished because the pitch of the mistuned harmonic “pops out” from the complex sound. In comparison, participants may take a longer time or sample the whole stimulus to decide whether two stimuli are present when the amount of mistuning is small. This postulates a self-terminating process in which duration would be determined by the perception of the mistuned harmonic as a separate tone. Within the framework proposed by Bregman (1990), the ORN may index early bottom-up processes, whereas the P400 may index top-down influences.

However, another explanation could also account for these results. For example, the differences in ERP amplitude might reflect the acoustical differences between the stimuli rather than the perception of the mistuned harmonic as a separate tone. It is possible that participants did not hear the mistuned harmonic as a separate tone but simply responded “two” for those stimuli that were perceptually different from the complex sounds with a 0%, 1%, or 2% mistuned harmonic. This would happen because a forced choice procedure was used in which participants had no alternative but to respond “one” or “two.” In other words, listeners could have made their judgments on the basis of other perceptual cues, such as roughness or unevenness, not related to the segregation of the mistuned harmonic as a separate object. However, previous studies using similar stimuli have shown that listeners can match the pitch of the segregated component with the pitch of a sine wave (e.g., Hartmann et al., 1990; Roberts & Brunstrom, 1998), indicating that the mistuned harmonic was heard as a separate entity.

Although the ORN was present in both listening conditions, its amplitude was larger during active than passive listening. This finding indicates that the processes indexed by the ORN can be influenced by controlled processes. Given that the 200-Hz and the 400-Hz stimuli were presented in separate blocks of trials and that the second harmonic was always the one that was varied, participants may have focused their attention either on a particular frequency region or on the expected frequency of the second harmonic for that block of trials. Behavioral studies have shown that attending to a particular frequency can improve detection of auditory signals (Hubner & Hafner, 1995) and duration judgments (Mondor & Bregman, 1994). Thus, focusing attention on a particular frequency region may have helped listeners segregate the mistuned harmonic from the complex sound and could have caused the enhanced ORN amplitude observed during the active listening condition.

¹ The term *mismatch* in this context is used differently from the way in which it has been used in relation to the ERP component labeled *mismatch negativity*. The mismatch between the mistuned harmonic and the other components of a complex stimulus is based on a representation of the harmonic content of the single stimulus being perceived. Mismatch negativity is an ERP component that occurs when the present stimulus deviates from a representation of what is considered standard on the basis of multiple preceding stimuli. We return to this later.

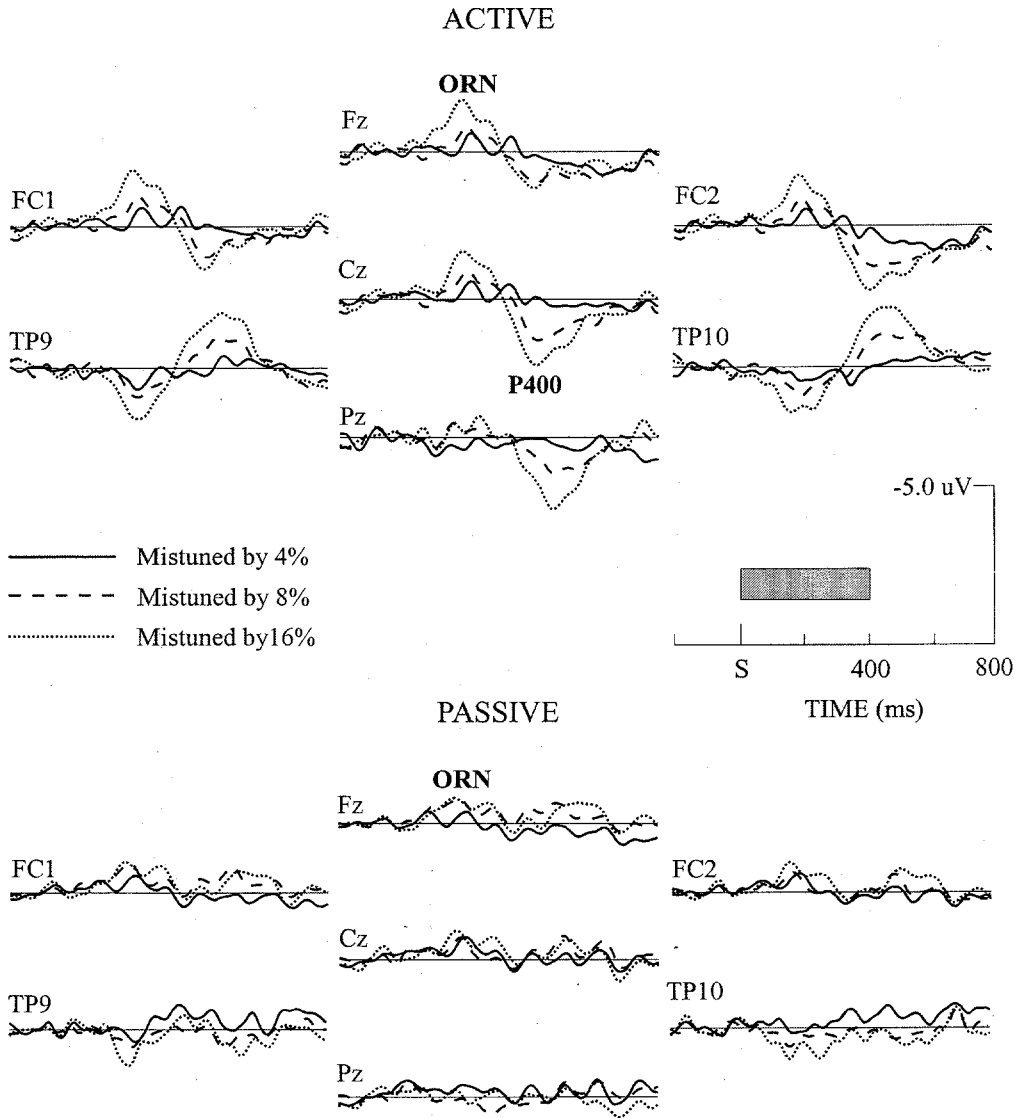


Figure 4. Experiment 1: Difference waves between the event-related brain potentials elicited by sounds with the second harmonic mistuned and the event-related brain potentials elicited by complex sounds without a mistuned harmonic obtained in both active and passive listening conditions. ORN = object-related negativity. Negativity is plotted upward.

Experiment 2

The aim of the second experiment was to see whether the ORN would occur when participants were unable to focus attention on one particular frequency. To this end, we varied randomly the number of the mistuned harmonic from trial to trial. Previous studies have shown that, even when mistuning is a constant proportion of component frequency, individuals are more likely to hear a low rather than a high mistuned harmonic as a separate tone (e.g., Hartmann et al., 1990; Lee & Green, 1994). (The mistuning of a high harmonic within the complex is usually associated not with a clearly perceived additional tone but with roughness or unevenness in the complex sound.) For each fundamental frequency tested, we kept the mistuning constant in frequency (Hz) at all harmonic numbers, thereby making it even more likely that

high harmonic numbers would be less well segregated than low ones. Hence, if the neural activity observed in Experiment 1 indexed concurrent sound segregation, the changes in neural activity in response to mistuning should be larger for the lower than for the higher harmonics. To minimize selective attention to a particular frequency, we intermixed stimuli with 200- and 400-Hz nominal frequencies within the blocks of trials. Under such conditions, listeners' judgments could not be based on a particular pitch, because both the fundamental frequency and the number of the mistuned harmonic would vary randomly from trial to trial.

Method

Participants. Twelve right-handed young adults (between 19 and 34 years of age; 4 men) participated in the second experiment. Four had

participated in Experiment 1. All participants reported normal hearing and provided informed consent.

Stimuli, task, and EEG recording. The duration, rise-fall time, and intensity of the stimuli were identical to those of Experiment 1. The fundamental frequency was either 200 or 400 Hz. The independent variable was the number of the mistuned harmonic (i.e., 2nd, 4th, 6th, 8th, or 10th harmonic) within the complex. For example, for the 200-Hz stimuli, the frequency of the 2nd, 4th, 6th, 8th, or 10th harmonic was shifted upward by 64 Hz (e.g., 16% of the original value of the 2nd harmonic). For the 400-Hz stimuli, the same harmonics were shifted upward by 128 Hz. The 200- and 400-Hz stimuli were intermixed together within a block of trials. Thus, within a block of trials, 12 different (2 harmonic and 10 inharmonic) and equiprobable stimuli were presented in a random order. As in Experiment 1, participants took part in both active and passive listening conditions. The order of conditions was counterbalanced across participants. The ERP recording and data processing were identical to those of Experiment 1.

Results

Behavioral data. Figure 5 shows the probability of reporting two auditory stimuli as a function of harmonic number. An ANOVA with harmonic number (2nd, 4th, 6th, 8th, or 10th) and fundamental frequency as variables revealed a main effect of harmonic number, $F(5, 55) = 108.63$, $p < .001$, $\epsilon = .55$, and a significant interaction between harmonic number and fundamental frequency, $F(5, 55) = 10.55$, $p < .001$, $\epsilon = .592$. Participants usually reported hearing two stimuli when the 2nd, 4th, or 6th harmonic was mistuned and a single sound when the 8th or 10th harmonic was mistuned. They also reported hearing two auditory stimuli more often for the 4th and 6th harmonics when the nominal frequency was 200 rather than 400 Hz.

ERP data. Figure 6 shows the group mean ERPs recorded during the active listening condition, averaged over the 200- and 400-Hz stimuli, for sounds with all harmonics in tune or with their second or eighth harmonic mistuned. All stimuli elicited a large N1-P2 complex over midline central and frontal electrodes, followed by a sustained negativity over frontal regions and a P3b potential over the parietal regions.

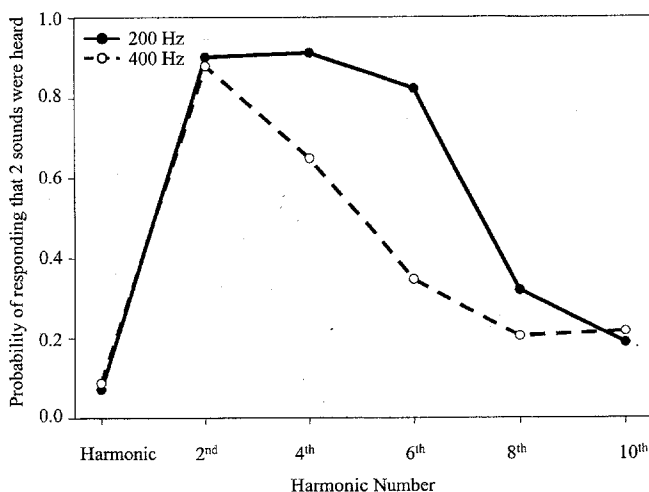


Figure 5. Probability of signaling two auditory objects as a function of harmonic number in Experiment 2: Group mean participant subjective reports for the 200-Hz and 400-Hz stimuli.

An ANOVA on mean amplitudes between 150 and 250 ms poststimulus with harmonic number, fundamental frequency, and electrode as variables revealed a main effect of harmonic number, $F(5, 55) = 4.68$, $p < .01$, $\epsilon = .625$, and a main effect of fundamental frequency, $F(1, 11) = 41.04$, $p < .001$. The complex sound with the 2nd harmonic mistuned generated more negative ERPs than did the harmonic or inharmonic stimuli with the 8th or the 10th harmonic mistuned ($p < .01$ in all cases). The complex sound with the 4th harmonic mistuned also generated more negative ERPs than the harmonic complex ($p < .01$). Moreover, ERPs to complex sounds with the 4th harmonic mistuned tended to be more negative than ERPs elicited by stimuli with the 8th or 10th harmonic mistuned ($ps = .052$ and $.076$, respectively). Harmonic and inharmonic stimuli with the 8th or 10th harmonic mistuned generated similar ERP amplitudes during the 150–250-ms interval. The interaction between harmonic number and fundamental frequency was not significant, $F(5, 55) = 2.11$.

As in the first experiment, the effects of mistuning on ERP amplitude paralleled the perceptual judgment (see Figure 7). A significant negative correlation was found between ERP amplitude and perceptual judgment, $r = -.56$, $t(11) = 6.09$, $p < .001$. That is, the greater the negative displacement during the 150–250-ms interval, the more likely participants reported hearing two concurrent stimuli.

The P400 amplitude (350–450 ms) at the midline electrodes (Fz, Cz, and Pz) also varied as a function of harmonic number, $F(5, 55) = 4.67$, $p < .01$, $\epsilon = .634$. Stimuli with the 2nd harmonic mistuned generated more positive ERPs than the other sounds ($p < .05$ in all cases). Complex sounds with the 4th harmonic mistuned also generated more positive ERPs than those with the 10th harmonic mistuned ($p < .05$). The interaction between harmonic number and fundamental frequency was not significant, nor was the interaction between harmonic number and electrode.

As for the ORN, the mean P400 amplitude at frontocentral sites paralleled the effects of mistuning on perceptual judgment (Figure 7). The correlation between ERP amplitude and perceptual judgment was significant, $r = .51$, $t(11) = 5.75$, $p < .001$, with amplitude increasing with increasing probability of reporting two distinct stimuli.

The P3b peak amplitude and latency measured at Pz were not affected by the number of the mistuned harmonic within the complex, $F_s(5, 55) = 1.85$ and $.497$, respectively. However, the P3b mean amplitude (500–600 ms) did vary with harmonic number, $F(5, 55) = 3.89$, $p < .05$, $\epsilon = .529$. Planned comparisons revealed that the P3b to the harmonic complex and inharmonic sounds with the 2nd harmonic mistuned was larger than the one obtained when the 8th or 10th harmonic was mistuned ($p < .05$ in all cases). There was no significant difference in P3b amplitude or latency elicited by harmonic and inharmonic complexes with the 2nd harmonic mistuned. The interaction between harmonic number and fundamental frequency was not significant.

Figure 8 shows the ORN and P400 responses isolated in the difference wave between the ERP to harmonic and inharmonic stimuli with a second, fourth, or eighth mistuned harmonic during the active and passive listening conditions. In both listening conditions, the ORN peaked approximately 180 ms after stimulus onset and reversed in polarity at inferior temporal sites. ORN amplitude was not affected by listening condition, $F(5, 55) = 0.44$. As in the first experiment, the ORN was followed by a P400

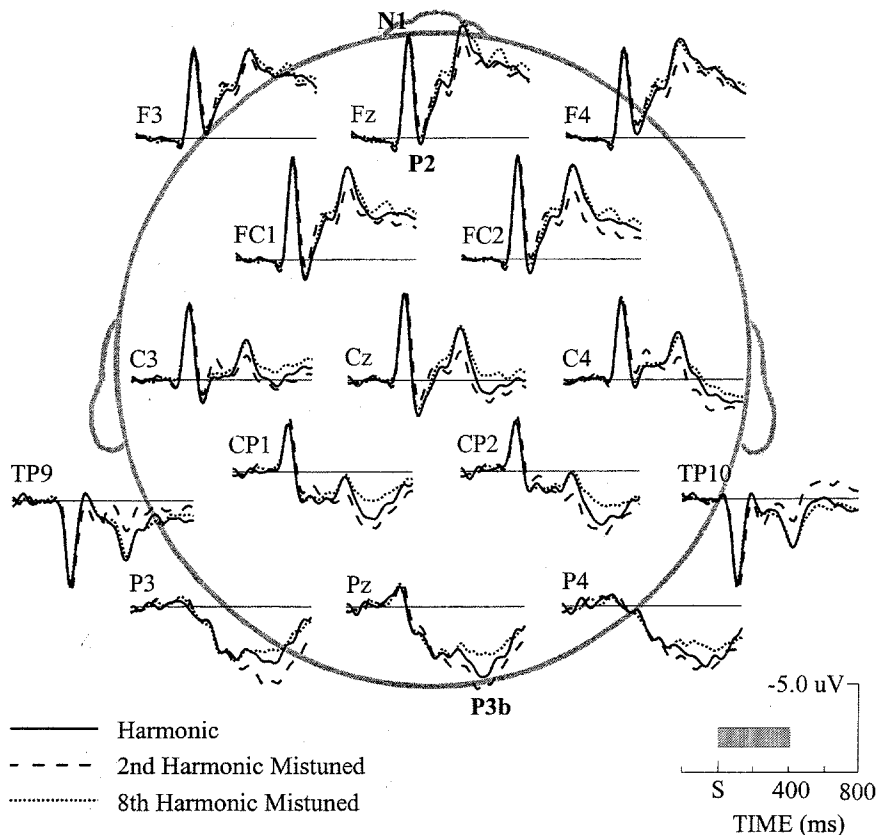


Figure 6. Experiment 2: Group mean event-related brain potentials for complex sounds with all partials at harmonic frequencies or with the second or eighth harmonic mistuned. Negativity is plotted upward.

component only when participants were required to decide whether one or two sounds were present.

The ORN amplitude recorded during passive listening varied as a function of harmonic number. The ANOVA yielded a main effect of harmonic number, $F(5, 55) = 4.57, p < .01, \epsilon = .629$, and a main effect of fundamental frequency, $F(1, 11) = 29.42, p < .001$. The interaction between harmonic number and fundamental frequency was also significant, $F(5, 55) = 5.00, p < .01, \epsilon = .637$. The effects of mistuning on ERPs were larger for the 400-Hz than for the 200-Hz fundamental frequency. Inharmonic complexes with a 2nd mistuned harmonic generated larger ORNs than harmonic or inharmonic complexes with the 6th, 8th, or 10th harmonic mistuned ($p < .01$ in both cases). Stimuli with the 4th harmonic mistuned also generated a larger ORN than harmonic or inharmonic complexes with the 6th, 8th, or 10th harmonic mistuned ($p < .05$ in all cases). The ORN amplitude recorded during passive listening correlated with perceptual judgment and the ORN recorded during active listening, $r_s = -.51$ and $.34, t_s(11) = 6.09$ and 5.09 , respectively ($p < .01$ in both cases). For the 350–450-ms interval, the effect of mistuning on ERP amplitude was not significant.

As in the first experiment, the P400 component appeared much more affected by listening conditions than the ORN component. This difference in attentional sensitivity was quantified through a MANOVA with component, listening condition, stimulus type (harmonic or complex sounds with second harmonic mistuned),

and electrode as variables. The interaction among ERP component, listening condition, and stimulus type was significant, $F(1, 11) = 5.38, p < .05$, reflecting larger attention effects on the P400 than on the ORN generation.

Discussion

Participants were more likely to report hearing two auditory stimuli when the mistuned harmonic within the complex was lower rather than higher in frequency. The listeners' judgments could not be based on hearing a particular pitch because both the fundamental frequency and the harmonic number varied randomly from trial to trial. This finding is consistent with previous studies showing that listeners can hear lower mistuned harmonics (sixth or lower) more easily as a separate tone than higher mistuned harmonics (e.g., Hartmann et al., 1990; Lee & Green, 1994). The difficulty in parsing out the high harmonics from the complex may be related to the fact that the low and high harmonics were mistuned by the same frequency amount (e.g., 64 Hz for the 200-Hz sound). Thus, the percentage of mistuning relative to the in-tune harmonic decreased with increasing harmonic number. Although the experimental design does not allow the effects of harmonic number and the effects of inharmonicity to be distinguished, it significantly affected the participants' ability to detect concurrent auditory stimuli. This suggests that perception of simultaneous auditory objects does not depend entirely on the listener's attention to a

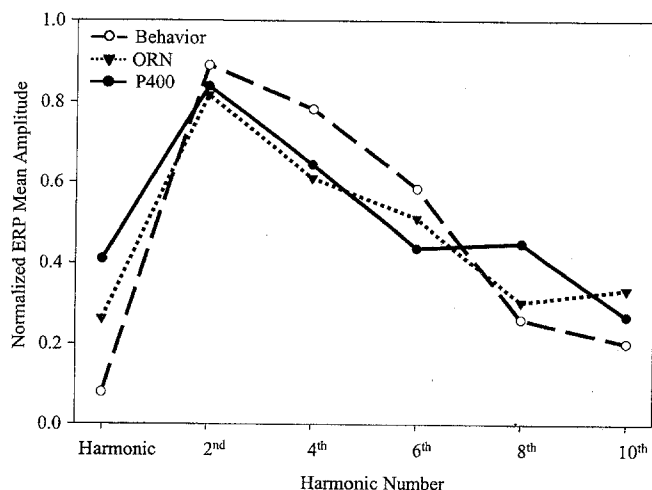


Figure 7. Effects of harmonic number on the object-related negativity (ORN) and P400 responses recorded during the active listening condition in Experiment 2. As a means of easing the comparison between event-related brain potentials (ERPs) and behavior, the ERP amplitudes were normalized for each participant, with 1.0 reflecting the largest displacement in negativity or positivity for the 150–250-ms and 350–450-ms intervals, respectively. Perceptual judgments for the 200-Hz and 400-Hz stimuli were averaged together. For the behavioral data, the scale reflects the percentage of trials in which participants reported hearing two auditory objects.

particular frequency region, consistent with the proposal that parsing of concurrent sounds can occur automatically.

Experiment 2 replicated and extended the main finding of Experiment 1. Perception of simultaneous auditory objects was associated with ORN and P400 responses. Both the ORN and P400 amplitudes varied with perceptual judgment, being larger when participants were more likely to report hearing two distinct auditory objects.

As in the first experiment, the ORN was present during both passive and active listening, whereas the P400 potential was present only when participants were attending to the stimuli. Moreover, ORN generation was less affected by attentional manipulation than P400 generation. This is consistent with the proposal that these two ERP components index distinct stages of processing, with the ORN reflecting an early automatic detection of the mistuned harmonic and the P400 indexing the attentional, controlled processes of the complex sound for an eventual response. In Experiment 2, the ORN amplitude was similar in both active and passive listening conditions. This finding provides converging evidence that ORN generation does not depend on selective attention to a particular frequency region. It also suggests that the larger amplitude observed in the active listening condition of the first experiment was due to the experimental design that encouraged listeners to focus their attention on a particular frequency region. Overall, the results from the two experiments suggest that the auditory system automatically groups together spectral regions that exhibit the same periodicity but that selective attention can facilitate this process.

As mentioned earlier, the ORN may reflect the mismatch between the mistuned harmonic and the harmonic template extrapolated from the incoming stimulus. However, the ORN may also

reflect an automatic mismatch response between the incoming stimulus and a representation of what is considered standard on the basis of the preceding stimuli in short-term memory. Many studies have shown that sounds that deviate from standard stimuli elicit a mismatch negativity (MMN) wave that is similar in latency and topography to the ORN component observed in our experiments (Näätänen, 1992). Although in the current study, the stimuli were equiprobable (i.e., there was no standard or deviant per se), it is possible that the participants sorted them into two groups. In the first experiment, tones considered harmonic (0%, 1%, 2%, and some of the 4% mistuned harmonic) could have been considered standard, and the less probable mistuned stimuli could have been considered deviant. Similarly, more than half of the stimuli in the second experiment were considered harmonic. Thus, the ORN could index a mismatch process that depends on either a harmonic template extrapolated from the present stimulus or a standard representation derived from the preceding stimuli.

Experiment 3

The aim of the third experiment was to examine whether ORN generation depended on information within the present stimulus or whether it depended, like the MMN, on representations of previously presented stimuli in short-term memory. To test these alternatives, we varied the proportion of stimuli with a mistuned harmonic within a block of trials. MMN amplitude has been shown to increase with decreasing probability of the deviant stimuli (for a review, see Näätänen, 1992; Picton, Alain, Otten, Ritter, & Achim, 2000). Thus, if the ORN is like the MMN and relies on a representation of the preceding stimuli, then it should vary as a function of the stimulus probability, being larger when stimuli with the mistuned harmonic are presented infrequently. However, if the detection of the mistuned harmonic depends on the harmonic template derived from the incoming stimulus, then the ORN should not be affected by the probability of the stimulus.

Method

Participants. Sixteen young adults (between 19 and 35 years of age; 9 men) participated in the third experiment. Four had participated in Experiment 1, Experiment 2, or both. All participants reported normal hearing and provided informed consent. The results from 2 participants were excluded because of excessive eye movement and muscle artifacts.

Stimuli, task, and EEG recording. The stimulus duration, rise-fall time, and intensity were identical to those of Experiment 1. The fundamental frequency was either 200 or 400 Hz. Within a block of trials, participants were presented with complex sounds that had either all partials at harmonic frequencies or the second harmonic mistuned by 16% of its original value. The independent variable was the percentage of complex sounds with a mistuned harmonic within a block of trials. In separate blocks of trials, the percentage of complex sounds with a mistuned harmonic was 30%, 50%, or 70%. As in Experiment 2, the 200- and 400-Hz stimuli were intermixed together within a block of trials to discourage selective attention to a particular frequency. Thus, within a block of trials, four different stimuli (two harmonic and two inharmonic) were presented in a random order. Again, participants took part in both active and passive listening conditions. The order of conditions was counterbalanced across participants. The ERP recording and data processing were identical to those of Experiment 1.

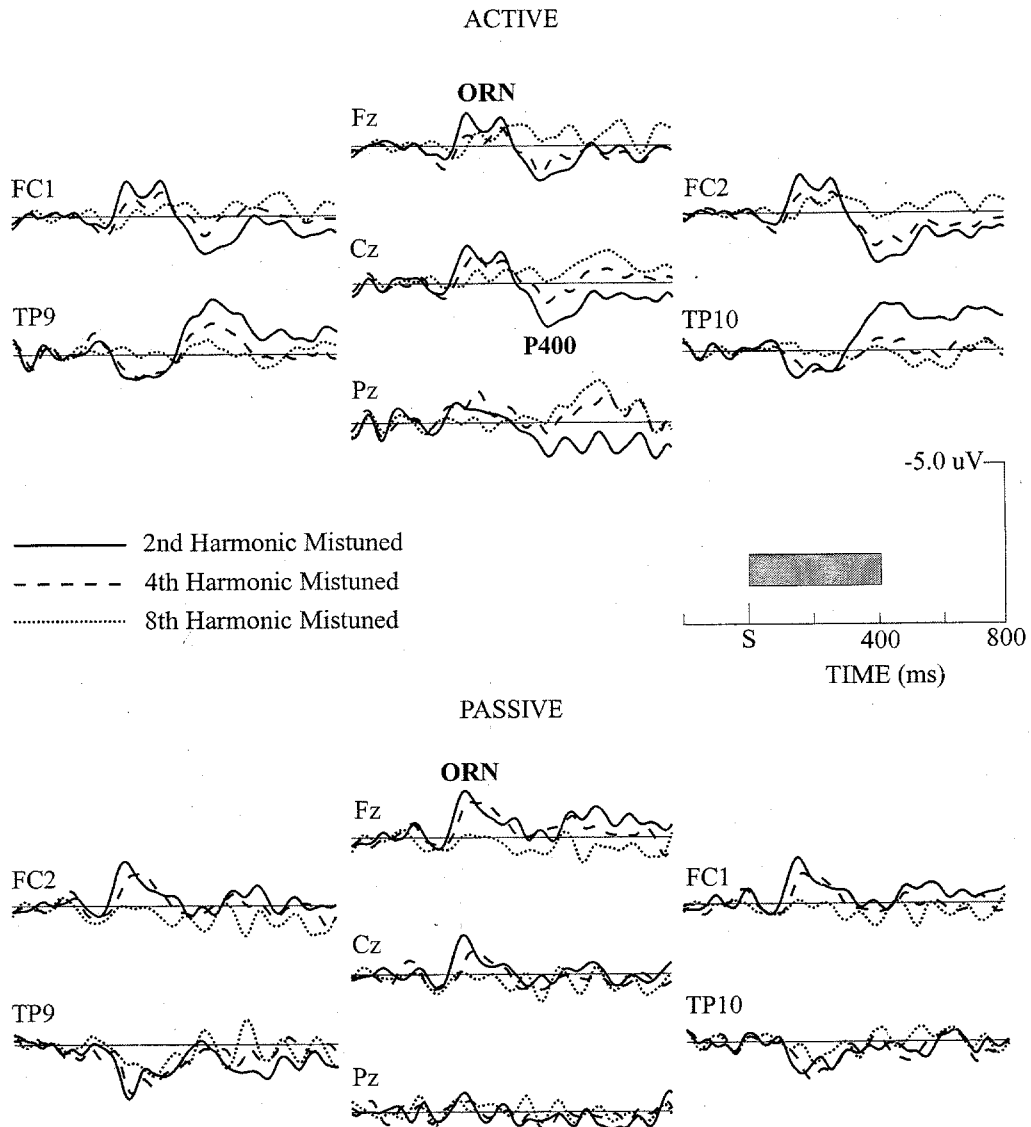


Figure 8. Experiment 2: Difference waves between the event-related brain potentials elicited by sounds with a mistuned harmonic and the event-related brain potentials elicited by complex sounds without a mistuned harmonic obtained in both active and passive listening conditions. ORN = object-related negativity. Negativity is plotted upward.

Results

Behavioral data. Overall, participants were more likely to report hearing two stimuli when the complex sound included a mistuned harmonic. The perception of the harmonic stimuli was not affected by the probability manipulation (Figure 9). In comparison, participants were less likely to report hearing two stimuli when the complex sounds with the mistuned harmonic were presented frequently, resulting in a significant Stimulus Type \times Probability interaction, $F(2, 26) = 6.99, p < .01, \epsilon = .753$.

ERP data. Figure 10 shows the group ERPs recorded during the active listening condition, averaged over the 200- and 400-Hz stimuli, for sounds with all partials at harmonic frequencies or with their second harmonic mistuned by 16%. All stimuli elicited a

large N1-P2 complex over midline central and frontal electrodes, followed by a sustained negativity over frontal regions and a P3b potential over the parietal regions. The N1 and P2 amplitude and latency elicited by harmonic or inharmonic stimuli did not vary as a function of stimulus probability. Whereas the P3b elicited by harmonic stimuli did not vary in amplitude as a function of stimulus probability, the P3b elicited by complex sounds with the second harmonic mistuned increased in amplitude with decreasing stimulus probability, resulting in a significant Stimulus Type \times Probability interaction, $F(2, 26) = 12.04, p < .01, \epsilon = .690$. P3b latency was not affected by the probability manipulation and was similar for both harmonic and inharmonic stimuli.

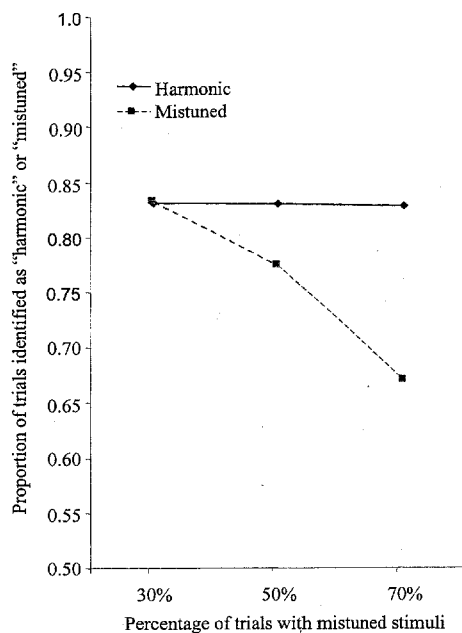


Figure 9. Experiment 3: Probability of reporting hearing one sound or two sounds as a function of stimulus probability within a block of trials.

In comparison with the harmonic stimulus, the ERPs elicited by complex sounds with a mistuned harmonic were associated with an ORN at frontocentral sites peaking between 125 and 225 ms after stimulus onset, $F(1, 13) = 30.88, p < .001$. The effect of mistuning on ERPs can be seen in the difference waves between the ERPs elicited by harmonic stimuli and the ERPs elicited by inharmonic stimuli (Figure 11). The ORN mean amplitude (125–225 ms) during the active listening condition was not affected by the probability manipulation: Probability \times Stimulus Type interaction, $F(2, 26) = 1.51$.

A similar pattern of results was observed during the passive listening condition (Figure 11). At frontocentral sites, inharmonic stimuli generated more negative ERPs than harmonic stimuli, $F(1, 13) = 24.68, p < .001$. The ORN was slightly larger when the stimuli with the mistuned harmonic were presented infrequently, but the interaction between probability and mistuning failed to reach significance, $F(2, 26) = 3.07, p = .085, \epsilon = .708$. The interaction between fundamental frequency and mistuning was not significant.

The ORN amplitude recorded at frontocentral sites did not differ significantly as a function of attention, $F(1, 13) = 0.10$. However, visual inspection of the data suggested that the ORNs recorded at the mastoid sites were larger in the active than passive listening conditions. The effect of attention on the ORN amplitude recorded at the mastoid was tested in a separate ANOVA. The analysis revealed a significant effect of listening condition on ORN recorded at the mastoid sites, $F(1, 13) = 5.14, p < .05$. At both frontocentral and mastoid sites, the interaction among probability, stimulus type, and listening condition was not significant, $F_s(2, 26) = 1.12$ and 1.35 , respectively.

In the active listening condition, the P400 amplitude varied as a function of stimulus probability. An ANOVA with probability, stimulus type, and electrode (Fz, Cz, or Pz) as variables yielded a

main effect of stimulus type, $F(1, 13) = 5.83, p < .05$. The interaction between probability and mistuning was also significant, $F(2, 26) = 4.05, p < .05, \epsilon = .892$, revealing a larger P400 amplitude when the stimuli with the mistuned harmonic were presented infrequently than when they were presented frequently. The interaction among stimulus type, probability, and electrode was not significant.

A MANOVA was used in examining the effect of the listening condition on the ORN and P400 amplitudes recorded at frontocentral sites. As in Experiments 1 and 2, the interaction among component, listening condition, and stimulus type was significant, $F(1, 13) = 5.47, p < .05$, indicating that P400 generation is more sensitive to attentional manipulation than ORN generation.

Discussion

As previously observed in Experiments 1 and 2, participants were more likely to report hearing two distinct auditory stimuli when complex sounds included a mistuned harmonic. However, participants were less likely to report hearing the mistuned harmonic as a separate tone when those stimuli were presented frequently. This finding suggests that the sequential context influences perception of the mistuned harmonic as a separate tone. This is consistent with a study showing that the pitch changes produced by a mistuned harmonic can be modulated by varying the frequency of a pure tone preceding the complex sound with a mistuned harmonic (Darwin, Hukin, & al-Khatib, 1995).

The probability of reporting hearing two auditory stimuli was associated with ORN and P400 responses. As in Experiments 1 and 2, the ORN was present in both active and passive listening conditions, whereas the P400 was present only during the active listening condition. ORN generation was also less affected by attentional manipulation than P400 generation. This finding is consistent with the proposal that these two components index distinct stages of processing, with ORN generation being more automatic than P400 generation. The effects of stimulus probability on the ORN and P400 provide further evidence that these two components index different aspects of auditory scene analysis. Whereas the ORN was not affected by probability during the active listening condition, the amplitude of the P400 increased with the decreasing probability of presenting a complex sound with a mistuned harmonic. The effects of probability manipulation on the P400 parallel the phenomenological experience of the participants; that is, the larger the P400 amplitude, the more likely participants were to report perceiving two auditory objects. Given that the ORN was not significantly affected by the probability manipulation, the effects of probability indicate that controlled processes may overwrite the outcome from the early, preattentive stage of processing.

The ORN was larger at the mastoid electrodes during the active than passive listening, suggesting that the ORN amplitude can be modulated by attention. As in the first experiment, participants may have focused their attention on a particular frequency because it was always the second harmonic that was mistuned. However, the attention effect was somewhat smaller than in the first experiment. This could be attributed to the mixed design that included both fundamental frequencies within the same block of trials. This design may have partly prevented participants from focusing their attention on a particular frequency. Still, it may not have discour-

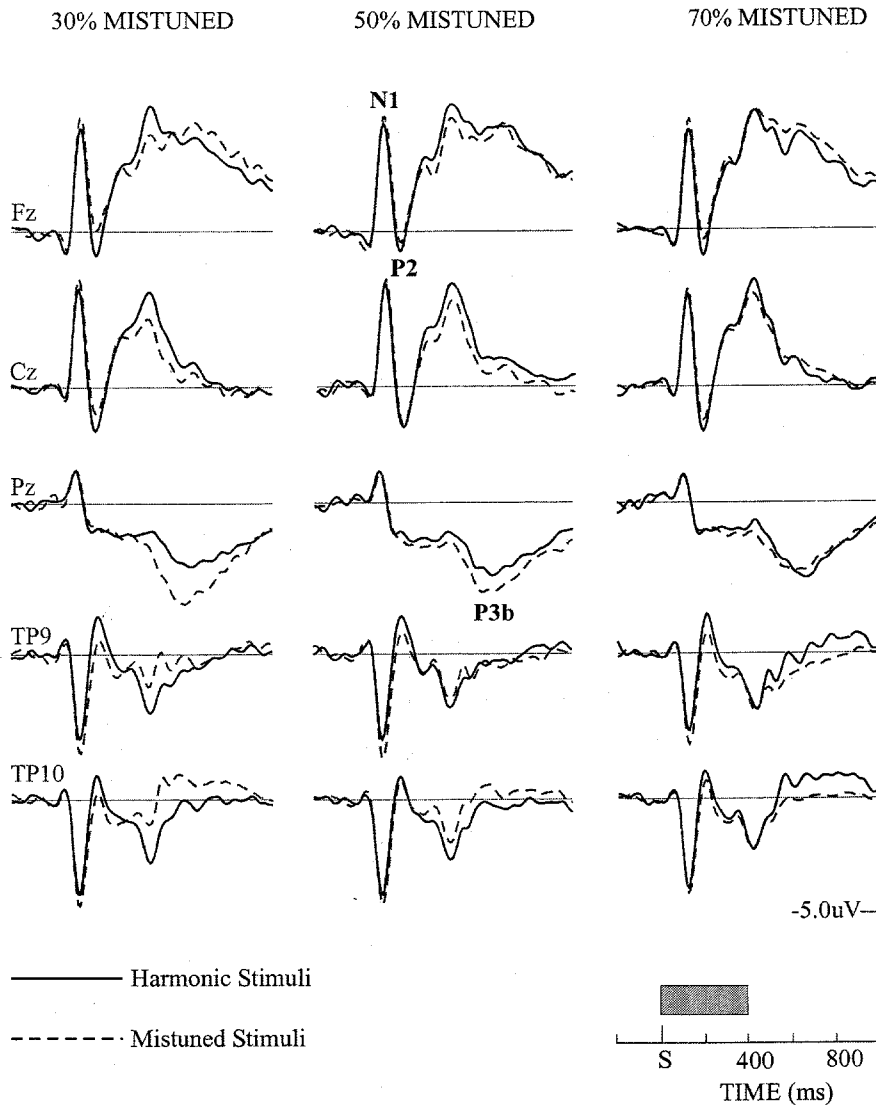


Figure 10. Experiment 3: Group mean event-related brain potentials recorded at the midline frontal (Fz), central (Cz), and parietal (Pz) sites, as well as at the left (TP9) and the right (TP10) mastoids, for complex sounds with all partials at harmonic frequencies or with the second harmonic mistuned by 16% of its original value. Negativity is plotted upward.

aged participants completely from focusing their attention on a particular frequency region.

The main goal of the third experiment was to determine whether ORN generation depended on a representation of the previously presented stimulus (as in the MMN case) or whether it depended on a representation of the harmonic template extrapolated from the incoming acoustic input. In the passive listening condition, the ORN amplitude tended to be smaller when the mistuned harmonic was presented frequently. The reduced ORN amplitude for the frequently presented mistuned harmonic may have been caused by the superimposition of the MMN wave elicited by the infrequent harmonic stimuli. The subtraction procedure used in the present study would have resulted in two negative potentials, one associated with detecting simultaneous auditory objects and one associated with detecting deviance from what is considered standard on

the basis of the preceding stimuli. Although the P3b could occur for both probable and improbable stimuli, the MMN should not be elicited by probable stimuli (which would be considered standard rather than deviant). The presence of the ORN in the probable condition shows that this new ERP component cannot be entirely explained as an MMN. It also indicates that ORN generation does not depend on representations of the previously presented stimuli but, rather, reflects "on-line" detection of simultaneous objects.

Brain Electrical Source Analysis

In all three experiments described so far, the ORN wave inverted in polarity between the frontocentral and the mastoid sites, consistent with generators in the supratemporal plane along the Sylvian fissure. The P400 also inverted between the central and

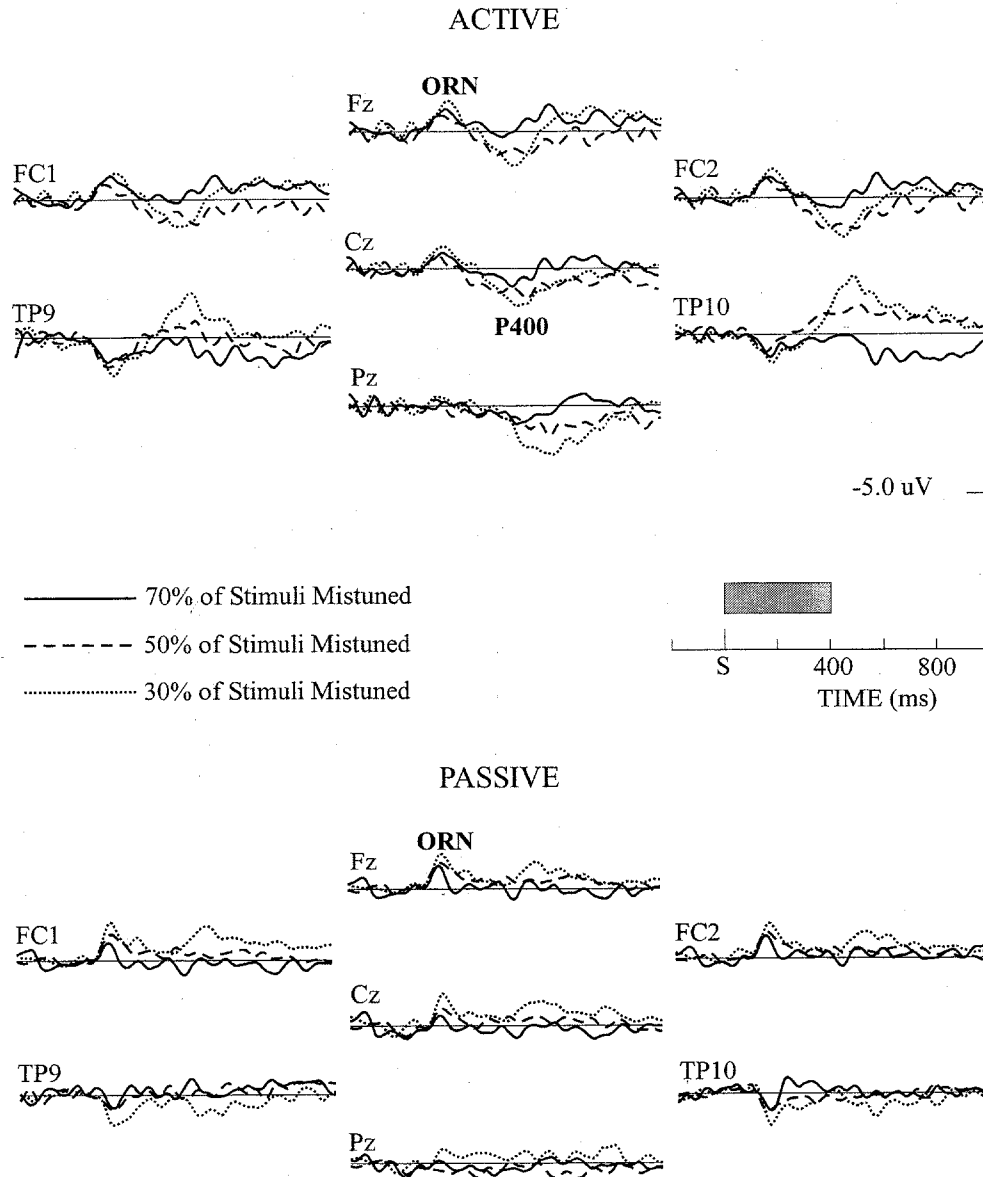


Figure 11. Experiment 3: Difference waves between the event-related brain potentials elicited by sounds with a mistuned harmonic and the event-related brain potentials elicited by complex sounds without a mistuned harmonic obtained in both active and passive listening conditions. ORN = object-related negativity. Negativity is plotted upward.

mastoid sites but showed a more widespread distribution than the ORN, suggesting different generators. We used BESA software (Scherg, 1989) to identify the underlying generator associated with distinguishing simultaneous auditory objects.

Method

Data set. The group mean ERPs used in the source analysis were obtained by combining the ERPs from the three experiments to optimize signal-noise ratio. The source analyses were restricted to ERPs elicited by harmonic stimuli and complex sounds with the second harmonic mistuned by 16% of its original value. The ERPs elicited by the 200- and 400-Hz stimuli were averaged together. In Experiment 3, the harmonic and inhar-

monic complexes were presented at different probabilities. To minimize variability, we included only the ERPs recorded from the block of trials in which the mistuned harmonic was presented infrequently. This block of trials was the most similar in terms of probability with the first two experiments. The ERPs recorded during active and passive listening were analyzed separately. The group mean ERPs included data from 38 participants.

Procedure. The analysis was carried out through a four-shell head model with relative conductivities of 1.0, 3.0, 0.0125, and 1.0 for the brain, cerebrospinal fluid, skull, and scalp, respectively, and radii of 71, 72, 79, and 85 mm.

A principal-components analysis was used to estimate the number of sources that might be found in the data. Given that patterns of activity are

often similar between generators in the two hemispheres, each of the component waveforms might have been generated by two sources, one in each hemisphere. Therefore, the number of possible sources was estimated at twice the number of components needed to account for 99% of the variance.

Results and Discussion

A principal-components analysis of the scalp-recorded waveform elicited by the harmonic stimuli during the passive listening conditions showed two main components, one involving the N1, P2, and sustained potential (95% of the variance) and another involving a peak slightly later than the N1 wave (3% of the variance). The active-passive difference waveforms showed two additional components, a slow wave beginning at approximately 200 ms and becoming larger as the sweep progressed, with a smooth peak near 600 ms (86% of the variance), and another wave peaking between 300 and 400 ms (9% of the variance). Analysis of the responses to the mistuned harmonic stimulus showed almost identical waveforms. We determined sources using pairs of symmetrical dipoles and fitted these dipoles to various regions of the recorded ERPs or the difference waveforms.

The final BESA solution contained eight sources (Figure 12). The residual variances for this solution over the latencies 50–700 ms were 3.6% for the passive conditions and 9.7% for the active conditions. Sources 1 and 2 were located in the region of the supratemporal plane and were directed toward the vertex. These sources showed a major peak near 105 ms poststimulus, followed by a sustained wave. The waveform of these sources was similar to that of the first principal component. Initially, we were able to distinguish the N1 source from a sustained potential source that was more anterior (cf. Scherg, Vajsar, & Picton, 1989). However, this distinction was not stable when we attempted to model the later P3b waves in the active conditions, and we therefore decided to model both the N1 and the sustained potential using one pair of dipoles. Sources 3 and 4 were close to Sources 1 and 2 but were oriented laterally. They showed source waveforms similar to that of the second principal component with the major peak at 140 ms poststimulus. The locations of these sources are consistent with results from intracerebral recordings in humans (Halgren, Baudena, Clarke, Heit, Liegeois, et al., 1995; Richer, Alain, Achim, Bouvier, & Saint Hilaire, 1989) and nonhuman primates (Stein-schneider, Arezzo, & Vaughan, 1980; Steinschneider, Reser, Fishman, Schroeder, & Arezzo, 1998), as well as with lesion studies in humans (Knight, Hillyard, Woods, & Neville, 1980; Woods, Clayworth, Knight, Simpson, & Naeser, 1987; Woods, Knight, & Scabini, 1993).

Sources 5 and 6 were derived to fit the slow positive deflection that occurred in the active conditions. Fitting was constrained to provide a radial orientation, and the main fitting was based on the active-passive difference waveforms with a residual variance between 450 and 750 ms. Sources 7 and 8 were used to fit the residual variance in the active-passive waveforms that was not fit through Sources 5 and 6. Fitting was based on the residual variance between 300 and 550 ms, and the BESA variance constraint was used to make the source waveform most active within the same latency region. Sources 7 and 8 provided similar fits to the data for a broad region of the medial temporal area. The final locations of these sources were chosen so that the sources did not interfere

significantly with Sources 1 and 2 when all of the sources were used together to fit the ERP waveforms. Sources 5, 6, 7, and 8 were active only when the stimuli were attended. The BESA solution is consistent with results from intracerebral recordings in humans (Alain, Richer, Achim, & Saint Hilaire, 1989; Halgren, Baudena, Clarke, Heit, Marinkovic, et al., 1995; Halgren, Marinkovic, & Chauvel, 1998; Halgren et al., 1980). These studies showed that, relative to task-irrelevant materials, task-relevant stimuli generated large potentials in medial temporal structures and posterior auditory association cortices.

The major difference between the response to the mistuned stimulus and the response to the harmonic stimulus was a change in the activity of the first two sources during the time of the ORN. Directly modeling the ORN in the harmonic minus inharmonic difference waveforms resulted in sources located slightly inferior and medial to Sources 1 and 2. However, attempts to obtain stable solutions for the unsubtracted waveforms that included Sources 1 and 2 and separate sources for the ORN were not successful.

In a separate source analysis, we compared the dipole source location for the N1 and ORN recorded during active and passive listening conditions. For each participant, two symmetrical regional sources were fit on a 40-ms interval centered on the peak of the N1 deflection elicited by the harmonic stimulus and on the mean 40-ms interval centered on the peak of the ORN. The *x* (left–right), *y* (front–back), and *z* (up–down) coordinates for the N1 and ORN source locations were then compared in a MANOVA.

Figure 13 shows the group mean location for the N1 and ORN recorded during the active listening condition. The residual variances for the N1 and ORN were 3.0% and 3.6%, respectively. The N1 sources (i.e., 1 and 2) were located near the transverse temporal gyri along the supratemporal plane. The ORN sources (i.e., 3 and 4) were inferior and medial to the N1 sources, consistent with neural activity in the medial Heschl's gyri or in the medial planum temporale. The MANOVA on the coordinates yielded a significant difference in source location between the N1 and ORN recorded during active and passive listening conditions, $F_{s(1, 37)} = 18.71$ and 3.44, respectively ($p < .05$ in both cases). This finding indicates that neurons activated by co-occurring auditory stimuli are different from those activated by stimulus onset (see also Imaizumi, Mori, Kiritani, & Yumoto, 1997).

General Discussion

The combination of behavioral and electrophysiological techniques reveals for the first time a human brain response associated with the discrimination of simultaneous auditory objects. In three experiments, the ORN began 100 ms after stimulus onset and peaked at about 180 ms. The amplitude was maximum over the frontocentral region and inverted in polarity at inferior temporal sites, consistent with generators in auditory cortices within the Sylvian fissure. When the stimuli were attended, the ORN was followed by a widely distributed P400 response. This component was maximum over the central and posterior regions and inverted in polarity at mastoid sites. The scalp topography was consistent with concurrently active generators in the medial temporal lobe and posterior auditory association cortices.

In all three experiments, ORN and P400 generation varied systematically as a function of the attentional manipulation.

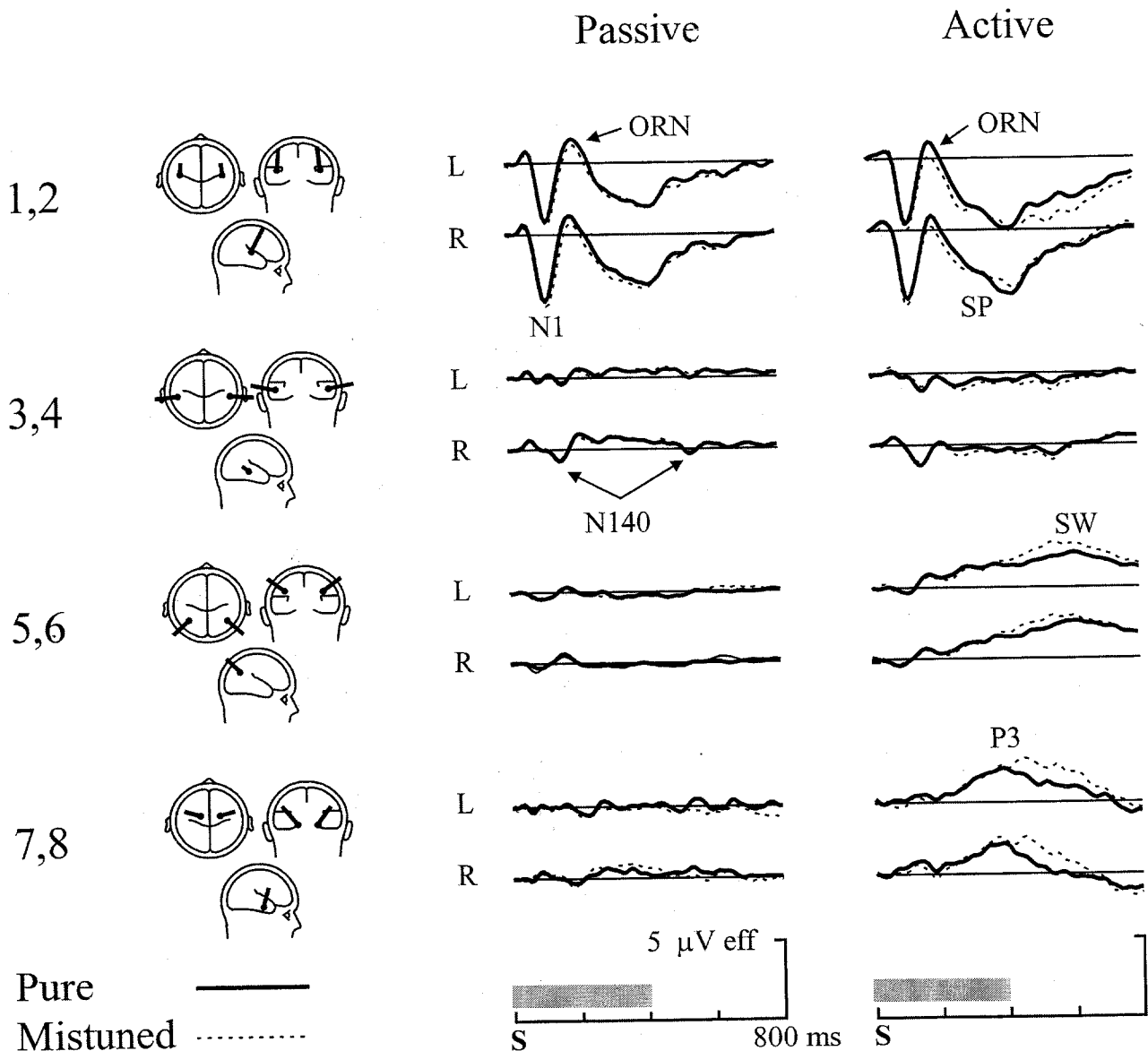


Figure 12. Source potentials for the Brain Electrical Source Analysis (BESA) solution. On the left are shown the locations and orientations of the sources. Each pair of sources was constrained to symmetry. The unit vectors of the BESA solution have been doubled in size from the original BESA figures to make the sources more legible. The source waveforms in the passive conditions are shown at center, with the left-hemisphere source (L) of each pair plotted above the right-hemisphere source (R). Only the first four sources show clear waveforms in the passive condition (N1, sustained potential [SP], and N140). In the active condition, shown at right, Sources 6 to 8 are also active, showing waveforms that correspond to the slow wave (SW) and P3 waves. Object-related negativity (ORN) is shown as a difference between the superimposed waveforms. This difference is larger in the active condition. *eff* = effective. Negativity is plotted downward.

Whereas the ORN was little affected by the listening condition, the P400 deflection was absent when the stimuli were not task relevant. We propose that the ORN indexes largely automatic processes because it is elicited even when participants are not required to attend and respond to the auditory stimuli. Furthermore, the ORN was present in a complex listening situation in which identifying the mistuned harmonic could not be based on attention to a particular frequency (Experiment 2). Together, these findings sug-

gest that detection of the mistuned harmonic as a separate tone must be preceded by a preliminary analysis of the frequency pattern.

The sequence of neural events underlying the perception of simultaneous auditory objects is consistent with Bregman's account of auditory scene analysis. Within this model, the acoustic wave is first decomposed into perceptual groups (i.e., objects) according to Gestalt principles. In the current study, partials that

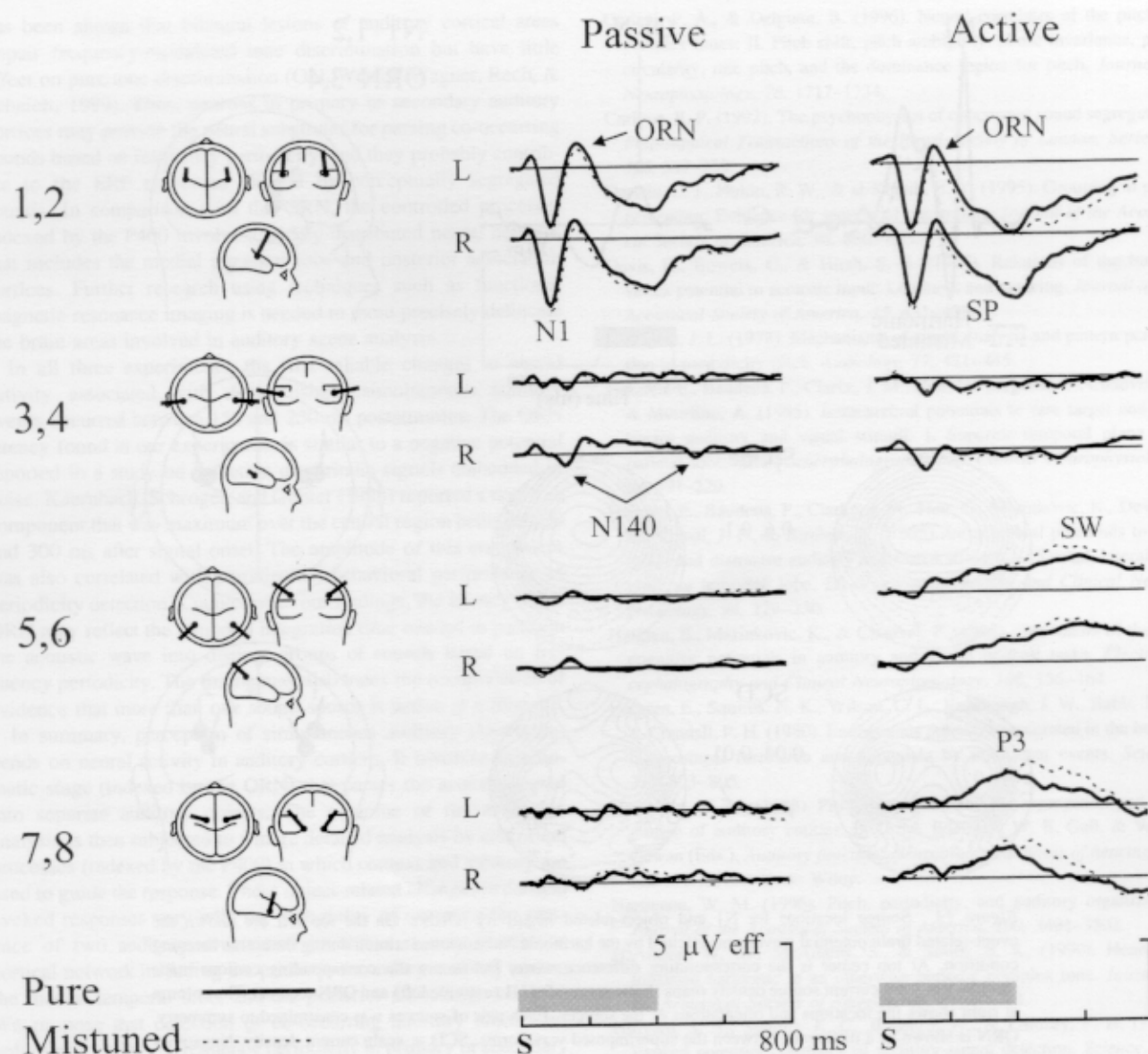


Figure 12. Source potentials for the Brain Electrical Source Analysis (BESA) solution. On the left are shown the locations and orientations of the sources. Each pair of sources was constrained to symmetry. The unit vectors of the BESA solution have been doubled in size from the original BESA figures to make the sources more legible. The source waveforms in the passive conditions are shown at center, with the left-hemisphere source (L) of each pair plotted above the right-hemisphere source (R). Only the first four sources show clear waveforms in the passive condition (N1, sustained potential [SP], and N140). In the active condition, shown at right, Sources 6 to 8 are also active, showing waveforms that correspond to the slow wave (SW) and P3 waves. Object-related negativity (ORN) is shown as a difference between the superimposed waveforms. This difference is larger in the active condition. *eff* = effective. Negativity is plotted downward.

Whereas the ORN was little affected by the listening condition, the P400 deflection was absent when the stimuli were not task relevant. We propose that the ORN indexes largely automatic processes because it is elicited even when participants are not required to attend and respond to the auditory stimuli. Furthermore, the ORN was present in a complex listening situation in which identifying the mistuned harmonic could not be based on attention to a particular frequency (Experiment 2). Together, these findings sug-

gest that detection of the mistuned harmonic as a separate tone must be preceded by a preliminary analysis of the frequency pattern.

The sequence of neural events underlying the perception of simultaneous auditory objects is consistent with Bregman's account of auditory scene analysis. Within this model, the acoustic wave is first decomposed into perceptual groups (i.e., objects) according to Gestalt principles. In the current study, partials that

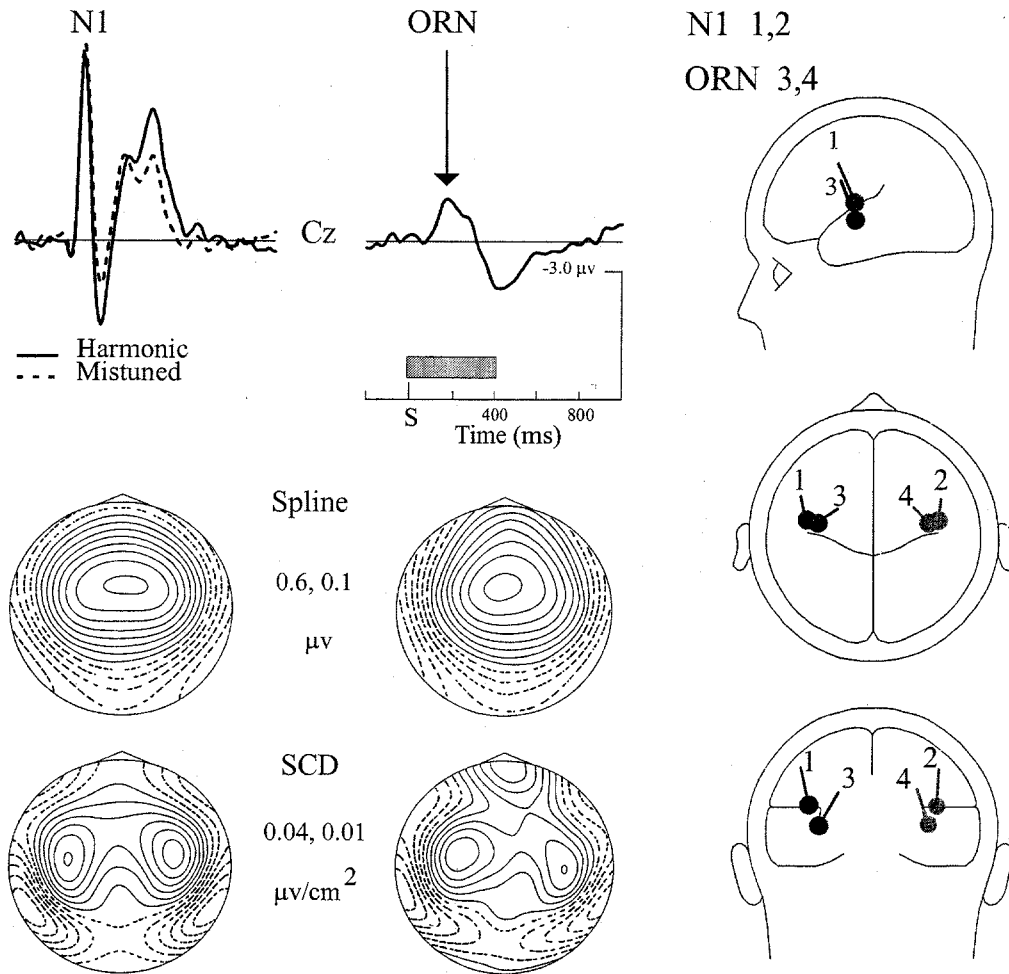


Figure 13. Source locations for N1 and object-related negativity (ORN). On the top left are shown the event-related brain potential waveforms elicited by the harmonic and mistuned stimuli during the active listening condition. At top center is the corresponding difference wave. Below are the corresponding contour maps (middle row) and current source density maps (bottom row) for N1 (extreme left) and ORN (center). The column at right shows the locations and orientations of the sources. Each pair of sources was constrained to symmetry. ORN is shown as a difference between the superimposed waveforms. SCD = scalp current density. Negativity is plotted upward.

were harmonically related could be grouped together into one entity, whereas the one that was sufficiently mistuned stood out of the complex as a separate object. The ORN may reflect an automatic mismatch process between a representation of the mistuned harmonic and the one expected on the basis of the harmonic template derived from the other harmonics within the stimulus. The purpose of the mismatch process would be to signal to higher auditory centers that more than one auditory object might be simultaneously present in the environment. This implies a pattern-matching process that occurs after the harmonics are grouped into distinct objects. This pattern matching would be more efficient when attention could be allocated to the mistuned harmonic, as evidenced by the attention effects on the ORN in Experiments 1 and 3. The output of the matching process may then be subjected to a more detailed analysis from the controlled (schema-driven) process. However, it appears that the output of the matching process can be overwritten by the information provided by the

sequential context, as shown by the effects of stimulus probability on both perceptual judgment and the P400 response.

The electrophysiological results also suggest that distinguishing simultaneous auditory objects involves a widely distributed neural network that includes the auditory cortex, the medial temporal lobe, and posterior association auditory cortices. The scalp distribution of the ORN is consistent with generators in the medial planum temporale. This suggests that the auditory cortex plays an important role in parsing and representing multiple auditory objects.

The segregation of co-occurring auditory events may depend on auditory neurons that are sensitive to frequency periodicity. Evidence from single-cell recordings in nonhuman primates (for a review, see Rauschecker, 1997) and neuromagnetic recordings in humans (Langner, Sams, Heil, & Schulze, 1997) suggests that frequency periodicity is represented within the primary or secondary auditory cortices, or both (Langner et al., 1997). In animals, it

has been shown that bilateral lesions of auditory cortical areas impair frequency-modulated tone discrimination but have little effect on pure tone discrimination (Ohl, Wetzels, Wagner, Rech, & Scheich, 1999). Thus, neurons in primary or secondary auditory cortices may provide the neural substrates for parsing co-occurring sounds based on frequency periodicity, and they probably contribute to the ERP responses elicited by perceptually segregated sounds. In comparison with the ORN, the controlled processes indexed by the P400 involve a widely distributed neural network that includes the medial temporal lobe and posterior association cortices. Further research using techniques such as functional magnetic resonance imaging is needed to more precisely delineate the brain areas involved in auditory scene analysis.

In all three experiments, the first reliable changes in neural activity associated with distinguishing simultaneous auditory events occurred between 150 and 250 ms poststimulus. The ORN latency found in our experiments is similar to a negative potential reported in a study on detection of periodic signals embedded in noise. Kaernbach, Schroger, and Gunter (1998) reported a negative component that was maximum over the central region between 200 and 300 ms after signal onset. The amplitude of this component was also correlated with participants' behavioral performance in periodicity detection. Coupled with our findings, the latency of the ORN may reflect the temporal integration time needed to partition the acoustic wave into distinct groups of sounds based on frequency periodicity. The timing may also index the accumulation of evidence that more than one sound source is active at a time.

In summary, perception of simultaneous auditory objects depends on neural activity in auditory cortices. It involves an automatic stage (indexed by the ORN) that parses the acoustic signal into separate auditory objects. The outcome of the automatic analysis is then subjected to a more detailed analysis by controlled processes (indexed by the P400) in which context and memory are used to guide the response. These object-related changes in cortical evoked responses vary with the probability of reporting the presence of two auditory events and involve a widely distributed cortical network including the primary-secondary auditory cortex, the medial temporal lobe, and the posterior association cortices. We propose that detection of co-occurring auditory objects involves processing of frequency periodicity in primary or secondary auditory cortices, which then provides the basis for building a perceptual representation of multiple objects or images in an acoustic scene.

References

- Alain, C., Richer, F., Achim, A., & Saint Hilaire, J. M. (1989). Human intracerebral potentials associated with target, novel, and omitted auditory stimuli. *Brain Topography*, *1*, 237-245.
- Beavios, M. W., & Meddis, R. (1991). A computer model of auditory stream segregation. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *43A*, 517-541.
- Berg, P., & Scherg, M. (1994). A multiple source approach to the correction of eye artifacts. *Electroencephalography and Clinical Neurophysiology*, *90*, 229-241.
- Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sounds*. London: MIT Press.
- Brown, G. J., & Wang, D. (1997). Modelling the perceptual segregation of double vowels with a network of neural oscillators. *Neural Network*, *10*, 1547-1558.
- Cariani, P. A., & Delgutte, B. (1996). Neural correlates of the pitch of complex tones: II. Pitch shift, pitch ambiguity, phase invariance, pitch circularity, rate pitch, and the dominance region for pitch. *Journal of Neurophysiology*, *76*, 1717-1734.
- Carlyon, R. P. (1992). The psychophysics of concurrent sound segregation. *Philosophical Transactions of the Royal Society of London, Series B*, *336*, 347-355.
- Darwin, C. J., Hukin, R. W., & al-Khatib, B. Y. (1995). Grouping in pitch perception: Evidence for sequential constraints. *Journal of the Acoustical Society of America*, *98*, 880-885.
- Davis, H., Bowers, C., & Hirsh, S. K. (1967). Relations of the human vertex potential to acoustic input: Loudness and masking. *Journal of the Acoustical Society of America*, *43*, 431-438.
- Goldstein, J. L. (1978). Mechanisms of signal analysis and pattern perception in periodicity pitch. *Audiology*, *17*, 421-445.
- Halgren, E., Baudena, P., Clarke, J. M., Heit, G., Liegeois, C., Chauvel, P., & Musolino, A. (1995). Intracerebral potentials to rare target and distractor auditory and visual stimuli: I. Superior temporal plane and parietal lobe. *Electroencephalography and Clinical Neurophysiology*, *94*, 191-220.
- Halgren, E., Baudena, P., Clarke, J. M., Heit, G., Marinkovic, K., Devaux, B., Vignal, J. P., & Biraben, A. (1995). Intracerebral potentials to rare target and distractor auditory and visual stimuli: II. Medial, lateral and posterior temporal lobe. *Electroencephalography and Clinical Neurophysiology*, *94*, 229-250.
- Halgren, E., Marinkovic, K., & Chauvel, P. (1998). Generators of the late cognitive potentials in auditory and visual oddball tasks. *Electroencephalography and Clinical Neurophysiology*, *106*, 156-164.
- Halgren, E., Squires, N. K., Wilson, C. L., Rohrbach, J. W., Babb, T. L., & Crandall, P. H. (1980). Endogenous potentials generated in the human hippocampal formation and amygdala by infrequent events. *Science*, *210*, 803-805.
- Hartmann, W. M. (1988). Pitch, perception and the segregation and integration of auditory entities. In G. M. Edelman, W. E. Gall, & W. M. Cowan (Eds.), *Auditory function: Neurobiological bases of hearing* (pp. 623-645). New York: Wiley.
- Hartmann, W. M. (1996). Pitch, periodicity, and auditory organization. *Journal of the Acoustical Society of America*, *100*, 3491-3502.
- Hartmann, W. M., McAdams, S., & Smith, B. K. (1990). Hearing a mistuned harmonic in an otherwise periodic complex tone. *Journal of the Acoustical Society of America*, *88*, 1712-1724.
- Hillyard, S. A., Squires, K. C., Bauer, J. W., & Lindsay, P. H. (1971). Evoked potential correlates of auditory signal detection. *Science*, *172*, 1357-1360.
- Hubner, R., & Hafer, E. R. (1995). Cuing mechanisms in auditory signal detection. *Perception & Psychophysics*, *57*, 197-202.
- Imaizumi, S., Mori, K., Kiritani, S., & Yumoto, M. (1997). Observation of neural processes of auditory scene analysis by magnetoencephalography. *Acta Oto-Laryngologica*, *532*(Suppl.), 106-108.
- Kaernbach, C., Schroger, E., & Gunter, T. C. (1998). Human event-related brain potentials to auditory periodic noise stimuli. *Neuroscience Letters*, *242*, 17-20.
- Knight, R. T., Hillyard, S. A., Woods, D. L., & Neville, H. J. (1980). The effects of frontal and temporal-parietal lesions on the auditory evoked potential in man. *Electroencephalography and Clinical Neurophysiology*, *50*, 112-124.
- Koffka, K. (1935). *Principles of Gestalt psychology*. New York: Harcourt, Brace, & World.
- Langner, G., Sams, M., Heil, P., & Schulze, H. (1997). Frequency and periodicity are represented in orthogonal maps in the human auditory cortex: Evidence from magnetoencephalography. *Journal of Comparative Physiology*, *181*, 665-676.
- Lee, J., & Green, D. M. (1994). Detection of a mistuned component in a

- harmonic complex. *Journal of the Acoustical Society of America*, 96, 716–725.
- Lin, J. Y., & Hartmann, W. M. (1998). The pitch of a mistuned harmonic: Evidence for a template model. *Journal of the Acoustical Society of America*, 103, 2608–2617.
- Lins, O. G., Picton, T. W., Berg, P., & Scherg, M. (1993a). Ocular artifacts in EEG and event-related potentials: I. Scalp topography. *Brain Topography*, 6, 51–63.
- Lins, O. G., Picton, T. W., Berg, P., & Scherg, M. (1993b). Ocular artifacts in recording EEGs and event-related potentials: II. Source dipoles and source components. *Brain Topography*, 6, 65–78.
- Martin, B. A., Sigal, A., Kurtzberg, D., & Stapells, D. R. (1997). The effects of decreased audibility produced by high-pass noise masking on cortical event-related potentials to speech sounds /ba/ and /da. *Journal of the Acoustical Society of America*, 101, 1585–1599.
- McCabe, S. L., & Denham, M. J. (1997). A model of auditory streaming. *Journal of the Acoustical Society of America*, 101, 1611–1621.
- McCarthy, G., & Wood, C. C. (1985). Scalp distributions of event-related potentials: An ambiguity associated with analysis of variance models. *Electroencephalography and Clinical Neurophysiology*, 62, 203–208.
- Mondor, T. A., & Bregman, A. S. (1994). Allocating attention to frequency regions. *Perception & Psychophysics*, 56, 268–276.
- Moore, B. C., Glasberg, B. R., & Peters, R. W. (1986). Thresholds for hearing mistuned partials as separate tones in harmonic complexes. *Journal of the Acoustical Society of America*, 80, 479–483.
- Näätänen, R. (1992). *Attention and brain function*. Hillsdale, NJ: Erlbaum.
- Näätänen, R., & Picton, T. W. (1987). The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. *Psychophysiology*, 24, 375–425.
- Ohl, F. W., Wetzel, W., Wagner, T., Rech, A., & Scheich, H. (1999). Bilateral ablation of auditory cortex in Mongolian gerbil affects discrimination of frequency modulated tones but not of pure tones. *Learning and Memory*, 6, 347–362.
- Palmer, A. R. (1990). The representation of the spectra and fundamental frequencies of steady-state single- and double-vowel sounds in the temporal discharge patterns of guinea pig cochlear-nerve fibers. *Journal of the Acoustical Society of America*, 88, 1412–1426.
- Parasuraman, R., & Beatty, J. (1980). Brain events underlying detection and recognition of weak sensory signals. *Science*, 210, 80–83.
- Parasuraman, R., Richer, F., & Beatty, J. (1982). Detection and recognition: Concurrent processes in perception. *Perception & Psychophysics*, 31, 1–12.
- Pichora-Fuller, M. K., Schneider, B. A., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *Journal of the Acoustical Society of America*, 97, 593–608.
- Picton, T. W., Alain, C., Otten, L., Ritter, W., & Achim, A. (2000). Mismatch negativity: Different water in the same river. *Audiology and Neuro-Otology*, 5, 111–139.
- Rauschecker, J. P. (1997). Processing of complex sounds in the auditory cortex of cat, monkey, and man. *Acta Oto-Laryngologica*, 532(Suppl.), 34–38.
- Richer, F., Alain, C., Achim, A., Bouvier, G., & Saint Hilaire, J. M. (1989). Intracerebral amplitude distributions of the auditory evoked potential. *Electroencephalography and Clinical Neurophysiology*, 74, 202–208.
- Roberts, B., & Brunstrom, J. M. (1998). Perceptual segregation and pitch shifts of mistuned components in harmonic complexes and in regular inharmonic complexes. *Journal of the Acoustical Society of America*, 104, 2326–2338.
- Scherg, M. (1989). *Fundamentals of dipole source analysis*. Basel, Switzerland: Karger.
- Scherg, M., Vajsar, J., & Picton, T. W. (1989). A source analysis of the late human auditory evoked potentials. *Journal of Cognitive Neuroscience*, 1, 326–355.
- Steinschneider, M., Arezzo, J., & Vaughan, H. G. Jr. (1980). Phase-locked cortical responses to a human speech sound and low-frequency tones in the monkey. *Brain Research*, 198, 75–84.
- Steinschneider, M., Reser, D. H., Fishman, Y. I., Schroeder, C. E., & Arezzo, J. C. (1998). Click train encoding in primary auditory cortex of the awake monkey: Evidence for two mechanisms subserving pitch perception. *Journal of the Acoustical Society of America*, 104, 2935–2955.
- von der Malsburg, C., & Schneider, W. (1986). A neural cocktail-party processor. *Biological Cybernetics*, 54, 29–40.
- Woods, D. L. (1995). The component structure of the N1 wave of the human auditory evoked potential. *Electroencephalography and Clinical Neurophysiology*, 44(Suppl.), 102–109.
- Woods, D. L., Clayworth, C. C., Knight, R. T., Simpson, G. V., & Naeser, M. A. (1987). Generators of middle- and long-latency auditory evoked potentials: Implications from studies of patients with bitemporal lesions. *Electroencephalography and Clinical Neurophysiology*, 68, 132–148.
- Woods, D. L., Knight, R. T., & Scabini, D. (1993). Anatomical substrates of auditory selective attention: Behavioral and electrophysiological effects of posterior association cortex lesions. *Brain Research: Cognitive Brain Research*, 1, 227–240.

Received February 29, 2000
 Revision received July 19, 2000
 Accepted December 11, 2000 ■