

# Consonance and Dissonance of Musical Chords: Neural Correlates in Auditory Cortex of Monkeys and Humans

YONATAN I. FISHMAN,<sup>1</sup> IGOR O. VOLKOV,<sup>2</sup> M. DANIEL NOH,<sup>2</sup> P. CHARLES GARELL,<sup>2</sup> HANS BAKKEN,<sup>2</sup> JOSEPH C. AREZZO,<sup>1</sup> MATTHEW A. HOWARD,<sup>2</sup> AND MITCHELL STEINSCHNEIDER<sup>1</sup>

<sup>1</sup>Department of Neuroscience, Albert Einstein College of Medicine, Bronx, New York 10461; and <sup>2</sup>Department of Surgery, Division of Neurosurgery, University of Iowa College of Medicine, Iowa City, Iowa 52242

Received 4 April 2001; accepted in final form 27 August 2001

**Fishman, Yonatan I., Igor O. Volkov, M. Daniel Noh, P. Charles Garell, Hans Bakken, Joseph C. Arezzo, Matthew A. Howard, and Mitchell Steinschneider.** Consonance and dissonance of musical chords: neural correlates in auditory cortex of monkeys and humans. *J Neurophysiol* 86: 2761–2788, 2001. Some musical chords sound pleasant, or consonant, while others sound unpleasant, or dissonant. Helmholtz's psychoacoustic theory of consonance and dissonance attributes the perception of dissonance to the sensation of "beats" and "roughness" caused by interactions in the auditory periphery between adjacent partials of complex tones comprising a musical chord. Conversely, consonance is characterized by the relative absence of beats and roughness. Physiological studies in monkeys suggest that roughness may be represented in primary auditory cortex (A1) by oscillatory neuronal ensemble responses phase-locked to the amplitude-modulated temporal envelope of complex sounds. However, it remains unknown whether phase-locked responses also underlie the representation of dissonance in auditory cortex. In the present study, responses evoked by musical chords with varying degrees of consonance and dissonance were recorded in A1 of awake macaques and evaluated using auditory-evoked potential (AEP), multiunit activity (MUA), and current-source density (CSD) techniques. In parallel studies, intracranial AEPs evoked by the same musical chords were recorded directly from the auditory cortex of two human subjects undergoing surgical evaluation for medically intractable epilepsy. Chords were composed of two simultaneous harmonic complex tones. The magnitude of oscillatory phase-locked activity in A1 of the monkey correlates with the perceived dissonance of the musical chords. Responses evoked by dissonant chords, such as minor and major seconds, display oscillations phase-locked to the predicted difference frequencies, whereas responses evoked by consonant chords, such as octaves and perfect fifths, display little or no phase-locked activity. AEPs recorded in Heschl's gyrus display strikingly similar oscillatory patterns to those observed in monkey A1, with dissonant chords eliciting greater phase-locked activity than consonant chords. In contrast to recordings in Heschl's gyrus, AEPs recorded in the planum temporale do not display significant phase-locked activity, suggesting functional differentiation of auditory cortical regions in humans. These findings support the relevance of synchronous phase-locked neural ensemble activity in A1 for the physiological representation of sensory dissonance in humans and highlight the merits of complementary monkey/human studies in the investigation of neural substrates underlying auditory perception.

## INTRODUCTION

Despite the ubiquity and importance of music in human culture, our understanding of the physiological bases of music

perception is still in its infancy. A fundamental feature of music is harmony, which refers to characteristics of simultaneous note combinations or "vertical" musical structure (i.e., chords). It has been recognized since antiquity that certain chords sound more pleasant than others (Pythagoras, ca. 600 BC, in Apel 1972). Chords composed of tones related to each other by simple (small-integer) frequency ratios, e.g., octave (2:1) and perfect fifth (3:2), are typically judged to be harmonious, smooth, or consonant, whereas chords composed of tones related to each other by complex (large-integer) ratios, e.g., minor second (256:243) and major seventh (243:128), are considered unpleasant, rough, or dissonant.

In considering consonance and dissonance, it is important to distinguish between musical consonance/dissonance, i.e., of a given sound evaluated within a musical context, and psychoacoustic, or sensory consonance/dissonance, i.e., of a given sound evaluated in isolation (see Plomp and Levelt 1965; Terhardt 1974b, 1977). Musical consonance/dissonance is culturally determined, as evidenced by its variation across cultures and historical periods (see Apel 1972; Burns and Ward 1982). In contrast, judgments of sensory consonance/dissonance are culturally invariant and largely independent of musical training (Butler and Daston 1968). Moreover, rodents, birds, monkeys, and human infants discriminate isolated musical chords on the basis of sensory consonance and dissonance similarly to expert human listeners and experienced musicians (Fannin and Braud 1971; Hulse et al. 1995; Izumi 2000; Schellenberg and Trainor 1996; Zentner and Kagan 1996). These findings indicate that sensory consonance/dissonance is likely shaped by relatively basic auditory processing mechanisms that are not music specific and that can be studied in experimental animals.

Several psychoacoustic theories have been proposed to explain why musical intervals characterized by simple frequency ratios sound more consonant than intervals characterized by complex frequency ratios (see Plomp and Levelt 1965 for review). The most prominent of these theories, first promoted by Helmholtz (1954), states that dissonance is related to the sensation of "beats" and "roughness." These perceptual phenomena occur when two or more simultaneous components of a complex sound are separated from one another in frequency by less than the width of an auditory filter or "critical bandwidth" (10–20% of center frequency) (Zwicker et al. 1957)

Address for reprint requests: Y. I. Fishman, Kennedy Building, Rm. 322, Albert Einstein College of Medicine, 1300 Morris Park Ave., Bronx, NY 10461 (E-mail: yfishman@aecom.yu.edu).

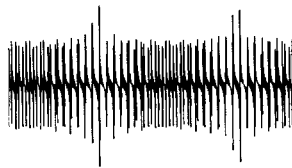
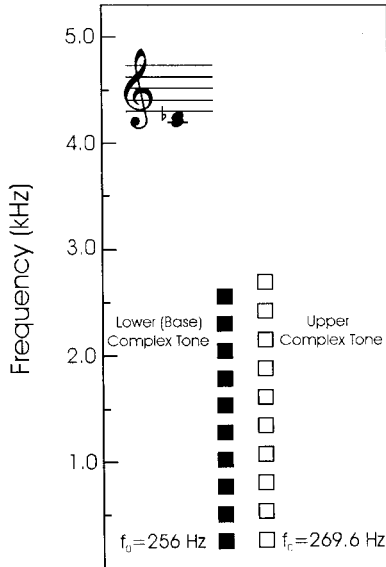
The costs of publication of this article were defrayed in part by the payment of page charges. The article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

# MUSICAL INTERVAL STIMULI

Base Tone: 256 Hz (Middle C)

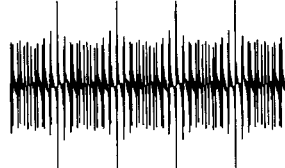
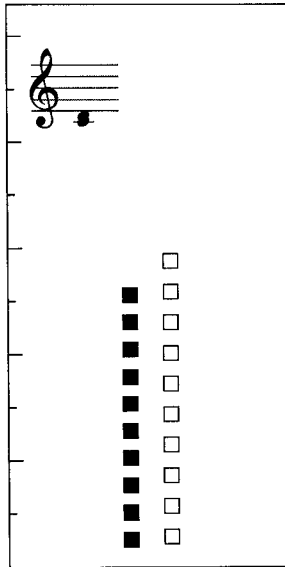
## Minor 2nd

Interval Ratio  
256:243



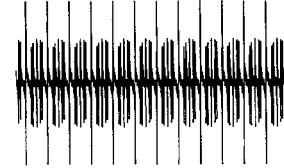
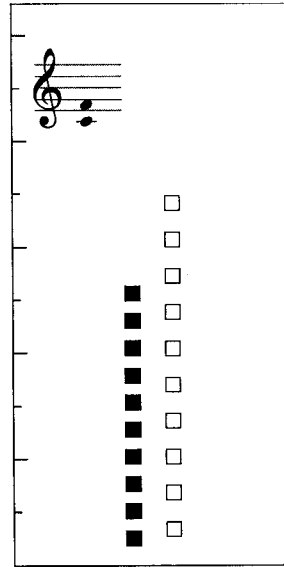
## Major 2nd

Interval Ratio  
9:8



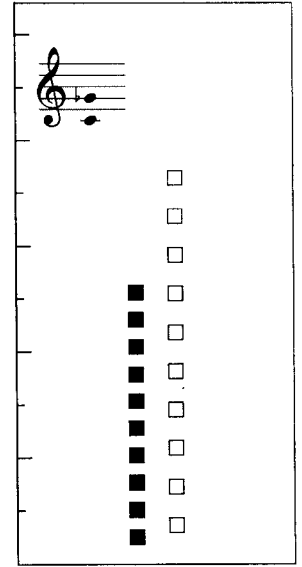
## Perfect 4th

Interval Ratio  
4:3



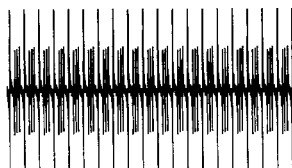
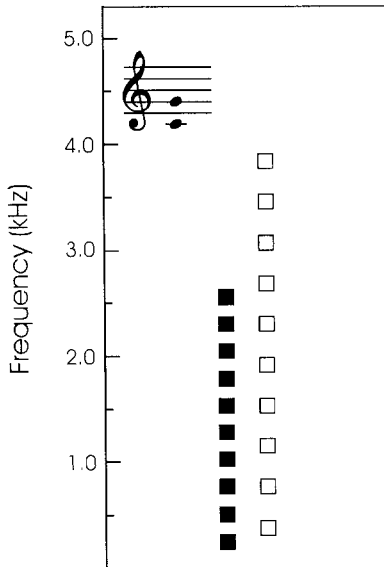
## Augmented 4th

Interval Ratio  
729:512



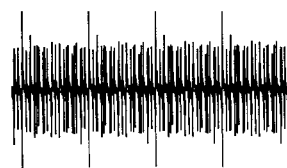
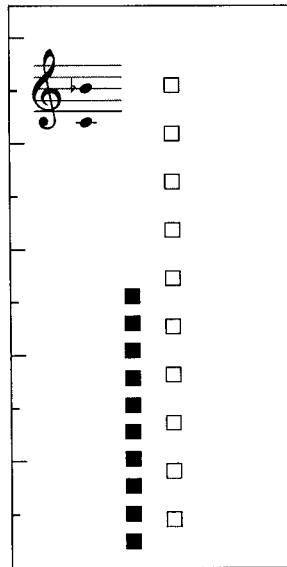
## Perfect 5th

Interval Ratio  
3:2



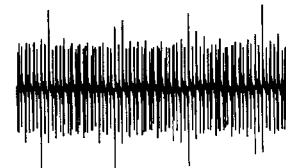
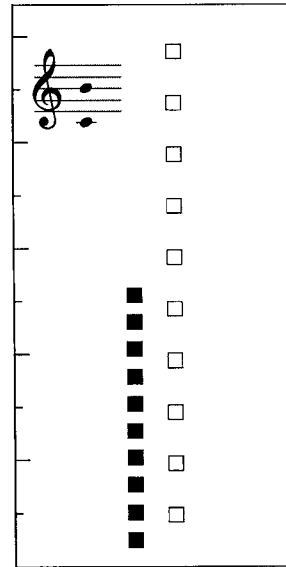
## Minor 7th

Interval Ratio  
16:9



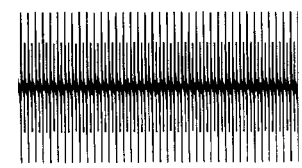
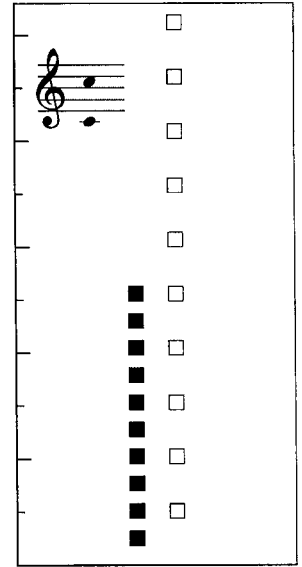
## Major 7th

Interval Ratio  
243:128



## Octave

2:1



and are hence unresolved by the auditory system. Unresolved frequency components interact in the auditory periphery, producing fluctuations in the amplitude of their composite waveform envelope that are perceived as beats (fluctuations below 20 Hz) or roughness (fluctuations from 20 to 250 Hz) (Kameoka and Kuriyagawa 1969a,b; Plomp and Levelt 1965; Plomp and Steeneken 1968; Terhardt 1968a,b, 1974a,b, 1978). The rate of these amplitude fluctuations equals the difference in frequency between the components. The disappearance of roughness for stimuli with amplitude fluctuation rates exceeding  $\sim 250$  Hz is thought to be due to the low-pass characteristic of the auditory nervous system (Plomp and Steeneken 1968; Terhardt 1974a, 1978).

The beats/roughness theory is impressive in its ability to predict the perceived dissonance of musical intervals on the basis of a relatively low-level psychoacoustic phenomenon. For intervals composed of harmonic complex tones, as produced by most musical instruments, dissonance depends on the ratio of the fundamental frequencies ( $f_0$ s) of the tones: dissonance is maximal when the  $f_0$ s of the complex tones form large-integer ratios and minimal when they form small-integer ratios (Kameoka and Kuriyagawa 1969b; Plomp and Levelt 1965). This pattern arises because chords composed of complex tones forming large-integer  $f_0$  ratios have fewer harmonics in common and more harmonics lying within the same critical band than chords composed of complex tones forming small-integer  $f_0$  ratios. Of these unresolved pairs of harmonics, the number with difference frequencies below 250 Hz is greater for intervals characterized by large-integer  $f_0$  ratios than for intervals characterized by small-integer  $f_0$  ratios. The summation of roughness contributed by each unresolved pair of frequencies separated by  $>20$  Hz and by  $<250$  Hz determines the overall perceived dissonance of musical intervals composed of complex tones (Kameoka and Kuriyagawa 1969b; Plomp and Levelt 1965; Terhardt 1974a, 1978). Consequently, musical intervals with large-integer  $f_0$  ratios produce more roughness and therefore more dissonance.

The neurophysiological basis of sensory consonance/dissonance perception is largely unknown. Bilateral lesions of auditory cortical areas in humans and animals are associated with deficits in pitch perception (Whitfield 1980; Zatorre 1988) and a range of music perception impairments (e.g., Liegeois-Chauvel et al. 1998; Peretz et al. 1994), including aberrant consonance/dissonance perception (Peretz et al. 2001; Tramo et al. 1990). Several physiological studies have suggested that roughness may be represented in primary auditory cortex (A1) by neuronal responses phase-locked to the amplitude-modulated temporal envelope of complex sounds (Bieser and Muller-Preuss 1996; Schulze and Langner 1997; Steinschneider et al. 1998). This hypothesis is supported by the correlation found between the magnitude of neuronal ensemble phase-locking to the AM frequency (= difference frequency) of harmonic complex tones in A1 of the awake monkey and the degree of roughness perceived by human listeners. Specifically, phase-locking is maximal at stimulus modulation frequencies at

which roughness is maximal and dissipates at stimulus modulation frequencies at which roughness disappears (Fishman et al. 2000a). Given the involvement of A1 in music perception and assuming the validity of Helmholtz's beats/roughness theory of sensory dissonance, it follows that if the hypothesized mechanism underlying the physiological representation of roughness is correct, then the perceived dissonance of musical chords should correlate with the magnitude of A1 activity phase-locked to the difference frequencies. The present study tests this hypothesis by examining phase-locked neuronal ensemble activity evoked by musical chords with varying degrees of consonance and dissonance in A1 of the awake macaque monkey. Macaques share similarities in basic auditory cortical anatomy and physiology with humans (Galaburda and Pandya 1983; Galaburda and Sanides 1980; Steinschneider et al. 1994, 1999) and are able to discriminate musical chords on the basis of sensory consonance/dissonance (Izumi 2000), making them appropriate animal models for investigating neural representation of sensory consonance and dissonance in the central auditory system.

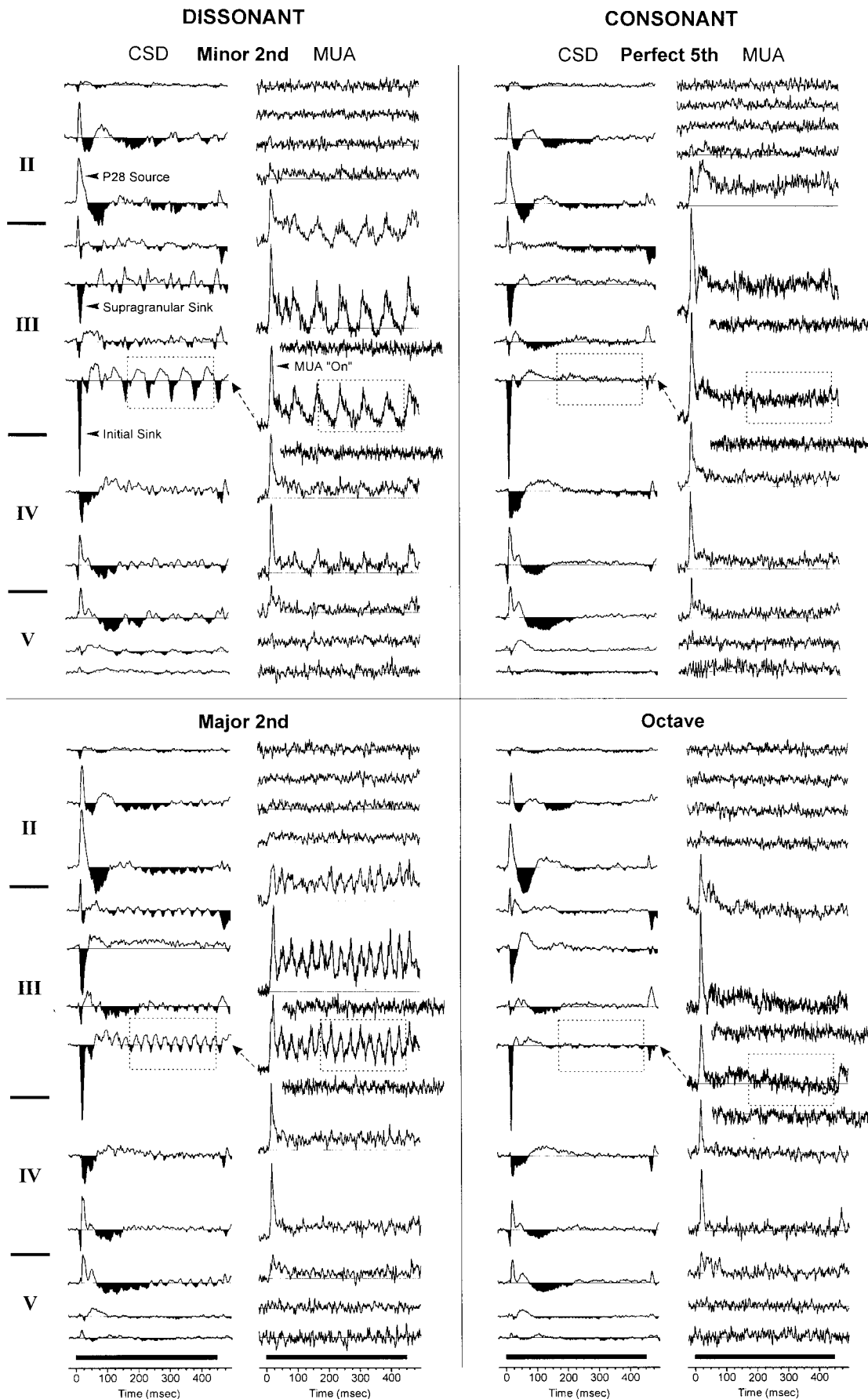
Correlation between patterns of cortical activity in an animal model and psychoacoustical features of consonance/dissonance perception leaves in question, however, whether these neural response patterns are applicable to the human brain. A stronger argument for the relevance of these physiological responses could be made if physiological findings similar to those obtained in the animal model are observed in human neural responses. Therefore, in parallel to the studies in monkeys, auditory-evoked potentials (AEPs) evoked by musical chords were also recorded directly from the auditory cortex of two patients undergoing surgical evaluation for medically intractable epilepsy. This cross-species approach has already been used to advantage in the study of auditory cortical representation of the voice onset time phonetic feature (Steinschneider et al. 1999) and offers several significant benefits. Clearly, it bolsters the relevance of the animal results by testing the suitability of the macaque as a model in which to examine neural correlates of higher perceptual processes. Furthermore, if a similarity between human and animal physiological response patterns can be demonstrated, the more refined sampling and analysis inherent in animal physiological studies can help to characterize the detailed mechanisms underlying the neural representation of the perceptual process under study.

## METHODS

### *Monkey surgery and electrophysiological recordings*

Three adult male monkeys (*Macaca fascicularis*) were studied using previously reported methods (Steinschneider et al. 1992, 1994, 1998). Animals were housed in our Association for Assessment and Accreditation of Laboratory Animal Care-accredited Animal Institute under daily supervision by veterinary staff. All experiments were conducted in accordance with institutional and federal guidelines governing the experimental use of primates. Briefly, using aseptic surgical techniques under general anesthesia (pentobarbital, initial and

FIG. 1. Waveforms and spectral content of the 8 musical interval stimuli presented in the study (Pythagorean tuning, only 150 ms displayed). Stimuli with base tones of 256 Hz are shown; all frequencies are doubled for the 512-Hz base tone stimuli. Each stimulus is composed of 2 simultaneous harmonic complex tones. Each complex tone contains the fundamental frequency ( $f_0$ ) and the 2nd–10th harmonic, all at equal amplitude. ■ and □, frequencies comprising the lower base tone and upper interval-defining tone, respectively. Corresponding musical notation is shown in the *top left corner* of each stimulus box. Ratio of the  $f_0$  of the higher tone to that of the lower tone is indicated *above* each stimulus box.



supplementary doses of 20 and 5 mg/kg iv, respectively), holes were drilled in the exposed skull to accommodate epidural matrices consisting of adjacent 18-gauge stainless steel tubes. Matrices were stereotaxically positioned to target A1 and were oriented at an angle of 30° from normal to approximate the anterior-posterior tilt of the superior temporal plane. This orientation guided electrode penetrations roughly perpendicular to the cortical surface, thereby fulfilling one of the major technical requirements of one-dimensional current-source density (CSD) analysis (Vaughan and Arezzo 1988). Matrices and Plexiglas bars used for painless head immobilization during the recording sessions were held in place by a pedestal of dental acrylic fixed to the skull by inverted screws keyed into the bone. Animals were given peri- and postoperative analgesic, antibiotic, and anti-inflammatory medications. Recordings began 2 weeks after surgery and were conducted in an electrically shielded, sound-attenuated chamber with the animals awake and comfortably restrained.

Intracortical recordings were obtained using linear-array multi-contact electrodes containing 14 recording contacts, evenly spaced at 150- $\mu$ m intervals (Barna et al. 1981). Individual contacts were constructed from 25- $\mu$ m-diameter stainless steel wires, each with an impedance of  $\sim$ 200 k $\Omega$ . An epidural stainless steel guide tube positioned over the occipital cortex served as a reference electrode. Field potentials were recorded using unity-gain headstage preamplifiers, and amplified 5,000 times by differential amplifiers with a frequency response down 3 dB at 3 Hz and 3 kHz. Signals were digitized at a sampling rate between 2 and 4 kHz (depending on the analysis time used) and averaged by computer (Neuroscan software and hardware, Neurosoft) to yield AEPs. To derive multiunit activity (MUA), signals were simultaneously high-pass filtered above 500 Hz, amplified an additional eight times, and full-wave rectified prior to digitization and averaging. MUA is a measure of the summed action potential activity of neuronal aggregates within a sphere of about 50–100  $\mu$ m in diameter surrounding each recording contact (Brosch et al. 1997; Vaughan and Arezzo 1988). For some electrode penetrations, raw data were stored on a 16-channel digital tape recorder (Model DT-1600, MicroData Instrument; sample rate: 6 kHz) for off-line analyses. Due to limitations of the acquisition computer, the sampling rates used were below the Nyquist frequency corresponding to the 3 kHz upper cutoff of the amplifiers. However, empirical testing revealed negligible signal distortion due to aliasing, as most of the spectral energy in the MUA lies below 1 kHz. Using shorter analysis windows and fewer channels, raw data re-digitized at 6 kHz, yielded nearly identical averaged waveforms as data sampled at the lower rate. Absence of aliasing was also confirmed by low-pass filtering the MUA at 800 Hz (96 dB/octave roll-off) following rectification and prior to digitization at 2 kHz, using digital filters (RP2 modules, Tucker Davis Technologies) acquired after the completion of this study. Differences between unfiltered and low-pass filtered MUA signals were negligible (see Fig. 2). To further confirm the validity of MUA measures, off-line multi-unit cluster analyses of unrectified high-pass filtered data were performed for some sites. Peristimulus time histograms (PSTHs) were constructed with a binwidth of 1 ms. Triggers for spike acquisition were set at 2.5 times the amplitude of the background "hash" of lower-amplitude, high-frequency activity.

One-dimensional CSD analyses characterized the laminar pattern of

net current sources and sinks within A1 generating the AEPs. CSD was calculated using a three-point algorithm that approximates the second spatial derivative of voltage recorded at each recording contact (Freeman and Nicholson 1975; Nicholson and Freeman 1975). Current sinks represent net inward transmembrane current flow associated with local depolarizing excitatory postsynaptic potentials or passive, circuit-completing current flow associated with hyperpolarizing potentials at adjacent sites. Current sources represent net outward transmembrane currents associated with active hyperpolarization or passive current return associated with adjacent depolarizing potentials. The corresponding MUA profile is used to help distinguish these possibilities: current sinks coincident with increases in MUA likely reflect depolarizing synaptic activity, whereas current sources associated with concurrent reductions in MUA from baseline levels likely reflect hyperpolarizing events rather than passive current return for adjacent synaptic depolarization.

Electrodes were manipulated with a microdrive and positioned using on-line examination of click-evoked potentials as a guide. Pure tone and chord stimuli were delivered when the electrode channels bracketed the inversion of early AEP components and the largest MUA, typically occurring during the first 50 ms within lamina IV (LIV) and lower lamina III (LLIII), was situated in the middle channels. Evoked responses to 75 presentations of the stimuli were averaged with an analysis window (including a 25-ms prestimulus baseline interval) of 300 ms for pure tones and 520 ms for musical chord stimuli.

#### Human electrophysiological recordings

Intracranial AEPs were recorded in one man (*subject 1*) and one woman (*subject 2*). Both subjects had medically intractable epilepsy, were right-handed, and required placement of multiple temporal lobe electrodes to determine the location of seizure onsets. Experimental procedures were approved by the University of Iowa Human Subjects Review Board and the National Institutes of Health. Informed consent was obtained from the subjects prior to their participation. Subjects underwent surgical implantation of intracranial electrodes (Radionics, Burlington, MA) to acquire diagnostic electroencephalographic (EEG) data required for planning subsequent surgical treatment. Subjects did not undergo any additional risk by participating in this study.

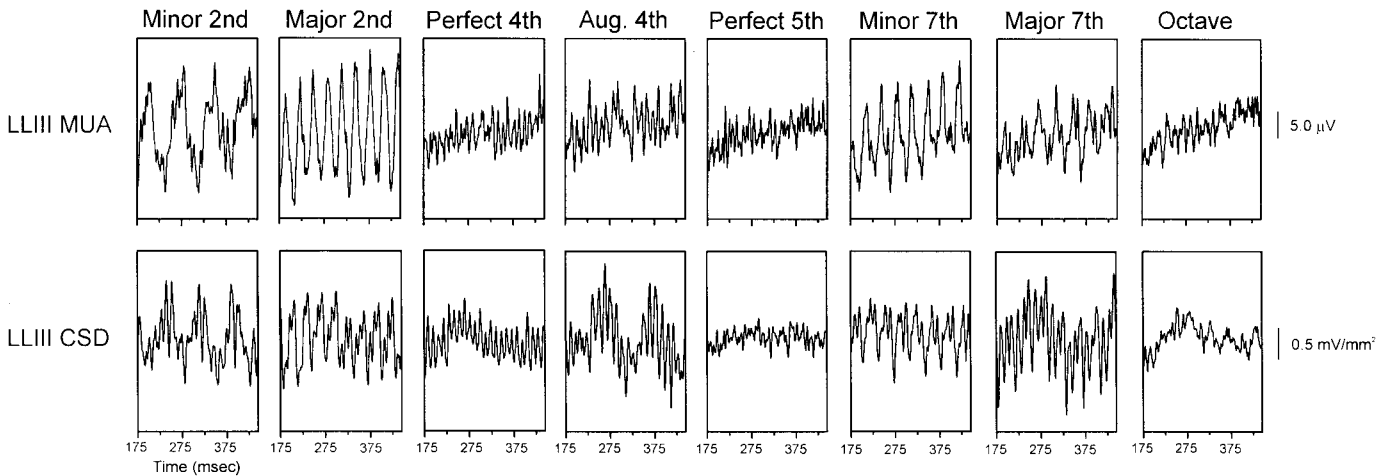
*Subject 2* had depth electrodes (Howard et al. 1996a,b) implanted in the right Heschl's gyrus and planum temporale. Data from this subject using different stimulus protocols have been reported (Steinschneider et al. 1999). Bipolar recordings at three locations were obtained from closely spaced recording contacts (impedance, 200 k $\Omega$ , 2.5–4.2 mm inter-contact distance) placed stereotaxically along the long axis of Heschl's gyrus. Spectral sensitivity of two of these sites, site 1 (the most posteromedial site) and site 3 (the most anterolateral site), was assessed via independent analysis of multiple unit responses. Maximal tone responses of units at sites 1 and 3 were  $2,125 \pm 252$  and  $736 \pm 91$  (SD) Hz, respectively, consistent with findings that higher frequencies are represented at more posteromedial locations in human A1 (Howard et al. 1996a; Steinschneider et al. 1999). *Subject 1* had three depth electrodes implanted in the right superior temporal gyrus: the first in Heschl's gyrus, the second in the planum temporale, and the

FIG. 2. Representative examples of current-source density (CSD) and multiunit activity (MUA) laminar response profiles evoked by the 2 most dissonant chords (*left*: minor and major 2nd) and by the 2 most consonant chords (*right*: perfect 5th and octave) presented in the study [best frequency (BF) = 1,600 Hz, base tone = 256 Hz]. Approximate laminar boundaries are shown on the *left* of the figure. Stimulus duration is represented by the black bar above the time axes. Consonant and dissonant stimuli evoke similar early response components reflecting initial cortical activation in lower lamina III (LLIII) and lamina IV (LIV; initial sink, MUA ON) and delayed activation in upper lamina III (ULIII; supragranular sink). Later activity differs between consonant and dissonant interval-evoked responses: dissonant stimuli evoke oscillatory activity phase-locked to the difference frequencies, whereas consonant stimuli evoke little or no oscillatory activity. The present study examines the degree of phase-locking in LLIII MUA, LLIII CSD, and ULIII CSD from 175 to 445 ms poststimulus onset (portion of the response enclosed by the dashed box) as a function of musical interval. Waveforms of LLIII MUA low-pass filtered at 800 Hz (96 dB/octave roll-off) prior to digitization are shown superimposed on waveforms of unfiltered MUA data to illustrate absence of significant signal aliasing, as demonstrated by the nearly flat difference waveforms shown below the superimposed waveforms.

**Base Tone: 256 Hz**

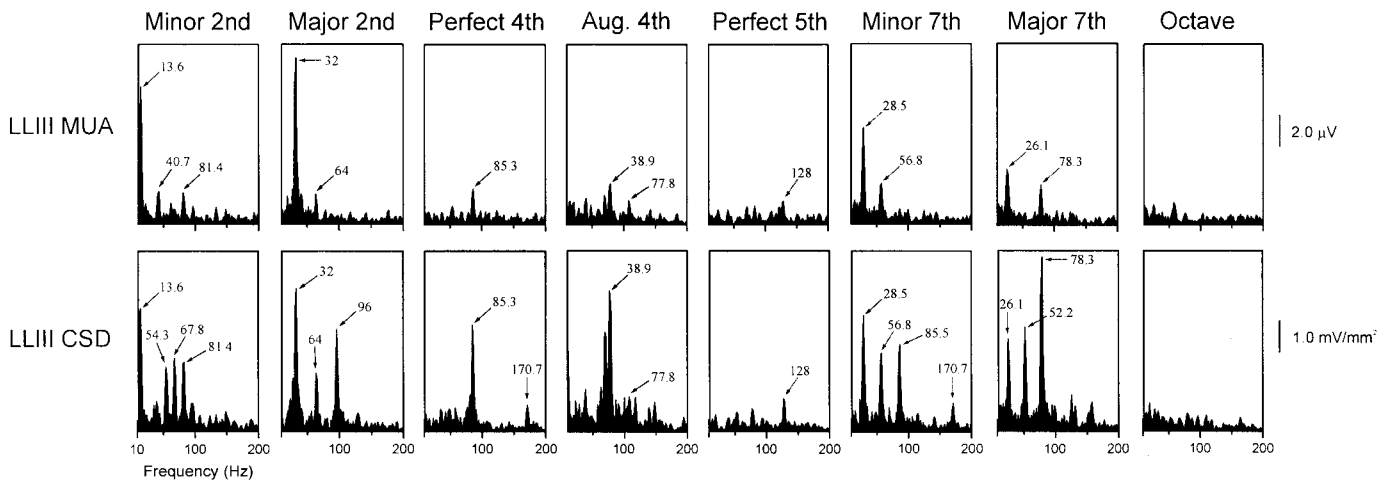
**A**

**Waveforms**

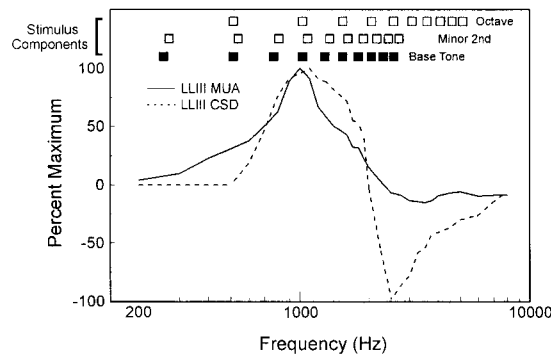


**B**

**Amplitude Spectra**



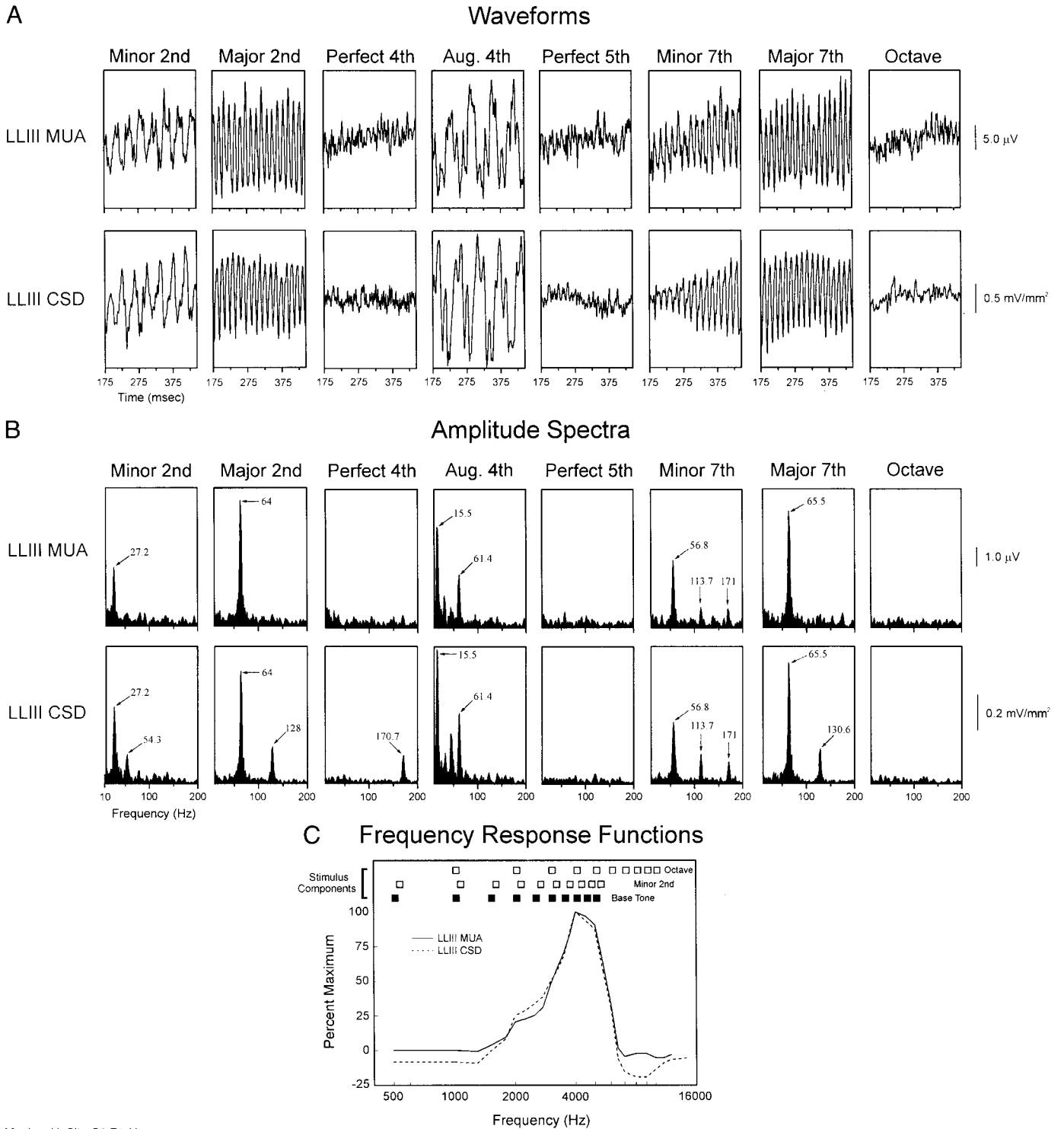
**C Frequency Response Functions**



Monkey H; Site C5 Rt. Hem.

FIG. 3. Representative waveforms (A) and corresponding amplitude spectra (B) of LLIII MUA and CSD (175–445 ms poststimulus onset) evoked by musical chords with base tones of 256 Hz. For clarity, only the frequency range from 10 to 200 Hz is displayed because no peaks in the amplitude spectra were evident above 200 Hz. C: frequency response functions (FRFs) based on peak amplitude of LLIII MUA and CSD within the first 50 ms poststimulus onset (BF = 1,000 Hz). Frequency components of the smallest interval (minor 2nd) and of the largest interval (octave) are schematically represented above the FRFs to illustrate overlap between stimulus spectra and the excitatory frequency response area of the neuronal ensemble. Dissonant intervals (e.g., minor and major 2nd) evoke oscillatory phase-locked responses, manifested as peaks in amplitude spectra. Arrows in the spectra indicate major peaks corresponding to predicted difference frequencies (values, in Hz, next to arrows). In contrast, the most consonant intervals (e.g., octave and perfect 5th) evoke little or no phase-locked activity, leading to comparatively flat amplitude spectra.

Base Tone: 512 Hz



Monkey H; Site C6 Rt. Hem.

FIG. 4. Representative waveforms (A) and corresponding amplitude spectra (B) of LLIII MUA and CSD responses (175–445 ms poststimulus onset) evoked by musical chords with base tones of 512 Hz recorded at a different site from that shown in Fig. 3. C: LLIII MUA and CSD FRFs (BF = 4,000 Hz). Same conventions as in Fig. 3. Dissonant intervals evoke oscillatory phase-locked responses manifested as peaks in amplitude spectra. In contrast, consonant intervals evoke little or no phase-locked activity, leading to comparatively flat amplitude spectra.

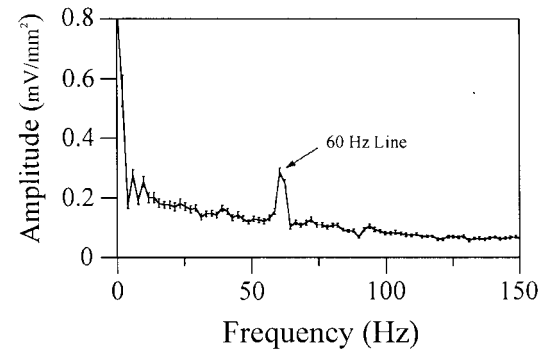
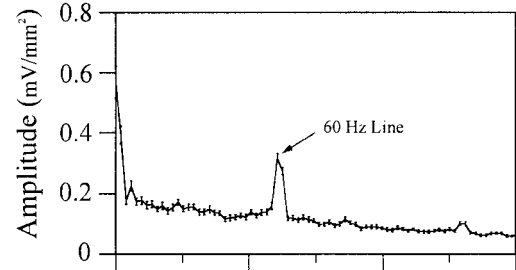
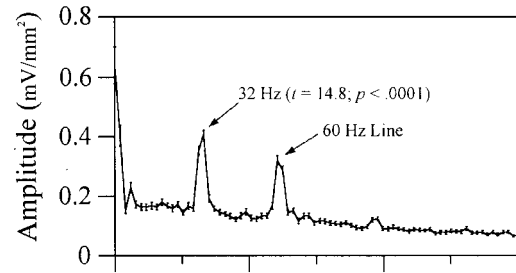
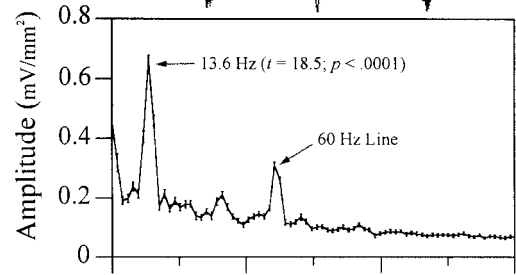
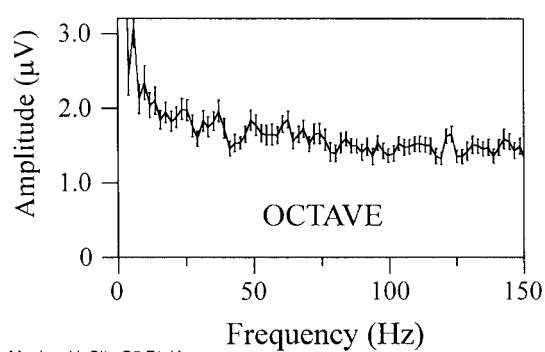
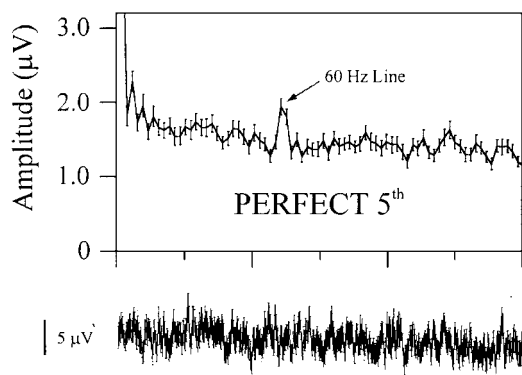
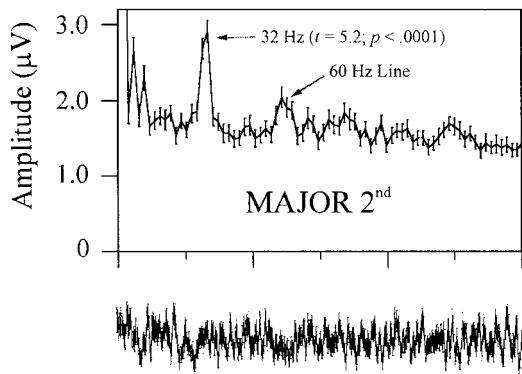
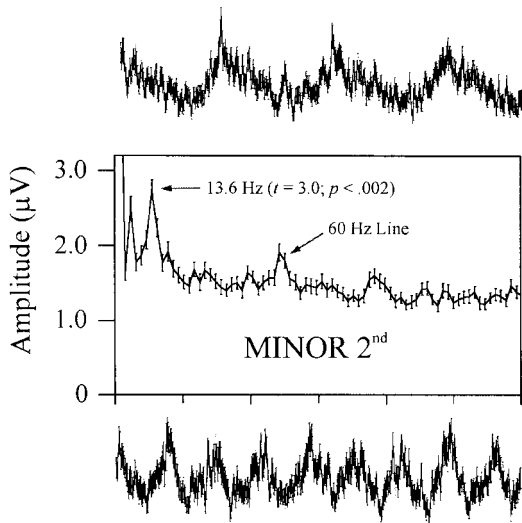
third in a more posterior location within the planum temporale. Click-evoked responses recorded at the location of the most posterior electrode were of low amplitude, and, consequently, musical chord-

evoked responses were not recorded at this electrode. Responses at the Heschl's gyrus and planum temporale electrodes were recorded from two higher-impedance (200 k $\Omega$ ) and one lower-impedance (30 k $\Omega$ )

Base Tone: 256 Hz

LLIII MUA

LLIII CSD





recording contacts (2.5–4.2 mm inter-contact distance). Spectral sensitivities of sites in *subject 1* were not determined. The reference electrode was a subdural electrode located on the ventral surface of the ipsilateral, anterior temporal lobe.

Recording sessions took place in a quiet room in the Epilepsy Monitoring Unit of the University of Iowa Hospitals and Clinics with the subjects lying comfortably in their hospital beds. Subjects were awake and alert throughout the recordings. For both subjects, sweeps exhibiting high-amplitude epileptic spikes at any time point within the analysis window were rejected by the acquisition computer or discarded following visual inspection of the data.

AEPs were recorded at a gain of 5,000 using headstage amplification followed by differential amplification (BAK Electronics). Field potentials were filtered (band-pass, 2–500 Hz; roll-off, 6 dB/octave), digitized (1.0- or 2.050-kHz sampling rate), and averaged, with an analysis window of 500 ms (including a 25-ms prestimulus baseline interval) in the case of *subject 2* and 1,000 ms (including a 325-ms prestimulus baseline interval) in the case of *subject 1*. Averages were generated from 50 to 75 stimulus presentations. Raw EEG and timing pulses were stored on a multi-channel tape recorder (Racal) for off-line analysis.

### Stimuli

**MONKEY RECORDINGS.** Frequency response functions (FRFs), based on pure tone responses, were used to characterize the frequency tuning of the cortical sites. Pure tones ranging from 0.2 to 17.0 kHz were generated and delivered at a sampling rate of 100 kHz by a PC-based system using SigGen and SigPlay (Tucker Davis Technologies). Pure tones were 175 ms in duration with 10-ms linear ramps. Stimulus onset asynchrony (SOA) for pure tone presentation was 658 ms. All stimuli were monaurally delivered at 60 dB SPL via a dynamic headphone to the ear contralateral to the recorded hemisphere. Sounds were introduced to the ear through a 3-in-long, 60-ml plastic tube attached to the headphone. Sound intensity was measured with a Bruel and Kjaer sound level meter (type 2236) positioned at the opening of the plastic tube. The frequency response of the headphone was flattened ( $\pm 3$  dB) from 0.2 to 17.0 kHz by a graphic equalizer (Rane).

Musical chords were synthesized by summation of appropriate pure tone components (all in sine phase) using Turbosynth sound-synthesizing software on a Macintosh computer, edited using SoundDesigner software, and presented in pseudorandom order using ProTools (Digidesign) or SigGen and SigPlay (Tucker Davis Technologies) software and hardware. Each chord was composed of two simultaneous harmonic complex tones, each containing the  $f_0$  and the second through the tenth harmonic (all of equal amplitude). The  $f_0$  of one of the complex tones defined the base tone (root) of the two-tone chord, while that of the second complex tone defined the musical interval. Intervals were presented in three different octave ranges (forming 3 stimulus sets), such that the  $f_0$  of the base tone was 128, 256, or 512 Hz, corresponding to C one octave below middle C, middle C, and C one octave above Middle C, respectively. Each stimulus set presented in a given electrode penetration was composed of eight different musical intervals with varying degrees of dissonance. Intervals were confined to one octave and were constructed according to the Pythagorean, or “pure fifth,” system of tuning (interval ratios obtained from Apel 1972). Spectral content and temporal waveforms of the

musical interval stimuli are shown in Fig. 1. The particular base tone used in a given electrode penetration was chosen so that at least one harmonic from each of the two complex tones comprising the chord overlapped the excitatory frequency response area of the sampled neuronal population. For some penetrations, more than one stimulus set was presented. Musical interval stimuli were 450 ms in duration, were gated on and off with 5-ms linear ramps, and were presented at a total intensity of 60 dB SPL with a SOA of 992 ms.

**HUMAN RECORDINGS.** In the case of *subject 1*, stimuli were delivered to the left ear (contralateral to the recorded hemisphere) by an insert earphone (Etymotic Research). In the case of *subject 2*, stimuli were delivered to the left ear by an external headphone (Koss, model K240DF) coupled to a 4-cm cushion. Stimuli were presented at a comfortable listening level (60–70 dB SPL). In the case of *subject 1*, musical interval stimuli were identical to those used in the monkey recordings and were presented in pseudorandom order with a SOA of 2,000 ms. Due to time constraints, only a subset of the chords presented in the monkey studies was presented in the human studies. In the case of *subject 2*, two-tone chords were generated using a keyboard synthesizer (Roland, model JV-35) in organ mode. Keyboard-generated sounds were edited and presented in the same manner as sounds created by addition of frequency components, except that their total duration was 375 ms. Spectral analyses of the organ sounds indicated the presence of multiple harmonic components (see Fig. 18). In contrast to the chords constructed from sine wave addition, keyboard generated chords were based on equal temperament tuning (the tuning system conventionally used in modern Western music), thereby allowing qualitative comparison between neural responses evoked by intervals derived from different systems of tuning (Pythagorean vs. equal temperament). Keyboard-generated chords were presented in pseudorandom order with a SOA of 658 ms. Subjects were informally asked to relate their impression of the musical chords following the acquisition of a block of electrophysiological responses (e.g., “Did the chord sound pleasant or unpleasant?”). Patients’ subjective evaluations of the chords were consistent with those reported in psychoacoustic studies on consonance and dissonance (Butler and Daston 1968; Kameoka and Kuriyagawa 1969b; Malmberg 1918).

### Monkey histology

At the end of the recording period, monkeys were deeply anesthetized with pentobarbital sodium and transcardially perfused with 10% buffered formalin. Tissue was sectioned (80  $\mu$ m thickness) and stained for acetylcholinesterase and Nissl substance to reconstruct the electrode tracks and to identify A1 according to previously published criteria (Hackett et al. 1998; Merzenich and Brugge 1973; Morel et al. 1993; Wallace et al. 1991a). Field R was demarcated from A1 by a reversal in the best frequency gradient (Merzenich and Brugge 1973; Morel et al. 1993). The earliest sink/source configuration was used to locate LIV (Steinschneider et al. 1992). Other laminar locations were then determined by their relationship to LIV and the measured widths of laminae within A1 for each electrode penetration histologically identified.

### Data analysis

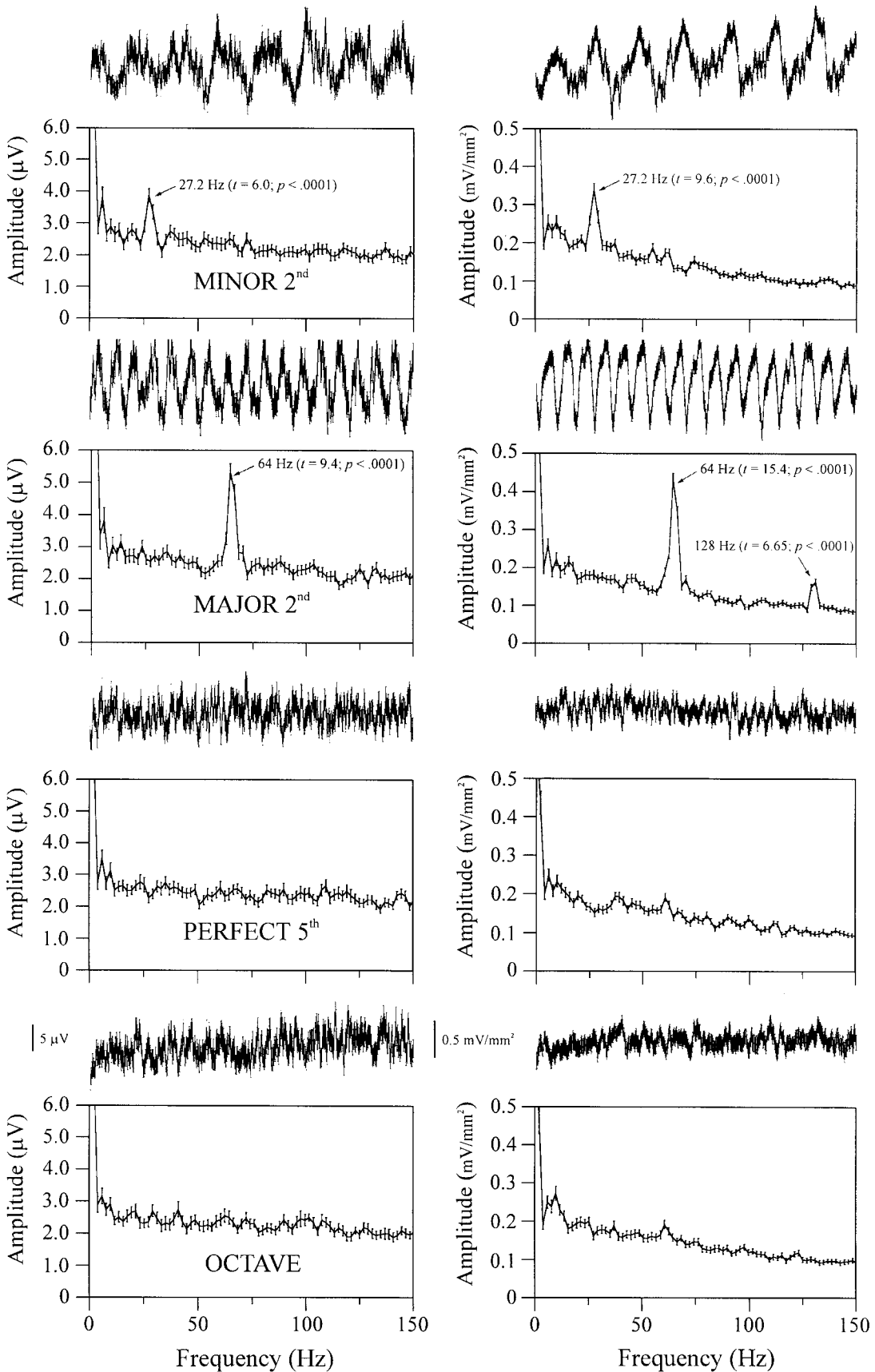
**MONKEY RECORDINGS.** The best frequency (BF) of the cortical site sampled in a given electrode penetration was defined as the pure tone

FIG. 5. Mean ( $\pm$ SE;  $n = 70$ ) LLIII MUA and CSD waveforms and corresponding mean ( $\pm$ SE) amplitude spectra of responses (175- to 445-ms poststimulus onset) evoked by the two most dissonant chords (minor and major 2nd) and by the two most consonant chords (octave and perfect 5th) at the same site as that shown in Fig. 2 (base tone = 256 Hz). Major peaks in the mean spectra of responses evoked by the minor 2nd and the major 2nd occur at predicted difference frequencies ( $\leftarrow$ ). Means at the peaks are significantly larger than means at corresponding frequencies in the spectrum of octave-evoked responses (one-tailed  $t$ -test;  $t$  and  $P$  values are shown in the figures). In contrast, the mean spectrum of perfect 5th-evoked responses (above 10 Hz) is not significantly different from that of octave-evoked responses ( $P > 0.05$ ). No significant differences between mean spectra are observed at frequencies  $>150$  Hz. A peak at 60 Hz corresponds to line noise, which disappears with time domain averaging.

Base Tone: 512 Hz

LLIII MUA

LLIII CSD



frequency eliciting the largest peak amplitude MUA within LLIII during the first 50 ms following stimulus onset. Determination of BF was generally based on MUA averaged across two to three LLIII electrode contacts. Use of peak amplitude initial MUA as a measure of BF yielded the expected anterolateral to posteromedial topographic gradients of low- to high-frequency representation in all animals (Merzenich and Brugge 1973; Morel et al. 1993; Recanzone et al. 2000).

Neuronal phase-locking to the difference frequencies relevant for sensory dissonance was quantified by spectral analysis of averaged responses using a fast Fourier transform (FFT) algorithm (ProStat, Poly Software International; Matlab, Mathworks). Spectral analysis has been used by the present authors and by other investigators to quantify stimulus phase-locked and non-phase-locked (e.g., gamma-band) oscillatory activity in auditory cortex and other cortical areas (e.g., Brosch et al. 1997; Crone et al. 1998; Eckhorn et al. 1993; Fishman et al. 2000a; Gray and Singer 1989; Schreiner and Urbas 1986; Steriade et al. 1996). Responses in the thalamorecipient zone (LIV and LLIII) and supragranular upper lamina III (ULIII) were analyzed separately. LLIII MUA and CSD are of interest because they reflect both initial cortical activation and activity at the location of cell bodies whose output is sent to other cortical areas potentially involved in further auditory processing (Galaburda and Pandya 1983; Pandya and Rosene 1993; Rouiller et al. 1991). ULIII responses, on the other hand, largely represent later polysynaptic activation of pyramidal cell elements by inter-laminar, intra-laminar, and cortico-cortical inputs (Galaburda and Pandya 1983; Matsubara and Phillips 1988; Mitani et al. 1985; Ojima et al. 1991; Rouiller et al. 1991; Steinschneider et al. 1994; Wallace et al. 1991b). The FFT was applied only to the "steady-state" phase of the response: 175–445 ms following stimulus onset. This time window isolated the portion of the response exhibiting phase-locked activity (when present), while excluding the initial ON response, major early response components, and potential OFF responses. The amplitude of the dominant frequency component in the amplitude spectrum within the frequency range from 10 to 300 Hz was used as a measure of phase-locked activity. This upper frequency boundary was chosen based on the fact that spectral peaks at frequencies >300 Hz were never observed, consistent with limits reported in previous studies of phase-locked activity in A1 of awake macaques (Fishman et al. 2000a; Steinschneider et al. 1998). Once the maximum of the spectrum was determined, it was counted as a peak only if the slope of the spectrum changed from positive to negative across 6 ( $\pm 3$ ) surrounding frequency bins. Otherwise, the next highest amplitude point in the spectrum was considered, and so on. This conservative criterion ensured that peaks corresponded to clear perturbations in the spectrum rather than merely to a point on the uniformly falling edge of a lower-frequency component. In the case of monkey data, the signal-to-noise ratio of oscillatory activity was sufficiently high that results were independent of whether or not this criterion was adopted. Because the musical intervals had multiple difference frequencies to which cortical neurons could potentially phase-lock, the peak of the amplitude spectrum provided an automatic and unbiased measure of oscillatory activity, free of a priori assumptions regarding the expected frequencies of phase-locked oscillations. However, as will be apparent, the vast majority of spectral peaks occurred at predicted difference frequencies calculated by pair-wise subtraction of stimulus frequency components.

In addition to spectral analyses of averaged response waveforms,

spectral analyses of responses to individual stimulus presentations were performed to assess variability of oscillatory activity across single trials and to evaluate the statistical significance of spectral peaks. It is important to distinguish between the spectrum of averaged response waveforms and the average of spectra of responses evoked by individual stimulus presentations. Whereas the former reflects only oscillatory components phase-locked to the stimulus, the latter reflects the combination of both phase-locked and non-phase-locked oscillations (including 60-Hz line noise), with non-phase-locked activity disappearing with appropriate signal averaging in the time domain. Statistical significance of spectral peaks was assessed by comparing mean spectra of non-octave chord-evoked responses with mean spectra of octave-evoked responses, based on the a priori assumption that the octave, being the most consonant of the intervals, should evoke the least amount of oscillatory activity out of all the intervals presented. Phase-locked responses were occasionally evident at more than one electrode contact located in LLIII or ULIII. For these penetrations, measures were based on the average of the amplitude spectra of responses recorded at two adjacent electrode contacts.

Based on averaged results of three psychoacoustic studies on consonance and dissonance (Butler and Daston 1968; Kameoka and Kuriyagawa 1969b; Malmberg 1918), the intervals used in the present study are ranked from most consonant to most dissonant as follows: octave (O), perfect fifth (P5), perfect fourth (P4), minor seventh (m7), augmented fourth (A4), major seventh (M7), major second (M2), and minor second (m2). Despite the fact that stimuli presented in these studies differed in several respects, e.g., chord base tones, relative amplitude of harmonics, and overall intensity (in 1st and 3rd of these studies, details of stimulus spectra were not reported), rank orders of dissonance were highly consistent (at least for the chords considered in the present study), provided each complex tone of the chord contained more than four lower harmonics (see Kameoka and Kuriyagawa 1969b). Thus although the absolute dissonances of the intervals may have differed across studies, their relative dissonances (rank orders) were similar. On the basis of these psychoacoustic ranks, we examined the degree to which the magnitude of oscillatory phase-locked activity in A1 correlates with the perceived dissonance of the chords.

**HUMAN RECORDINGS.** Spectral analysis (FFT) was also used to quantify phase-locked activity in AEPs recorded in human auditory cortex. For *subject 1*, the FFT was applied to the same window as that examined in the analysis of electrophysiological data obtained from monkey A1 (175–445 ms). This analysis window excluded ON and OFF response components. A shorter FFT analysis window (175–370 ms), consistent with shorter stimuli, was applied to the data from *subject 2*. Phase-locked activity was quantified using two complementary measures. The first measure was similar to that used to quantify monkey data: the peak of the amplitude spectrum from 10 to 300 Hz (no significant oscillatory activity was observed at frequencies >150 Hz). The second measure was the area under the amplitude spectrum from 10 to 150 Hz, which thereby includes spectral peaks corresponding to multiple difference frequencies or harmonics of a single difference frequency. Accordingly, an increase in oscillatory phase-locked activity at multiple frequencies, and potentially relevant for the encoding of roughness and sensory dissonance, would thus be represented by an increase in the integral of the amplitude spectrum of the "steady-state" response within the 10- to 150-Hz frequency range. As in the analysis

FIG. 6. Mean ( $\pm$ SE;  $n = 70$ ) LLIII MUA and CSD waveforms and corresponding mean ( $\pm$ SE) amplitude spectra of responses (175–445 ms poststimulus onset) evoked by the two most dissonant chords (minor and major 2nd) and by the two most consonant chords (octave and perfect 5th) at the same site as that shown in Fig. 4 (base tone = 512 Hz). Major peaks in the mean spectra of responses evoked by the minor 2nd and the major 2nd occur at predicted difference frequencies (indicated by arrows). Means at the peaks are significantly larger than means at corresponding frequencies in the spectrum of octave-evoked responses (one-tailed  $t$ -test;  $t$  and  $P$  values are shown in the figures). In contrast, the mean spectrum of perfect 5th-evoked responses (above 10 Hz) is not significantly different from that of octave-evoked responses ( $P > 0.05$ ). No significant differences between mean spectra are observed at frequencies >150 Hz.

Minor 2<sup>nd</sup>

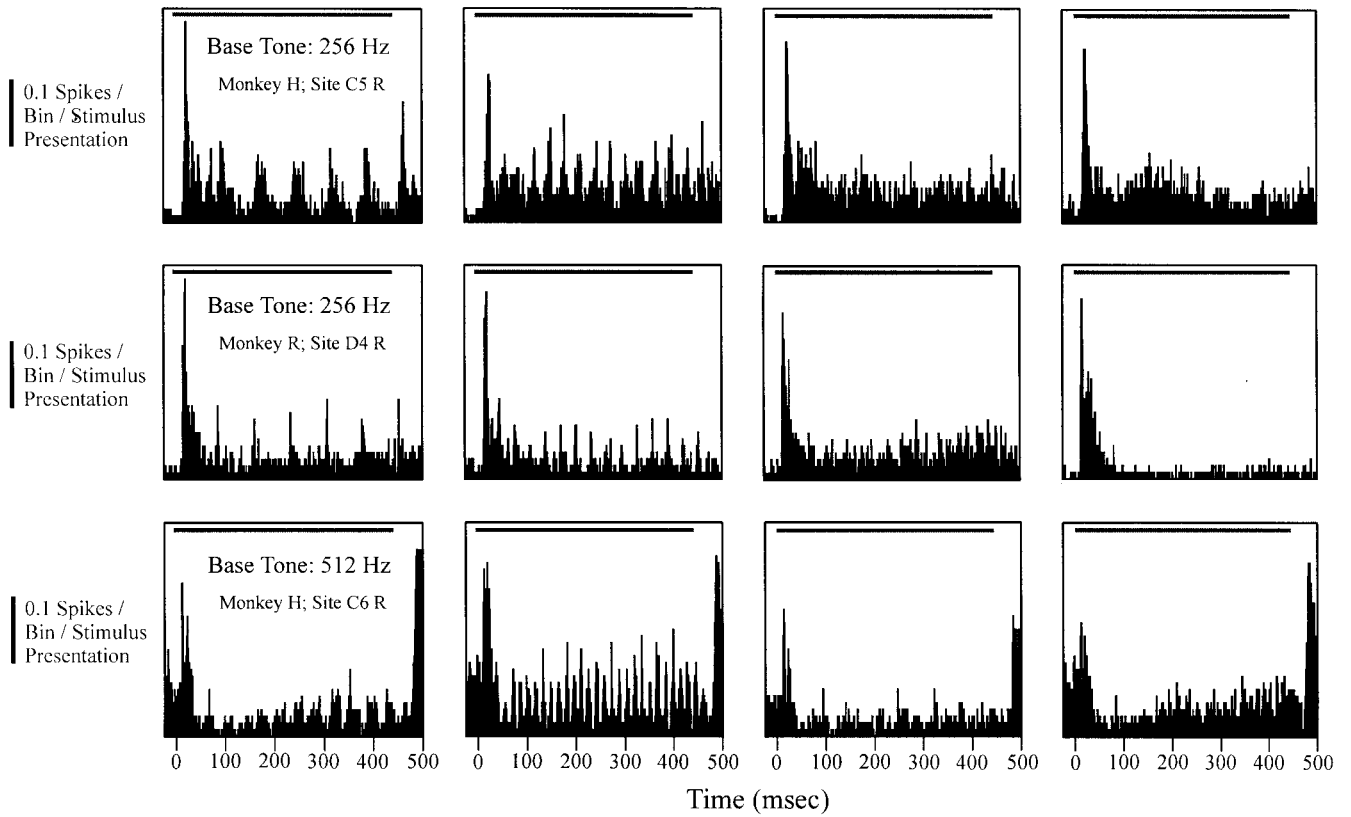
Major 2<sup>nd</sup>

Perfect 5<sup>th</sup>

Octave

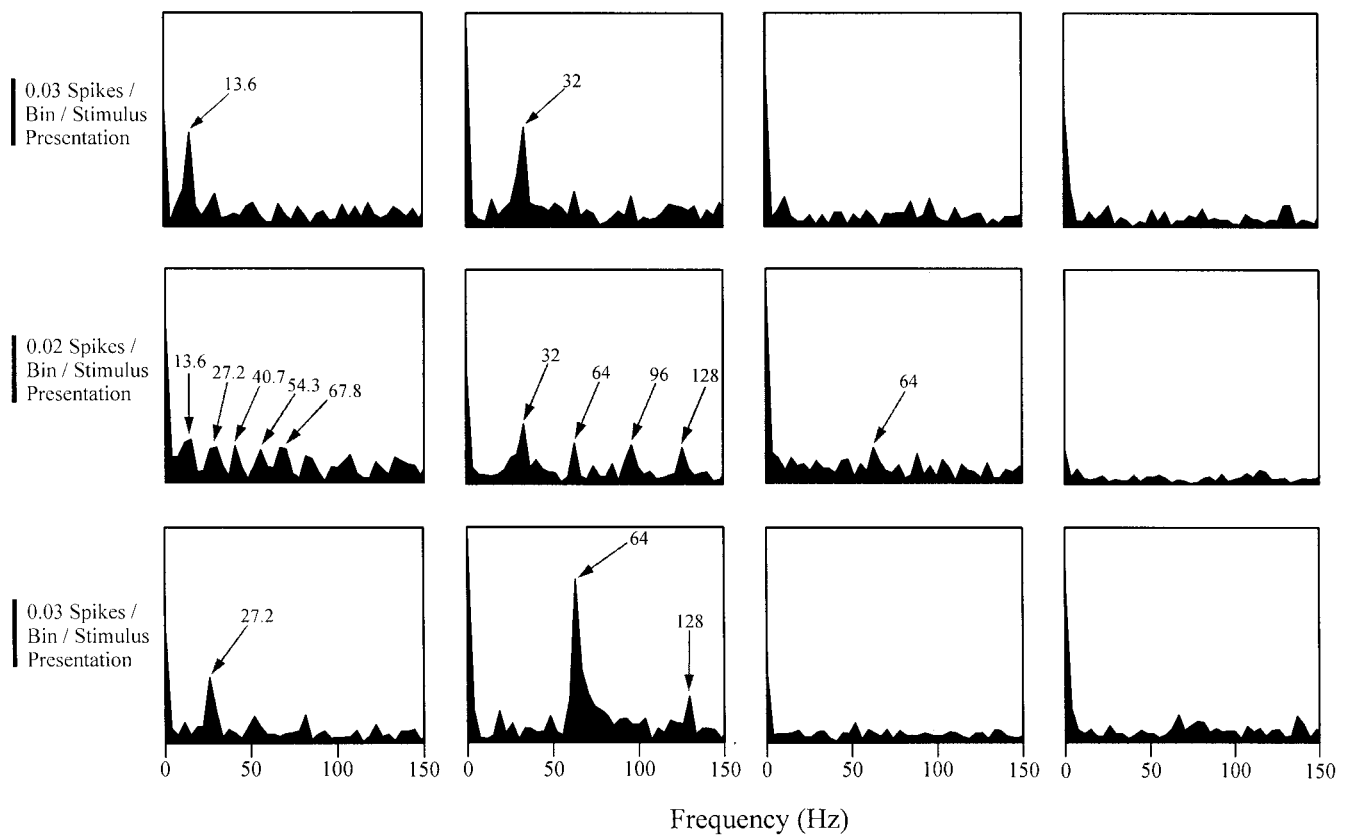
A

Peristimulus Time Histograms



B

Amplitude Spectra (175-445 msec)



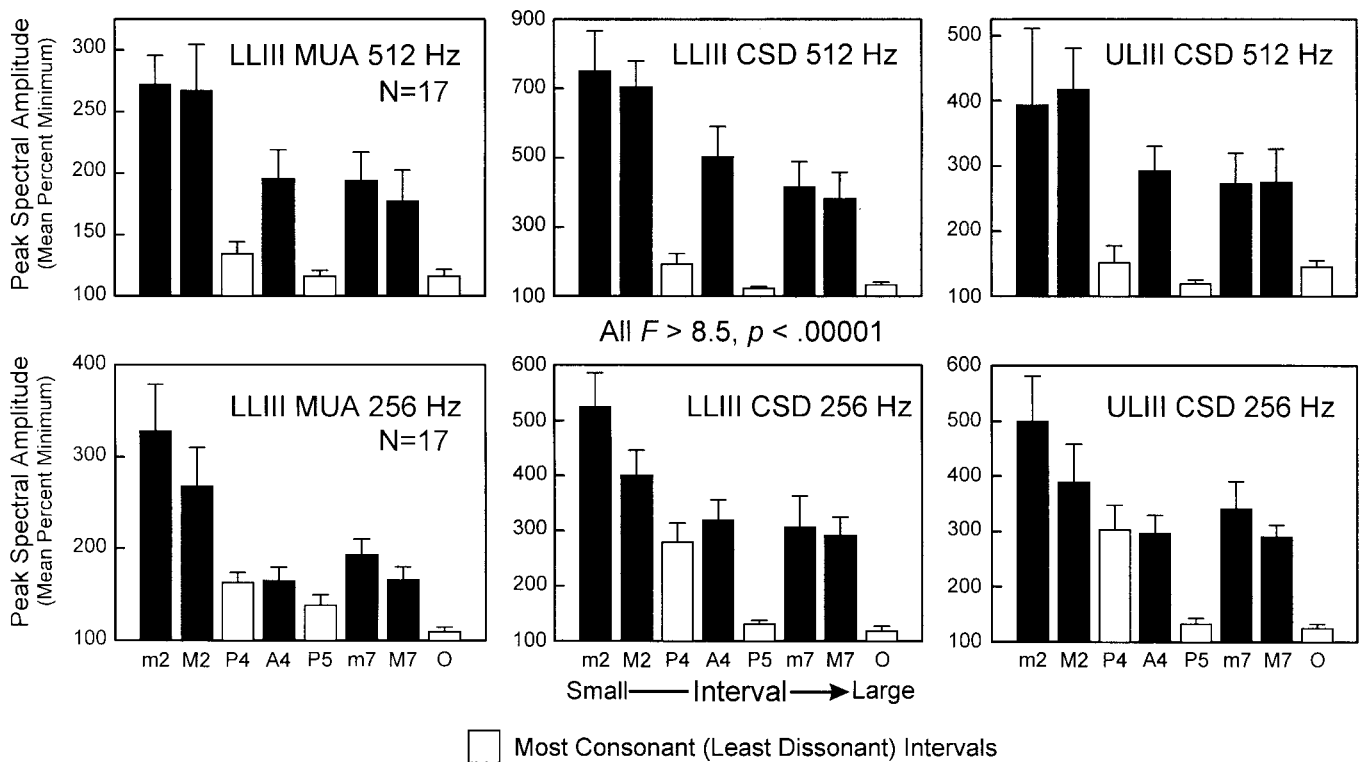


FIG. 8. Mean normalized peak spectrum amplitude ( $\pm$ SE) from 10 to 300 Hz as a function of musical interval (ordered from small to large). LLIII MUA and CSD and ULIII CSD data for 256- and 512-Hz base tone intervals are represented in separate histograms as indicated. Means are based on data from 17 electrode penetrations. Data corresponding to the 3 most consonant intervals (octave and perfect 5th and 4th) are represented by the white bars. Differences in mean peak spectrum amplitude across stimulus conditions are statistically significant (one-way ANOVA:  $F > 8.5, P < 0.00001$ ).

of monkey data, statistical significance of spectral peaks was assessed by comparing mean spectra of AEPs evoked by non-octave intervals with mean spectra of octave-evoked AEPs.

## RESULTS

### Neural ensemble activity evoked by musical chords in monkey A1

Results are based on 32 perpendicularly oriented (error,  $<20\%$ ) electrode penetrations into A1. 256 and 512 Hz base tone stimulus sets (each comprised of 8 intervals) were each presented in 17 penetrations. Due to the comparatively low number of sampled cortical sites with BFs  $<1,280$  Hz (the highest frequency component of the base tone in the 128-Hz octave stimulus), chords with base tones of 128 Hz were presented in only four electrode penetrations. As this small sample size precluded meaningful interpretation of statistical measures, data based on 128 Hz base tone interval-evoked responses are not discussed.

Figure 2 shows representative CSD and MUA laminar response profiles evoked by the two most consonant and the two most dissonant musical chords presented: octave and perfect

fifth, minor second and major second, respectively (base tone = 256 Hz; BF = 1,600 Hz). CSD and MUA waveforms in each quadrant of the figure represent neuronal activity recorded simultaneously at 150- $\mu$ m interval depths within A1. The dashed boxes superimposed on the LLIII responses delineate the temporal window subjected to spectral analysis.

All of the musical chords elicit a stereotypical laminar pattern of activity characterized by short-latency current sinks (below-baseline deflections in the CSD) located in the thalamorecipient zone (lamina IV and LLIII, Initial sink), and slightly later supragranular sinks located in mid- and ULIII (supragranular sink). These sinks are coincident with above-baseline bursts of MUA in lamina III and IV (MUA ON), indicating that they primarily represent current flow associated with depolarizing synaptic potentials. The LLIII and ULIII sinks are balanced by deeper and more superficial current sources (above-baseline deflections in the CSD, e.g., P28 source) that, together with the sinks, form current dipole configurations consistent with initial activation of pyramidal cells in LLIII and delayed polysynaptic activation of pyramidal cell elements in ULIII.

While initial components of responses evoked by the con-

FIG. 7. A: peristimulus time histograms (PSTHs) of multiunit cluster activity recorded in LLIII at 3 A1 sites (binwidth = 1 ms). BFs of the sites shown in the 1st and 3rd rows are 1,600 and 4,000 Hz, respectively. The site shown in the second row did not display a clear BF and exhibited broad frequency tuning ranging from 200 to 15,000 Hz. Black bars above the PSTHs indicate stimulus duration. Note that chords presented at the site shown in the 3rd row had base tones of 512 Hz. B: amplitude spectra of PSTHs in A from 175 to 445 ms poststimulus onset. Spectra of minor- and major-2nd-evoked responses display peaks at predicted difference frequencies and their harmonics, whereas major peaks are absent in spectra of perfect 5th- and octave-evoked responses. No significant peaks are observed at frequencies  $>150$  Hz.

FIG. 9. Maxima of amplitude spectra of chord-evoked responses (175–445 ms poststimulus onset) in each electrode penetration were ranked (from lowest to highest amplitude). Mean ranks ( $n = 17$  penetrations) are shown as a function of the perceived dissonance of the musical chords (ordered from least dissonant to most dissonant). Error bars indicate SE. LLIII MUA and CSD and ULIII CSD data for 256- and 512-Hz base tone intervals are represented in separate histograms as indicated. For all response measures and for both chord octave ranges, physiological ranks are significantly correlated with perceptual ranks (Spearman correlation analysis:  $r$  and one-tailed  $P$  values are indicated in the figure). A linear regression line superimposed on the histograms emphasizes this relationship.

sonant and dissonant stimuli are qualitatively very similar, later portions of the responses differ considerably. Both of the dissonant intervals evoke prominent oscillations in the MUA and CSD that are phase-locked to the predicted difference frequencies (minor 2nd: 13.6 Hz; major 2nd: 32 Hz). Neuronal beating patterns are evident both in the thalamorecipient zone and in ULIII, with maximal phase-locked activity typically occurring in LLIII. This laminar distribution is consistent with that observed in previous investigations of phase-locked neural ensemble activity in macaque A1 (Fishman et al. 2000a; Stein-schneider et al. 1998). In contrast, little or no oscillatory activity is evoked by the consonant intervals. Waveforms of LLIII MUA low-pass filtered at 800 Hz (96 dB/octave roll-off) prior to digitization are superimposed on those of unfiltered data. MUA waveforms generated under these two conditions are virtually identical, confirming the absence of significant signal aliasing. This is further demonstrated by the nearly flat difference waveforms, shown below the superimposed waveforms.

Figure 3 shows representative chord-evoked responses (base tone = 256 Hz) recorded at a site with a BF of 1,000 Hz. LLIII MUA and CSD waveforms of chord-evoked responses (from 175 to 445 ms poststimulus onset) and associated amplitude spectra are depicted in Fig. 3, A and B, respectively. The most dissonant intervals (minor and major 2nd) evoke robust phase-locked oscillations in both the MUA and CSD, which are manifested as prominent peaks in the associated amplitude

spectra at predicted difference frequencies (indicated by arrows). In contrast, responses evoked by the most consonant intervals (octave and perfect 5th) display little or no oscillatory activity and are characterized by comparatively flat amplitude spectra. Intervals of intermediate dissonance (e.g., major 7th, minor 7th, and augmented 4th) also evoke oscillatory phase-locked responses. Even the perfect fourth evokes oscillatory activity, consistent with the theoretical prediction that perfect fourths become disproportionately more dissonant, compared with octaves and perfect fifths, when the base tone of the chord is lower than  $\sim 300$  Hz (see Fig. 12 in Plomp and Levelt 1965).

Figure 4 shows representative chord-evoked responses (base tone = 512 Hz) recorded at a site with a BF of 4,000 Hz. Similar to the pattern of responses evoked by chords with base tones of 256 Hz, the most dissonant 512-Hz base tone chords (minor and major 2nd) evoke phase-locked oscillations in both the MUA and CSD, which are represented as prominent peaks in the associated amplitude spectra at predicted difference frequencies (indicated by arrows). In contrast, responses evoked by the most consonant chords (octave and perfect 5th) are characterized by a virtual absence of rapid oscillations and by comparatively flat amplitude spectra. Intervals of intermediate dissonance also evoke oscillatory responses phase-locked to predicted difference frequencies.

Oscillatory phase-locked activity is visible not only in the averaged responses but also in responses evoked by individual stimulus presentations. Statistical significance of phase-locked

Subject 1; Site 1 Heschl's Gyrus

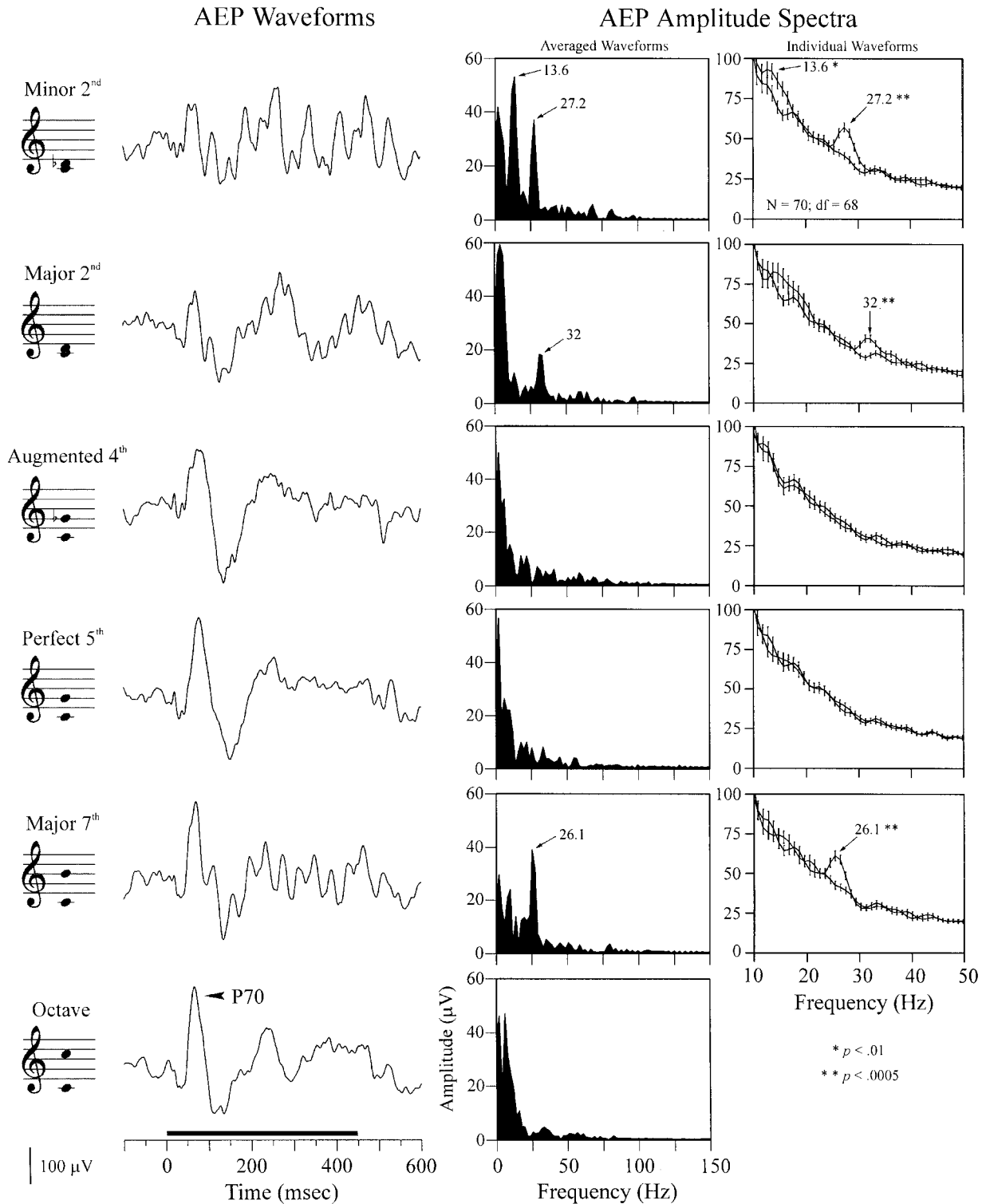


FIG. 10. Waveforms and corresponding amplitude spectra of chord-evoked AEPs recorded at a single site in Heschl's gyrus of *subject 1* (base tone = 256 Hz). *Far left*: musical notation representation of the chords. Stimulus duration is represented by the black bar above the time axis. The "P70" component of the AEP is indicated in the octave-evoked response. AEPs evoked by dissonant chords (e.g., minor 2nd) display oscillations phase-locked to the predicted difference frequencies, whereas AEPs evoked by consonant chords (e.g., octave and perfect 5th) display little or no oscillatory activity. Arrows in the amplitude spectra indicate major peaks occurring at predicted difference frequencies (values, in Hz, next to arrows). Mean spectra of AEPs evoked by non-octave chords in individual stimulus presentations are shown superimposed on mean spectra of octave-evoked AEPs in the *right-hand side* of the figure. Error bars indicate SE. Means at peaks are significantly larger than means at corresponding frequencies in the spectrum of octave-evoked responses (*P* values indicated in figure). No significant differences between mean spectra are observed at frequencies >50 Hz.

## Subject 1; Mean Site 1, 2, and 3 Heschl's Gyrus

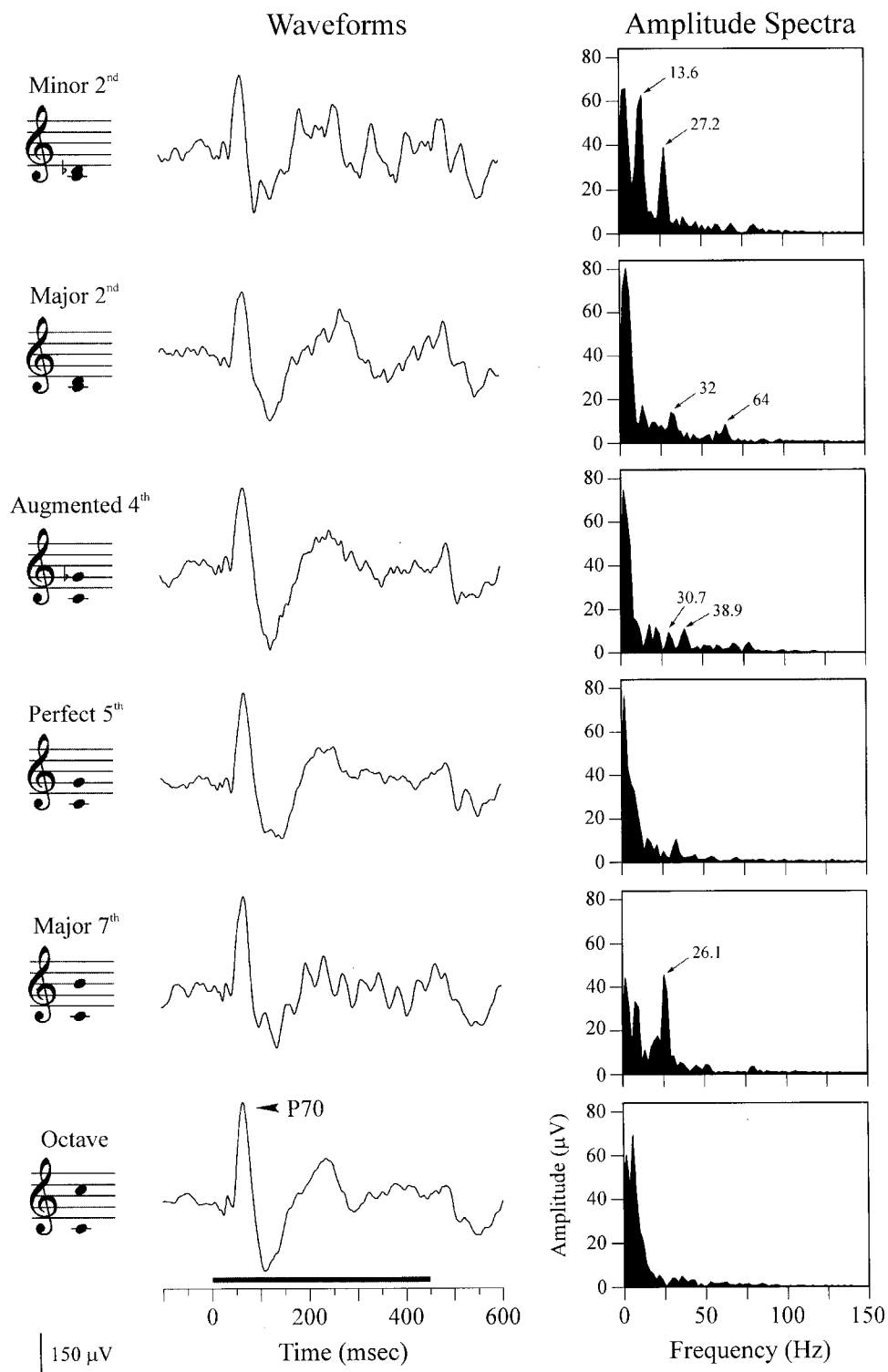


FIG. 11. Waveforms and amplitude spectra of chord-evoked AEPs averaged across 3 Heschl's gyrus recording sites in *subject 1* (base tone = 256 Hz). Same conventions as in Fig. 10.

activity in the dissonant-chord-evoked responses relative to octave-evoked responses is demonstrated in Figs. 5 and 6 for two representative A1 sites (the same as those shown in Figs. 2 and 4, respectively). The figures show mean ( $\pm$ SE) LLIII MUA and CSD waveforms and corresponding mean ( $\pm$ SE) amplitude spectra of responses evoked by the two most disso-

nant chords (minor and major 2nd) and by the two most consonant chords (octave and perfect 5th). Major peaks in the mean spectrum of responses evoked by the minor second and the major second occur at predicted difference frequencies ( $\leftarrow$ ). Means at the peaks are significantly larger than means at corresponding frequencies in the spectrum of octave-evoked



## Human Subject 1 Intracortical AEPs: Heschl's Gyrus

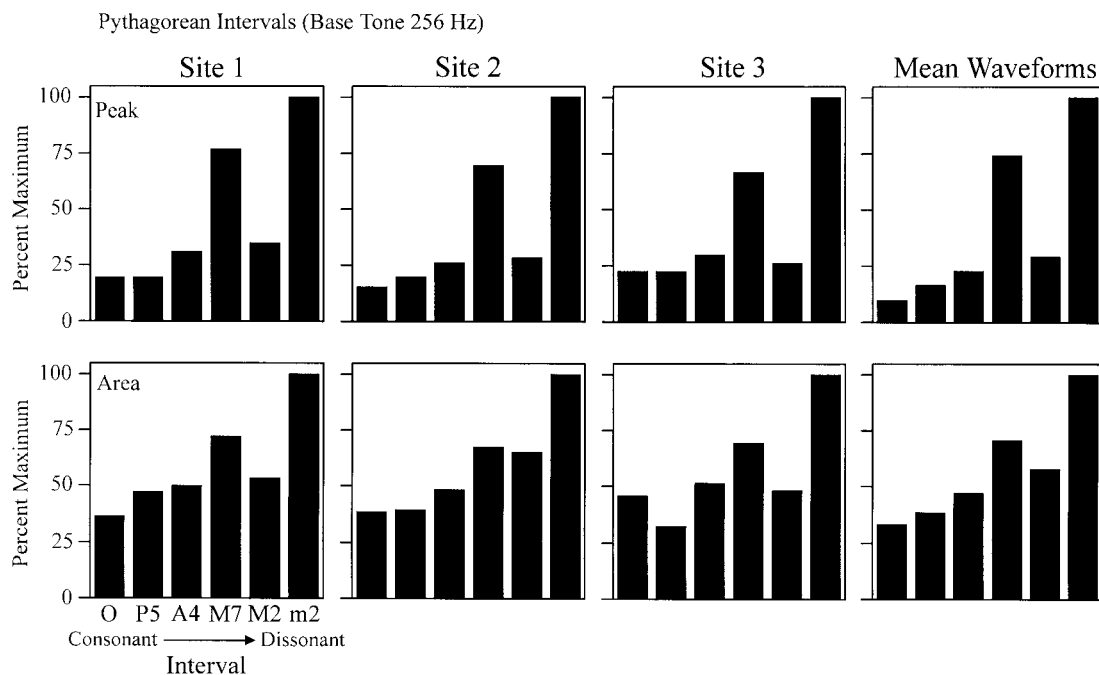


FIG. 12. Normalized (percent maximum) peak and area measurements of amplitude spectra (10–150 Hz) of chord-evoked AEPs recorded in Heschl's gyrus of *subject 1* as a function of the dissonance of the chords (base tone = 256 Hz). Data from each of the 3 recording sites are represented in separate histograms as indicated. Spectrum peaks and areas tend to increase with increasing dissonance of the chords. This trend is also apparent for AEP waveforms averaged across the 3 recording sites (*right-most histogram*).

responses (one-tailed *t*-test; *t* and *P* values are shown in the figures). In contrast, the mean spectrum of perfect fifth-evoked responses (above 10 Hz) is not significantly different from that of octave-evoked responses ( $P > 0.05$ ). No significant differences between mean spectra are observed at frequencies  $>150$  Hz. Peaks at 60 Hz are present in the mean spectra of Fig. 5 due to the fact that mean spectra of responses to individual stimulus presentations include 60-Hz line noise, which disappears with time domain averaging.

Similar chord-evoked oscillatory response patterns are observed when responses are analyzed using more conventional neurophysiological techniques. Figure 7A shows PSTHs based on multiunit spike activity recorded in LLIII at three representative sites in A1. Data from two of these sites are represented in Figs. 2–6. Similarly to MUA results, PSTHs of the minor and the major-second-evoked responses display periodic oscillations that are absent in the PSTHs of the octave- and perfect-fifth-evoked responses. Dissonant-chord-evoked oscillations are manifested as peaks in corresponding amplitude spectra at predicted difference frequencies or their harmonics (Fig. 7B). In contrast, spectra of consonant-chord-evoked responses are characterized by a general absence of significant peaks at frequencies  $>10$  Hz.

Figures 2–7 illustrate a general pattern of musical chord-evoked responses in A1: dissonant intervals evoke oscillatory phase-locked activity, whereas consonant intervals evoke comparatively little or no phase-locked activity. However, the relative magnitude of phase-locked activity differs across sites. According to the roughness theory of sensory dissonance, the total dissonance of a musical chord reflects the sum of roughness contributed by each pair of unresolved frequency compo-

nents (Kameoka and Kuriyagawa 1969b; Plomp and Levelt 1965; Terhardt 1974a, 1978). Accordingly, the overall dissonance of a musical chord should be represented by response patterns averaged across the tonotopic map in A1.

To quantify the average relative phase-locked activity across the cortical sites sampled, the peak of the amplitude spectrum of each chord-evoked response was first expressed as a percentage of the minimum peak amplitude of the eight chord-evoked response spectra obtained in each electrode penetration. Normalized peak spectrum amplitudes were subsequently averaged across penetrations. The resultant mean normalized amplitudes, plotted as a function of musical interval (ordered from left to right according to interval width) for each of the three response measures examined are shown in Fig. 8. On average, the octave and perfect fifth evoke comparatively little phase-locked activity, whereas the minor and major second generally evoke the highest amplitude phase-locked responses. The perfect fourth, the third most consonant interval, also yields comparatively little phase-locked activity when presented within the octave above middle C (i.e., with a base tone of 512 Hz) but becomes physiologically more “dissonant” when presented within the octave of middle C. Differences among mean normalized amplitudes across interval conditions are statistically significant (repeated-measures ANOVA: all  $F > 8.5$ ,  $P < 0.00001$ ).

To examine the extent to which the magnitude of phase-locked neuronal ensemble activity in A1 correlates with the perceived dissonance of the musical intervals, spectra of the eight musical interval-evoked responses from each electrode penetration were ranked according to their peak amplitude (1 = lowest amplitude, 8 = highest amplitude). Physiological

## Subject 1; Site 1 Heschl's Gyrus

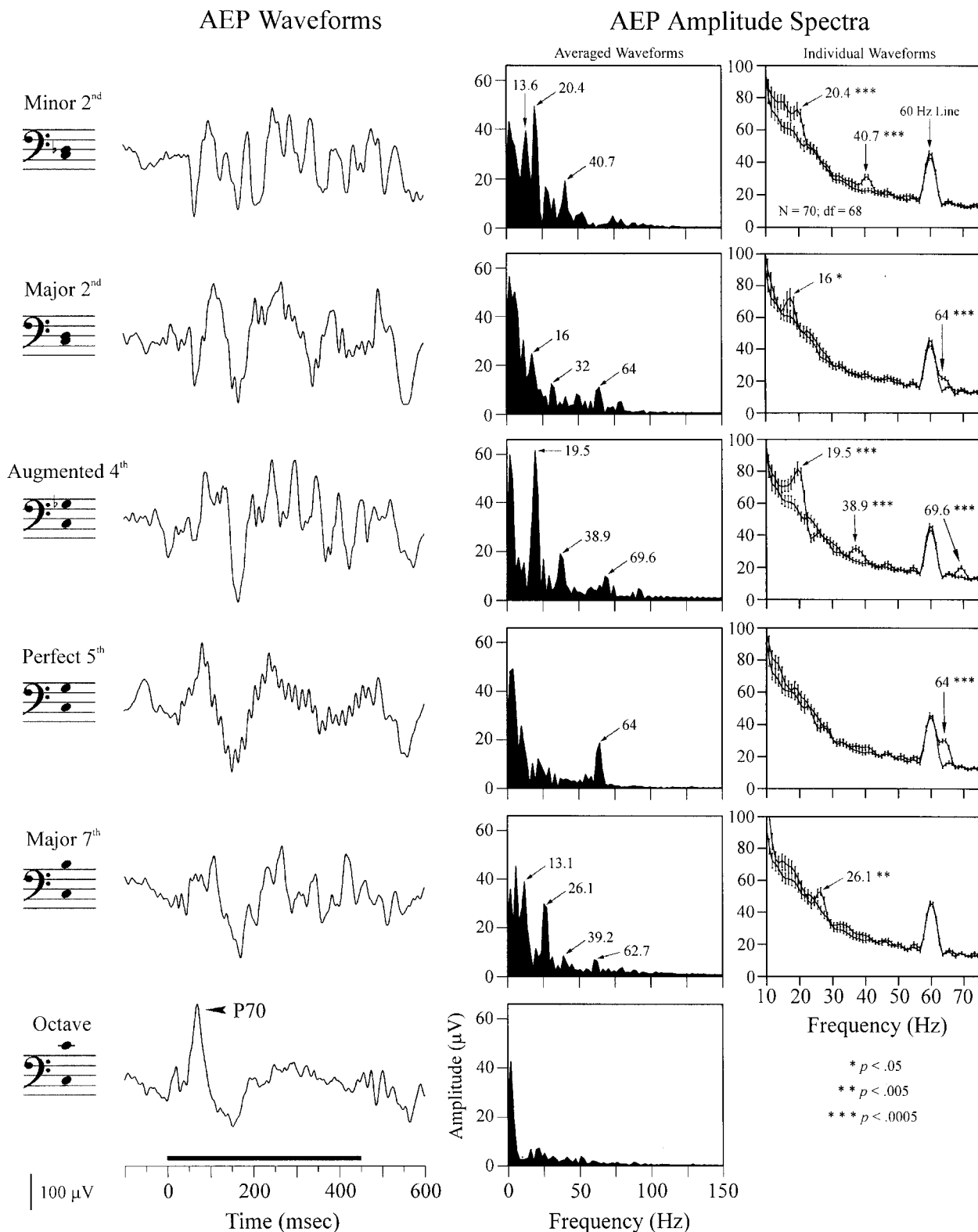


FIG. 13. Waveforms and amplitude spectra of chord-evoked AEPs recorded at a single site in Heschl's gyrus of *subject 1* (base tone = 128 Hz). Same site and conventions as in Fig. 10. AEPs evoked by all chords, except for the octave, display oscillations phase-locked to the predicted difference frequencies. Arrows in the amplitude spectra indicate major peaks corresponding to predicted difference frequencies (values, in Hz, next to arrows). Most of the spectral peaks are statistically significant, relative to mean spectra of octave-evoked AEPs ( $P$  values indicated in the figure). Peaks at 60 Hz correspond to line noise, which disappears with time domain averaging. No significant differences between mean spectra are observed at frequencies >75 Hz.

ranks from each penetration were then compared with perceptual ranks of the intervals (1 = least dissonant, 8 = most dissonant; see details in METHODS). For all three of the response measures examined, mean rank of spectral amplitude tends to increase with the perceived dissonance of the chords (Fig. 9). This relationship, quantified by Spearman rank-order correlation analysis based on raw data ( $n = 17$  penetrations) and emphasized by the superimposed linear regression lines, is statistically significant ( $r$  values are indicated in the figure;  $P < 0.00001$ ) for all response components and octave ranges examined. The strongest correlation between neural and perceptual measures is seen for LLIII CSD, while the weakest association is seen for ULIII CSD.

#### *AEPs evoked by musical chords in human auditory cortex*

AEPs recorded directly from human auditory cortex display strikingly similar response patterns to those observed in monkey A1. Figure 10 shows musical chord-evoked AEPs recorded at a single site within Heschl's gyrus (*subject 1*; base tone = 256 Hz). AEPs evoked by the most dissonant chords (minor 2nd, major 2nd, and major 7th) display prominent oscillations phase-locked to the predicted difference frequencies, which are manifested as peaks (indicated by arrows) in the corresponding amplitude spectra (*middle*). In contrast, responses evoked by the octave and by the perfect 5th are characterized by an absence of rapid oscillations and by comparatively flat, low-amplitude spectra (above  $\sim 10$  Hz).

Statistical significance of the spectral peaks was assessed by comparing the mean spectrum of AEPs evoked by the non-octave chords with that of AEPs evoked by octaves (shown superimposed in Fig. 10, *right*; error bars represent SE). Means at the peaks of the dissonant chord-evoked response spectra are significantly larger than means at corresponding frequencies in the mean spectrum of octave-evoked responses (one-tailed  $t$ -test;  $P$  values are indicated in the figure). Mean spectra of AEPs evoked by the perfect fifth and the augmented fourth are not significantly different from the mean spectrum of octave-evoked AEPs ( $P > 0.05$ ). No significant differences between the mean spectrum of non-octave chord-evoked responses and that of octave-evoked responses were observed at frequencies  $> 50$  Hz. Similar phase-locked response patterns are displayed by AEPs averaged across the three recording sites in Heschl's gyrus of *subject 1* (Fig. 11), indicating that oscillatory activity is synchronized (i.e., displays phase coherence) over a considerable distance across the cortical tissue. As quantified in Fig. 12, the magnitude of oscillatory phase-locked activity at each of the three sites and in the averaged data tends to increase with increasing dissonance of the chords. The only major deviation from this trend is the greater oscillatory activity evoked by the major seventh relative to that evoked by the major second.

Figure 13 depicts AEPs evoked by chords with base tones of 128 Hz recorded at the same site in Heschl's gyrus as that shown in Fig. 10. In this case, with the exception of the octave, all intervals, including the perfect fifth, evoke oscillatory responses. This pattern is consistent with the observation that in octave ranges below middle C, all intervals, except octaves with base tones  $> 100$  Hz, sound rougher and more dissonant than their higher octave counterparts (Plomp and Levelt 1965). This may explain why, in lower octave ranges, intervals smaller than octaves, including perfect fifths, tend to be

avoided in music composition (Plomp and Levelt 1965). Oscillatory phase-locked responses are manifested as peaks in corresponding amplitude spectra at predicted difference frequencies (indicated by arrows). Many of these spectral peaks are statistically significant relative to the mean spectrum of octave-evoked responses (*right-hand column*; one-tailed  $t$ -test;  $P$  values are indicated in the figure; same conventions as in Fig. 10). No significant differences between mean spectra are observed at frequencies  $> 75$  Hz ( $P > 0.05$ ).

As in the case of responses evoked by intervals with base tones of 256 Hz, similar phase-locked response patterns are evident when AEPs evoked by intervals with base tones of 128 Hz are averaged across the three Heschl's gyrus recording sites (Fig. 14), indicating that oscillatory activity is synchronized over a considerable distance across the cortical tissue. Responses are quantified in Fig. 15, which again shows that, with the exception of the comparatively high values obtained for augmented fourth-evoked responses, the magnitude of oscillatory activity tends to increase with increasing dissonance of the chords. This trend is not apparent, however, for AEPs recorded at the most medial location in Heschl's gyrus, Site 3.

Chord-evoked AEPs recorded simultaneously at posterior electrode contacts located in the planum temporale display little or no oscillatory activity, even when elicited by the most dissonant chords. This is illustrated in Fig. 16, which shows representative AEPs recorded at a single site in the planum temporale (base tone = 128 Hz). Phase-locked activity is largely absent except for low-amplitude (but statistically significant, relative to octave-evoked responses;  $P < 0.005$ ) oscillations at 64 Hz in the perfect fifth-evoked response. Above 10 Hz, mean spectra of AEPs evoked by all non-octave intervals are not significantly different from the mean spectrum of octave-evoked AEPs ( $P > 0.05$ ), except for the mean spectrum of the perfect fifth-evoked response. AEPs evoked by chords with base tones of 256 Hz (whose difference frequencies are double those of the 128-Hz base tone chords) in the planum temporale are characterized by a similar absence of oscillatory activity, even in the case of perfect fifth-evoked responses (data not shown). Correspondingly, mean spectra of all non-octave 256-Hz base tone chord-evoked responses are not significantly different from mean spectra of octave-evoked responses ( $P > 0.05$ ; data not shown). This markedly diminished sensitivity to temporal features of the stimuli is remarkable given the comparatively large amplitude of the AEPs (e.g., the P70 component—indicated in the octave responses—recorded in the planum temporale is approximately twice the amplitude of that recorded in Heschl's gyrus).

In both Heschl's gyrus and the planum temporale, the amplitude of the P70 component tends to increase with increasing width of the musical intervals (i.e., as interval ratio increases from the minor 2nd, the smallest interval, to the octave, the largest interval), as illustrated in Fig. 17A, *left*. Previous physiological studies in monkey A1 have shown that the amplitude of intracortical AEP components increases with increasing frequency separation between the harmonics of a complex tone spectrally centered at the BF, consistent with a manifestation of critical band masking phenomena (Fishman et al. 2000b). Because the number of pairs of resolved harmonics in the musical chords increases with interval width (Fig. 17A, *right*), we hypothesized that the enhancement in P70 amplitude with

## Subject 1; Mean Site 1, 2, and 3 Heschl's Gyrus

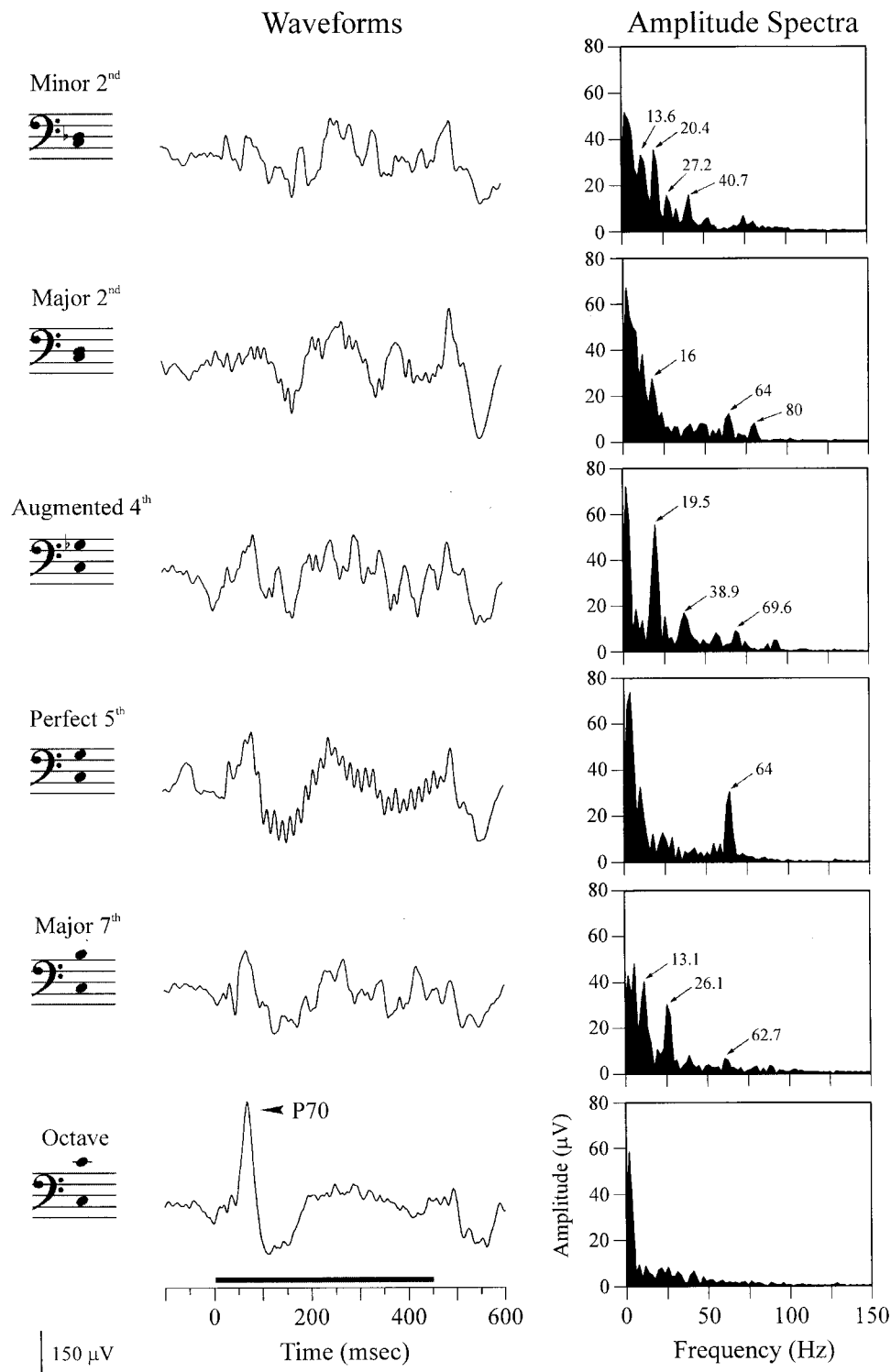


FIG. 14. Waveforms and corresponding amplitude spectra of chord-evoked AEPs averaged across the 3 Heschl's gyrus recording sites in *subject 1* (base tone = 128 Hz). Same conventions as in Fig. 11. AEPs evoked by all chords, except for the octave, display oscillations phase-locked to the predicted difference frequencies. Arrows in the amplitude spectra indicate major peaks corresponding to predicted difference frequencies (values, in Hz, next to arrows).

increases in interval width may reflect similar critical band masking effects. In support of this hypothesis, P70 amplitudes in both Heschl's gyrus and the planum temporale are correlated with the number of pairs of spectrally resolved harmonics in the chords (Fig. 17B). Linear regression lines superimposed on the scatter-plots emphasize this relationship. Spearman corre-

lation coefficients for AEPs recorded at each of the three Heschl's gyrus and three planum temporale electrode contacts are indicated in the scatter-plot insets. Coefficients  $>0.83$  are statistically significant ( $n = 6$ ,  $P < 0.05$ ). Amplitudes of chord-evoked responses in A1 of the monkey are unrelated to interval width (data not shown). A possible explanation for

## Human Subject 1 Intracortical AEPs: Heschl's Gyrus

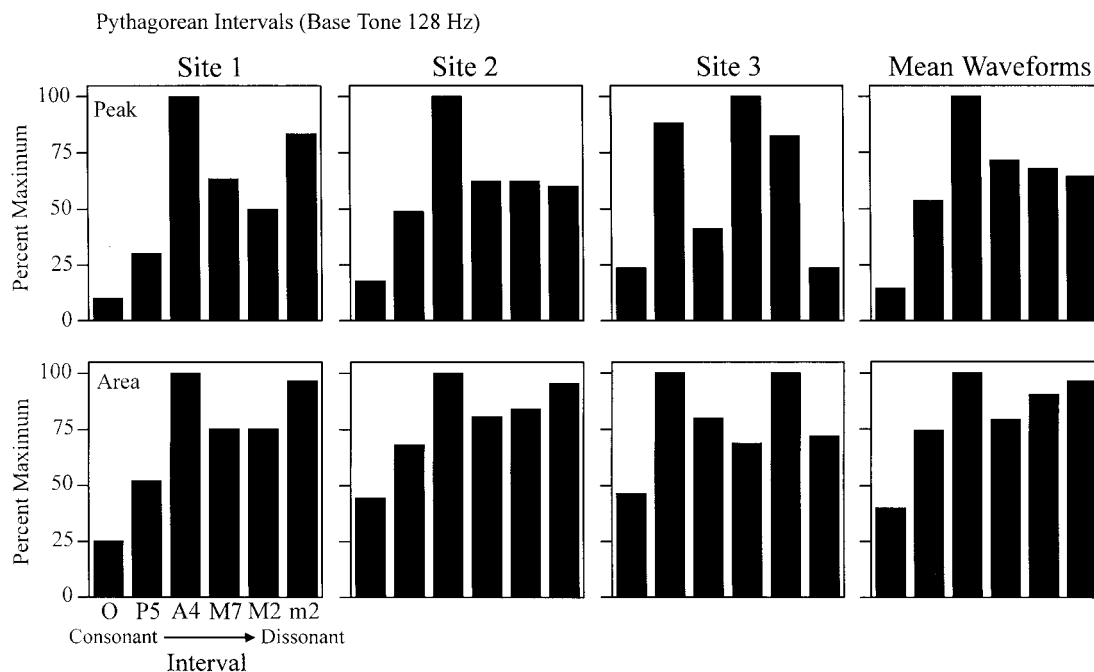


FIG. 15. Normalized (percent maximum) peak and area measurements of amplitude spectra (10–150 Hz) of chord-evoked AEPs recorded in Heschl's gyrus of *subject 1* as a function of the dissonance of the chords (base tone = 128 Hz). Same conventions as in Fig. 12.

why such effects in monkey A1 are not observed in the present study is considered in the DISCUSSION.

The relative perceived consonance/dissonance of musical intervals tuned according to the Pythagorean or pure fifth tuning system does not differ substantially from that of intervals tuned according to the equal temperament system, which divides the octave into 12 equal semitones. Accordingly, patterns of oscillatory activity evoked by equal temperament intervals were similar to those evoked by Pythagorean intervals. Figure 18B shows AEPs evoked by chords played an octave below middle C on an electronic keyboard (tuned in equal temperament) at three sites within Heschl's gyrus of *subject 2*. AEPs evoked by the minor and major second display oscillations that are manifested as peaks (indicated by arrows) in the corresponding amplitude spectra. In contrast, AEPs evoked by the perfect fifth and by the octave display comparatively little or no oscillatory activity. Observations are quantified in Fig. 19. The bipolar nature of these recordings between closely spaced electrode contacts ensures that these AEPs represent locally generated potentials reflecting synaptic activity of neuronal populations within Heschl's gyrus.

## DISCUSSION

Musical chords evoke oscillatory responses in monkey A1 that are phase-locked to the predicted difference frequencies, observed both in the thalamorecipient zone and in more superficial cortical layers, and manifested by synchronous synaptic and action potential activity of neuronal populations, as reflected by CSD, MUA, and PSTH measures. The magnitude of oscillatory activity correlates with the dissonance of the chords generally perceived by human listeners. Chord-evoked AEPs recorded directly from auditory cortex in humans display re-

markably similar oscillatory response patterns that are synchronized over a considerable distance across the cortical tissue. The observed parallels between monkey and human data and their correlation with perception strongly suggest that phase-locked oscillatory responses in monkey A1 are not epiphenomena but likely represent an important component of the auditory cortical representation of sensory dissonance.

In contrast to the prominent oscillatory activity displayed by AEPs recorded in Heschl's gyrus, AEPs recorded in the planum temporale were comparatively insensitive to rapid temporal features of the stimuli. These differential response characteristics may reflect differences in functional time constants between the neuron populations comprising these two auditory cortical regions or differences in their inputs. These findings support the functional segregation of primary and secondary auditory cortical fields in humans, whereby each field is biased toward processing specific aspects of the stimuli. Diminished ability of sites in the planum temporale to represent rapid acoustic transients embedded in consonant-vowel syllables was also reported by Steinschneider et al. (1999). However, given that only three planum temporale sites were examined in a single subject, these observations must be viewed as preliminary findings that will require physiological data from additional subjects before a more general conclusion can be drawn regarding differences in temporal processing between Heschl's gyrus and the planum temporale.

The fact that components of the AEPs (e.g., "P70") generated in these functionally distinct cortical areas overlap in time has important implications for noninvasive studies that utilize single current dipole models of auditory cortical activation to characterize the topographic and functional organization of auditory cortex. The present results, as well as those of other

## Subject 1; Site 1 Planum Temporale

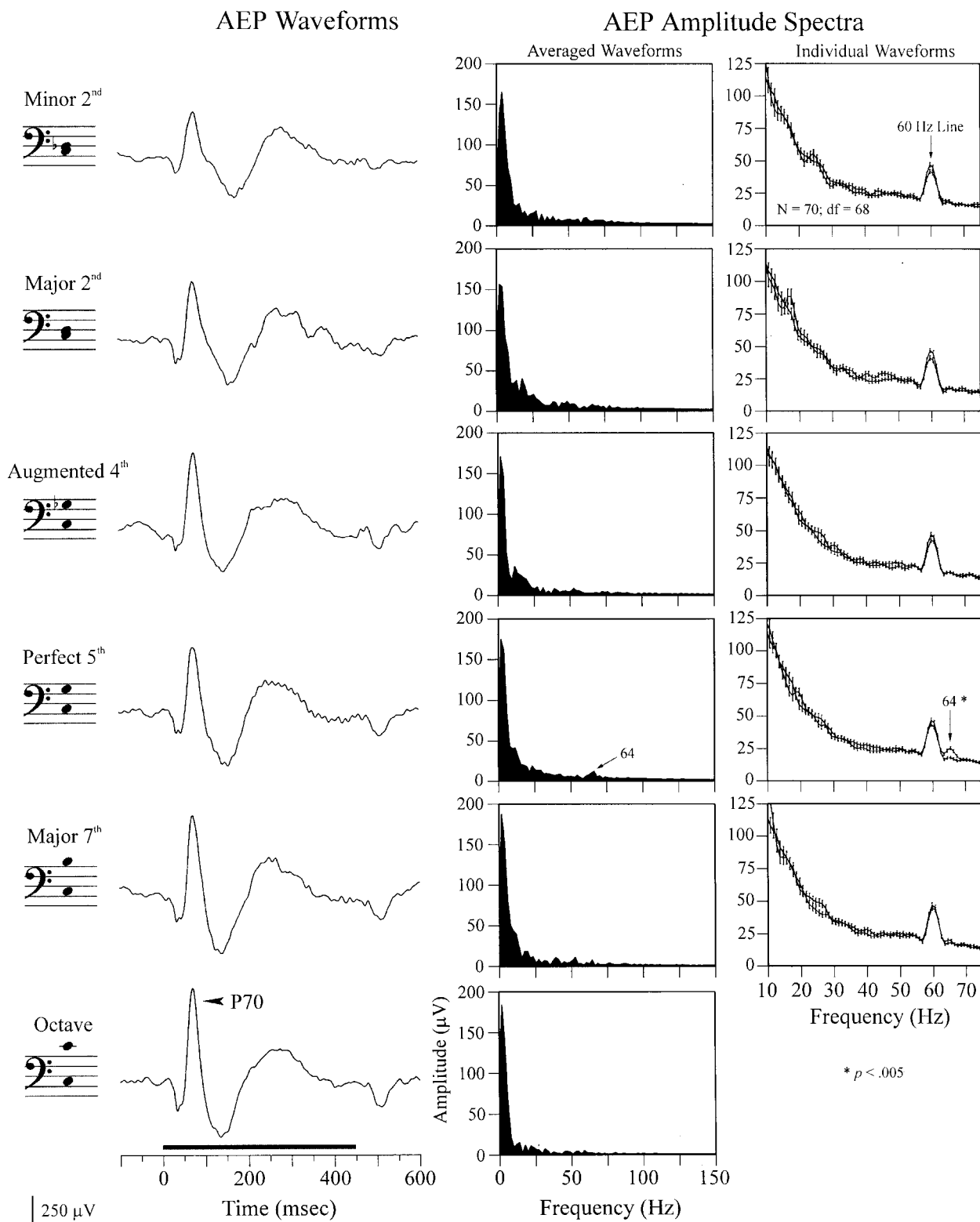


FIG. 16. Representative waveforms and corresponding amplitude spectra of chord-evoked-AEPs recorded in the planum temporale of *subject 1* (base tone = 128 Hz). Same conventions as in Fig. 10. Note the larger amplitude of planum temporale responses (compared with Heschl's gyrus responses) and the absence of significant oscillatory activity in AEPs evoked by even the most dissonant chords. The only exception is the low-amplitude (but statistically significant) oscillatory activity at 64 Hz in the perfect 5th-evoked response. A complete absence of statistically significant phase-locked activity is observed for AEPs evoked by chords with base tones of 256 Hz (data not shown).

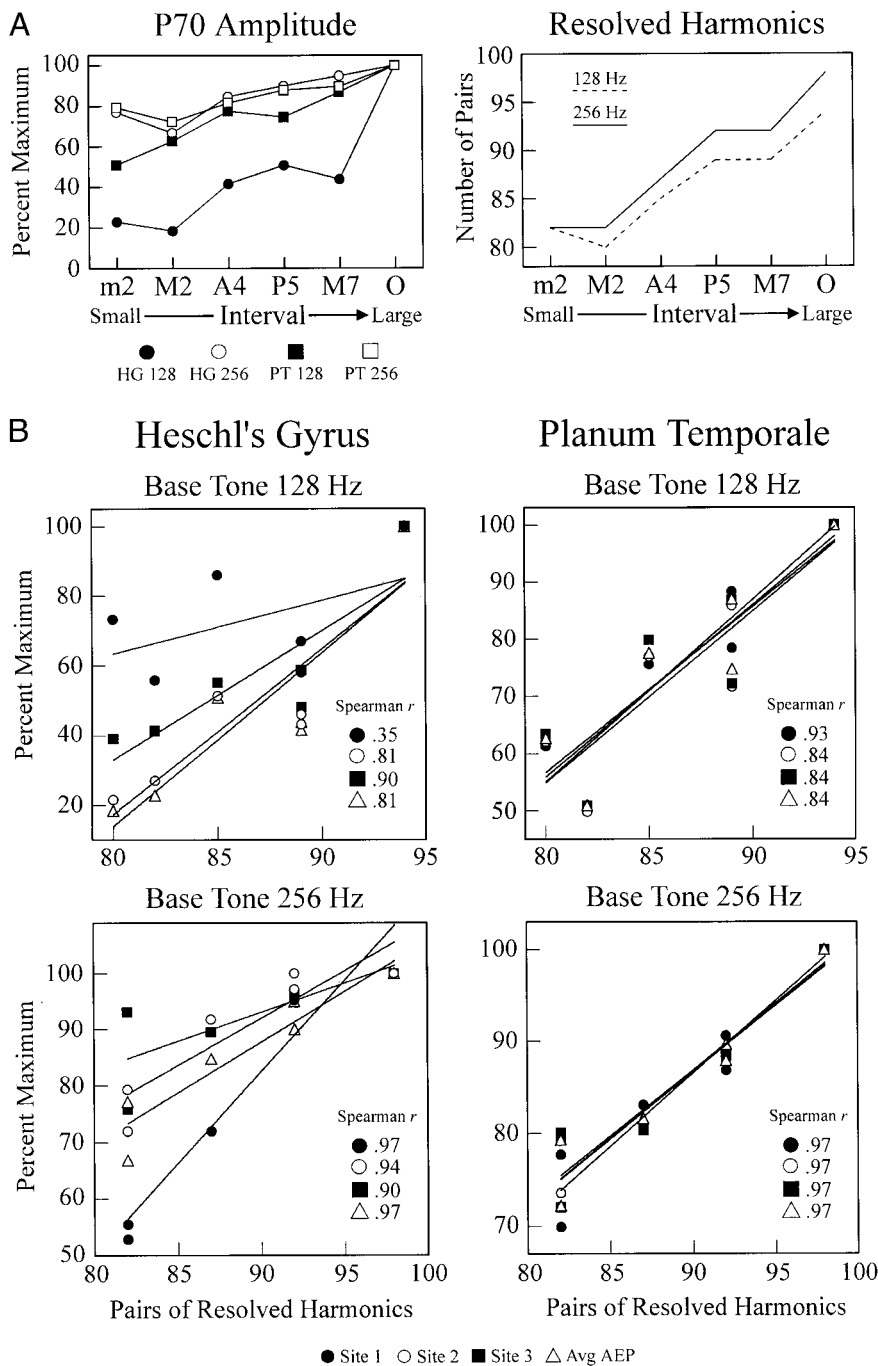


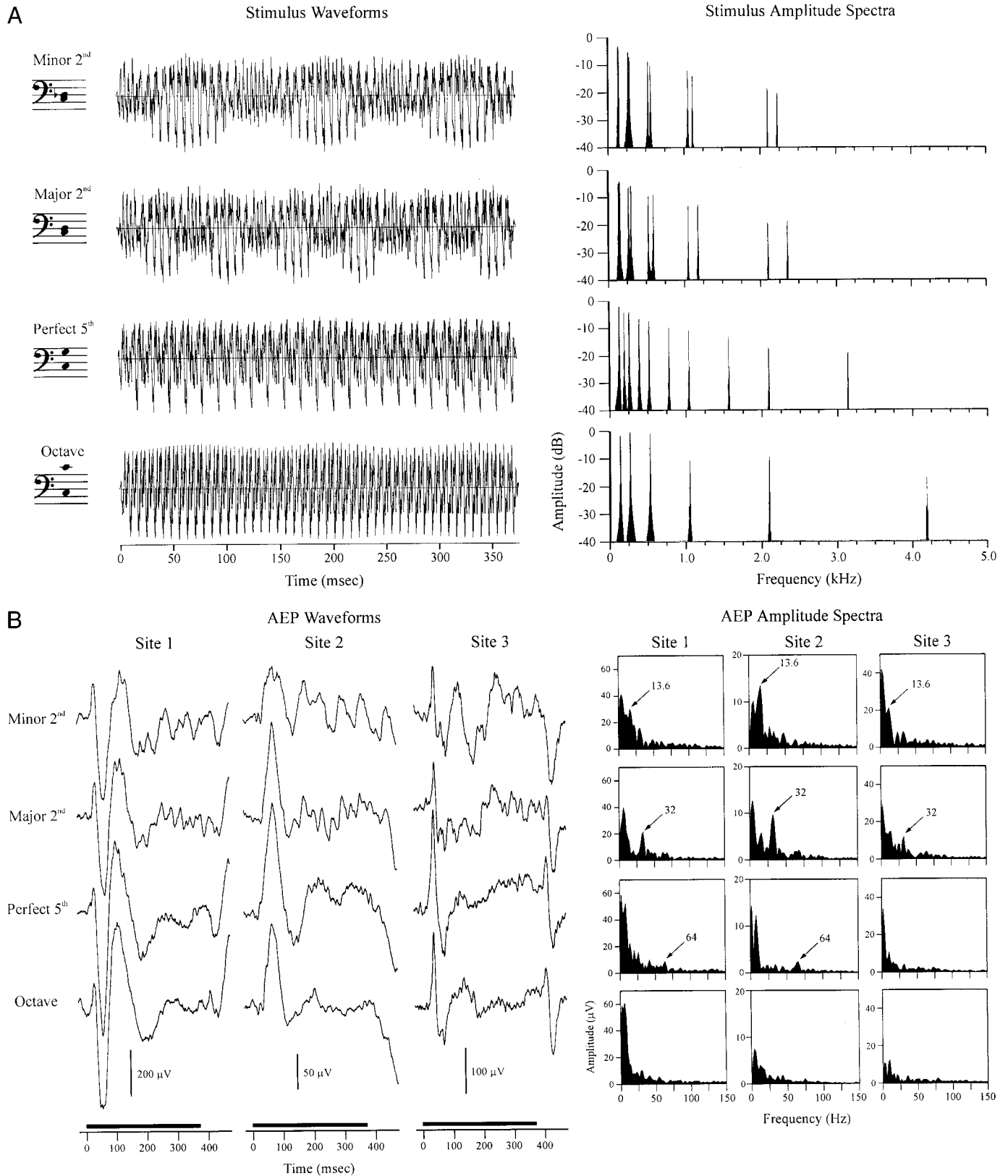
FIG. 17. *A, left*: amplitude of the P70 component of human intracranial AEPs (*subject 1*) as a function of musical interval width (only data for AEPs averaged across 3 recording sites are shown). Symbols representing data corresponding to AEPs evoked by chords presented in the 2 octave ranges and recorded in Heschl's gyrus and in the planum temporale are identified in the legend below. Amplitudes tend to increase with increasing interval width. *Right*: number of pairs of resolved harmonics in the chords as a function of interval width (---, intervals with base tones of 128 Hz; —, intervals with base tones of 256 Hz). *B*: normalized amplitude (percent maximum) of P70 as a function of the number of pairs of resolved harmonics in the chords. Heschl's gyrus and planum temporale data for the 2 octave ranges are represented in separate scatter plots, as indicated. Symbols representing data corresponding to AEPs recorded at each of the 3 electrode contacts located in Heschl's gyrus and the planum temporale are identified in the legend at the bottom of the figure.  $\Delta$ , data corresponding to AEPs averaged across the 3 recording sites. P70 amplitude tends to increase with the number of pairs of resolved harmonics comprising the chords. Spearman correlation coefficients are shown in the insets. Coefficients  $>0.83$  are statistically significant ( $P < 0.05$ ). Superimposed linear regression lines emphasize this relationship.

human intracranial investigations (Howard et al. 2000; Liegeois-Chauvel et al. 1994; Steinschneider et al. 1999) render untenable the often used assumption that auditory cortical organization can be elucidated by modeling auditory cortical activity as a single dipole generator situated within the superior temporal gyrus. This conclusion, also emphasized by other investigators (Lütkenhöner and Steinsträter 1998; Schreiner 1998), highlights the necessity of direct intracranial recordings for the valid interpretation of results obtained using noninvasive physiological techniques.

Taken together, these findings provide further evidence for the involvement of A1 in the representation of roughness (Fishman et al. 2000a) and sensory dissonance via synchronized phase-locked activity. Consequently, this study offers

physiological support for Helmholtz's beats/roughness theory of sensory dissonance. Similar response patterns are evoked both by Pythagorean and equal temperament chords. As a result of dividing the octave into 12 equal semitones, in the equal temperament system none of the intervals (except for the octave) are defined by simple  $f_0$  ratios and are thus slightly "out-of-tune," compared with their  $f_0$  ratios in the Pythagorean system (see Apel 1972). For example, whereas the  $f_0$ s of tones comprising a fifth in the Pythagorean system are related by the ratio 3:2, this ratio becomes 2.996:2 in the equal temperament system. As a result, additional beats are introduced. However, in the case of the fifth, for instance, these beats are so slow ( $<3$  Hz) that they are barely perceptible (for short-duration sounds) or at least do not contribute to roughness, thus allowing for the

### Equal Temperament Intervals





## Human Subject 2 Intracortical AEPs: Heschl's Gyrus

Equal Temperament Intervals (Base Tone 128 Hz)

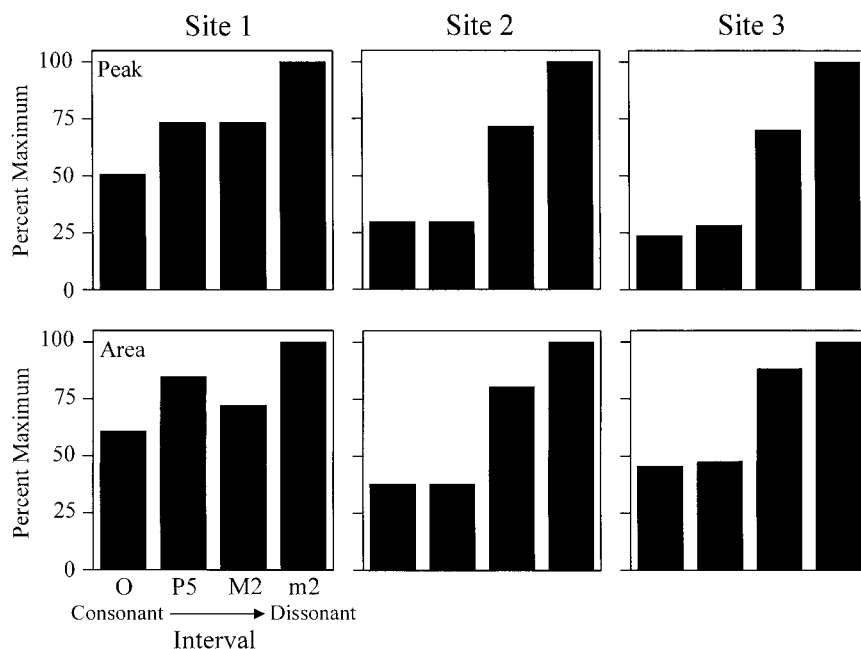


FIG. 19. Normalized peak and area of the amplitude spectra (10–150 Hz) of AEPs evoked by equal temperament chords at the 3 sites in Heschl's gyrus of *subject 2* shown in Fig. 18. Both spectral measures tend to increase with increasing dissonance of the intervals.

standard use of the equal temperament system in Western music for the past 200 years.

#### Relationship to other physiological studies

Several other investigators (e.g., Bieser and Muller-Preuss 1996; Schulze and Langner 1997; Steinschneider et al. 1998) have proposed that neural activity in A1 phase-locked to the amplitude-modulated temporal envelope of complex sounds may represent a physiological correlate of roughness. Using neuronal ensemble recording techniques identical to those used in the present study, Fishman et al. (2000a) tested this hypothesis and demonstrated a correlation between the magnitude of phase-locking to the AM frequency (= difference frequency) of harmonic complexes in A1 of the awake monkey and the perceived roughness of the stimuli as measured in human psychoacoustic experiments (Terhardt 1968a,b, 1974a,b). The upper limit of detectable phase-locking corresponded with the upper perceptual limit for the detection of roughness (200–300 Hz). These limiting rates are comparable to those obtained in other physiological studies of A1 using unanesthetized animals (e.g., de Ribaupierre et al. 1972; Goldstein et al. 1959; Steinschneider et al. 1998).

An additional finding of the present study was that the amplitude of the P70 component of the intracranial AEP recorded both in Heschl's gyrus and in the planum temporale tended to increase with increasing width of the musical intervals. This amplitude enhancement correlated with the number of pairs of resolved frequency components comprising the stimuli and is likely a result of critical band filtering. While analogous response enhancements were not observed in monkey A1 in the present study, such critical band-related amplitude increases were previously observed in monkey A1 responses to three-component harmonic complex tones with center frequencies fixed at the BF (Fishman et al. 2000b). It is

possible that the variable spectral distance between the partials of the chords and the BF of the site prevented such effects from being observed in monkey A1. In contrast, the comparatively diminished spectral specificity of the human AEPs, combined with the fact that they reflect the activity of larger and more widespread neuronal populations, may have facilitated the appearance of critical band effects.

The near linear relationship observed between stimulus resolvability and the amplitude of the P70 component of AEPs recorded in the planum temporale, coupled with their comparatively large amplitude, suggests that such critical band effects are likely to be visible in AEPs recorded from the scalp. Indeed, critical band-related increases in the amplitude of middle latency components of the scalp-recorded AEP evoked by two-tone stimuli have been previously demonstrated (Burrows and Barry 1990). P70 amplitude may reflect the perceived loudness of the chords since, due to critical band masking, complex sounds with spectrally resolved components are perceived as louder than sounds with unresolved components (Zwicker et al. 1957).

#### Proposed mechanisms

Since all of the (Pythagorean) musical chords used in the present study contained the same number of harmonics, and hence the same total number of difference frequencies, the question remains why the magnitude of oscillatory phase-locked activity in A1 correlates with the perceived dissonance of the intervals. The present findings can be partly explained by comparing the number of difference frequencies <250 Hz in consonant intervals with that in dissonant intervals. These represent difference frequencies capable of producing phase-locked responses in A1, as reported in previous investigations using awake macaques (Fishman et al. 2000a; Steinschneider et al. 1998). Conversely, difference frequencies above this range rarely, if ever, produce phase-locked responses in A1.

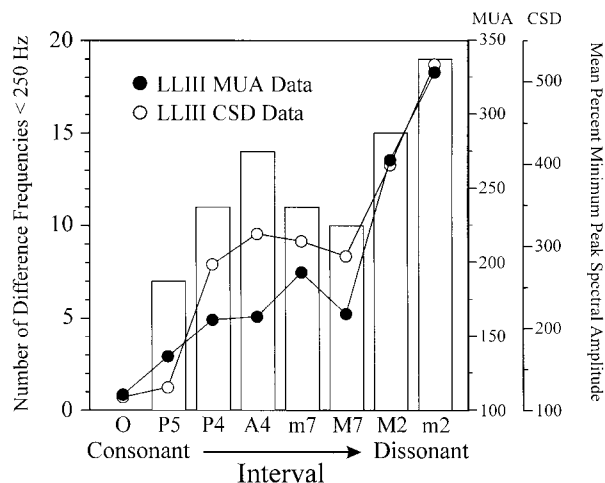


FIG. 20. Hypothesized basis for the correlation observed between the magnitude of oscillatory phase-locked activity in A1 and the perceived dissonance of the musical chords (base tone = 256 Hz). To generate phase-locked responses in A1, stimulus difference frequencies must be <250 Hz. The number of chord difference frequencies <250 Hz is shown as a function of the dissonance of the chords (histogram). Dissonant intervals tend to have more difference frequencies <250 Hz than consonant intervals. LLIII MUA and CSD data from Fig. 8 are shown superimposed on the histogram. Mean normalized peak spectrum amplitudes are significantly correlated with the number of difference frequencies in the chords <250 Hz (MUA:  $r = 0.85$ ;  $P < 0.01$ ; CSD:  $r = 0.93$ ;  $P < 0.001$ ).

Consonant intervals have far fewer difference frequencies <250 Hz than dissonant intervals (see Fig. 20). Thus the probability of cortical phase-locking to the difference frequencies of consonant intervals is considerably lower than that of phase-locking to the difference frequencies of dissonant intervals. Accordingly, the magnitude of chord-evoked oscillatory phase-locked activity in monkey A1 (as represented by the peak of the response amplitude spectrum) is correlated with the number of difference frequencies in the chords lying below 250 Hz (Pearson correlation: 256-Hz data from Fig. 8, LLIII MUA:  $r = 0.85$ ;  $P < 0.01$ ; LLIII CSD:  $r = 0.93$ ;  $P < 0.001$ ; see Fig. 20). The same reasoning can be used to explain why phase-locked oscillations are evoked by the perfect fourth when the base tone of the chord is at 256 Hz but largely disappear when the base tone is at 512 Hz. When the base tone is at 256 Hz, the two lowest difference frequencies of the perfect fourth are 85.35 and 170.7 Hz, whereas these difference frequencies are doubled when the base tone is at 512 Hz (170.7 and 341.4 Hz), values close to and exceeding the upper frequency limit of cortical phase-locking.

Given that subcortical auditory neurons within the inferior colliculus and medial geniculate nucleus are capable of phase-locking to stimulus periodicities well in excess of 250 Hz (e.g., Langner 1992; Langner and Schreiner 1988; Rouiller and de Ribaupierre 1982; Rouiller et al. 1979), it is possible that the correlation observed between oscillatory activity and sensory dissonance is determined by phase-locking limitations of auditory cortical neurons, i.e., the difference frequencies of consonant intervals are generally too high to be represented by phase-locked discharges at the cortical level. The marked deterioration in temporal encoding of rapid stimulus amplitude fluctuations at the cortical level may not only account for the consonance of musical intervals characterized by small-integer  $f_0$  ratios, but may also explain why some human subjects with

lesions of auditory cortex report that music that sounded pleasant prior to the lesions, sounds dissonant or out-of-tune (Tramo et al. 1990). In the absence of a functional auditory cortex, such brain-damaged subjects may be evaluating musical intervals on the basis of phase-locked activity generated in subcortical auditory structures. As a result, perfect fifths and octaves, which would be capable of eliciting robust phase-locked responses subcortically, could sound dissonant. An examination of chord-evoked responses in subcortical auditory areas is required to test this hypothesis and thereby assess the relative contribution of cortical and subcortical structures to the physiological representation of sensory consonance and dissonance.

### Concluding comments

While the present results provide strong support for a physiological representation of sensory dissonance in A1, they cannot account for other fundamental aspects of consonance and dissonance perception. First, the present study provides no information regarding the neural substrates underlying emotional responses to consonant versus dissonant chords (e.g., Blood et al. 1999). Second, owing to the limited applicability of the beats/roughness theory of dissonance, our findings do not explain why melodic intervals (sequential tone pairs), which do not produce beats or roughness owing to their non-simultaneity, are also evaluated along the dimension of consonance/dissonance according to their frequency ratios (e.g., Ayres et al. 1980; Schellenberg and Trehub 1994). This capacity has also been demonstrated in human infants, suggesting an innate predisposition to discriminate harmonically related from harmonically unrelated sounds (Demany and Armand 1984; Schellenberg and Trehub 1994, 1996). Moreover, the beats/roughness theory, and by extension our physiological data, are unable to explain why intervals extending beyond an octave, for which the large frequency separation between components precludes the generation of roughness, are judged with respect to their consonance and dissonance similarly to within-octave intervals. As proposed by Terhardt (1974b, 1977, 1978), these abilities, based on an evaluation of harmonicity, may have their origin in pitch analysis mechanisms, whereby a complex tone is assigned a global pitch corresponding to its  $f_0$  via harmonic pattern recognition. Alternatively, as suggested by Schellenberg and Trainor (1996), these capacities may arise from learning which simultaneous combinations of sounds give rise to sensory dissonance and generalizing these acquired rules to nonsimultaneous and supra-octave contexts. Having identified potential neural substrates underlying a fundamental perceptual attribute of music, the present study offers the possibility that these more complex features of music perception may also be amenable to physiological investigation.

We thank Dr. Steven Walkley, M. Huang, L. O'Donnell, Dr. Elena Zotova, and S. Seto for providing excellent technical, secretarial, and histological assistance. We also thank Dr. John Brugge and two anonymous reviewers for providing helpful comments on an earlier version of the manuscript.

This research was supported by National Institutes of Health Grants DC-00657, DC-042890-02, and NS-07098.

### REFERENCES

APEL W. *Harvard Dictionary of Music* (2nd ed.). Cambridge, MA: Harvard Univ. Press, 1972.

- AYRES T, AESCHBACH S, AND WALKER EL. Psychoacoustic and experiential determinants of tonal consonance. *J Aud Res* 20: 31–42, 1980.
- BARNA J, AREZZO JC, AND VAUGHAN HG JR. A new multicontact array for the simultaneous recording of field potentials and unit activity. *Electroencephalogr Clin Neurophysiol* 52: 494–496, 1981.
- BIESER A AND MULLER-PREUSS P. Auditory responsive cortex in the squirrel monkey: neural responses to amplitude-modulated sounds. *Exp Brain Res* 108: 273–284, 1996.
- BLOOD AJ, ZATORRE RJ, BERMUDEZ P, AND EVANS AC. Emotional responses to pleasant and unpleasant music correlate with activity in paralimbic regions. *Nat Neurosci* 2: 382–387, 1999.
- BROSCHE M, BAUER R, AND ECKHORN R. Stimulus-dependent modulations of correlated high-frequency oscillations in cat visual cortex. *Cereb Cortex* 7: 70–76, 1997.
- BURNS EW AND WARD WD. Intervals, scales, and tuning. In: *The Psychology of Music*, edited by Deutsch D. New York: Academic, 1982, p. 241–269.
- BURROWS DL AND BARRY SJ. Electrophysiological evidence for the critical band in humans: middle-latency responses. *J Acoust Soc Am* 88: 180–184, 1990.
- BUTLER JW AND DASTON PG. Musical consonance as musical preference: a cross-cultural study. *J Exp Psychol* 79: 129–142, 1968.
- CRONE NE, MIGLIORETTI DL, GORDON B, SIERACKI JM, WILSON MT, UEMATSU S, AND LESSER RP. Functional mapping of human sensorimotor cortex with electrocorticographic spectral analyses. I. Alpha and beta event-related desynchronization. *Brain* 121: 2271–2299, 1998.
- DEMANY L AND ARMAND F. The perceptual reality of tone chroma in early infancy. *J Acoust Soc Am* 76: 57–66, 1984.
- DE RIBAUPIERRE F, GOLDSTEIN MH JR, AND YENI-KOMSHIAN G. Cortical coding of repetitive acoustic pulses. *Brain Res* 48: 205–225, 1972.
- ECKHORN R, FRIEN A, BAUER R, WOELBERN T, AND KEHR H. High frequency (60–90 Hz) oscillations in primary visual cortex of awake monkey. *Neuroreport* 4: 243–246, 1993.
- FANNIN HA AND BRAUD WG. Preference for consonant over dissonant tones in the albino rat. *Percept Mot Skills* 32: 191–193, 1971.
- FISHMAN YI, RESER DH, AREZZO JC, AND STEINSCHNEIDER M. Complex tone processing in primary auditory cortex of the awake monkey. I. Neural ensemble correlates of roughness. *J Acoust Soc Am* 108: 235–246, 2000a.
- FISHMAN YI, RESER DH, AREZZO JC, AND STEINSCHNEIDER M. Complex tone processing in primary auditory cortex of the awake monkey. II. Pitch versus critical band representation. *J Acoust Soc Am* 108: 247–262, 2000b.
- FREEMAN JA AND NICHOLSON C. Experimental optimization of current source density technique for anuran cerebellum. *J Neurophysiol* 38: 369–382, 1975.
- GALABURDA AM AND PANDYA DN. The intrinsic architectonic and connectional organization of the superior temporal region of the Rhesus monkey. *J Comp Neurol* 221: 169–184, 1983.
- GALABURDA A AND SANIDES F. Cytoarchitectonic organization of the human auditory cortex. *J Comp Neurol* 190: 597–610, 1980.
- GOLDSTEIN MH JR, KIANG NY-S, AND BROWN RM. Responses of the auditory cortex to repetitive acoustic stimuli. *J Acoust Soc Am* 31: 356–364, 1959.
- GRAY CM AND SINGER W. Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex. *Proc Natl Acad Sci USA* 86: 1698–1702, 1989.
- HACKETT TA, STEPNIWSKA I, AND KAAS JH. Subdivisions of auditory cortex and ipsilateral cortical connections of the parabelt auditory cortex in macaque monkeys. *J Comp Neurol* 394: 475–495, 1998.
- HELMHOLTZ HLF. *On the Sensations of Tone as a Physiological Basis for the Theory of Music*, edited by Ellis AJ. New York: Dover, 1954 (original work published 1863).
- HOWARD MA, VOLKOV IO, ABBAS PJ, DAMASIO H, OLLENDIECK MC, AND GRANNER MA. A chronic microelectrode investigation of the tonotopic organization of human auditory cortex. *Brain Res* 724: 260–264, 1996a.
- HOWARD MA, VOLKOV IO, GRANNER MA, DAMASIO HM, OLLENDIECK MC, AND BAKKEN HE. A hybrid clinical-research depth electrode for acute and chronic in vivo microelectrode recording of human brain neurons. *J Neurosurg* 84: 129–132, 1996b.
- HOWARD MA, VOLKOV IO, MIRSKY R, GARELL PC, NOH MD, GRANNER M, DAMASIO H, STEINSCHNEIDER M, REALE RA, HIND JE, AND BRUGGE JF. Auditory cortex on the human posterior superior temporal gyrus. *J Comp Neurol* 416: 79–92, 2000.
- HULSE SH, BERNARD DJ, AND BRAATEN RF. Auditory discrimination of chord-based spectral structures by European starlings (*Sturnus vulgaris*). *J Exp Psychol* 124: 409–423, 1995.
- IZUMI A. Japanese monkeys perceive sensory consonance of chords. *J Acoust Soc Am* 108: 3073–3078, 2000.
- KAMEOKA A AND KURIYAGAWA M. Consonance theory. I. Consonance of dyads. *J Acoust Soc Am* 45: 1451–1459, 1969a.
- KAMEOKA A AND KURIYAGAWA M. Consonance theory. II. Consonance of complex tones and its calculation method. *J Acoust Soc Am* 45: 1460–1469, 1969b.
- LANGNER G. Periodicity coding in the auditory system. *Hear Res* 60: 115–142, 1992.
- LANGNER G AND SCHREINER CE. Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanisms. *J Neurophysiol* 60: 1799–1822, 1988.
- LIEGEOIS-CHAUVEL C, MUSOLINO A, BADIÉ JM, MARQUIS P, AND CHAUVEL P. Evoked potentials recorded from the auditory cortex in man: evaluation and topography of the middle latency components. *Electroencephalogr Clin Neurophysiol* 92: 204–214, 1994.
- LIEGEOIS-CHAUVEL C, PERETZ I, BABAI M, LAGUITTON V, AND CHAUVEL P. Contribution of different cortical areas in the temporal lobes to music processing. *Brain* 121: 1853–1867, 1998.
- LÜTKENHÖNER B AND STEINSTRÄTER O. High-precision neuromagnetic study of the functional organization of the human auditory cortex. *Audiol Neurootol* 3: 191–213, 1998.
- MALMBERG CF. The perception of consonance and dissonance. *Psychol Monogr* 25: 93–133, 1918.
- MATSUBARA JA AND PHILLIPS DP. Intracortical connections and their physiological correlates in the primary auditory cortex (AI) of the cat. *J Comp Neurol* 268: 38–48, 1988.
- MERZENICH MM AND BRUGGE JF. Representation of the cochlear partition on the superior temporal plane of the macaque monkey. *Brain Res* 50: 275–296, 1973.
- MITANI A, SHIMOKOUCHI M, ITOH K, NOMURA S, KUDO M, AND MIZUNO N. Morphology and laminar organization of electrophysiologically identified neurons in the primary auditory cortex in the cat. *J Comp Neurol* 235: 430–447, 1985.
- MOREL A, GARRAGHTY PE, AND KAAS JH. Tonotopic organization, architectonic fields and connections of auditory cortex in macaque monkeys. *J Comp Neurol* 355: 437–459, 1993.
- NICHOLSON C AND FREEMAN JA. Theory of current source density analysis and determination of conductivity tensor for anuran cerebellum. *J Neurophysiol* 38: 356–368, 1975.
- OJIMA H, HONDA CN, AND JONES EG. Patterns of axon collateralization of identified supragranular pyramidal neurons in the cat auditory cortex. *Cereb Cortex* 1: 80–94, 1991.
- PANDYA DN AND ROSENE DL. Laminar termination patterns of thalamic, callosal, and association afferents in the primary auditory area of the Rhesus monkey. *Exp Neurol* 119: 220–234, 1993.
- PERETZ I, BLOOD AJ, PENHUNE V, AND ZATORRE R. Cortical deafness to dissonance. *Brain* 124: 928–940, 2001.
- PERETZ I, KOLINSKY R, TRAMO M, LABRECQUE R, HUBLET C, DEMEURISSÉ G, AND BELLEVILLE S. Functional dissociations following bilateral lesions of auditory cortex. *Brain* 117: 1283–1301, 1994.
- PLOMP R AND LEVELT WJM. Tonal consonance and critical bandwidth. *J Acoust Soc Am* 38: 548–560, 1965.
- PLOMP R AND STEENEKEN HJM. Interference between two simple tones. *J Acoust Soc Am* 43: 883–884, 1968.
- RECANZONE GH, GUARD DC, AND PHAN ML. Frequency and intensity response properties of single neurons in the auditory cortex of the behaving macaque monkey. *J Neurophysiol* 83: 2315–2331, 2000.
- ROULLER EM AND DE RIBAUPIERRE F. Neurons sensitive to narrow ranges of repetitive acoustic transients in the medial geniculate body of the cat. *Exp Brain Res* 48: 323–326, 1982.
- ROULLER EM, DE RIBAUPIERRE Y, AND DE RIBAUPIERRE F. Phase-locked responses to low frequency tones in the medial geniculate body. *Hear Res* 1: 213–226, 1979.
- ROULLER EM, SIMM GM, VILLA AEP, DE RIBAUPIERRE Y, AND DE RIBAUPIERRE F. Auditory cortico-cortical interconnections in the cat: evidence for parallel and hierarchical arrangement of the auditory cortical areas. *Exp Brain Res* 86: 483–505, 1991.
- SHELLENBERG EG AND TRAINOR LJ. Sensory consonance and the perceptual similarity of complex-tone harmonic intervals: tests of adult and infant listeners. *J Acoust Soc Am* 100: 3321–3328, 1996.
- SHELLENBERG EG AND TREHUB SE. Frequency ratios and the discrimination of pure tone sequences. *Percept Psychophys* 56: 472–478, 1994.

- SCHELLENBERG EG AND TREHUB SE. Natural musical intervals: evidence from infant listeners. *Psychol Sci* 7: 272–277, 1996.
- SCHREINER CE. Spatial distribution of responses to simple and complex sounds in the primary auditory cortex. *Audiol Neurootol* 3: 104–122, 1998.
- SCHREINER CE AND URBAS JV. Representation of amplitude modulation in the auditory cortex of the cat. I. The anterior auditory field (AAF). *Hear Res* 21: 227–241, 1986.
- SCHULZE H AND LANGNER G. Periodicity coding in the primary auditory cortex of the Mongolian gerbil (*Meriones unguiculatus*): two different coding strategies for pitch and rhythm? *J Comp Physiol [A]* 181: 651–663, 1997.
- STEINSCHNEIDER M, RESER D, FISHMAN F, AND AREZZO J. Click train encoding in primary auditory cortex of the awake monkey: evidence for two different mechanisms subserving pitch perception. *J Acoust Soc Am* 104: 2935–2955, 1998.
- STEINSCHNEIDER M, SCHROEDER CE, AREZZO JC, AND VAUGHAN HG JR. Speech-evoked activity in primary auditory cortex: effects of voice onset time. *Electroencephalogr Clin Neurophysiol* 92: 30–43, 1994.
- STEINSCHNEIDER M, TENKE CE, SCHROEDER CE, JAVITT DC, SIMPSON GV, AREZZO JC, AND VAUGHAN HG JR. Cellular generators of the cortical auditory evoked potential initial component. *Electroencephalogr Clin Neurophysiol* 84: 196–200, 1992.
- STEINSCHNEIDER M, VOLKOV IO, NOH D, GARELL PC, AND HOWARD MA III. Temporal encoding of the voice onset time phonetic parameter by field potentials recorded directly from human auditory cortex. *J Neurophysiol* 82: 2346–2357, 1999.
- STERIADE M, AMZICA F, AND CONTRERAS D. Synchronization of fast (30–40 Hz) spontaneous cortical rhythms during brain activation. *J Neurosci* 16: 392–417, 1996.
- TERHARDT E. Über akustische rauhigkeit und schwankungsstärke. *Acustica* 20: 215–224, 1968a.
- TERHARDT E. Über die durch amplitudenmodulierte sinustöne hervorgerufene hörempfindung. *Acustica* 20: 210–214, 1968b.
- TERHARDT E. On the perception of periodic sound fluctuations (roughness). *Acustica* 30: 201–213, 1974a.
- TERHARDT E. Pitch, consonance, and harmony. *J Acoust Soc Am* 55: 1061–1069, 1974b.
- TERHARDT E. The two component theory of musical consonance. In: *Psychophysics and Physiology of Hearing*, edited by Evans EF and Wilson JP. London: Academic, 1977, p. 381–390.
- TERHARDT E. Psychoacoustic evaluation of musical sounds. *Percept Psychophys* 23: 483–492, 1978.
- TRAMO MJ, BHARUCHA JJ, AND MUSIEK FE. Music perception and cognition following bilateral lesions of auditory cortex. *J Cognit Neurosci* 2: 195–212, 1990.
- VAUGHAN HG JR AND AREZZO JC. The neural basis of event-related potentials. In: *Human-Event Related Potentials, EEG Handbook, Revised Series*, edited by Picton TW. New York: Elsevier, 1988, vol. 3, p. 45–96.
- WALLACE MN, KITZES LM, AND JONES EG. Chemoarchitectonic organization of the cat primary auditory cortex. *Exp Brain Res* 86: 518–526, 1991a.
- WALLACE MN, KITZES LM, AND JONES EG. Intrinsic inter- and intralaminar connections and their relationship to the tonotopic map in cat primary auditory cortex. *Exp Brain Res* 86: 527–544, 1991b.
- WHITFIELD IC. Auditory cortex and the pitch of complex tones. *J Acoust Soc Am* 67: 644–647, 1980.
- ZATORRE RJ. Pitch perception of complex tones and human temporal-lobe function. *J Acoust Soc Am* 84: 566–572, 1988.
- ZENTNER MR AND KAGAN J. Perception of music by infants. *Nature* 383: 29, 1996.
- ZWICKER E, FLOTTORP G, AND STEVENS SS. Critical bandwidth in loudness summation. *J Acoust Soc Am* 29: 548–557, 1957.