ELECTROPHYSIOLOGICAL STUDIES OF AUDITORY CONTEXTS

by

PETR JANATA

A DISSERTATION

Presented to the Department of Biology
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

August 1996

"Electrophysiological Studies of Auditory Contexts," a dissertation prepared by Petr Janata in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Biology. This dissertation has been approved and accepted by:

Dr. Janis C. Weeks, Chair of the Examining Committee

Date

Committee in charge:	Dr. Janis C. Weeks, Chair Dr. Terry T. Takahashi Dr. Daniel Udovic Dr. Nathan J. Tublitz Dr. William M. Roberts Dr. Michael I. Posner

Accepted by:

Vice Provost and Dean of the Graduate School

An Abstract of the Dissertation of

Petr Janata for the degree of Doctor of Philosophy

in the Department of Biology to be taken August 1996

Title: ELECTROPHYSIOLOGICAL STUDIES OF AUDITORY CONTEXTS

Approved: ______________________________________
Dr. Terry T. Takashashi

Memories of previous sensory input and accumulated knowledge of how the sensory environment behaves are capable of shaping our perceptions of incoming sensory information. Similarly, moment-to-moment sensory input is capable of reshaping stored representations, especially when the recent information doesn't match our expectations. In this dissertation, I studied neural correlates of perceptual and cognitive processes in the auditory system to (a) understand how harmonic sounds are represented by populations of neurons in the auditory midbrain and how these representations are influenced by the contexts in which the sounds appear and (b) find neurophysiological correlates of the processes by which stored representations of harmonic sounds are activated to generate expectations, or mental images, of specific physical information.

To address the first issue, neural representations of incoming auditory information were studied by recording the responses of neurons in the central nucleus of the barn owl's inferior colliculus (ICc) to harmonic sounds. While consisting of multiple frequencies, a harmonic sound is perceived as having a single pitch corresponding to its fundamental frequency, even when the fundamental frequency is physically absent. Recordings of single neurons and populations of neurons showed that the ICc neurons encode the fundamental frequencies of harmonic sounds and combinations of harmonic sounds in their temporal pattern of firing. Additional experiments showed that the neural representation of a harmonic sound depended on the context in which the sound occured. If a particular sound was embedded in a rapidly repeating sequence of different sounds, the neural response was altered relative to the response observed when the sound was played in isolation or when it formed the stream of rapidly repeating sounds.

The second topic was approached by recording the brain electrical activity at the scalps of musicians as they listened to simple eight note melodies and imagined portions of these melodies. The patterns of evoked brain activity were very similar for heard and imagined musical events, suggesting that the act of imagining a specific musical event activates auditory cortex. Additionally, the results implicate auditory cortex as a site at which external and internal representations of pitch are compared.

CURRICULUM VITA

NAME OF AUTHOR: Petr Janata

PLACE OF BIRTH: Ann Arbor, Michigan

DATE OF BIRTH: September 2, 1967

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon
Reed College

DEGREES AWARDED:

Doctor of Philosophy in Biology, 1996, University of Oregon
Bachelor of Arts in Interdisciplinary Studies (Biology/Psychology), 1990, Reed College

AREAS OF SPECIAL INTEREST:

Neural Basis of Expectation. Auditory/Music Perception and Cognition

PROFESSIONAL EXPERIENCE:

Research assistant, Department of Biology, University of Oregon, Eugene, 1992–96

Teaching assistant, Department of Biology, University of Oregon, Eugene, 1991–92

Fulbright Fellow, Institute of Neurophysiology, University of Vienna, Austria, 1990–91

AWARDS AND HONORS:

Co-Recipient of Ed Arbas Memorial Award for Best Graduate Student Paper Presented at Western Nerve Net, Eugene, Oregon, 1996
Reviewer for Journal of the Acoustical Society of America, 1996
Reviewer for Journal of Cognitive Neuroscience, 1995
McDonnell Summer Institute in Cognitive Neuroscience Fellow, University of California, Davis, 1994
Fulbright Fellowship to Vienna, Austria, 1990–91

Summer Fellow. R. S. Dow Neurological Sciences Institute, Good Samaritan Hospital & Medical Center, Portland, Oregon, 1989

Deutscher Akademischer Austauschdienst (DAAD) Scholarship. Univ. der Bundeswehr, Munich, Germany, 1987-88

International Association for the Exchange of Students for Technical Experience (IAESTE) Scholarship. German Cancer Research Institute, Heidelberg, Germany, 1987

PUBLICATIONS:

Janata, P. (1994) Differential processing of musical expectancies as assayed by measures of brain-electrical activity. *Soc. Cognitive Neuroscience Abstracts* 1: pg.21

Janata, P. (1995) ERP measures assay the degree of expectancy violation of harmonic contexts in music. *J. Cog. Neurosci.* 7(2): 153-164.

Janata, P. (1996) Electrophysiological imaging of imagined musical events. *Soc. Cognitive Neuroscience Abstracts* 3:122.

Janata, P. & Petsche, H. (1993) Spectral analysis of the EEG as a tool for evaluating expectancy violations of musical contexts. *Music Perception* 10: 281-304.

Janata, P. & Reisberg, D. (1988) Response-time measures as a means of exploring tonal hierarchies. *Music Perception.* 6(2): 163-174.

Janata, P. & Takahashi, T.T. (1995) Neural responses to spatio-temporal pitch contexts in the barn owl. *Soc. Neurosci. Abst.* 21:673.

Janata, P. & Winn, T. (1995) Event-related brain potentials in response to imagined auditory events. *Soc. Music Perception & Cognition Abst.* pg. 20.

Janata, P.; Keller, C.H. & Takahashi, T.T. (1994) Pitch coding in the barn owl's inferior colliculus: Responses to complex harmonic sounds. *Soc. Neurosci. Abstr.* 20:319.

Keller, C.H.; Takahashi, T.T. & Janata, P. (1993) A precedence effect in the owl's auditory space map? *Soc. Neurosci. Abst.* 19:531.

Potje-Kamloth, K.; Janata, P.; Janata, J. & Josowicz, M. (1989) Electrochemical encapsulation for sensors.*Sensors and Actuators.* 18:613-623.

Potje-Kamloth, K.; Janata, P. & Josowicz, M. (1989) Carbon fiber microelectrodes. In Book of Proceedings of ElectroFinnAnalysis Conference. Plenum Press.

Takahashi, T.T.; Keller, C.H. & Janata, P. (1993) Resolving multiple sound sources in the owl's midbrain. *Soc. Neurosci. Abst.* 19:531.

ACKNOWLEDGMENTS

I thank Katie Henry and Terry Takahashi for suggestions on Chapter I, Terry Takahashi, Kip Keller and Peter Cariani for their criticisms and discussions of Chapter II, Terry Takahashi for his criticisms of Chapter III, and Terry Takahashi, Michael Posner, and Ramesh Srinivasan for comments on Chapter IV. Tom Winn was helpful in the acquisition of data for Chapter IV. I am deeply grateful to Terry Takahashi for causing me to continually question the meaning of my work, Michael Posner for many helpful suggestions and encouragement, and Ramesh Srinivasan for intriguing discussions about the EEG and help with spectral analysis methods. I am particularly indebted to Kip Keller for his tireless, patient and expert assistance in preparation and execution of experiments contributing to Chapters II and III, not to mention his many valuable insights.

The research was supported by NIH Training Grant 5T32 GM07257. Support was also provided by NIDCD Grant DC 02050 and a Whitehall Foundation grant to Terry Takahashi, and the McDonnell/Pew Center for the Cognitive Neuroscience of Attention.

TABLE OF CONTENTS

Chapter Page

LIST OF FIGURES

Page

CHAPTER I

ON THE IMPORTANCE OF DETECTING AND UNDERSTANDING TRANSIENT CHANGES IN THE ENVIRONMENT

Central to an organism's survival is its ability to detect, orient to, process, and respond to unexpected events in its surrounding environment. For instance, sudden snaps of twigs, screeching tires and screams of many types all attract our attention. These sudden transitions from a steady-state of auditory affairs alert our attention so that a finer-grained analysis of the situation giving rise to the unexpected sound can be performed. The results of the finer analysis interact with systems that specify and activate behaviorally-appropriate motor programs. Without neural systems that are capable of detecting, orienting to, analyzing sudden events with respect to ongoing contexts, and responding appropriately to these events, organisms would be helpless in the task of organizing information about the world around them.

Analyses of the meanings (directions) of detected sudden changes are implied in the procedural description of what an organism has to do to survive. If such analyses did not take place, organisms would respond to sudden alterations in the physical world indiscriminately. There would be no knowledge of whether the sudden change represents an alteration that is consonant with, or predictable by ongoing physical processes, or instead represents the entry of a new process into the organism's environment. A task of a nervous system, therefore, is to organize observations about discrete sensory transitions into patterns (contexts) that are capable of predicting with some degree of certainty the time and place of the next transition.

If similar environmental circumstances arise frequently, the organism benefits from possessing a mental model of the environment's behavior so that a once unexpected stimulus can be classified as normal and not elicit an orienting response or a reallocation of attentional resources. The acquired model generates expectations about changes in both the quality and quantity of future sensory input. One of the best studied examples of a rather complex neural model of an expectation to date, is that of the "corollary discharge" in mormyrid electric fish (Bell, 1989). These fish image their environments by sensing the electrical field distribution across their bodies that arises both as a result of their emitting regular electrical pulses as well as the presence of other electrical field generators in their environment. In order to maintain discriminability of objects in the environment, these fish send a "negative" copy (the corollary discharge) of the input pattern that they expect to receive in response to their own emitted electrical field to the regions of the midbrain involved in processing the electrosensory information. Thus, expectations are created to filter sensory input, and these expectations are dynamically modified as electric field-deforming objects in the fish's environment come and go.

Expectations are defined as predictions of the future state or sequence of states of some sensory parameter. Although they might be rather complicated, as in the electric fish example, they can be as simple as the expectation to hear a low frequency tone following a high frequency tone in a series of alternating high and low tones. Expectations can also signal invariance, such as the repetition of a single tone. Since the number of auditory patterns that we and other organisms learn to associate with physical events in our environments is vast, it can be assumed that we elaborate many intricate mental models that are capable of integrating incessant fluctuations in air pressure into a managable flow of auditory information. Examples of such models are species specific vocalizations that serve in communication such as speech, music, or bird song. Additionally, we form representations of other auditory patterns such as the sounds of our car and other machines. All of these representations serve to signal sequences of state transitions in the physical processes giving rise to the sounds.

The overall goal of my research is to understand the neural mechanisms underlying the prediction of auditory events and detection of auditory events that do not conform to expectations. As suggested above, the successful prediction of auditory events requires knowledge about sequences (patterns) of change that typify specific auditory situations. These situations range from "textures" that are created by physical processes such as "gurgling" streams or "humming" engines to arrangements of musical notes into simple rhythmic patterns, melodies or quartets. While these auditory situations may seem disparate, they all involve patterning of auditory information in time. The question then becomes, "How are patterns of auditory information represented and stored by the brain?" The role of this introductory chapter is to illuminate key issues that bridge the investigations of context-dependent neural representations of harmonic sounds reported in Chapters 2 and 3 with the study of musical processing reported in Chapter 4. The notion will be developed that the temporal properties of auditory input are represented in the brain on a number of time scales, and that the temporal structure in combination with the mapped neural representations of physical parameters such as frequency or spatial location serves to perceptually organize the auditory environment.

The Role of Time in Shaping the Perception of Sound Arrangements

Three issues are of particular importance in discussing the processing of auditory objects and their relationships by the brain. The first pertains to the different time scales across which naturally-occurring auditory events are distributed. The second is a consideration of perceptual correlates of patterns of discrete events transpiring on these different time scales. Finally, the sensitivity of different anatomical levels of the brain to events transpiring on different time scales may help define how organization of information into hierarchies of percepts is achieved. Thus, time is of the essence. Other essentials are the concepts of pitch and pitch patterns. These factors are important because they provide a common language for describing the rather different experiments that will be presented in Chapters 3 and 4. Even though Chapter 4 investigates brain processes associated with imagining specific notes in melodies on a time scale of seconds, while Chapter 3 looks at the processing of sequences of harmonic sounds in the hundreds of milliseconds range, both experimental designs can be formulated as studies of brain responses to actual or inferred pitch patterns that occur on defined time scales.

Throughout this dissertation, different time scales will be referred to in terms of the inter-stimulus interval (ISI), defined as the amount of time separating onsets of successive events. The definition of pitch is a little more confusing. Pitch itself is a percept rather than a physical entity (Moore, 1989), meaning that different configurations of frequencies can give rise to a qualitatively unitary percept. Most commonly, pitch is identified with single frequencies, i.e., pure tones such as 440 Hz. However, the percept of a 440 Hz pitch can also be elicited by a set of frequencies that are higher harmonics (integer multiples) of the 440 Hz fundamental frequency. In fact, a pitch of 440 Hz can be perceived even if the fundamental is absent from the set of harmonics in a phenomenon known as "perception of the missing fundamental," "periodicity pitch," or "pitch of the residue" (Licklider, 1951; deBoer, 1976; Houtsma & Smurzynski, 1990). Nonetheless, a pitch is identified by its fundamental frequency, be it created by a single pure tone or by a harmonic complex. It is extremely important to remember, however, that the neural representation of "pitch" is likely to be very different when it is created by a pure tone or by a set of related higher harmonics of a fundamental frequency. This is because different frequencies maximally stimulate different regions of the basilar membrane and are consequently represented in different locations on tonotopic maps in the central nervous system. Thus, a 400 Hz pure tone will activate those portions of tonotopic maps that best respond to 400 Hz. However, a harmonic sound consisting of 4000, 4400, and 4800 Hz will activate a higher-frequency region of the tonotopic map, yet it will still give rise to the same pitch percept as the 400 Hz tone. Similarly, a harmonic sound consisting of 5600, 6000, and 6400 Hz will activate yet a different region of the tonotopic map. Only in the event that a neural map of "pitch", as opposed to tonotopy, exists in any given brain area, might the three stimuli be expected to activate the same neurons. The ability to experimentally "place" harmonic sounds onto different regions of the tonotopic map will be exploited in the experiments of Chapters 2 and 3.

Substantial evidence has accumulated that perceptions of the pitch of complex sounds are based on analyses of the temporal properties of neural activity (for review, see Langner, 1992; Cariani & Delgutte, 1996a,b). The principle tenet of this view is that the temporal patterning of the neural response is governed by the stimulus waveform. Periodicities in neural firing patterns correlate closely with periodicities (the times between major peaks) in the stimulus's waveform. Thus, temporal intervals in the range of 1 to 10 msec are important for encoding information that is relevant to the formation of pitch percepts. The neural representations of the temporal patterns formed by individual harmonic sounds and combinations of harmonic sounds are the central theme of Chapter 2. Chapter 2 demonstrates that the midbrain of the barn owl represents the pitch associated with harmonic sounds in the temporal pattern of firing of populations of neurons. Information about the response properties can then be used to investigate whether the neural representations of harmonic sounds are modifiable by the context in which they occur. In this case, contexts are sequences of harmonic sounds that are defined by their fundamental frequency, spectrum and time interval between successive sounds in the sequence.

Pitch patterns are arrangements of individual pitch elements in time. The perception of pitch patterns, as well as patterns of non-pitch sounds such as noise bursts, depend strongly on the relative timing among pattern elements as well as the overall rate at which the patterns are presented. For example, the elements in a pattern might be separated by 50 msec (relative timing), and an overall 4-element pattern might be presented at an overall rate of one pattern every 500 msec. This is an example of a hierarchical organization of sounds into auditory objects and sequences of auditory objects. A particularly useful synthesis of the psychological literature on the perceptual correlates of pitch patterns occurring on different time scales is provided by Warren (1993). When each element is shorter than 150–200 msec, the order of sounds in a set of sounds that continuously repeats cannot be identified (Warren et al., 1969; Warren & Obusek, 1972), even though patterns with different orderings of elements can be discriminated from each other. Thus, when ISIs are in the range of 100–200 msec, patterns of sounds are perceived as unified objects that might be discriminated based on their "texture", but not by direct identification of their individual elements (Warren & Ackroff, 1976). When harmonic speech sounds (vowels) of different lengths are arranged in sequences, 3 distinct percepts appear at different ranges of item durations (Warren et al., 1990; Warren, 1993). Below 30 msec, different arrangements of vowels differ in their timbre but the order of the vowels can not be determined. Above 100 msec, subjects succeed in naming vowel order. Between vowel durations of 30 and 100 msec, however, different vowel arrangements are heard as different words. In Chapter 3, we will see that neurons of the barn owl inferior colliculus are most sensitive to transitions among harmonic stimuli when the interstimulus intervals are under 200 msec. Even though the experiments in Chapter 3 do not address the behavioral discrimination of harmonic sound objects by barn owls, the commonalty between time scales over which transitions are readily signaled by the owl midbrain and time scales over which specific auditory percepts are formed in humans is suggestive that perceptual organization of auditory events occurring on these time scales may derive in part from patterns of inferior colliculus output.

The relative timing of individual elements in a pitch pattern plays a critical role in specification of the pattern identity. This is demonstrated by performance decrements in judging pitch pattern similarity in the face of altered temporal relations among the pitches when ISIs range between 80 and 130 msec (Sorkin, 1987). A salient example of the importance of timing information about auditory relations comes from phoneme perception in language-learning impaired children (Merzenich et al., 1996).

At longer ISIs, individual elements can be easily identified and distinguished from each other. More importantly, knowledge about relations among the pitches is preserved and guides expectations and attention that are specific both for pitch and time of occurrence (Jones et al., 1982; Jones & Yee, 1993). Melodies are a simple and familiar example of pitch patterns that create expectations for specific notes over time-scales ranging from a few hundred milliseconds to several seconds. Chapter 4 will look at brain processes underlying the generation of images (expectations) of pitches in simple melodies.

In summary, psychophysical experiments have amply demonstrated that patterns in auditory information are organized into relational structures on a number of time scales. In order to establish the relational structure, it is imperative to know the nature of the transitions, e.g. increases or decreases in frequency, intensity, duration, pitch, or changes in spectral envelope, and the times at which they occur. What, then, is known about the neural mechanisms underlying the perception of pitch patterns or the detection of auditory transitions that occur on various time scales?

Neurophysiological Studies of Auditory Transience Detection

Most studies to date of neural correlates of auditory transience detection have used relatively simple sounds and patterns (contexts). The simplest auditory context that incorporates a stimulus transition is the repetition of a single frequency (pure tone) at a fixed ISI with an occasional interposition of a tone that is higher or lower in frequency. Behavioral and neurophysiological studies of responses to such sequences have been cast in terms of habituation to repeating stimuli and orientation to "deviant" stimuli (Sokolov, 1960; Näätänen, 1992), where deviance is usually defined in terms of stimulus probability. It is within this tradition that the experiments described in Chapter 3 were formulated and will be discussed. Although the language used throughout the literature is in terms of "deviant" stimuli and the detection of and orienting to potentially threatening circumstances in the environment, the studies can just as easily be interpreted within the more neutral and extensible framework of transition detection in auditory pattern perception.

The size of responses in the cochlear nucleus, inferior colliculus, and medial geniculate nucleus of cats decreases in response to successive clicks or tones in a repeating series. This is true for ISIs of up to 10 seconds (Simons et al, 1966; Webster, 1971; Kitzes & Buchwald, 1969), although the size of the response decrement depends both on the ISI and the brain region from which the response is recorded. Webster (1971) found 10% and 35% response decrements to clicks presented at ISIs of 5 sec and 100 msec, respectively, in the inferior colliculus of cats and a 50% response decrement in the 100 msec ISI condition in the thalamus. Responses to such sequences are also readily measured at the human scalp, with the first stimulus in the sequence eliciting a large negative deflection in the event-related potential occurring at around 85 msec (N100). The responses to successive presentations of the same tone diminish by 75 - 90% at ISIs of 1 or 3 seconds (Ritter, Vaughan & Costa, 1968; Frühstorfer, Soveri & Järvilehto, 1970). The N100 diminishes in amplitude at shorter ISIs (Hari et al, 1982). Typical ISIs used in human studies are 500 msec and longer.

It is important to note that the time-course and magnitude of response habituation at longer ISIs, such as 5 or 10 sec, differs across brain levels. Evoked responses in the brain stem tend to show a shallow and slow linear decay at the these ISIs, while auditory cortex shows a rapid and large response decrement to a steady plateau. At ISIs of 100 msec, however, the response decrement function of cat inferior colliculus across successive events (Webster, 1971) appears more similar to the N100 decay function for human cortex, suggesting that some of the same neural principles underlying the response decrement might operate in both cases, but with different time frames. Given this observation, it is tempting to infer that given their differential sensitivities to auditory events transpiring at different rates, different brain levels may be responsible for organizing different auditory percepts of pitch patterns that depend on stimulus presentation rate.

The decrement in neural response to a repeating stimulus is typically thought of in terms of a developing memory trace for the physical parameters of the repeating stimulus against which incoming stimuli are matched and against which deviances are readily detected and can be oriented to (Sokolov, 1960; Näätänen, 1992). Habituation of responses to repetitive auditory stimuli is observed at all stages of auditory processing and for a wide range of ISIs, suggesting that each of these brain areas forms a "memory trace", or is at least influenced by such a representation of the repeating stimulus. How might a deviation from a repetitive sequence affect the neural response?

Neural responses to auditory transitions have been studied primarily with the auditory evoked response recorded at the human scalp using the "oddball paradigm." In this paradigm, random sequences are constructed from two sounds that differ along a physical dimension, in our case frequency, and are assigned a global probability of occurrence, typically 90% for the "standard" and 10% for the "deviant." Deviant stimuli make their mark on averaged scalp-recorded evoked potentials recorded in two ways. The lower the probability of a stimulus, the larger a positive wave that occurs around 300 msec after the stimulus onset (Tueting et al., 1971). Countless studies in many cognitive and perceptual domains have demonstrated that contextual deviances within attended informational streams profoundly affect the morphology of the later phases of averaged evoked potentials. Among these studies are several illustrating sensitivity to violations of predictions created by musical contexts formed by patterns of single pitches (Besson & Faita, 1995; Johnston, 1994; Cohen & Erez, 1991; Paller et al, 1992; Verleger, 1990; Besson & Macar, 1987) or pitch combinations (Janata, 1995) at ISIs ranging from 200 msec to 1000 msec.

Other studies have focused on an earlier component (100–250 msec post stimulus onset) of the event-related potential that assumes more negative values when a physical mismatch is registered. The mismatch negativity (MMN) is of particular importance to us because it has been interpreted as reflecting the comparison of an incoming sensory stimulus with a memory trace of the stimulus/stimulus sequence preceding that stimulus (Näätänen, 1992). If the physical parameters of the sensory stimulus don't match the physical parameters represented in the memory trace, a response is generated to reflect the detection of a mismatch, and this response is manifested at the human scalp as an enhanced negativity. This process is postulated to occur in the auditory cortex and to occur automatically and in the complete absence of attention (see Näätänen, 1992, for an extensive review and discussion of these processes within an attentional framework). Within the framework of the standard auditory oddball paradigm, in which two tones of different frequency are used, a steady-state in the neural response is achieved by the repeating stimulus. A physically-differing stimulus constitutes a deviation from the steady-state, and results in an evoked response that signals the mismatch. The MMN has been demonstrated for ISIs of as long as 10 seconds in instances of rather large changes in frequency (Böttcher & Ullsperger, 1992). Common properties of the MMN are that regular intervals, larger physical deviances, and short ISIs facilitate a larger MMN (Näätänen, 1990, 1992).

The exact neural mechanisms underlying the comparison process indexed by the MMN are not well understood. In fact the purview of the MMN process has been forced to extend beyond simple physical comparisons of successive stimuli to include comparisons among temporal and other featural relationships of pitch patterns consisting of 2 to 8 elements presented at inter-element intervals of 40 to 400 msec (Saarinen et al, 1992; Winkler & Schröger, 1995, Tervaniemi et al., 1994; Woods & Alain, 1993). This means that the auditory cortex does not seem to be obligatorily involved in the detection of physical deviation between sequential elements provided that the sequence has an invariant higher-level organization. In other words, the cortex is freed from the task of detecting physical change and can instead occupy itself with detecting differences among patterns of physical change.

Direct neurophysiological evidence for sensitivity to auditory patterns comes from studies of single neuron responses to sequential combinations of sounds. Some neurons in cat auditory cortex are sensitive to patterns of pure tones in the sense that the response to a pure tone depends on the tone preceding it at ISIs of 300–900 msec (Weinberger & McKenna, 1988). Sensitivity to combinations of auditory elements in a behaviorally relevant context has been demonstrated for neurons in the neostriatum of songbirds (Margoliash, 1983; Margoliash & Fortune, 1992; Sutter & Margoliash, 1994).

The MMN itself has been studied in the cortex of monkeys (Javitt et al., 1992) and cats (Csépe et al., 1987). More recent studies suggest that MMN-like processes also occur in the medial geniculate region of the thalamus in guinea pigs in response to tones and some speech-like sounds (Kraus et al., 1994 a,b), and in the medial geniculate and inferior colliculus of cats in response to deviant tones (Csépe et al., 1993). However, the parameter range explored in the study of inferior colliculus response to deviants in an oddball paradigm was rather restricted.

The goal of Chapters 2 and 3 is to establish a means of quantifying neural activity recorded from the inferior colliculus in barn owls in response to complex harmonic sounds and to examine how these responses are altered when the harmonic sounds are imbedded in longer sequences of similar and dissimilar sounds. As such, these studies establish that certain types of transitions in auditory parameters affect neural responses in the midbrain and provide a framework for future investigations of how higher auditory centers such as the auditory thalamus or primary telencephalic auditory areas might use such information to form representations of more complicated auditory objects that involve sets of specific transitions. One example of auditory objects that extend over long periods of time are melodies, which can be defined as patterns of pitch changes. When we hear melodies that are familiar to us, we are able to imagine their continuation if they are abruptly truncated, indicating that we have stored knowledge of the next pitch or set of pitches are. The brain processes that are invoked by the mental continuation/prediction of auditory pattern elements (notes in melodies) are the topic of Chapter 4.

CHAPTER II

REPRESENTATIONS OF MULTIPLE HARMONIC SOUNDS IN THE MIDBRAIN OF THE BARN OWL: COMPARISON OF EVOKED-POTENTIAL AND SINGLE-UNIT RECORDINGS.

Abstract

Representations of harmonic sounds in the central nucleus of the inferior colliculus (ICc) of the barn owl, _Tyto_ _alba_, were studied using evoked-potential and single-unit recording techniques. Frequency spectra of the evoked potential recordings showed that the neural response robustly represented the periodicities in the harmonic stimuli. The periodicities observed included the fundamental frequency of a stimulus, even though it was absent from the stimulus itself, and several combination tones when pairs of harmonic sounds were presented simultaneously. The periodicities observed in the response of any given neuronal population depended on the position of the stimulus's constituent frequencies relative to the population's frequency tuning curve. The evoked potential was found to be a better indicator of the periodicity structure in the stimulus than were recordings of single neurons. A possible relationship between evoked potential recordings of population activity and the combined activity of many single neurons is discussed.

Introduction

Most natural sounds consist of energy in multiple frequency bands, with the energy in any given frequency band changing through time. Auditory systems are confronted with the task of representing the spectro-temporal properties of an auditory stimulus with sufficient accuracy so that subtle changes in pitch, timbre, loudness and other psychophysical dimensions can be discriminated.

Substantial physiological evidence suggests that the time-varying pattern of amplitude fluctuations present in the waveforms of harmonic stimuli such as speech sounds or amplitude-modulated pure tones is robustly represented in the synchronized firing of neurons in the auditory nerve, cochlear nucleus and midbrain of amphibians, fish, birds and mammals (for review, see Langner, 1992). A recent pair of studies (Cariani & Delgutte, 1996a,b) demonstrates strong similarities between the autocorrelation functions of auditory nerve firing and the autocorrelation functions of several types of periodic stimulus waveforms, thus providing compelling evidence that several pitch perception phenomena are correlated with the temporal patterning in auditory nerve discharges.

In the present report, we investigate the representations of multiple, spatially overlapping, harmonic sounds at the level of the ICc in the barn owl, Tyto alba. We compare the abilities of single neurons and neuronal populations to represent the temporal fine structure present in harmonic stimuli. Additionally, we probe the ICc's distributed representation of pairs of concurrently presented harmonic sounds.

The barn owl has served as a model system for the study of neurophysiological mechanisms underlying spatial hearing. The ICc, with its orthogonally mapped representations of inter-aural time difference and tonotopy (Wagner et al., 1987), is a particularly important structure in which to study the neural encoding of auditory objects. Since almost all auditory information ascending the nervous system passes through the inferior colliculus (Wild, 1987), the ICc can be thought of as a display of the complete auditory scene. This display provides the information from which subsequent auditory centers determine the spatial and spectral identities of, and relationships among, multiple auditory objects. Recent studies in our laboratory have investigated conditions under which neurons in the external nucleus of the inferior colliculus (ICx) that receive input directly from the ICc can discriminate spatially separated sources of noise (Takahashi & Keller, 1994; Keller & Takahashi, 1996a,b), and it has been proposed that the ability of ICx neurons to segregate concurrent auditory objects depends on the pattern of ICc activation (Takahashi & Keller, 1994). Harmonic sounds constitute stimuli of intermediate complexity between pure tones and white noise in their temporal and spectral properties, and as such provide a useful means for investigating the spatio-temporal representation of auditory objects on the space-frequency map in the ICc.

Methods

Harmonic Sounds

Complex harmonic sounds were digitally synthesized by summing sets of higher harmonics of fundamental frequencies ranging from 73 Hz to 1600 Hz. The number of harmonics (5–21 spectral components/sound) and the frequency range (1 kHz to 10 kHz) into which the harmonics fell were manipulated within and across experiments in 14 barn owls. For any given harmonic sound, the set of frequencies composing the sounds was selected to fall into a particular frequency band. For example, 3 harmonic sounds, all with a fundamental frequency (F_0) of 329 Hz, were synthesized with harmonics 7–12, 10–15 and 16–21. Consequently, these sounds fall into spectral bands of 2.3–4.0 kHz, 3.3–5.0 kHz and 5.2–6.9 kHz, respectively. The implication of falling into different spectral bands is that the three sounds will activate different populations of neurons distributed across the ICc's tonotopic axis. In no case was the fundamental frequency present in the stimulus itself. Note that the fundamental

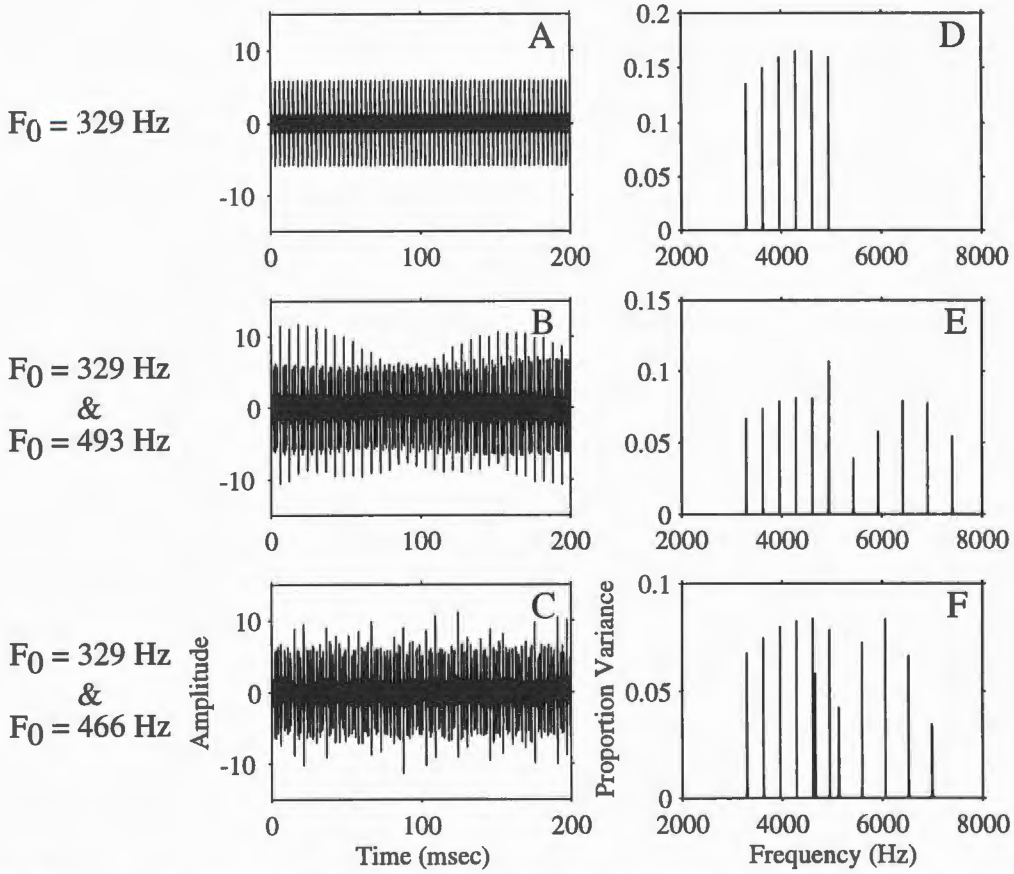


FIGURE 1. Examples of complex harmonic waveforms showing variation in their temporal fine structure. All stimuli in this experiment were harmonic complexes consisting of several higher harmonics of a particular fundamental frequency (F_0). The fundamental frequency itself was always omitted. The exact spectral components of any given harmonic complex were selected to fall into a desired spectral range, e.g. 4000 Hz to 6000 Hz. A single harmonic complex has a waveform with a period corresponding to that of the missing fundamental frequency. The temporal patterning in the waveform can be made more complicated by adding a second harmonic complex. (A) $F_0 = 329$ Hz. Harmonics 10–15. (B) Combination of two harmonic sounds with different fundamental frequencies. $F_0(1) = 329$ Hz & $F_0(2) = 493$ Hz. Harmonics 10–15 for both sounds. (C) A combination of harmonic sounds which creates a more complicated fine structure than that seen in B. $F_0(1) = 329$ Hz & $F_0(2) = 466$ Hz. Harmonics 10–15 for both sounds. (D-F) Normalized power spectra corresponding to the sounds in A–C. Proportion variance indicates the amount of variance in the overall waveform explained by each frequency bin.

frequency is the first harmonic. Higher harmonics are integer multiples of the fundamental frequency.

Several sets of harmonic sounds were constructed to form musical scales starting at fundamental frequencies corresponding to C3 (261Hz), E3 (329 Hz), or G3 (392 Hz). Harmonic sounds in the musical scales were all synthesized with a fixed set of harmonics (either harmonics 5–15 or 10–15). Harmonic sounds based on the fundamental frequencies of notes from the musical scale were either played alone (unisons) or in successive combinations of the first note in the scale with every other note (harmonic intervals). The frequency range of harmonics present in sounds corresponding to the bottom notes in a scale was typically lower and sometimes did not overlap the frequency range covered by the harmonics of a note higher in the scale. Examples of stimuli are shown in Fig. 1. Figs. 1A and 1D show the time and frequency domain representations of a harmonic sound consisting of harmonics 10–15 of a 329 Hz F_0. This stimulus represents a harmonic sound played in unison. Panels 1B and 1C show the waveforms that result when the 329 Hz F_0 harmonic sound is paired with harmonic sounds with F_0s of 493 Hz and 466 Hz respectively.

Digitized stimuli were converted to analog form at a rate of 50,000 samples/sec. The sounds were 200 msec in duration with linear 1.5-msec ramps at sound onset and offset. Following D/A conversion (Data Translations), attenuation (Tucker-Davis Technologies, PA4) and amplification (McIntosh, M754), the sounds were emitted from a single speaker (Alpine 6020HX) situated 70 cm in front of the owl. The intensity of the emitted sound was 20 to 30 dB above neuronal thresholds.

Electrophysiological Recording

Anesthetic and surgical procedures used in these experiments were approved by the Institutional Animal Care and Use Committee at the University of Oregon and have been described previously (Takahashi & Keller, 1994). An anesthetized (0.05 ml/h, 100 mg/ml Vetalar, Parke-Davis and 0.025 ml/h, 5 mg/ml Diazepam, C-IV, LyphoMed) adult owl from our aviary was maintained in a fixed position within a stereotaxic device by a stainless steel plate cemented to the skull. Recordings were made within a sound attenuating booth (Industrial Acoustics Co.) lined with acoustic foam (Ilbruck Sonex). Recordings of neural activity were made using tungsten microelectrodes (5-10 MΩ, Frederick Haer).

Electrodes were placed stereotactically relative to a center position located between ear bars placed in the owl's ear canals. Bursts of broadband noise (100 ms) were used to locate populations of auditory neurons. Neurons were identified as being located in the ICc if they showed distinct peaks in their frequency tuning curves, as well as some degree of spatial selectivity (determined by playing sounds from different positions along the azimuth).

Evoked neural activity was bandpass filtered (10 Hz–2.5 kHz, fourth order Butterworth filter) and digitized with 12-bit resolution at a rate of 8000 samples/second. Each stimulus was presented 10–30 times at interstimulus intervals of 1 or more seconds.

Data Analysis

For population recordings, the digitized traces were averaged offline. For single unit recordings, the entire digitized trace was stored and spike times were computed offline. Spike times were recorded as the time at which an arbitrary threshold value was crossed, and were viewed either as a raster or post-stimulus time histogram (PSTH). All analyses were performed using Matlab (Mathworks). The specific details of the analyses for each figure are presented in the text and figure captions. Briefly, power spectra of evoked potential recordings were computed using a fast-Fourier transform (FFT) routine for each individual repetition of the stimulus. The mean value of the time-series (DC offset) was subtracted from the time-series prior to the FFT computation. A

256 msec (2048 point) window encompassing the response to the stimulus and silent periods on either end was used for FFT computations. Power spectra from individual trials were then averaged across repetitions and normalized to the total power in the power spectrum. The value in each frequency bin of these normalized power spectra indicates the proportion of the overall variance contained in that frequency bin. For single unit recordings, power spectra were computed both as the average of power spectra of single-trial rasters and as the power spectrum of the PSTH derived from summing individual trial rasters.

In order to have a quantitative measure of how well single neuron or population evoked potential responses represented periodicities related to the fundamental frequency of the harmonic sound, the power in frequency bins corresponding to the fundamental frequency of the stimulus and up to 9 higher harmonics of the fundamental frequency was summed to determine the total amount of variance directly related to a representation of the stimulus. When two simultaneously-presented harmonic complexes served as the stimulus, the power in frequency bins corresponding to the 1st through 9th order "distortion products" was summed. The set of frequency bins corresponding to distortion products was determined by the formula $m \cdot F_0(1) \pm n \cdot F_0(2)$, where $F_0(1)$ and $F_0(2)$ are the two fundamental frequencies and m and n are integers ranging from one to nine.

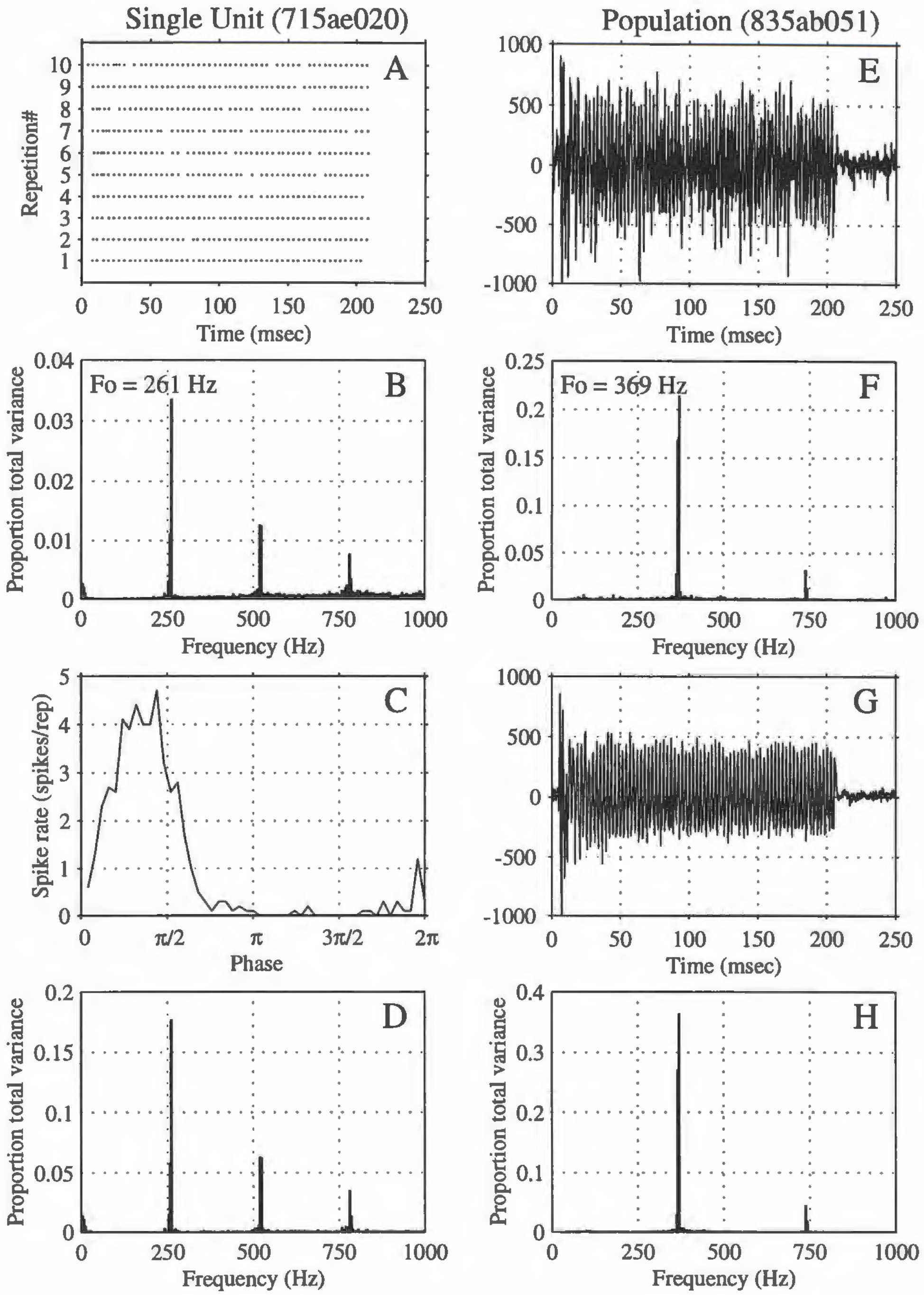
In order to determine the maximum values that might be expected to arise from this particular measure, the stimuli were half-wave rectified (to introduce the distortion products related to the fundamental frequency), low-pass filtered at 2.5 kHz, resampled at 8000 Hz, and subjected to the same measure of proportion of variance explained by frequency bins corresponding to the distortion products. The amount of variance explained for the stimuli ranged around 90%.

Results

Comparison of Single Unit and Population Responses

Both single neurons and populations of neurons showed periodicities in their firing patterns that could be related to periodicities in the stimulus. This was the case even when the stimulus periodicities were relatively complex, as was the case for combinations of harmonic sounds. Fig. 2 compares the responses of a single neuron (panels A-D) with a population of neurons (panels E-H) to harmonic stimuli with fundamental frequencies (F_0s) of 261Hz and 369Hz, respectively.

FIGURE 2. Comparison of single neuron and population responses evoked by harmonic complexes. Panels A–D show the response for a single unit in the ICc to a harmonic sound with a 261 Hz fundamental frequency (F_0). This neuron's response was amongst the most highly synchronized with the periodicities in the stimulus. (A) Raster plot of the neuron's responses on 10 repetitions of the stimulus. Each row of dots represents a single trial, and each dot marks the time of occurrence of a spike. Responses have been collapsed into 0.5 ms bins. (B) Average power spectrum of single-trial post-stimulus time histograms (PSTHs). A power spectrum was computed for each row in the raster plot and the 10 spectra were averaged. The ordinate is expressed as the proportion of the total variance in the PSTH explained by each frequency component in the spectrum. Approximately 12% of the single unit activity on a single repetition of the stimulus is related directly to the periodicities in the stimulus. (C) Period histogram of the single neuron response summed across 10 repetitions of the stimulus. The phase angle of the 261 Hz fundamental frequency period is plotted on the abscissa. The emergence of a peak around one phase angle indicates that the neuron is reasonably successful in representing the periodicity in the stimulus. The fact that only 12% of the activity is directly related to the stimulus means that although the neuron tends to fire at the same point in every cycle of the stimulus, it does not fire on every cycle. (D) The power spectrum of the PSTH derived by summing the raster plot in A across the 10 stimulus repetitions. Note the increase in the overall proportion of the variance explained (to ~45%) when single unit activity is first pooled across several repetitions. The increase in "stimulus related" variance is most easily explained by the presence of a spike at each cycle of the stimulus. Panels E–H show the response of an ICc neural population to a harmonic sound with a 369 Hz F_0. (E) Population response to a single presentation of the stimulus. (F) Average of power spectra computed on each of 20 repetitions. (G) Average evoked potential of 10 repetitions. (H) Power spectrum of averaged evoked potential shown in G. Note the larger proportion of variance explained in both the single-trial (50%) and multi-trial average (76%) power spectra for the population activity compared to the single neuron activity.


Single Unit (715ae020)
Population (835ab051)
A
Repetition#
10
9
8
7
6
5
4
3
2
1
Time (msec)
0
50
100
150
200
250
E
1000
500
0
-500
-1000
Time (msec)
B
Fo = 261 Hz
Proportion total variance
0.04
0.03
0.02
0.01
0
Frequency (Hz)
250
500
750
1000
F
Fo = 369 Hz
Proportion total variance
0.25
0.2
0.15
0.1
0.05
Frequency (Hz)
C
Spike rate (spikes/rep)
5
4
3
2
1
0
Phase
π/2
π
3π/2
2π
G
Time (msec)
D
Proportion total variance
0.2
0.15
0.1
0.05
Frequency (Hz)
H
Proportion total variance
0.4
0.3
0.2
0.1
Frequency (Hz)

It is important to note that although we refer to the stimuli used in this study by their fundamental frequencies, none of the stimuli in this study actually possessed energy at the fundamental frequency, i.e., they were all "missing fundamental" stimuli.

The raster plot in Fig. 2A shows that this neuron was relatively adept at firing on several successive cycles of the stimulus's fundamental frequency. The power spectrum was computed for each horizontal line of the raster and the average of these spectra plotted in Fig. 2B. The height of each spectral peak corresponds to the proportion of total variance in the single trial raster that falls into that frequency bin. The appearance of peaks in the 261 Hz frequency bin and higher harmonics thereof indicated that the neuron was able to follow the fundamental frequency of the stimulus on a single trial basis, although not perfectly. Integration of the power in the frequency bins that were related to the stimulus, and those bins immediately adjacent, showed that the total variance explained by stimulus-related frequency bins was 12% of the total variance in any given trial. The neuron's ability to phase-lock to the fundamental frequency is illustrated in the period-histogram shown in Fig. 2C. The ordinate indicates the number of times the neuron fired at any given phase angle of the fundamental frequency of the stimulus over the course of 10 stimulus repetitions. When a power spectrum is computed on a PSTH constructed from responses to 10 stimulus repetitions, the total variance explained by energy in frequency bins related to the stimulus increased to 45% (Fig. 2D). This increase in variance is explained most easily by the presence of some number of spikes at the same point on every cycle of the stimulus in the summed PSTH.

Panels E–H in Fig. 2 show the same set of analyses performed on a population evoked potential recorded in a different owl. The response to a single repetition of the stimulus is shown in Fig. 2E, and the average of single-trial power spectra in Fig. 2F. The average amount of variance related to the stimulus in the evoked potential on a single trial was 50%. In the evoked potential average of 10 repetitions of the stimulus (Fig 2G), 76% of the variance in the average evoked potential was attributable to stimulus-related frequencies.

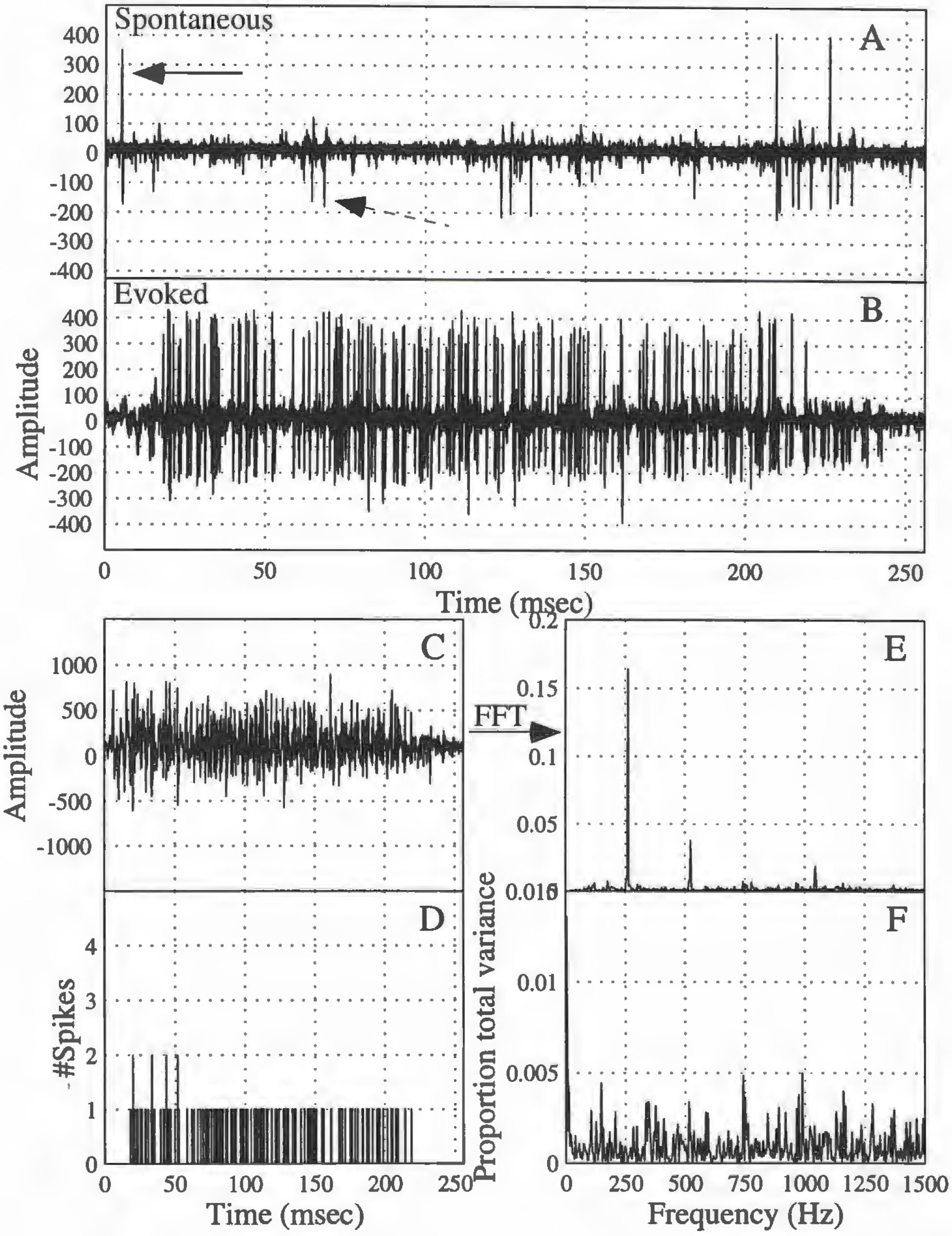
Typically, single neurons did not respond as well as the one shown in Fig. 2. When they did display a robust sensitivity to the periodicities in the stimulus, they would do so for only a narrow range of fundamental frequencies. In single unit recordings from 37 cells tested with harmonic stimuli ranging in fundamental frequency from 73 Hz to 783 Hz, the mean proportion of the overall variance related to the stimulus was 11.7% (± 2.6% s.d.) when spectra were calculated on single-repetition PSTHs (as in Fig. 2b). When the PSTH was constructed from 10–40 repetitions, and the frequency spectra computed on these summed PSTHs, the amount of overall variance explained increased to 19.1% (± 7.2%).

A larger proportion of the variance in population recordings could be assigned to frequency bands related to the stimulus than was the case for single unit recordings. For harmonic stimuli ranging in F_0 from 220 to 783 Hz the mean proportion of variance explained in frequency spectra of single trials of population recordings was 22.9 ± 7.8% (across 17 recording sites). As in the single unit recordings, if the single trial responses were first averaged together to form an average evoked potential, the amount of variance explained increased to 50.9 ± 11.5%. Thus, the proportion of variance explained by population recordings was more than twice the amount explained by single unit recordings. Furthermore, the proportion of variance explained by single trial evoked potentials (23%) was similar to the amount explained in single-unit PSTHs constructed from up to 40 repetitions (20%). Considering only those stimuli that were used in both single-unit and population recording experiments (F_0s between 466 and 783 Hz), the average amount of variance related to the stimulus explained by single neurons on individual trials was ~2%, while that explained by single trial evoked potentials was 21%. The highest F_0 in the evoked potential recordings for which a peak in the power spectrum was discernable above background noise was 1,359 Hz.

Derivation of "Evoked Potentials" From Single-Neuron Recordings

At several recording sites, the activity of several neurons could be seen in traces of spontaneous activity. This allowed us to compare the information about a harmonic stimulus contained in the activity of a single, easily discriminable neuron, with the aggregate activity of all the neurons detectable by the electrode. Fig. 3A shows ten superimposed segments of spontaneous activity in which the activity of a single neuron with large, positively-deflecting action potentials is clearly visible (solid arrow). The activity of several "smaller" units, most with negatively-deflecting action potentials, is also visible (e.g., dashed arrow). A harmonic sound with a fundamental frequency of 261 Hz (harmonics 7–15) evoked a robust response (Fig. 3B). We found that as the threshold settings were lowered to capture a larger number of positively and negatively deflecting action potentials, the energy in the spectra of the PSTHs redistributed into frequency bins related to the stimulus and ultimately resembled closely the spectrum of the averaged "evoked potential".

FIGURE 3. Comparison of information contained in evoked potentials and PSTHs. (A) Ten superimposed segments of spontaneous activity at a single recording site. The trace is dominated by a single neuron with a large spike (solid arrow), but the activity of numerous other individual neurons (e.g. dashed arrow) is evident as lower-amplitude and often negatively deflecting spikes. (B) Ten superimposed traces of the neurons' responses to a stimulus consisting of harmonics 7–15 (1827–3915 Hz) of a 261 Hz fundamental frequency. (C) Average of the 10 traces shown in B. (D) PSTH generated by detecting threshold crossings set to capture only the large positive-going deflection in the individual traces in B and summing the resulting spike rasters. (E) Normalized power spectrum of the voltage trace shown in C. The peaks fall at frequencies related to the fundamental frequency of the stimulus. (F) Normalized power spectrum of the PSTH shown in D. Energy is not preferentially concentrated at stimulus-related frequency bins.


Spontaneous
A
400
300
200
100
0
-100
-200
-300
-400
Evoked
B
Amplitude
0
50
100
150
200
250
Time (msec)
C
1000
500
-500
-1000
FFT
D
#Spikes
1
2
3
4
E
0.2
0.15
0.1
0.05
0.015
F
0.01
0.005
Proportion total variance
0
250
500
750
1000
1250
1500
Frequency (Hz)

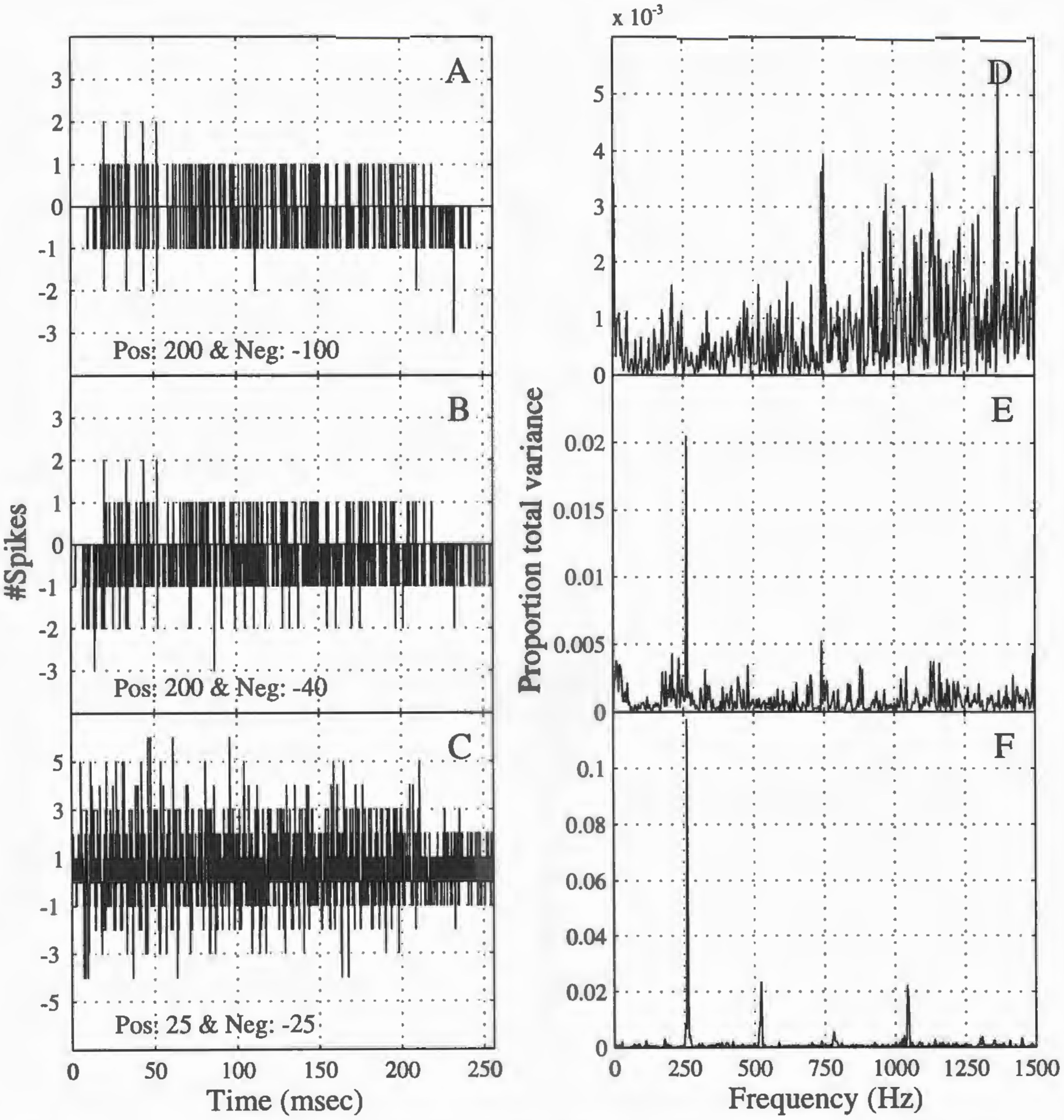


FIGURE 4. Derivation of an evoked potential from modified post-stimulus time histograms. Spike times were determined for the superimposed traces shown in Fig. 3B by setting threshold levels at several different positive and negative values to capture both positively and negatively deflecting spikes, respectively. Through the successive lowering of the positive and negative spike-detection thresholds, the activity of an increasing number of neurons is reflected in the PSTH. PSTHs were constructed by subtracting the PSTH for negative spikes from the PSTH of positive spikes, and power spectra were computed for these modified PSTHs. (A) Positive threshold = 200; Negative threshold = -100. (B) Positive threshold = 200; Negative threshold = -40. (C) Positive threshold = 25; Negative threshold = -25. (D–F) Normalized power spectra corresponding to the PSTHs in A–C, respectively.

The activity evoked by the 261 Hz stimulus was analyzed in two ways. In an evoked potential approach, the digitized traces recorded on ten repetitions of the stimulus were simply averaged together. The resulting waveform (Fig. 3C) was qualitatively similar to many evoked potentials evoked on single presentations of harmonic stimuli at recording sites where no clear single-unit activity was discriminable (e.g., Fig. 2E). The major peaks in the power spectrum of the averaged "evoked potential" fell at frequencies related to the fundamental frequency of the stimulus (Fig. 3E).

In contrast, the spectrum of a PSTH derived from isolating the activity of the large positively-deflecting spike on several repetitions of the stimulus (Fig. 3D) did not show discriminable peaks in frequency bins related to the fundamental frequency of the stimulus. Thus, on the basis of the activity of this single neuron, one would not necessarily infer that coding of periodicities related to the fundamental frequency is directly reflected in periodicities in neural firing.

If the unit activity of the background units present at this location was taken into account, aggregate PSTHs could be constructed that reflected the stimulus periodicity and resembled the "evoked potential". Fig. 4 shows the PSTHs and corresponding spectra that resulted as spike detection thresholds were set at several different positive and negative values to capture both positively and negatively deflecting spikes, respectively. PSTHs were constructed by subtracting the PSTH for negative spikes from the PSTH of positive spikes. Through successive lowering of the spike-detection threshold, the activity of an increasing number of neurons is reflected in the PSTH, and the corresponding spectrum resembles more closely that of the averaged "evoked potential" (Fig 4F).

Population Responses to Combinations of Harmonic Sounds

The complexity of the temporal patterning in the sound and, consequently, the neural response, could be manipulated by presenting combinations of harmonic sounds. The relationship of the two fundamental frequencies determined the spacing among the spectral components belonging to the two sounds. Sounds with a simple ratio relationship between the constituent F_0s, such as 2:1 (658 Hz:329 Hz) or 3:2 (493 Hz:329 Hz) have more harmonics (partials) in common and display more even spacings among spectral components than pairs of fundamental frequencies related by more complicated ratios, e.g. 8:5 (523 Hz:329 Hz). A series of harmonic sound pairs with a broad range of complexity in their ratio relationships is embodied in the set of harmonic intervals used in western tonal music. Intervals that form an octave (ratio of 2:1; 329 Hz and 658 Hz) or a fifth (ratio of 3:2; 329 Hz and 493 Hz, Figs. 1B,E) have a large number of overlapping or evenly spaced spectral components (harmonics), compared to intervals such as the major 2nd (329 Hz and 369 Hz) or tritone (329 Hz and 466 Hz, Figs. 1C,F) whose spectra show a variety of spacings among the component frequencies of the two individual sounds. It is expected that the larger the number of difference frequencies (spectral component spacings) present among spectral components of two harmonic sounds, the more complicated the temporal patterning of the sound waveform.

FIGURE 5. Population responses to a harmonic sound consisting of harmonics 10–15 of a 329 Hz fundamental frequency, played in isolation (A) or paired with a similarly constructed sound with a 493 Hz F_0 (B) or 466 Hz F_0 (C). (D–F) Power spectra corresponding to panels A–C. The proportion of the overall variance in the neural response explained by each spectral component is plotted on the ordinate. The smaller peaks in the spectra (such as the peak at 164 Hz in E) are the result of interactions of spectral components of the 2 harmonic sounds.

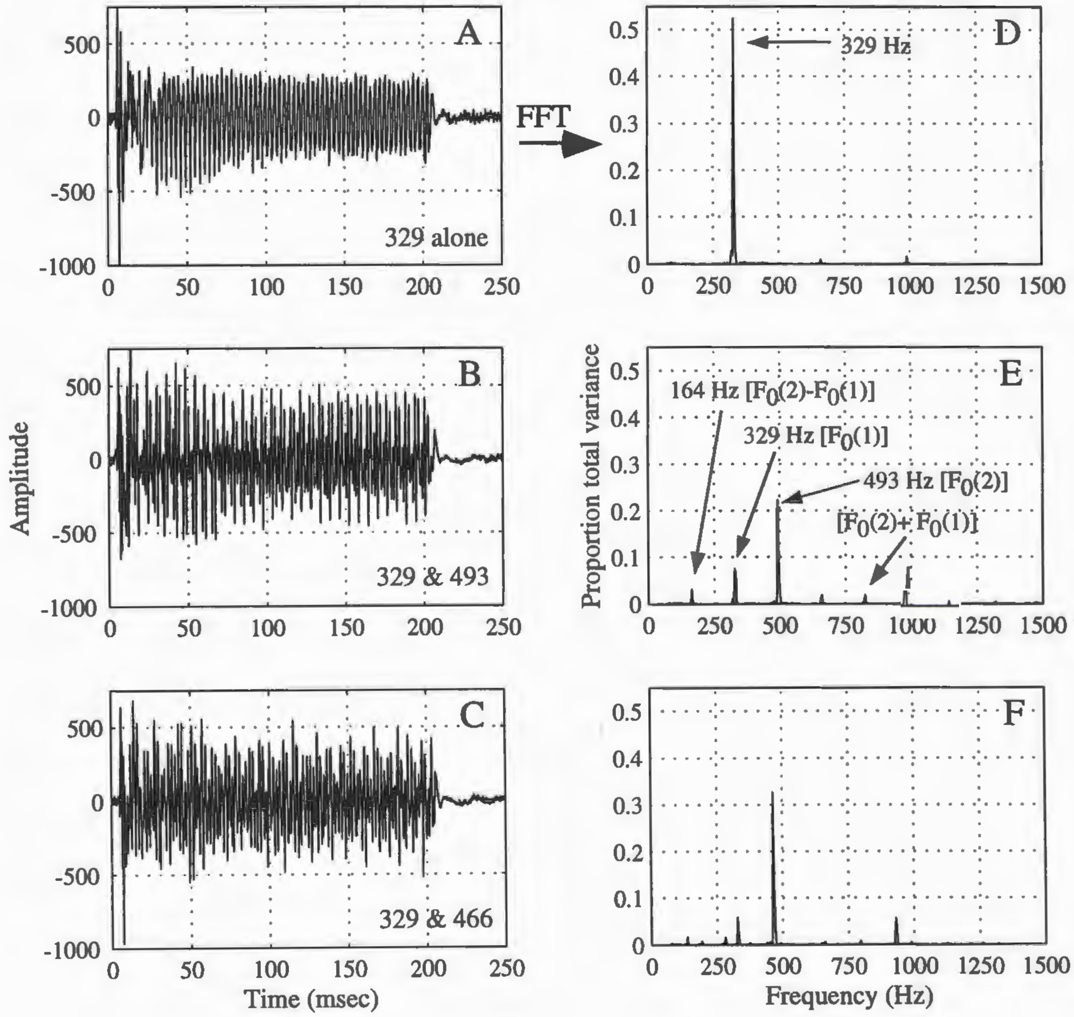
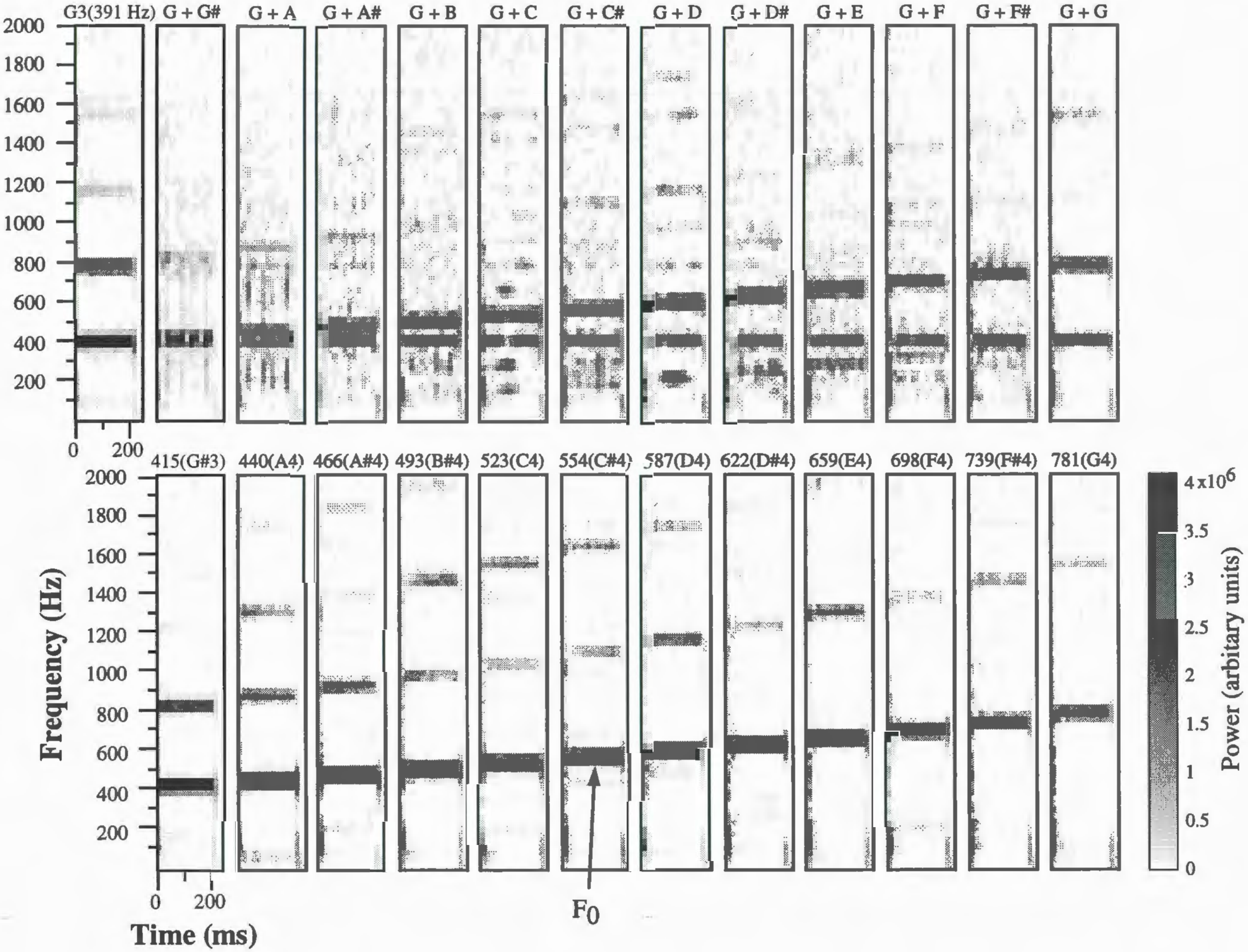

A
329 alone
FFT
D
329 Hz
B
329 & 493
E
164 Hz [$F_0(2)$-$F_0(1)$]
329 Hz [$F_0(1)$]
493 Hz [$F_0(2)$]
[$F_0(2)$+$F_0(1)$]
C
329 & 466
F
Amplitude
Proportion total variance
Time (msec)
Frequency (Hz)
500
0
-500
-1000
0
50
100
150
200
250
0.5
0.4
0.3
0.2
0.1
0
250
500
750
1000
1250
1500

Fig. 5 shows the response of a neural population to a 329 Hz F_0 stimulus presented in unison, i.e. alone, (Fig. 5A) and in combination with a closely related (F_0 = 493 Hz, ratio of F_0s = 3:2) harmonic sound (Fig. 5B) and with a distantly related (F_0 = 466 Hz, ratio of F_0s = 7:5) harmonic sound (Fig. 5C). The neural response is characterized by an onset response (6 msec post-stimulus) that is larger in amplitude than the tonic portion of the response. The sustained response shows an oscillation whose frequency is that of the fundamental frequency of the harmonic stimulus (Fig. 5D). When two harmonic sounds were presented together, frequency components corresponding to each of the fundamental frequencies were present in the neural response (Figs. 5E,F), together with frequency components that correspond to combination tones formed by the two F_0s. Note the differences in temporal fine structure in the time-domain traces of the neural responses to stimuli consisting of two harmonic sounds with different F_0s.

The different patternings in temporal fine structure that arise when harmonic sounds are presented in multiple combinations are readily observed in the spectrograms shown in Fig. 6. The spectrograms shown in Fig. 6 were computed from the average of evoked potentials recorded at five sites [center frequencies (CFs): 4,400, 4800, 5500, 6700, 7200 Hz] distributed along the tonotopic axis of the ICc, and as such represent the aggregate activity that would be observed in the ICc in response to these different stimulus situations. The leftmost panel in the top row is a spectrogram of the evoked potential response to a broad-band harmonic sound (harmonics 5–15) with the fundamental frequency of the musical note G3 (391 Hz). The bottom row of spectrograms shows the neural responses to harmonic sounds whose fundamental frequencies are based on the notes of an ascending chromatic scale, i.e., each successive key on a piano keyboard, terminating an octave above (G4, 782 Hz) the starting point.

FIGURE 6. Spectrograms of neural responses to harmonic sounds presented alone (unisons, bottom row) and as pairs (harmonic intervals, top row). Each panel is a spectrogram of the average of evoked potentials recorded from 5 locations along the tonotopic axis of the ICc. Spectrograms were computed using a 32 msec sliding window, with 50% overlap between successive windows. 13 harmonic sounds were used, each consisting of harmonics 5–15 of fundamental frequencies starting at 391 Hz and continuing through 12 equally spaced steps to an octave (782 Hz) above the original pitch. The 13 pitches form an ascending chromatic scale beginning on the musical note G3 and ending on G4. The leftmost panel in the top row shows the spectrogram of the neural response to the 391 Hz pitch (G3) played alone. The responses to the other unisons are seen in the bottom row. The fundamental frequency of each harmonic sound is given at the top of the panel. The upper row of spectrograms shows the responses to G3 paired with the harmonic sound corresponding to each of the spectrograms in the bottom row. The musical note names of the harmonic sound pairs are provided above the spectrograms.

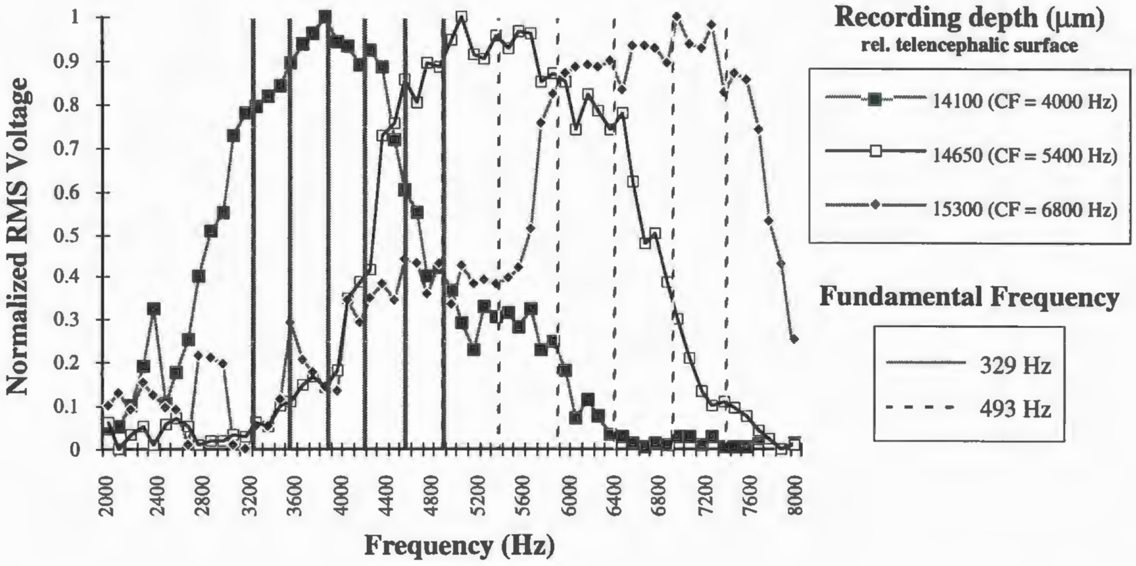

G3(391 Hz)
G + G#
G + A
G + A#
G + B
G + C
G + C#
G + D
G + D#
G + E
G + F
G + F#
G + G
2000
1800
1600
1400
1200
1000
800
600
400
200
0
200
415(G#3)
440(A4)
466(A#4)
493(B#4)
523(C4)
554(C#4)
587(D4)
622(D#4)
659(E4)
698(F4)
739(F#4)
781(G4)
Frequency (Hz)
Time (ms)
F0
4 x10^6
3.5
3
2.5
2
1.5
1
0.5
0
Power (arbitary units)

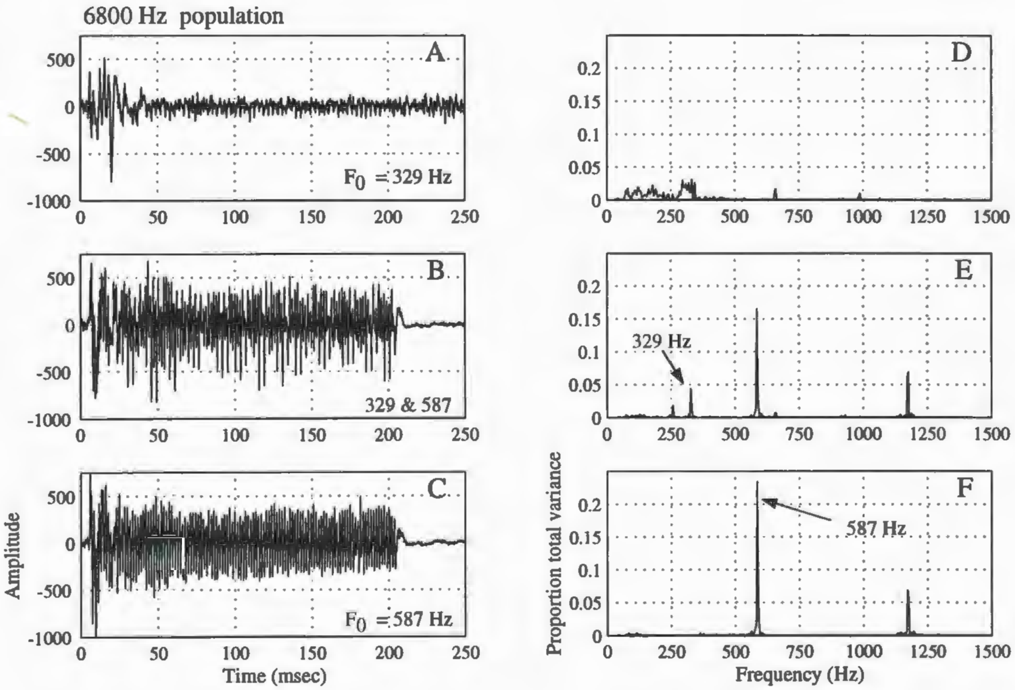
The continuous horizontal lines in each of these spectrograms indicates that the neural response codes the fundamental frequency of the harmonic sounds, and as the fundamental frequency increases over the range of an octave, so does the frequency of oscillation in the neural response.

The top row of spectrograms in Fig. 6 illustrates the consequence of pairing G3 with each of the other harmonic sounds in the chromatic scale. In addition to both of the fundamental frequencies being present in the neural response (continuous horizontal lines), energy now appears in a number of other frequency bands for brief moments during the response to the stimulus. Several features of the spectrograms are noteworthy. The first is the presence of energy in frequency bands corresponding to combination tones formed from the two F_0s. The combination tones, e.g. f2–f1, are seen most clearly in panels labelled G+C and G+D, as brief horizontal lines beneath the line corresponding to the 391 Hz stimulus occurring in the middle of the stimulus epoch (~100 msec). Note also that the distribution of energy in the spectrograms is more constrained to fewer frequency bands for the intervals with the simplest frequency ratios, e.g., G+C (4:3), G+D (3:2) and G+G (2:1), than for intervals with more complicated relationships such as G+G# (8:9) or G+C# (5:7). For example, the vertical lines in the panels for G+G# and G+A indicate energy is present in multiple frequency bands for those particular 32 msec windows. Stated more simply and independently of a musical framework, the spectro-temporal pattern of the population evoked response reflects the interactions among the spectral components present in the two sounds.

Other interesting features of the neural spectrograms which appear to be relatively independent of the spectral properties of the stimulus are the onset and offset responses. Onset responses tended to have a broad spectral distribution, while offset responses were characterized by energy below 150 Hz (best seen in the bottom row of spectrograms).

FIGURE 7. Idealized stimulus spectra of two harmonic complexes superimposed on three frequency tuning curves recorded at different depths along the tonotopic axis of the ICc are shown at the top of the figure. The sounds consist of harmonics 10 through 15, and, given their different fundamental frequencies, fall into different spectral regions. Consequently, the amount of energy from each stimulus within any given tuning curve varies across the tuning curves, with both harmonic sounds contributing energy to the middle tuning curve. Panels A–F: Neural population response to harmonic sounds falling within and outside of the rightmost tuning curve shown above. (A) Average response to a sound consisting of harmonics 10–15 of a 329 Hz fundamental frequency. (B) Response of the same population to the 329 Hz sound played in conjunction with a harmonic sound consisting of harmonics 10–15 of a 587 Hz fundamental frequency. (C) Response to the 587 Hz sound played alone. D–F Frequency spectra of the responses shown in A–C, respectively.

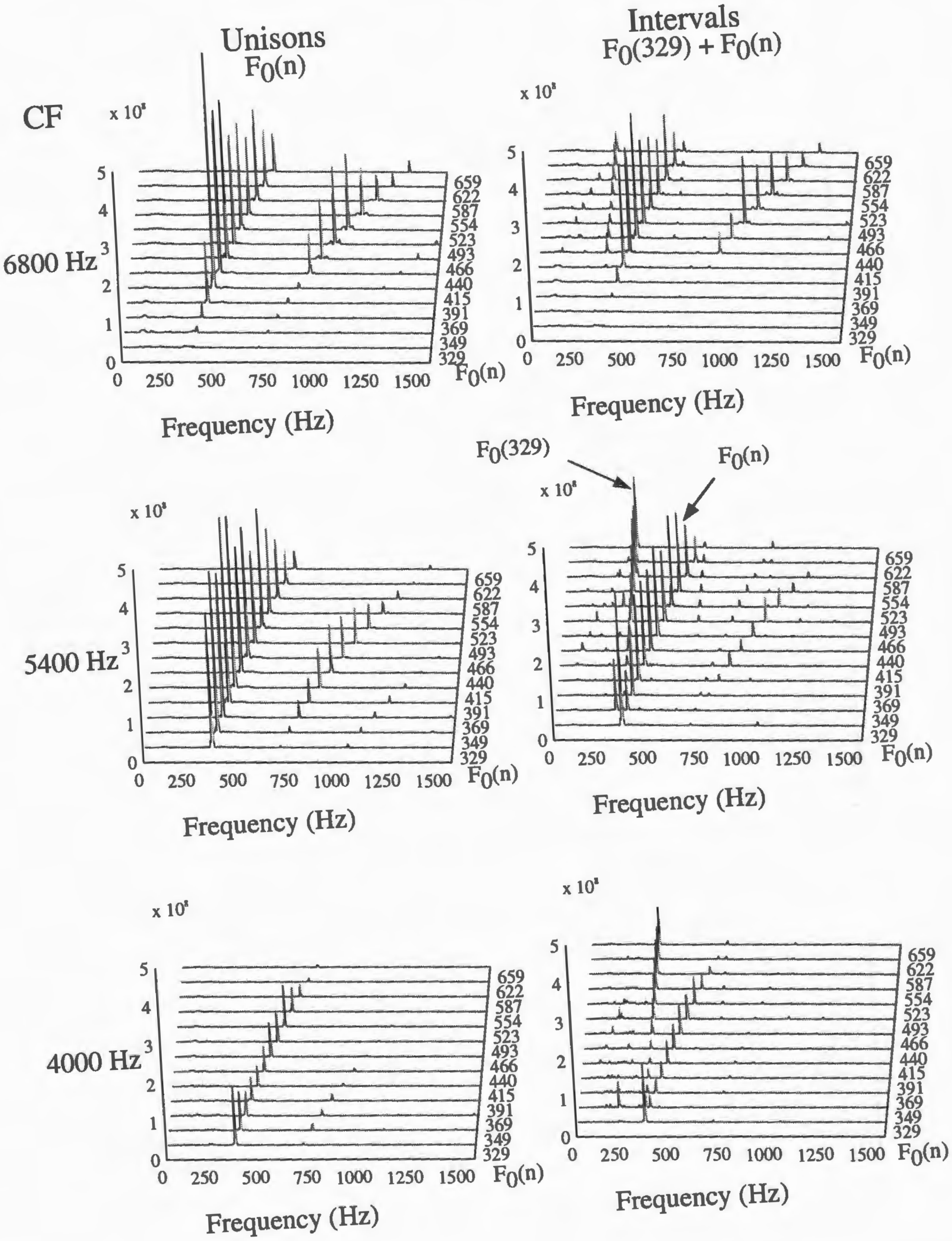

Normalized RMS Voltage
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
2000
2400
2800
3200
3600
4000
4400
4800
5200
5600
6000
6400
6800
7200
7600
8000
Frequency (Hz)
Recording depth (µm)
rel. telencephalic surface
14100 (CF = 4000 Hz)
14650 (CF = 5400 Hz)
15300 (CF = 6800 Hz)
Fundamental Frequency
329 Hz
493 Hz


6800 Hz population
A
F0 = 329 Hz
B
329 & 587
C
F0 = 587 Hz
Amplitude
500
0
-500
-1000
0
50
100
150
200
250
Time (msec)
D
E
329 Hz
F
587 Hz
Proportion total variance
0.2
0.15
0.1
0.05
0
250
500
750
1000
1250
1500
Frequency (Hz)

Distribution of Neural Activity Across the Tonotopic Map

The temporal pattern of activity of any single neuronal population with a unique frequency tuning curve depends on the relationships among the spectral components of the harmonic sounds that fall into the population's frequency tuning curve. Fig. 7 illustrates the positioning of two harmonic sound spectra relative to the tuning curves of three ICc neuronal populations that are separated by 650 μm along the dorsoventral axis (i.e., tonotopic axis). While energy from both harmonic sounds falls into the population whose tuning curve has a center frequency of 5400 Hz, the population with the higher center frequency (6800 Hz) receives energy mainly from the higher-pitched harmonic sound. Responses of the neuronal populations whose tuning curves are shown at the top of Fig. 7 (CF = 5400 Hz; CF = 6800 Hz) are shown in Figs. 5 and 7, respectively. The top panels (A, D) in each figure show the response to the same harmonic sound ($F_0 = 329$ Hz). When the stimulus falls into the tuning curve, the population shows a sustained response to the harmonic sound (Fig. 5A). When the sound is placed outside the population's tuning curve it evokes a rather strong onset response, but not a sustained response (Fig. 7A). A higher pitched (587 Hz) harmonic sound whose harmonics fall into the tuning curves of both populations evokes both the onset and sustained responses in both populations. The response of the 6800 Hz population when the two sounds are played together is influenced by both sounds (Fig. 7B,E), but only weakly by the lower-pitched sound. Note that the 329 Hz peak is enhanced when the 329 Hz pitch is played in conjunction with a higher pitch (Fig 7E), compared to when the 329 Hz stimulus is played alone (Fig 7D). Onset responses of the type shown in Figs. 6 and 7A were ubiquitous and appeared consistently.

FIGURE 8. Distribution of responses across different regions of the ICc tonotopic map in response to harmonic sounds played in isolation or in combination. The 3 rows correspond to the population tuning curves shown in Fig. 7 (top panels = rightmost tuning curve). The center frequency (CF) of each population's tuning curve is listed on the left. Harmonic sounds consisted of harmonics 10–15 of fundamental frequencies ranging from 329 to 659 Hz. The left column shows responses to single harmonic sounds (Unisons) while the right column shows responses to combinations of a 329 Hz F_0 stimulus with each of the 12 other F_0s in the octave spanning 329 and 659 Hz (Intervals). Each line of the individual waterfall plots is the power spectrum for an individual stimulus condition. The representations of the stimulus set differ across the three neuronal populations.


Unisons
F0(n)
Intervals
F0(329) + F0(n)
CF
6800 Hz
5400 Hz
4000 Hz
x 10⁸
0
1
2
3
4
5
0 250 500 750 1000 1250 1500
Frequency (Hz)
F0(n)
659
622
587
554
523
493
466
440
415
391
369
349
329
F0(329)
F0(n)

This was the case even if the population did not respond tonically to the stimulus, implying that a signal is propagated across a broader region of the ICc than is ultimately involved in a fine-grained representation of the temporal properties present in the stimulus.

Fig. 8 shows the distribution of activity in the part of ICc represented by the tuning curves shown in Fig. 7 in response to harmonic complexes presented in isolation (unisons) or paired with another harmonic complex (intervals). Since each harmonic complex consists of harmonics 10–15 of any given fundamental frequency, the spectral region occupied by a sound's carrier frequencies shifts to higher frequencies and broadens as the fundamental frequency increases. Thus, the population with a center frequency of 4000 Hz responds to the 329 F_0 stimulus but does not respond to the 659 Hz F_0 stimulus (Fig. 8, bottom left). In contrast, the population with a 6800 Hz CF shows no response to the 329 Hz stimulus played in isolation, but responds well to the 659 Hz stimulus. The center population (CF = 5400 Hz) responds well to all of the stimuli. When stimulated with harmonic intervals, the population shows responses at a range of difference frequencies (distortion products) corresponding to f1+f2, 2f1-f2 and f2-f1 (Fig. 8, middle row, right).

An enhancement of the neural representation of the lower-pitched sound, whose carrier frequencies fall out of the dominant region of the population's tuning curve, is clearly seen in the upper right panel of Fig. 8. The top panels show that the 6800 Hz neuronal population responds primarily to the harmonic sounds with higher F_0s, and exhibits a vanishingly small response to the 329 Hz stimulus played alone (Fig. 8, top left). When paired with the higher F_0 stimuli, a neural response to the 329 Hz pitch can be observed (Fig. 8, top right).

Discussion

Temporal Patterning Across Tonotopic Regions of the ICc

Fundamental Frequency Coding

We have found that populations of neurons in the barn owl's inferior colliculus are capable of representing the patterns of amplitude modulation present in harmonic stimuli of varying complexity. These findings are consistent with reports that midbrain neurons in other species are capable of representing the temporal structure in harmonic stimuli (for review, see Langner, 1992). For example, single neurons distributed across the tonotopic map in the cat inferior colliculus can track the periodicities of amplitude modulated pure tones (harmonic sounds with three spectral components) to modulation frequencies as great as 1000 Hz (Langner and Schreiner, 1988). The strongest synchronizations to the stimulus envelopes occur in units with center frequencies < 5 kHz and modulation rates < 300 Hz.

The temporal patterning in the neural response depends on the presence of multiple spectral components from the stimulus in the neuronal population's tuning curve and also on the relative spacings of the spectral components. The relative spacings of the spectral components determine the frequencies of the amplitude modulations (difference frequencies) that are measured in the neural response. For example, Fig. 5A shows the response of a neural population with a center frequency of ~5400 Hz to a harmonic sound consisting of harmonics 10–15 of a 329 Hz fundamental frequency. Since the highest spectral component within this harmonic complex is 4935 Hz, a portion of the population's tuning curve is filled with a portion of the sound's spectrum (Fig. 7), causing the population to respond. The response itself shows an amplitude modulation at the frequency difference between the spectral components (329

Hz). In the case of harmonic sounds where the spectral components are integer multiples of the fundamental frequency, the frequency difference coincides with the fundamental frequency of the stimulus (Fig. 5A). The same harmonic sound is less effective in evoking a response from a neuronal population with a tuning curve centered around 6800 Hz because less energy from the stimulus falls into the tuning curve of this population (Fig. 7A,D).

The pattern of amplitude modulations across the tonotopic map becomes more complicated when multiple harmonic complexes with different fundamental frequencies are presented simultaneously. Regions of the map that receive energy exclusively from one of the harmonic complexes show an amplitude modulation primarily at the F_0 of that harmonic complex and are negligibly influenced by the other sound. However, in regions of the tonotopic map where the carrier frequencies (spectral components) of the individual harmonic complexes overlap, the population's firing pattern displays energy at frequencies corresponding to each of the fundamental frequencies as well as to various combinations of the fundamental frequencies. The most prevalent such combination is a subtraction of the lower F_0 from the higher F_0. Thus, the map supports images of multiple harmonic sounds, where information about the harmonic nature of the sounds' spectra is encoded in the distribution of amplitude modulation patterns across the map (compare the response to the harmonic interval 329 + 587 across the three populations in Fig. 8, right column). A higher brain center observing the activity across the ICc's tonotopic map could in principle segregate auditory objects based on differences in temporal modulation patterns emanating from different regions of the tonotopic map. Such models have been proposed for the segregation of concurrent vowels with different fundamental frequencies (Assmann & Summerfield, 1990).

In a few instances we observed an improvement in the representation of one harmonic complex when a second harmonic complex was added. The top panels in Fig. 8 show that the neural population responded preferentially to harmonic stimuli consisting of harmonics 10–15 of higher F_0s and not very well to a 329 Hz F_0 stimulus whose carrier frequencies were largely outside this population's tuning curve (Fig. 8, top left; see also Fig. 7A). When accompanied by higher F_0 stimuli ($F_0 > 440$), the response to the 329 Hz F_0 stimulus was enhanced (Fig 8, top right). This enhancement might arise from lateral interactions along the tonotopic axis. If a region responds robustly to both harmonic complexes (Fig 8, middle right), the strongly patterned neural discharges in the regions sensitive to both harmonic complexes could influence the activity of adjacent regions through these lateral connections. Alternatively, if a region is receiving sub-threshold input from a stimulus, the introduction of a stronger input might bring the responses to the weaker stimulus above threshold.

Onset Responses

Our evoked potential recordings demonstrated that phasic and tonic activations of ICc neuronal populations have different spatial extents of activation along the tonotopic axis. The tonic response to the amplitude modulation pattern in the stimulus depended on superposition of the stimulus' carrier frequencies with the population's tuning curve. However, a phasic onset response, characterized by energy in several bands below 400 Hz, was seen at regions of the tonotopic axis outside those activated directly by stimulus carrier frequencies. Neurons within the ICc respond to stimulus onsets even when the frequency or interaural time difference (ITD) of the stimulus is not the preferred frequency or ITD of the neuron (Wagner, 1990). The extension of onset responses beyond regions of preferred frequency and spatial tuning suggests that onset responses may signal the entrance of a new auditory object into the space/frequency map in the ICc. The transient activation of a larger portion of the map may serve to summon attention at higher auditory centers.

Detection of sound onsets is an important factor in the detection and identification of sounds. Recordings from the barn owl's auditory space map in ICx have shown that the onset response of a single neuron to a second, spatially displaced sound is suppressed if it follows a first sound by fewer than 5 msec (Keller and Takahashi, 1996a). Under such stimulus conditions, the owl neglects to orient to the second sound, illustrating the psychophysical phenomenon observed in humans known as the "precedence effect." The neural responses to such stimuli raise the possibility that detection of stimulus onsets is critical for the execution of a behavioral response. Another example of the importance of stimulus onset coding/detection comes from studies of the physiological representations of voice onset time (VOT) which is used for differentiating voiced and unvoiced consonants. The presence of onset responses in monkey auditory cortex in response to the onset of the consonant and vowel portions of a syllable indicates that a timing comparison of the two onset responses could recover the VOT present in the original stimulus (Steinschneider et al., 1994).

Single Neuron vs. Evoked-Potential Representations of Temporal Fine Structure

A goal of this study was to compare single-unit and evoked potential recordings. Given a limited set of neurons/locations at which recordings are made, the evoked-potential approach yields inferences about the representations of harmonic sounds within the ICc that might not arise from examining PSTHs derived from single-unit data. As illustrated in Figs. 3 and 4, the averaged evoked activity, although biased strongly by the activity of a single neuron, reflected the fundamental frequency of the stimulus, whereas the activity of a single neuron analyzed with traditional single-unit techniques did not. In general, we found that the strength of stimulus-related periodicities in PSTHs of single-unit activity constructed from multiple repetitions of a stimulus equalled the strength of stimulus-periodicities present in single-trial population recordings. This observation suggests that an ensemble of ICc neurons is capable of representing the temporal fine structure in the portion of the stimulus spectrum whose energy it senses. Evoked potential recordings also allowed us to infer that ensembles of ICc neurons may represent higher amplitude-modulation rates better than could be represented by single neurons alone. The range of observed modulation frequencies extends to higher frequencies (300–1000 Hz) in the cat IC when multi-unit activity is compared to single-unit activity (Langner and Schreiner, 1988).

A major problem in interpreting the functional significance and information content of evoked potential recordings is an inadequate understanding of the source of the measured electrical field. The morphology of the field potential depends on both the anatomical arrangement of the active neurons as well as the position of the recording electrode relative to the group of neurons (Llinas & Nicholson, 1974). On the basis of the examples in Figs. 3 and 4, we suggest that the field potentials recorded in the ICc reflect the sum of the spiking activity of multiple neurons. Since the polarity of any given neuron's spikes depends on the position of the electrode relative to that neuron, it is possible to record spikes of both polarities at most recording locations. The summation of spikes of both polarities results in excursions of the field potential about both sides of baseline (Figs. 2E,G; 3C; 5A-C), and the resulting frequency spectra more accurately describe the periodicities present in the stimulus.

The bipolar nature of the field potential is potentially important with regard to coding the entire stimulus waveform. Determination of mechanisms by which a neural representation of both positive and negative components of a stimulus waveform is formed has posed a problem for neurophysiologists because the activity of single neurons when analyzed by PSTHs is inherently unipolar. Javel et al. (1988) found that compound PSTHs (PSTHs with excursions on both sides of the baseline) provide more accurate matches between the temporal structures in the stimulus waveform and the neural response. To estimate the PSTH corresponding to both the positive and negative portions of the waveform, the same stimulus is played twice but with a phase angle difference of 180° between the two repetitions. Evoked potential recordings bypass the need to present a stimulus twice with different starting phases to acquire a neural waveform whose frequency spectrum is devoid of half-wave rectification artifacts associated with spike-thresholding. Thus, recording of evoked potentials seems an adequate method for quickly assaying the functional representations of complex stimuli in the inferior colliculus.

Temporal Patterning in the Auditory System and Implications for Mechanisms of Pitch Perception and Auditory Object Formation

Perception of pitch largely underlies our ability to successfully orient to and identify complex sounds in our surroundings. When listening to sounds that we say possess a specific pitch, information from many different frequency channels is combined into a unified perception of a single pitch (Bregman, 1990). This grouping allows the perceived pitch to be characterized by a single frequency, even though many individual frequencies may be physically present in the sound. The sensation of pitch is strongest if the individual frequencies composing the sound are harmonically related, i.e., integer multiples of the fundamental frequency (Hartmann, 1988). This observation holds true even for sounds in which the fundamental frequency is physically absent from the stimulus, and has been termed 'residue pitch', 'periodicity pitch' or the 'pitch of the missing fundamental' (deBoer, 1976).

Psychophysical studies in humans have given rise to two primary theories of pitch perception (see Houtsma and Smurzynski, 1990 and Javel and Mott, 1988 for discussions). In one view, the temporal patterning of the auditory information provides the necessary cues for extracting the F_0, while the other view maintains that the F_0 is determined by a pattern analysis on the activations of a tonotopic structure. At the core of the argument is the issue of whether spectral components of the sound are resolved vs. unresolved with regard to the cochlear filters that are stimulated.

The pattern analysis/place-rate models depend on the resolution of spectral components into separate channels or on a comparison of response strength across frequency channels with different center frequencies. The relative spacing of active channels, which can be analysed by a higher order pattern analyzer, determines the harmonic relationship among the spectral components (e.g., Terhardt et al., 1982). The temporal coding model, on the other hand, relies on some form of analysis of inter-spike intervals that correspond to the periodicities in the stimulus. This analysis transpires both within individual frequency channels and across multiple frequency channels. Cariani & Delgutte (1996 a,b) have demonstrated a striking similarity in the autocorrelation functions of a stimulus and the inter-spike interval histograms averaged across a population of auditory nerve fibers, indicating that the temporal structure of a stimulus is maintained in the ensemble of auditory nerve fibers.

Periodicities in the inter-spike intervals may simply reflect phase-locking to individually resolved spectral components at low frequencies, or they can reflect the amplitude modulation that arises from the inclusion of multiple spectral components in a single cochlear filter (unresolved harmonics). Since the bandwidth of cochlear filters increases with increasing frequency, higher harmonics of a F_0 will tend to fall into the same tuning curves and thereby give rise to an amplitude modulation whose period corresponds to the F_0 of the stimulus. Although there is no actual energy at the amplitude modulation frequency in a linear transduction mechanism, non-linearities in the hair cell transduction channel gating mechanism introduce components into the gating currents at the distortion frequencies described by the equation $m \cdot f_1 \pm n \cdot f_2$ (Jaramillo et al., 1993; Markin & Hudspeth, 1995). When multiple spectral components (regardless of harmonic relatedness) are present in the tuning curve of an individual hair cell, periodicities should be evident in the hair cell's currents that correspond to a sum of numerous distortion products, which then may give rise to the periodicities recorded in auditory nerve fiber activity.

Our data pertain primarily to the unresolved harmonics condition described above because the frequency content of our stimuli was consistently in a frequency range from 2000 to 8000 Hz. Our experiments sought to characterize patterns of activation in the barn owl's ICc in response to harmonic sound objects that fall into this frequency range, rather than directly address issues of pitch perception. Nonetheless, our data do correlate with pitch percepts associated with such harmonic sounds in humans. In a study using harmonic complexes similar to ours (200 Hz F_0; 11 harmonics starting at harmonics >7), Houtsma and Smurzynski (1990) found pitch percepts in human subjects that were adequate for identifying harmonic intervals and detecting direction of pitch movement. They concluded from their data that two mechanisms exist for determination of pitch. Although the pitch percepts associated with harmonic sounds with more than five unresolved harmonics were sufficient for generating robust pitch percepts, they were not as strong as when the carrier frequencies of their harmonic sounds fell below 2000.

Conclusions

Ultimately, the perception of pitch and temporal modulations in acoustic signals in general depends on the pattern of activity across the various tonotopic maps that are encountered as auditory information traverses the cochlea to the cortex. For any given neural structure within the auditory pathway, the pattern of neural activity evoked by an acoustic signal represents the cues that are available to the nervous system for forming percepts of that signal. The form of the evoked pattern of activity is necessarily constrained by the interaction of a sound's spectral components with the filtering properties of the cochlea, non-linearities in the mechanosensory transduction process and the filtering properties of all synaptic stations that lie between the auditory periphery and more centrally-located structure of interest. We suggest that the pattern of temporal modulations on the tonotopic map in the ICc, induced by harmonic sounds played either in isolation or in various harmonic relations, may provide a basis for binding together a set of frequencies into auditory objects by causing populations of neurons to fire coherently.

CHAPTER III

ENHANCED RESPONSES OF MIDBRAIN NEURONS TO STIMULUS TRANSITIONS EMBEDDED IN SEQUENCES OF HARMONIC SOUNDS

Abstract

The hypothesis that neural representations of harmonic sounds in the midbrain are shaped by the context in which they occur was tested by recording the activity of populations of neurons in the inferior colliculus of barn owls in response to sequences of rapidly repeating harmonic sounds. In addition to overall increases or decreases in response strength as a function of the repetition rate (inter-stimulus intervals ranging from 40 msec to 480 msec), responses to infrequent "deviant" stimuli were generally stronger than responses to the standard stimuli surrounding them in the sequence or to the same stimulus when it served as a standard. Stimulus deviance was manipulated by altering the fundamental frequency of the harmonic sound, spectral envelope or both. Spectral deviants affected the neural response most strongly. The results are discussed with respect to the "mismatch negativity" and human psychoacoustic studies of auditory pattern perception and auditory stream segregation.

Introduction

Organisms are continuously confronted with changing sensory input, and the ability to detect and respond to these changes when appropriate underlies an organism's ability to orient within its sensory environment. The neural mechanisms and brain regions involved in the detection of and orientation to novel stimuli or rare stimuli (deviant stimuli) are poorly understood. Correlates of a deviance detection process are readily found in measures of cortical activity (Näätänen, 1992). The question arises whether the processing of a sensory stimulus in sub-cortical structures is sensitive to the overall context in which the stimulus occurs or if, instead, the neural representation of a stimulus remains invariant until it reaches the cortex.

Contributions of sub-cortical structures to the mismatch negativity (MMN), an evoked potential index of deviance detection routinely described in recordings from the human scalp (Näätänen, 1992), have been observed in the cat (Csépe et al., 1993) and the guinea pig (Kraus et al., 1994a, 1994b). The non-primary auditory thalamus in guinea pigs shows enhanced responses to deviant tones and some speech-like sounds, suggesting that some information about transitions in steady-state auditory stimulation is extracted prior to reaching the auditory telencephalon.

The goal of this study was to elucidate conditions under which midbrain neurons in the barn owl (Tyto alba) respond to sudden changes in stimulus features such as fundamental frequency (pitch) or spectral envelope (timbre) in sequences of discrete harmonic sounds. Stimulus presentation rates were varied to determine the temporal extent of any observed context effects as well as to assess general increments and decrements in response strength across sequences of regularly and rapidly repeating stimuli. The barn owl's central nucleus of the inferior colliculus (ICc) contains orthogonally oriented maps of tonotopy and inter-aural time difference (Wagner et al., 1987), and as such constitutes a useful structure for studying neural representations of complex auditory scene dynamics. As an almost-obligatory waystation for auditory information ascending the brainstem (Wild, 1987), the pattern of ICc activity represents the cues that are available to higher auditory centers for parsing the auditory scene. Thus, understanding ICc dynamics during presentations of streams of rapidly repeating and spectrally complex stimuli should provide clues about the perceptual organization of complex auditory situations.

Harmonic (periodic) sounds were used in these experiments because of their ability to stimulate populations of ICc neurons in a sustained and temporally predictable manner. In addition, many naturally-occurring sounds, such as speech sounds, are harmonic stimuli that occur in quick succession, e.g., sequences of phonemes. Evoked potential recordings from populations of barn owl ICc neurons effectively capture the temporal patterning in harmonic sounds presented alone or as various pairs (Janata et al., 1994). Specifically, when presented with a harmonic sound consisting of higher harmonics of a fundamental frequency, the evoked potential contains significant power at the fundamental frequency, even if the fundamental frequency is missing. Although single neurons also encode the temporal modulation patterns in harmonic stimuli, it is often necessary to repeat the stimuli many times for the encoding of the temporal patterning to become evident in post-stimulus time histograms. Therefore, evoked potential measures provide for an expedited assessment of a neuronal population's representation of a stimulus with moderate spectral complexity. The goal of this study was to determine if these representations can be modified by contexts in which individual harmonic sounds appear. A portion of this work has been presented elsewhere in preliminary form (Janata & Takahashi, 1995).

Methods

Electrophysiological Recording

Anesthetic and surgical procedures used in these experiments were approved by the Institutional Animal Care and Use Committee at the University of Oregon and have been described previously (Takahashi & Keller, 1994). An anesthetized owl (0.05

ml/h, 100 mg/ml Vetalar, Parke-Davis and 0.025 ml/h, 5 mg/ml Diazepam, C-IV, LyphoMed) was maintained in a fixed position within a stereotaxic device by a stainless steel plate cemented to the skull. Recordings were made within a sound attenuating booth (Industrial Acoustics Co.) lined with acoustic foam (Ilbruck Sonex). Recordings of neural activity were made using either tungsten microelectrodes (5-10 MΩ, Frederick Haer) or glass micropipette electrodes filled with physiological saline (0.9% NaCl; ~1-2 MΩ impedance).

Bursts (100-msec) of broadband noise were used to locate populations of auditory neurons. Neurons were identified as being located in the central nucleus of the inferior colliculus (ICc) if they showed peaks in their frequency tuning curves, as well as some degree of spatial selectivity (determined by playing sounds from different positions along the azimuth).

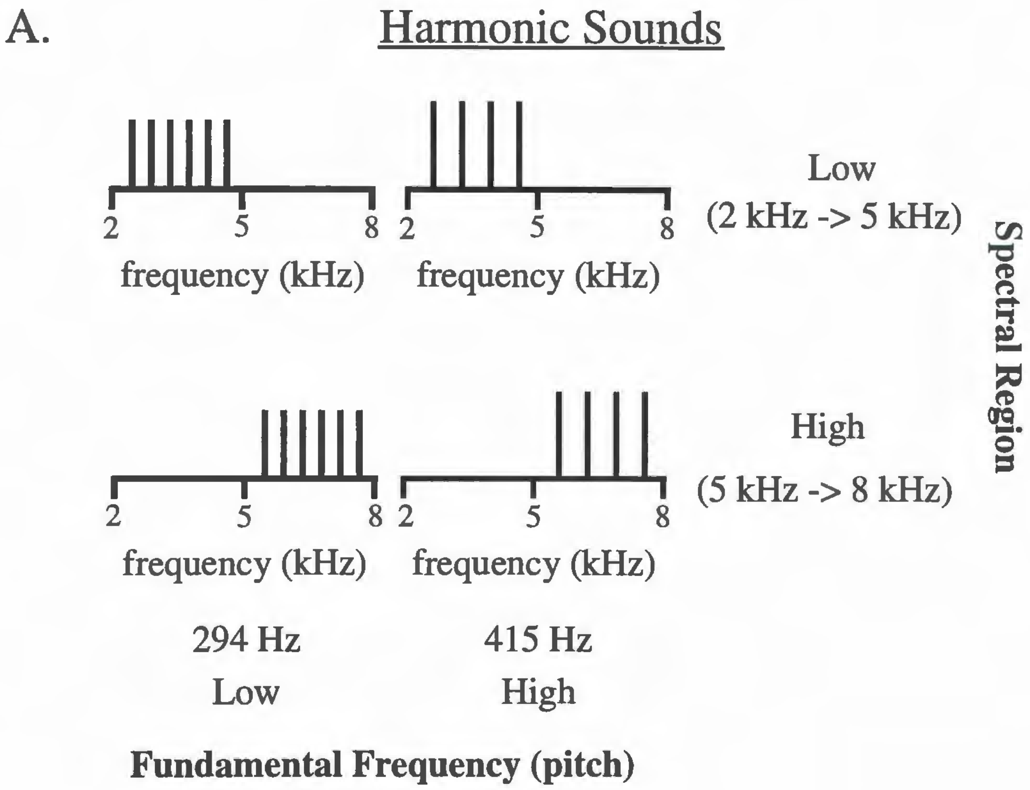
Evoked neural activity was amplified and bandpass filtered (1 Hz to 2.5 kHz, Mike Walsh Electronics), digitized with 12-bit resolution at a rate of 6400 samples/second (PD1, Tucker Davis Techonologies) and written to disk.

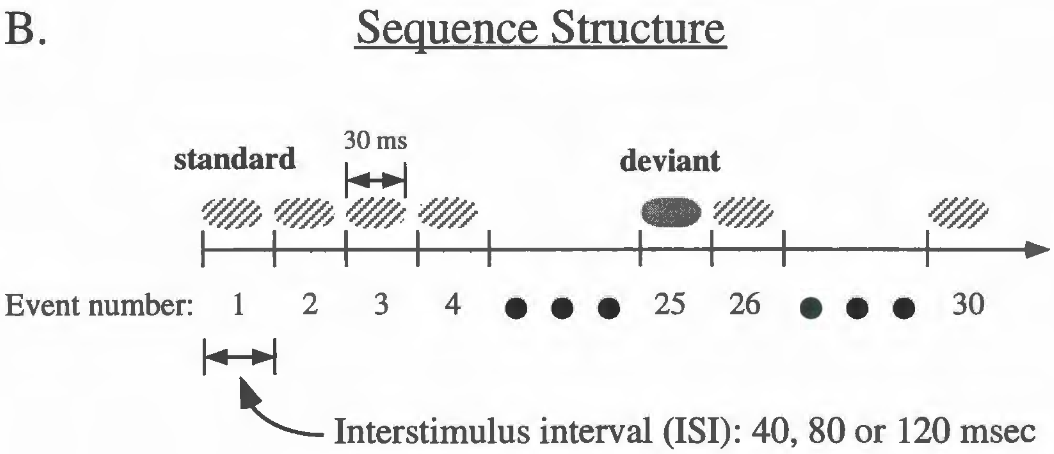
Stimuli

Harmonic Complexes

All stimuli were harmonic complexes consisting of several harmonics of a fundamental frequency. Fundamental frequencies (F_0s) ranging from 200 to 450 Hz were used. The harmonics present in any given stimulus were selected to fall into a circumscribed spectral region that was expected to activate a discrete population of neurons along the tonotopic axis in the ICc. The stimuli could therefore be manipulated independently along pitch (fundamental frequency) and spectral dimensions (Fig. 9A).

FIGURE 9. Experimental design. (A) All stimuli were harmonic sounds consisting of several harmonics of a fundamental frequency (F_0). Schematized are four harmonic sounds varying in F_0 and spectral envelope. For any given F_0 (e.g., 294 Hz), a sound was created with all higher harmonics of the F_0 falling into the 2,000 to 5,000 Hz band (low spectral region) and another with all harmonics falling between 5,000 and 8,000 Hz (high spectral region). (B) Harmonic sounds were presented in sequences in which the structure was fixed with a "deviant" stimulus falling on event #25, or randomly generated given a global deviant probability of 10%. In any given sequence, one of the four stimuli was selected as a standard (e.g., 294 Hz pitch in high spectral region), and one of the remaining three stimuli was the deviant. Thus, deviants were characterized by a change in F_0, spectral region or both. The experiment was completely counterbalanced meaning that each stimulus served as a standard and each of the deviant types once. Individual harmonic sounds were 30 msec long and sequences were presented at inter-stimulus intervals (measured stimulus onset to stimulus onset) ranging from 40 to 480 msec.


A.
Harmonic Sounds
2
5
8
frequency (kHz)
2
5
8
frequency (kHz)
Low
(2 kHz -> 5 kHz)
Spectral Region
2
5
8
frequency (kHz)
2
5
8
frequency (kHz)
High
(5 kHz -> 8 kHz)
294 Hz
Low
415 Hz
High
Fundamental Frequency (pitch)


B.
Sequence Structure
standard
30 ms
deviant
Event number:
1
2
3
4
25
26
30
Interstimulus interval (ISI): 40, 80 or 120 msec

Individual harmonic sounds were constructed by summing the component frequencies, applying either 2.5 msec linear or cosine ramps at onset and offset, arranging the sequence in which they were to be played, and converting them from digital to analog form at 51200 samples/second (Tucker Davis Technologies, PD1). The sounds were then attenuated (Tucker Davis Technologies, PA4), amplified (McIntosh, M754) and played from a loudspeaker (Alpine, 6020HX) situated on a hoop located on the horizontal meridian, 70 cm in front of the owl. The azimuthal location of the speaker was set to lie in the spatial receptive field of the neuronal population. In most of the experiments, individual bursts were 30 msec in duration and presented at 20–30 dB above neuronal thresholds.

Stimulus Sequences

Sequences of individual 30 msec harmonic sound events were presented at several rates. Stimulation rates are described in terms of inter-stimulus interval (ISI): the time between onsets of successive events. ISIs of 40, 60, 70, 77, 80, 100, 120, 143, 160, 240, 250, and 480 msec were used across several experiments, although only the 40 and 80 msec ISIs were used consistently in all experiments.

Two types of stimulus sequences were presented. In each sequence, one stimulus was defined as the "standard" and one or more of the remaining stimuli were defined as "deviants". "Fixed" sequences consisted of 24 standards followed by a single deviant, and in some cases terminated with five more occurrences of the standard (Fig. 9B). Fixed sequences were repeated from 5 to 25 times. In "random" sequences, standard and deviant stimuli were each assigned a probability of occurrence, and each stimulus event in the sequence was determined given this probability structure. No attempt was made to adjust to local sequence structure to prevent immediate repetition of low-probability stimuli. Random sequences consisted of several hundred stimulus events and lasted 10 sec or longer, depending on the ISI. When sequences consisted of two stimuli, the standard was assigned a 90% probability of occurrence. In some experiments, sequences were constructed with three deviants, each set to a 5% probability of occurrence. Any given random sequence was presented only once.

As described above, deviants could be of several types: (1) they could differ in pitch (F_0) from the standard but consist of component frequencies from the same spectral region (pitch deviants); (2) they could be of the same F_0 but differ in spectral region (spectral deviants); or (3) they could differ along both dimensions (pitch/spectral deviants). The role of each stimulus was counterbalanced across conditions such that it served as a standard on some trials and a deviant on others (a total of 3 trial types). For example, assume the sound schematized in the top left corner of Fig. 9A (low pitch & low spectral region) served as the standard. Each of the other pitches served once as a deviant, resulting in three separate sequences with the low pitch/low spectral region serving as the standard. Next, the role of standard was assigned to one of the other pitches, and three sequences were constructed with each of the remaining pitches serving as a deviant. When this procedure was reiterated for all the stimuli at any given recording site, a total of 12 sequence types were presented with each of the four pitches serving in every contextual capacity. In the fixed sequence conditions, the counterbalancing resulted in each stimulus being presented an equal number of times at each sequence position.

Data Analysis

Fixed Sequences

A total of 7 owls were tested with stimuli from fixed sequence or random sequence conditions. Both time and frequency domain analyses were used to assess the effect of each deviance type in the sequence on its ICc evoked potential representation. Changes in response strength over the course of a sequence were visualized by averaging the traces recorded in all sequence conditions. Although averages can be constructed for every standard/deviant combination at every recording site, the data are most easily viewed in terms of averages created by deviance type, e.g. pitch deviant or spectral deviant.

The overall strength of the evoked response to each event was quantified by computing the root-mean square (RMS) value of a 40 msec window. The beginning of the window was delayed by 6 msec relative to the onset of each event to account for travel time of the sound from the speaker to the owl's ears (2.3 msec) and neural conduction time to the IC. Changes in RMS values over the course of a sequence were expressed as a proportion of the RMS amplitude in response to the first stimulus in the sequence.

Random Sequences

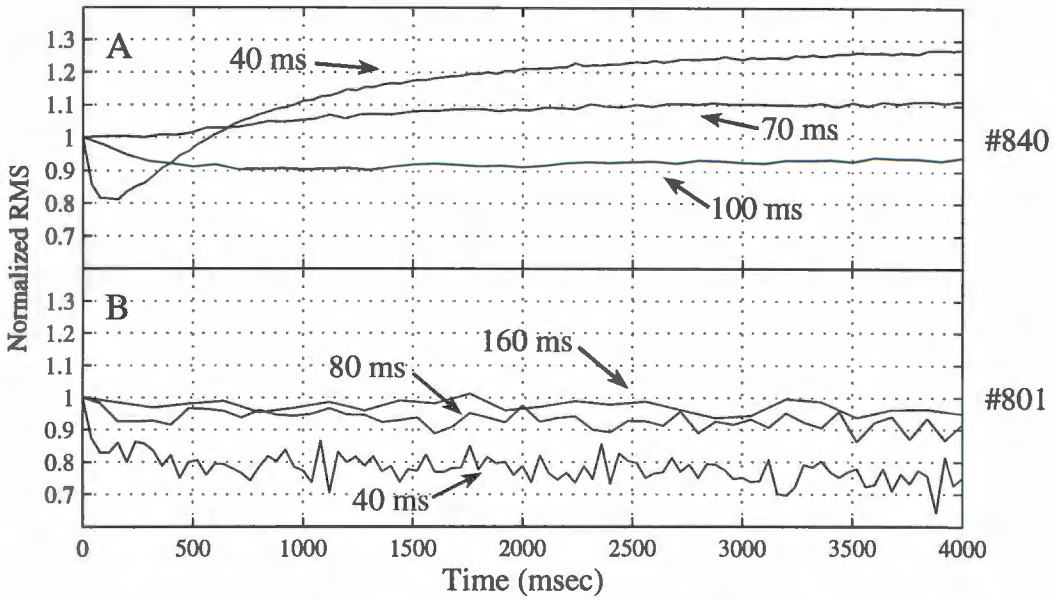
The evoked activity recorded over the course of a "random" sequence, in which individual stimuli were assigned probabilities of occurrence, was parsed into 40-msec windows encompassing each stimulus event. RMS magnitudes were determined for each 40 msec window, and these were averaged separately for standards and deviants. To facilitate comparison across recording sites, the average response to the deviants in the particular sequence condition at that recording site was normalized to the average response to the stimulus defined as the standard. The normalized values from every recording site were averaged and entered into a 2-way repeated measures analysis of variance (ANOVA, SAS software) to test for significant effects of deviance type (pitch deviant, spectral deviant, etc) and ISI.

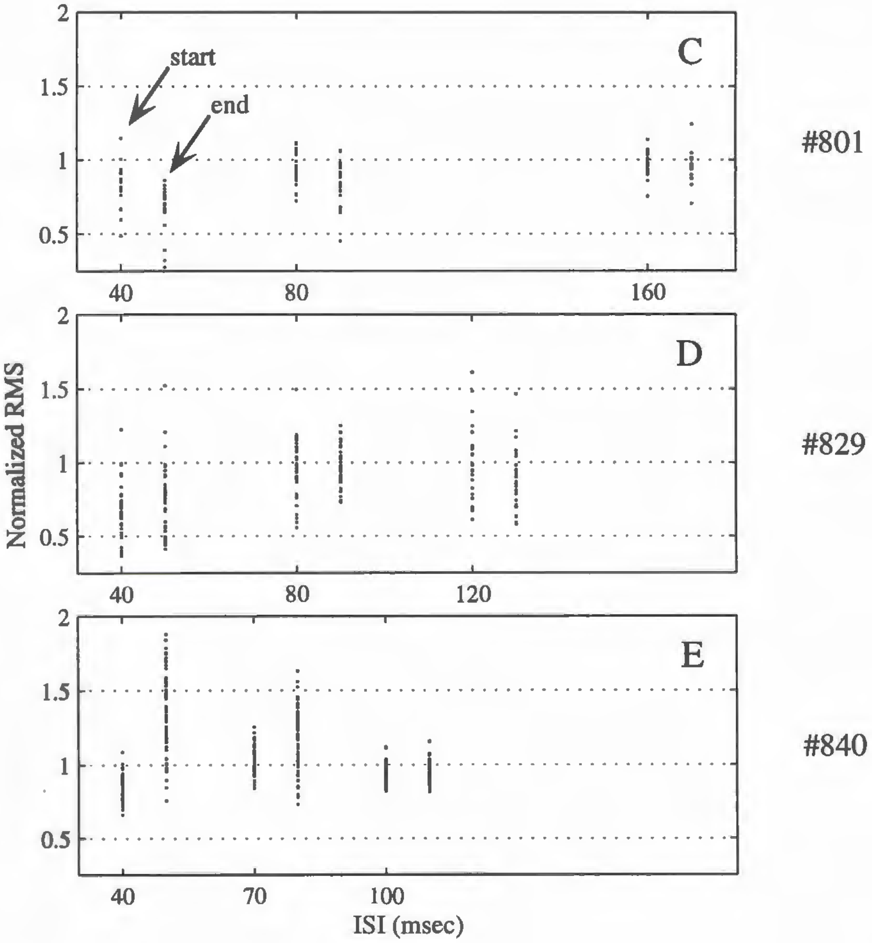
Results

Effects of Varying Stimulus Repetition Rate

Before examining the effects of stimulus deviance on neural responses, let us examine changes in response strength that occur simply as a function of repeated stimulation at a variety of ISIs. Stimulus repetition rate significantly affected the RMS magnitude of the neural response to individual stimuli during a sequence of stimuli. Across different owls, both decreases and increases in response strength were observed and were always most pronounced at the fastest repetition rates. Figures 10A and 10B illustrate the time course of such global changes in the RMS measure of response strength for three ISIs in two owls. In the case of owl #840 (Fig. 10A), the traces are in response to a single repeating stimulus. The traces are constructed from the RMS values of the 40 msec windows encompassing each auditory event, but not including the intervening periods of silence. RMS values were normalized to the RMS value elicited by the first event (harmonic sound burst) in the sequence. For each owl, the traces are the average of RMS values computed in multiple sequence conditions at several ICc recording sites. The traces in Fig. 10B (owl #801) have responses to deviants embedded at random locations.

FIGURE 10. Changes in ICc neuronal response strength over the course of sequences of repeating harmonic sounds. (A) Normalized root mean square (RMS) voltage of ICc neuronal responses in one owl (#840) to sequences of harmonic sounds repeating at three different inter-stimulus intervals (ISIs). The RMS value for each event in the sequence is divided by the RMS value of the response to the first event in the sequence. The ISI of each sequence is indicated by the arrows. Sequences consisted of a single repeating sound. Each trace is the average across many sequences of harmonic sounds with different fundamental frequencies and across multiple ICc recording sites. (B) Data from a different owl (#801). Four different harmonic sounds occurred in each sequence with different probabilities. Traces are averages across multiple four harmonic sound sequences and recording sites. (C–E) Distributions of average normalized RMS values observed at the beginnings and ends of sequences in three owls. The left set of dots in each pair of vertical pair of distributed dots represents the averaged normalized RMS value of the 2nd through 6th events in a sequence of harmonic sounds at the particular ISI noted on the abscissa (labelled "start" for the 40 msec ISI condition in C), while the right set represents the average RMS value of the last 10 events in the sequence. Clustering of values around 1.0 implies little change in the strength of the neural response relative to the 1st stimulus in the sequence, whereas values below 1.0 indicate a decrease and above 1.0 an increase.

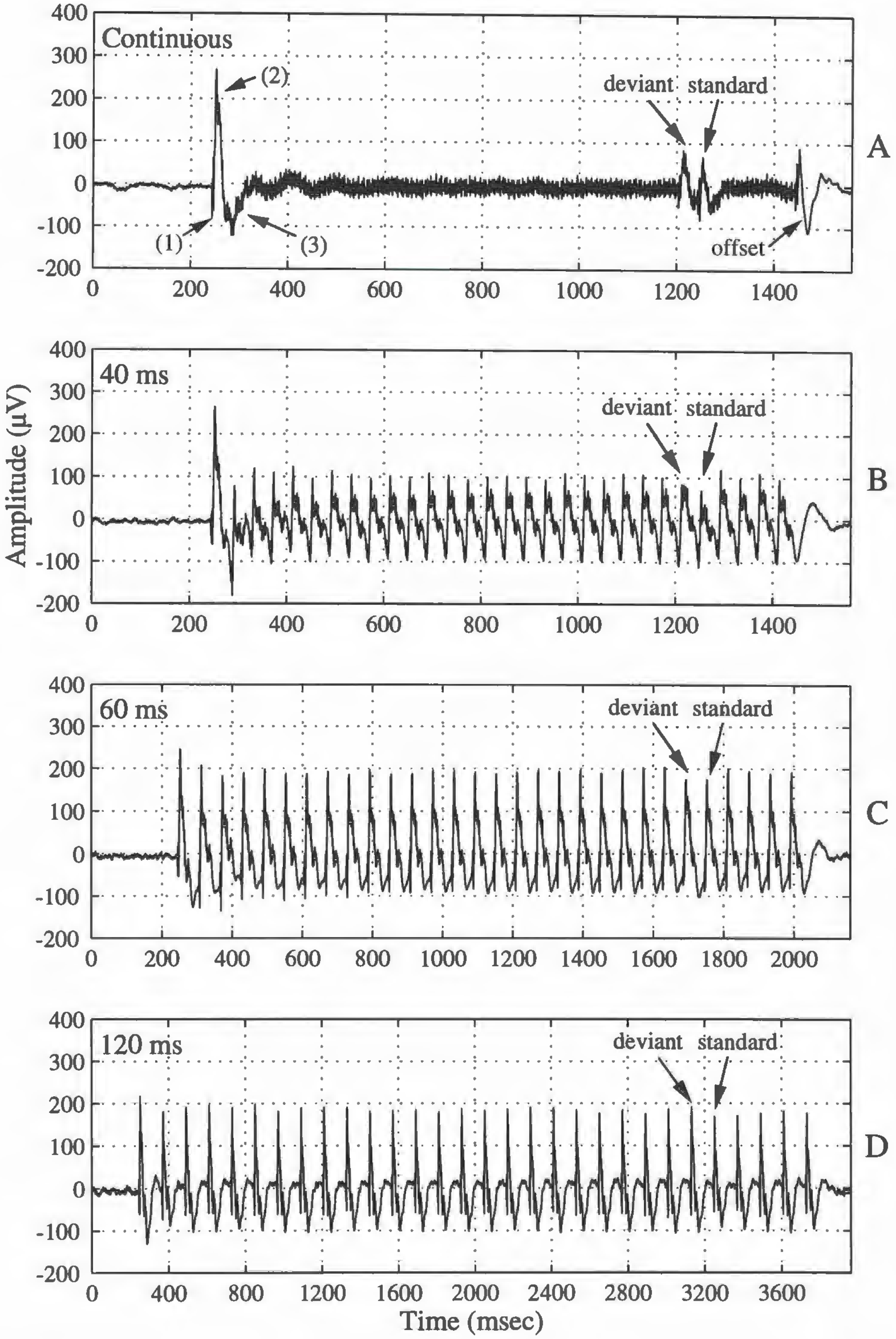

A
40 ms
70 ms
100 ms
#840
B
160 ms
80 ms
40 ms
#801
Normalized RMS
1.3
1.2
1.1
1
0.9
0.8
0.7
0
500
1000
1500
2000
2500
3000
3500
4000
Time (msec)


C
start
end
#801
40
80
160
D
#829
40
80
120
E
#840
40
70
100
Normalized RMS
2
1.5
1
0.5
ISI (msec)

Increases in response strength above the initial level were observed for the 40 msec ISI condition in two owls. The example shown in Fig. 10A illustrates a rapid 20% decrement in response strength during the initial 250 msec followed by an average increase of 30%, with most of the increase taking place in the initial 2-3 sec. In this owl, a substantial RMS increase was noted at the 70 msec ISI also. At the longest ISI tested (100 msec), the response strength decreased slightly within the initial 500 msec and then remained at this level. Overall decreases of the type seen in Fig. 10B were observed in 5 of 7 owls. Responses to sequences in the remaining two owls showed either increases or decreases followed by increases of the type seen in Fig. 10A.

The distributions of normalized RMS values recorded at the beginnings and ends of sequences in all stimulus conditions and recording sites are shown for 3 owls in Fig. 10C–E as a function of ISI. For each ISI, two sets of distributions are plotted. The distribution on the left shows the average normalized RMS value of the 2nd through 6th events at the beginning of the sequence (labelled "start" in Fig. 10C), while the distribution on the right shows the assymptotic RMS level as determined by averaging the RMS values in response to the final 10 events in the sequence (labelled "end" in Fig. 10C). In each case, the RMS is expressed as a proportion of the RMS value for the response to the first event in the sequence. After the first stimulus of the sequence in the 40 msec ISI, the response weakened rapidly by approximately 25% (Fig. 10C-E). At longer ISIs, changes were less pronounced. After several seconds of the rapidly repeating sounds bursts, the response strength was often quite different from that observed at the beginning of the sequence.

FIGURE 11. Average evoked potentials recorded in the ICc of a barn owl in response to sequences of harmonic sounds. Each sequence consisted of 24 repetitions of a single, "standard" harmonic sound, followed by a single instance of a "deviant" harmonic sound, and terminated with five more repetitions of the "standard". Deviant harmonic sounds differed in their fundamental frequency, spectral distribution of the component frequencies or both. Traces show responses averaged across all sequence types and 5 recording sites. A baseline of 240 msec was recorded in every condition. Each harmonic sound event was 30 msec in duration with exception of the "Continuous" condition. Transitions to a deviant event and the subsequent return to the standard event are marked by arrows. (A) Individual harmonic sound events of 40 msec duration linked together to form a single continuous sound that was 1200 msec long. The numeric labels (1-3) point to different aspects of the onset response. An offset response to cessation of stimulation is labelled on the right. (B) Sequences of 30 msec harmonic sound events separated by 10 msec (ISI = 40 msec). (C) ISI = 60 msec. (D) ISI = 120 msec. The time scale differs across the four panels because of the different ISIs.


Continuous
(1)
(2)
(3)
deviant
standard
offset
A
40 ms
deviant
standard
B
60 ms
deviant
standard
C
120 ms
deviant
standard
D
Amplitude (µV)
Time (msec)
400
300
200
100
0
-100
-200
0
200
400
600
800
1000
1200
1400
1600
1800
2000
2400
2800
3200
3600

Context Effects

Fixed Sequences

Fig. 11 shows traces of average evoked activity that illustrate decrements in response strength as a function of ISI, as described in the previous section, together with altered responses to deviants in fixed sequence conditions. In these sequences, the deviant occurred as the 25th event in the sequence. All traces are averages across all deviance conditions at all recording sites in one owl. Fig. 11A shows an average across conditions in which individual events were linked to form a 1200-msec continuous stimulus. The continuous stimulus condition is identical to the 40 msec ISI condition except that each event consists of an unramped, 40 msec sound burst instead of a 30 msec sound burst followed by 10 msec of silence. Because 40 msec was an integer multiple of the periods of the F_0s (275 Hz and 450 Hz), sound bursts could be aligned into a continuous stimulus without abrupt transitions at the event borders. Fig. 11A shows that the onset of the continuous stimulus was characterized by a strong tri-phasic response consisting of a rapid downward deflection (labelled 1), followed by a positive wave lasting approximately 15 msec (2) which transitioned into a somewhat longer (~50 msec) negative wave (3). The intial negative deflection and sharp beginning of the positive component will be referred to collectively as the rapid onset response. The remainder of the positive component and negative component will be subsequently referred to as the slow-potential components. Following the overall onset response, a steady-state was achieved which was characterized by rapid amplitude fluctuations that corresponded to a following of the F_0s of the stimuli (Fig. 3A, between 500 and 1000 msec). The steady-state was maintained until the arrival of a deviant stimulus (see arrow in Fig. 11A) which elicited a response similar to the one seen at the beginning of the sequence. An immediate reversion to the standard stimulus also elicited a similar response. Termination of the stimulus sequence evoked a triphasic waveform deviating from baseline for approximately 80 msec. The negative aspect of this waveform lasted approximately 35 msec, while the initial positive portion lasted about 10 msec.

When short, 10 msec silent periods were inserted between successive 30-msec events (Fig. 11B), each event evoked rapid onset and slow-potential components. Note that successive events arrived before the negative slow-potential evoked by the previous event had a chance to return to baseline. The onset response underwent a rapid decrement in size and continued at a constant level. The deviance and return to standard evoked a noticeable perturbance in the evoked potential. Similar effects were seen at ISIs of 60 msec and 120 msec (Figs. 11C & 11D, respectively), although the decrement in response strength across the first several events in the sequence was smaller at longer ISIs, and the responses to the deviant stimulus did not appear to differ significantly.

Fig. 12 presents the time-domain averages shown in Fig. 11 in terms of RMS amplitude. The four panels are arranged according to deviance type. At each recording site, the RMS value associated with each event was normalized to the response magnitude of the first event in the sequence. Deviants were presented at event #25. The normalized RMS time-series from 5 recording sites in one owl are plotted as a function of event number in the sequence at several ISIs. The same pattern of results was observed in each of the 3 owls tested with this exact paradigm.

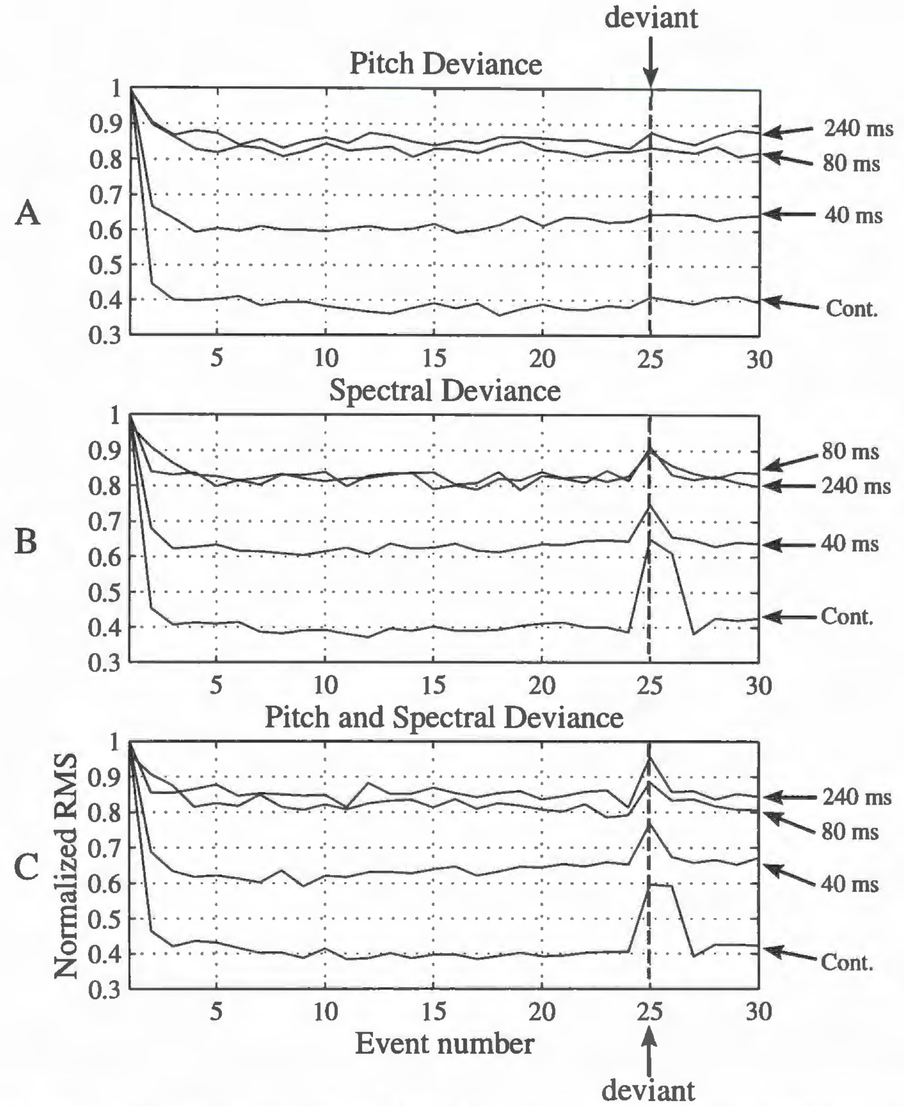


FIGURE 12. Time-series of ICc evoked potential magnitudes as a function of deviation type. Each trace shows changes in RMS across the extent of a harmonic sound sequence, averaged across five recording sites in one owl. The traces are labelled with the ISI between successive events to the right of each panel. Cont. = continuous sound without separation between harmonic sound events. A standard harmonic sound consists of higher harmonics of one of two F_0s (275 and 450 Hz), referred to as pitches. A set of harmonics is chosen to fall into one of two spectral regions (2kHz -> 5kHz or 5kHz to 8 kHz). A deviant stimulus was inserted at event number 25. A "pitch deviance" means that the F_0 of the sound changes, but not the spectral region into which its harmonics fall. A "spectral deviance" means that the F_0 is left unchanged, but the harmonics are taken from the other spectral region. A "pitch and spectral deviance" means that both parameters are changed. Sequences are completely counterbalanced, so that each pitch/spectral region combination serves as the standard in some cases and as the deviant in others. Thus, when added across all of the possible sequence combinations, an equal number of all stimuli are presented at every position in the sequence.

At the four ISIs shown, a rapid decrease in response magnitude was registered at the beginning of the sequence and was strongest for the continuous stimulus (60% decrement) and 40 msec ISI (40% decrement) conditions. A weaker decrease (15–20%) was seen at longer ISIs.

Pitch deviants (changes in F_0) failed to evoke a stronger response compared to the standard stimuli surrounding them in the sequence (Fig. 12A). In contrast, spectral deviants (changes in spectral envelope) evoked robust increases in RMS magnitude at all ISIs (Fig. 12B). At ISIs of 40 msec and greater, the RMS magnitude of the standard immediately following the deviant returned to a similar level as was observed for the standard stimulus immediately preceding the deviant event. In the continuous stimulus condition, in which the deviance effect was strongest, RMS magnitude remained enhanced for the standard stimulus immediately following the deviant event before returning to pre-deviance levels. Combined pitch/spectral deviants showed the same pattern of results as spectral deviants alone (Fig. 12C).

It is important to point out that, averaged across all sequence conditions, the stimulus conditions were identical at every position in the sequence (see Methods for details). In other words, due to counterbalancing across conditions, every stimulus served 3 times as the standard in event position 24, and 3 times as a deviant in position 25. Thus, any changes in features of the evoked response reflect an effect of the stimulus context and not simply an effect of differences in the aggregate physical characteristics of all stimuli presented at one sequence position vs. the next.

Fig. 13 shows similar deviance effects averaged across four single neurons in the same owl. Normalized spike rate (proportion of spikes elicited by each event relative to the number of spikes evoked by the first event) is plotted across the extent of the sequence. Although the inter-event variability in response magnitude is greater in the single-unit activity than in the RMS measure of the population response, the results

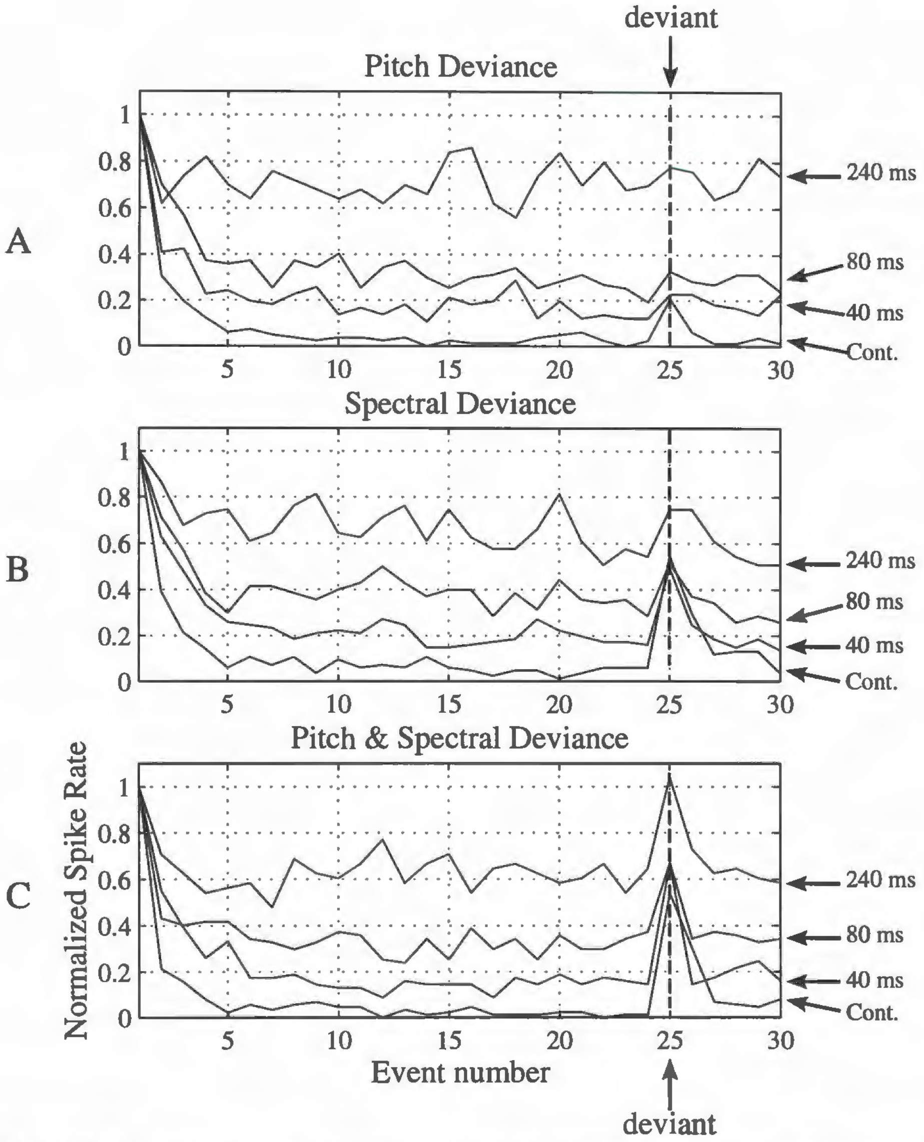


FIGURE 13. Average responses of single neurons in the ICc in response to deviations in sequences as a function of deviation type. Spike rate is expressed as the number of spikes during a 40-msec event epoch (30-msec stimulus & 10-msec silence) divided by the number of spikes evoked by the first event. Traces reflect the averaged activity of 4 single units. Stimulus sequences were constructed as in Figures 12.

are similar. Spike rate decreases as a function of the ISI, and drops close to zero in the continuous stimulus condition. In contrast to the population activity measure, the single units differentiate between the 80 msec and 240 msec ISIs more strongly. As was the case for the evoked potentials, pitch deviants do not produce an enhancement in spike

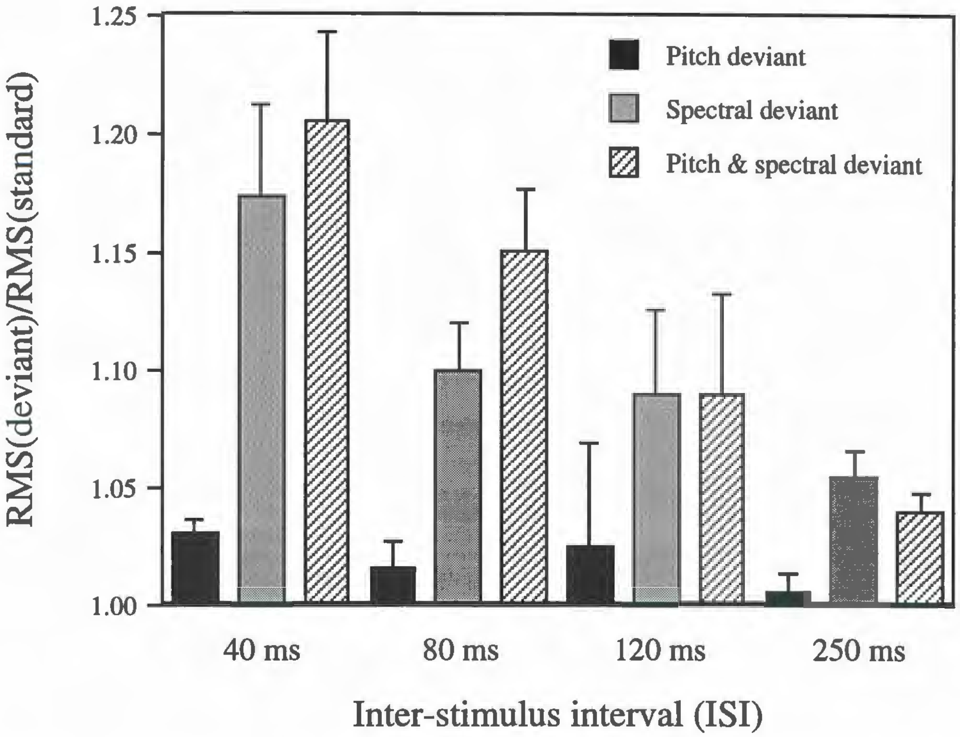


FIGURE 14. Responses to 3 different types of context-deviants randomly embedded in sequences of standards presented at varying ISIs. Sequences consisted of two stimulus types randomly strung together. The stimulus defined as the "standard" occurred 90% of the time. The deviant stimulus was 1 of 3 types (see Fig. 9 for details): (1) a pitch deviant, meaning that its constituent frequencies fell into the same spectral range as the standard, but it differed in F_0 (solid bars); (2) a spectral deviant, meaning that its F_0 was the same as that of the standard, but its constituent frequencies fell into a different spectral range (shaded bars); (3) a pitch and spectral deviant (striped bars), indicating that both F_0 and spectral region differed. Each stimulus event was 30 msec in duration. Sequences consisted of 400 to 500 stimulus events and were presented at multiple ISIs (shown along the abscissa). The response strength to the deviants is expressed as the RMS value for the deviant divided by the RMS value for the standard. Thus, a value of 1.1 represents a 10% enhancement in the response to the deviant relative to the standard. Error bars indicate standard error of the mean.

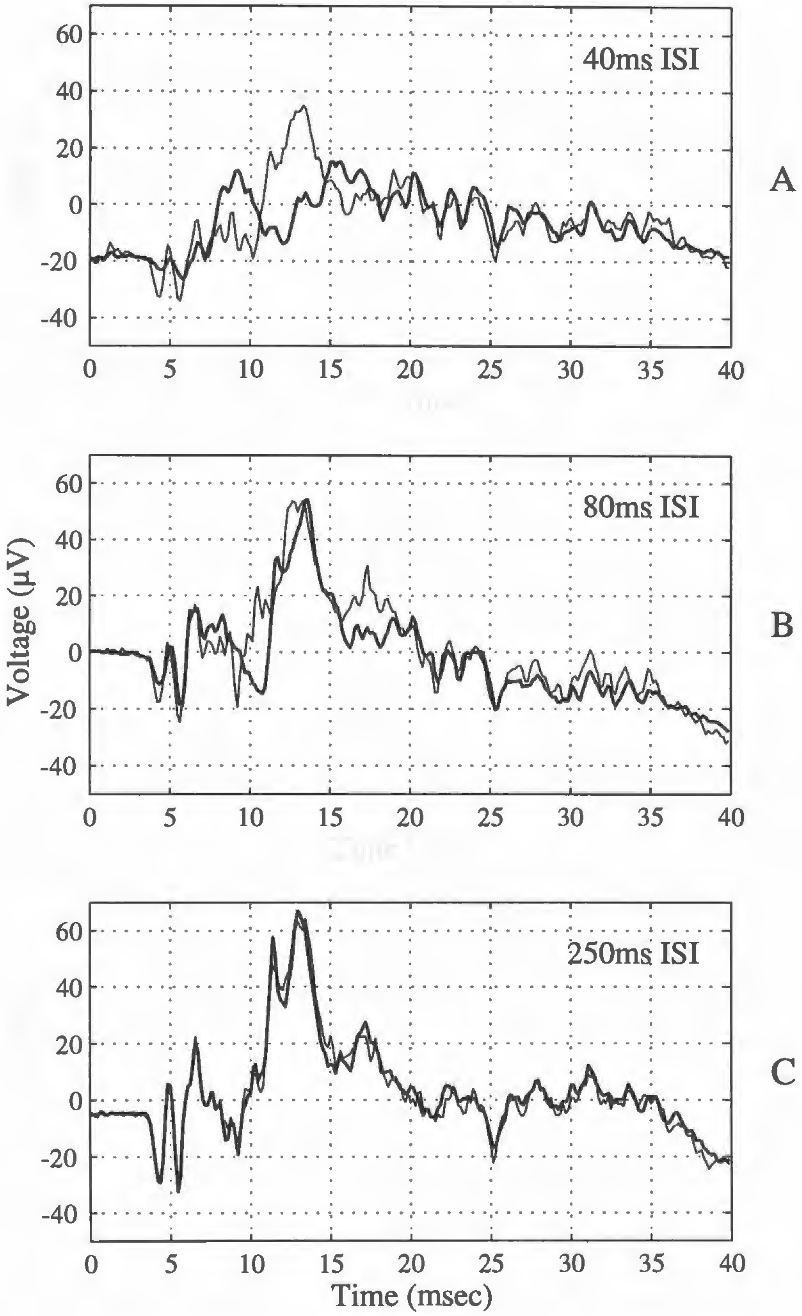
rate, with the exception of the deviant embedded in a continuous stimulus (Fig. 13A). The conjunction of a pitch and spectral deviation (Fig. 13C) produces a marked increase in spike rate at all ISIs and larger responses compared to spectral deviants alone (Fig. 13B).

Global Probability Effects

Responses to harmonic sounds that occurred with 10% probability (deviants) in randomly generated sequences consisting of 400 to 500 stimulus events differed significantly from responses to harmonic sounds that occurred with 90% probability (standards). Fig. 14 shows that averaged over 10 recording sites in two owls, the response strength depended on the ISI [$F(2,18)=6.62$, $p < 0.01$] and also on deviance type [$F(2,18)=9.56$, $p < 0.01$]. A relative RMS value of 1.0 means that the magnitude of the response to deviants matched the response to standards, whereas a value larger than 1.0 indicates that the response to the deviant was stronger. Spectral deviants and combined pitch and spectral deviants evoked the largest responses at all ISIs (Fig. 14). The strength of the response to the deviant relative to the standard decreased at longer ISIs. Spectral deviants at 40 msec ISIs showed an increase of ~18% in their response strength over standards, but only a 4% increase when the ISI was 250 msec. An ISI of 480 msec was tested in one owl. There was a rapid decrease in response strength over the initial 2-4 events to levels approximating those in the 240 msec ISI conditions shown in Fig. 12. However, none of the deviants elicited enhanced responses at an ISI of 480 msec, while responses to ISIs of 40 and 80 msec in this bird were comparable to those described here.

Fig. 15 compares the waveforms (average of 5 sites in one owl) elicited by standard harmonic sounds with those elicited by deviant sounds interspersed randomly among standards in sequences presented at ISIs of 40, 80 and 250 msec. The thin lines show the averaged responses to all spectral and combined pitch/spectral deviants, and the thick lines the average responses to standards. Due to the counterbalancing of harmonic sounds in the role of standard and each of the deviance types across the set of all sequences, the temporal fine structure (pattern of peaks and troughs) in the average waveforms, e.g. between 20 and 35 msec, was the same for standards and deviants. In fact, if there was no effect of sequential context, the response waveforms for standards and deviants should appear to be identical at all ISIs. This was largely true at an ISI of 250 msec (Fig. 15C): the responses to the standards and deviants were almost identical and the onset response (in the range from 5–15 msec) was well developed. At an ISI of 80 msec (Fig. 15B), the morphology of standard and deviant waveforms diverged, primarily between 10 and 20 msec. The onset response was weakest at the 40 msec ISI, and the two waveforms differed between 7.5 and 20 msec (Fig. 15A). The shape of the onset response (peak at 13 msec) in the deviance waveform at the 40 msec ISI resembled more closely the onset responses observed for both standards and deviants at the longer ISIs.

FIGURE 15. Average response waveforms evoked by standard and deviant stimuli in random sequences of harmonic sound events presented at interstimulus intervals of 40 msec (top), 80 msec (middle) and 250 msec (bottom). In each panel, the thin line denotes the response to spectral and combined pitch/spectral deviants and the thick line the response to the standards.

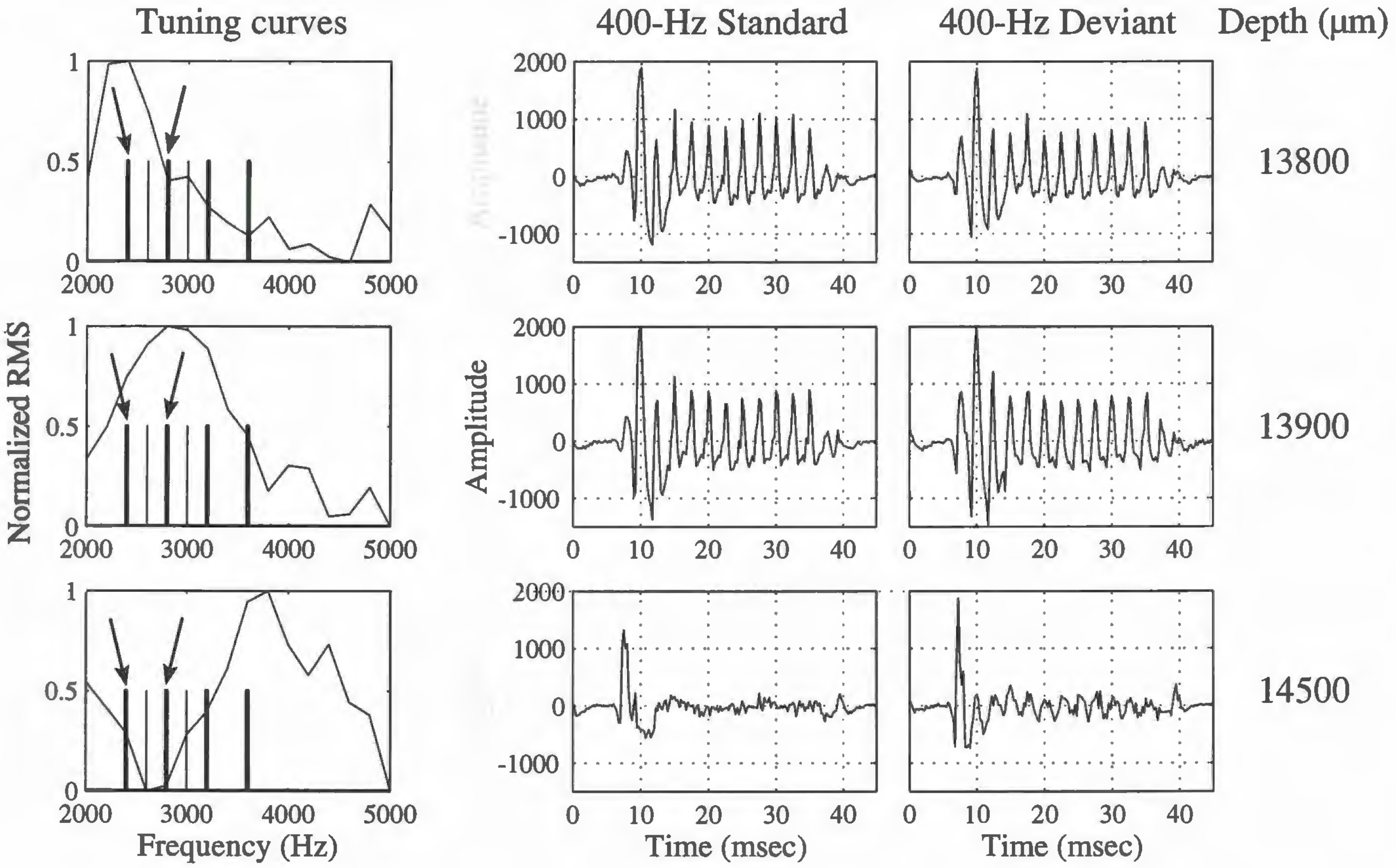

40ms ISI
A
80ms ISI
B
250ms ISI
C
Voltage (μV)
Time (msec)
60
40
20
0
-20
-40
0
5
10
15
20
25
30
35
40

Localization of Sequence Effects on the Tonotopic Map

Sequence effects were demonstrated in the previous sections by averaging responses from multiple recording sites and across all stimuli. This was done so that the results could not be attributed solely to fortuitous interactions of a stimulus, particularly spectral deviances, with the frequency tuning characteristics of the recording site. For example, the response to a sequence consisting of a repeating standard whose component frequencies fell outside the population's tuning curve followed by a deviant whose component frequencies fell into the tuning curve would not be expected to be the same as to a sequence in which the roles of the two harmonic sounds were reversed.

Figures 16 and 17 illustrate how sequence effects were rather localized along the ICc's tonotopic axis and depended on the relative positioning of the stimuli's spectra in the frequency response curve of the neuronal population from which the recording was made. The three rows in Fig. 16 show recordings taken from three neuronal populations with different tuning curves (Fig. 16, left column) in response to the terminal harmonic sound in a sequence of harmonic sounds repeating with a 40 msec ISI. In the "standard" condition (Fig. 16, middle column), every sound in the 24-event sequence had a F_0 of 400 Hz and consisted of the four component frequencies diagrammed by the thick, widely spaced, vertical lines in the left column of Fig. 16. The "deviant" condition consisted of 23 200-Hz F_0 events followed by a single 400-Hz F_0 deviant (Fig. 16, left column, thin closely spaced lines). The reverse conditions were not tested.

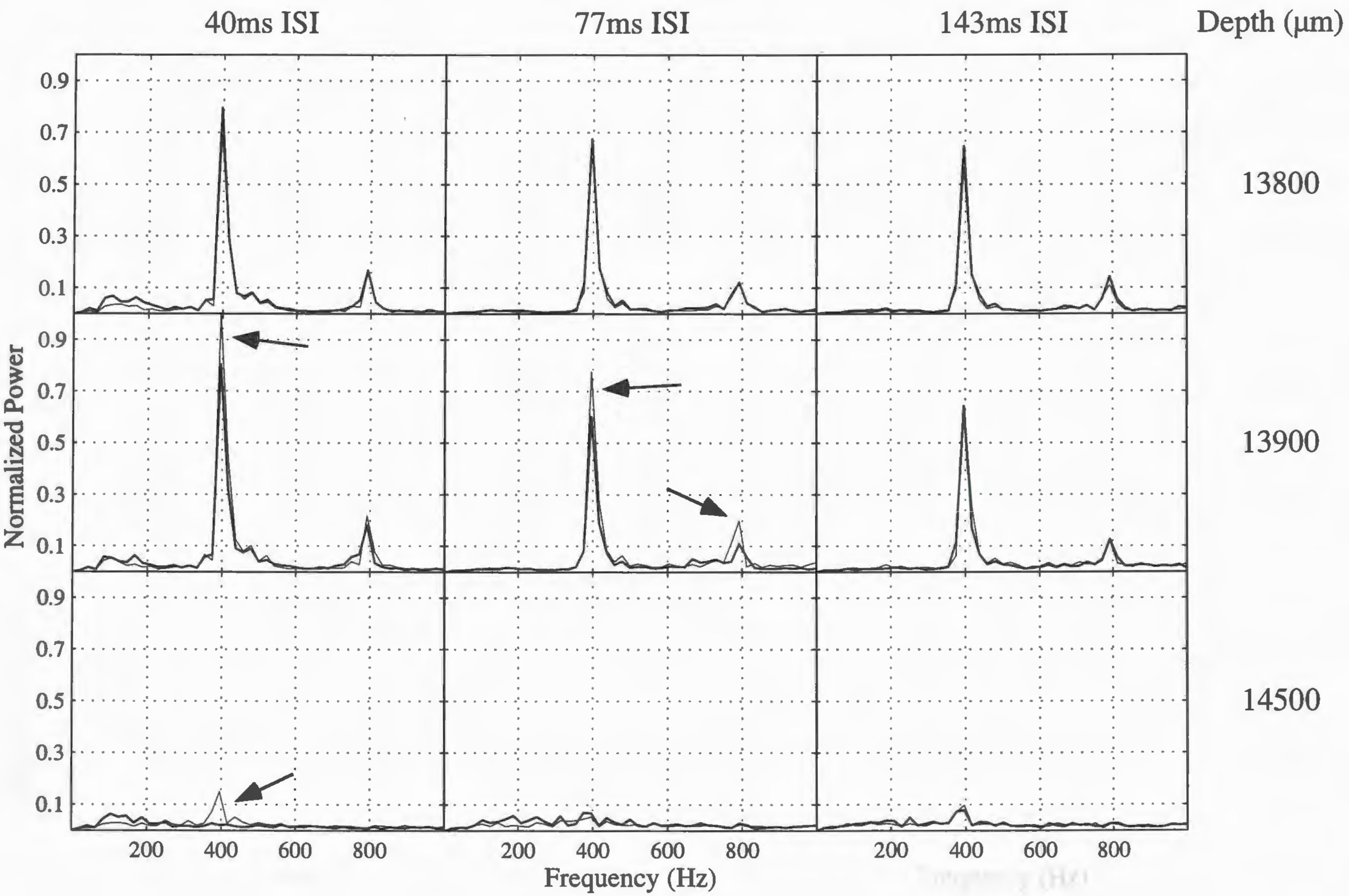
FIGURE 16. Averages (N=25) of potentials evoked by a harmonic sound with a fundamental frequency of 400 Hz when it was the 24th stimulus in a sequence of like-stimuli, or the 1st 400-Hz F_0 stimulus following 23 200-Hz F_0 harmonic sounds. Each row of panels corresponds to data recorded from populations of neurons at different depths relative to the surface of the brain. The frequency response curve for each population is shown on the left with schematic representations of the stimulus spectra superimposed. The thin lines are spaced 200 Hz apart while the thicker lines are spaced 400 Hz apart. The arrows indicate frequency components shared by the two stimuli. The center column of panels shows the responses to 24th sequential presentation of the 400-Hz F_0 stimulus (congruous condition). The ISI was 40 msec. The right column shows the response when the 400 Hz stimulus constituted an incongruous completion to the preceding sequence of 200-Hz F_0 harmonic sounds. The power spectra corresponding to the time-domain averages are shown in the leftmost column of Figure 9.


Tuning curves
400-Hz Standard
400-Hz Deviant
Depth (μm)
13800
13900
14500
Normalized RMS
1
0.5
0
2000
3000
4000
5000
Frequency (Hz)
Amplitude
2000
1000
0
-1000
10
20
30
40
Time (msec)

Additionally, the data were recorded with the high-pass filter set at 10 Hz, thereby eliminating the low-frequency onset and offset responses described in the previous sections. The most apparent difference in the time-domain averages to the terminal 400-Hz F_0 stimulus in both conditions was recorded from the population located at a depth of 14,500 μm (Fig. 16, bottom row). This population had a best frequency of 3,800 Hz. When terminating a sequence of 400-Hz F_0 stimuli, the response to the missing F_0 was suppressed and only an onset response was registered (Fig. 16, middle column, bottom row). If, however, the 400-Hz F_0 stimulus was preceded by repetitions of the 200-Hz F_0 stimulus whose component frequencies fell entirely into the weak and possibly inhibitory region of the frequency response curve, the response to the 400-Hz F_0 stimulus showed an oscillation at the F_0 . The oscillation at the F_0 was also enhanced (larger peak to peak amplitude) in the deviant condition at a recording depth of 13,900 μm (middle panels) during the first 10 msec of the response. As was observed in the data from a different owl shown in Fig. 15A, the difference between standard and deviant stimuli appears to be coded in the initial 5–15 msec of the response to the stimulus.

The relative response magnitudes to the 400-Hz F_0 stimulus serving as standard and deviant are most easily viewed in terms of the average power spectra computed for responses on each trial (Fig. 17). The three rows show power spectra recorded from the neural populations corresponding to the three rows in Fig. 16. Thin lines denote responses to the 400-Hz F_0 stimulus in the deviant condition and thick lines denote responses to the same stimulus serving as the standard. Arrows point to differences in the spectra. Responses were largest at an ISI of 40 msec (Fig. 17, left column).

FIGURE 17. Power spectra comparing responses to the same stimulus in two different contexts. The thick lines show the response to the stimulus when it was the 24th event in a sequence of like stimuli (standard), and the thin lines show when it was a deviant continuation of a sequence of harmonic sounds with a different fundamental frequency. The sequence was presented at a 40 msec ISI (left column), 77 msec ISI (center column) and 143 msec ISI (right column). The three rows correspond to different recording depths, and consequently neural populations, with different tuning curves within the ICc. The positioning of the stimuli relative to the tuning curves is seen in Fig. 16. The power spectra in the leftmost column correspond the time-domain waveforms seen in the middle and right columns of Fig. 8. Arrows point to differences in the power spectra of the same stimulus in congruous and incongruous contexts.


40ms ISI
77ms ISI
143ms ISI
Depth (μm)
13800
13900
14500
Normalized Power
0.9
0.7
0.5
0.3
0.1
200
400
600
800
Frequency (Hz)

At the 13,800 μm recording depth, the response magnitude at the 400-Hz F_0 was the same in the two contexts at all ISIs (overlapping spectra in top row of Fig. 17), whereas in a neighboring group of neurons (13,900 μm, Fig. 17, middle row), the stimulus evoked a significantly larger response as a deviant [$F(1,24)=110.93$, $p < 0.0001$] and at shorter ISIs [$F(2,48)=101.14$, $p < 0.0001$]. The difference between conditions in the weakly responsive population at 14,500 μm that was evident in the time-domain averages in Fig. 16 (bottom row) is clearly seen in the power spectrum at 400 Hz also (Fig. 17, bottom left corner). Thus, the response magnitude to transitions in sequences of harmonic stimuli appears to vary locally and depend on the relationships of the individual stimulus spectra to the frequency response curves of the neuronal populations involved.

Discussion

The results demonstrate that populations of neurons in the inferior colliculus of the barn owl are sensitive to transitions in sequences of rapidly repeating stimuli. An infrequent occurrence of a stimulus elicits a larger response than when the same stimulus serves as the frequently occurring, standard stimulus. This effect is strongest at interstimulus intervals shorter than 100 msec and when the defining characteristic of the deviant stimulus relative to the standard is a shift in the constituent frequencies (spectral distribution) of the harmonic sound (see Figures 12–14). In other words, rapid shifts in the distribution of energy along the tonotopic dimension of the ICc result in an enhanced response to the novel stimulus relative to the ambient response level.

The strength of neural responses to a sequence of harmonic sounds can show either adaptation or potentiation and is sensitive to the repetition rate. At ISIs of 240 msec (approx. 4 stimuli/sec), the response strength may change approx. 15% from the response to the first stimulus in the sequence, whereas at a shorter ISI of 40 msec (25 stimuli/sec) the magnitude of the response can change from 25 to 50% (Figures 10–13). At some recording sites, the response magnitude increased over the course of 1–2 sec. Overall, these results replicate earlier findings by Webster (1971), who found that cat IC responses to sequences of 20 msec pulse bursts presented at 10 bursts/sec (100 msec ISI) also potentiate or adapt. Decrements in response strength are rapid, typically reaching a constant level within 2 or 3 stimulus events.

Studies of habituation-like phenomena in the ascending auditory pathway using single repeating tones, clicks or noise bursts found decrements in response magnitude at interstimulus intervals of up to 10 sec in the cochlear nucleus, inferior colliculus and medial geniculate nucleus of the thalamus in cats (Simons et al, 1965; Kitzes & Buchwald, 1969, Webster, 1971). Response decrements were usually larger at successively higher stations of the auditory pathway (Simons et al, 1965; Webster, 1971). In the present experiment, initial decrements in response strength were observed at all ISIs up to the longest tested (480 msec).

With regard to models of stimulus mismatch detection, the results of this study indicate that information about changes in sequences of auditory stimuli is available at relatively early stages of sensory processing and need not necessarily be extracted by telencephalic structures, as has been broadly implied by the literature on the mismatch negativity (MMN) in humans (see Näätänen, 1992 for review). This conclusion is supported by findings of deviance sensitivity of the non-primary auditory thalamus in guinea pigs (Kraus et al., 1994a,b) and MMN-like potentials recorded from the IC in cats (Csépe et al., 1993). The extraction of this information seems to occur automatically insofar as the recordings were made from an anaesthetized preparation. Cortical and collicular mechanisms of change detection may, however, differ in one critical regard: the time window over which successive stimuli can be compared and changes detected. In the present study, alterations in responses to a stimulus in the role of spectral deviant compared to the role of standard diminished significantly at ISIs longer than 120 msec and were absent at 480 msec. In humans, the MMN has been demonstrated for ISIs of up to 10 sec given a sufficiently large physical deviance (Böttcher & Ullsberger, 1992), although other studies have set the limit at 4 sec to 8 sec (Mäntysalo & Näätänen, 1987). Csépe et al. (1993) found correlates of the MMN in cat IC in response to deviant (10% global probability) 1 msec tone bursts presented at ISIs of 500 msec, but did not test other ISIs or stimuli.

Extraction of Information About Complex Auditory Patterns?

Although it is natural to construe extraction of information about stimulus transitions in sequences of rapid auditory events in terms of deviance detection and orienting, the responses to deviants reported in the present study may serve a different purpose from providing orienting cues. After all, our auditory environments are filled with stimuli that fluctuate rapidly in their intensity, pitch or spectral envelope, yet we do not orient our attention to a stimulus every time such transitions occur. Instead, patterns of transitions might be used to represent auditory objects whose individual elements (short-term steady-states) might be too brief for individual identification. For example, psychoacoustic studies have shown that sounds consisting of several elements are perceived as a unitary pattern, instead of a series of individually identifiable sounds, when the duration of individual elements is less than 200 msec (Warren & Ackroff, 1976). Repeating patterns consisting of ten 40 msec speech sounds, tone bursts or fixed-structure noise bursts have discriminable identities as a whole, even though the temporal ordering of individual elements cannot be determined (Warren et al, 1990; Warren & Bashford, 1993). Similarly, repeating event patterns consisting of brief tone bursts interspersed with silences of equivalent duration as the tone bursts are perceived as integrated units instead of individual tone bursts if the inter-event intervals are under

100 msec (Royer & Robin, 1986). Thus, perception of acoustic events whose properties vary on a time scale under 200 msec may be based on patterns of stimulus onset times or enhanced responses to stimulus transitions. The results from the present study suggest that information about these stimulus transitions is already present at the collicular level.

Neurophysiological evidence that stimulus patterns might be encoded as units that can be compared across successive iterations of the pattern comes from recent studies of the mismatch negativity in humans that have shown the MMN process to be sensitive to deviations in tone patterns ranging in complexity from alternation between two tones (Nordby et al, 1988; Alain et al., 1994) to a change in the frequency or duration of one tone in a sequence of short tones (Näätänen et al., 1993; Tervaniemi et al., 1994; Winkler & Schröger, 1995). Thus, the cortical mismatch detection system is not sensitive to a physical change from the immediately preceding sensory item, but rather to a change in the overall expected pattern.

Fig. 11A demonstrates how such patterns might be represented at the collicular level, showing that the stimulus onset, offset and transitions during the continuous stimulus are marked by large evoked responses. Similarly, when the overall stimulus consists of a rapid sequence of discrete harmonic sounds, the onset, offset and transitions are clearly evident in the evoked potentials (Fig. 11B,C) and in the RMS measure (Figs. 12 & 13). In this view, a complex pattern is coded as a sequence of stimulus transition times, that may even exist independently of topographical maps of physical attributes (e.g. tonotopy) or be used in conjunction with activity in neural maps to demarcate salient times at which to analyze activity on the map. For example, Fig. 16 shows an enhancement of the neural response at the fundamental frequency of a harmonic sound during the initial 10 msec of the stimulus when it serves as a deviant stimulus. Telencephalic sensitivity to specific combinations of discrete auditory events has been demonstrated for contours of tone-sequences in cats (Weinberger & McKenna, 1988) and bird song syllable combination specificity in songbirds (Margoliash & Fortune, 1992). Another piece of evidence that stimulus onsets subserve auditory object detection comes from a study of the "precedence effect" in the barn owl which shows a suppression both of neural responses as well as behavioral orienting to the 2nd sound in a pair of sounds if the onsets of the sounds occur within 5–10 msec of each other (Keller & Takahashi, 1996).

A Neural Basis for Auditory Stream Segregation?

The range of ISIs over which the most robust IC effects were observed also corresponds closely to that observed in auditory stream segregation phenomena described in humans. In its simplest form, auditory stream segregation refers to the perceptual grouping of discrete auditory events, such as tone bursts, into distinct streams (coherent sequences) as a function of their frequency difference and repetition rate (Bregman, 1990). At slow repetition rates, alternation between two tones of different frequency will be perceived as a single sequence jumping up and down in pitch, while at shorter ISIs the alternating sequence will split into two individual streams consisting of single, repeating tones. For example, van Noorden (1975, as cited in Bregman, 1990) found that a series of alternating tones lying 4 semitones apart will obligatorily segregate into two perceptually separate streams of events at inter-stimulus intervals of 50–150 msec. Although studied primarily using pure tones, streaming has been observed using missing fundamental stimuli such as those used in this study. Complex harmonic sounds with non-overlapping spectra, but possessing the same F_0 (spectral deviants in this study) will form separate streams. Harmonic sounds with overlapping spectra (pitch deviants) will not (Bregman, 1990). These observations match well the effects observed in this study (Figs. 12 & 13). Qualitatively, the spectral and combined pitch/spectral deviants in this study "popped out" more strongly in both the fixed or random sequences than did the simple changes in F_0, and the rapid random sequences with spectral deviants sounded like a continuous stream of the standard pitch, punctuated with distinct events of a different timbre. In contrast, when the deviants were pitch deviants with the same spectral envelope, the sequence sounded like a single warbling stream.

In summary, inferior colliculus responses to harmonic sounds are sensitive to the context in which a harmonic sound event occurs and may serve to provide information about physical transitions in sequences of such sounds to higher auditory centers. Barn owl ICc responses to such transitions are strongest in the range of interstimulus intervals at which perception in humans shifts from identification of individual acoustic events in a series to a holistic representation of the acoustic pattern or a splitting into separate event streams.

CHAPTER IV

BRAIN ELECTRICAL ACTIVITY EVOKED BY IMAGINED MUSICAL AND NON-MUSICAL EVENTS

Introduction

Brain imaging studies in the visual domain, utilitizing positron emission tomography (PET) and functional magnetic resonance imaging (fMRI), have suggested that brain areas, including primary cortical areas, that are activated during the perception of externally presented visual stimuli may also be activated when subjects imagine visual stimuli (Kosslyn et al., 1993; Le Bihan et al., 1993). A recent PET study (Zatorre et al., 1996) and neuropsychological evidence (Zatorre & Halpern, 1993) suggest that the auditory cortex may be involved in musical imagery.

To test the hypothesis that imagined auditory events can activate brain areas in much the same way as heard auditory events, event-related potentials (ERPs) were recorded from the scalp as subjects with moderate to high levels of musical training listened to short, eight-note melodies and then heard the first five notes and continued imagining the completions of these melodies. ERPs from these conditions were compared to each other and to two control conditions in which subjects a) heard five noise bursts and continued imagining 3 noise bursts, and b) heard a series consisting of three to eight noise bursts but performed no imagery task. While the ERP method provides poor localization of active brain regions when compared to PET and fMRI brain imaging methods, it enjoys superior temporal resolution, allowing one to trace sequences of brain activations triggered by a stimulus. Ideally, the sequences of activations can be constrained to loci of activity noted in PET and fMRI studies (Heinze et al., 1994; Snyder et al., 1995). The ERP methodology has been used extensively to study the perceptual and cognitive processes associated with the processing of external stimulus events but it has not found application in the study of purely internal stimulus events because the stimulus onset times of internal events cannot readily be inferred or ascertained by the experimenter. Given these constraints, it has been difficult to separate those components of the auditory ERP waveform that are attributable to activation stemming from the processing of external (exogenous) stimulus events vs activation resulting from an internal (endogenous) cognitive process.

Specifically, it has been proposed that several external and internal processes may give rise to a negative peak in the ERP waveform which has a latency of approximately 100 msec (Näätänen & Picton, 1987; Näätänen, 1990). A "true" N100 component of the auditory evoked potential arises from activation of the superior plane of the temporal lobe, site of the primary and secondary auditory cortices (Vaughan Jr. and Ritter, 1970; Hari et al., 1980; Pantev et al., 1988, 1995), and represents an analysis of physical parameters of exogenous auditory stimuli (Näätänen and Picton, 1987; Näätänen, 1990). While the size and shape of the N100 component may be influenced by endogenous factors, such as selective attention or stimulus mismatch detection (Näätänen and Picton, 1987; Näätänen, 1990; Woldorff and Hillyard, 1991), its existence has not been demonstrated under conditions of explicit imagery. Evidence is presented here that the N100, the positive component that precedes it, and the P200 can be evoked by imagining a pitch in a melody, leading to the inference that auditory cortices are activated by this act.

Materials and Methods

Subjects and Stimuli

Seven subjects participated in this task after giving informed consent. The data from one subject were excluded for failure to adequately perform the behavioral component of the task. The remaining subjects had all received formal musical training (mean = 10.7 years; std. dev. 6.3 yrs). A set of five simple 8-note melodies, e.g. ascending and descending scales and arpeggios, were presented in 3 keys in both major and minor modes from two loudspeakers (65 dB SPL) situated approximately six feet in front of the subject. The notes, which had the timbre of a trumpet, were generated by a Kurzweil K2000 synthesizer, and stored in Macintosh 'snd' resources for presentation. Each note was 385 msec in duration and the inter-note interval was 500 msec, resulting in a tempo of 120 beats/minute.

The experimental conditions are schematized in Figure 18. In the heard melody condition, subjects listened to the melody and pressed a response-key synchronously with the onset of the last note. One second after the onset of the last note, subjects heard a probe note and responded as quickly and accurately as possible whether this note was the same as the last note in the melody. Shortly thereafter, subjects heard the first five notes of the same melody and were asked to continue imagining the final three notes (imagined melody condition), pressing a key when they thought the last note would have occurred. They then heard a probe note and responded whether it was the same as the last note they imagined. Interspersed with the melody trials were trials in which subjects heard 5 noise bursts (150 msec duration, 500 msec ISI) and were asked to continue imagining 3 noise bursts, pressing a key when they believed the eighth noise burst would have occurred. To control for ERPs that might be generated by stimulus omissions (Simson et al., 1976), subjects also received a condition in which they were

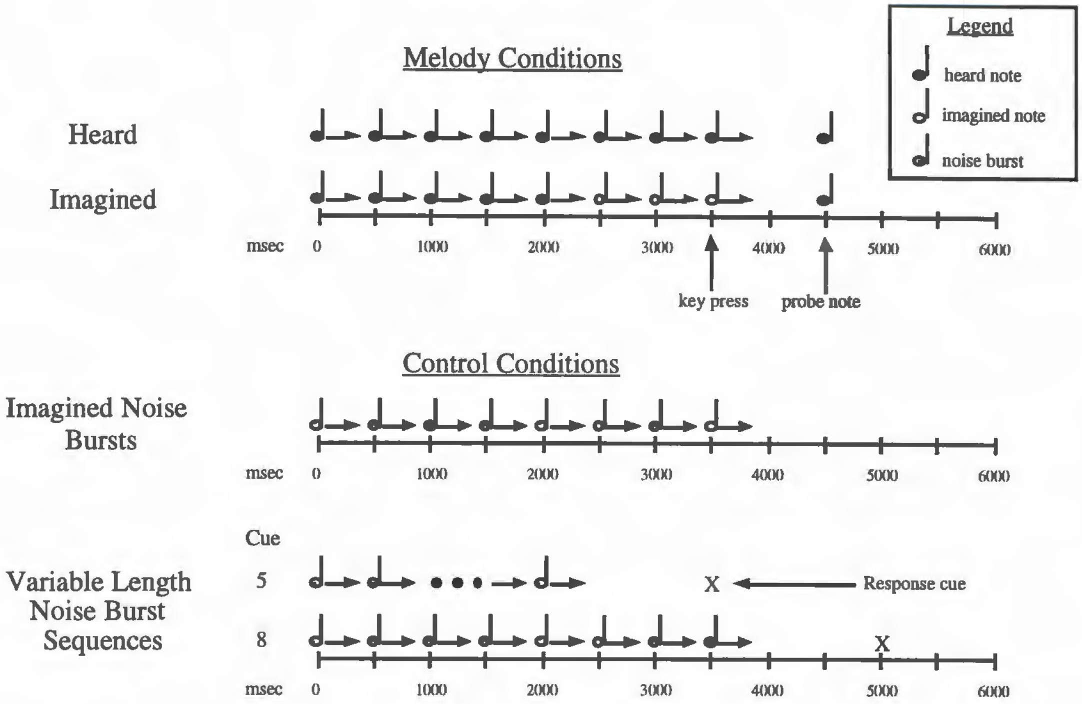


FIGURE 18. Experimental design. Brain electrical activity was recorded in each of 4 experimental conditions. In the Heard Melody (HM) condition, subjects were presented with a simple 8 note melody and were asked to press a key synchronously with the last note in the melody. Following a brief pause, they heard a probe note and were to determine as quickly and accurately as possible whether the probe note was the same as the last note in the melody. Immediately thereafter, the first five notes of the same melody were played in the Imagined Melody (IM) condition and subjects were to continue imagining the last three notes, pressing a key when they thought the last note would have occurred. As in the HM condition, a probe note was played and subjects were to respond whether it was the same as the last note they had imagined. Interspersed in blocks of the melody conditions, subjects received trials on which they heard a series of five identical noise bursts and were to continue imagining three more noise bursts (Imagined Noise Bursts condition), pressing a key when they thought the last noise burst would have occurred. In a final control condition (Variable Length Noise Burst Sequences), subjects were given a cue that predicted with 90% accuracy the number of noise bursts they would hear in the next sequence of noise bursts. Two and one-half seconds following the last noise burst in the sequence, they received a prompt to respond whether the number of noise bursts they heard was the same as the number predicted. There were no imagery or other response requirements in this condition.

told the number of noise bursts they would hear in a train of noise bursts, before the start of each trial.

On 10% of the trials, the number of noise bursts that was predicted did not match the number actually heard. 2500 msec following the onset of the final noise burst in the train, subjects received a cue to which they responded if the number predicted matched the actual number heard.

EEG recording

The EEG, recorded from 129 scalp electrodes using a Geodesic Electrode Net (Electrical Geodesics, Inc.), was low-pass filtered (50 Hz) and digitized at 125 samples/second. All voltage values are expressed relative to the average voltage across all sites. In the melody and imagined noise burst conditions, the EEG was recorded for 4 seconds starting 50 msec prior to the 5th note or noise burst. In the variable-length noise burst sequence condition, the EEG was recorded for 4 seconds starting 200 msec prior to the onset of the penultimate noiseburst. Subjects received 120 trials in each of the melody conditions and 66 trials in the noise burst conditions.

Data analysis

Data were visually inspected for eye-movement and electrode artifacts and trials with excessive artifacts were rejected from further analysis. ERPs were constructed for each of the experimental conditions by averaging all good trials. EEG data from the heard and imagined melody conditions were also analyzed in the frequency domain in order to assay the time-course of power changes in a 5.5 to 9.2 Hz frequency band. FFTs were computed for a sliding 256 msec window across the length of the trial. Each successive window overlapped 87.5% with the previous window. The rationale for looking at this particular frequency band is presented with the results.

Topographical mapping

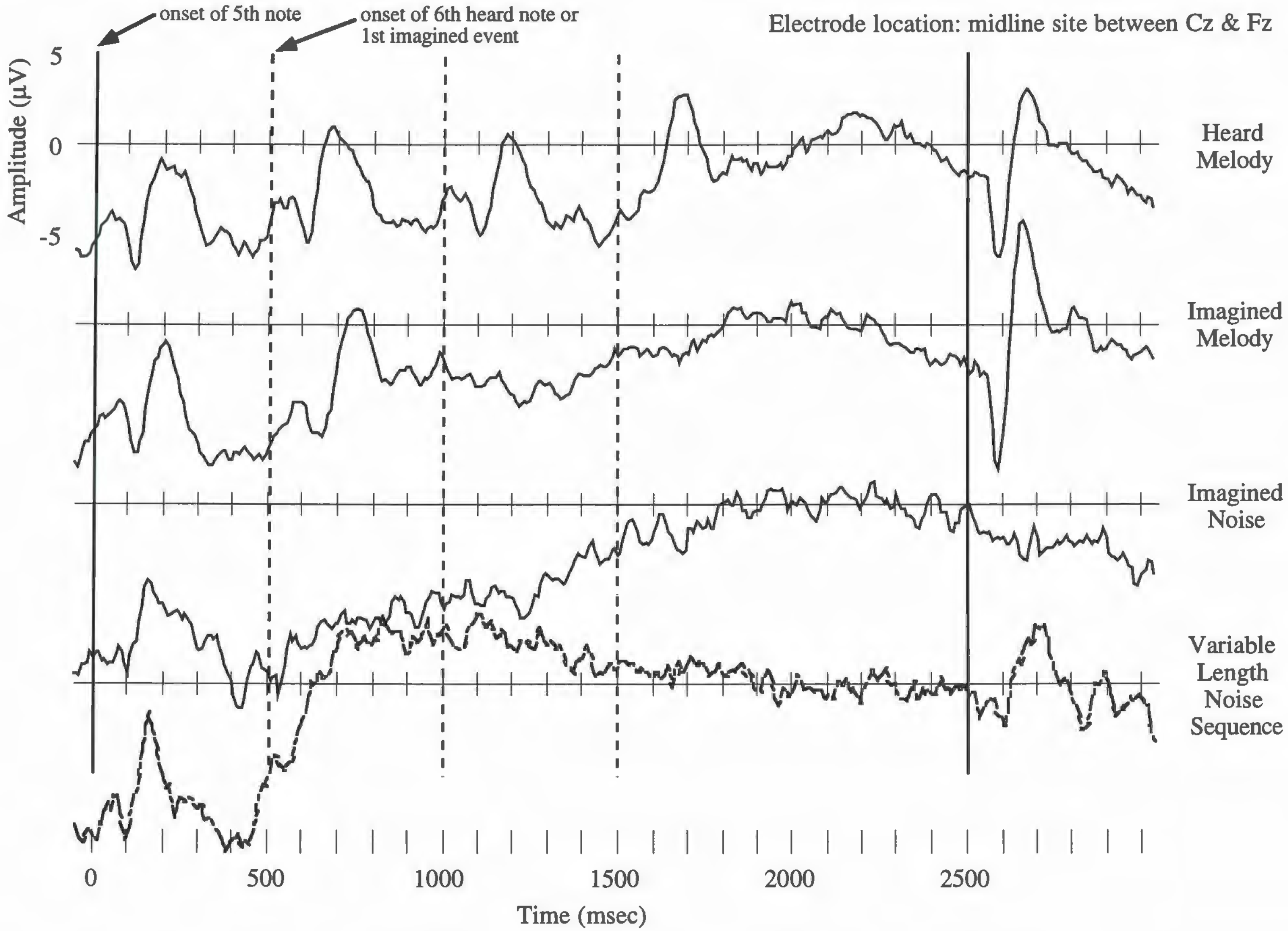
In order to compare the scalp topographies of the waveform components that were similar across the heard and imagined melody conditions, interpolated images were computed for samples corresponding to 24 msec windows surrounding the peaks of the N100 and P200. Interpolations were based on data from all 129 electrodes and were computed using a 3-D spline algorithm (Srinivasan et al., 1996).

Results

ERPs to Heard and Imagined Musical Events

The grand-average waveform of individual subjects' ERP waveforms (N=6 subjects) recorded from a midline electrode site, located 6 cm in front of the vertex (Cz), in each of the experimental conditions is shown in Figure 19. The main points of comparison among the four conditions are the evoked potentials to events marked by the solid and dashed vertical lines at the left of the figure. The solid line demarcates the onset of the 5th note in the melody conditions and the last noise burst in the noise burst conditions. In all conditions, except the heard melody condition, this was the last heard event.

FIGURE 19. Grand average ERPs (N=6) in each of the experimental conditions, recorded from an electrode site situated along the midline, 6 cm anterior to the vertex (Cz). The leftmost solid vertical line marks the onset of the 5th note in the melody conditions and the last noise burst in the noise burst conditions. The dashed vertical lines indicate the onset times of the 6th, 7th and 8th notes in the complete heard melody condition (open circles) and the predicted onset times of imagined events in both the imagined melody (open triangles) and imagined noise burst (solid line) conditions. There were no external or anticipated events during these times in the variable length noise burst condition (dashed line). The solid vertical line on the right side of the graph indicates the onset of the probe note about which subjects had to make a decision in both the heard and imagined melody conditions.


onset of 5th note
onset of 6th heard note or 1st imagined event
Electrode location: midline site between Cz & Fz
Amplitude (µV)
5
0
-5
Heard Melody
Imagined Melody
Imagined Noise
Variable Length Noise Sequence
0
500
1000
1500
2000
2500
Time (msec)

The auditory ERPs evoked in each of these conditions consist of the positive/negative/positive waveform (P1/N100/P200) complex that typifies auditory evoked potentials. The leftmost dashed line indicates the onset of the 6th note in the heard melody condition, the first of three imagined notes in the imagined melody condition, the first of three imagined noise bursts in the imagined noise burst condition, and the momentary cessation of any task-specific activity in the variable length series of noise bursts condition. The task instruction to imagine the next three notes in the melody resulted in the appearance of a positive/negative/positive waveform similar to the heard P1/N100/P200. Both the negative and positive peaks in this waveform were delayed by 48 msec relative to their counterparts in the heard melody condition. This delay may reflect a cost associated with switching from the more passive perceptual portion of the task to the cognitively more engaging portion of the task, which involved both a timed key-press at the onset of the 8th heard note/3rd imagined note and a subsequent forced choice judgement about a probe tone (solid vertical line on the right). Figure 20 shows the replicability of the imagery effect in several single-subject average ERPs. In most subjects, the evoked potential was larger for the imagined note than for the heard note, possibly reflecting an attentional modulation of the ERP amplitude.

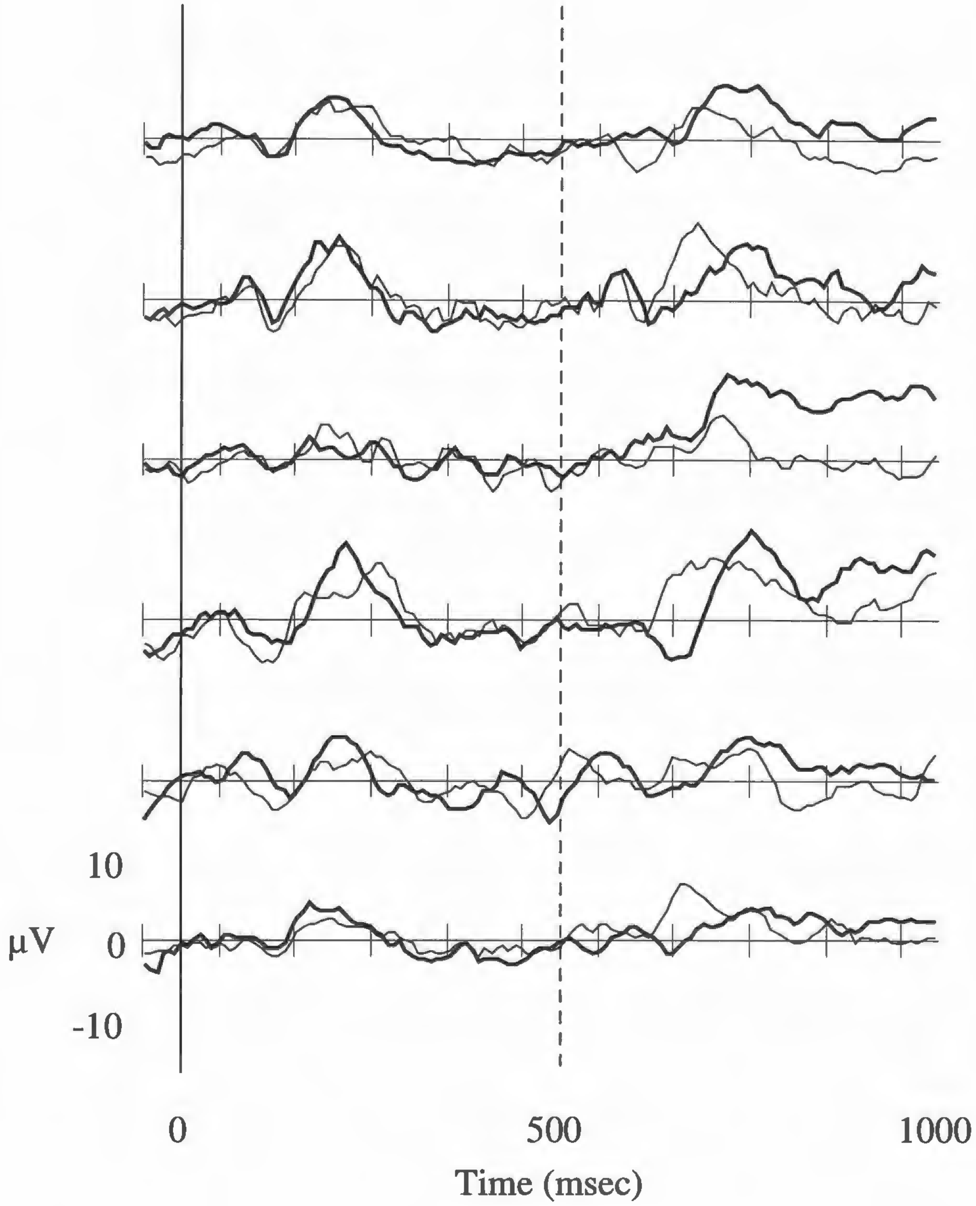


FIGURE 20. Average ERPs for 6 individual subjects in response to the 5th and 6th heard notes in the heard melody condition (thick line), and the 5th heard note and 1st imagined note in the imagined melody condition (thin line). Same electrode site as in Figure 19.

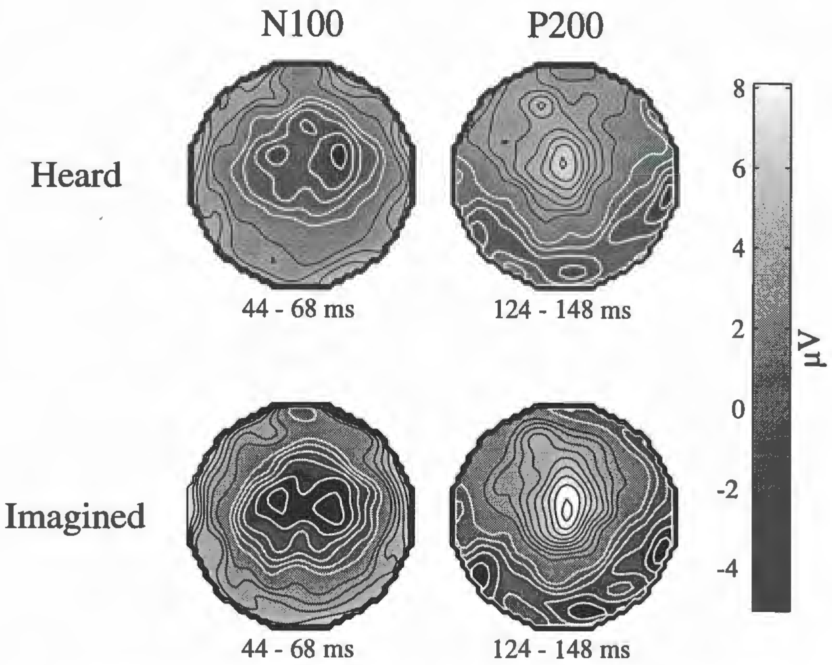


FIGURE 21. Topographical distribution of the scalp potential, averaged across subjects, of the N100 and P200 components in response to heard probe notes about which subjects had to make a judgement in the heard and imagined melody conditions. Each topographical image is a view down on the center of the head and is the average of topographies generated for three successive samples spanning a 24 msec time-window.. The nose would be located at the top of the panel and the left and right ears on the left and right sides, respectively. Dark and light colors represent negative and positive voltages, respectively. Contour interval: 1 μV. The top row shows the topographical distributions of the N100 and the P200 in the heard melody condition, respectively. The N100 topography is characterized by a negativity recorded dorsally and a positivity recorded ventrally, while the P200 shows a maximum at central and frontal sites along the midline and a minimum at temporo-occipital sites. The bottom row shows largely similar N100 and P200 topographies for the probe note in the imagined melody condition.

Scalp topography

The relationship between what appear to be the same N100/P200 waveform components evoked in response to heard and imagined notes (Figure 19) is best depicted in a topographic analysis of the scalp surface potential distribution at each of the waveform peaks (Figures 21–23). Each topographical plot is a view down onto the top of the head (the nose would be at the top of the circle), and should be thought of as wrapping down over the sides of the head. The interpolations are truncated slightly before the outer ring of electrodes which extend approximately 3 cm beyond the outer ring of the 10–20 electrode placement standard. For reference purposes, the topographical distributions of N100 and P200 components to the probe notes at the end of the heard and imagined melody conditions are shown in Figure 21. These topographies reflect the pattern of brain activity that is evoked by a target note to which a speeded judgement must be made. The N100 is characterized by a negative field distributed around the vertex, and a positive field recorded around the bottom rings of electrodes. Such a polarity inversion is the hallmark of the auditory N100 component, and it has been used to infer that the neuronal generators of this component lie in the superior plane of the temporal lobe (Vaughan Jr. and Ritter, 1970; Hari et al., 1980), site of the primary and secondary auditory cortices. The topography of the P200 shows a distribution of the positive portion of the potential field across central and frontal sites, and the negative portion across the back of the head.

Most pertinent to the present study is a comparison of the topographies at the time points corresponding to the apparent N100 evoked by the 1st imagined note in the imagined melody condition and its counterpart (6th heard note) in the heard melody condition (Figure 22). Although delayed by 48 msec, the topography of the negative peak in the imagined waveform is very similar to the heard N100's topography, suggesting that the underlying neuronal generators responsible for these waveform components are largely the same. In both cases, the maximal negativity is recorded at the vertex, and the maximal positivity at inferior electrode sites. In the imagery condition, the N100 distribution is accompanied by a strong infero-frontal positivity that is lateralized primarily to the right side.

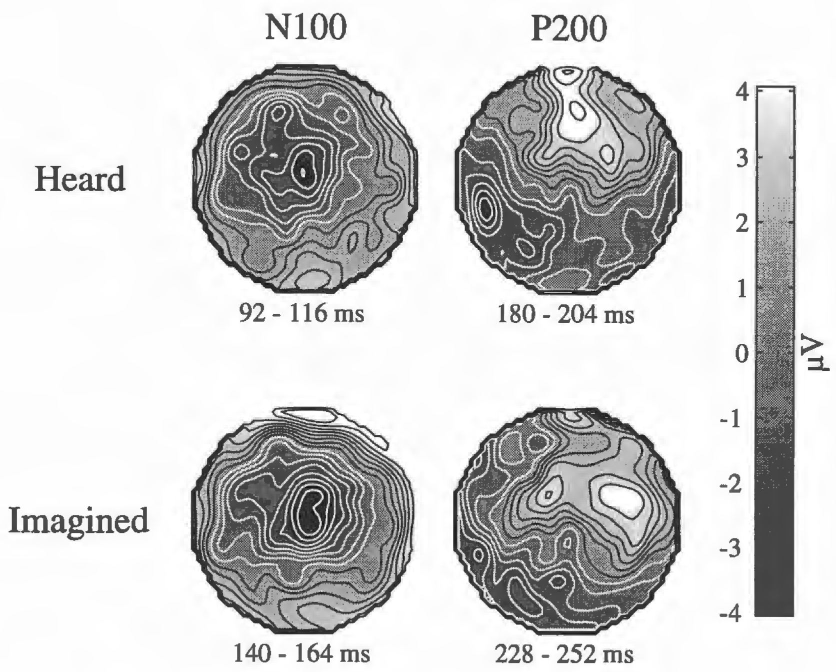


FIGURE 22. Topographical distribution of scalp potentials of N1/P2 complexes evoked by heard and imagined musical events. The left and right images in the top row show the N100 and P200 evoked by the 6th heard, respectively. Note the shift of the positive focus to more frontal sites compared to the P200 topographies shown in Figure 21. The bottom left plot shows the N100 evoked by the 1st imagined note, and the bottom right image the P200 evoked by the 1st imagined note. The N1/P2 complex was delayed by 48 msec in the imagined melody condition relative to its counterpart in the heard melody condition. The imagined N100 is accompanied by a strong positive focus at extreme frontal and right infero-frontal locations. The P2 in the imagined melody condition is distinctly more lateralized to the right than its heard melody counterpart. Contour interval: 0.67 μV. Construction and orientation of images as in Figure 21.

In contrast to the N100, the scalp distribution of the P200 diverges between the two melody conditions. The heard P200 is characterized by an anterior positivity and posterior negativity, with the positivity largest along the frontal midline. Note the more anterior distribution of the P200 to the 6th note in the melody relative to the P200 evoked in response to the probe note. The scalp topography to the imagined note at the corresponding peak is distinctly more right lateralized and shows several focal positive loci above frontal, central and temporal sites. This finding supports results from EEG and PET studies which show right hemisphere activation during various pitch discrimination and musical judgement tasks (Zatorre et al., 1992, 1994, 1996; Janata and Petsche, 1993).

Responses to the 2nd and 3rd imagined notes

Evoked potentials to the 2nd and 3rd imagined notes are not evident in the imagined melody's time waveform (Figure 19). Several factors may account for this result. First, subjects displayed a significant degree of variability in the time at which they pressed the key as an indication of their expectancy for when the last imagined note should have occurred. On average, key presses were made 258 msec (± 89 msec std. dev.) after the expected onset of the final note. If the key-press data reflect a similar variability in the timing of subjects' mental imagery, the temporal variability in the onset of the imagined evoked potential from trial to trial would have the effect of attenuating, broadening and/or cancelling the N100/P200 peaks in the averaged ERP. Second, the strict task demand of pressing a key synchronously with the last imagined note may have induced subjects to switch from a strategy of precisely imagining the proper musical pitch to focussing on making a properly timed response, thus failing to instantiate the representation of the note in auditory cortex as strongly as they were able to in response to the first imagined note. Finally, the evoked responses to imagined events might be expected to show habituation similar to that seen in response to repeated sounds.

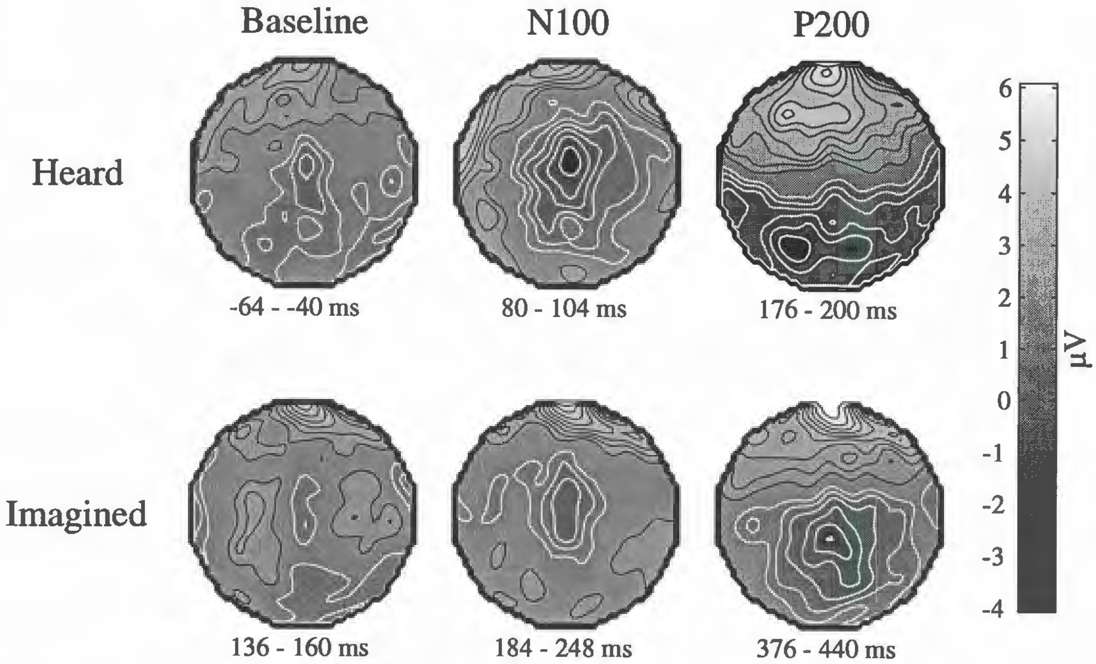


FIGURE 23. Topography of scalp potentials in time-windows related to the 7th heard event and 2nd imagined event. The top left image shows the average topography of the time window 64 to 40 msec preceding the onset of the 7th heard note. No strong foci are present. The N100 and P200 to the 7th heard note are shown in the top-middle and top-right panels, respectively. Unlike the P200 topography for the probe note or 6th heard note, the voltage distribution of the P200 following the 7th heard note is positive at all anterior sites and negative at all posterior sites. The scalp voltage surrounding the 2nd imagined note undergoes similar transitions in topographical organization as that seen for the heard melody condition. Bottom left: average topography of a 24 msec window preceding the approximate time at which the 7th note would be imagined. Bottom middle: voltage distribution observed 184 msec following the putative onset of the 7th note and lasting approximately 64 msec. Bottom right: voltage distribution 376–440 msec following the putative onset of the 7th note. This topography was sustained for approximately 250 msec, and was established following that shown in bottom middle plot with a very similar time-course as that observed for the N100 to P200 transition shown in the top row. Contour interval: 0.67 μV.

Two pieces of evidence suggest that subjects were, in fact, imagining the 2nd note and activating auditory cortex in the process. The first is a similarity in the transitions of scalp voltage topography at the time of the 7th heard note and putative imagining of its counterpart in the imagined melody condition. The top row of images in Figure 23 shows the scalp voltage distribution beginning 64 msec prior to the onset of the 7th heard note and continuing through the N1/P2 time windows. The N100 topography is very similar to that observed for the other heard events and 1st imagined note (Figs. 21 & 22). The P200 topography differs from the other P200 topographies primarily in its strict positive/negative division across anterior and posterior sites, respectively. The scalp topography in the imagined melody condition undergoes a very similar set of state transitions, albeit with a protracted and delayed time-course (Figure 23, bottom row). Two weakly positive foci at centro-temporal sites, a positive focus at extreme frontal and right frontal positions and a weak posterior negativity characterize the 150 msec preceding the suspected onset of the 2nd imagined note (Fig. 23, bottom left). Approximately 184 msec after the theoretical (zero-lag) onset of the 2nd imagined note, a voltage distribution arises that is typical of the N100 topographies observed for the heard and 1st imagined events. Although of smaller overall magnitude, the central negative focus and positive outer ring last for approximately 100 msec and then rearrange to yield a topographical distribution resembling that of a P200 (Fig. 23, bottom right). The P200's negative focus was more centrally positioned in the imagined melody condition. Extreme midline frontal and right inferior frontal sites show a positive focus throughout the imagery but not the heard condition.

Another approach was taken to mitigate the effects of inter-trial variability in the onset of presumed imagined auditory evoked potentials to the 2nd and 3rd imagined notes: computation of time-varying frequency power spectra. Figure 24 shows the changes in power of the 5.5–9.2 Hz frequency band, averaged across all subjects, at the anterior midline electrode site for which the time-domain data are shown in Figure 19. This frequency band captures the power in the slow wave components (N1/P2) of the auditory ERP (Pantev et al., 1991; Makeig, 1993). Consequently, the dashed line shows three peaks whose latencies correspond to the occurrence of notes five through

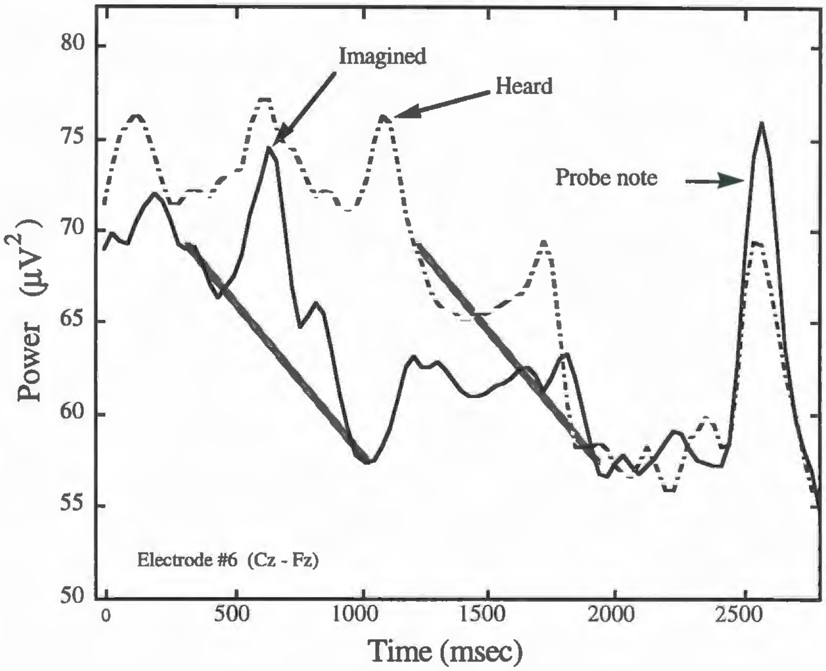


FIGURE 24. Average power across six subjects in the 5.5 to 9.2 Hz frequency band over the course of the EEG recording epoch in the heard and imagined melody conditions. Same electrode site as in Figure 19. The onset of the 5th note in the melody is at time = 0 msec, and the onset time of the probe note is at t = 2500 msec. The dashed and solid lines correspond to the heard and imagined melody conditions, respectively. The two superimposed parallel lines of equal length show changes in baseline power within each of the conditions. The decrease in power may reflect alpha-desynchronization in response to the more cognitively demanding portions of the trial. Inter-trial variability in the phase angle (onset) of an evoked potential will not affect the estimate of average power. To the extent that the N1/P2 waveform can be characterized by power within a narrow frequency band, tracing the power in this frequency band through time will give an indication of when auditory evoked potentials occurred.

seven of the melody. The large peak at the right is the response to the probe note. The plot for the imagined melody condition (Figure 24, solid line) also shows each of these peaks, though the peak believed to correspond to the 2nd imagined note is delayed (200 - 300 msec) and somewhat obscured due to an overall drop in power within this band. The decrease in power is best explained as alpha-desynchronization in response to the switch from the more passive and perceptual task of listening to notes in the melody to the attentionally demanding tasks of imagining notes, making precise timing judgements, and making a forced-choice decision about the probe note (Pfurtscheller and Aranibar, 1977; Van Winsum et al., 1984). In this interpretation, the 500 msec period before the probe note in both the heard and imagined melody conditions reflects the baseline power level during the cognitively demanding portion of the trial. Since task demands increase earlier in the imagined melody condition, the drop in alpha power to the new baseline also occurs earlier (diagonal line on left, Figure 24). Viewed against this baseline, the peaks in the imagined melody condition between 1000 and 2000 msec may reflect responses to the 2nd and 3rd imagined notes. As in the time domain waveforms (Figure 19), the second (first imagined note) and last peaks (probe note) are larger in the imagined melody condition than in the heard melody condition, supporting the assumption that the N100/P200 waveform is characterized by power within this frequency band.

Discussion

The ability of an imagined musical event to generate the N100 and P200 components of the auditory evoked potential in this study is in agreement with the results of an earlier ERP study of musical expectancy which showed an N1/P2 and sustained positive waveform during a period of silence in which subjects were to imagine the best possible resolution to a sequence of chords prior to the occurrence of the actual resolution (Janata, 1995). In this earlier study, however, the possibility of an N1/P2 waveform indexing auditory imagery was confounded with a possible offset response to the chord immediately preceding the period of imagining. Stimuli in the present study were separated by silence, rendering an interpretation of the N1/P2 waveform evoked in the imagined melody condition as an offset response implausible.

Given the similarity of the N100 topographical distributions to the 6th heard and 1st imagined notes, as well as the results of MEG dipole localization studies of the N100 component that localize the source of the N100 component to the superior surface of the temporal lobe (Pantev et al., 1988, 1995), we infer that secondary and possibly primary auditory cortices are activated when subjects are asked to imagine specific pitches in a melody. The recent PET study by Zatorre and colleagues (1996) demonstrated the activation of secondary auditory cortex, but not primary auditory cortex, during a musical imagery task in which subjects had to compare the pitch of the notes associated with two words in the lyric of a familiar melody. Subjects engaged in this task play the melody in their minds until they reach the 2nd of the two words and complete the pitch judgement (Halpern, 1988). When subjects heard the actual melody and lyric, a large portion of the superior temporal gyrus encompassing both primary and secondary auditory cortices showed increased blood flow, while imagining the melody and lyric resulted in sustained activation of secondary auditory cortical regions as well as inferior frontopolar regions.

Given difficulties in estimating the exact locus/loci of N100 generators within the superior temporal plane, and the inability of PET to resolve brain activations resulting from individual stimulus events, it is too early to rule out the possibility that the auditory evoked potential to the 1st imagined note is not activating primary auditory cortex, even though a more conservative interpretation would suggest secondary auditory cortex activation. The strong positive foci at frontal midline and right frontal sites in the imagery portions of the electrophysiological data presented in this paper are highly consistent with the results from PET studies of pitch judgement tasks and the interpretation that several right frontal cortical areas, including anterior cingulate, mid-frontal and inferior frontopolar sites, are activated by pitch retrieval from memory (Zatorre et al, 1994,1996).

Relationship of "Imagined" and "Omitted" or "Emitted" Stimulus Potentials

Other researchers have investigated the scalp potentials that are generated in response to stimulus omissions in sequences of repeating auditory stimuli (Simson et al., 1976). These potentials have been called either "omitted" or "emitted" stimulus potentials. When subjects are asked to count the number of stimulus omissions, a negative component is typically evoked at 230 msec post-stimulus by the absence of a stimulus at the expected time followed by a large positive component distributed across posterior electrode sites at 465 msec post-stimulus (Simson et al., 1976). A recent study of melodic processing reports the occurrence of ERP components that resemble the auditory evoked potential at the time a completion of the melody was expected to occur, but failed to occur (Besson and Faïta, 1995). The negative component of these "emitted potentials" peaked around 200 msec post-stimulus, and the positive component peaked between 300 and 700 msec.

Given these earlier finding, the results of the present study might be interpreted as scalp potentials that are involuntarily elicited by the cessation of a regularly repeating sequence of stimuli, rather than potentials that are voluntarily evoked as a consequence of a task that explicitly requires the generation of a specific image at a specific moment in time. Several aspects of the ERP data shown in Figure 19 speak against the omitted stimulus potential explanation: 1) The absence of an N1/P2 component in the imagined noise burst condition, 2) the absence of an N1/P2 in the condition with no explicit imagery requirement, 3) the absence of an N1/P2 during the silent period following the last note in the heard melody condition, and 4) the absence of a P300 component to the absent, but imagined, stimulus. Additionally, the imagination of a second note in a sequence of imagined notes seems to be indexed by a transition through a N1/P2 topography with a concurrent spectral power peak in a frequency band that captures the N1/P2 waveform, suggesting that the subjects' continued imagination is providing spatio-temporal organization of the evoked brain activity that reflects the mental processes associated with the imagination of discrete musical events at discrete times.

Perhaps the omitted stimulus potential findings in the earlier studies can be explained in terms of the mechanisms that are enabling proper performance in the musical imagery task. The musical imagery task requires that subjects imagine a specific pitch at a precise moment in time. Cues that enable them to do this are familiarity with the melody, intrinsic expectancy-guiding properties of western tonal music that allow for predictions of what the pitch of the next note will/should be, and the establishment of a regular tempo (inter-stimulus interval) over the course of the first five notes in the melody (Jones, 1993). Thus, the expectation to hear a particular note at a specific point in time may facilitate the formation of an image of the note in the auditory cortex at the proper moment. A similar explanation was proposed by Simson et al (1976) to explain their omitted potential results, in which they suggested that the regular timing of their stimuli led to anticipatory activation of the cortex in the supratemporal plane. Likewise, the melodies used by Besson and Faïta (1995) generated strong rhythmic and melodic expectations for the occurrence of a specific pitch at a specific time, about which a congruency judgement had to be made. Therefore, their results can also be explained in terms of activation of brain areas that are responsible for comparing internal and external representations by expectancies created by the preceding musical context. To the extent that the auditory cortex is involved in such a comparative process (as has been suggested by the mismatch negativity literature, see Näätänen, 1990), the salience of the internal stimulus might determine the strength, coherence and time-course with which neuronal populations in the auditory cortex are activated, thereby determining the appearance of such activation at the scalp surface in terms of latency, amplitude, and possibly topographical distribution. Conditions of explicit and highly directed imagery, such as those employed in this study, could then account for a closer match in the recorded latencies of the imagined and heard N1/P2 components than has been observed previously.

Finally, the ability of subjects to evoke an imagined stimulus potential may depend on the specific type of auditory imagery. No N100/P200 component was observed when subjects were asked to imagine bursts of white noise. The absence of these components from the imagined auditory ERPs may reflect the qualitative difference in the noise burst and trumpet sound stimuli the subjects were being asked to imagine. All of the melody notes have specific pitches and a single timbre associated with them. Consequently, their representations, in terms of spectral content, are stable through time. Noise bursts, on the other hand, are a collection of random frequencies occurring at random times, and therefore lack a stable representation. Given the tonotopic organization of the auditory cortex (Merzenich and Brugge, 1973), these two classes of stimuli should possess different neural representations within the cortex. Noise bursts will activate numerous regions of the tonotopic maps sporadically for short periods of time, while pitches will activate more circumscribed regions for the entire duration of the stimulus. In addition, mounting evidence for specific maps of pitch in the auditory cortex (Pantev et al., 1989; Langner et al., 1993) suggests that notes in a melody may have a more elaborate representation within auditory cortex than do noise bursts. The presence of an ERP to an imagined note, but not to an imagined noise burst, may indicate a special ability of simple musical tasks and musical contexts to engage a network of brain areas whose activity is capable of instantiating highly specific mental images in auditory cortex at very specific times.

BIBLIOGRAPHY

Alain, C.; Woods, D.L. & Ogawa, K.H. (1994) Brain indices of automatic pattern processing. *Neuroreport* 6:140-144.

Assmann, P.F. & Summerfield, Q. (1990) Modeling the perception of concurrent vowels: Vowels with different fundamental frequencies. *J. Acoust. Soc. Amer.* 88:680-697.

Bell, C.C. (1989) Sensory coding and corollary discharge effects in mormyrid electric fish. *J. Exp. Biol.* 146:229-153.

Besson, M. & Faïta, F. (1995) An event-related potential (ERP) study of musical expectancy: Comparison of musicians with nonmusicians. *J. Exp. Psychol. Hum. Percept Perform* 21:1278-1296.

Bregman, A.S. (1990) Auditory scene analysis: the perceptual organization of sound. MIT Press, Cambridge.

Cariani, P. & Delgutte, B. (1996a) Neural correlates of the pitch of complex tones. I Pitch and pitch salience. *J. Neurophysiology* in press.

Cariani, P. & Delgutte, B. (1996b) Neural correlates of the pitch of complex tones. II Pitch shift, pitch ambiguity, phase-invariance, pitch circularity, rate-pitch and the dominance region for pitch. *J. Neurophysiology* in press.

Cohen, D. & Erez, A. (1991) Event-related-potential measurements of cognitive components in response to pitch patterns. *Music Perception* 8:405-430.

Csépe, V.; Karmos, G. & Molnar, M. (1987) Evoked potential correlates of stimulus deviance during wakefulness and sleep in cat: Animal model of mismatch negativity. *Electroenceph. clin. Neurophysiol.* 66:571-578.

Csépe, V; Karmos, G. & Molnar, M. (1993) Animal model of mismatch negativity. In: New Developments in Event-Related Potentials (Heinze, H.-J.; Münte, T.F. & Mangun, G.R., eds.). Birkhäuser. Boston. pp.29-33.

de Boer, E. (1976) On the "residue" and auditory pitch perception. In: *Handbook of Sensory Physiology*. Keidel, W.D. & Neff, W.D. (Eds.). Springer Verlag. Berlin. pp 479-583.

Frühstorfer, H.; Soveri, P., & Järvilehto, T. (1971) Habituation and dishabituation of hte human vertex response. *Electroenceph. clin. Neurophysiol.* 28: 153-161.

Halpern, A.R. (1988) Mental scanning in auditory imagery for tunes. *J. Exp Psychol. Learn. Mem. and Cogn.* 14:434-443.

Hari, R.; Aittoniemi, K.; Järvinen, M.L.; Katila, T. & Varpula, T. (1980) Auditory transient and sustained magnetic fields of the human brain. Localization of neural generators. *Exp. Brain Res.* 40: 237-240.

Hari, R.; Kaila, K.; Katila, T.; Tuomisto, T., & Varpula, T. (1982) Interstimulus interval dependence of the auditory vertex response and its magnetic counterpart: Implications for their neural generation. *Electroenceph. clin. Neurophysiol.* 54: 561-569.

Hartmann, W.M. (1988) Pitch perception and the segregation and integration of auditory entities. In: *Auditory Function: Neurobiological Bases of Hearing,* Edelman, G.M., Gall, E.W., and Cowan, M.W., eds., J. Wiley & Sons, Inc., New York, pp. 623-645.

Heinze, H. J.; Mangun, G. R.; Burchert, W.; Hinrichs, H.; Scholtz, M.; Muntel, T. F.; Gosel, A.; Scherg, M.; Johannes, S.; Hundeshagen, H.; Gazzaniga, M. S. & Hillyard, S. A. (1994) Combined spatial and temporal imaging of brain activity during visual selective attention in humans. *Nature* 372: 543-546.

Houtsma, A.J.M. & Smurzynski, J. (1990) Pitch identification and discrimination for complex tones with many harmonics. *J. Acoust. Soc. Amer.* 87: 304-310.

Janata, P. (1995) ERP measures assay the degree of expectancy violation of harmonic contexts in music. *J. Cog. Neurosci.* 7: 153-164.

Janata, P.; Keller, C.H. & Takahashi, T.T. (1994) Pitch coding in the barn owl's inferior colliculus: Responses to complex harmonic sounds. *Soc. Neurosci. Abstr.* 20:319.

Janata, P. & Petsche, H. (1993) Spectral analysis of the EEG as a tool for evaluating expectancy violations of musical contexts. *Music Perception* 10:281-304.

Janata, P. & Takahashi, T.T. (1995) Neural responses to spatio-temporal pitch contexts in the barn owl. *Soc. Neurosci. Abst.* pg 673.

Jaramillo, F.; Markin, V.S & Hudspeth, A.J. (1993) Auditory illusions and the single hair cell. *Nature 364,*527-529.

Javel, E; McGee, J.A.; Horst, J.W. & Farley, G.R. (1988) Temporal mechanisms in auditory stimulus coding. In: *Auditory Function: Neurobiological Bases of Hearing,* Edelman, G.M., Gall, E.W., and Cowan, M.W., eds., J. Wiley & Sons, Inc., New York, pp. 515-558.

Javel, E. & Mott, J.B. (1988) Physiological and psychophysical correlates of temporal processes in hearing. *Hearing Res.34*, 275-294.

Javitt, D.C.; Schroeder, C.E.; Steinschneider, M.; Arezzo, J.C. & Vaughan, H.G. Jr. (1992) Demonstration of mismatch negativity in the monkey. *Electroenceph. clin. Neurophysiol.* 83:87-90.

Johnston, P.A. (1994) Brain Physiology and Music Cognition. Unpublished doctoral dissertation. University of California, San Diego.

Jones, M.R. (1993) Dynamics of musical patterns: How do melody and rhythm fit together? In: T.J. Tighe, W.J. Dowling (Eds.), Psychology and Music: The Understanding of Melody and Rhythm. pp. 67-92.

Jones, M.R.; Boltz, M. & Kidd, G. (1982) Controlled attending as a function of melodic and temporal context. *Percept. & Psychophysics* 32:211-218.

Jones, M.R. & Yee, W. (1993) Attending to auditory events: the role of temporal organization. In Thinking in sound: the cognitive psychology of human audition, S. McAdams & E. Bigand, Eds. Oxford University Press. Oxford. pp 69-112.

Keller, C.H. & Takahashi, T.T. (1996a) A precedence effect in the owl's auditory space map? *J. Comp. Physiol. A 178*,499-512.

Keller, C.H. & Takahashi, T.T. (1996b) Binaural cross-correlation predicts the responses of neurons in the owl's auditory space map under conditions simulating summing localization. *J. Neurosci.* 16:4300-4309.

Kitzes, M. & Buchwald, J. (1969) Progressive alterations in cochlear nucleus, inferior colliculus, and medial geniculate responses during acoustic habituation. *Exp. Neurol.* 25,85-105.

Kosslyn, S.M.; Alpert, N.M.; Thompson, W.L.; Maljkovic, V.; Weise, S.B.; Chabris, C.F.; Hamilton, S.E.; Rauch, S.L. & Buonanno, F.S. (1993) Visual mental imagery activates topographically organized visual cortex: PET investigations. *J. Cogn. Neurosci.* 5:263-287..

Kraus, N.; McGee, T.; Carrell, T.; King, C.; Littman, T. & Nicol, T. (1994). Discrimination of speech-like signals in auditory thalamus and cortex. *J. Acoust. Soc. Amer.* 96: 2758-2768.

Kraus, N.; McGee, T.; Littman, T.; Nicol, T. & King, C. (1994) Encoding of acoustic change involves non-primary auditory thalamus. *J. Neurophysiol.* 72: 1270-1277.

Langner, G. (1992) Periodicity coding in the auditory system. *Hearing Res.* 60: 115-142.

Langner, G.; Sams, M.; Heil, P.; McEvoj, L.K.; Hari, R. & Ahonen, A. (1993) Spatial representation of periodicity pitch in the human auditory cortex. *Soc. Neurosci. Abst.* 19:1423.

Langner, G. & Schreiner, C.E. (1988) Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanisms. *J. Neurophysiol.* 60:1799-1822.

Le Bihan, D.; Turner, R.; Zeffiro, T.A.; Cuénod, C.A.; Jezzard,P. & Bonnerot-,V. (1993) Activation of human primary visual cortex during visual recall: a magnetic resonance imaging study. *Proc. Nat. Acad. Sci. USA* 90:11802-11805.

Licklider, J.C.R. (1951) A duplex theory of pitch perception. *Experientia* 7:128-134.

Llinas, R. & Nicholson, C. (1974) Analysis of field potentials in the central nervous system. In Handbook of Electroencephalography and Clinical Neurophysiology. Amsterdam, Elsevier. pp 61-83.

Makeig, S. (1993) Auditory event-related dynamics of the EEG spectrum and effects of exposure to tones. *Electroenceph. clin. Neurophysiol.* 86: 283-293.

Mäntysalo, S. & Näätänen, R. (1987) The duration of a neuronal trace of an auditory stimulus as indicated by event-related potentials. *Biological Psychology* 24:183-195.

Margoliash, D. (1983) Acoustic parameters underlying the responses of song-specific neurons in the white-crowned sparrow. *J. Neurosci.* 3:1039-1057.

Margoliash, D. & Fortune, E.S. (1992) Temporal and harmonic combination-sensitive neurons in the zebra finch's HVc. *J. Neurosci.* 12:4309-4326.

Markin, V.S. & Hudspeth, A.J. (1995) Gating-spring models of mechanoelectrical transduction by hair cells of the internal ear. *Annu. Rev. Biophys. Biomol. Struct. 24*,59-83.

Merzenich, M.M. & Brugge, J.F. (1973) Representation of the cochlear partition on the superior temporal plane of the macaque monkey. *Brain Res.* 50:275-296.

Moore, B.C.J. (1989) Introduction to the Psychology of Hearing, Third Edition. Academic Press. London.

Näätänen, R. (1990) The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function. *Behav. Brain Sci.* 13: 201-288.

Näätänen, R. (1992) Attention and brain function. Lawrence Erlbaum Associates. Hillsdale.

Näätänen, R. & Picton, T. (1987) The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. *Psychophysiology* 24:375-425.

Näätänen, R.; Schröger, E.; Karakas, S.; Tervaniemi, M. & Paavilainen, P. (1993) Development of a memory trace for a complex sound in the human brain. *Neuroreport* 4:503-506.

Nordby, H.; Roth, W.T. & Pfefferbaum, A. (1988) Event-related potentials to breaks in sequences of alternating pitches or interstimulus intervals. *Psychophysiology* 25:262-268.

Pantev, C.; Bertrand, O.; Eulitz, C.; Verkindt, C.; Hampson, S.; Schuierer, G. & Elbert, T. (1995) Specific tonotopic organizations of different areas of the human auditory cortex revealed by simultaneous magnetic and electric recordings. *Electroenceph. clin. Neurophysiol.* 94:26-40.

Pantev, C.; Hoke, M.; Lehnertz, K.; Lutkenhoner, B.; Anogianakis, G. & Wittowski, W. (1988) Tonotopic organization of the human auditory cortex revealed by transient auditory evoked magnetic fields. *Electroenceph. clin. Neurophysiol.* 69:160-170.

Pantev, C.; Hoke,M.; Lutkenhoner, B. & Lehnertz, K. (1989) Tonotopic organization of the auditory cortex: pitch versus frequency representation. *Science* 246:486-488.

Pantev, C.; Makeig,S.; Hoke, M.; Galambos, R.; Hampson, S. & Gallen, C. (1991) Human auditory evoked gamma-band magnetic fields. *Proc. Nat. Acad. Sci. USA* 88,:8996-9000.

Pfurtscheller, G. & Aranibar, A. (1977) Event-related cortical desynchronization detected by power measurements of scalp EEG. *Electroenceph. clin. Neurophysiol.* 42: 817-826.

Ritter, W.; Vaughan, Jr. H.G. & Costa, L.D. (1968) Orienting and habituation to auditory stimuli: A study of short term changes in average evoked responses. *Electroenceph. clin. Neurophysiol.* 25: 550-556.

Royer, F.L. & Robin, D.A. (1986) On the perceived unitization of repetitive auditory patterns. *Perception & Psychophysics* 39:9-18.

Saarinen,J.; Paavilainen,P.; Schoger,E.; Tervaniemi,M. & Naatanen,R. (1992) Representation of abstract attributes of auditory stimuli in the human brain. *Neuroreport.* 3:1149-1151.

Schreiner, C.E. & Langner, G. (1988) Periodicity coding in the inferior colliculus of the cat. II. Topographical organization. *J. Neurophysiol.* 60:1823-1840.

Simons, L.A.; Dunlop, C.W.; Webster, W.R., & Aitkin, L.M. (1966) Acoustic habituation in cats as a function of stimulus rate and the role of temporal conditioning of the middle ear muscles. *Electroenceph. clin. Neurophysiol.* 20:485-493.

Simson, R.; Vaughan Jr., H.G. & Ritter, W. (1976) The scalp topography of potentials associated with missing visual or auditory stimuli. *Electroenceph. clin. Neurophysiol.* 40: 33-42.

Snyder, A.; Abdullaev, Y. G.; Posner, M. I. & Raichle, M. E. (1995) Scalp electrical potentials reflect regional cerebral blood flow responses during processing of written words. *Proc. Nat. Acad. Sci. USA* 92:1689-1693.

Sokolov, E.N. (1960) Neuronal models and orienting reflex. In M.A.B. Brazier (Ed.), The central nervous system and behaviour. Madison, N.J.: Madison Printing. pp. 187-276.

Srinivasan, R.; Nunez, P.L.; Tucker, D.M.; Silberstein, R.B. & Cadusch, P.J. (1996) Spatial sampling and filtering of EEG with spline laplacians to estimate cortical potentials. *Brain Topogr.* 8:1–12.

Steinschneider, M.; Schroeder, C.E.; Arezzo, J.C. & Vaughan, Jr., H.G. (1994) Speech-evoked activity in primary auditory cortex: effects of voice onset time. *Electroenceph. clin. Neurophysiol.* 92: 30-43.

Sutter, M.L. & Margoliash, D. (1994) Global synchronous response to autogenous song in zebra finch HVc. *J. Neurophysiol.* 72:2105-2123.

Takahashi, T.T. & Keller, C.H. (1994) Representation of multiple sound sources in the owl's auditory space map. *J. Neurosci.* 14: 4780-4793.

Terhardt, E.; Stoll, G. & Seewann, M. (1982) Pitch of complex signals according to virtual-pitch theory: test, examples and predictions. *J. Acoust. Soc. Amer.* 71:671-678.

Tervaniemi, M.; Maury, S. & Näätänen, R. (1994) Neural representations of abstract stimulus features in the human brain as reflected by the mismatch negativity. *Neuroreport,* 5:844-846.

Tervaniemi,M.; Saarinen,J.; Paavilainen,P.; Danilova,N. & Naatanen,R. (1994). Temporal integration of auditory information in sensory memory as reflected by the mismatch negativity. *Biological Psychology* 38:157-167.

Tueting, P.; Sutton, S. & Zubin, J. (1971) Quantitative evoked potential correlates of the probability of events. *Psychophysiology* 7:385-394.

Van Winsum, W.; Sergeant, J. & Geuze, R. (1984) The functional significance of event-related desynchronization of alpha rhythm in attentional and activating tasks. *Electroenceph. clin. Neurophysiol.* 58: 519-524.

Vaughan Jr., H.G. & Ritter,W. (1970) The sources of auditory evoked responses recorded from the human scalp. *Electroenceph. clin. Neurophysiol.* 28: 360-367.

Wagner, H. (1990) Receptive fields of neurons in the owl's auditory brainstem change dynamically. *Europ. J. Neurosci.* 2, 949-959.

Wagner, H.; Takahashi, T.T. & Konishi, M. (1987) Representation of interaural time difference in the central nucleus of the barn owl's inferior colliculus. *J. Neurosci.* 7,3105-3116.

Warren, R.M. (1993) Perception of acoustic sequences: global integration versus temporal resolution. In Thinking in sound: the cognitive psychology of human audition, S. McAdams & E. Bigand, Eds. Oxford University Press. Oxford. pp 37-68.

Warren, R.M. & Ackroff, J.M. (1976) Two types of auditory sequence perception. *Perception & Psychophysics* 20:387-394.

Warren, R.M. & Bashford, J.A., Jr. (1993) When acoustic sequences are not perceptual sequences: The global perception of auditory patterns. *Perception & Psychophysics* 54:121-126.

Warren, R.M. & Obusek, C.J. (1972). Identification of temporal order within auditory sequences. *Perception & Psychophysics*, 12:86-90.

Warren, R.M.; Bashford, J.A. Jr. & Gardner, D.A. (1990) Tweaking the lexicon: organization of vowel sequences into words. *Perception & Psychophysics* 47(5): 423-32.

Warren, R.M.; Obusek, C.J.; Farmer, R.M. & Warren, R.P. (1969). Auditory sequence: confusion of patterns other than speech or music. *Science* 164:586-587.

Webster, W.R. (1971) The effects of repetitive stimulation on auditory evoked potentials. *Electroenceph. clin. Neurophysiol.* 30:318–330.

Wild, J.M. (1987) Nuclei of the lateral lemniscus project directly to the thalamic auditory nuclei in the pigeon. *Brain Research 408*, 303-307.

Winkler, I. & Schroger, E. (1995) Neural representation for the temporal structure of sound patterns. *Neuroreport* 6(4):690-694.

Woldorff, M.G. & Hillyard, S.A. (1991) Modulation of early processing during selective listening to rapidly presented tones. *Electroenceph. clin. Neurophysiol.* 79:170-191.

Woods, D.L. & Alain,C. (1993) Feature processing during high-rate auditory selective attention. *Percept. & Psychophys.* 53(4):391-402.

Zatorre, R.J.; Evans, A.C. & Meyer, E. (1994) Neural mechanisms underlying melodic perception and memory for pitch. *J. Neurosci.* 14:1908-1919.

Zatorre, R.J.; Evans, A.C.; Meyer, E. & Gjedde, A. (1992) Lateralization of phonetic and pitch discrimination in speech processing. *Science* 256: 846-849.

Zatorre, R.J. & Halpern, A.R. (1993) Effect of unilateral temporal-lobe excision on perception and imagery of songs. *Neuropsychologia* 31: 221-232.

Zatorre, R.J.; Halpern, A.R.; Perry, D.W.; Meyer, E. & Evans, A.C. (1996) Hearing in the mind's ear: A PET investigation of musical imagery and perception. *J. Cogn. Neurosci.* 8:29-46.