Title
Cortical processing of speech-like sounds

Author
Keeling, Diane M.

Publication Date
1993

Peer reviewed|Thesis/dissertation

CORTICAL PROCESSING OF SPEECH-LIKE SOUNDS

by

M. Diane Keeling

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Speech and Hearing Sciences

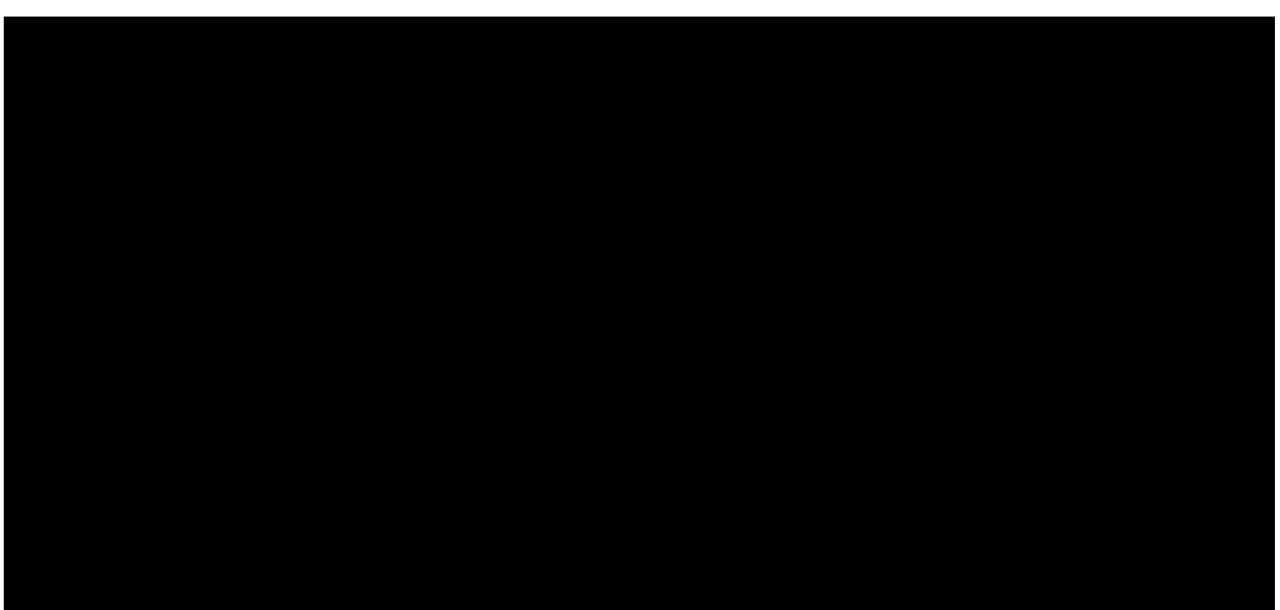
in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA

San Francisco



ACKNOWLEDGEMENTS

I acknowledge the invaluable contribution of the animals that provided information for this research. I would like to thank my advisor Dr. Christoph Schreiner for his guidance, enthusiasm, patience, and support. His knowledge of the neurophysiology of audition is remarkable; his teaching ability is greatly appreciated by his many students. I would also like to thank Dr. Michael Merzenich and Dr. William Jenkins. Their advice, assistance, and encouragement were vital. Dr. Earl Schubert provided valuable input. Dr. Walter Bruszewski helped me get the show on the road by artfully creating the animal response test apparatus; Marshall Fong directed the electronics end of the test system. Tricia Chambers and Barbara Calhoun provided intellectual and moral support. Steven Brown was a good friend. Finally, thanks to my husband, Barry Rahn, for his ever constant love and support; yes, Barry, I'm finally done.

Abstract

The objective of this study was to investigate the perception of formant spacing in vowel-like stimuli, by measuring the ability of humans and cats to detect frequency shifts in formant locations. A second objective was to see whether the training and overlearning in cats would modify the cortical representation of these stimuli. Perception of a vowel-like ripple stimulus was investigated in both a psychophysical procedure and, in animals, in electrophysiological recording of multiple unit responses in primary auditory cortex. Six humans were tested in a modified 2-Alternative-Forced-Choice paradigm for their ability to discriminate envelope phase shifts in vowel-like ripple stimuli. Thresholds were lowest for ripple densities below approximately 4 ripples/octave - the peak spacings that correspond to the majority of formant ratios in English vowels. Three cats were trained and tested in a similar procedure for their ability to discriminate envelope phase shifts at a ripple density of 1 ripple/octave. Their behavioral thresholds correlated with the Q-40 values of overall neural responses, indicating that a sharpening in tuning was related to the degree of proficiency in the auditory discrimination. Cortical ripple density representations and envelope phase shift representations were significantly different in trained cats as compared to naive controls. These data imply that language acquisition involves dynamic cortical reorganization of speech feature representations and that human linguistic capabilities evolved from basic mammalian auditory processing capacities.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

Background

By means of the senses, the brain extracts information from the world around us. Audition connects us with our fellow creatures and provides the avenue for human communication. The production and perception of speech are the bedrock of language, thought, indeed cultural evolution. Speech is intimately linked with our sense of our selves and our conscious behavior.

It is generally believed that the vocal and auditory systems have co-evolved so that "the vocal system has adapted to produce sounds suitable for detection and processing by the auditory system, and the auditory system has evolved to detect and process these sounds" (Suga, 1988, p.684). The development of language in humans may have taken advantage of those acoustic parameters that accentuate certain differences in a manner analogous to the way in which vowels are discriminated using the formants - those frequencies which are boosted in the particular conformations of the vocal tract. It may also be the case that certain acoustic cues common to speech and to other biologically significant communicative sounds are processed by basic neural mechanisms that are similar in humans and in other animals.

Speech perception is one of the most elaborate feats performed by the human brain, and while a number of models of speech perception have been posited and electrophysiological recording of cortical neuronal response to various acoustic signals has been performed, only speculative theories about the general physiological processes underlying speech perception have so far been possible. The issues involved are not simple.

The first complexity to be appreciated is the fact that the correspondence between the speech unit perceived and its acoustic representation is not invariant, that is, the acoustic representation of a phoneme (the smallest acoustic segment in speech, e.g. the /d/ sound in the word "dog") varies according to the phonemic context in which it occurs. See for example Figure 1, page 4. Nor do the acoustic segments present themselves for processing in a discrete, sequential fashion, as one who has learned reading by phonics (the sounding out of words) might expect. The phoneme may have psychological and physiological reality but it does not have any obvious clear-cut physical reality. As Borden & Harris (1984) point out: "the imposition of the idea of 'phonemes' is a linguistic device, useful in constructing an alphabet or in describing a language, but artificial and one step removed from the flow of speech itself" (p.196). Also, because of the nature of the articulatory process, speech sounds are rarely produced in isolation; the acoustic stream is dynamic, constantly adjusting as the vocal tract anticipates its next movement, resulting in what has been termed co-articulation, meaning that the cues for successive speech sounds overlap. Finally, speech is understood across speakers, whose voice pitch and other vocal acoustic characteristics may vary a great deal.

Partly because of these apparently monumental encoding and decoding problems, Liberman, Cooper, Shankweiler, & Studdert-Kennedy, (1967) and others (e.g. Chomsky, 1972) have suggested that "speech is special", i.e. that humans have unique linguistic capacities designed specifically for producing, perceiving, and understanding speech and language. Furthermore, it has been suggested that the perception and production of speech are intimately related in that the perception of speech is directly dependent upon neural "knowledge" of its production. In this

respect, speech perception is considered to be qualitatively different from other forms of auditory communication.

Indeed communication among other animals is thought by some to be restricted to almost reflexive, sign-like "manifestations of different levels of arousal that lack clearly defined external referents" (Seyfarth, Cheney, & Marler, 1980, referring to Smith, 1977). Thus, evidence presented by Kuhl and Miller (1975) demonstrating speech sound discrimination by the chinchilla was quite surprising. Evidence of speech-like perception among other animals followed (e.g. Morse & Snowden, 1975; Waters & Wilson, 1976).

In the meantime, human infant studies have indicated that the ability to produce speech is not necessary in order to perceive speech sounds (Eimas et al., 1971); however experience does seem to affect subsequent discrimination ability (Eilers et al., 1977). The issue of speech perception is thus, far from resolved.

In this section, the following topics will be presented: theories of speech perception, vowel perception, ripple stimuli, animal communication, critical bands, auditory cortex, plasticity of auditory cortex, and the rationale for and hypotheses of the research carried out.

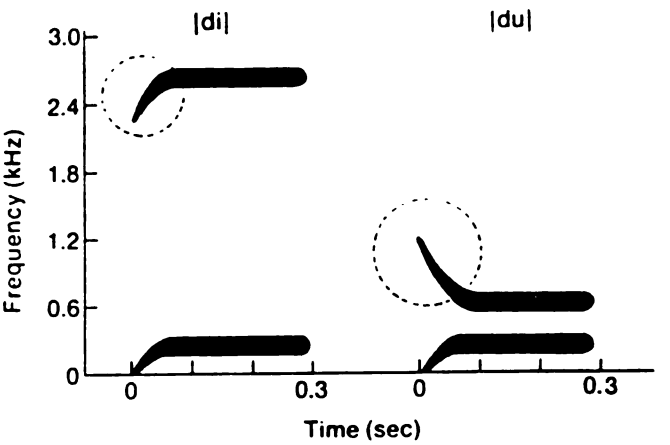


Figure 1. Two formant patterns producing the sounds /di/ and /du/. Second formant transition is encircled. (From Liberman, 1970.)

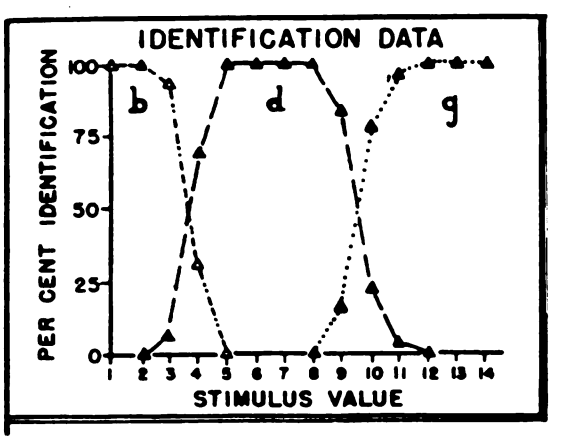


Figure 2. Identification test results showing the percentage of time each stimulus was identified as /b/, /d/, or /g/. (From Liberman, Harris, Hoffman, & Griffith, 1957.)

Theories of Speech Perception

Speech has been described as "meaningful sound strung out in time" (Borden & Harris, 1984, p.2) It is a means of communicating or signaling to others, and it is supposed that selective pressure to communicate information more efficiently led to speech and language, which in turn, aided early humans in their cognitive development. There is currently controversy about the vocal vs. gestural origin of speech. Lieberman (1984) has presented evidence that pre-human primates were physically incapable of producing the sounds that are used in most languages, because of the position of the larynx in the throat in relation to the oral cavity (not deep enough). This interesting research implies that lower primates do not communicate vocally to the extent that humans do, but not because they are not intelligent enough to form the mental symbolic constructs which stand for elements of experience. Rather, vocal interaction is not an efficacious means of communication for these creatures.

Gestural communication by means mainly of the hands may have preceded the evolution of vocal communication. This is especially interesting since it has been shown that gestural communication is capable of conveying a rich amount of linguistic information. Indeed, manual babbling has now been reported to occur in deaf children who had been exposed to sign languages from birth (Petitto & Marentette, 1991). This implies that the building of an abstract linguistic structure of language is not dependent upon the speech modality.

The perception of sensory information has some intriguing facets. In 1957, Liberman, Harris, Hoffman, & Griffith reported categorical perception of acoustic speech parameters. Categorical perception is not unique to speech or to audition. It "occurs when the continuous, variable,

and confusable stimulation that reaches the sense organs is sorted out by the mind into discrete, distinct categories whose members somehow come to resemble one another more than they resemble members of other categories. The best-known example is color categories: Physically speaking, colors differ only in their wavelengths, which gradually get shorter across the spectrum of visible colors. What we see, however, are qualitative changes, from red to orange to yellow to green, and so forth" (Harnad, 1987, p.ix).

In the perception of most auditory stimuli, subjects are able to discriminate more steps than they are able to label (identify); e.g. one knows there is a difference between two tones even though one may not be able to name either. In the processing of three different consonant vowel patterns (/ba/, /da/, /ga/) differing only in the direction and extent of their second formant transition, Liberman et al. observed that although the stimuli were presented along a continuum of 14 equidistant steps, subjects perceived them as three different stimuli representative of the three different phonemic labels. See Figure 2, page 4. Consonant perception is highly categorical; steady state vowel perception is much less categorical (Fry, Abramson, Eimas, & Liberman, 1962; Schouten & van Hessen, 1992). When it was observed that some non-speech stimuli were not perceived in a categorical manner (Liberman, Harris, Kinney, and Lane, 1961), it was speculated that speech processing was performed differently from other types of auditory processing.

The motor theory of speech perception (Liberman, Shankweiler, & Studdert-Kennedy, 1967) attempted to explain this processing by suggesting that in perceiving speech, the listener makes reference to her/his own articulations. The problem of lack of invariance is not solved, however,

since apparently a lot of variability in how speech sounds can be formed is also allowed before recognition is impaired (Harris, 1974). Nevertheless, it seems obvious that speech production and perception are closely related temporally during both the acquisition and everyday use of speech as communication. Pickett (1980) points outs that the motor theory of speech perception was derived from the "functional" school of psychology, which itself arose from evolutionism; these theories maintain that all perception is organized to serve as a basis for behavioral action.

The idea that perception was dependent upon production was questioned when Eimas, Siqueland, Jusczyk, & Vigorito (1971) reported that 1- and 4-month-old infants were able to discriminate speech sounds and appeared to perceive these differences categorically. Using a non-nutritive sucking device, these investigators measured rates of sucking to determine when the infant dishabituated, and thus presumably noticed that a new sound had been presented. Not only were infants able to make fine discriminations, but they were "perceiving speech sounds along the voicing continuum in a manner approximating categorical perception, the manner in which adults perceive these same sounds" (p.306).

Studies showing similar discrimination of other speech parameters by infants have been reported (Eimas, 1975; Jusczyk, 1977). Interestingly infants appear able to distinguish three voicing contrasts even when the parental language has fewer voicing contrasts (Lasky, Syrdal-Lasky, & Klein, 1975; Streeter, 1976), indicating the presence of some innate categorical boundaries that will be modified with exposure to the prevailing acoustical environment. It has also been shown that infants perceive some non-speech sounds categorically (Jusczyk, Rosner, Cutting, Foard, & Smith, 1977); adult reports of this phenomenon have also been

made, in contrast to the earlier finding of Liberman et al., 1961 - e.g. Cutting & Rosner, 1974; Miller, Wier, Pastore, Kelly, & Dooling, 1976; Pisoni, 1977; Pastore, Li, & Layer, 1990). There also appear to be developmental changes in speech discrimination in infants (Eilers & Minifie, 1975; Eilers, Wilson, & Moore, 1977). Infants were unable to discriminate some fricative sounds, e.g. /s/ vs. /z/ and /f/ vs. the /th/ sound in "thin". Some sounds may be more difficult to discriminate (and to produce) than others and "infants may require some listening experience with the sounds to show evidence of discrimination" (Eilers et al., 1977, p.767). Kuhl, Williams, Lacerda, Stevens, & Lindblom (1992) showed that linguistic experience alters phonetic perception in infants by 6 months of age, earlier than had been thought. It was suggested that the infants' language-specific phonetic categories may emerge from a "capacity and proclivity to store in memory biologically important stimuli" (p.608).

Aside from the motor theory, the other main class of speech perception theory invokes the operation of auditory feature detectors. Abbs & Sussman (1971) proposed that "the phonological attributes of human speech are decoded by neurosensory receptive fields operating as 'feature detectors'. These fields are held to be innately structured to detect, and respond to, the various distinguishing physical parameters of the acoustic sound stream" (p.23). This has been likened to a "pandemonium" model, in which the signaling function of the detectors is analogous to yelling, with the feature/s yelling the loudest "calling the shots" (Harnad, 1987, p.11).

Eimas & Corbit (1973) tested this model by repetitively presenting a stimulus so that the sensitivity of a particular detector was selectively reduced. As a consequence, the boundary for that feature appeared to

shift towards the feature signaling the bordering category. This model supports the notion that the perceptual analysis is based upon distinctive features; also, this form of analysis would be analogous to the complex analysis of features apparently occurring in the visual system (Hubel & Wiesel, 1962).

Such an instructive feature detector model would presumably require a top in the hierarchy , i.e. a central unit that signals one specific linguistic percept "created" by the lower feature detectors, a mechanism which would theoretically require a huge amount of "memory space".

What appears more likely is that a pattern of firing signals the percept. The pattern would be spatial or temporal, i.e. the percept would be signaled by which cells are firing, or by the temporal pattern of the firing of one or more cells, or by some combination, so that a rhythm or oscillation of firing is set up among an ensemble of cells which are linked spatially.

Lenneberg (1974) has said that the dynamic nature of speech "casts doubt on the possibility of explaining the neurology of language by simply postulating a system of tuned feature detectors (say single units tuned to distinctive phonological or semantic features)... systems of that sort could merely perform isomorphic transformations of one pattern into another pattern; they could not explain the nature of either understanding or using language" (p.621). In this regard one might profitably distinguish between auditory feature detection and linguistic processing, the former dealing with acoustic representations, and the latter with higher-level (including top-down) cognitive strategies. The auditory feature detector model would also be more useful and plausible if the tuning of the detectors was considered to be "plastic" and the operation of the detectors context-

dependent. Although extant feature detector models are simplistic, detection of speech parameters /features must occur at some level, and that detection must have a neural substrate.

Let us now look at speech, vowels in particular, a little more closely, in an effort to elucidate the acoustic parameters of speech and to identify those features which are important for the formation of linguistic percepts.

Vowel Perception

Expiration of air by the lungs is made audible by the vibration of the vocal folds; the energy is further modified in the supra-laryngeal vocal tract and speech is produced. Speech is composed of a series of sounds, which are called phonemes, and of which there are approximately 40 in the English language. One group of these sounds are periodic, i.e. a periodic, regular modulation of the breath stream is produced by the glottis (the opening between the vocal folds); these sounds are more commonly known as vowels. The other group of sounds are characterized as aperiodic, i.e. they have a turbulent or transient sound source produced at a constriction within the vocal tract; they are more commonly known as consonants. Vowels are always voiced, i.e. the vibration of the vocal folds occurs during production; consonants may be either voiced or unvoiced, i.e. in addition to their main aperiodicity, they may also contain a periodic component. An example of the effect of voicing in consonants is the difference between the unvoiced phoneme [k] and its voiced counterpart [g]. Speech is a sequence of consonants alternating with vowels, i.e. sequences of acoustic contrast. This pattern makes for efficient and rapid production and perception of speech.

Figure 3 shows a spectrum of the glottal pulses, produced by the vibration of the vocal folds of the larynx. The frequency of the vocal fold vibration is called the fundamental frequency (also indicated by F0). The average fundamental frequency for men is about 120 Hz; that for women is 220 Hz, and for children 265 Hz. The difference is due mainly to the difference in size of the vocal folds. Upon the spectrum of the glottal pulses is imposed the spectrum of the acoustic features of the supra-laryngeal tract. The tract can be compared to a tube, open at one end. Such

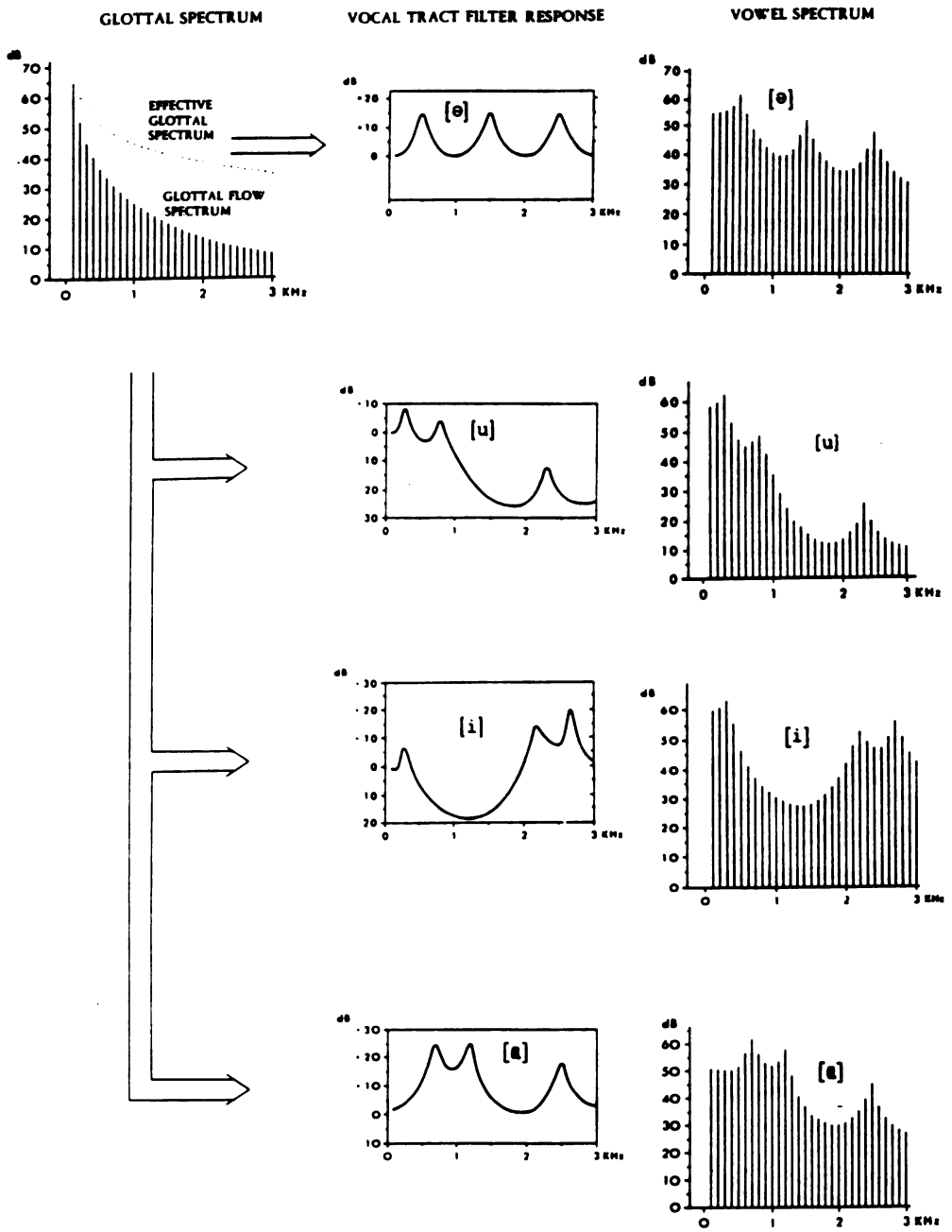


Figure 3. The production of vowels according to the source-filter theory. Onto the glottal spectrum is superimposed the vocal tract filter response to create the vowel spectrum.(From Pickett, 1980.)

a tube, according to its length and its configuration will be seen to have certain natural acoustic resonances - frequencies which are boosted in amplitude compared to the other frequencies of the spectrum. When the configuration/shape of the tube is changed, different frequencies are boosted in amplitude. This is essentially how vowels are produced.

Vowels are characterized by their formants - frequencies which are elevated in amplitude relative to other frequencies and which are produced by the natural resonances of the vocal tract activated during production of the vowel (indicated as F1, F2, F3, etc.). Formant values show considerable variability across speakers, however, and even overlap for some vowels (Peterson & Barney, 1952). The variability across speakers is due to physical differences in vocal tract size but also to a speaker's past experience, mainly his/her particular dialectical background, but the values may vary within context and even across repetitions as well.

In the Peterson & Barney study, 76 speakers, including 33 men, 28 women, and 15 children each recorded two lists of 10 words, each list containing ten monosyllabic words, each beginning with [h] and ending with [d] and differing only in the vowel. Most of the speakers spoke General American. The words were presented to a group of 70 listeners (men and women only) who were asked to identify the words. "The ease with which the observers classified the various vowels varied greatly" (p.177). A device was created at Bell Labs in Connecticut, by R. Potter and colleagues, to produce a visual display, called a spectrogram, that enabled investigators to analyze the frequencies represented in speech across time. The frequency and the amplitude of the formants within the vowels were measured. Figure 4 illustrates the scatter of the first and second formant frequency locations characterizing ten vowels.

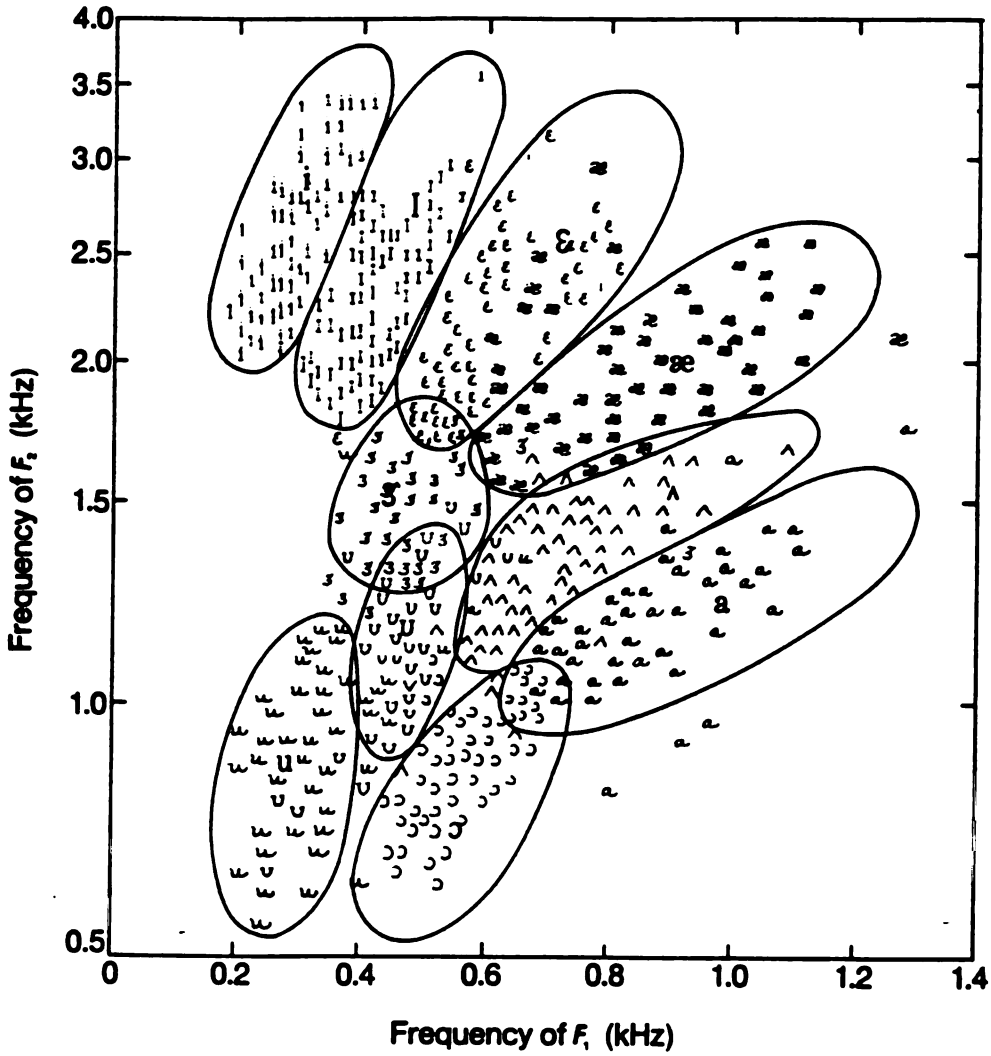


Figure 4. Classification of vowel sounds in terms of their first and second formant frequencies. Vowels cluster but also overlap. (From Peterson & Barney, 1952.)

In light of this variability it is amazing that listeners are so proficient at identifying vowels. The time course of a dialogue between two speakers is usually quite rapid and only rarely marked by mislabeling and misunderstanding of words. If there is no constant code between absolute formant frequency values and a psychological linguistic percept, how are vowels perceived?

It is clear that in spite of speaker variation, listeners are somehow able to adjust for this. Nearey (1989) has looked at two types of speaker normalization that have been proposed: 1) an "intrinsic" method based upon relationships among the steady-state properties of the formants within individual vowels and 2) an "extrinsic" procedure involving the relationships among the formant frequencies of the entire vowel system of a single speaker. Intrinsic normalization implies that information contained within the vowel, e.g. formant ratios or differences, or fundamental frequency-to-formant relationships, provide the necessary vowel identity information. Extrinsic normalization assumes that a frame of reference is established from information distributed across the speaker's vowels, e.g. that there is a "transsyllabic specification of vocal tract size or formant ranges" (p.2091). As pointed out, however, the two factors seem to be related; both provide important information. Nearey concludes that all factors should be considered in a full account of English vowel perception, not any one set of proposed mechanisms to the exclusion of others.

Another problem related to variability within the acoustic signal is termed "target undershoot". Strange (1989) discusses the contextual variation due to coarticulation. That is, when vowels are co-articulated with consonants in continuous speech spoken at normal to rapid rates, the formant target values (those measured for steady-state vowels) are often not reached. Having observed however that vowel discriminability remains high in these contexts, Strange and colleagues hypothesized and confirmed that additional dynamic sources of information are available in coarticulated syllables. One such source appears to be formant transitions into and out of the syllable nucleus, also called formant trajectories. These time-varying acoustic parameters are in effect continuously varying formants patterns; "the acoustic information for vowel identity resides in the changing spectral structure"(p.285).

Miller (1989) has proposed what he calls an "auditory-perceptual theory of phonetic recognition". The theory is descended from the formant-ratio theory of vowel quality, originally proposed by Lloyd (1890), who stated that vowel quality depends on the intervals between the formants, not their absolute values. The theory relies heavily on the logarithmic frequency scale and Miller gives a detailed description of the effectiveness of various scales (e.g. mel, Bark, Koenig) in clustering vowel data (i.e. plotting formant values against each other as Peterson & Barney (1952) had done). The best grouping was observed by using the logs of ratios of formant center frequencies measured either in hertz or mels. Miller further justifies the use of a log frequency scale by citing Weber's law (namely that as stimuli are increased by multiplication, sensations increase by addition) and the octave basis of musical scales. The theory makes use of formant ratios and a "sensory reference" which is based on

the speaker's average fundamental frequency. Sensory and perceptual "paths" describe changes in spectral patterns of the formants during the course of an utterance. "It is proposed that a segmentation mechanism based on the dynamics of these spectral patterns causes perceptual target zones to issue neural symbols or category codes that correspond to the vowel sounds" (p.2129).

Other researchers have suggested that vowels are recognized on the basis of spatial patterns of excitation of the basilar membrane regardless of the specific locations along its dimension. Syrdal and Gopal (1986) transformed vowel formant values to a critical band scale. The critical band scale has been determined from a wide variety of psychoacoustical experiments and is proposed to act as a series of bandpass filters. (It will be discussed further in a later section). Acoustic energy falling within the bandwidth of the critical band is integrated. One critical band represents a relatively constant length of about 1.3 mm along the basilar membrane and about 1300 cochlear neurons. The critical band scale increases with frequency linearly up to about 500 Hz and approximately logarithmically thereafter. It constitutes about one-third of an octave. Zwicker (1961) supported the use of the critical band scale, which divides the human auditory range below 16 kHz into 24 critical band units, or Barks, named after Barkhausen, the creator of the unit of loudness measure. Syrdal and Gopal, using Zwicker and Terhardt's (1980) formula, proposed a quantitative model of vowel recognition in which they converted formant values into Barks and then observed F1-F0, F2-F1, and F3-F2 differences. They concluded that their model "provides excellent vowel clustering while it greatly reduces between-speaker variability" (p.1086).

Chistovich (1985) has reported a difference between stimuli with closely spaced formants and those with widely spaced formants when amplitude relations between the two formants are changed. Closely spaced formants were considered to be those with a distance between them of less than 3 - 3.5 bark (roughly equivalent to one octave). Their perception was closely related to the one-formant stimulus phonetically most similar to the two-formant stimulus, implying a loss of ability to discriminate the two peaks within this critical distance. She suggested therefore that spatial integration of spectral information may be an important factor in the processing of vowels. Chistovich speculates that "there exists a set of detectors tuned to specific spectral configurations...nearly simultaneous occurrence of two peaks in the dynamic spectrum is needed to excite the detectors tuned to two-formant shapes" (p.803). She laments the absence of neurophysiological data concerning the reaction of neurons in AI of auditory cortex to stimuli with complex spectrum shapes.

Ripple Stimuli

Based upon the observation that formant ratios provide important information about vowel identity, a complex stimulus with a frequency spectrum containing various regular peaks, superimposed upon a periodic fundamental frequency, may be a good compromise to simulate vowels, and yet retain manipulable parameters and general known acoustic qualities.

There is some information available regarding psychophysical processing of ripple stimuli. Van Veen and Houtgast (1983, 1985) varied the spectral sharpness (depth of modulation) of ripple stimuli, including synthetic vowels; they observed that the perceptual differences were mainly determined by the lower ripple densities (up to a limit of about 2 ripples/octave).

Hillier (1991) suggests that the auditory system selectively analyzes spatial frequency patterns in a manner analogous to the visual system; "there are neurons in the lateral geniculate nucleus and in the striate cortex that respond maximally to visual gratings of a particular spatial frequency" (p.3). It is suggested that the auditory system in effect performs a Fourier analysis of the spectral envelope of the signal; this has the advantage of shift-invariance, i.e. if a particular signal is present, "a particular set of envelope frequency channels will respond no matter where the signal occurs along the basilar membrane" (p.10). He presented ripple stimuli of various envelope frequencies to human subjects and measured thresholds of detection of a modulated signal vs. a flat spectrum. Thresholds were generally lowest at the envelope frequency of 2 cycles/octave; threshold curves generally manifested a U-shape, with best performance at central envelope frequencies, and worst performance at the extremes. It was also

observed that performance improved as the bandwidth of the signal increased.

Shamma, Vranic, and Wiser (1992) measured the sensitivity of human subjects to changes in the symmetry and bandwidth of spectral peaks. Figure 5 illustrates changes in symmetry (b) and in bandwidth (c) as compared to the standard stimulus (a). They found that detection of a change in peak symmetry was independent of the peak shape of the standard, and that detection of a change in peak bandwidth was monotonically dependent on the standard's bandwidth. These studies were extended, varying peak amplitude, spectral density, and pitch (Shamma & Vranic, 1993). Thresholds for both symmetry and bandwidth changes increased monotonically with increasing peak amplitudes and increasing spectral densities. There was a larger pitch cue contribution for the detection of symmetry than bandwidth peak changes.

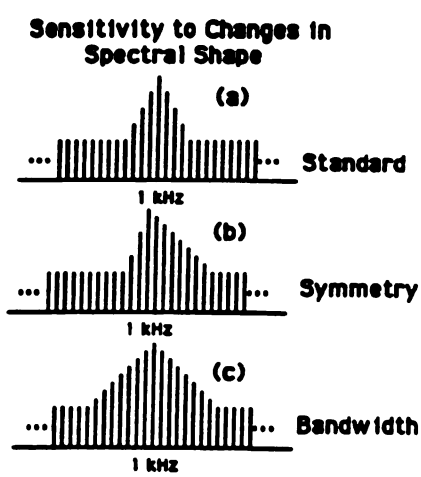


Figure 5. Complex waveform (a) consisting of an equal amplitude base and a peak added to it. Peak takes different symmetries (b) and bandwidths (c).

In order to obtain the neurophysiological data necessary to directly observe neural responses to ripple stimuli, it will be necessary to use an animal subject. Let us therefore turn to the issue of animal vocalization and perception, to determine whether information about speech perception in humans can be expected to be gleaned from such an approach.

Animal Communication

Although the 'raison d'etre' of human speech and language would seem to be communication, there has been considerable discussion about whether communication among humans is analogous to what is thought to be communication among animals. The discussion has revolved around defining what is meant by "communication" vs. "language" (see for example Tartter, 1986, chapter 8; and Lieberman, 1984). Attempts have been made to train animals (mainly apes, gorillas, and chimpanzees) to use human language. These studies were unable however, to differentiate between cognitive abilities, that is, abilities to form associations between concepts/objects and an ability to form a truly linguistic symbolic representation of an object/idea. (See the following references: Gardner & Gardner, 1969; Patterson, 1978 a, b; Premack, 1971; Rumbaugh, Gill, & von Glaserfield, 1973; and Terrace, Pettito, Sanders, & Bever, 1979.)

Field data for vervet monkeys in their natural environment in Kenya were presented by Seyfarth, Cheney, & Marler in 1980. The animals gave different alarm calls to different classes of predators and reacted to play-backs in different ways, indicating that the calls carried distinctive semantic meanings. Cheney & Seyfarth (1988) further showed that monkeys who had learned to ignore an unreliable signaler (a monkey "crying wolf"?) also ignored an acoustically different but same referent call from the same individual. Such transfer did not occur, however, if the identity of the signaler or the referent of the call changed. The authors suggest that vervet monkeys, "like humans, process information at a semantic, and not just an acoustic level" (p.477).

Discrimination of specific speech features was reported by Kuhl and Miller in 1975, in the chinchilla (a mammal with auditory capabilities

fairly similar to those of humans - see Miller, 1970). Using a shock avoidance procedure, they trained animals to respond differently to a variety of spoken /t/ and /d/ consonant-vowel syllables. Once trained, the animals generalized to other speakers, other vowel contexts, and to computer-synthesized /ta/ and /da/ syllables.

In another series of investigations, Kuhl & Padden (1982, 1983) tested monkeys, using a positive reinforcement technique, wherein the monkey was trained on a same-different task, i.e. the monkey was rewarded for one behavior when stimuli were the same, and for another when the stimuli were different. Thus experiments could be conducted which were similar to work done with humans, and animals could be tested on a variety of speech stimuli. Voice-onset-time (VOT) and place-of-articulation discrimination were investigated. In both cases, discriminability was poor when the stimulus pairs were from the same phonetic category and good when they were from different phonetic categories.

Other investigators have reported discrimination of speech parameters in animals. For example, Morse & Snowdon (1975) reported categorical discrimination of speech sounds by rhesus monkeys using the cardiac component of the orienting response. The monkeys significantly differentiated stimuli across phonetic boundaries, but also discriminated within phonetic categories, although to a lesser degree. Sinnott, Beecher, Moody, & Stebbins (1976) concluded that monkeys and humans had similar sensory capacities; however, while humans showed a longer latency of response when both stimuli were within the same phonemic category, monkeys did not. They argued that this was support for unique speech processing capacities in humans. Sinnott & Adams (1987) measured

difference limens to various VOT stimuli in monkeys and humans and observed that again, monkeys were much less sensitive to speech cues than were humans. Waters & Wilson (1976) reported perceptual boundaries in rhesus monkeys that were close to the human boundaries between voiced and voiceless consonants.

Baru (1975) discussed the parameters involved in discrimination of vowels in the dog. Dogs were able to discriminate /a/ from /i/ regardless of fundamental frequency (male voice vs. female voice), and discrimination of vowels appeared to rely on the same cues as those for humans (mainly the first two formant frequencies of the signal). Following either unilateral or bilateral ablation of primary auditory cortex, discrimination was not impaired except when signal duration was 30 msec (as opposed to 75, 150, or 300 msec).

Dewson (1964) reported vowel discrimination in the cat. Cats were trained to discriminate between /u/ and /i/ but following bilateral ablation of the ventral insular-temporal cortex (but not all of primary auditory cortex) did not retain or were unable to relearn the discrimination. The animals could still discriminate frequency and intensity differences.

Cat vocalizations have been analyzed and appear to contain spectral peaks of energy similar to formants in human vowels (Brown, Buchwald, Johnson, & Mikolich, 1978). The average fundamental frequency for cats is within the range of 400-600 Hz (Shipley, Carterette, & Buchwald, 1990); it was also observed that the frequency with largest energy is usually a resonance of the vocal tract. The authors state: "the presence of formants in cat vocalizations and the magnitude of energy in these frequencies suggests that in processing conspecific vocalizations the auditory system of

the cat needs to perform extensive identification and tracking of formant patterns" (p.15).

Thus, animals do seem to have fairly sophisticated auditory processing abilities. It seems plausible that the temporal cortex in cats codes complex stimuli. Before turning to cortical function however, let us review the concept of critical bands, especially as it pertains to perception in cats.

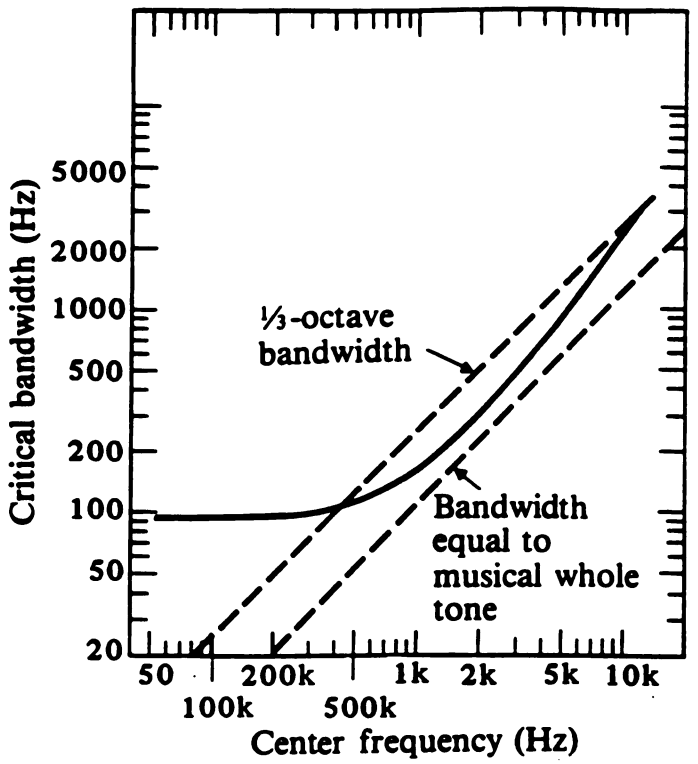
Critical Bands

The concept of the critical band has been proposed in order to explain several puzzling psychoacoustic results. Scharf (1970) says: "as a purely empirical phenomenon, the critical band is that bandwidth at which subjective responses rather abruptly change" (p.159). It refers to a frequency interval bounding a stimulus; it acts as a filtering process assumed to take place within the auditory system. "At the critical bandwidth, listeners' responses to complex sounds change. In seven different kinds of experiments - on loudness summation, narrow-band masking, two-tone masking, threshold, phase sensitivity, musical consonance, and harmonic discrimination - listeners react one way when the stimuli are wider than the critical band and another way when the stimuli are narrower" (Scharf, 1970, p.196).

Figure 6 (taken from Rossing, 1982, p.86) illustrates measurements of the critical bandwidth as a function of the center frequency of the stimulus used. As can be seen, critical bandwidths are approximately one-third of an octave. Recently, Glasberg and Moore (1990) reanalyzed several earlier sets of data, and also analyzed some current data obtained for very low and high center frequencies, using a modified fitting procedure, and leading to a revised formula describing the variation of critical bandwidths as a function of frequency. This function is shown in Figure 7. The computation of critical bandwidths using this formula (ERB (equivalent rectangular bandwidth) = 24.7 (4.37F + 1), where F is frequency in kHz) results in the following values:

Center frequency (Hz)	Critical bandwidth (Hz)
100	36
200	46
500	79
1000	133
2000	241
5000	564
10000	1104

Comparing the above values to those in Figure 6, we see that the more recently calculated critical bandwidths are generally smaller, especially those for center frequencies below 500 Hz.



Center frequency (Hz)	Critical bandwidth (Hz)
100	90
200	90
500	110
1,000	150
2,000	280
5,000	700
10,000	1,200

Figure 6. Critical bandwidth as a function of the critical band center frequency (from Rossing, 1982).

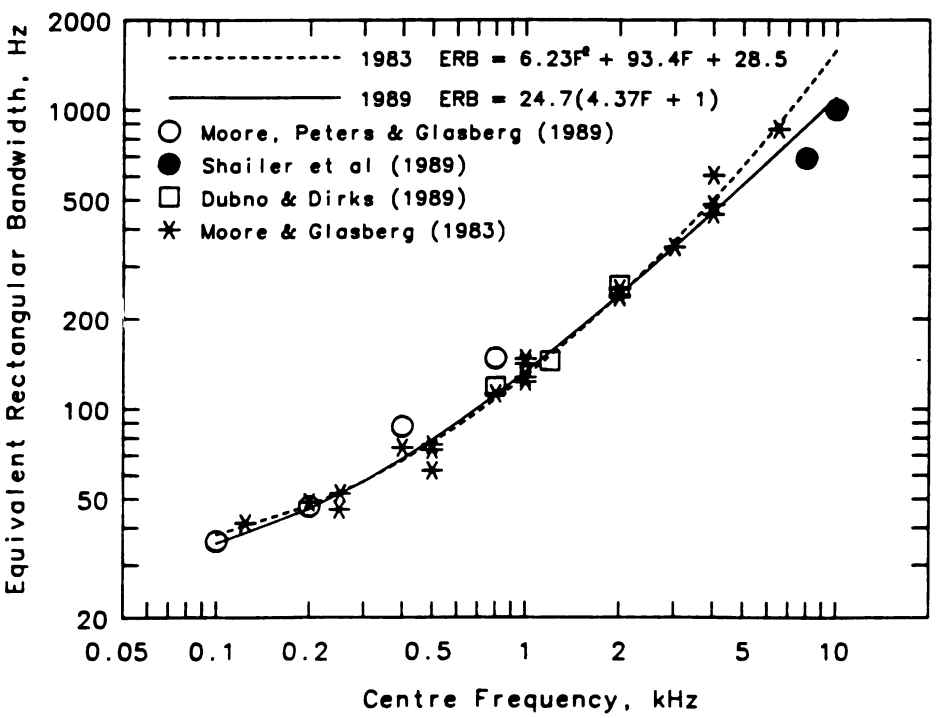


Figure 7. Summary of data relating the "equivalent rectangular bandwidth" of the auditory filter to center frequency (from Glasberg & Moore, 1990).

In Figure 8, (from Rossing, 1982, p.106), the critical band is shown in relation to the just-noticeable difference (jnd) in frequency; the jnd at each frequency is almost a constant percentage of the critical bandwidth. Wier, Jesteadt, & Green (1977) measured frequency discrimination as a function of frequency and sensation level. Their results are shown in Figure 9. They concluded that "discrimination of change in the frequency of pulsed sinusoids is dependent on both signal intensity and frequency; changes in signal intensity have a greater effect at low frequencies than at high" (p.183). The slope of the function for the 5-dB SL data was very similar to the slope for critical band data, while the data for higher sensation levels were fitted by lines with progressively steeper slopes, suggesting that the relationship between the critical band and the jnd is not as constant as had been thought.

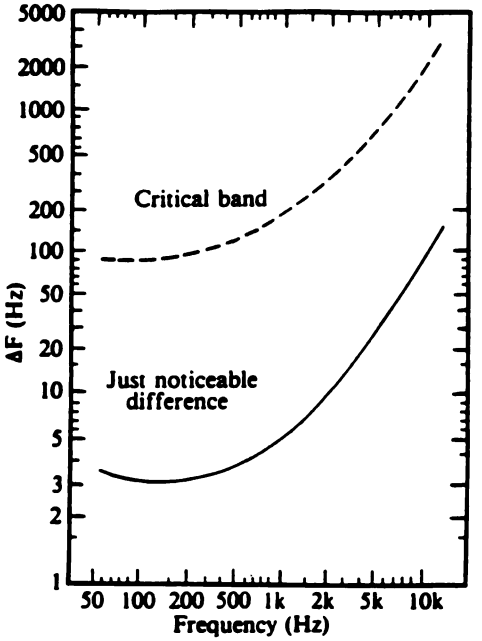


Figure 8. Critical band compared to just-noticeable difference in frequency determined by modulating the frequency of a tone at 4 Hz (from Rossing, 1982, p.106, originally from Zwicker, Flottorp, & Stevens, 1957).

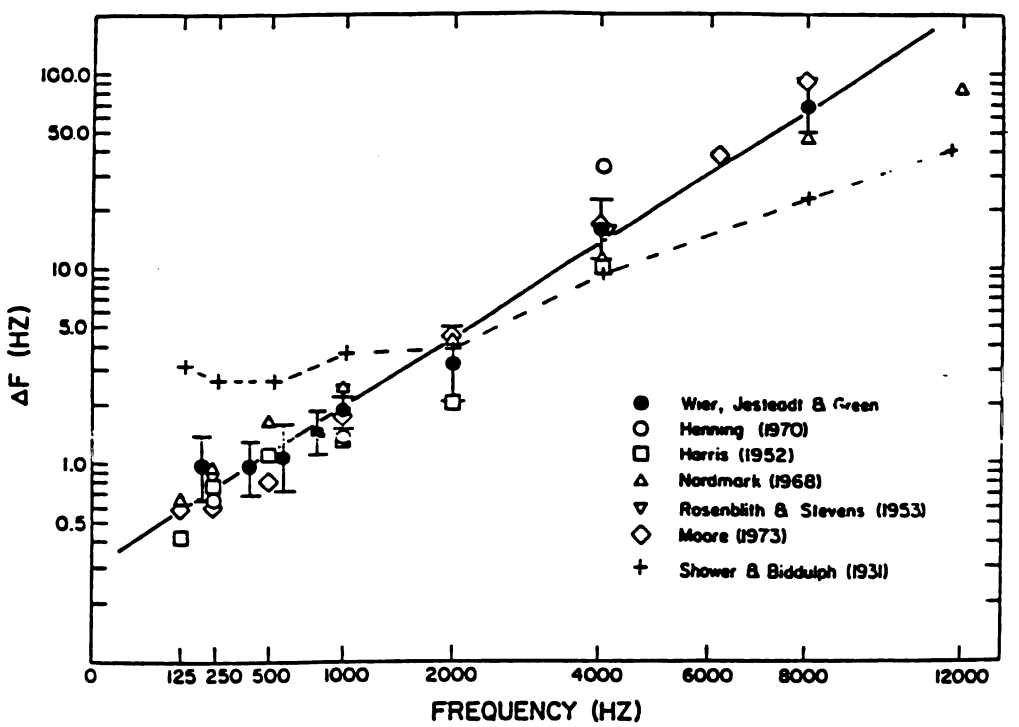


Figure 9. Frequency difference limen as a function of frequency. Stimulus levels 30 to 50 dB SL. (From Wier, Jesteadt, & Green, 1977).

Critical band measurements have been made in cats as well as other animals (Scharf, 1970; Pickles, 1975; Greenwood, 1974). Scharf calculated from reported sensitivity measurements that the critical band at center frequencies above 500 Hz is two to three times wider in the cat than in human; for both human and cat, at 1000 Hz, the critical band is estimated to cover 1 to 2 mm of basilar membrane. Pickles measured critical bands in the cat using a behavioral psychophysical method. He reported a critical band at 1000 Hz of 350 Hz; the human equivalent has been reported to be 150 Hz. At 2000 Hz, the critical band in cat was 700 Hz, as compared to 280 Hz for human. Thus, these data support Scharf's estimate that the critical band in cats is two to three times wider in frequency than in humans.

Auditory Cortex

The auditory cortex in the human and in many nonhuman primates is located primarily on the superior temporal gyrus and is buried for the most part within the Sylvian fissure (Lass, 1988). In the cat, the auditory cortex represents a large area mostly on the lateral surface of the brain. The following descriptions will refer to the cat.

On the basis of constant cellular characteristics seen with Nissl stain, Rose (1949) defined primary auditory cortex (termed AI), secondary cortex (AII), and an auditory area on the posterior ectosylvian gyrus (Ep). Primary auditory cortex was described as cytoarchitecturally similar to other primary sensory cortex, with six layers and a high density of pyramidal and granule cells in layers II, III and IV. The ventral division of the medial geniculate body projects to primary auditory cortex, the afferent fibers ending mainly in layer IV (Winer, Diamond, & Raczkowski, 1977).

Using surface evoked potential recordings, Woolsey and Walzl (1942) first reported a tonotopic organization in auditory cortex. In 1975, Merzenich, Knight, & Roth refined this description employing microelectrode mapping techniques. They reported that "any given frequency band (or sector of the cochlear partition) is represented across a belt of cortex of nearly constant width that runs on a nearly straight axis across AI "(p.247) - see figure 8, page 30. Also, best frequency was constant within vertical penetrations into AI in the active middle and deep cortical layers, i.e. the cells appear to be organized in columns. Further, "there is an orderly representation of the cochlea within the field rostral to AI, with a reversal in best frequencies across its border with AI". Finally,

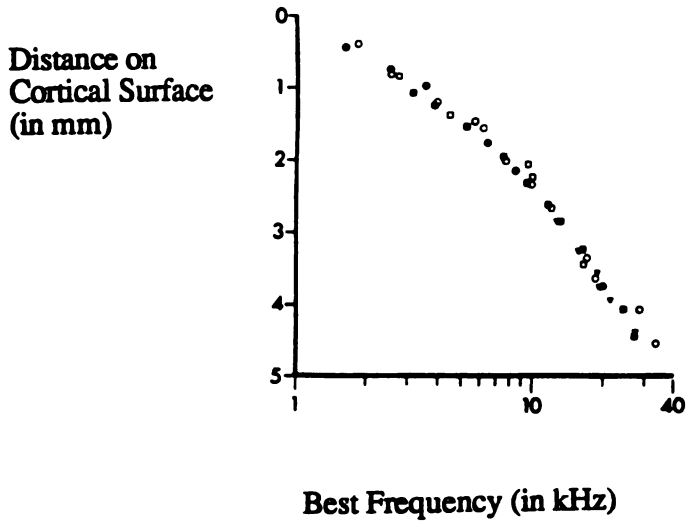


Figure 10. An example of a cortical map showing best frequency as a function of distance across the cortical surface (from Merzenich et al.,1975).

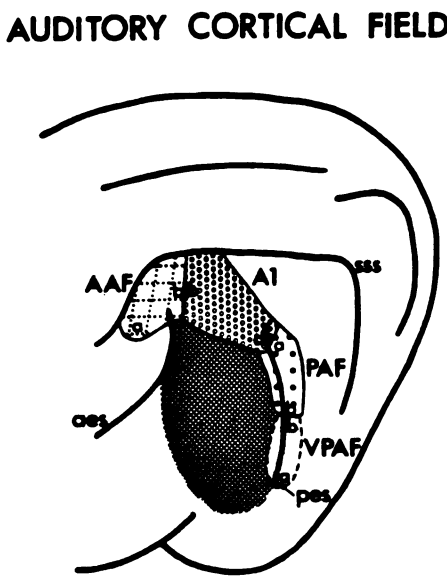


Figure 11. Schematic representation of the basic organization of auditory cortical fields in the cat. AI - primary auditory field; AAF - anterior auditory field; PAF - posterior auditory field; VPAF - ventroposterior auditory field. The region of representation of the cochlear apex (a) and cochlear base (b) are indicated for all four of these cochleotopically organized fields. (from Merzenich et al.,1979).

"physiological definitions of AI boundaries are consistent with their cytoarchitectonic definition" (ibid).

Further mapping studies (Knight, 1977; Merzenich, Roth, Knight & Colwell, 1977; and Reale & Imig, 1977) have shown that there are at least three other cochleotopically organized cortical fields in cats, besides AI, namely, the anterior auditory field (AAF), which borders AI rostrally, the posterior auditory field (PAF), and the ventro-posterior field (VPAF) - see figure 9, p.30, from Merzenich, Andersen, & Middlebrooks (1979).

Another organizing feature of the auditory cortex is the existence of binaural bands roughly orthogonal to the iso-frequency contours (Imig & Adrian, 1978; Imig & Brugge, 1978; Middlebrooks, Dykes, & Merzenich, 1980). In alternating bands, "neurons are predominantly 'excitatory-inhibitory' (driven by contralateral stimulation with driven responses inhibited by ipsilateral stimulation); or, in the second set of bands, 'excitatory-excitatory' (driven by both contralateral and ipsilateral stimulation; or driven by one or the other ear, with the response strongly facilitated by stimulation of the other)" (Merzenich et al., 1979, p.65). It has also been shown that discrete frequency-band-specific lesions in unilateral primary auditory cortex produced profound deficits of sound localization ability in the contralateral hemifield (Jenkins & Merzenich, 1984).

There are however, some features of signals other than frequency and laterality which appear to be represented in a systematic way within auditory cortex. For example, sharpness of tuning is distributed along the iso-frequency domain of AI. Schreiner & Cynader (1984) have demonstrated low Q-10 dB values in area AII. Moving dorsally, the Q-10 dB values increased gradually, reaching a maximum about 2 mm dorsal to

the AI/AII border. From the central portion of AI toward the supra-Sylvian sulcus, the Q-10 dB values gradually decreased, although they remained higher there than in AII. This gradient in sharpness of tuning along the dorso-ventral extent of AI was confirmed by Schreiner & Mendelson (1990), the most dorsal and most ventral regions of AI being the most broadly tuned, with sharpest tuning in the central area. Amplitude modulated signals also appear to be represented somewhat systematically within the anterior auditory cortical area (Schreiner & Urbas, 1986, 1988). Also, non-monotonicity of rate/level functions, i.e. the degree of change of firing rate as a function of level increase, has been investigated, and it appears that primary auditory cortex contains a topographic representation of intensity information in the iso-frequency domain (Schreiner, Mendelson, & Sutter, 1992; Heil & Scheich, 1991). Finally the direction and velocity of logarithmic frequency sweeps are systematically represented in AI (Mendelson, Schreiner, Grasse, & Sutter, 1988; Mendelson & Grasse, 1992). The most dorsal region of AI showed a preference for sweeps from low to high frequencies; moving ventrally, the preferred sweep direction gradually changed to sweeps from high to low frequencies, while most of the ventral portion of AI showed no specific sweep direction selectivity. FM-speed selective responses also displayed a systematic spatial organization, again progressing as a function of the dorso-ventral position within AI.

Suga (1984) has eloquently demonstrated the systematic representation of different types of biosonar information in the auditory cortex of the mustached bat. He shows how "complex acoustic signals are processed by specialized neurons that are tuned to particular information-bearing parameters (IBPs) or combinations of IBPs" (p.315).

Furthermore, "the biologically more important values of an IBP are overrepresented by the large number of IBP filters tuned to them in order to achieve higher resolution" (ibid). Thus the highly refined localization ability of this animal has a corresponding highly refined cortical topographical representation. He has hypothesized (1988) that speech recognition is similarly based upon a spatio-temporal pattern of neural activity occurring in various cortical areas.

Wollberg & Newman (1972; Newman & Wollberg, 1973) looked at the responses of cells in the superior temporal gyrus (STG) of awake squirrel monkeys to a variety of different species-specific vocalizations. Eighty-nine percent of the cells sampled responded to over half of the vocalizations, suggesting that "many neurons in the STG do not select between different classes of vocalizations according to presence or absence of simple acoustic features" (Newman & Wollberg, 1973, p.287), and that vocalizations are encoded in a highly complex way.

Langner, Bonke, & Scheich (1981) explored neuronal discrimination of natural and synthetic vowels in field L of trained mynah birds. Field L is a layered and tonotopically organized primary auditory projection area in the bird. Only a minority of units responded selectively to one particular vowel; many units responded to several vowels. It was therefore speculated that vowel recognition may be based on populations of simultaneously activated units.

Steinschneider et al. (1990) recorded activity in AI of an awake monkey to 3 consonant-vowel syllables, and to the syllables' isolated formants and formant pairs. They reported that response features were related to the tonotopic organization of AI and that formant interactions seem to modulate the response to whole syllables.

Shamma, Fleshman, Wiser, and Versnel (1993) studied responses in primary auditory cortex of the ferret to direction of frequency-modulated tone sweeps and to spectrally shaped noise. The excitatory and inhibitory portions of the response area were described in terms of the asymmetry of the excitation and inhibition around the best frequency and were shown to correlate with the preferred direction of FM sweeps and with spectral shapes. The authors concluded that "cortical responses encode the locally averaged gradient of the acoustic spectrum by their differential distribution along the isofrequency planes. This enhances the representation of such features as the symmetry of spectral peaks and edges and the spectral envelope" (p.367). They suggest a possible analogy in AI with spatial frequency "channels" in the primary visual cortex.

Schreiner, Calhoun, and Keeling (1993) presented ripple stimuli to cortical neurons in the cat. The ripple spectra that were generated had the following characteristics: the carrier consisted of a harmonic series (F0 ranging from 50 to 200 Hz) with a 6 dB/octave decline of the component amplitudes (120 to 255 components); the bandwidth of the stimulus was 3 octaves; the spectral envelope of the signal was represented by a sinusoid on a logarithmically scaled frequency axis, the frequency of the envelope sinusoid was referred to as ripple density (ripples/octave); the modulation depth of the envelope (ripple depth) was linear on a dB scale. The geometrical center of the band-limited signal was always at a maximum of the sinusoidal spectral envelope. The center was positioned at the characteristic frequency (cf) of each cortical neuron and the ripple density or the frequency distance between spectral peaks was systematically varied. The resulting 'ripple transfer function' was reconstructed for different modulation depths and overall intensities. For the majority of neurons, the

ripple transfer function was a bandpass and a 'best' ripple density could be defined. The remaining transfer functions appeared to have a lowpass characteristic, at least for those ripple densities used (0.3 to 8 ripples/octave). Best ripple densities ranged from 0.6 to 4 ripples/octave with a mean between 1 and 2 ripples/octave. Spatial mapping of responses to ripple spectra along the isofrequency domain of AI revealed a systematic shift of the best ripple density in multiple unit responses from central AI to dorsal AI. The shift paralleled the previously described variation of integrated excitatory bandwidth with sharply tuned locations and high best ripple densities near the dorso-ventral center and broader tuning curves with lower best ripple densities toward the dorsal end of AI. Ventral AI showed a less systematic distribution of ripple densities. The hypothesis of a systematic 'spatial frequency' or 'spectral envelope frequency' representation in AI, oriented orthogonal to the frequency axis, was supported.

Thus, one sees at the primary auditory cortical level a rudimentary representation of spectral and temporal features of speech or speech-like signals. What is <u>not</u> seen is selectivity to particular calls, or syllables, or vowels. Rather neural responses seem to be evoked by discrete stimulus components. What is supported is the idea that percepts are created by **patterns** of neural firing across the length and breadth and probably depth of the sensory cortical area.

Plasticity of Auditory Cortex

Merzenich and colleagues have provided evidence that, in adult mammalian somatosensory cortex, the topographic map of the body surface is not static, and that receptive fields at particular cortical sites can change in size and location throughout adult life; the cortical locus at which a given skin surface is represented can shift several hundreds of microns across the cortex (Clark, Allard, Jenkins, & Merzenich, 1988; Jenkins, Merzenich, Ochs, Allard, & Guic-Robles, 1990; Allard, Clark, Jenkins, & Merzenich, 1991). They have suggested that "inputs are selected on the basis of temporal correlation" (Clark et al, 1988, p.444), i.e. mechanisms which are correlated in time may underlie cortical representations and their modification. Such changes also appear to underlie the recovery seen following damage to the brain (Jenkins & Merzenich, 1987). It has also been shown that extended use of a monkey digit, in a discriminative task for which reward is contingent upon behavior, induces reorganization/modification of somatosensory cortical representation (Recanzone, Jenkins, Hradek, & Merzenich, 1992; Recanzone, Merzenich, Jenkins, Grajski, & Dinse, 1992).

In auditory cortex, Robertson and Irvine (1989) have shown that frequency organization in guinea pig reorganizes following partial unilateral deafness. In cats, the cochleotopic representation in primary auditory cortex (AI) was extensively reorganized following neonatal, bilateral high frequency cochlear damage and hearing loss (Harrison, Nagasawa, Smith, Staton, & Mount, 1991). In owl monkeys, Recanzone, Schreiner, and Merzenich (1993) observed plasticity in the frequency representation of primary auditory cortex following discrimination training. An increase in the cortical area of representation of those

frequencies which were behaviorally relevant to the monkeys was observed. The sharpness of tuning was increased, as was the latency of the cortical responses.

Looking at plasticity from another angle, some investigators have classically conditioned alert animals with auditory stimuli, recorded single and/or multiple unit responses at various levels, and reported changes in neuronal response concomitant with conditioning or "learning". Weinberger & Diamond (1988) discussed learning-induced discharge plasticity in the thalamocortical auditory system of the cat. Using a pupillary dilation conditioned response, they observed systematic changes in neuronal responses to the acoustic conditioned stimulus (CS) in the magnocellular, but not the ventral, area of the medial geniculate nucleus of thalamus. About two-thirds of primary auditory cortical cells also changed their responses during conditioning, while "virtually all (21/22; 95%) of the neurons recorded in AII/VE [Ventral Ectosylvian field] were affected by learning" (p.497). Further, frequency-specific effects were observed in AII/VE, i.e. the frequency receptive field (the range of iso-intensity frequencies which evoked a response) changed, either showing increased or decreased evoked activity at the CS frequency. These authors propose that "cortical neurons are concerned not only with the physical characteristics of sound; they may also be said to be concerned with the perceptual aspects of sound. This is reasonable given the fact that perception is well known to be jointly determined by the physical and psychological parameters of stimuli.... In this sense auditory cortex may be viewed not simply as a sensory cortex but as a perceptual cortex. It appears that cortex is far more plastic and dynamic than is generally recognized" (p.505).

Disterhoft & Stuart (1977) observed an increase in firing of inferior colliculus multiple units following conditioning (CS signaling food) in alert rats; the increases were specific to the conditioned stimulus. Kitzes, Farley, & Starr (1978) similarly observed a change in neuronal response following conditioning, in their case in auditory cortex of paralyzed cat, after pairing white noise (the CS) with tail shock. Augmentation and suppression of both spontaneous and evoked activity were found. Since there were such a variety of changes, these authors felt that modulation of cortical activity did not arise from a simple gain control mechanism but rather that the influences were dynamically biasing neuronal activity. They pointed out that "the behavioral context in which acoustic signals occur is an important determinant of the ... activity of a large proportion of auditory cortex neurons" (p.694).

Edeline and Weinberger (1993) recently reported the effects of discrimination training on receptive field plasticity in a classical conditioning paradigm. Responses to the frequency of the positive conditioned stimulus (tone signaling shock) were increased, whereas responses to the negative conditioned stimulus were reduced. There was no difference in effect between an easy and a difficult discrimination task.

Thus, auditory cortical representation of acoustic stimuli is not static, but changes with changing relevance of the stimuli. This dynamic adjustment of cortical response appears to underlie the learning process, as well as recovery following injury.

Rationale and Hypotheses

Thus, from the preceding information and the concepts presented, it can be concluded that the perception of vowels involves the processing of peaks of spectral energy, the frequencies of which are located at various distances or spacings. In order to elucidate the precision of this processing, it was proposed that vowel-like stimuli be presented to humans in a psychophysical testing procedure, varying several parameters of the stimulus, and measuring the discriminability of subjects in each condition.

It was decided to create a vowel-like "ripple" stimulus, which could be easily manipulated to allow the presentation of various parameters of the basic stimulus. The main parameter of interest was to be the peak spacing or ripple density. Holding the ripple density constant, in each test condition, the phase of the envelope of the stimulus was to be shifted so that the locations of the peaks changed. The precision with which the peak shifts could be determined, and the variation of this precision across several ripple densities was expected to shed light upon the importance of small changes in formant location for vowel processing.

The ability of animals, specifically cats, was considered to be relevant to vowel processing since cat vocalizations contain spectral peaks, similar to the formants of vowels. It may also be that processing of this basic feature, namely peak spacing, is an ability common across the mammalian species. Testing of the discrimination capabilities of cats for vowel-like stimuli will show how well this species can process such sounds, and will be of interest as a comparison to human capabilities. An operant conditioning paradigm was therefore employed to obtain a threshold for discriminability of phase shifts of a ripple stimulus in cats.

The main impetus for behavioral training and testing of cats, however, was to determine whether such training forces changes in the auditory cortex of the adult animals. Electrophysiological recording of neurons in response to ripple stimuli may be compared for "untrained" naive cats and trained cats. It is hypothesized that the cortical representation of these complex stimuli will be magnified by learning effects. Changes seen at the cortical level in trained cats would be expected to be correlated with behavioral discrimination measures. Such changes would reflect the effect of learning, and may be expected to be similar to the processes that are in play during human acquisition of language.

Hypotheses

It was hypothesized that discrimination training, testing, and the overlearning of ripple spectra would ensure that the stimuli were behaviorally relevant to cats. It was speculated that, according to critical band data for humans and cats, cats would be able to discriminate between shifts of the peak location (ripple phase shifts), but with thresholds two to three times higher than humans. It was postulated that at the primary auditory cortical level, a vowel-like ripple spectrum would be represented on the basis of the frequency of the dominant peaks in conjunction with a representation of formant ratio or peak spacing as processed by the filter function (sharpness of tuning) of the driven neurons. It was further postulated that, as for plastic changes seen in primate cortex consequent to behavioral training, the cortical representation of behavioral stimuli would be different, reflecting more efficient processing of these relevant stimuli, in the cortex of trained animals as compared to that in untrained animals.

METHODS

The Methods Section is divided into three parts: (1) Human Psychophysical Testing; (2) Animal Behavioral Testing; and (3) Electrophysiological Recording.

HUMAN PSYCHOPHYSICAL TESTING

Stimulus generation and presentation and response tabulation were controlled by computer program, using a Macintosh IIx computer. The LabVIEW application (National Instruments Corporation) that was used automates procedures for data acquisition and instrument control. LabVIEW uses "virtual instruments" (VIs) which are software constructions having the characteristics of an instrument. It has a front panel that is displayed on the computer screen and operated via the keyboard and mouse, a program that performs the virtual instrument's function, and a calling interface for communication with other VIs.

A program was created using a DSP board and output via a 16bit Audio board, that allowed the investigator to present ripple stimuli to the subject in the sound booth, and to record and tabulate on-line the responses of the subject. An input/output board served as the interface between the computer and the subject. A Crown D-75 amplifier amplified the signals.

Subjects and Procedure

Six persons (two men and four women; mean age: 35 years, range 22 to 44 years) with normal or nearly normal hearing as determined by professional audiological assessment, were tested using a modified 2-Alternative Forced Choice (2-AFC) procedure. The procedures were approved by the University of California San Francisco Committee on

Human Research (protocol # H2016-06911-02). The subject read and signed a university-approved consent form. The subject was seated in a sound-attenuating chamber; acoustic foam covered the walls of the booth. The subject was told that upon initiating a trial by a center button press on a panel held on the lap, s/he would hear through headphones (AKG -K240DF) three sounds. The stimulus presentation was monaural, to the left ear. The task of the subject was to indicate which of the three sounds was different from the other two. The first stimulus served as a standard stimulus and was always the same for a single test session. Either the second or the third stimulus was different from the other two; the subject was to indicate the odd stimulus by a button press: left button if the second stimulus was different from the other two, right button if the third stimulus was different. A green light was illuminated if the response was correct. In each test session there were 147 trials (21 different triads, each presented 7 times). A session usually took about ten minutes. Subjects usually completed three or four sessions per sitting.

Stimulus

The formant values for ten English vowels (values takes from Peterson & Barney, 1952) were converted into ratios for the first formant and second formant (F1/F2) and for the second formant and third formant (F2/F3). These values were then converted to a logarithmic scale, giving the number of peaks or ripples per octave, also termed the ripple density. Figure 12 (p.49) shows the ripple densities for F1/F2 and for F2/F3. Values range from 0.1 to 4.5 ripples per octave, with the majority below 3 ripples/octave.

In order to test the discriminability of humans for these various formant spacings, a ripple stimulus was created with the following characteristics: a) the carrier consisted of a harmonic series with a fundamental frequency of approximately 45 Hz (to give closely spaced harmonics) and a 6 dB/octave decline of the component amplitudes; b) the bandwidth of the stimulus was 3 octaves; c) the spectral envelope of the signal was represented by a sinusoid on a logarithmically scaled frequency axis, the frequency of the envelope sinusoid was referred to as ripple density (ripples/octave), and varied from 0.5 to 8 ripples/octave; and d) the modulation depth of the envelope (ripple depth) was linear on a dB scale, and was set at 20 dB.

The standard ripple stimulus (Figure 13, p.50) was centered at 1.38 kHz. This frequency is approximately centered within the range of formant frequencies for English vowels: 270 Hz to 3010 Hz (Peterson & Barney, 1952). The stimulus, being three octaves wide, extended from 340 Hz to 2,672.6 Hz. Varying the density of the ripple varied the spacing between the frequency peaks; for a higher density stimulus, the peaks were closer together. Within each trial, it was deemed necessary that the energy content of each stimulus be kept constant; this was done by varying the phase of the envelope of the stimulus, instead of the ripple density.

The ripple phase of the standard stimulus was 0 degrees; the signal (or odd) ripple stimulus had a phase varying from 1 degree to 180 degrees. Figure 14 (p.51) illustrates a ripple stimulus, with ripple density of 2, shifted 45 degrees. As can be seen, the frequency values of the peaks are shifted. The phase was shifted in both a positive and a negative direction relative to the standard stimulus; for the positive phase shift, frequency

peaks shifted to lower frequencies, for the negative phase shift, frequency peaks shifted to higher frequencies.

The intensity of the stimulus was 60 dB SPL. All three stimuli within each trial were presented at the same level, as it was found to be too difficult for subjects to ignore intensity variations of more than 2-3 dB while discriminating phase shifts. However, the intensities of the stimuli varied randomly by +/- 3 dB around the standard 60 dB SPL stimulus across trials.

Threshold computation

The minimum phase shift of the envelope that could be detected with this modified 2-AFC method was computed in the following way. In each testing session, one-hundred forty-seven trials were presented. In each trial, either the second or third stimulus was different from the other two. The subject was forced to make a choice and was thus either correct or incorrect on each trial. At the beginning of the session, the range of phase shift values was set, with 1 to 180 degrees being the largest range. The ideal was to present a range of values that challenged the subject in order to define the true threshold, but that was not so difficult as to produce extreme frustration for the subject (Green, 1987). Twenty-one values within this range were then randomly presented as the "different stimulus", each being presented 7 times (thus equaling a total of 147 trials). The number of responses correct at each phase shift value were tabulated on-line by computer. On screen could be seen the overall percent of responses correct. In addition, the percent correct on the left side (left button, indicating second stimulus different) and the percent correct on the right side (right button, indicating third stimulus different) were shown, in order to detect response bias for either stimulus. At the end of the test session,

stimulus and response values were stored, and the thresholds computed at a later time. A not discriminated stimulus resulted in a correct response percentage of 50%. Discrimination thresholds were considered as 75% correct, and were calculated by a linear regression formula applied to the psychometric function. The threshold was thus the minimum phase shift, in degrees, detectable at 75% correct. Ripple densities from 0.5 ripples/octave to 8 ripples/octave were tested. Each ripple density was tested at least twice, totalling 294 presentations. A threshold was obtained for each subject at each ripple density value. Further conditions were presented as described below.

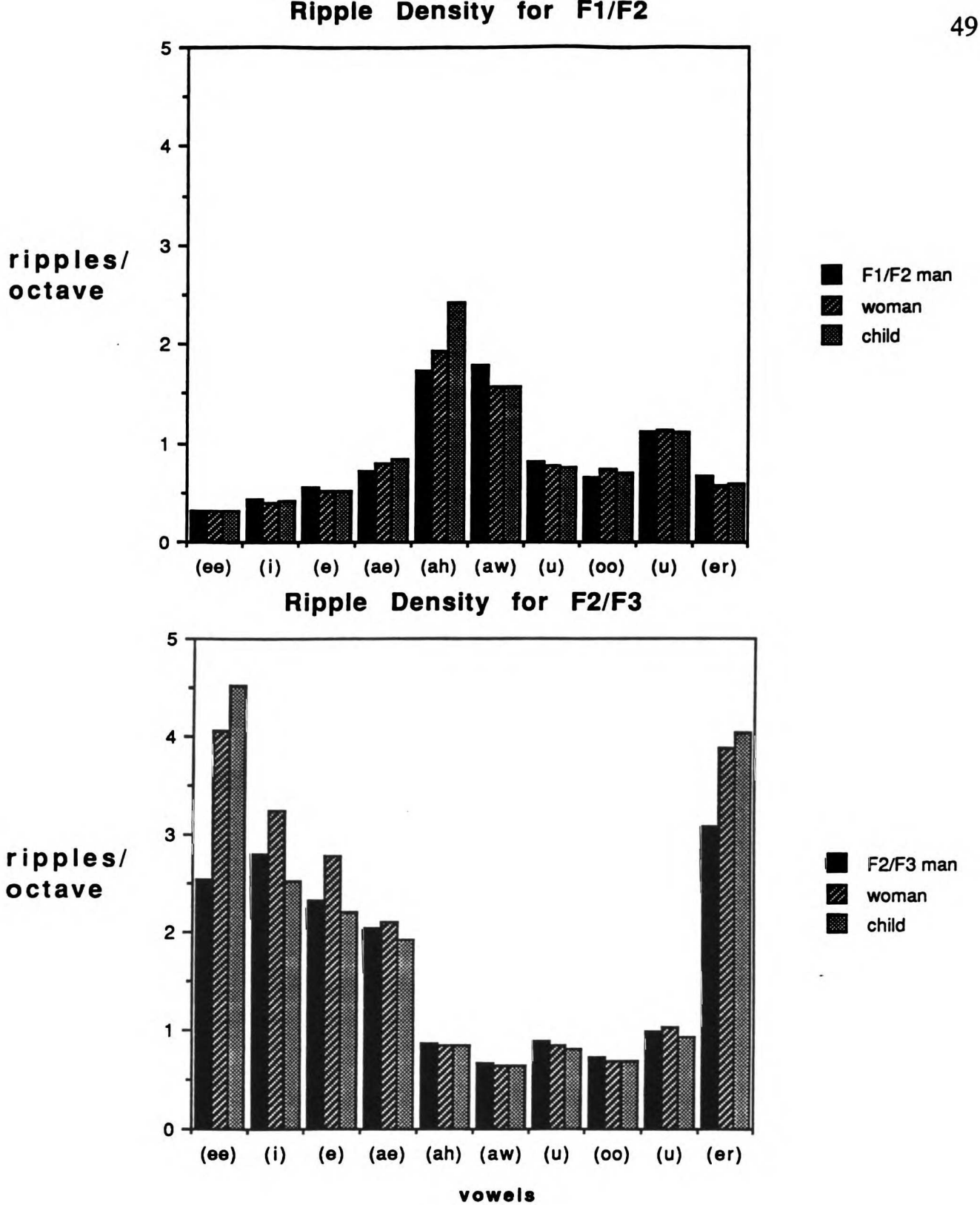


Figure 12. Ripple Density for F1/F2 and for F2/F3. Values range from 0.3 to 4.5 ripples per octave, with the majority below 3 ripples/octave.

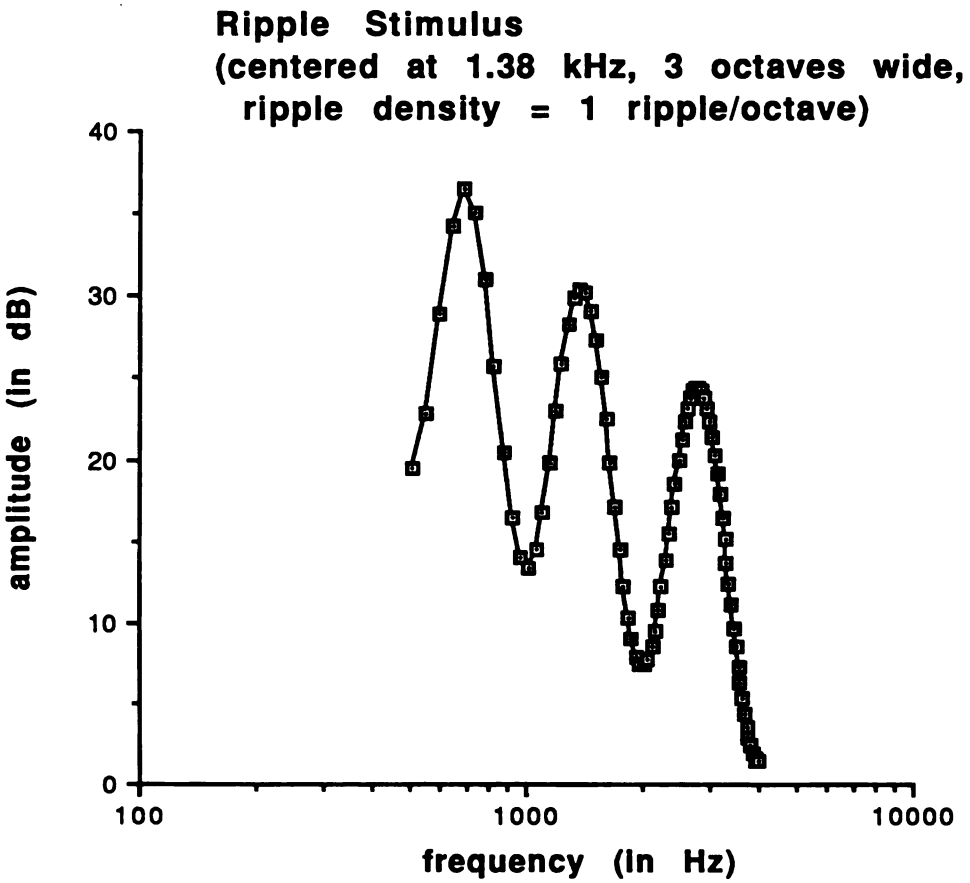


Figure 13. Ripple Stimulus for Human Testing

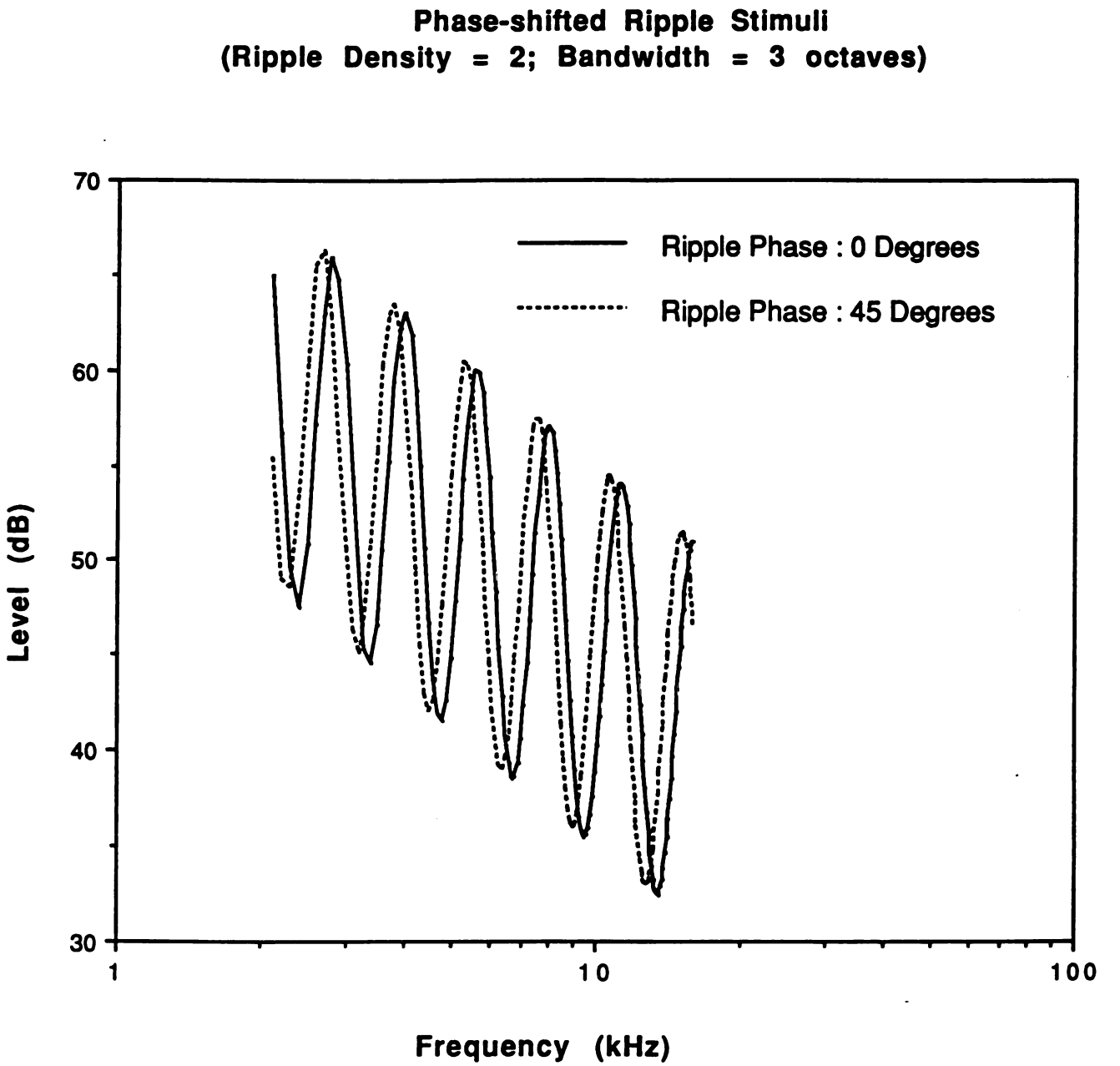


Figure 14. Phase-shifted ripple stimuli, with ripple density of 2, shifted 45 degrees. As can be seen, the frequency values of the peaks are shifted.

a. Ramped vs. Unramped Condition

In order to evaluate whether the subjects used spectral envelope information across the entire width of the spectrum or were biased by changes at the edges of the spectrum, a ripple stimulus was created in which the spectral envelope was ramped at the low- and high-frequency end of the signal. In the ramped portions, the depth of modulation changed gradually over 0.5 octaves from a modulation depth of 0 dB at the high and low-frequency ends of the spectrum to the full value of 20 dB. Since the total signal width was 3 octaves, the fully modulated envelope was realized over 2 octaves.

Two subjects of the originally described six subjects were tested on the ramped ripple stimulus at 7 different ripple densities, ranging from 0.5 ripples/octave to 6 ripples/octave.

b. Depth of Modulation

The ripple stimulus had been created with the possibility of varying several of the its parameters. One such parameter which was varied for one set of tests was depth of modulation of the envelope of the ripple stimulus. It was varied across the values 2.5, 5, 7, 10, 20, and 30 dB, for each of the ripple densities of 0.5, 1, 2, and 4 ripples/octave. Two subjects were tested.

c. Intensity

The intensity of the stimulus had been 60 dB SPL throughout all previous testing. For this set of tests, the overall intensity levels were varied across 30, 40, 50, 60, and 70 dB SPL. Thresholds were obtained at 3 different ripple densities (one, three and five ripples/octave) for three subjects.

ANIMAL BEHAVIORAL TESTING

Three cats were trained to detect shifts in the phase of the stimulus envelope. A two-alternative forced choice paradigm was used. All procedures were approved by the University Committee on Animal Research (protocol # A886-05785-03). The stimulus generation and response tabulation were controlled and presented with the LabVIEW application.

The animals were food deprived to 90- 95% of their pre-fast weight; reward during testing was diluted canned cat food (Hill's Prescription Diet feline c/d). The food was dispensed from a large dose syringe using a stepping motor, which pumped a small amount of food onto a fiberglass plate, about 6 cm in diameter, which the cat licked. The food reward system was modelled after a dispensing system devised by Thompson, Porter, O'Bryan, Heffner & Heffner (1990). The cats were weighed daily and received a supplement of dry cat food after testing in order to obtain their total daily nutritional requirement. The test cage was located within a sound-attenuating chamber. Behavior was monitored continuously with a video camera.

Training procedure

Initially, a hungry cat was placed in the test cage, allowed to explore its new environment, and received food rewards under the investigator's control on a small food reward plate. In the next stage, the cat was trained to press a nose key located at mid-line to indicate an observing ("ready") response, for which it received food. The observing response not only ensured that the cat was attending and could initiate trials at its own pace and volition, but also forced the cat's head position to be constant from trial to trial. A center key press automatically triggered a food reward.

Alternatively, if the cat required "shaping", the investigator could produce a food reward as the cat progressively came closer and closer to pressing the center key on its own. For instance, if the cat oriented its head toward the center key, or moved toward the center key, or lightly pressed the center key, the investigator would deliver a food reward, requiring over trials behavior that came closer and closer to the desired autonomous behavior.

Once the cat was reliably pressing the center key in order to receive the food reward, auditory stimuli were introduced; reward then required a center key press, which was followed by auditory stimulus presentation, to which the cat was required to make a response either to the left or to the right side. Responses were indicated on two spatially separated nose-press keys, one to each side of the middle observing key. (See Figure 15.) To facilitate this stage of learning, the investigator usually remained present in the sound booth and baited the side keys with a small amount of food, by inserting a small narrow plastic tube with food on the end into the cage and onto the response key, which the cat readily followed. Eventually, by indicating the correct response with only a touch to or movement toward the correct key, the investigator could help the cat acquire this behavior. Shaping, by giving reward as the animal approached the correct response, on the control panel outside the booth, was also later used to assist the learning of this stage by the animal.

Initially, stimuli were presented from two small lateralized speakers, which were located at approximately 60 to 80 degrees to the left and right of midline. At the beginning of training the stimulus on the left side was either two or three stimuli, of the same phase (always 0), (designated AAA or AA); the stimulus on the right was maximally different, i.e. 180 degrees

different in phase from the left stimulus (designated BBB or BB). Over time the stimuli were changed to be AA on the left, and AB on the right. The requirement that the animal go left for reward to a left-side stimulus presentation and go right to a right-side stimulus presentation was usually fairly easily learned by the cats and took approximately 2 to 3 weeks of training.

As the animal learned to go left or go right based upon the stimulus presented, the speakers were moved in to the center location gradually until finally the ripple stimuli were presented from a single speaker overhead. Stimulus pair AA (i.e. two identical stimuli) required Response A (left-side key press) in order to receive reward; Stimulus pair AB (i.e. second stimulus different from the first) required Response B (right-side key press) in order to receive reward. The switch from discrimination based upon side of stimulus presentation to discrimination based upon the quality of the stimuli, that is, two stimuli the same vs. a second stimulus different from the first, was quite difficult for the cats to acquire. Some animals were unable to learn this task, and their training was discontinued. The three cats which were able to learn this discrimination varied in length of time required to learn the procedure from a minimum of 3 weeks to a maximum of two months. Performance, i.e. percent correct overall, and percent correct on the left and percent correct on the right, was calculated on-line, as for human subjects. Unlike the human testing however, there was no pre-set number of trials per session; cats continued working until they were sated and did not initiate any further trials. They usually performed anywhere from 100 to 250 trials per day. At the end of the test session, the stimulus phase and response values were stored. Thresholds were computed later, in the same manner as that used for the human

subjects, i.e. by a linear regression formula applied to the psychometric function which calculated the phase shift correctly detected on more than 75% of the trial presentations.

Stimulus

Stimuli were ripple spectra centered at 3.675 kHz, with a ripple density of 1.00 ripple/octave, a fundamental frequency of 69 Hz and a ripple bandwidth of 3 octaves. The stimulus extended from 910 Hz to 7,153.2 Hz. The center frequency of 3.675 kHz was not so low as to preclude the possibility of recording, since frequencies lower than approximately 500 Hz to 1 kHz are usually located within the posterior ectosylvian sulcus and difficult to record from. Also, cat vocalizations appear to contain peaks of spectral energy, the largest being usually a resonance of the vocal tract and located between about 1 and 2 kHz, with spectral peaks extending up to approximately 4 kHz (Shipley, Carterette, & Buchwald, 1990). Thus, the center frequency of 3.675 kHz is a frequency likely to be of interest to cats, and possible to record from on the cortical surface. The phase of the envelope of the stimulus was shifted, originally by 180 degrees, in a positive direction only (a downward shift in the frequency of the center peak); the range of values was narrowed and lowered gradually as the animal's performance improved. The ripple density of the ripple stimuli was constant at one ripple/octave; thus, the distance between the ripple peaks remained the same, however the frequency locations of the peaks changed with the shift in ripple phase. Depth of modulation of the envelope was 30 dB. The intensity of the stimuli varied randomly by +/- 3 dB around approximately 64 dB SPL. Figure 16 illustrates the ripple stimulus used for animal testing. At the end of testing, the cats underwent electrophysiological recording.

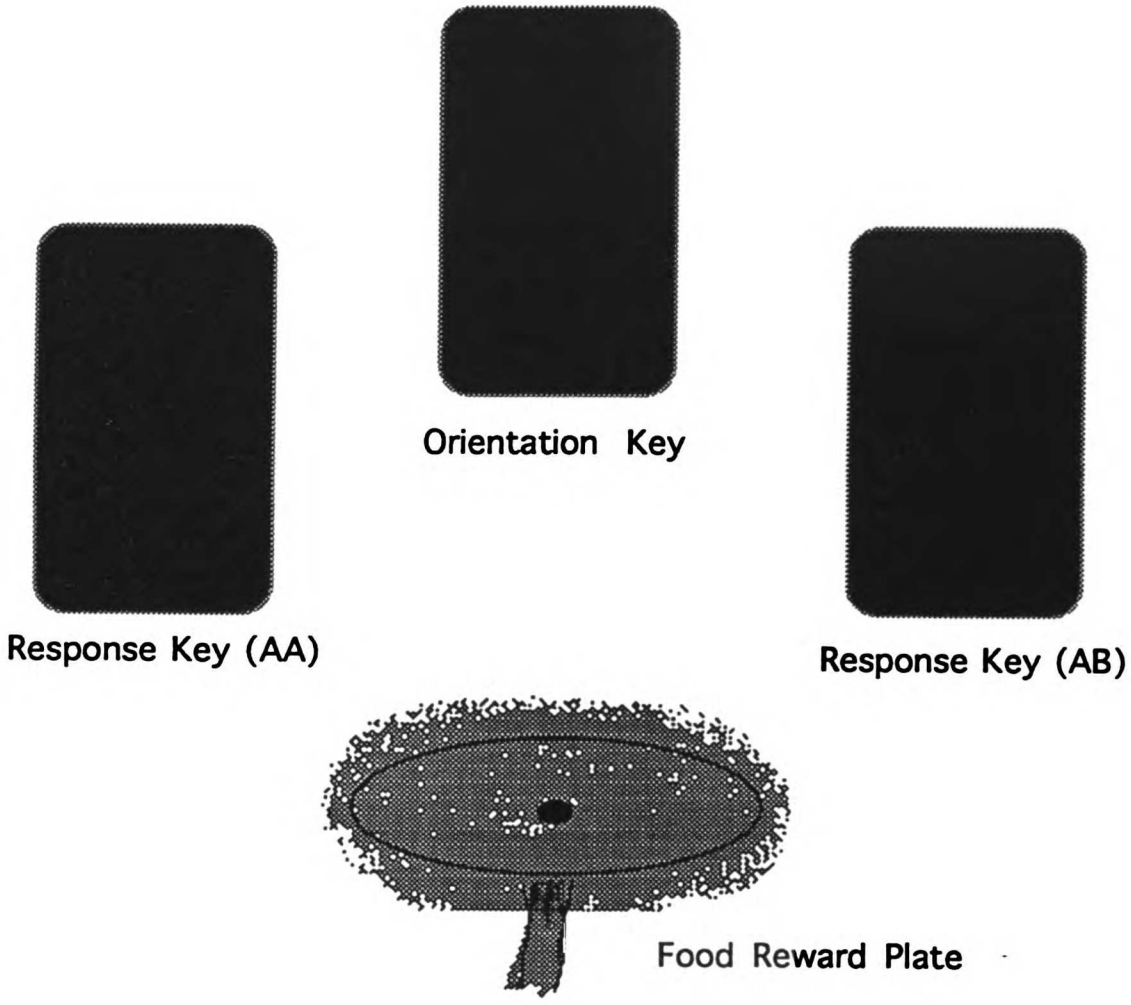


Figure 15. Configuration of Animal Response Keys and Reward Plate. If the two stimuli were the same, the cat pressed the left response key to receive reward; if the second stimulus was different, the cat pressed the right response key to receive reward.

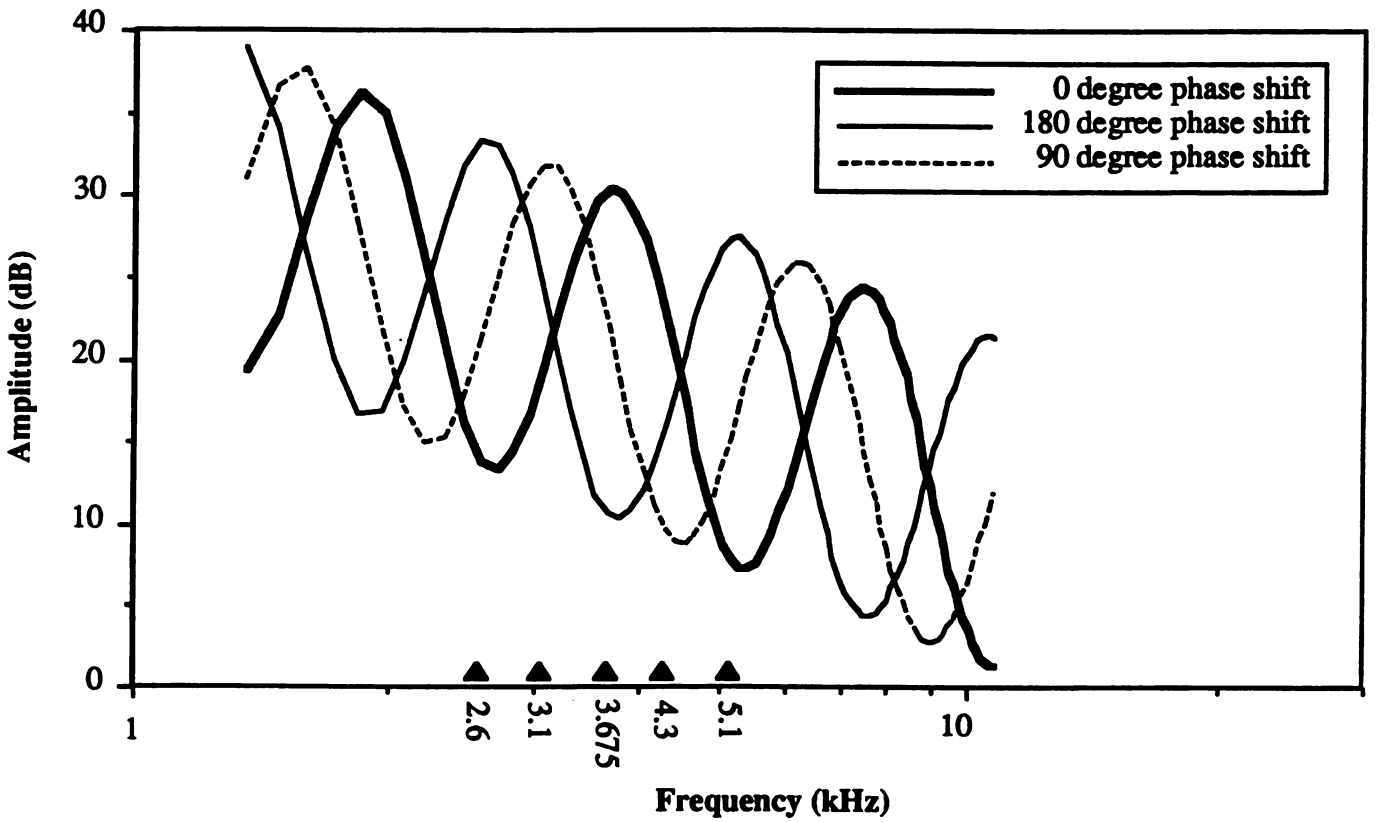


Figure 16. Ripple Stimulus for Animal Testing, showing 0 degree, 180 degree, and 90 degree phase shift. Corresponding frequency locations are shown: 3.675, 2.6, and 3.1 kHz. Frequencies to the right represent phase shifts of negative 90 and negative 180 degrees, at locations of 4.3 and 5.1 kHz.

ELECTROPHYSIOLOGICAL RECORDING

Electrophysiological recording from primary auditory cortex was performed in the three trained cats, as well as in several normal un-trained cats. The cats were injected intra-muscularly with acetylpromazine maleate (.09 mg/Kg) and ketamine hydrochloride (10 mg/kg). When a cat was sufficiently sedated, it was shaved along the inside of the two forelimbs (for IV cannulation), along the throat area (for insertion of tracheal tube), and over the posterior and lateral area of the head (for temporal cortex exposure). Venous cannulation was performed; anesthesia was induced with an initial dose of pentobarbital sodium (30 mg/kg). Animals were maintained at a surgical level of anesthesia with a continuous infusion of pentobarbital sodium (2 mg $\cdot$ Kg^{-1} $\cdot$ h^{-1}) in lactated Ringer solution (infusion volume: 3.5 ml/h) and, if necessary, with supplementary intravenous injections of pentobarbital sodium. The state of anesthesia was monitored and the rate of infusion was adjusted to maintain an a-reflexive state. The cats were also given dexamethasone sodium phosphate (0.14 mg/kg im) to prevent brain edema and atropine sulfate (1 mg im) to reduce salivation. A tracheotomy was performed and a tracheal tube inserted to ease breathing and to reduce breathing noises. The temperature of the animal was monitored with a rectal temperature probe and maintained at 37.5 C with the use of a heating water blanket with feedback control. The EKG was continuously displayed on an oscilloscope and amplified through a loudspeaker.

The cat's head was placed in a standard rigid mouth-bar headholder, an incision made in the scalp down the center from forehead to occiput, the right temporalis muscle retracted, and the right cortical surface (right

side for convenience due to stimulation and recording set-up) exposed by drilling through the skull in the region of the middle ectosylvian gyrus. The dura overlying the middle ectosylvian gyrus was removed, the exposed cortex covered with silicone oil, and a video image of the surface vasculature was obtained with a CCD camera, an image capture board (Data Translation DT2255) and capture software (Image 1.4, NSCA). Electrode penetration sites were marked on the image of the cortical surface in a display program (Canvas, Deneba).

Experiments were conducted in a sound-shielded room (IAC). Auditory stimuli were presented via calibrated headphones (STAX 54) enclosed in small chambers that were connected to sound delivery tubes placed into the acoustic meati (Sokolich, US Patent 4,251,686; 1981). The sound delivery system was calibrated with a sound level meter (Bruel & Kjaer) and a waveform analyser (General Radio 1521-B). Auditory stimuli were digitally generated by a signal processing computer (TMS32010) and converted into analog signal by a 16-bit digital-to-analog converter running at a 60 kHz sampling rate. Additional attenuation was provided by a pair of passive attenuators (Hewlett Packard).

For each recording site responses were firstly recorded to at least 675 different tone bursts. Tone bursts were presented in a pseudorandom sequence of different frequency-level combinations selected from 15 level values and 45 frequency values. From the responses to all stimuli, frequency response areas (FRAs) were reconstructed. Steps between levels were 5 dB, resulting in a sampled dynamic range of 75 dB. The frequency range covered by the 45 frequency steps was geometrically centered around the estimated cf of the recording site and covered 3 to 5 octaves, depending on the estimated width of the FRA. Stimulus frequencies were

equidistant on a logarithmic frequency scale. Following the presentation of pure tones and the generation of the FRA, ripple stimuli with characteristics similar to the stimuli used in animal behavioral testing were created and presented: bandwidth was 3 octaves, fundamental frequency was approximately 69 Hz, ripple densities (10 to 14 values) of from 0.5 to 8 ripples/octave were presented, envelope phase shifts (10 values) of from 0 to 180 degrees were presented.

Primary auditory cortex (AI) was identified using cf information, high frequencies (20 kHz) being represented anteriorly and low frequencies (2 - 4 kHz) posteriorly (Merzenich et al., 1975) and using sulcal/gyral landmarks. Multiple unit recording was performed using parylene coated tungsten electrodes with impedances of 1.0 to 2.0 MOhm at 1 kHz; in a few cases, single unit responses were recorded with glass-coated tungsten electrodes (3-4 MOhm at 1 kHz). The electrode was introduced into the cortex approximately orthogonal to the surface (as viewed through a Zeiss operating microscope); penetrations were parallel to each other. The electrode was advanced through the cortical layers to a depth of 600 - 1000 μm, corresponding to cortical layers III and IV, using a hydraulic microdrive (Kopf) remotely controlled by a stepping motor. Activity of small groups of neurons was amplified, band-pass filtered (1-10 kHz, 12 dB octave), and monitored on an oscilloscope and an audio monitor. The discriminator (BAK DIS-1) level was set to exclude evoked potentials and to accept events that resembled action potentials of an amplitude at least 50% above the background signal. The number of events per presentation and the arrival time of the first event after the onset of the tone bursts were recorded and stored in a computer (DEC 11/73). The recording window had a duration of 50 ms, corresponding to the stimulus duration.

Poststimulus time histograms (PSTHs) were constructed in response to the ripple stimuli. Binwidth was 0.8 ms.

A general response profile of cells was assessed including spontaneous discharge rate, characteristic frequency (cf) (the stimulus frequency with the lowest sound-pressure level necessary to evoke neuronal activity), threshold (the lowest level evoking activity in the FRA), bandwidth approximately 10 dB above threshold (to compute Q-10, which is cf divided by bandwidth 10 dB above threshold), bandwidth 40 dB above threshold (to compute Q-40, which is cf divided by bandwidth 40 dB above threshold), and in some cases binaural interaction class. Following electrophysiological testing animals were sacrificed with an overdose of pentobarbital sodium; bilateral thoracotomy was performed.

RESULTS

The RESULTS Section is divided into three parts: (1) Human Psychophysical Testing; (2) Animal Behavioral Testing; and (3) Electrophysiological Recording.

HUMAN PSYCHOPHYSICAL TESTING

Thresholds were obtained for each subject at various ripple densities. The threshold represents the smallest envelope phase shift in degrees that the subject can detect with 75% or higher accuracy. Figure 17a shows a typical psychometric function for a representative subject. Performance (percent correct) is plotted as a function of degrees shifted negative and positive, in this case for a ripple density of 1 ripple (or peak)/octave. Figure 17b shows the performance of the same subject for a ripple density of 4 ripples/octave. In each test session, each value was presented 7 times; in most cases, each ripple density session was presented at least twice. Values shown represent the average performance over 14 stimulus presentations. The thresholds for a ripple density of 1 ripple/octave were +2.3 and -4.7 degrees. The thresholds for a ripple density of 4 ripples/octave were +13.5 and -22.5 degrees.

Figures 18 and 19 illustrate the performances for each of the six subjects, presenting thresholds (minimum phase shift detectable at 75% correct) in firstly positive, then negative directions, as a function of ripple density. Note that for positive phase shifts, the threshold generally increased with an increase in ripple density to a maximum at about 4 to 5 ripples/octave, then generally decreased. For negative phase shifts, there was greater variability in the shapes of the functions, and thresholds were highest at anywhere from 3 to 8 ripples/octave.

Figure 20 presents thresholds across all tested ripple densities pooled for all subjects: a) for positive phase shifts and b) for negative phase shifts. For positive phase shifts, thresholds were highest at 4.5 and 5 ripples/octave. The lowest threshold was at 1.5 ripples/octave, where variability among subjects was lowest. For the negative phase shifts, the highest threshold was at a ripple density of 4 ripples/octave. The lowest threshold was again at 1.5 ripples/octave, where variability among subjects was also lowest. Tables 1 and 2 present the means and standard deviations for each of the averaged values across all ripple densities, for envelopes shifted in both positive and negative directions. A repeated-measures ANOVA was carried out for both positive threshold and negative threshold data. For the positive data, the overall F was significant at $p=0.03$. Comparison among individual means using the Scheffe test for the various ripple densities showed differences significant at $p<0.05$ for the following comparisons: .5 vs. 4, 4.5, and 5; 1 vs. 4, 4.5, and 5; 1.5 vs. 4, 4.5, and 5; 2 vs. 4.5 and 5; 2.5 vs. 4, 4.5, and 5; 3.5 vs. 4.5, and 5; 4.5 vs. 8; and 5 vs. 8. For the negative threshold data, the overall F was not significant ($p=0.07$), although a similar statistical trend was observed. However, in comparing the peaks for the subjects in the positive vs. the negative phase shift condition (two-group paired t-test), there was a statistically significant ($p=0.01$) difference between the location of the peaks in the two conditions.

Thus, in general, what was observed for both positive and negative envelope phase shifts was an increase in threshold as ripple density increased up to approximately 4 to 5 ripples/octave, then a decrease, at least to the extent of the ripple densities tested here, i.e. 8 ripples/octave.

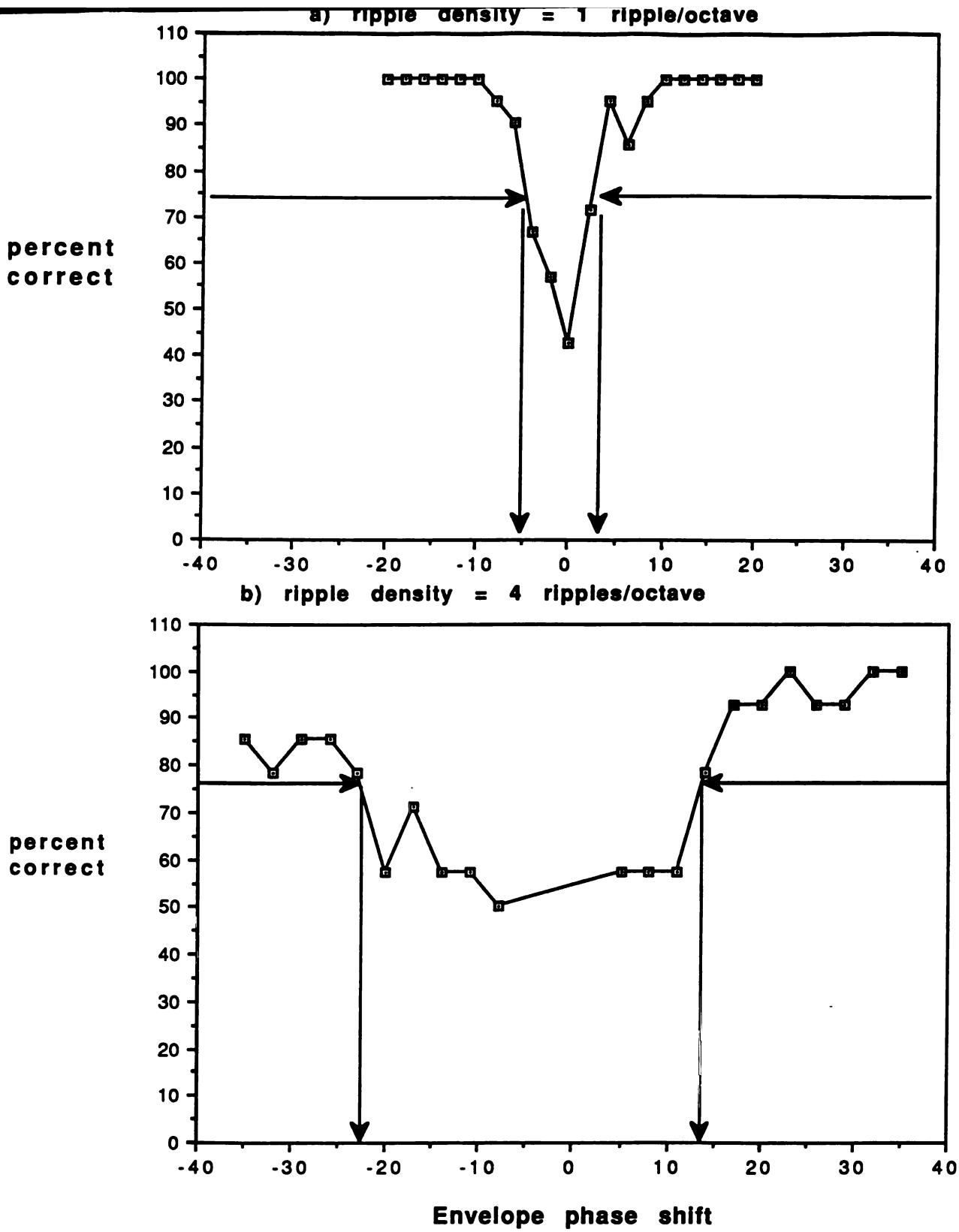


Figure 17. Performance of subject dk on a ripple stimulus of ripple density a) 1 ripple/octave and b) 4 ripples/octave. The thresholds for a ripple density of 1 ripple/octave were +2.3 and -4.7 degrees. The thresholds for a ripple density of 4 ripples/octave were +13.5 and -22.5 degrees.

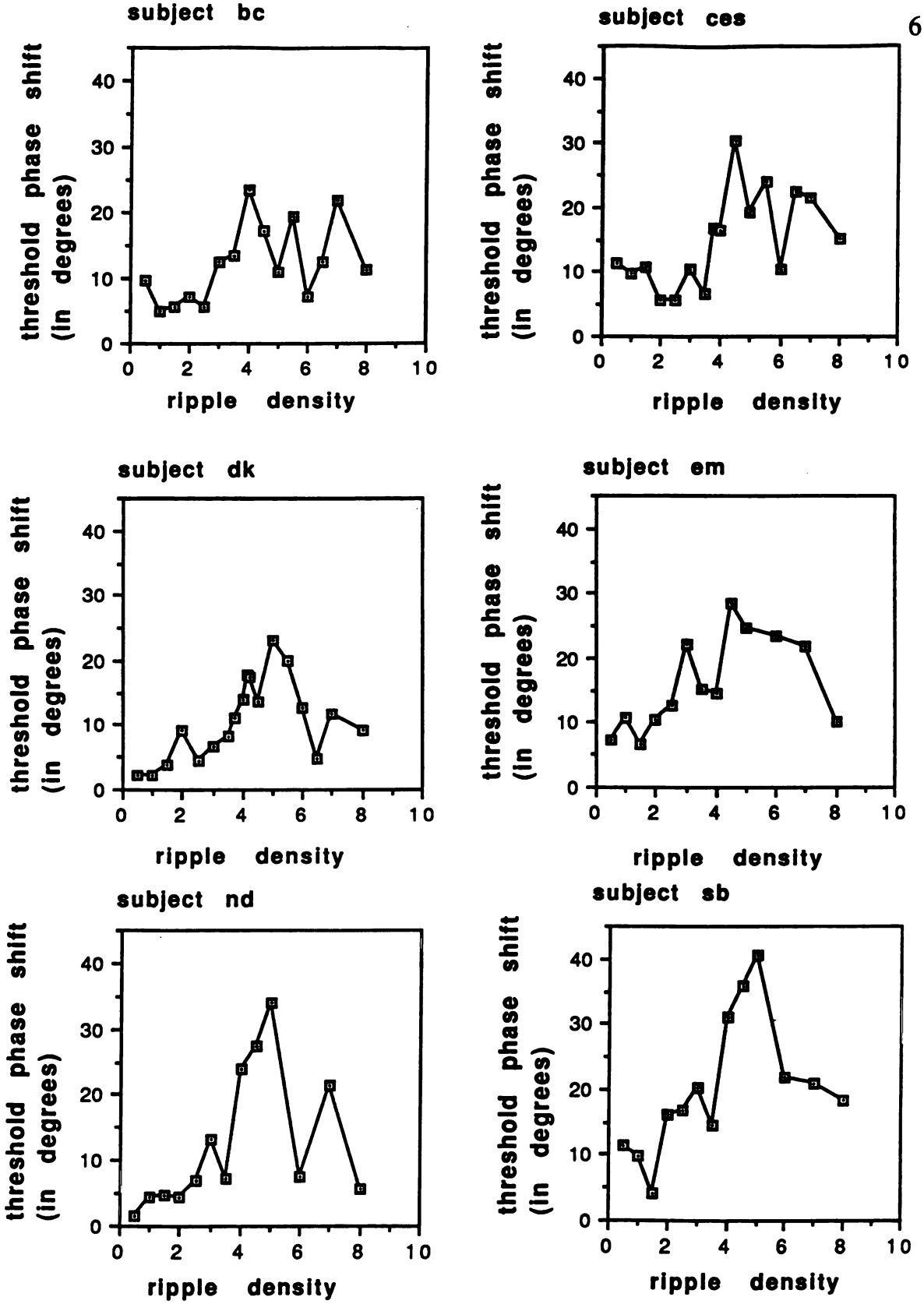


Figure 18. Performance of subjects across ripple densities, for positive envelope phase shifts.

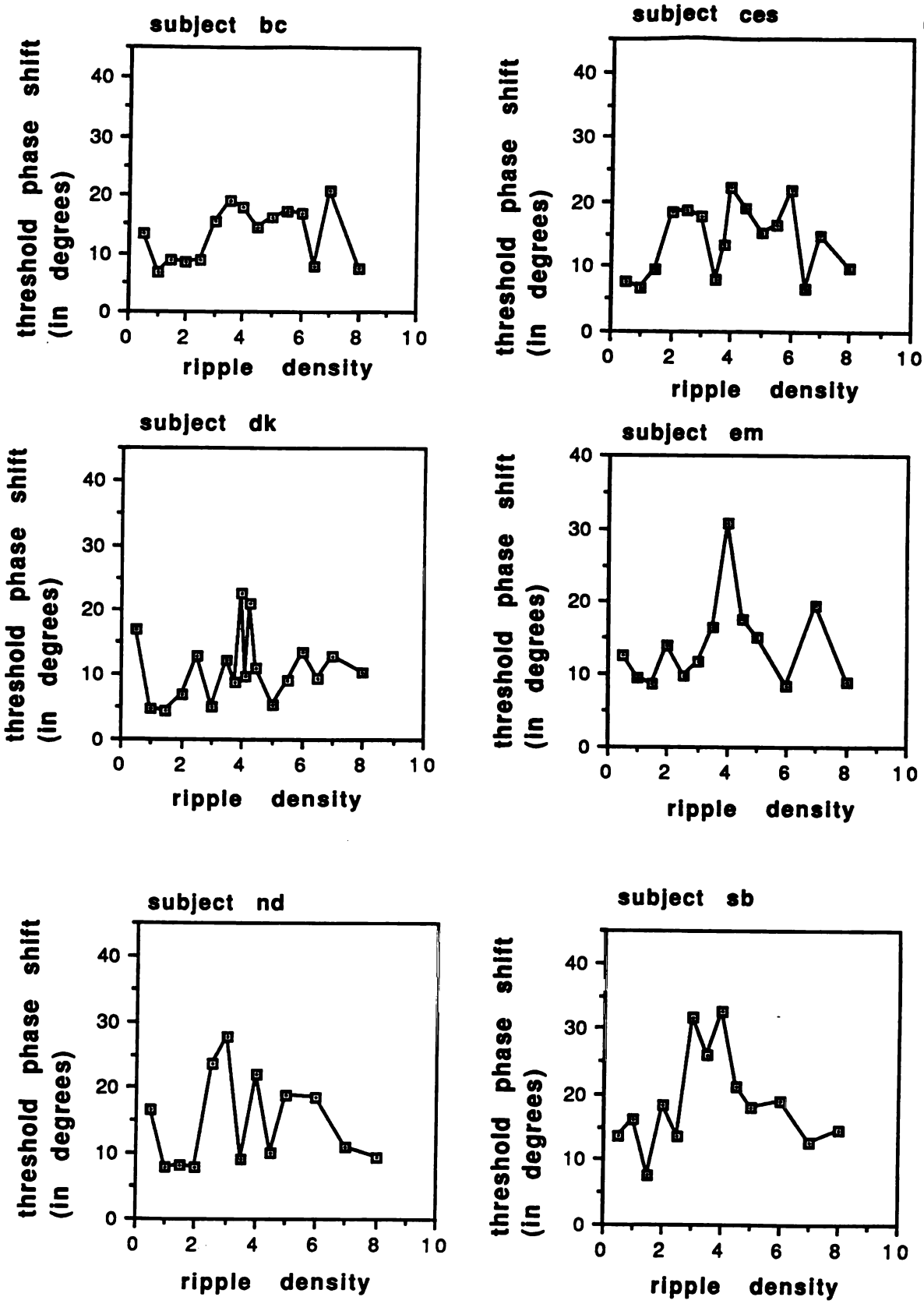


Figure 19. Performance of subjects across ripple densities, for negative envelope phase shifts.

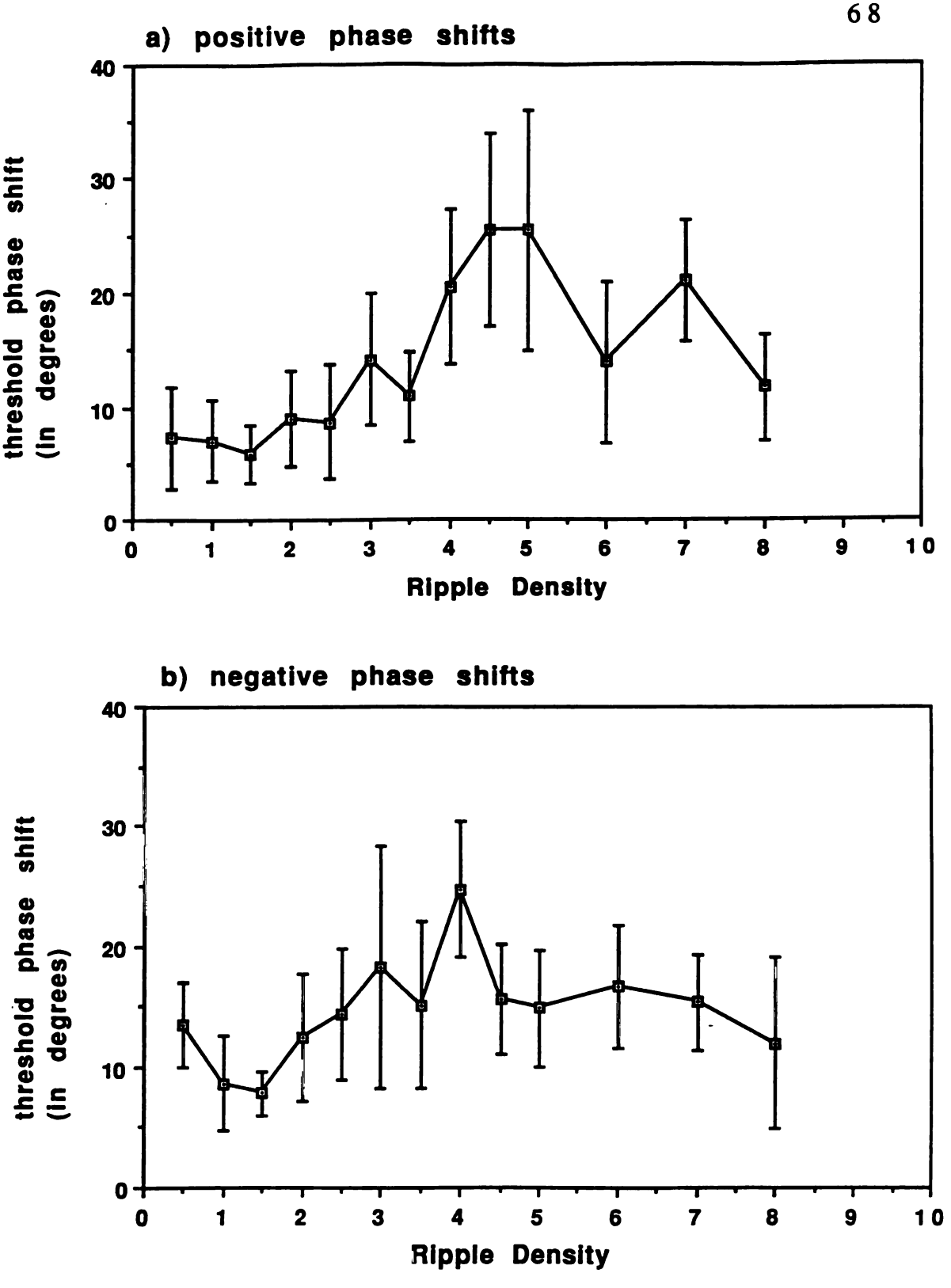


Figure 20. Thresholds pooled for all subjects a) for positive phase shifts and b) for negative phase shifts, showing means and standard deviations at each ripple density.

ripple density	mean threshold shift	standard deviation
0.5	7.3	4.4
1	7	3.6
1.5	5.9	2.5
2	9	4.2
2.5	8.7	5
3	14.2	5.9
3.5	11	3.9
4	20.5	6.9
4.5	25.5	8.4
5	25.5	10.6
6	13.9	7.1
7	21.1	5.3
8	11.7	4.7

Table 1. Mean threshold shift and standard deviation at each ripple density for positive phase shifts (N=6)

ripple density	mean threshold shift	standard deviation
0.5	13.5	3.4
1	8.6	4
1.5	7.8	1.8
2	12.4	5.3
2.5	14.4	5.4
3	18.3	10
3.5	15.1	6.9
4	24.7	5.6
4.5	15.6	4.5
5	14.8	4.8
6	16.6	5
7	15.3	4
8	11.9	7.1

Table 2. Mean threshold shift and standard deviation at each ripple density for negative phase shifts (N=6)

a. Ramped vs. Unramped Condition

In order to evaluate whether the subjects used spectral envelope information across the entire width of the spectrum or were biased by changes at the edges of the spectrum, a ramped ripple stimulus was created. The results of the tests on the ramped ripple stimulus as compared to the original un-ramped stimulus are shown in Figure 21. The positive phase shift thresholds are plotted as a function of different ripple densities for the ramped and un-ramped conditions. For subject (dk), the thresholds were somewhat higher in the ramped condition for all ripple densities; however, the shapes of the functions were very similar, with the highest thresholds at a density of 5 ripples/octave. Comparison of the two conditions (ramped vs. unramped) using a paired t-test revealed that they were different for this subject ($p=0.01$). For subject (bc), the thresholds were very similar in the ramped and un-ramped conditions. Thresholds were higher in the ramped condition for 5 of 8 ripple densities. Again, the shapes of the functions were very similar, this time with the highest thresholds at 4 ripples/octave. For this subject, when the two conditions were compared using a paired t-test, they did not significantly differ ($p=.89$). Thus, for each subject the global pattern of threshold distribution was the same. However, since there were differences between some of the means for one of the subjects, there is an indication that some subjects may use the edge of the spectrum in their discrimination processing.

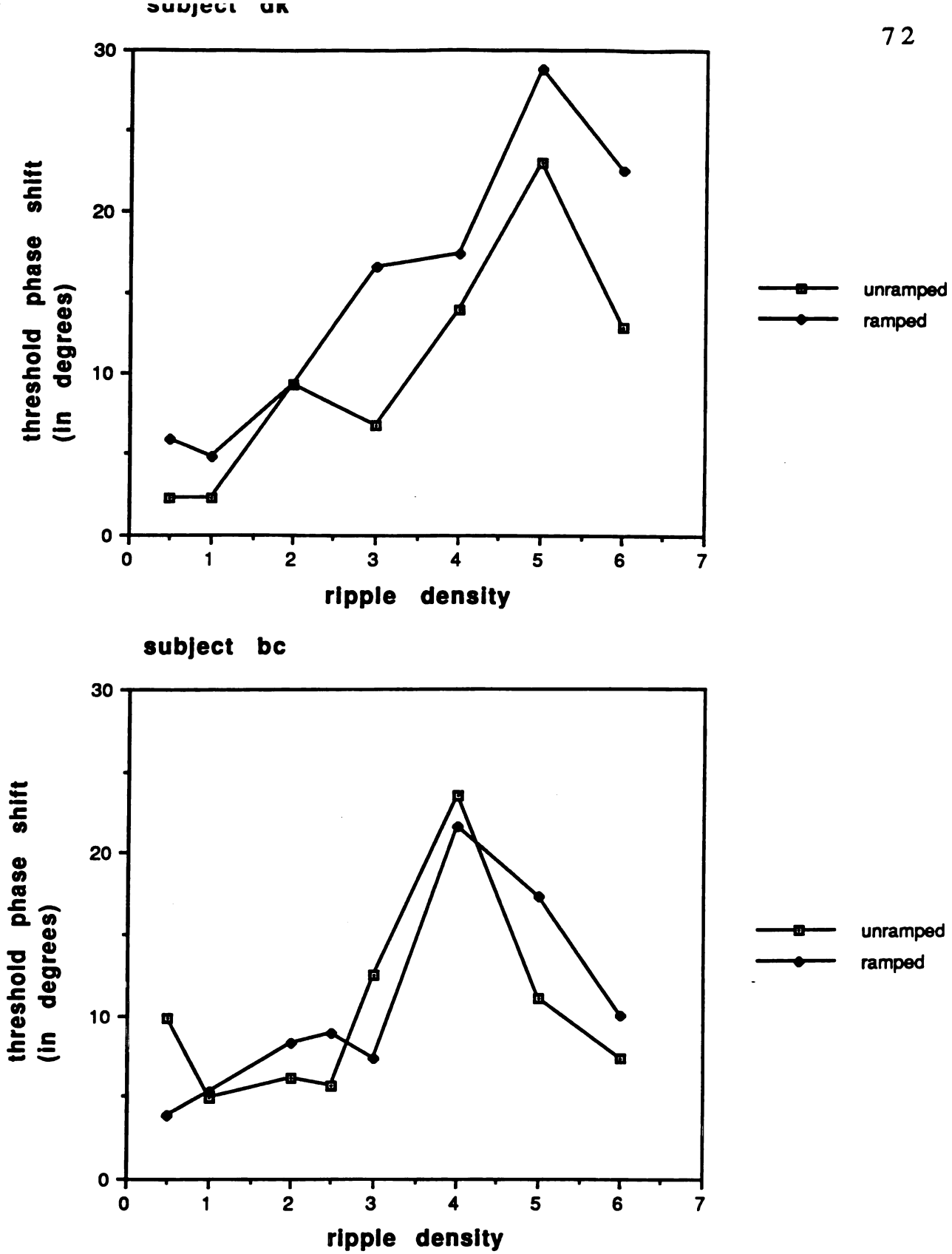


Figure 21. Thresholds for ramped and unramped stimulus conditions for two subjects. Some subjects may use the edge of the spectrum in their discrimination processing.

b. Depth of Modulation

The depth of modulation was varied, to shed light on the contrast required to discriminate spectral energy peaks. The results from these test sessions are shown in Figure 22. The positive phase shift thresholds are plotted as a function of the depth of modulation, for four different ripple densities. For both subjects, similar trends were observed, namely a) a rapid change of threshold with modulation depths between 2.5 and 7.5 dB, and b) a relatively shallow threshold curve for modulation depths of 10, 20 and 30 dB. In both subjects, thresholds were highest at a depth of modulation value of 2.5 dB for all four ripple densities tested. Thresholds then rapidly decreased at depth of modulation values of 5, and 7.5 dB. For the first subject (dk), there was a slight increase at a depth-of-modulation value of 10 dB across all 4 ripple densities, and then generally, a decline in threshold. The absolute lowest threshold for this subject was at a depth of modulation of 20 dB with ripple densities of 1 and 0.5 ripples/octave. For subject bc, the threshold decreased from 2.5 to 10 dB depth of modulation for all ripple densities except one (ripple density = 4 ripples/octave), where it was increased. The absolute lowest threshold for this subject was at a depth of modulation of 30 dB with a ripple density of 0.5 ripples/octave. For both subjects, lowest thresholds were at either 20 or 30 dB depth of modulation for all ripple densities.

It appeared from the results of Figure 22 that the thresholds at the lowest three depths of modulation: (2.5, 5, and 7.5 dB) grouped separately from the thresholds derived for the second three depths of modulation: (10, 20, and 30 dB). Therefore, threshold values were collapsed across ripple densities and divided into these two groups, and compared. A linear regression analysis was performed. The values for the two subjects were

combined, i.e. at each point, they were added and averaged. The regression for the first group of modulation depths was compared to the regression for the second group; the regression coefficients (slopes) were significantly different at $p<.01$. Figure 23 illustrates the comparison of the regression lines for the two groups of depth of modulation for the two subjects' combined values.

Thus, it appears that below a depth of modulation value of approximately 7.5 dB, the threshold for detecting ripple stimulus envelope phase shifts increases dramatically. Above the depth of modulation value of 7.5 to 10 dB, the threshold varies only slightly , at least for the depths of modulation presented herein, namely, 10, 20, and 30 dB.

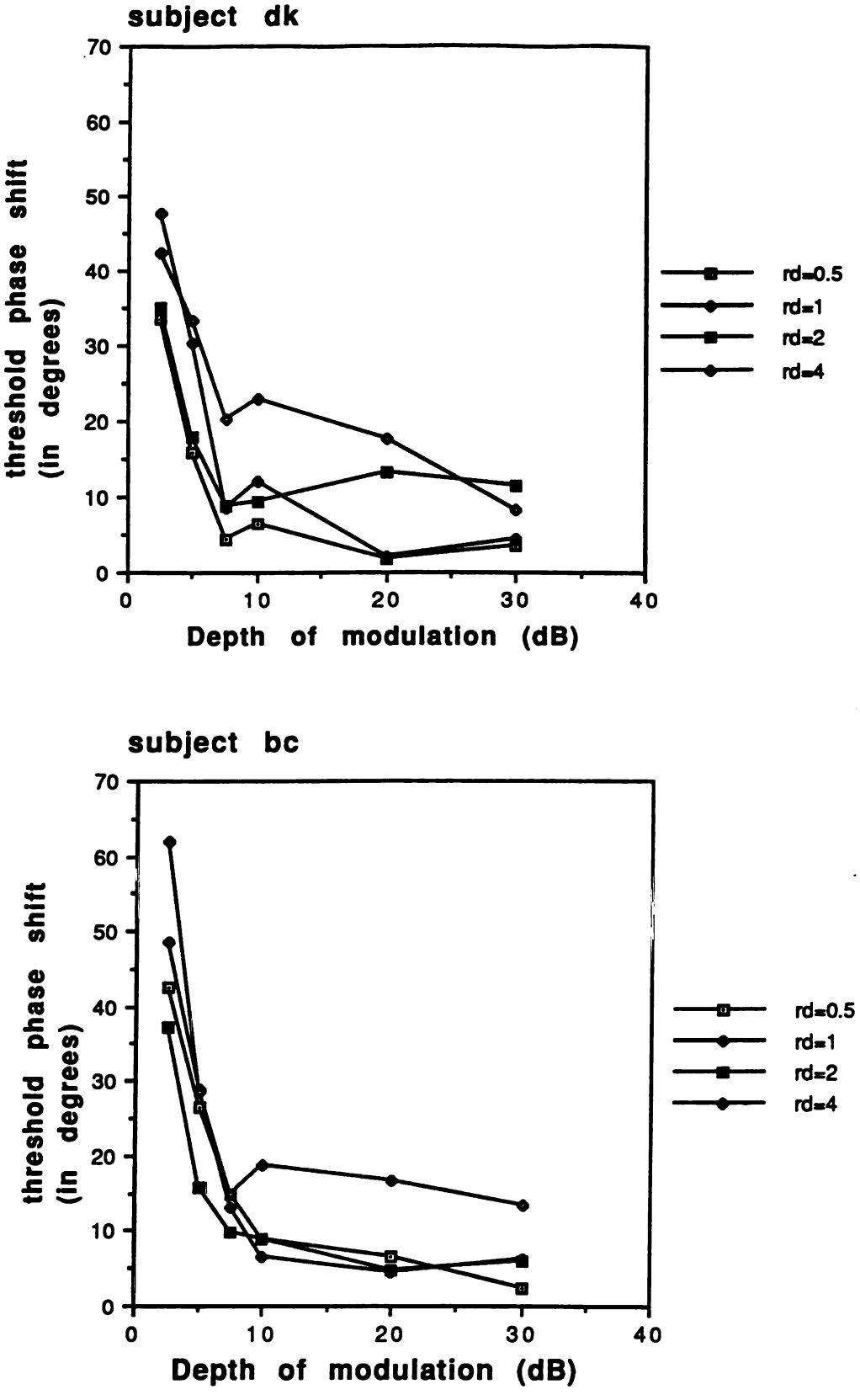


Figure 22. Threshold as a function of depth of modulation (two subjects). Two subjects were tested across four ripple densities at six different depths of modulation.

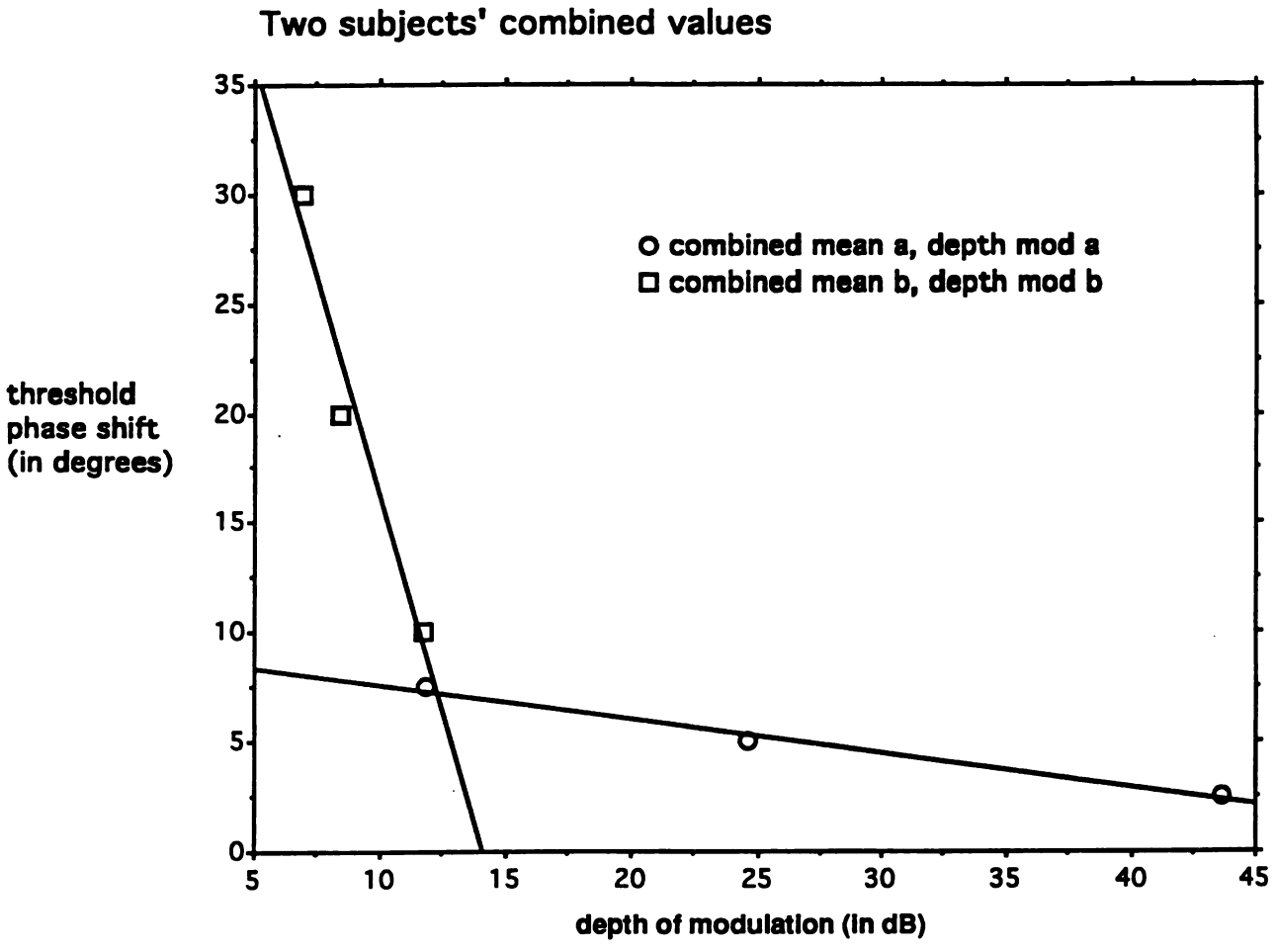


Figure 23. Comparison of the regression lines for the two groups of depth of modulation for the two subjects' combined values. Below a depth of modulation value of 7.5 dB, the threshold for detecting ripple stimulus envelope phase shifts increases dramatically.

c. Intensity

It is important to know whether the ripple phase thresholds measured are intensity independent, or whether they vary according to the intensity at which these stimuli are presented, since vowels are received and differentiated over a large range of levels. Results are shown in Figure 24. The positive phase shift thresholds are plotted as a function of ripple densities for five different intensity levels.

For subject kk, mean and standard deviation for the variability due to intensity at each ripple density, are shown in Table 3. When an ANOVA was computed comparing the values for three ripple densities across five intensities, the overall F was significant at p=.01, but there were no significant differences among the various intensities (for all post-hoc tests: Fisher PLSD, Scheffe, and Dunnett).

	mean threshold	standard deviation
rd=1	7.5	2.4
rd=3	10.7	2.4
rd=5	14.8	3.8

Table 3. Means and standard deviations across intensities for subject kk.

For subject dk, similar results were found. The mean and standard deviation for the variability due to intensity at each ripple density is shown in Table 4. When an ANOVA was computed comparing the values for three ripple densities across five intensities, the overall F was significant at $p<0.01$ (the curve was not flat), but there were no significant differences among the various intensities (for all post-hoc tests: Fisher PLSD, Scheffe, and Dunnett).

	mean threshold	standard deviation
rd=1	3.3	1.5
rd=3	9.5	2.0
rd=5	20.9	4.8

Table 4. Means and standard deviations across intensities for subject dk.

For subject bc, results were again similar. The mean and standard deviation for the variability due to intensity at each ripple density, is shown in Table 5. When an ANOVA was computed comparing the values for three ripple densities across five intensities, the overall F was not significant ($p=0.10$); there were no significant differences among the various intensities (for all post-hoc tests: Fisher PLSD, Scheffe, and Dunnett).

	mean threshold	standard deviation
rd=1	4.5	1.9
rd=3	5.9	3.0
rd=5	9.4	4.6

Table 5. Means and standard deviations across intensities for subject bc.

Thus, in general, thresholds appear to be independent of intensity over the tested range.

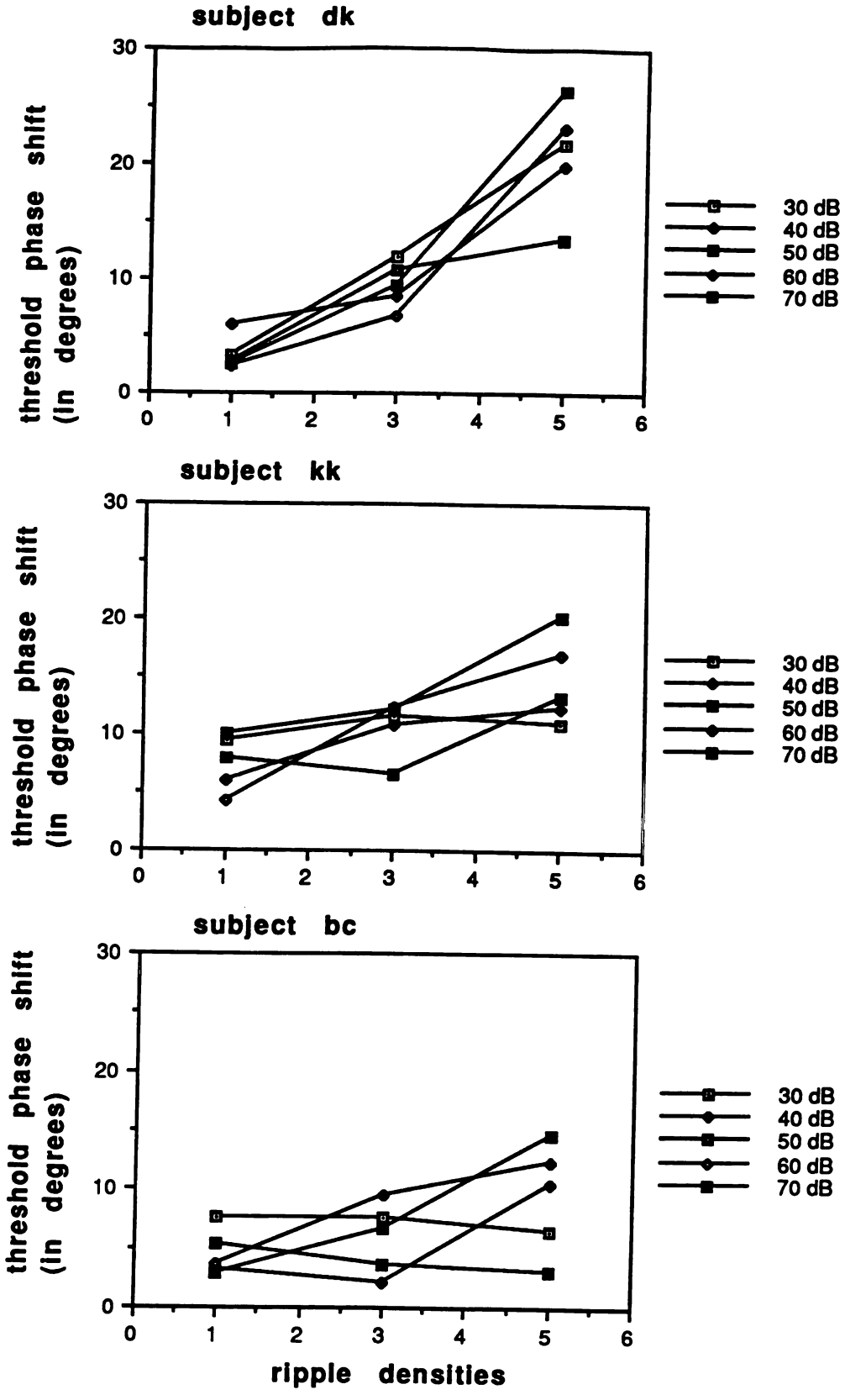


Figure 24. Threshold at various intensities (three subjects). Thresholds appear to be independent of intensity over the tested range.

Phase shifts compared to frequency differences

When the ripple stimulus is shifted, the frequency values at the location of the peaks change (actually the amplitude at the various constant frequencies changes so that different peaks occur). It is possible to calculate the frequency difference that would correspond to a particular ripple stimulus envelope phase shift using the formula: phase shift = (log (b/a))/log 2)*360* ripple density, where 'a' is the center frequency of the standard stimulus, in this case, 1.38 kHz, and 'b' is the "other" frequency. It was considered of interest to compare the frequency difference equivalent to the mean threshold envelope phase shift to frequency difference limens previously reported for pure tones. The frequency difference (in Hz) at the center frequency that corresponds to the average threshold for positive envelope phase shifts at each ripple density is listed in Table 6.

ripple density	mean threshold (in degrees)	difference in frequency (in Hz)
0.5	7.3	-38.25
1	7	-18.47
1.5	5.9	-10.41
2	9	-11.91
2.5	8.7	-9.22
3	14.2	-12.52
3.5	11	-8.33
4	20.5	-13.55
4.5	25.5	-14.97
5	25.5	-13.48
6	13.9	-6.14
7	21.1	-7.99
8	11.73	-3.89

Table 6. Frequency differences corresponding to threshold positive phase shifts.

Similar frequency differences for negative envelope phase shifts are listed in Table 7.

ripple density	mean threshold (in degrees)	difference in frequency (in Hz)
0.5	13.5	73.64
1	8.6	23.04
1.5	7.8	13.89
2	12.4	16.57
2.5	14.4	15.39
3	18.3	16.3
3.5	15.1	11.51
4	24.7	16.51
4.5	15.6	9.24
5	14.8	7.89
6	16.6	7.37
7	15.3	5.82
8	11.9	3.96

Table 7. Frequency differences corresponding to threshold negative phase shifts.

Wier, Jestead, and Green (1977) reported frequency difference limens (DLs) for humans, incorporating their own and previous data (see p.30). Fay (1988) reports that the frequency discrimination threshold (or frequency difference limen) at a frequency of 1000 Hz is 1.9 Hz and at 2000 Hz is 3.2 Hz for humans (p.458). Extrapolating from this information, one may estimate the frequency dl at 1.38 kHz to be approximately 2.5 Hz. From Tables 6 and 7, we observe that the smallest shift detected was equivalent to a frequency shift of 3.89 Hz (for a positive shift at 8 ripples/octave). The negative shift at 8 ripples/octave was quite

similar (3.96 Hz). Thus, the smallest discriminable frequency difference observed is larger than the frequency DL value reported by Fay, by almost a factor of 2. Although pure tone discrimination and ripple discrimination are clearly not the same, this comparison may be useful in the discussion of psychophysical processing mechanisms.

ANIMAL BEHAVIORAL TESTING

The first behavioral cat (#218) began training at age ten months. It required almost two months to shape the cat to reliably perform detection of ripple envelope phase shifts. The cat was subsequently tested over a period of five months before it underwent electrophysiological recording. Threshold was computed after each daily session as the smallest degree positive shift in phase of the envelope of the stimulus that the cat could detect with 75% or greater accuracy. An example of a psychophysical function for this cat is shown in Figure 25. Its threshold in this particular test session was approximately 57 degrees.

Figure 26a shows the performance of the animal over the time course of the test sessions. The average threshold over the first ten test sessions was about 127 degrees. The final threshold for this animal (its average over the last ten test sessions) was a phase shift of 80 degrees.

The second trained cat (#293) began training at age eight months; it required only 3 weeks to train to the procedure. It was tested over a six-month period, and then underwent electrophysiological recording. Figure 26b shows the improvement in threshold over time for this cat. The average threshold over the first ten test sessions for this cat was 76 degrees; its final threshold (the average of the last ten sessions) was 18 degrees.

The third trained cat (#176) began testing at age six months and required two months to train to the procedure. It was tested over a four month period, then underwent electrophysiological recording. Figure 26c shows the change over time of the threshold for this animal. The average threshold over the first ten test sessions for this cat was 66 degrees. Its final threshold (again, the average of the final ten test sessions) was 33 degrees.

A paired two-group one-tail t-test comparing the average threshold at the beginning of testing (averaged over ten consecutive testing days) to the final threshold for all three cats was performed, with the test hypothesis being that the latter threshold was smaller. The difference was significant at p=0.01.

A comparison of the three cats' performance over time is shown in Figure 26d. Note that the thresholds for the first trained cat (#218) were higher overall than the thresholds for the other two cats.

Two other cats (#122 and 40764) were unable to learn the discrimination procedure; they were able to respond to the stimulus on the basis of location of the sound, i.e. they were able to respond when two lateralized speakers were used. However, they were unable to make the switch to discriminating the stimuli coming from one overhead speaker, based on the quality of the sounds. It is interesting to note that these two cats, in contrast to the three successfully trained animals, were generally nervous, timid, and slow-moving. They also underwent electrophysiological recording, and were considered "exposed" control animals; i.e. they had been exposed to the stimuli, but had not performed operant discriminations based upon a perception of differences in these stimuli.

Conclusion

Three out of five cats were able to learn a two-alternative forced choice auditory discrimination procedure. The time required to train to the procedure varied from 3 weeks to 2 months. The threshold was lower for all three cats at the end of testing, as compared to the threshold earlier in training. Performance improvements were continuous and progressive.

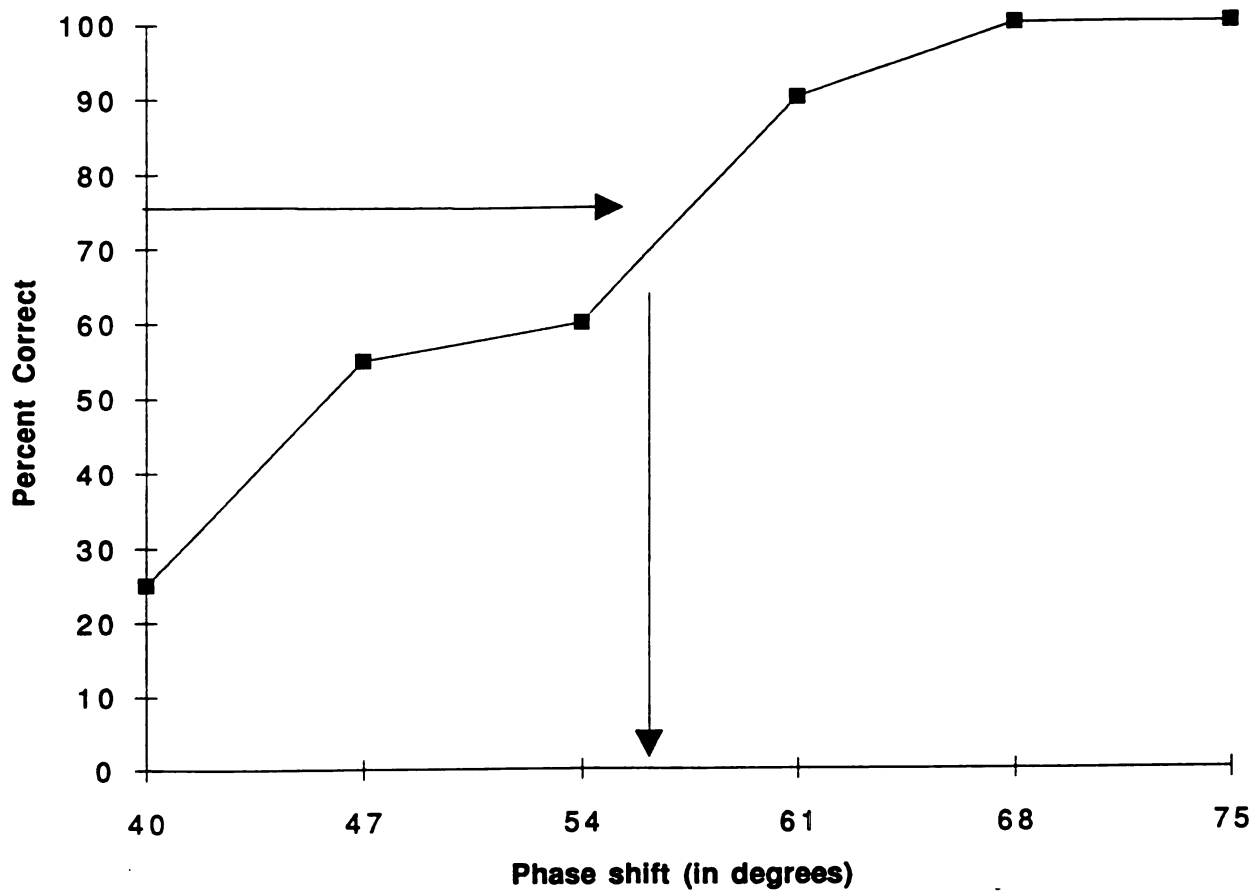


Figure 25. Typical psychophysical function for cat #218. Threshold envelope phase shift (75% correct) is approximately 57 degrees.

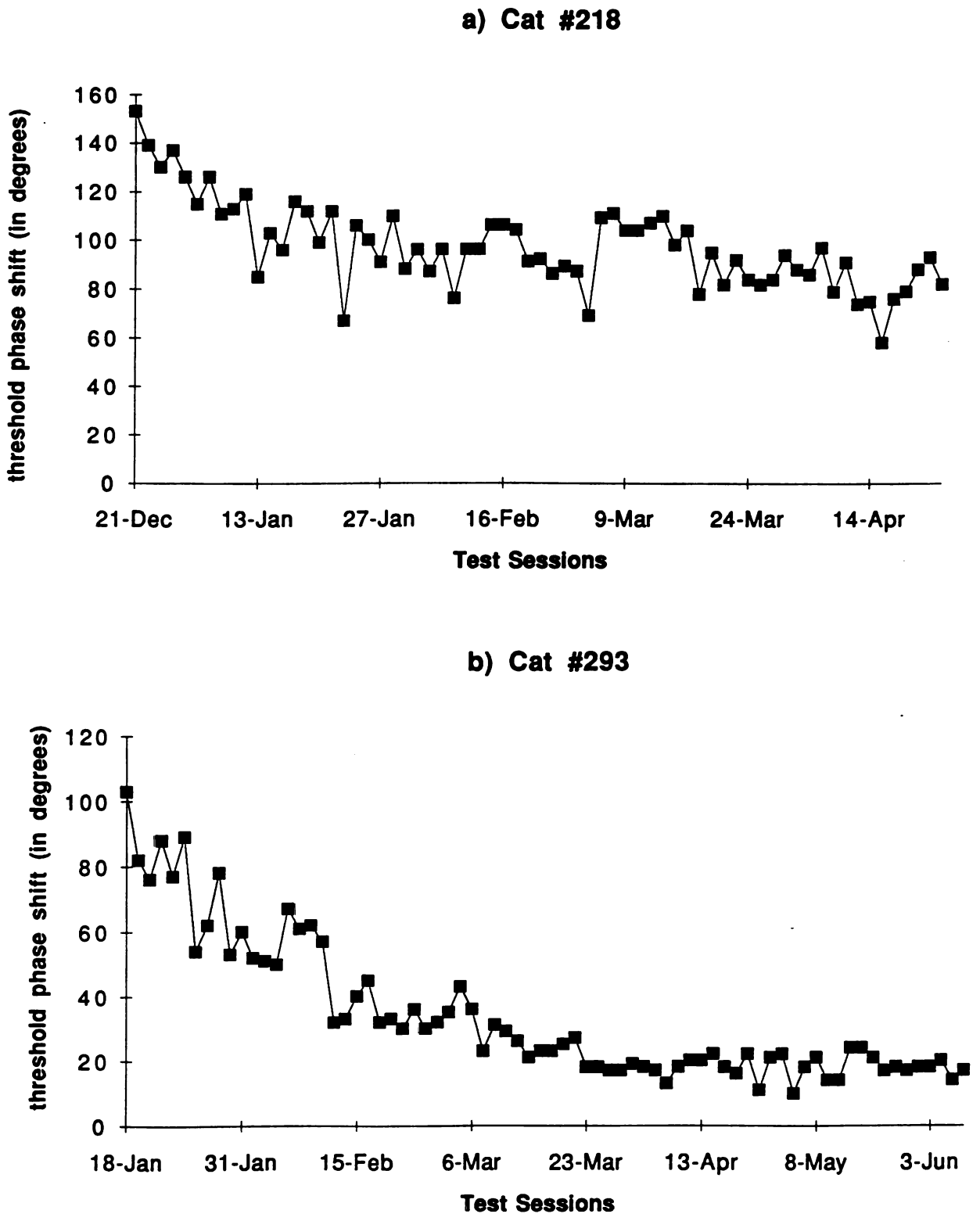
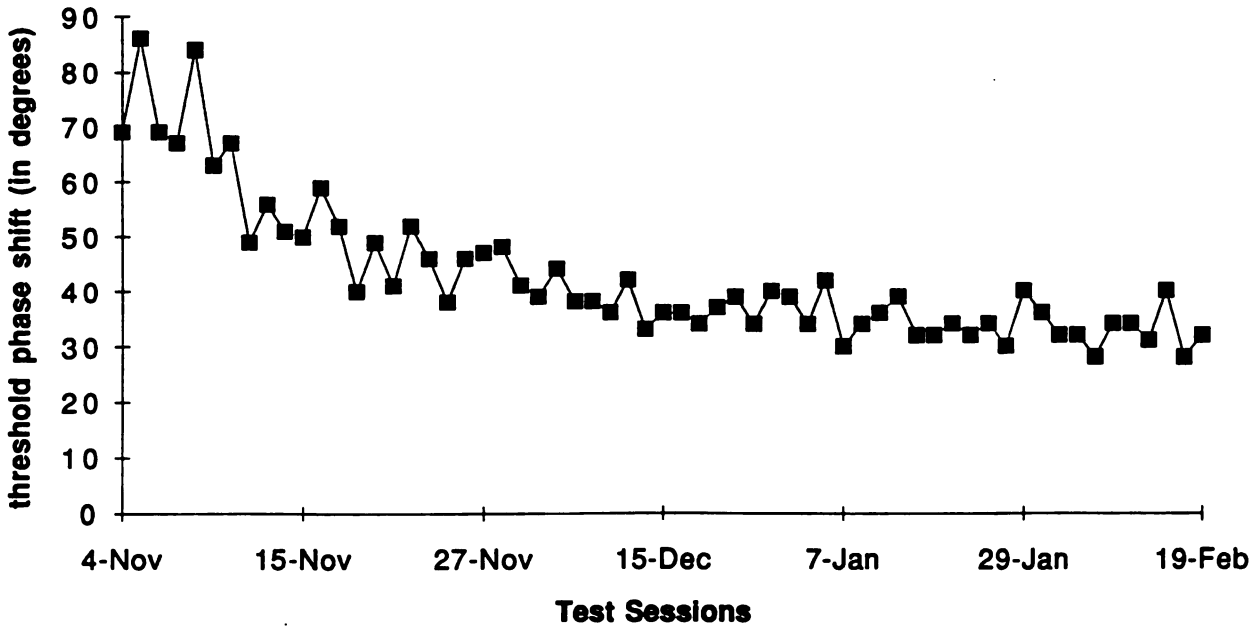


Figure 26 a) Performance over time of Cat #218, b) Performance over time of Cat #293.

c) Cat #176



d) Three trained cats

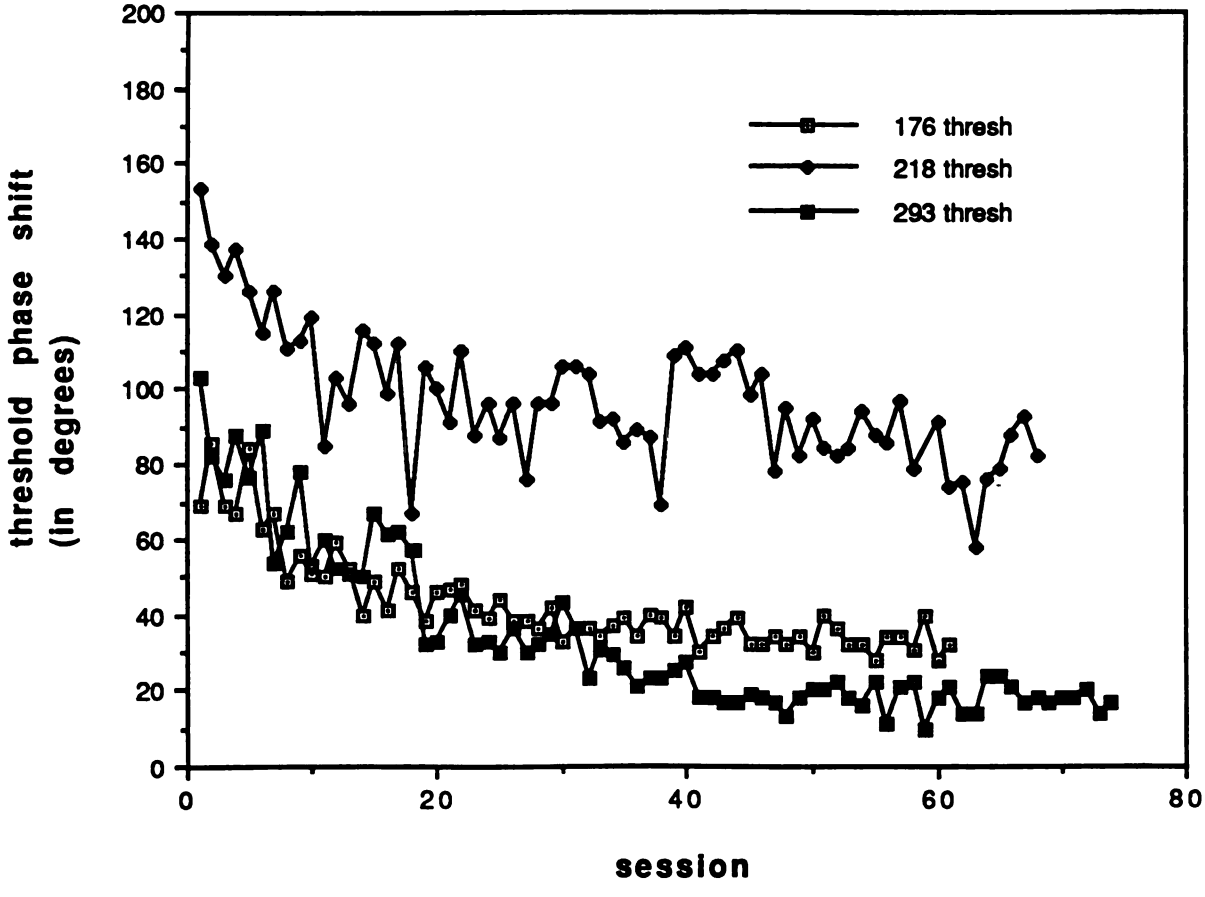


Figure 26 c) Performance over time of Cat #176, d) Comparison of thresholds over time of three trained cats.

ELECTROPHYSIOLOGICAL RECORDING

Several (eighteen) normal "untrained" cats provided data for this thesis. The recording protocol changed slightly over time as information about neural responses was gathered. Thus, not all untrained cat data are included in all analyses. The three trained cats, #218, 293, and 176 underwent recording at the end of behavioral testing. Two "exposed" cats, #122, and 40764, i.e. cats that had been placed on the training procedure but had been unable to learn the discrimination paradigm, also provided electrophysiological data.

Ripple Density Representation in Cortex

Ripple stimuli of various densities were presented during electrophysiological recording to both trained and untrained cats. In this section of the results, the spike counts recorded in the post-stimulus histograms were normalized, that is,. for the responses to the ripple stimulus, the best response was designated as 100% and the smallest 0; all other responses were computed as a percentage between 0 and 100. In this way, variation in the responses to the ripple could be more easily observed, and the responses could be compared to one another, and across animals. The "neural response" was defined as the normalized spike count averaged across all penetrations, and described as a function of the ripple density of the stimulus. For electrophysiological response recording to different ripple densities, the ripple stimulus was centered on the cf of the multi-unit penetration, then various ripple densities (0.5 to 8 ripples/octave) were presented.

The electrophysiological results for the trained cats will be discussed initially; untrained data will then be presented; the results from the trained and the untrained animals will then be compared.

Figure 27a shows spike counts averaged across all penetrations as a function of ripple density for the first trained cat, #218. The total number of spikes, recorded over thirty repetitions of the ripple stimulus presentation, during the 30 ms following response onset, across all penetrations, for each ripple density, were added and averaged. The best response occurred to a ripple density of 1 ripple/octave. This function may be described as a cosine ripple transfer function (cosine, because it represents the stimulus presented only at a phase of 0 degrees; a complete description of the ripple transfer function also requires the response to the stimulus at phase 90 degrees). The bandwidth halfway between the peak and the lowest value within the function may be measured. In this case, the bandwidth 50% down from the peak is 2.33 ripples/octave.

Figure 27b illustrates the neural response as a function of ripple density for the second trained cat, #293. In this animal the best response was at a ripple density of 0.33 ripples/octave. The bandwidth 50% down from the peak was 1.0 ripple/octave.

Figure 27c illustrates the neural response as a function of ripple density for the third trained cat, #176. The best response was at a ripple density of 1 ripple/octave (0.33 ripple density is quite close; the peak at 4.66 was the neural response to the ripple density presented first in the series). The bandwidth 50% down from the peak was 1.66 ripples/octave.

Figure 27d illustrates the neural response as a function of ripple density averaged across these three trained cats. The best response was at a ripple density of 1 ripple/octave. The bandwidth 50% down from the peak was 1.75 ripples/octave.

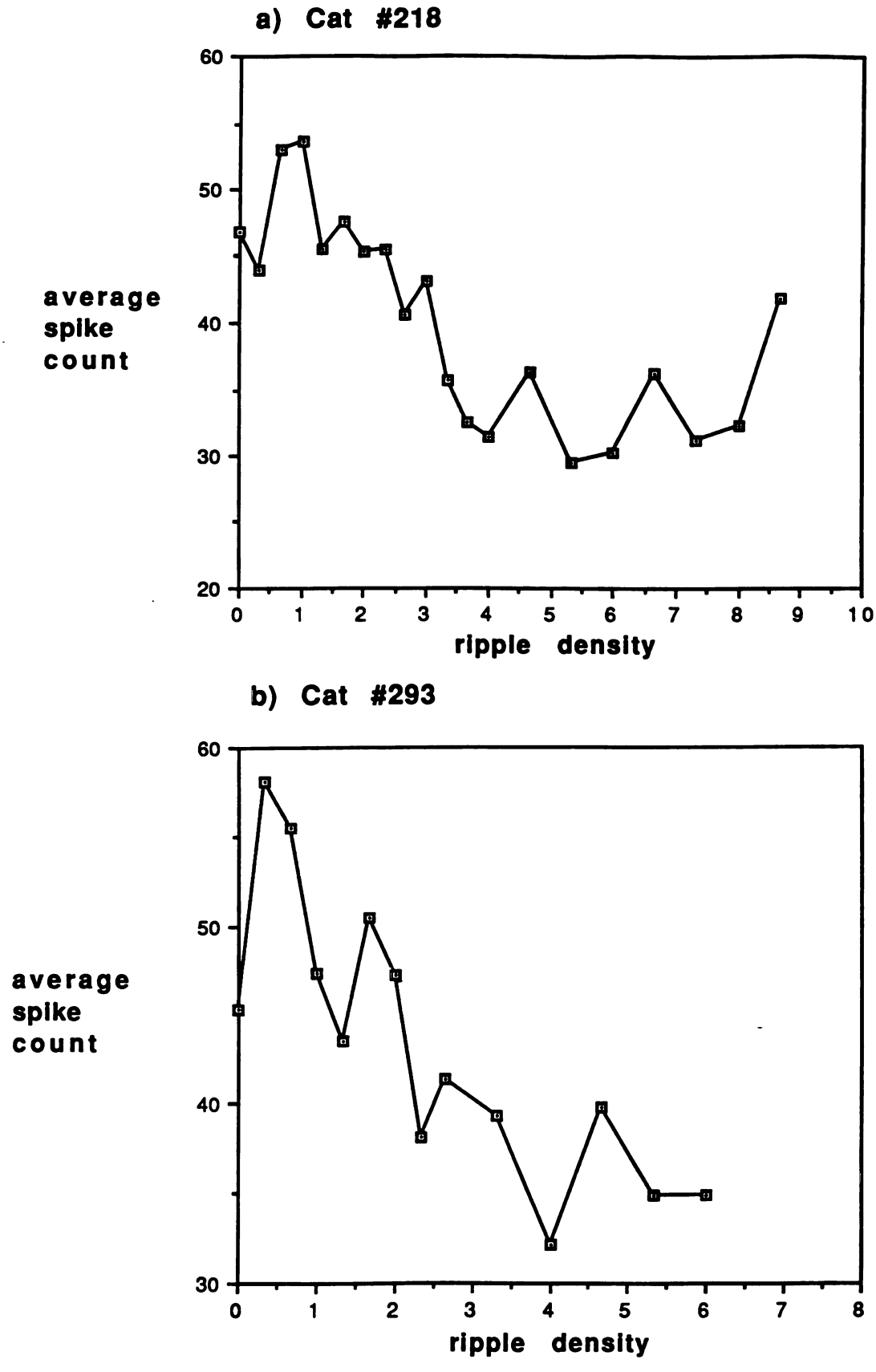
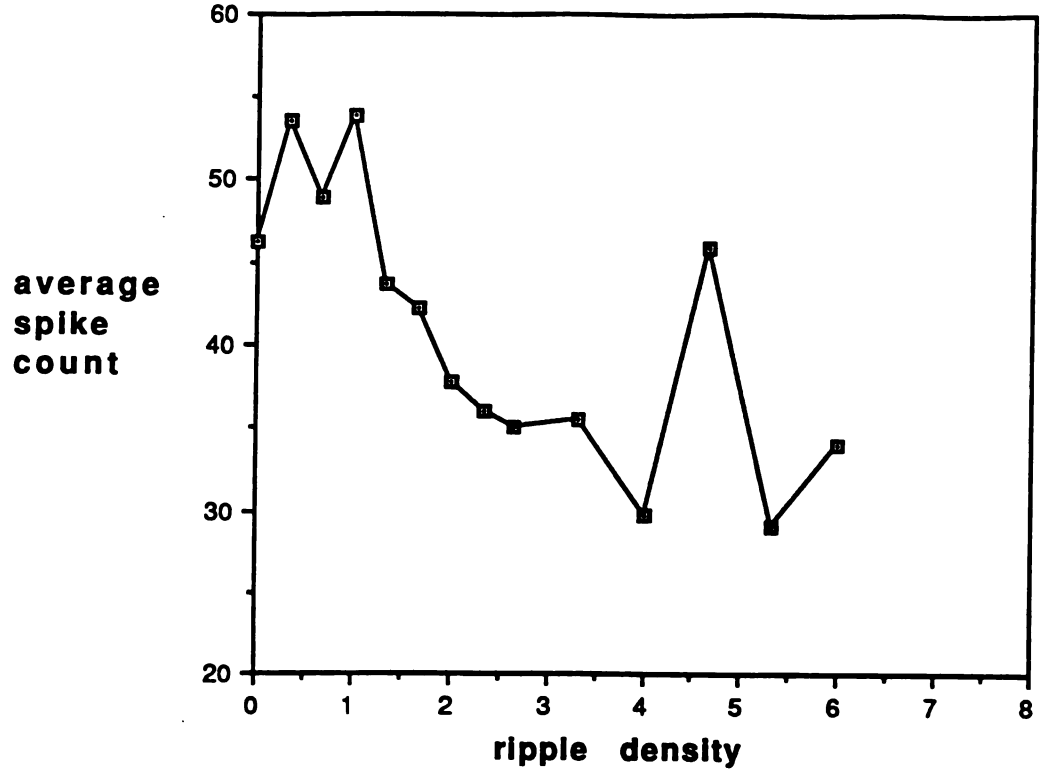


Figure 27 a) Neural response as a function of ripple density for Cat #218, and b) Neural response as a function of ripple density for Cat #293.

c) Cat #176



d) Average of three Trained cats

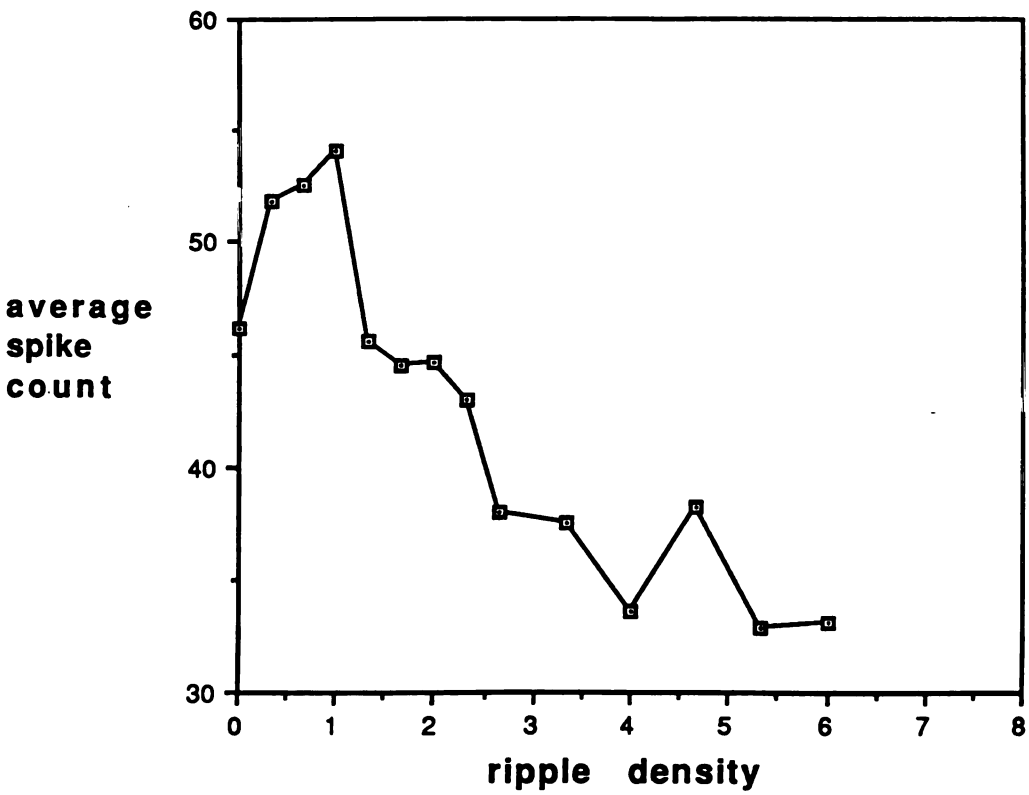


Figure 27 c) Neural Response as a function of ripple density for Cat #176, and d) Neural Response as a function of ripple density averaged across all three Trained Cats.

Figure 28 a,b,c,d,e, and f presents similar functions for five untrained cats (424, 188, 1681, 1673, and 92-1700) and for the average of the five untrained cats.

Table 8 presents the peak responses, bandwidths 50% down from the peak, number of microelectrode penetrations ('units'), and range of characteristic frequencies (cfs) included in the neural response for each of the three trained cats and for the five un-trained cats.

cat #	training status	r.d. peak response	bandwidth 50% down	number units	range of cfs (kHz)
218	Trained	1	2.33	45	2.3 to 9
293	Trained	0.33	1.2	77	2.3 to 11.8
176	Trained	1	1.66	50	2.7 to 11.7
681	Untrained	0.66	1.0	15	1.9 to 11.5
1673	Untrained	0.33	1.0	17	2.6 to 7.2
92-1700	Untrained	0.33	0.5	11	4.3 to 13.4
424	Untrained	0.66	1.33	18	1.5 to 7.9
188	Untrained	0.33	1.0	26	1.1 to 9.4

Table 8. Bandwidths at Best Ripple Response for Trained and Untrained Cats

Performing a Mann-Whitney U two-group un-paired comparison of the trained animals' bandwidths to those of the five untrained cats reported a significant difference ($p=0.05$). That is, the bandwidth 50% down from the best response to a particular ripple density was wider in trained cats than in untrained cats.

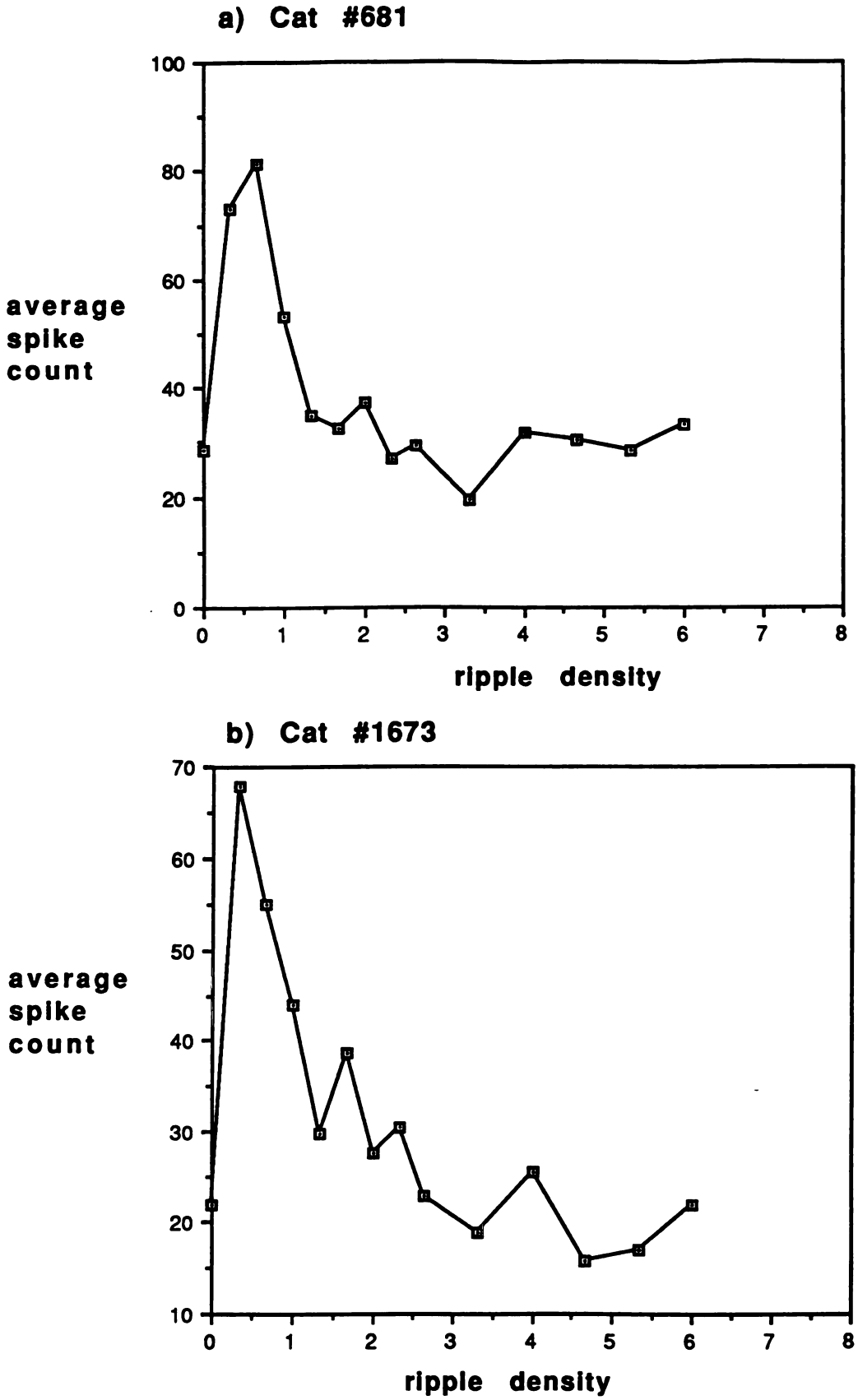
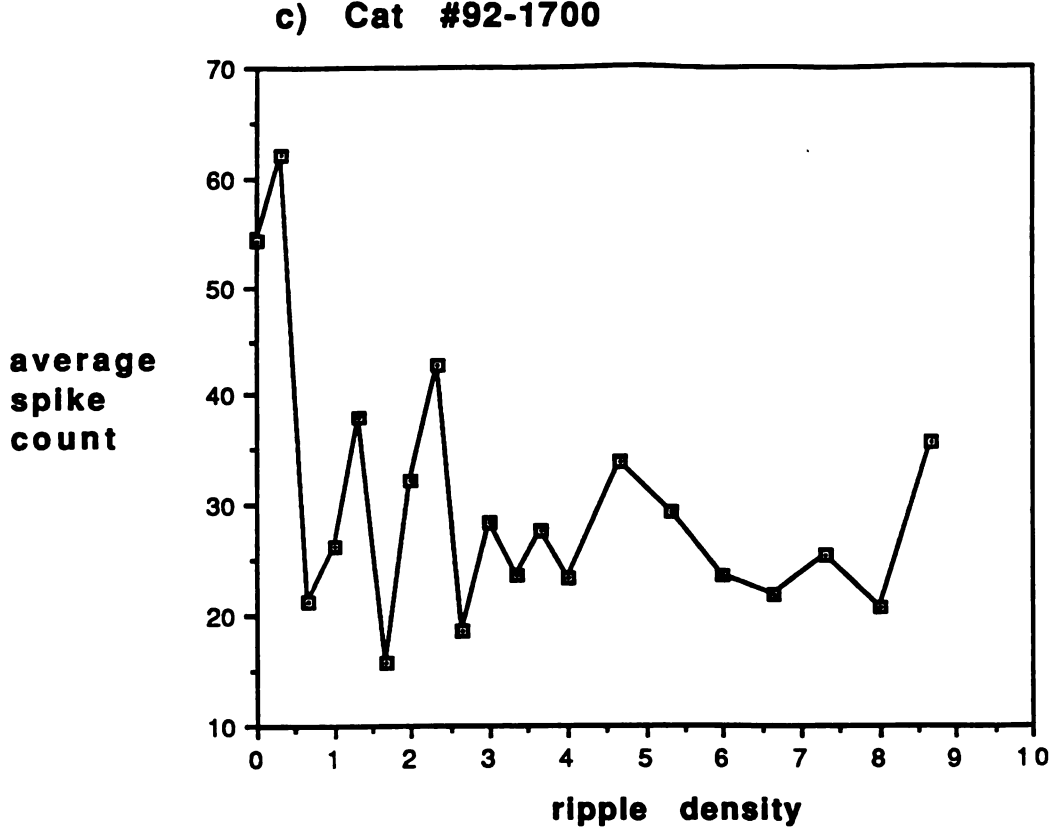


Figure 28 a) Neural response as a function of ripple density for Cat #681, and b) Neural response as a function of ripple density for Cat #1673.



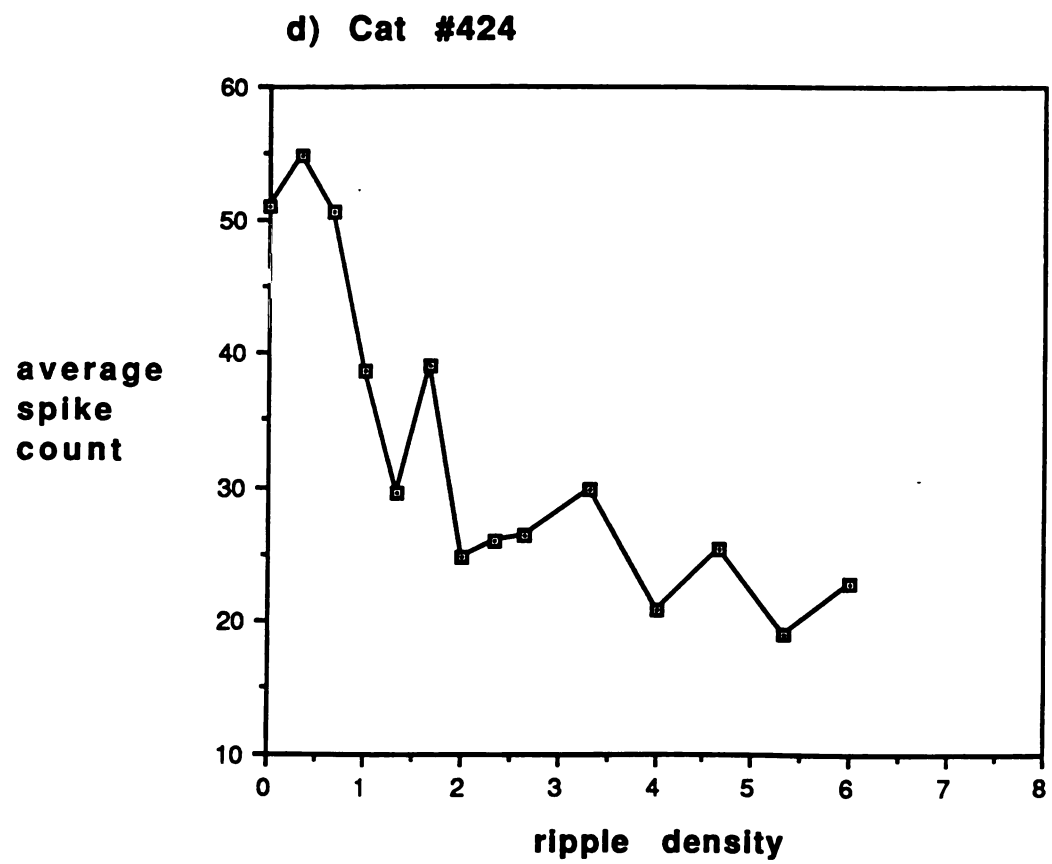
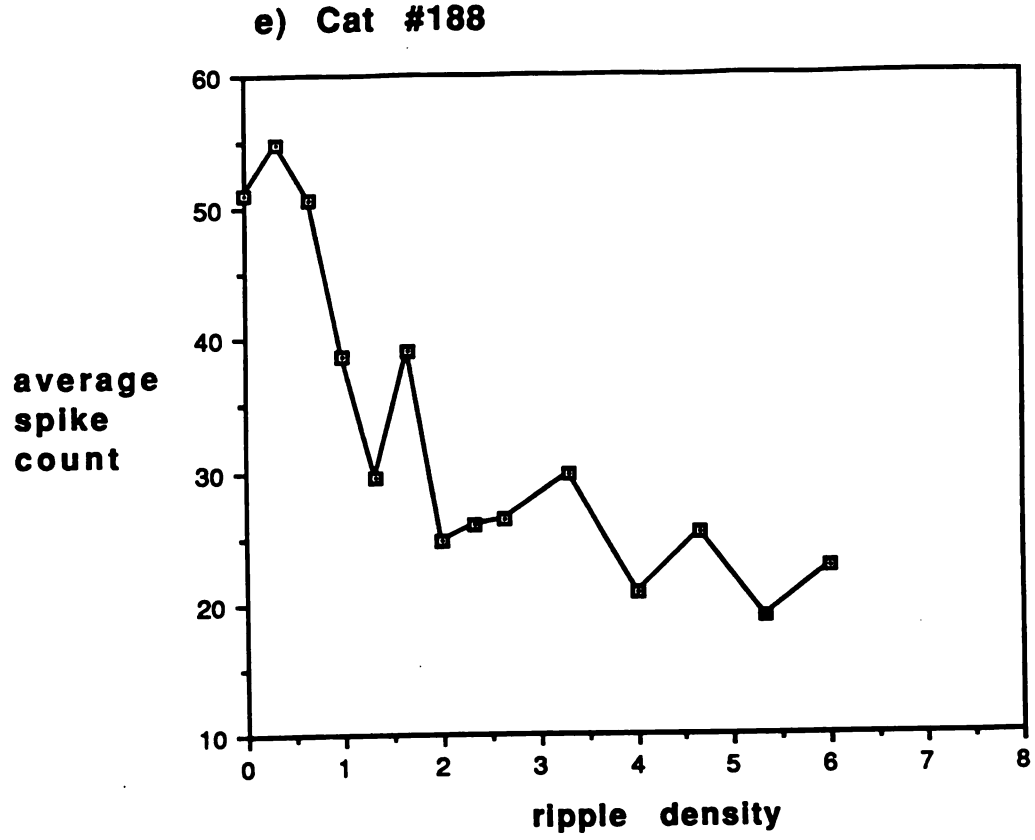


Figure 28 c) Neural response as a function of ripple density for Cat #92-1700, and d) Neural response as a function of ripple density for Cat #424.



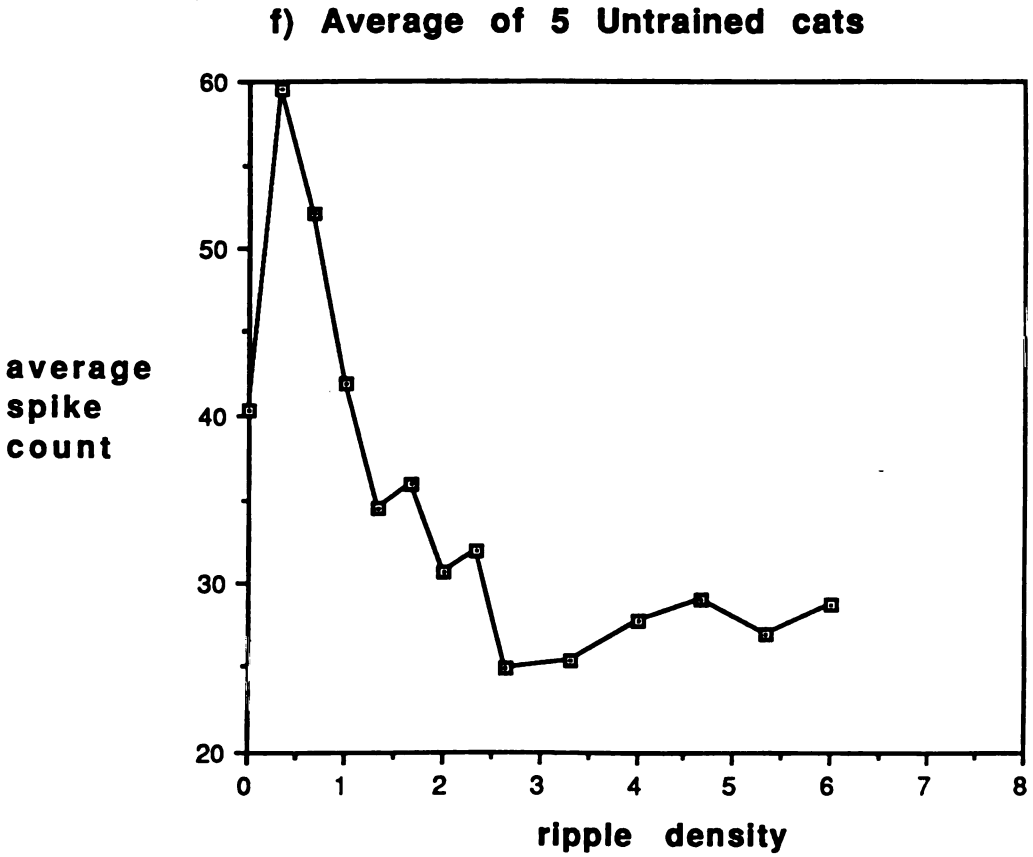


Figure 28 e) Neural Response as a function of ripple density for Cat #188, and f) Neural Response as a function of ripple density averaged across five Untrained Cats.

Figure 29 presents a comparison of the neural response as a function of ripple density for the average of the three trained cats and the average of five untrained cats. Performing a Wilcoxon Signed-Rank two-group paired comparison reported a difference significant at a level of p=0.001. Thus, the response to ripple densities was different in the trained cats as compared to the untrained cats. In trained cats, the best response was recorded in response to a higher ripple density than the best response in untrained cats. Overall, i.e., across all ripple densities presented, the responses were on average higher in the trained cats, except for the best response ripple density in untrained cats.

In conclusion, untrained cats showed the highest response to ripple densities of either 0.66 or 0.33 ripples/octave, with the mean best ripple response being at a ripple density of 0.33 ripples/octave. For three trained cats, the highest response was at either 1 or 0.33 ripples/octave, with the mean best ripple response being at a ripple density of 1 ripple/octave. The bandwidth of the ripple transfer function 50% down from the peak response was wider in trained cats as compared to trained cats.

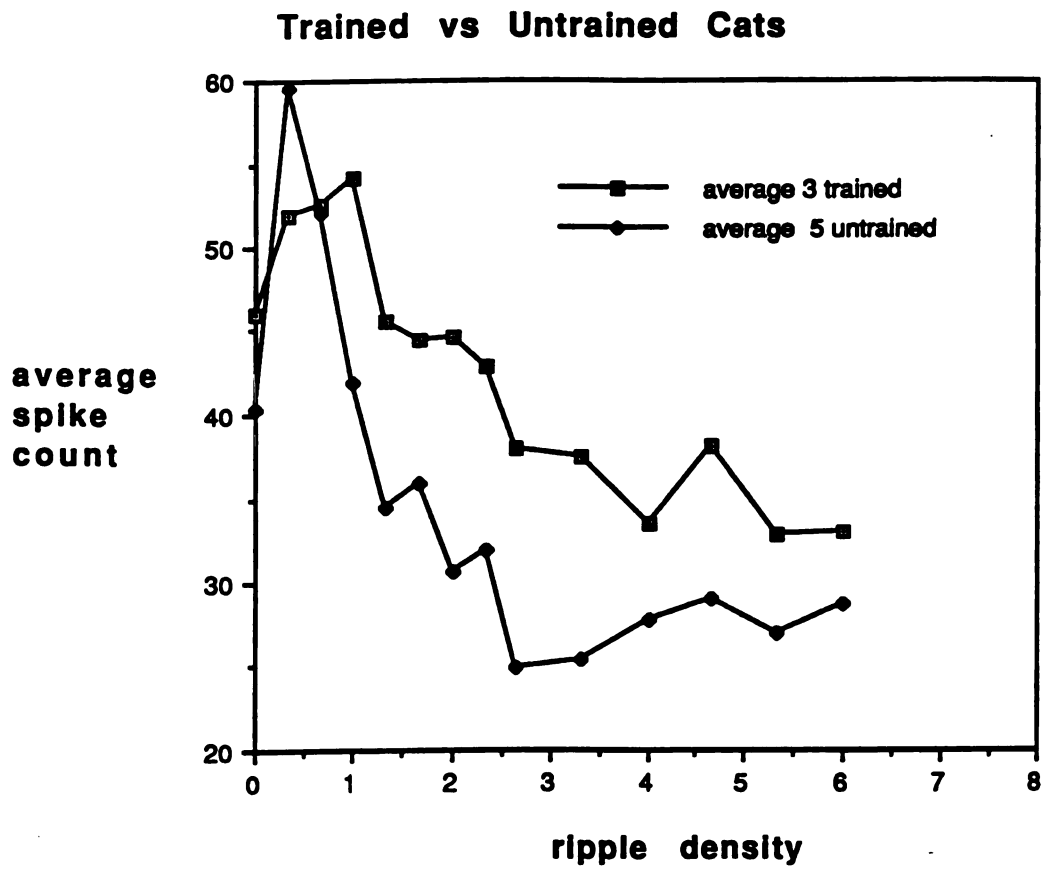


Figure 29. Comparison of Ripple Transfer Function for Trained Cats vs Untrained Cats

Phase Shift Representation in Cortex

Behavioral discriminability had been tested by shifting the phase of the envelope of the ripple stimulus and obtaining a threshold, i.e. the smallest phase shift in degrees that could be detected with 75% accuracy. During behavioral testing, the envelope of the stimulus was shifted only in a positive direction since it proved to be too difficult for the cats to learn to discriminate shifts in both directions. A positive envelope shift actually resulted in a downward change in the frequency of the peaks of the ripple stimulus. The ripple density of the behavioral stimulus and of the ripple stimulus that was phase-shifted in electrophysiological recording was one ripple per octave.

The trained cats discriminated a phase shifted stimulus that was always centered on 3.675 kHz. Therefore, for electrophysiological recording to phase shifts, the standard ripple stimulus was always centered on 3.675 kHz, even when the cf at that penetration was different.

In this first section of data analysis, average normalized responses across all penetrations were plotted against phase shift for each cat, irrespective of the cf of the location. Figure 30 displays this function for the three trained cats. Figure 31 displays the same function for four untrained cats. In both figures it is seen that for all cats, the best response occurs at either 0 or 10 degrees phase shift. The responses for the three trained cats were averaged together and compared to the average of the responses for four untrained cats. This function is shown in Figure 32. The best response for both groups of cats occurs at a phase of 0 degrees. It was possible to plot the spike counts only for the phase shift values that had been presented to all three trained and all four untrained animals, since an

average spike count had to be obtained at that value. Therefore, there are only 5 points for the trained cats, since the phase shift values had been selected to envelope the behavioral phase shift threshold, and they were different for each cat. Similarly, for the untrained animals, the phase shift values presented during electrophysiological recording were identical in 7 cases.

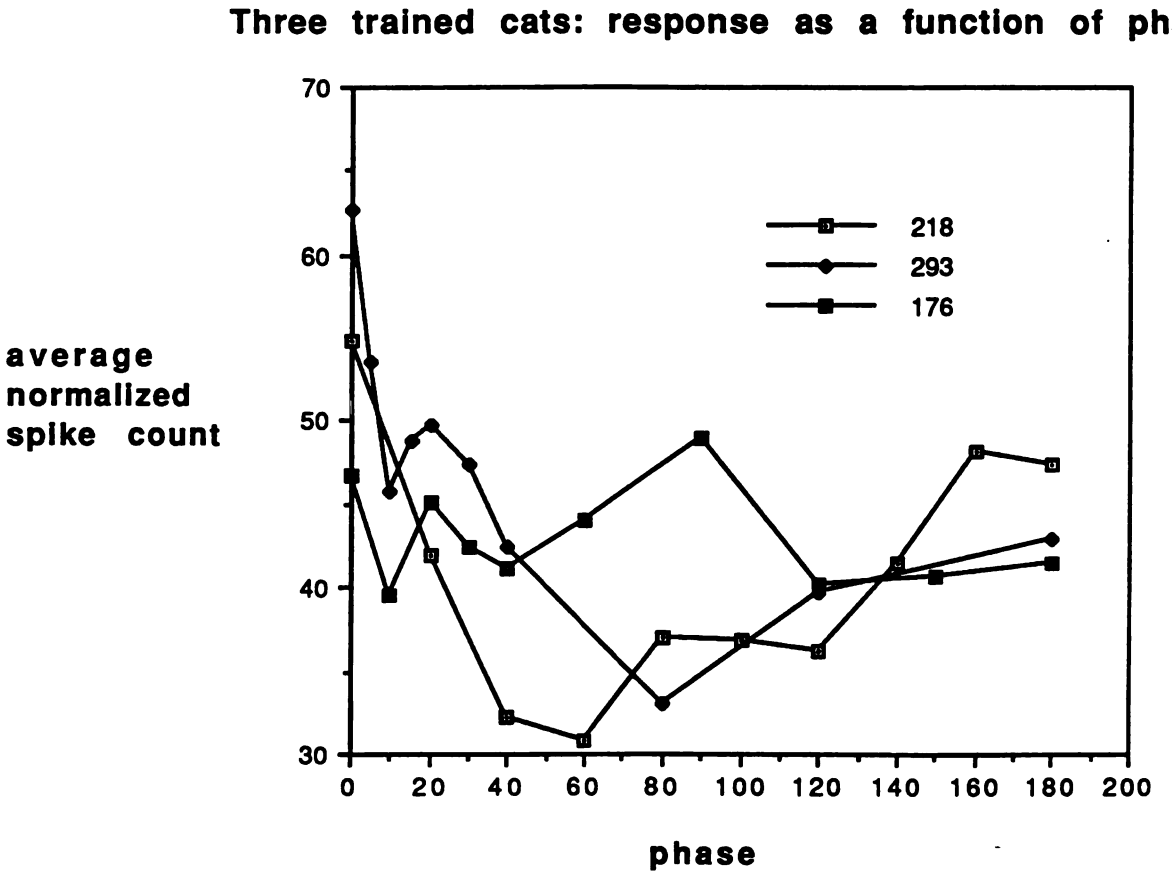


Figure 30. Neural response as a function of phase for three trained cats.

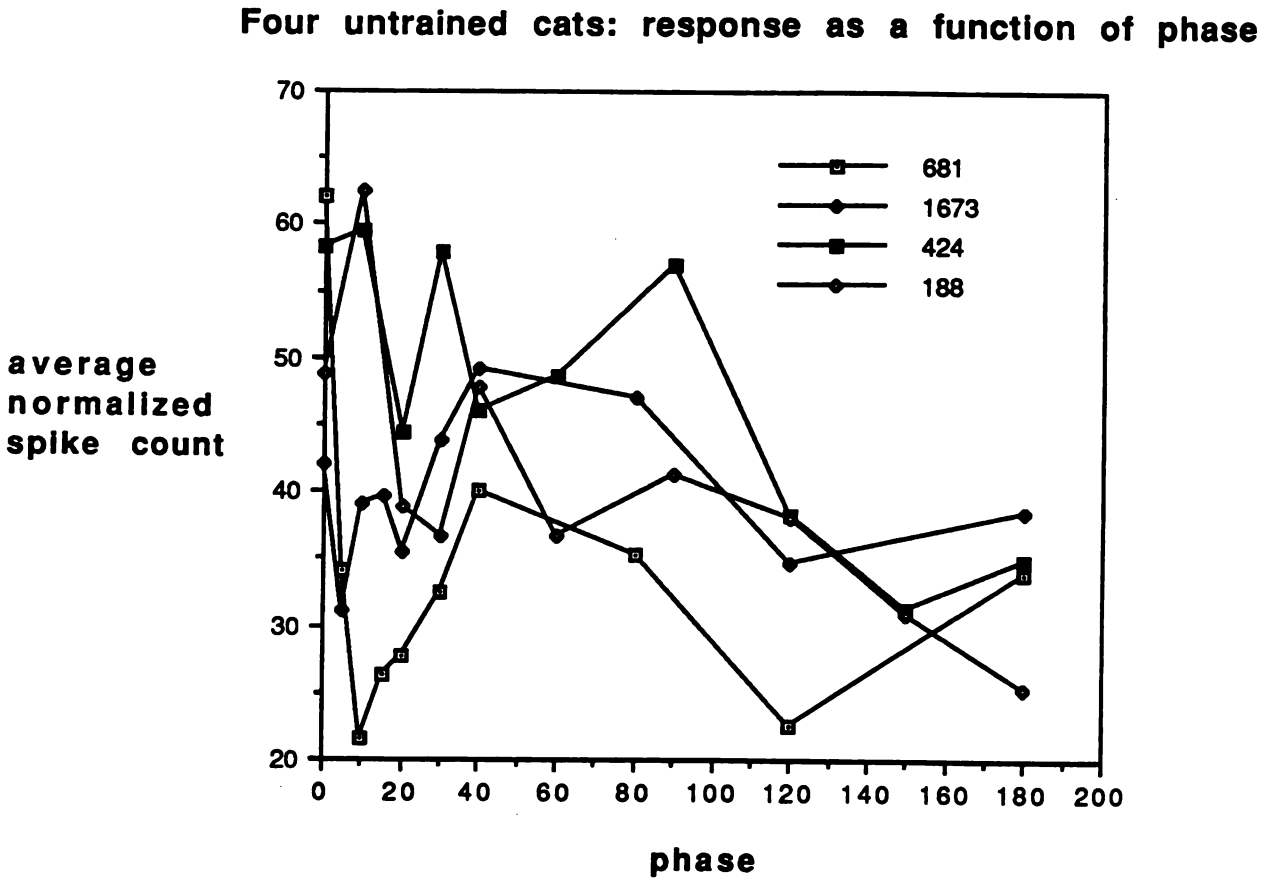


Figure 31. Neural response as a function of phase for four untrained cats.

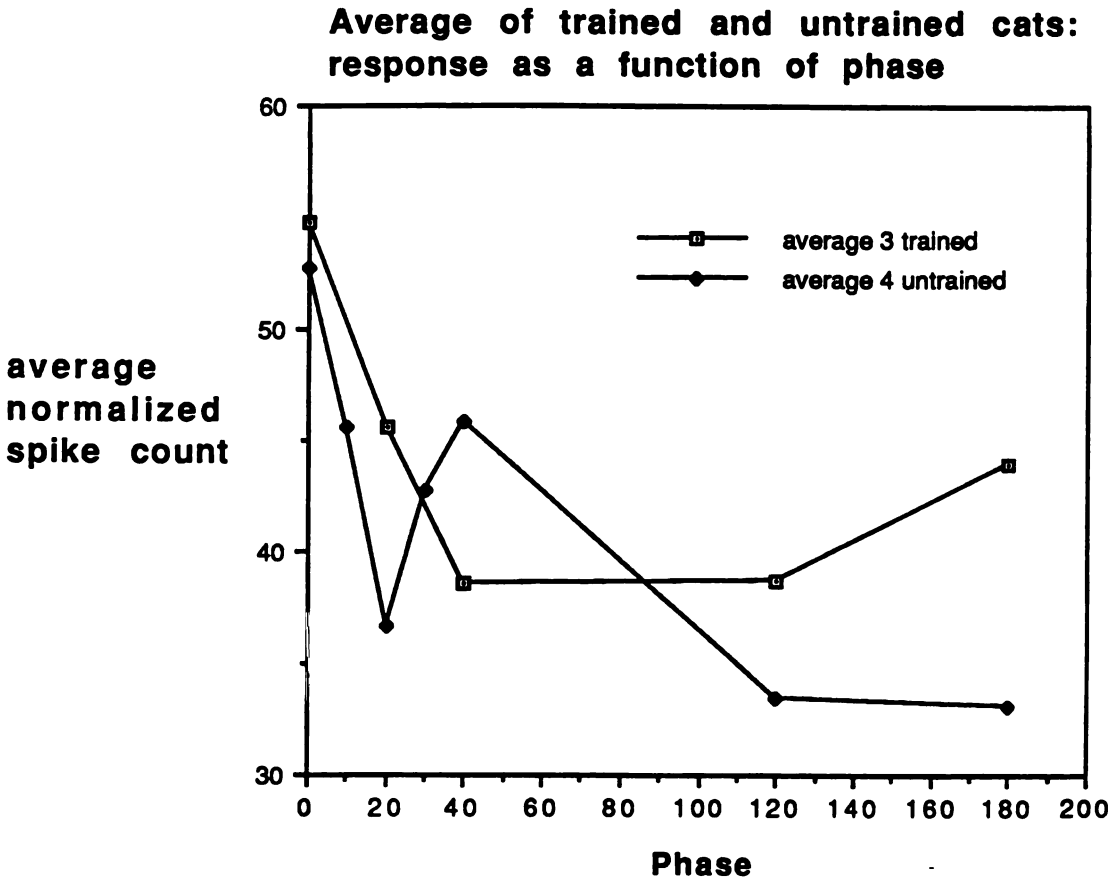


Figure 32. Comparison of Neural Response as a function of Phase for the Average of responses for Trained cats and Untrained cats.

Consideration of Penetration Response relative to Behavioral Stimulus

For the next section of the analysis of responses to phase shifts, the responses were divided into groups based upon 1) the location of the characteristic frequency of the penetration and 2) the phase shift that was presented to that penetration. The cf at each penetration was considered relative to the center frequency of the standard stimulus, namely 3.675 kHz; i.e with 3.675 kHz as the 0 degree location of the ripple phase shifts, the frequencies located equivalent to 0 to 180 degrees below and above this location were considered. A phase shift in degrees may be converted to frequency using the formula (log b= (phase shift/360 * log 2) + log a) where a is the frequency 3.675 kHz and b is the shift frequency. Those responses at cfs located from 0 to 90 degrees below the center frequency of the stimulus (3.675 kHz) (a positive phase shift), those located at from positive 90 to 180 degrees below 3.675 kHz, those located at from 0 to negative 90 degrees above the 3.675 kHz, and those located at from negative 90 to negative 180 degrees above 3.675 kHz were tabulated. The frequency below 3.675 kHz equivalent to a phase shift of 90 degrees is 3.1 kHz; the frequency below 3.675 kHz equivalent to a phase shift of 180 degrees is 2.6 kHz; the frequency above 3.675 kHz equivalent to a shift of negative 90 degrees is 4.3 kHz; the frequency above 3.675 kHz equivalent to a shift of negative 180 degrees is 5.1 kHz.

Ten phase shifts centered on 3.675 kHz were presented at each penetration. The phase shifts were positive only and ranged from 0 to 180 degrees. The responses were divided into the average spike counts in response to stimulus phase shifts of 0 to 90 degrees and 90 to 180 degrees.

The table below shows average normalized spike counts for the first trained cat #218 grouped by location of cfs relative to the center frequency of the test standard stimulus (3.675 kHz) and divided into 1) average responses to phase shifts 0 to 90 degrees, and 2) responses to phase shifts 90 to 180 degrees.

Cat	cfs located at	response at 0 to 90 shift	response at 90 to 180 shift
218	3.1 to 3.675 kHz (0 to 90 re 3.675 kHz)	43.2	34.21
	2.6 to 3.1 kHz (90 to 180 re 3.675 kHz)	35.92	62.74
	3.675 to 4.3 kHz (0 to neg. 90 re 3.675 kHz)	61.16	28.76
	4.3 to 5.1 kHz (neg. 90 to neg. 180 re 3.675 kHz)	29.87	39.6

Table 9. Response at Various cf Locations to Phase Shifts for Cat #218

One would expect the units located at 0 to 90 degrees relative to 3.675 kHz to respond well to a phase shift of the stimulus of from 0 to 90 degrees since the tips of their tuning curves are centered within this range. Similarly, those units whose cfs range from 2.6 to 3.1 kHz (located at 90 to 180 degrees relative to 3.675 kHz) are expected to respond best to a phase shift of the standard stimulus of from 90 to 180 degrees. Those units

located above 3.675 kHz are expected to respond progressively less well to a 0 to 90 degree shift and a 90 to 180 degree shift since the tips of the tuning curves of these penetrations are not located within the range of the shift of the center frequency. The units are however located within the frequency spectrum of the stimulus, which is three octaves wide; the highest (in frequency) peak of the three-octave-wide stimulus (one ripple/octave) at a phase shift of 180 degrees is 5.1 kHz.

The responses for all three trained cats were averaged (mean = 41.87, standard deviation = 12.84); they are presented in Table 10.

Three trained cats combined	**cfs located at**	**response at 0 to 90 shift**	**response at 90 to 180 shift**
	3.1 to 3.675 kHz (0 to 90 re 3.675 kHz)	47.65	37.17
	2.6 to 3.1 kHz (90 to 180 re 3.675 kHz)	39.42	56.36
	3.675 to 4.3 kHz (0 to neg. 90 re 3.675 kHz)	57.67	19.36
	4.3 to 5.1 kHz (neg. 90 to neg. 180 re 3.675 kHz)	38.57	38.78
Total		**183.31**	**151.67**

Table 10. Response at Various cf Locations to Phase Shifts for Three Trained Cats (combined)

Similar responses for three untrained cats were tabulated (mean = 40.00, standard deviation = 16.3). The responses combined across these cats (#188, 424, and 1673) are presented in Table 11.

Three untrained cats combined	**cfs located at**	**response at 0 to 90 shift**	**response at 90 to 180 shift**
	3.1 to 3.675 kHz (0 to 90 re 3.675 kHz)	50.36	29.33
	2.6 to 3.1 kHz (90 to 180 re 3.675 kHz)	38.79	37.69
	3.675 to 4.3 kHz (0 to neg. 90 re 3.675 kHz)	58.02	24.10
	4.3 to 5.1 kHz (neg. 90 to neg. 180 re 3.675 kHz)	40.45	41.23
Total		**187.62**	**132.35**

Table 11. Response at Various cf Locations to Phase Shifts for Three Untrained Cats (combined)

Over the course of training and testing, the trained cats were presented with phase shifts of from 0 to 180 degrees. Thus those frequencies of from 2.6 to 3.675 kHz were included as center frequencies of the phase-shifted stimulus that the cat was asked to discriminate as different from the standard stimulus centered on 3.675 kHz. One might ask whether there was a difference in the neural response of the trained cats at those frequencies that corresponded to center frequencies at which

they had learned discriminations, as compared to those frequencies that did not correspond to center frequencies at which they had made discriminations. Thus, the responses at penetrations whose cfs were located at from 2.6 kHz to 3.675 kHz were grouped, to compare them to penetrations whose cfs were located at from 3.675 to 5.1 kHz.

Within these two groups, those frequencies nearest in proximity to the center frequency of the standard stimulus (3.675 kHz) might be expected to show an increase in response relative to those frequencies farther away. Thus the locations grouped into 2.6 to 3.675 kHz were further grouped into 0 to 90 degrees from 3.675 (namely 3.1 to 3.675 kHz) and those farther away at 90 to 180 degrees (namely 2.6 to 3.1 kHz). Similarly, the cfs located at from 3.675 to 5.1 kHz were further grouped into 0 to negative 90 degrees (namely 3.675 to 4.3 kHz) and those at negative 90 to negative 180 degrees (namely 4.3 to 5.1 kHz).

It should be remembered that the behavioral threshold for the first trained cat #218 was 80 degrees, for the second trained cat, 18 degrees, and for the third trained cat, 33 degrees. Thus, behavioral thresholds for all three trained cats were located within the phase shift range of 0 to 90 degrees. It may be that the responses that would reflect the effect of training and testing are those occurring in response to a shift of from 0 to 90 degrees only.

A two-factor ANOVA was performed for the spike counts (Y dependent variable) of each of the two animal groups: the trained animals and the untrained animals. The first independent variable was the location of the cf of the penetration; there were four categories within this variable: 1) frequencies at 0 to 90 degrees relative to 3.675 kHz (3.1 to 3.675 kHz), 2) frequencies at 90 to 180 degrees (2.6 to 3.1 kHz), 3) frequencies at 0 to

negative 90 degrees relative to 3.675 kHz (3.675 to 4.3 kHz), and 4) frequencies at negative 90 to negative 180 degrees (4.3 to 5.1 kHz). The second independent variable was the phase shift presented, i.e., 0 to 90 degrees vs 90 to 180 degrees. The individual values for each of the three trained and each of three untrained cats were used (as opposed to the averages shown above).

For the untrained animals, the phase shift significantly affected the spike count ($p=0.04$). Neither the cf location nor an interaction between cf location and phase shift was significant. For the trained animals, the phase shift significantly affected the spike count ($p=0.01$) and there was a highly significant interaction effect ($p=0.0001$) of the location of the cf of the penetration and the phase shift presented at that penetration. This indicates that for the trained cats, the location of the cf of the penetration in combination with the phase shift presented affected the neural response. For untrained cats, the location of the cf of the penetration did not affect the response. This result implies that the training of the cats created an emergent interactive effect, whereby the phase shift presented affected the neural response, but only at certain frequency locations.

Phase Shift Response Adjusted for cf of Penetration

It is clear that neurons have receptive fields, i.e. excitatory and inhibitory response areas that jointly make up the tuning curve for that neuron or small cluster of neurons. It has been observed that a tuning curve contains a characteristic frequency - that frequency at which the lowest intensity stimulus evokes a neural response. This frequency is considered to be the frequency to which this neuron is most sensitive. For example, Figure 33 illustrates the average of 6 single unit responses to phase shifts in untrained cat #301. The peak is at a phase of 0 degrees with decreasing response on both sides of the peak.

Therefore, it was expected that a ripple stimulus with center frequency of 3.675 kHz would evoke the best neural response when presented to a cortical penetration whose cf was 3.675 kHz. The phase shift of such a stimulus is 0 degrees. Therefore, it was considered necessary to analyze the neural response to phase shifts, adjusting for the cf of the penetration.

A frequency (in Hz) may be converted to a phase shift value (in degrees) using the formula: (phase shift = (log (f/3.675))/log 2 * 360), where "f" is the cf of the penetration. For example, for a penetration with cf of 4 kHz, a ripple stimulus presented at phase 0 degrees with center frequency 3.675 kHz, is equivalent to a ripple stimulus presented at 44 degrees positive for 4 kHz. Similarly, a penetration with cf of 3 kHz, is equivalent to a phase shift of -105 degrees In this way, the phase shifts centered at 3.675 kHz were converted and adjusted with respect to the cf of each penetration. For each animal, a function could be created which showed the neural response at phase shift values from -180 to +180 degrees. Figure 34 shows the neural response as a function of phase

presentation for untrained cat #681. Spike counts were corrected to the strength of the response observed within the tuning curve, that is, the strength of response was characterized by taking an average spike count of five responses within the tuning curve; then the responses to ripple stimuli were divided by this strength of response correction factor at each penetration. The maximum response is at a phase value of -20 degrees, closely followed by the response at 0 degrees.

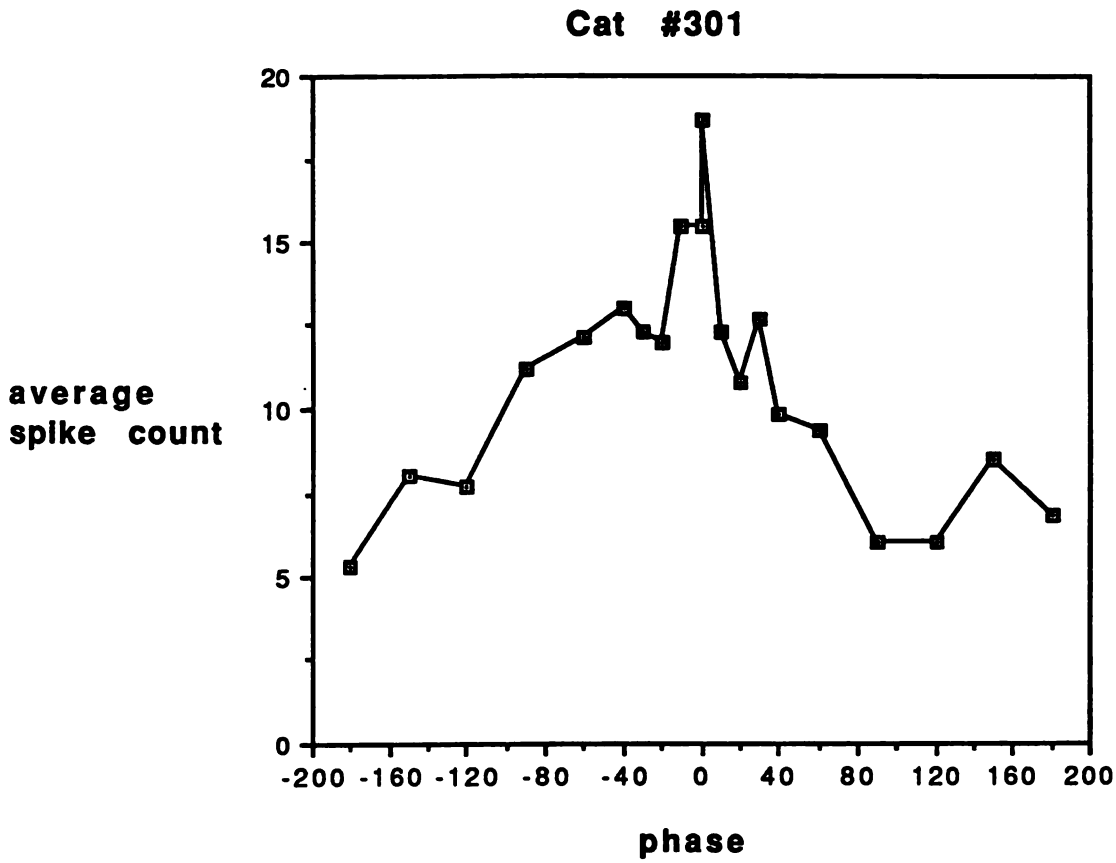


Figure 33. Neural response as function of phase presentation for cat #301. The best response is at a phase value of 0 degrees.

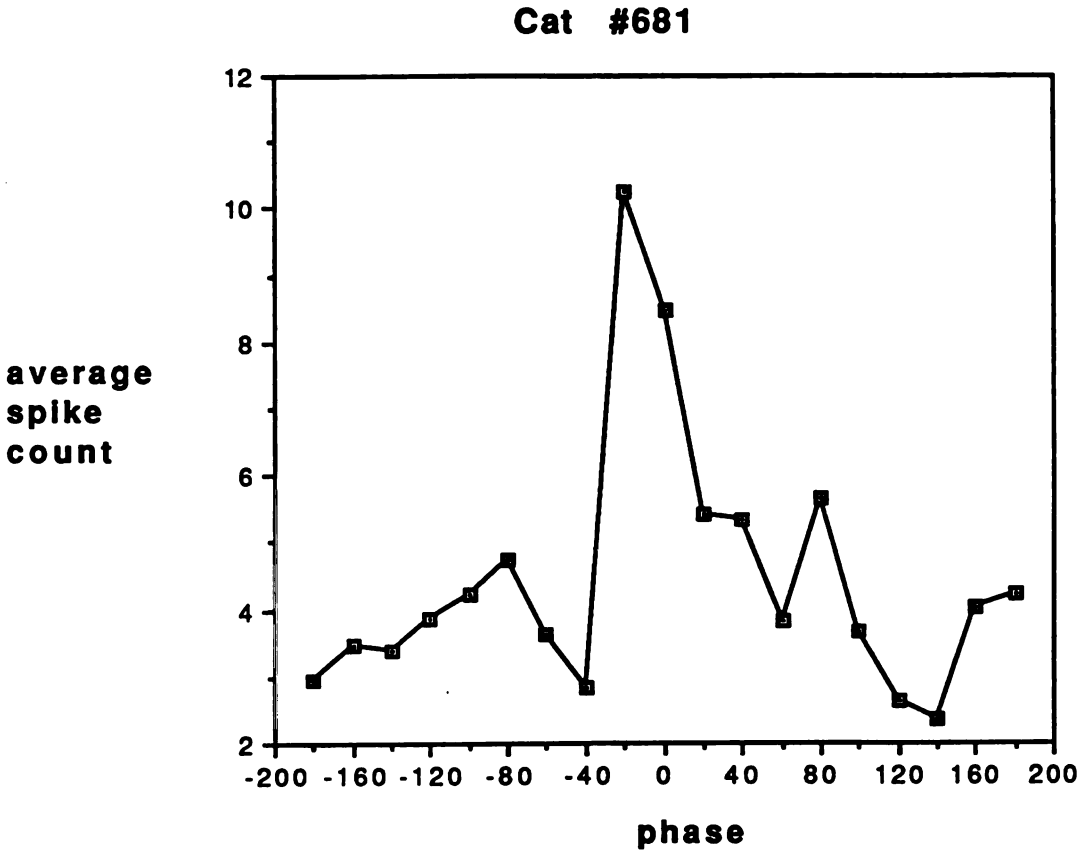


Figure 34. Neural response as a function of phase presentation for cat #681. The best response is at a phase value of -20 degrees, followed by 0 degrees.

In a similar manner, the responses as a function of phase presentation were computed for three other untrained animals and for the three trained animals. The neural response as a function of phase presentation for the three trained cats is illustrated in Figure 35. In Figure 36, the average neural response for four untrained cats is plotted against the average neural response for the three trained cats. The maximum spike count value for the untrained cats' average was at -20 degrees, closely followed by 0 degrees. The maximum spike count value for the trained cats' average was at +40 degrees. The spike count values for the trained cats were compared to those for the untrained cats. The values were different at a significance level of $p=0.0001$ (Mann-Whitney U test).

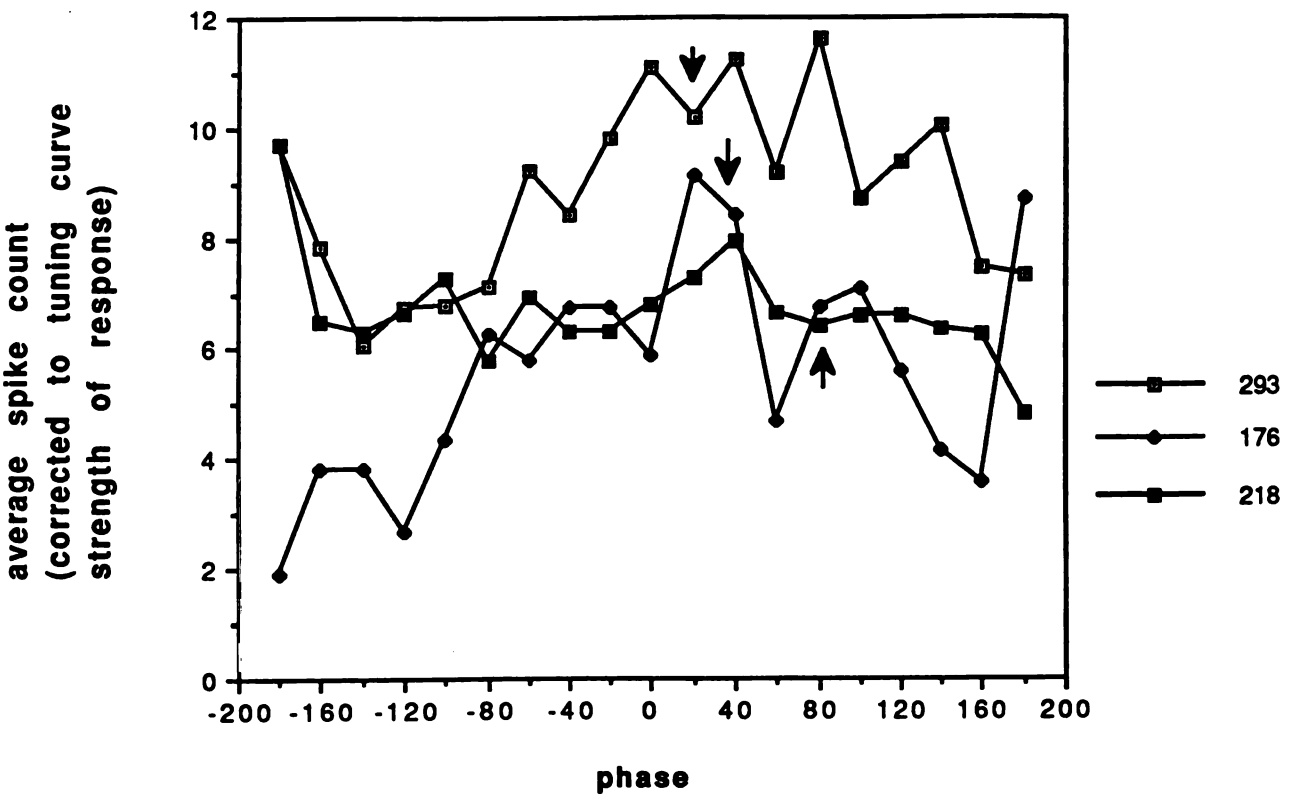


Figure 35. Neural response as a function of phase presentation for the three trained cats. Behavioral thresholds were for #293- 18 deg., #176- 33 deg., and #218- 80 deg., and are indicated by arrows.

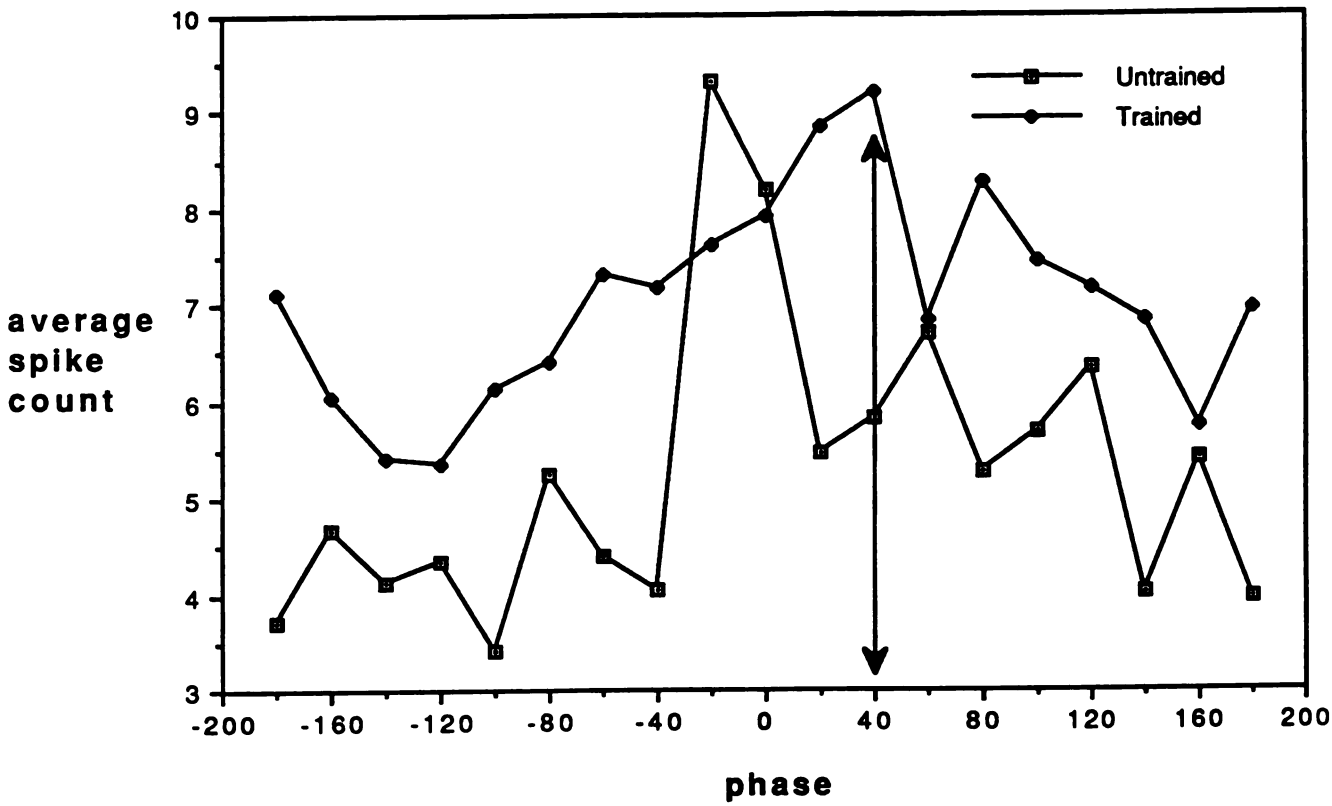


Figure 36. Neural response for four untrained cats plotted against the average neural response for the three trained cats. The maximum spike count value for the untrained cats' average was -20 degrees, closely followed by 0 degrees. The maximum spike count value for the trained cats' average was +40 degrees.

Tuning

The sharpness of tuning of the excitatory response areas of the penetrations for untrained and trained cats were measured and compared. The Q-40 value is calculated by dividing the characteristic frequency (cf) at the particular penetration by the bandwidth 40 dB above the threshold. Q-40 values for only penetrations with cfs below 11 kHz were included in this analysis.

Figure 37 displays the regression of the Q-40 values against the cf at each penetration for the untrained cats. Figure 38 displays the equivalent regression for the trained cats. In these figures, there are 139 values for the three trained cats, and 155 values for seven un-trained cats. In both cases, there is no relationship between the Q-40 value and the cf of the penetration, indicating that the sharpness of tuning is independent of the frequency of the penetration. .

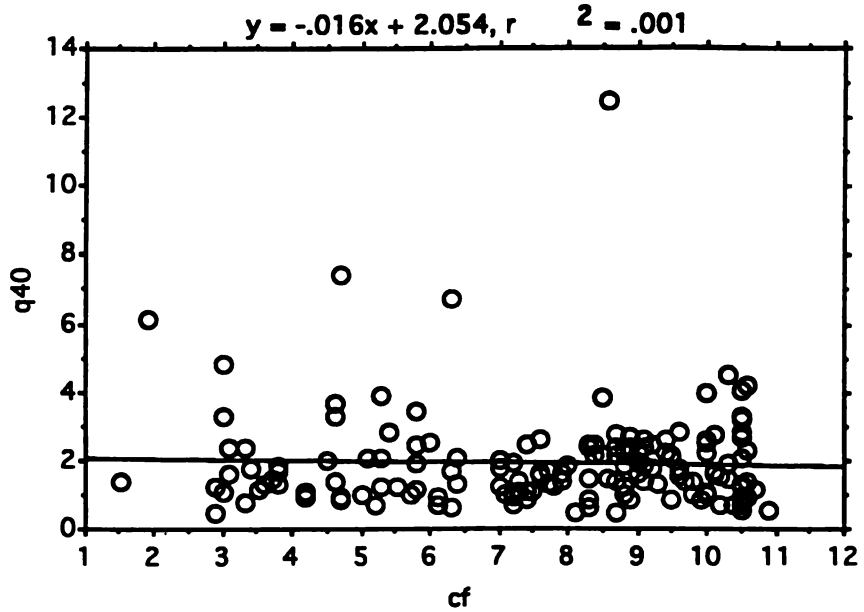


Figure 37. Regression of the Q-40 values against the cf at each penetration for the untrained cats.

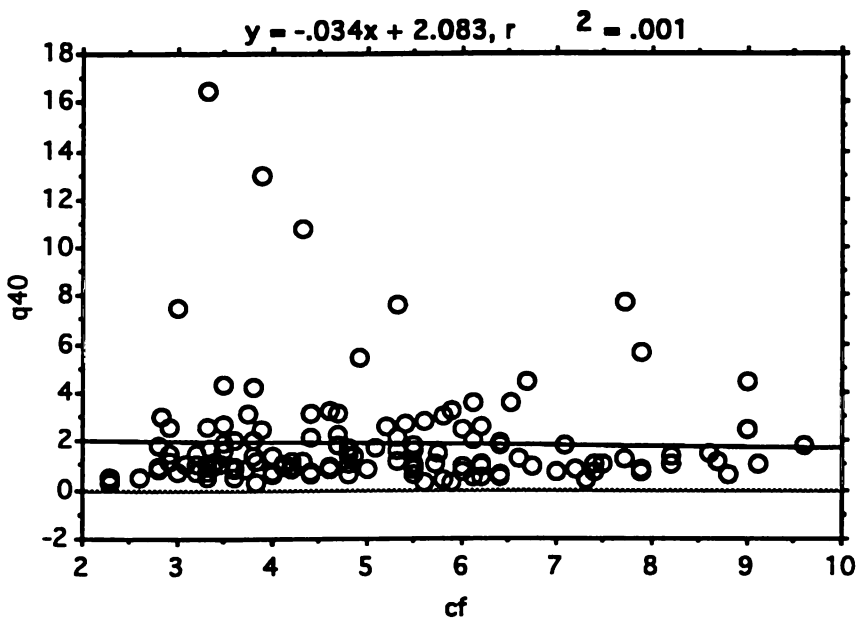


Figure 38. Regression of the Q-40 values against the cf at each penetration for the trained cats.

Figure 39 illustrates the distribution of the same Q-40 values for the three trained cats and for seven untrained cats, including 139 Q-40 values for the three trained cats, and 155 Q-40 values for seven un-trained cats. A Mann-Whitney U two-group un-paired comparison was performed; the two groups of animals, trained and untrained had different Q-40 values at a significance level of p=0.004. The mean Q-40 value for the untrained cats was 1.93, with standard deviation of 1.4; for the trained cats the mean Q-40 value was 1.91, with standard deviation of 2.2.

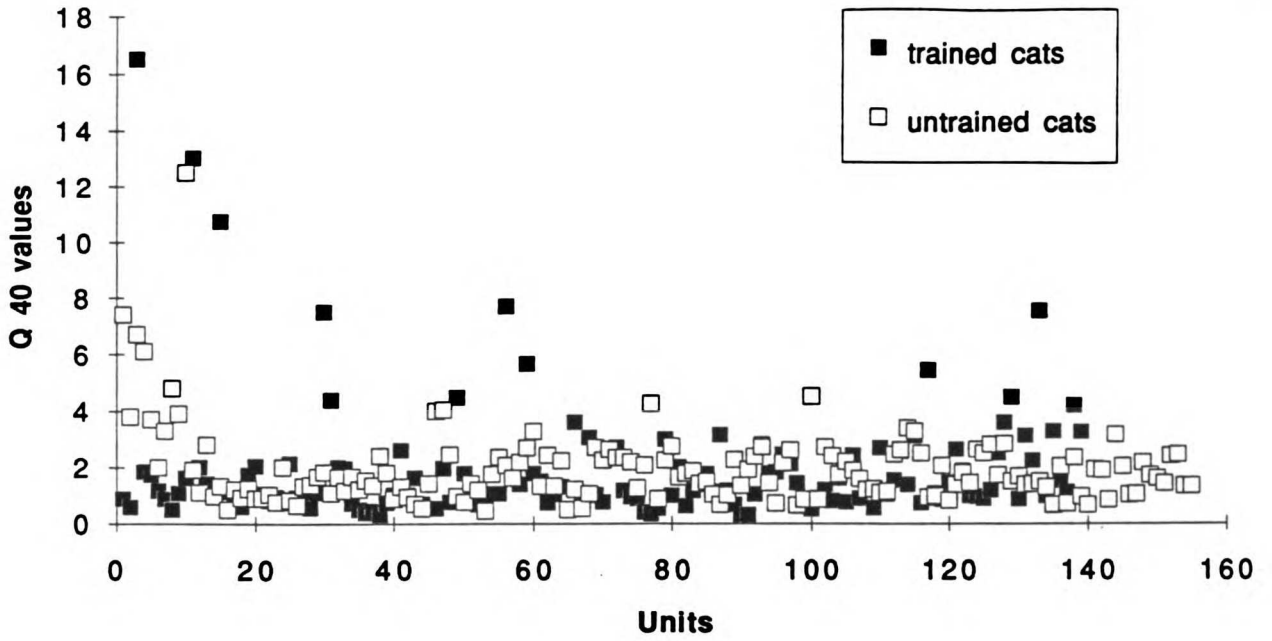


Figure 39. Distribution of Q-40 values for trained and untrained cats.

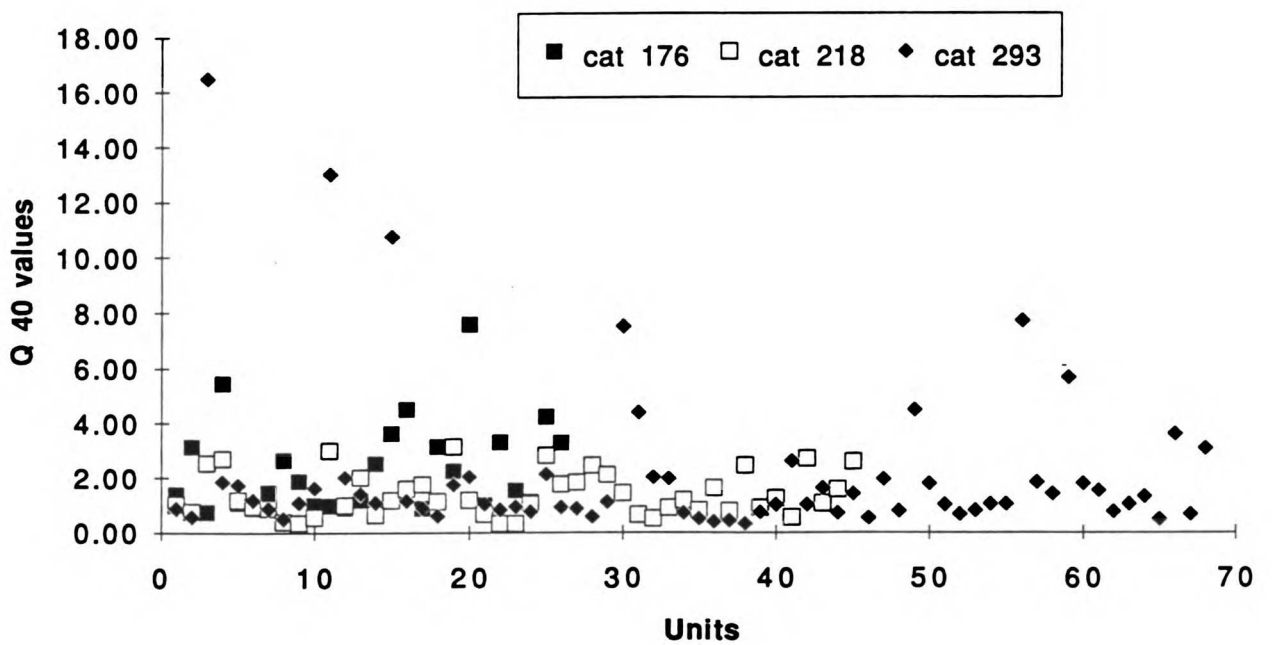


Figure 40. Distribution of Q-40 values for three trained cats.

Figure 40 illustrates the distribution of Q-40 values for the three trained cats. The cat with the best (lowest) ripple envelope phase shift threshold (18 degrees), #293, had the highest Q-40 values; the cat with a threshold of 33 degrees, #176, had the next highest Q-40 values; and the cat with the poorest threshold (80 degrees) had the lowest Q-40 values. The average Q-40 values and the behavioral thresholds for each of the trained cats are shown in the table below.

Cat	Q-40	behavioral threshold (in degrees)
218	1.37	80
293	2.09	18
176	2.36	33

Table 12. Q-40 values and behavioral thresholds for trained cats

The average Q-40 values for the two "exposed" cats, #122 and 40764, i.e. cats which had been exposed to the ripple stimulus for approximately two and four months respectively, but which were unable to learn the single speaker procedure, were 1.75 and 1.61. The correlation between the average Q-40 value for each of the three trained cats and its behavioral threshold was -.88, with an R-squared value of .77. A regression analysis of the trained cats' Q-40 values and their behavioral thresholds is shown in Figure 41. Due to the small sample size, the data do not allow a statistically secure p-value conclusion; however the R-squared value of 0.77 indicates that 77% of the variance in the data is explained by the relationship between the sharpness of tuning and the behavioral phase shift thresholds.

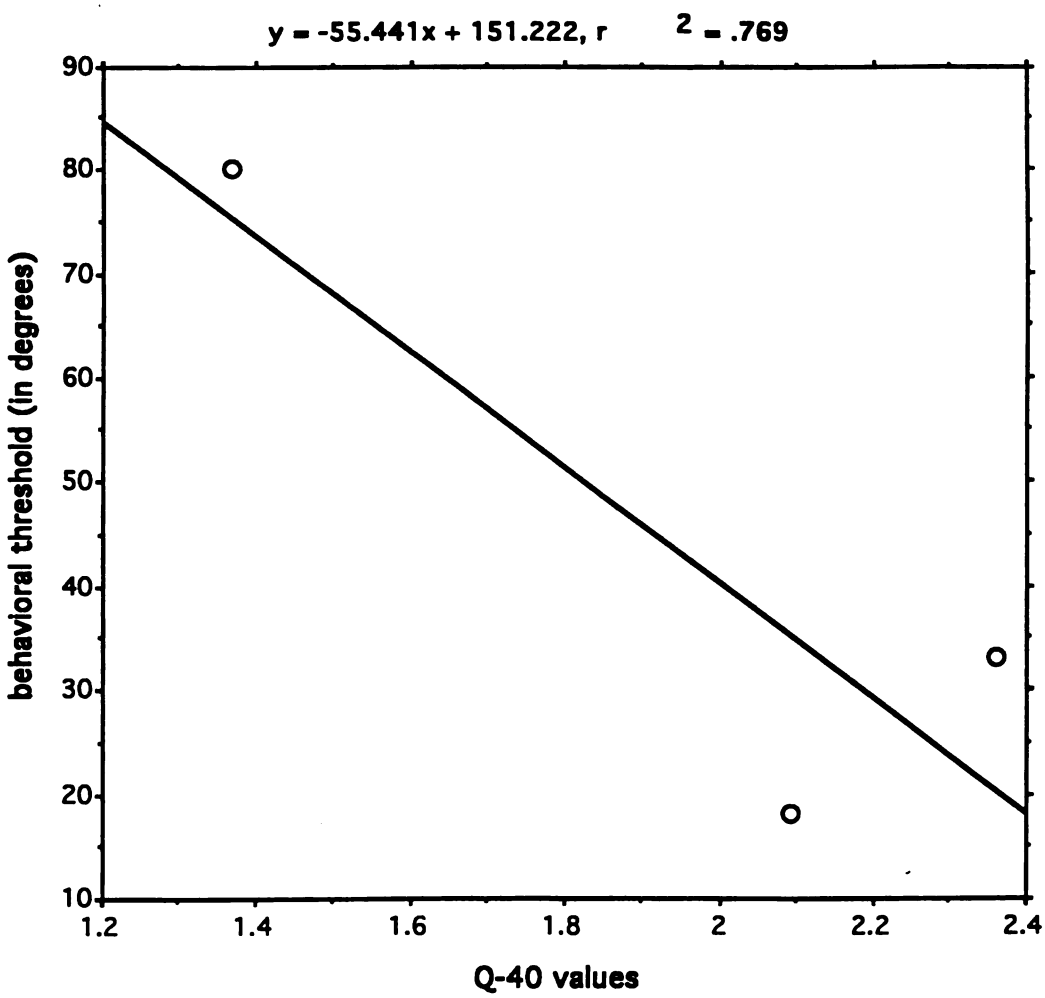


Figure 41. Regression analysis for the trained cats' Q-40 values and their behavioral thresholds.

Schreiner & Mendelson (1990) observed a frequency-independent maximum in sharpness of tuning for both Q-10 and Q-40 values located near the center of the dorsoventral extent of AI. A representative pseudo-three-dimensional projection of the spatial distribution of Q-40 values in AI, taken from their paper, is shown in Figure 42. In order to be certain that the Q-40 values of the trained cats were not penetrations recorded only at the very central portion of AI, three-dimensional representations of the spatial distribution of Q-40 values for the three trained cats were created. They are shown in Figure 43. The highest Q-40 values do not appear to be concentrated in the central portion of AI. This observation lends support to the idea that the training of the animals affected the sharpness of tuning, as measured by the Q-40 values.

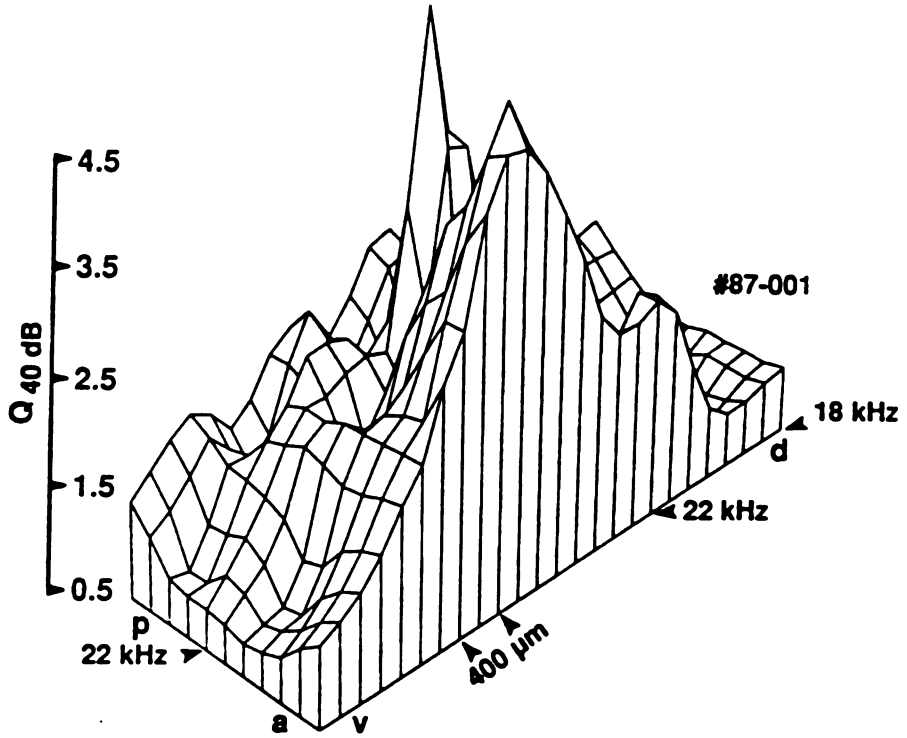


Figure 42. A representative pseudo-three-dimensional projection of the spatial distribution of Q-40 values in AI (from Schreiner & Mendelson, 1990). A maximum is seen in the center of the dorso-ventral extent.

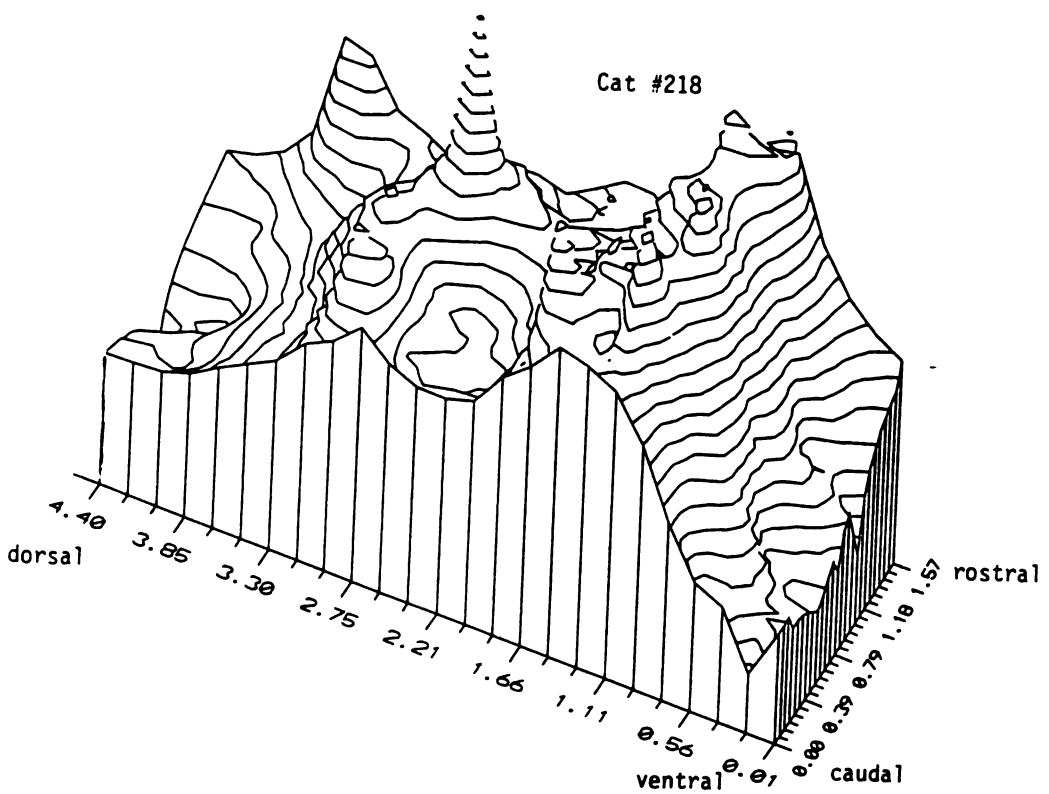


Figure 43 a). Three-dimensional representation of the spatial distribution of Q-40 values for trained cat #218.

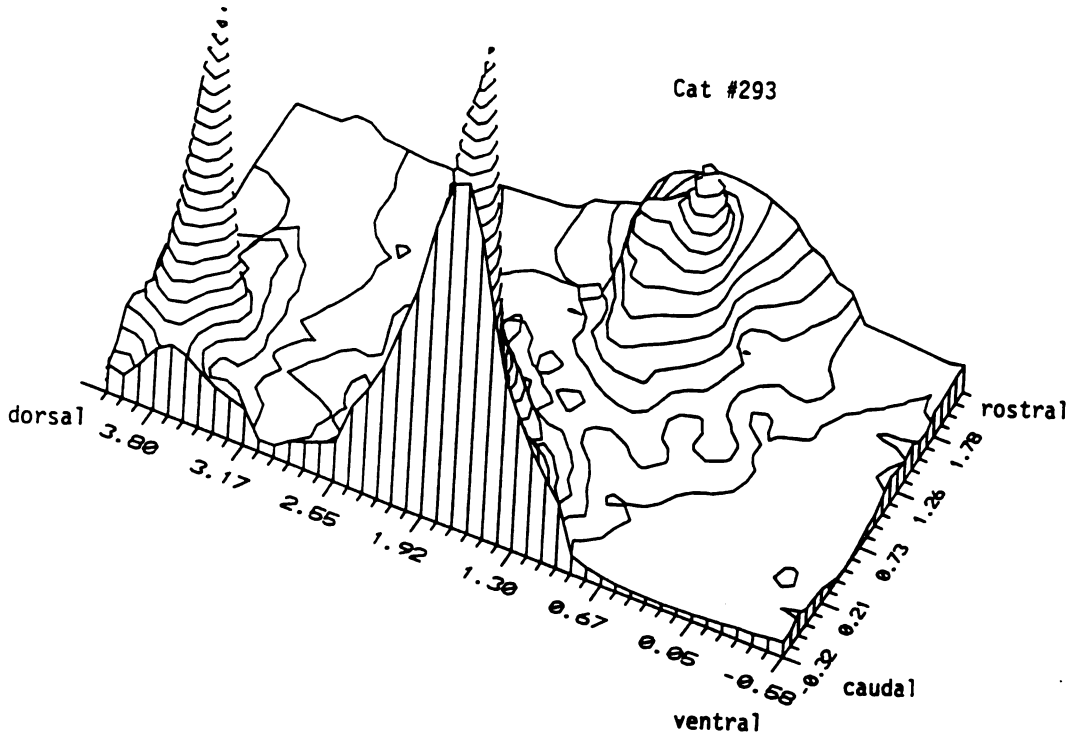


Figure 43 b) Three-dimensional representation of the spatial distribution of Q-40 values for trained cat #293.

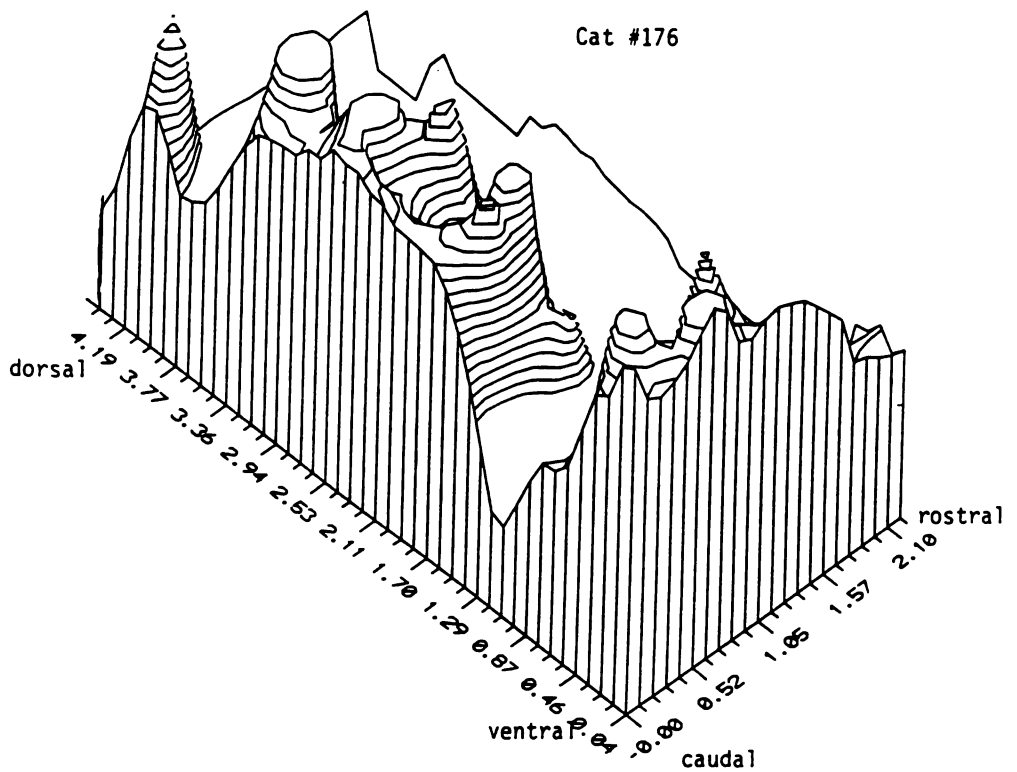


Figure 43 c) Three-dimensional representation of the spatial distribution of Q-40 values for trained cat #176.

DISCUSSION & CONCLUSIONS

The results of the research completed in this dissertation have several implications. The breadth of the subject matter bespeaks the need for an integration of several lines of investigation, to most effectively answer questions about brain function and perception and language. Thus, in the following sections, several issues will be addressed, including: a) how the results bear on language acquisition and language learning, b) what processes may underlie the physiological plasticity observed in these experiments, and c) what general mammalian auditory processing capacities have to tell us about the evolution of linguistic capabilities.

Speech Processing

Humans appear to be most adept at discriminating ripple envelope shifts at spacings of 3.5 or fewer peaks/octave. The spacings of English vowels are mostly of 3 or less peaks/octave (see Figure 44, p.127). Not surprisingly, perhaps, humans are best at discriminating changes within the spacings of English vowels. Since formants are determined by natural resonances of the human vocal tract, the implication arises that the human auditory system has either evolved to be best at discriminating those parameters that are important in speech, and/or through adaptive learning, has developed a formal representation of this most heavily practiced peak-spacing domain..

Flanagan (1955) first provided detailed measures of difference limens (DLs) for vowel formant frequencies. In his studies, for each synthesized vowel three out of four of the formants were held constant, while either F1 or F2 was varied. In the first set of tests, F2 was 1500 Hz, F3 measured 2500 Hz, F4 was 3550 Hz, and F1 was set at 300, 500, or 700

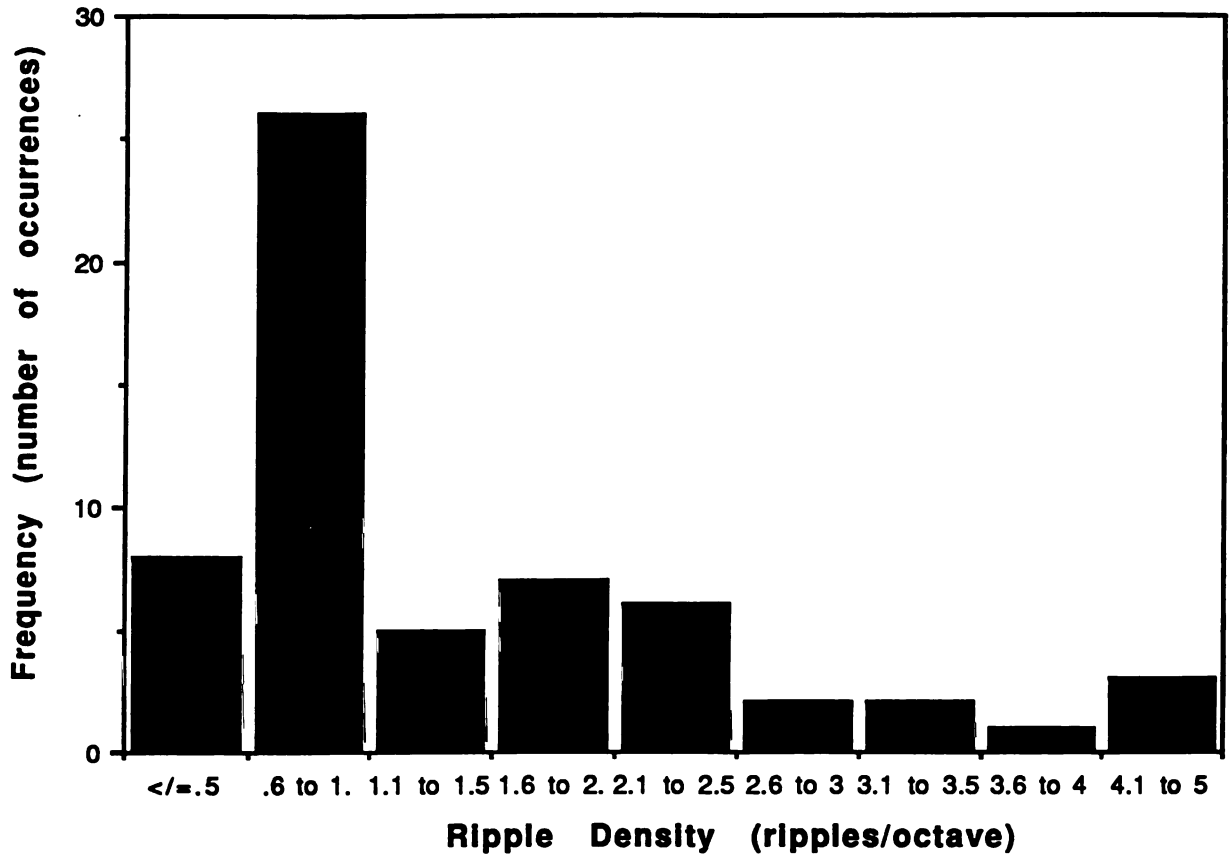


Figure 44. Frequency distribution of ripple density values for 10 English vowels (formant values from Peterson & Barney, 1952). The spacings are mostly of 3 or less peaks/octave.

Hz. Frequency variations used in determining the DLs for F1 were +/-10, +/-20, +/-30, +/-40, +/-50, +/-60, and +/-70 Hz.

In a second set of tests, F2 was set at 1000, 1500, or 2000 Hz; F1 was held constant at 500 Hz; the frequency variations used in determining the DLs for F2 were +/-25, +/-50, +/-75, +/-100, +/-125, +/-150, and +/-175 Hz. Subjects heard two stimuli and were asked to report whether the second sound was the same as or different from the first. The 50% point of stimuli judged different was taken as the DL. The mean just-discernible frequency shift was approximately 3-5% of the formant frequency.

Comparing the Flanagan results to the present ripple study results is complicated by the fact that the peak or formant spacings in synthesized vowels were not identical among all peaks within a vowel, as they were in our ripple study. However, ignoring for the moment the formants above F2, the various spacings between F1 and F2 studied may be considered. In the first set of tests in which F1 was varied, the frequency values of F1 and F2 were i) 300 Hz and 1500 Hz; ii) 500 Hz and 1500 Hz; and iii) 700 Hz and 1500 Hz, respectively. In the second set of tests, in which F2 was varied, the frequency values of F1 and F2 were i) 500 Hz and 1000 Hz; ii) 500 Hz and 1500 Hz; and iii) 500 Hz and 2000 Hz. The equivalent ripple densities are 0.4, 0.6, 0.9, 1, 0.6, and 0.5 ripples or peaks/octave.

Flanagan reported that human subjects could detect a 3 to 5% shift in spectral peak position. That would be equivalent to a 55 Hz shift at the center frequency of 1.38 kHz in the present human ripple envelope phase shift study. The equivalent envelope phase shift for a ripple density of 1 ripple/octave, is 20.3 degrees. For a ripple density of 0.5, the phase shift value equivalent to 55 Hz is 10.1 degrees. As will be recalled from Tables 6 & 7, human phase shift thresholds, at a ripple density of 1 ripple/octave,

were 7 degrees for positive envelope phase shifts and 8.6 degrees for negative envelope phase shifts, averaging 7.8 degrees. The human phase shift thresholds at a ripple density of 0.5 were 7.3 degrees for positive envelope phase shifts, and 13.5 degrees for negative envelope phase shifts, averaging 10.4 degrees. The equivalent frequency DLs were 21 Hz and 54 Hz for ripple densities of 1 and 0.5 ripples/octave. This information is presented in table form below.

	Ripple Density (ripples/octave)	**Average Frequency DL**	**Threshold Phase shift**
Flanagan study	0.5	55 Hz	10.1 degrees
	1.0	55 Hz	20.3 degrees
Ripple study	0.5	54 Hz	10.4 degrees
	1.0	21 Hz	7.8 degrees

Table 13. Comparison of Flanagan DL values and ripple study threshold phase shift values.

Thus the frequency DL values observed by Flanagan of approximately 3-5% of the frequency tested matched the measures of the detectable shifts for 0.5 ripples/octave stimuli, but distributions at 1 ripple/octave were poorer for vowel peak shift than would be expected from the current measures of ripple peak shifts in our human study.

Van Veen & Houtgast (1985) have earlier argued that the depth of modulation plays only a minor role in vowel discrimination. They also observed that the ripple densities that are most important for the perception of differences in spectral sharpness are around 2 ripples/octave. In the current study, it was observed that across four ripple densities (0.5, 1, 2, and 4 ripples/octave), depth of modulation changes affected discrimination

similarly, with depths below 7.5 dB producing a dramatic increase in threshold. It is difficult to compare these results directly with those of van Veen & Houtgast because they used a different filtering method to spectrally sharpen their stimuli. However, in contradistinction to their general conclusions, the present study reveals that for depths of modulation below about 7.5 dB, shifts in frequency peak positions rapidly come to be far more difficult to detect.

Shamma & Vranic (1993) (see p.20) reported that the detection of change in symmetry in spectral shape was more dependent on peak frequency (pitch effect) than was the detection of a change in bandwidth. They also reported that "the listeners required less training and exhibited higher sensitivity (lower thresholds) for the detection of symmetry than bandwidth factor peak changes" (p.47). Vranic, Versnel, & Shamma (1993) also presented both single and double spectral peaks, but did not report a covariance of neural responses of AI neurons in ferret cortex with the number of spectral peaks. An expanded report concerning the effect of single and double peaks on psychophysical discrimination of these complex stimuli would be revealing, and directly comparable to the present study. Their data imply that spectral shape is very informative to the nervous system, and that many parameters of complex stimuli provide discriminatory cues.

In Figure 45 (page 133), the average minimum envelope positive phase shifts detected across the ripple densities tested - 0.5 to 8 ripples/octave - are shown. Also shown are the frequency difference limen (dl) values for 4 different frequency values: 1 kHz, 3 kHz, 6 kHz, and 8 kHz, plotted against ripple density. As for formant ratios, frequency difference limen/frequency ratios can be converted to a logarithmic scale,

and termed "peaks" or "events per octave", or "ripple density". Thus, as can be seen, as ripple density increases, the number of events/octave that the frequency difference limen represents also increases. It should be noted that the slope of the envelope phase shift threshold and the slopes of the frequency difference limen functions are generally similar. This suggests that the discriminative functions involved in frequency discrimination and in ripple envelope shift discrimination may be similar. However as was earlier pointed out , the frequency dl for a frequency of 1.38 kHz (the center frequency of the human ripple stimulus) is approximately 2.5 Hz, while the smallest discriminable phase shift for ripple stimuli observed in this study was equivalent to a frequency value of approximately 4 Hz. Thus, the values do not coincide.

The critical band ratio converted to ripple density is also shown in Figure 44. Its value is approximately 4 ripples/octave, and corresponds to the transition to increased thresholds, suggestive of an interference of the filter (critical band) spacing with the spacing of major spectral peaks in the stimulus. Thus peaks of spacings of less than 4 per octave would span the critical band border, and discriminability of envelope phase shifts within the range from 0.5 to approximately 4 ripples/octave would be good and fairly constant, as it appears to be. Peaks spaced at 4 per octave or more would fall within a single critical band; therefore envelope peak shifts would not be as easily discriminable, and the threshold should increase.

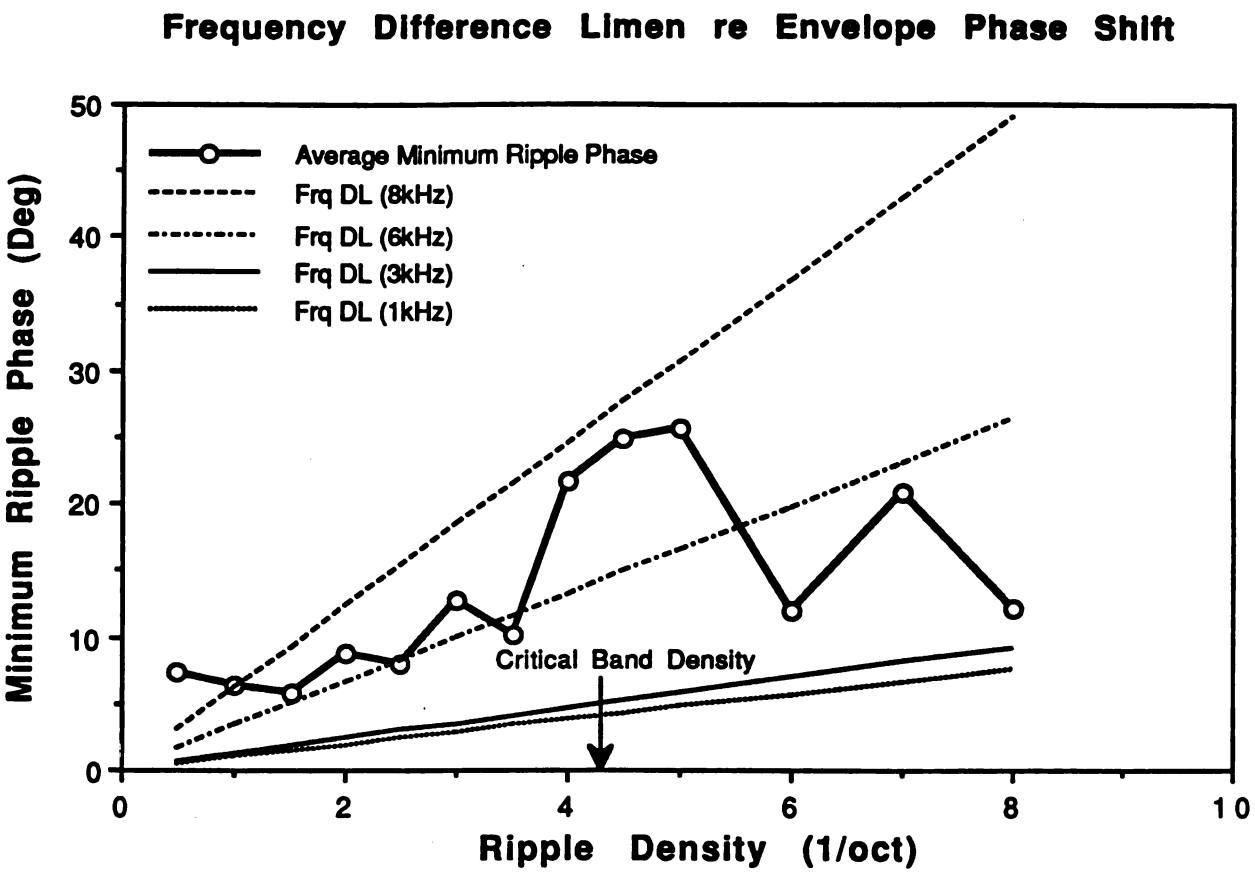


Figure 45. Frequency Difference Limen re Envelope Phase Shift. Human thresholds plotted against ripple density, frequency difference limens, and critical band density.

It can be observed that the minimum envelope phase shift function decreases after a peak at about 5 ripples/octave. There is an increase again at 7 ripples/octave and then another decrease at 8 ripples/octave. All subjects showed a decrease in threshold after a peak around 4 or 5 ripples/octave. Some did not show a second later increase in threshold. Why should the threshold decrease after approximately 5 ripples/octave? As mentioned, some subjects showed a second peak in threshold, while others did not. It may be that there is a second interference with peak spacing discrimination related to a function of the critical band.

Studies by Schreiner et al.(1993) have shown that "untrained" cats have the best neural response to ripple densities between 1 and 2 ripples/octave. This implies that there may be something inherently biologically appealing about the octave scale, in that one event per octave, or one event for every doubling of frequency, facilitates discrimination. With the critical band being approximately one-quarter to one-third octave wide, one might expect that spectral peak spacings of less than 3 or 4 would be fairly easily discriminated, since the peaks fell into different critical bands. It is also the case that most spacings of formants in English vowels are less than 4 peaks per octave. Thus the animal electrophysiological data and the human psychophysical data indicate that formants are located within a range of spacing values that coincide with maximally discriminable perceptual information elements.

Infant studies have shown that the young infant has a very sophisticated ability to perceive the basic sounds of language (Eimas & Tartter, 1979). Infants even below the age of 4 months are able to make discriminations among steady-state vowels. Trehub (1973) showed that [a] could be distinguished from [i] whether in isolation or in the context of a

preceding stop consonant. Infants could also distinguish [i] from [u]. Swoboda, Morse, & Leavitt (1976) showed that the vowel contrast [i] vs [I], as in "beet "vs "bit" was discriminable by infants. Trehub, Endman, & Thorpe (1990) presented complex stimuli that differed in spectral structure to 7 to 8.5 month-old infants. Infants successfully differentiated two spectral structures, even when they varied in fundamental frequency, intensity, or duration. It was concluded that infants classify tonal stimuli on the basis of timbre. Similarly, Clarkson, Clifton, & Perris (1988) presented sounds that differed in their spectral envelopes to 7-month-old infants, who successfully discriminated the sounds. They point out that vowels have multidimensional qualities which distinguish them individually; they argue that timbre is a multidimensional attribute, and point to evidence that shows that "adults' perception of tonal complexes reveals that the spectral envelope is the most important determinant of timbre (Plomp, 1970)" (p.20).

Thus, the ability to process vowel differences based upon changes in spectral envelope seems to be an innate ability in humans. From the animal behavioral data presented in this dissertation, discriminability of changes in spectral envelopes is also present in cats. It may be a basic mammalian auditory processing capacity.

Even though humans seem to have innate speech-processing capabilities, experience does seem to have an effect. Streeter (1976) investigated the discrimination of voicing onset by infants reared in a linguistic environment in which the English voiced/voiceless distinction was not used. The results suggested that "previous exposure to the prevoiced/voiced distinction in the language or perhaps exposure to the acoustic cues underlying this distinction is a prerequisite for discrimination

of the contrast. On the other hand, the voiced/voiceless discrimination can be made in the absence of relevant linguistic exposure, since Kikuyu infants do not hear this voicing distinction used in their language. Thus, the results suggest that there is an interaction between nature and nurture; some phonetic or acoustic discriminations may be universal, whereas others seem to require previous relevant exposure" (p.40).

Kuhl, Williams, Lacerda, Stevens, & Lindblom (1992) observed that exposure to a specific language in the first six months of life alters infants' phonetic perception. Infants in America and Sweden were tested with both native- and foreign-language vowel sounds. "Infants from both countries exhibited a language-specific pattern of phonetic perception" (p.606). As pointed out, this perceptual organizing occurs well before the age at which children begin to acquire word meanings, i.e. well before the acquisition of language. It was suggested that "linguistic experience shrinks the perceptual distance around a native-language prototype, in relation to a nonprototype, causing the prototype to perceptually assimilate similar sounds" (p.608). Thus phonetic prototypes were conjectured to be "fundamental perceptual-cognitive building blocks" (p.608).

The data in this dissertation support the idea that certain basic auditory processing capabilities are operative in humans and in other mammals as well, specifically at least in cats, and that exposure to or experience with the stimuli refines the processing capability and the cortical representation of that stimulus. In the same way that a perceptual organization was set up in the brains of the six-month-old infants, so categorization and learning occur during the acquisition of a first or second language. Learning occurred in these animals as they performed the auditory discrimination, and their brain representations of these

behaviorally important stimuli were changed in parallel. It seems plausible that, at birth, certain basic auditory perceptual capabilities are present, and that with linguistic exposure, certain contrasts are sharpened and certain other unused phonetic contrasts drop out of the discriminative repertoire. The cortical organization thus dynamically represents those stimuli to which the human/animal has been exposed and which are informationally and biologically relevant to that creature.

Cortical Representation

Ripple densities were represented differently in the auditory cortices of the trained cats as compared to untrained cats. The peak of the ripple transfer function was shifted slightly toward higher ripple densities (1 ripple/octave compared to 0.33). The best ripple density peak response for untrained animals was somewhat lower than the mean best ripple density reported in Schreiner et al. (1993): 0.33 as compared to 1 to 2 ripples/octave. That may be partly due to the fact that most of the Schreiner et al. (1993) data were collected as single units, whereas the majority of the data reported for the untrained cats in the present study was collected as the response of small clusters of units (2-6 neurons), which may be less precise.

In hindsight, it would have been more telling to have trained the animals on a ripple density other than 1 ripple/octave, as untrained cats already show their best response to ripple densities in the range of about .33 to approximately 1 ripple/octave. At the same time, preliminary data (Kruger) indicates that it may be more difficult for cats to learn the discrimination at higher ripple densities, and it may have been more

difficult or impossible to obtain a behavioral threshold at substantially higher ripple densities.

The bandwidth 50% down from the peak was wider in trained cats than in untrained cats (see Figure 29, p.98). A broadening of the filter function implies that a larger number of spatial frequencies (ripple densities) were processed more efficiently by the trained cats as compared to the untrained cats. Since the trained cats were only trained on a ripple density of 1 ripple/octave, this increased processing efficiency was not due to a direct learning effect. What may have happened was an improvement in peak spacing discrimination that generalized somewhat so that spacings close to 1 ripple/octave, though not challenged, were affected in a facilitory manner.

Q-40 values over all penetrations were higher in trained cats as compared to untrained cats, indicating a general sharpening of frequency response tuning. The behavioral threshold was correlated (-.87) with Q-40 values, indicating that the sharpening in tuning was related to the proficiency of auditory discrimination. Recanzone et al. (1993) reported a sharpening of tuning for behaviorally relevant frequencies in monkeys trained to discriminate small differences in the frequencies of sequentially presented tonal stimuli. In this study (Keeling), there were not enough penetrations with cf at the center frequency of the behavioral stimulus to compare the tuning there to the tuning at other frequency locations, nor would we expect a simple frequency contrast to code this discrimination. The fact that the overall tuning was sharpened may indicate a more global effect for this complex stimulus.

In the Recanzone study, the sharpness of tuning was not correlated with behavioral performance; however the cortical representation of the

behaviorally relevant frequencies was increased and its area was correlated with behavioral performance. In the current study, a best ripple density response at each penetration was measured, and an increase in representation of best ripple density of 1 ripple/octave was not consistently seen in all trained animals as compared to untrained animals. However, although there is some indication of a spatial organization for ripple density representation orthogonal to the frequency axis (Schreiner et al., 1993), this representation is not presently as circumscribed as is the representation for pure tones. A larger n of trained animals might reveal a reliable change in this representation.

Analysis of the cortical neural responses to envelope phase shifts revealed that 1) the training of the cats caused an interactive effect whereby the phase shift presented affected the neural response, but only at certain frequency representation locations, and 2) the maximum spike count value for the untrained cats' average was at -20 degrees phase shift, closely followed by 0 degrees, whereas the maximum spike count value for the trained cats' average was at +40 degrees phase shift. Generally, it is expected that a neuron will respond best to a ripple stimulus centered at the cf of the neuron when the phase is 0 degrees. This is because at that phase, the center frequency of the ripple coincides with the cf of the unit - that is, it is located in the center of the excitatory area of the frequency response curve. All three trained cats, however, showed a best neural response at a phase shift of other than 0 degrees. Cat #293 showed a best response at a phase shift of +80 degrees, cat #176 showed a best response at a phase shift of +20 degrees, and cat #218 showed a best response at a phase shift of -180 degrees. The phase shifts that the cats were exposed to and were trained to discriminate were in the range of 0 to +180 degrees. Their

behavioral thresholds were: cat #293 - 18 degrees, cat #176 - 33 degrees, and cat #218 - 80 degrees. The correlation between phase shift best response (considering -180 degrees as +180 degrees) and behavioral threshold was +.82, with an R-squared value of .67.

These data indicate that neural responses to envelope phase shifts of a ripple stimulus were changed as a result of behavioral training. It appears that neural responses were increased to phase shifts within the range of the trained cats' behavioral thresholds, except for cat #218, whose performance was much less adept than was that of the other two trained cats. The best response to envelope phase shifts did not coincide directly with these cats' behavioral thresholds, indicating that perhaps the changes occurred at a finer-grain level, for example, only at frequencies of 3.675 kHz (the center frequency of the behavioral stimulus). A larger sample of penetrations, particularly around the cf of 3 to 4 kHz in trained and untrained cats, should be able to answer this question. At this time, the specificity of the changes is unclear; it can only be observed that the average best response to envelope phase shifts did change, and did move to a higher location within the function.

Edeline & Weinberger (1993) determined frequency response receptive fields of neurons in the guinea pig auditory cortex before and after two-tone discrimination training. Responses to the frequency of the CS+ were increased, whereas responses to the frequency of the CS- and to the best frequency were reduced. In many cases, tuning was shifted such that the frequency of the CS+ became the new best frequency. This frequency-specific plasticity developed whether the animals discriminated the CS+ or not. That is, in both an easy and a difficult task, animals showed frequency-specific plasticity at the cortical level, even though they

failed to discriminate the CS+ (as measured by cardiac deceleration) in the difficult task. These authors suggested that "the auditory cortex is able to analyze, recode, and retain precise information about a significant stimulus independent of the expression of a behavioral response" and that "the [receptive field] plasticity observed in the auditory cortex is a support for memory" (p.102).

The data from this dissertation similarly suggest that a retuning or reorganization occurred, since the ripple density representation, the sharpness of tuning, and the representation of envelope phase shift were changed in the auditory cortices of the trained cats. However, since the stimulus is a complex harmonic whose representation in untrained animals is not a simple tonotopic sequence, it is not a straightforward assignment to describe learning-induced changes. The disadvantage of using a pure tone as a stimulus, i.e. its lack of relevance and generalizability to real-world acoustic events, is made up for, at least to some extent, by the fact that the interpretation of neural responses to this stimulus is fairly uncomplicated. Evidence suggests, however, that spectral envelope information is coded by a combination of frequency- and integrated excitatory bandwidth-responsive neurons. Since a sharpening of tuning and an interactive effect of envelope phase shift and frequency location were observed in these trained animals, the assumption of a combination of feature-selective responses or a pattern of neural responses coding this vowel-like percept is supported. The changes following training indicate a fine-tuning, an increased efficiency, and a sharpening of the precision of resolution of the relevant parameters of this discrimination.

Evolution of linguistic abilities

Clearly humans are more "intelligent" than cats. If in doubt, try training a cat on an operant discrimination procedure. Cats were able to learn the paradigm, however, and showed quite fine resolution abilities (at least two out of three cats did), but there was variability in performance, and two other cats were unable to learn the discrimination procedure. Thresholds for the two adept cats were about 3 - 4 times higher than humans for the equivalent ripple density. Their intelligence as well as their perceptual discrimination abilities for detecting envelope phase shifts are at a lower level but are on the same continuum as humans.

Watson (1991) has recently reported significant relationships between simple sensory, cognitive, and motor abilities and psychometric intelligence. She reminds us that Spearman (1904) and others reported relationships between pitch discrimination and intelligence. The suggestion is that working memory is a limited-capacity system, and that therefore individuals who can process information rapidly will have more working memory available to process new information. Thus, the sensory ability to perceive differences and make discriminations can be considered part of a processing capacity. Learning calls upon this information-processing ability.

The innate abilities with which humans are born interact with experiential input, especially during development but later as well, to produce a cerebral organization which reflects that interaction. The huge amount of neural modification that occurs within the brain during development allows us to "make sense of the world". As Edelman (1987) says: "it is difficult to imagine the world as it is presented to a newborn organism...the environment presented [to such an organism] is inherently

ambiguous; even to animals eventually capable of speech such as ourselves, the world is initially an unlabeled place" (p.3). Adaptive behavior requires initial categorization of salient aspects of the environment so that learning can occur (p.4). Thus, perceptual abilities are very important in order to form concepts. All animals are born with certain basic sensory and discriminative capacities; all must interact successfully with their environment in order to survive. This capacity for learning and for change has allowed Homo sapiens to evolve. In turn, Homo sapiens has developed speech and language, culture and community. Our capacity for speech has advanced our development of thought, which has changed the world. These faculties have evolved over many eons of time due to the mechanism of neural adaptation. How this cortical change occurs is unknown. Evidence that it occurs is amazing in itself.

REFERENCES

Abbs, J.H., & Sussman, H.M. Neurophysiological feature detectors and speech perception: a discussion of theoretical implications. *Journal of speech and hearing research,* 1971, 14, 23-36.

Abramson, A., & Lisker, L. Discriminability along the voicing continuum: cross language tests. *Proceedings of the Sixth International Congress of Phonetic Science, Prague, 1967.* Prague: Academia, 1970.

Allard, T., Clark, S.A., Jenkins, W.M., & Merzenich, M.M. Reorganization of somatosensory area 3b representations in adult owl monkeys after digital syndactyly. Journal of neurophsyiology, 1992, Sep., 66,3, 1048-1058.

Baru, A.V. Discrimination of synthesized vowels /a/ and /i/ with varying parameters (fundamental frequency, intensity, duration and number of formants) in dog. In G. Fant & M.A.A. Tatham (Eds.), *Auditory analysis and perception of speech.* London: Academic Press, 1975.

Borden, G.J., & Harris, K.S. *Speech Science Primer.* Baltimore, MD: Williams & Wilkins, 1984.

Brown, K.A., Buchwald, J.S., Johnson, J.R., & Mikolich, D.J. Vocalization in the cat and kitten. *Developmental Psychobiology*, 1978, 11, 6, 559-570.

Cheney, D.L. & Seyfarth, R.M. Assessment of meaning and the detection of unreliable signals by vervet monkeys. *Animal behavior,* 1988, 36, 477-86.

Chistovich, L.A. Central auditory processing of peripheral vowel spectra. *Journal of the Acoustical Society of America*, 1985, March, 77(3), 789-805.

Chistovich, L.A., and Lublinskaya, V.V. The 'center of gravity' effect in vowel spectra and critical distance between the formants: Psychoacoustical study of the perception of vowel-like stimuli. *Hearing Research*, 1979, 1, 185-195.

Chomsky, N. *Language and Mind,* (2nd ed.). New York: Harcourt Brace Janovich, 1972.

Clark, S.A., Allard, T., Jenkins, W.M., & Merzenich, M.M. Receptive fields in the body-surface map in adult cortex defined by temporally correlated inputs. *Nature,* 1988, 332, 31 March, 444-5.

Clarkson, M.G., Clifton, R.K, & Perris, E.e. Infant timbre perception: discrimination of spectral envelopes. *Perception & Psychophysics*, 1988, 43, 15-20.

Cutting, J.E., & Rosner, B.S. Categories and boundaries in speech and music. *Perception and psychophysics,* 1974, 16, 564-71.

Davis, H. & Silverman, S.R. *Hearing and Deafness,* (4th ed.). New York: Holt, Rinehart, & Winston, 1978.

Dewson, J.H. Speech sounds discrimination by cats. *Science,* 1964, 144, 555-6.

Disterhoft, J.F. & Stuart, D.K. Differentiated short latency response increases after conditioning in inferior colliculus neurons of alert rat. *Brain research,* 1977, 130, 315-33.

Edeline, J.-M., & Weinberger, N.M. Receptive field plasticity in the auditory cortex during frequency discrimination training: selective retuning independent of task difficulty. *Behavioral Neuroscience*, 1993, 107, 1, 82-103.

Edelman, G.M. *Neural Darwinism - the theory of neuronal group selection.* New York: Basic Books, 1987.

Eilers, R.E., & Minifie, F.D. Fricative discrimination in early infancy. *Journal of speech and hearing research*, 1975, 18, 158-67.

Eilers, R.E., Wilson, W.R., & Moore, J.M. Developmental changes in speech discrimination in infants. *Journal of speech and hearing research*, 1977, 20, 766-80.

Eimas, P.D. Auditory and phonetic coding of the cues for speech: discrimination of the (r-l) distinction by young infants. *Perception and psychophysics*, 1975, 18, 341-7.

Eimas, P.D., & Corbit, J.D. Selective adaptation of linguistic feature detectors. *Cognitive psychology*, 1973, 4, 99-109.

Eimas, P.D., Siqueland, E.R., Jusczyk, P., & Vigorito, J. Speech perception in infants. *Science,* 1971, 171, 303-6.

Eimas, P.D. & Tartter, V.C. On the development of speech perception: mechanisms and analogies. In *Advances in Child Development and Behavior*, H.W. Reese & L.P. Lipsitt, (Eds.), New York: Academic Press, 1979.

Fay, R.R. *Hearing in Vertebrates: a psychophysics databook.* Winnetka, Illinois: Hill-Fay Assoc., 1988.

Flanagan, J.L. A difference limen for vowel formant frequency. *Journal of the Acoustical Society of America*, 1955, May, 27, 3, 613-617.

Fry, D., Abramson, A., Eimas, P., & Liberman, A.M. The identification and discrimination of synthetic vowels. *Lang. Speech,* 1962, 5, 171-189.

Gardner, R.A., & Gardner, B.T. Teaching sign language to a chimpanzee. *Science*, 1969, 165, 664-72.

Glasberg, B.R. & Moore, B.C.J. Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, 1990, 47, 103-138.

Green, D.M. Stimulus selection in adaptive psychophysical procedures. *Journal of the Acoustical Society of America*, 1990, June, 87, 6, 2662-2674.

Greenwood, D.G. Critical bandwidth in man and some other species in relation to the traveling wave envelope. In H.R. Moskowitz et al. (eds.) *Sensation and Measurement.* Dordrecht-Holland: D. Reidel, 1974, 231-239.

Harnad, S. *Categorical perception - The groundwork of cognition.* Cambridge: Cambridge University Press, 1987.

Harris, K.S. Physiological aspects of articulatory behavior. In T. Sebeok (Ed.), *Current trends in linguistics, Vol.12.* The Hague: Mouton, 1974.

Harrison, R.V., Nagasawa, A., Smith, D.W., Stanton, S., & Mount, R.J. Reorganization of auditory cortex after neonatal high frequency cochlear hearing loss. *Hearing Research*, 1991, 54, 11-19.

Heffner, R.S., & Heffner, H.E. Hearing range of the domestic cat. *Hearing research*, 1985, 19, 85-8.

Heil, P. & Scheich, H. Functional organization of the avian auditory cortex analogue. I. Topographic representation of isointensity bandwidth. *Brain Research*, 1991, Jan.18, 539, 1, 110-120.

Hillier, D.A. *Auditory Processing of Sinusoidal Spectral Envelopes.* Doctoral Dissertation. Saint Louis, Missouri: Washington University, Sever Institute of Technology. 1991.

Hubel, D.H., & Wiesel, T.N. Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex. *Journal of Physiology*, 1962, 160, 106-54.

Imig, T.J., & Adrian, H.O. Binaural columns in the primary field (AI) of cat auditory cortex. *Brain research*, 1978, 138, 241-57.

Imig, T.J., & Brugge, J.F. Sources and terminations of callosal axons related to binaural and frequency maps in primary cortex of the cat. *Journal of comparative neurology*, 1978, 182, 637-60.

Jacobson, R., Fant, G.M., & Halle, M. *Preliminaries to speech analaysis.* Cambridge, Mass.: MIT Press, 1963.

Jenkins, W.M. & Merzenich, M.M. Reorganization of neocortical representations after brain injury: A neurophysiological model of the bases of recovery from stroke. *Prog. Brain Research*, 1987, 71, 249.

Jenkins, W.M. & Merzenich, M.M. Role of cat primary auditory cortex for sound-localization behavior. *Journal of neurophysiology*, 1984, 52, 5, 819-47.

Jenkins, W.M., Merzenich, M.M., Ochs, M.T., Allard, T., & Guic-Robles, E. Functional reorganization of primary somatosensory cortex in adult owl monkeys after behaviorally controlled tactile stimulation. *Journal of neurophysiology*, 1990, Jan, 63, 1, 82-104.

Jusczyk, P.W. Perception of syllable-final stop consonants by 2-month-old infants. *Perception & psychophysics*, 1977, 21, 450-4.

Kitzes, L.M., Farley, G.R., & Starr, A. Modulation of auditory cortex unit activity during the performance of a conditioned response. *Experimental neurology*, 1978, 62, 678-97.

Klatt, D. Software for a cascade/parallel formant synthesizer. *Journal of the acoustical society of America*, 1980, 67, 971-95.

Kluender, K.R., Diehl, R.L., & Killeen, P.R. Japanese quail can learn phonetic categories. *Science,* 1987, 4 Sept., 237, 1195-7.

Knight, P.L. Representation of the cochlea within the anterior auditory field (AAF) of the cat. *Brain research*, 1977, 130, 447-67.

Kuhl, P.K. Discrimination of speech by nonhuman animals: basic auditory sensitivities conducive to the perception of speech-sound categories. *Journal of the acoustical society of America,* 1981, 70, 340-9.

Kuhl, P.E. Speech perception in early infancy: perceptual constancy for spectrally dissimilar vowel categories. *Journal of the Acoustical Society of America*, 1979, Dec., 66, 6, 1668-1679.

Kuhl, P.K The special-mechanisms debate in speech research: Categorization tests on animals and infants. In *Categorical perception - the groundwork of cognition* , S. Harnad (Ed.). Cambridge: Cambridge University Press, 1987.

Kuhl, P.K., & Miller, J.D. Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli. *Journal of the acoustical society of America,* 1978, 63, 905-17.

Kuhl, P.K., & Miller, J.D. Speech perception by the chinchilla: voiced-voiceless distinction in alveolar plosive consonants. *Science*, 1975, 190, 69-72.

Kuhl, P.K., & Padden, D.M. Enhanced discriminability at the phonetic boundaries for the voicing feature in macaques. *Perception and psychophysics,* 1982, 32, 542-50.

Kuhl, P.K., & Padden, D.M. Enhanced discriminability at the phonetic boundaries for the place feature in macaques. *Journal of the acoustical society of America,* 1983, 73, 1003-10.

Kuhl, P.K. Williams, K.A., Lacerda, F., Stevens, K.N., & Lindblom, B. Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, 1992, Jan.31, 255, 606-608.

Langner, G. Bonke, D., & Scheich, H. Neuronal discrimination of natural and synthetic vowels in field L of trained mynah birds. *Exp. Brain Reseach*, 1981, 43, 11-24.

Lasky, R.E., Syrdal-Lasky, A., & Klein, R.E. VOT discrimination by four and six and a half month old infants from Spanish environments. *Journal of experimental child psychology*, 1975, 20, 215-25.

Lass, N.J. *Handbook of speech-language, pathology, and audiology.* Toronto: B.C. Decker, 1988.

Lenneberg, E.H. Language and Brain: developmental aspects. *Neurosciences research program bulletin,* 1974, 12, 4.

Liberman, A.M. The grammars of speech and language. *Cognitive psychology*, 1970, 1, 301-23.

Liberman, A.M., Cooper, F.S., Shankweiler, D.S., & Studdert-Kennedy, M. Perception of the speech code. *Psychological Review*, 1967, 74, 431-61.

Liberman, A.M., Harris, K.S., Hoffman, H.S., & Griffith, B.C. The discrimination of speech sounds within and across phoneme boundaries. *Journal of experimental psychology*, 1957, 54, 358-68.

Liberman, A.M., Harris, K.S., Kinney, J.H., & Lane, H.L. The discrimination of relative onset time of the components of certain speech and nonspeech patterns. *Journal of experimental psychology,* 1961, 61, 379-88.

Lieberman, P. *The biology and evolution of language.* Cambridge: Harvard University Press, 1984.

Lloyd, R.J. *Some Researches into the Nature of Vowel-Sound.* Liverpool, England: Turner and Dunnett, 1890.

Mendelson, J.R., Schreiner, C.E., Grasse, K., & Sutter, M. Spatial distribution of responses to FM sweeps in cat primary auditory cortex. *Association for Research in Otolaryngology Abstract*, 1988, 11, 36.

Mendelson, J.R., & Grasse, K.L. A comparison of monaural and binaural responses to frequency modulated (FM) sweeps in cat primary auditor cortex. *Experimental Brain Research*, 1992, 92, 1, 435-454.

Merzenich, M.M., Andersen, R.A., & Middlebrooks, J.H. Functional and topographic organization of the auditory cortex. In *Hearing mechanisms and speech, Experimental brain research, suppl. II,* O. Creutzfeld, H. Scheich, & Chr. Schreiner (Eds.). Berlin: Springer-Verlag, 1979.

Merzenich, M.M., Knight, P.L., & Roth, G.L. Representation of cochlea within primary auditory cortex in the cat. *Journal of neurophysiology,* 1975, 38, 231-49.

Merzenich, M.M., Roth, R.A., Knight, P.L., & Colwell, S.A. Some basic features of organization of the central auditory system. In *Psychophysics and physiology of hearing*, E.F. Evans & J.P. Wilson (Eds.). London; Academic Press, 1977.

Middlebrooks, J.C., Dykes, R.W., & Merzenich, M.M. Binaural response-specific bands in primary auditory cortex (AI) of the cat: topographical organization orthogonal to isofrequency contours. *Brain research*, 1980, 181, 31-48.

Miller, J.D. Audibility curve of the chinchilla. *Journal of the acoustical society of America,* 1970, 48, 2, 2, 513-23.

Miller, J.D. Auditory-perceptual interpretation of the vowel. *Journal of the Acoustical Society of America*, 1989, May, 85(5), 2114-2133.

Miller, J.D., Wier, C.C., Pastore, R.E., Kelly, W.J., & Dooling, R.J. Discrimination and labeling of noise-buzz sequences with varying noise-lead times: an example of categorical perception. *Journal of the acoustical society of America,* 1976, 60, 410-7.

Morse, P.A., & Snowdon, C.T. An investigation of categorical speech discrimination by rhesus monkeys. *Perception and psychophysics*, 1975, 17, 1, 9-16.

Nearey, T.M. Static, dynamic, and relational properties in vowel perception. *Journal of the Acoustical Society of America*, 1989, May, 85(5), 2088-2113.

Newman, J.D., & Wollberg, Z. Multiple coding of species-specific vocalizations in the auditory cortex of squirrel monkeys. *Brain research*, 1973, 54, 287-304.

O'Malley, M.H., & Kloker, D.R. Computer aided instruction in Speech Science. *Today's speech*, 1972, 17-24.

Pastore, R.E., Li, X.F., & Layer, J.K. Categorical perception of nonspeech chirps and bleats. *Perception and Psychophysics*, 1990, Aug, 48, 2, 151-156.

Patterson, F. The gestures of a gorilla: Language acquisition in another Pongid. *Brain and Language*, 1978a, 5, 72-97.

Patterson, F. Conversations with a gorilla. *National Geographic*, 1978b, 154, 438-65.

Peterson, G.E., & Barney, H.L. Control methods used in a study of vowels. *Journal of the Acoustical Society of America*, 1952, 24, 175.

Petitto, L.A., & Marentette, P.F. Babbling in the manual mode: evidence for the ontogeny of language. Science, 1991, 251, 1493-1496.

Pickles, J.O. *An introduction to the physiology of hearing*. London: Academic Press, 1982.

Pickles, J.O. Normal critical bands in the cat. *Acta Otolaryngol.*, 1975, 80, 245-254.

Pisoni, D.B. Identification and discrimination of the relative onset time of two-component tones: implications for voicing perception in stops. *Journal of the acoustical society of America*, 1977, 61, 1352-61.

Plomp, R. Timbre as a multidimensional attribute of complex tones. In R. Plomp & G.F. Smoorenburg (Eds.), *Frequency analysis and periodicity detection in hearing*. Leiden: Sijthoff, 1970.

Premack, D. Language in chimpanzees. *Science*, 1971, 172, 808-22.

Recanzone, G.H., Jenkins, W.M., Hradek, G. T., & Merzenich, M.M. Progressive improvement in discriminative abilities in adult owl monkeys performing a tactile frequency discrimination task. *Journal of Neurophysiology*, 1992, May, 67, 5, 1015-30.

Recanzone, G.H., Merzenich, M.M., Jenkins, W.M., Grajski, K.A., & Dinse, H.R. Topographic reorganization of the hand representation in cortical area 3b of owl monkeys trained in a frequency discrimination task. *Journal of Neurophysiology*, 1992, 67, 1031-56.

Recanzone, G.H., Schreiner, C.E., & Merzenich, M.M. Plasticity in the frequency representaton of primary auditory cortex following discrimination training in adult owl monkeys. *Journal of Neuroscience*, 1993, Jan. 13, 1, 87-103.

Reale, R.A., & Imig, T.J. An orderly frequency representation in the posterior ectosylvian sulcus of the cat. *Neuroscience Abstr*, 1977, 3, 10.

Robertson, D. & Irvine, D.R.F. Plasticity of frequency organizaton in auditory cortex of guinea pigs with partial unilateral deafness. *Journal of Comparative Neurology*, 1989, 282, 456-71.

Rose, J.E. The cellular structure of the auditory region of the cat. *Journal of comparative neurology*, 1949, 91, 409-39.

Rossing, T. D. *The Science of Sound.* Reading, Massachusetts: Addison-Wesley Publishing Co., 1982.

Rumbaugh, D.M., Gill, T.V., & von Glaserfield, E.C. Reading and sentence completion by a chimpanzee (Pan). *Science,* 1973, 182, 731-3.

Scharf, B. Critical Bands. In *Foundations of Modern Auditory Theory,* Vol I. J.V. Tobias (Ed.) New York: Academic Press, 1970.

Schouten, M.E. & van Hessen, A.J. Modeling phoneme perception. I: Categorical perception. *Journal of the Acoustical Society of America*, 1992, Oct., 92 (4 Pt 1),1841-1855.

Schreiner, C.E., Calhoun, B., & Keeling, D. Physiology and topography of cortical neurons explored with vowel-like ripple-spectra. *Association for Research in Otolaryngology Abstract*, 1993, 161.

Schreiner, C.E., & Cynader, M.S. Basic fundamental organization of second auditory cortical field (AII) of the cat. *Journal of Neurophysiology,* 1984, 51, 6, 1284-1305.

Schreiner, C.E., & Mendelson, J.R. Functional topography of cat primary auditory cortex: distribution of integrated excitation. *Journal of Neurophsyiology*, 1990, Nov., 64, 5, 1442-1459.

Schreiner, C.E., Mendelson, J.R., & Sutter, M.L. Functional topography of cat primary auditory cortex: representation of tone intensity. *Experimental Brain Research*, 1992, 92, 1, 105-122.

Schreiner, C.E., & Urbas, J.V. Representation of amplitude modulation in the auditory cortex of the cat. I. The anterior auditory field (AAF). *Hearing research,* 1986, 21, 227-41.

Schreiner, C.E., & Urbas, J.V. Representation of amplitude modulation in the auditory cortex of the cat. II. Comparison between cortical fields. *Hearing research,* 1988, 32, 49-64.

Seyfarth, R., Cheney, D., & Marler, P. Monkey responses to three different alarm calls: evidence of predator classification and semantic communication. *Science*, 1980, 210, 14 Nov., 801-3.

Shamma, S.A., Fleshman, J.W., Wiser,P.W., & Versnel, H. Organization of response areas in ferret primary auditory cortex. *Journal of Neurophysiology*, 1993, Feb. 69,2, 367-383.

Shamma, S.A. & Vranic, S. Detection of changes in spectral peak shape varies with peak amplitude, spectral density, and pitch. *Association for Reasearch in Otolaryngology Abstracts*, 1993, 187, p.47.

Shamma, S.A., Vranic, S., & Wiser, P. Spectral gradient columns in primary auditory cortex: physiological and psychoacoustical correlates. In *Auditory Physiology and Perception*, Y.Cazals, K. Horner, & L. Demany (Eds.) Great Britain: Pergamon Press, 1992.

Shipley, C., Carterette, E.C., & Buchwald, J.S. The acoustic phonetics of the cat's meow. Personal communication. Accepted *Journal of the Acoustical Society of America.*

Sinex, D.G., & McDonald, L.P. Average discharge rate representation of voice onset time in the chinchilla auditory nerve. *Journal of the acoustical society of America,* 1988, 83 (5), May, 1817-27.

Sinex, D.G., & McDonald, L.P. Synchronized discharge rate representation of voice-onset time in the chinchilla auditory nerve. *Journal of the acoustical society of America,* 1989, 85 (5), May, 1995-2004.

Sinnott, J.M., & Adams, F.S. Differences in human and monkey sensitivity to acoustic cues underlying voicing contrasts. *Journal of the acoustical society of America*, 1987, 82, 5, 1539-47.

Sinnott, J.M., Beecher, M.D., Moody, D.B., & Stebbins, W.S. Speech sound discrimination by monkeys and humans. *Journal of the acoustical society of America*, 1976, 60, 3, 687-95.

Slawson, A.W. Vowel quality and musical timbre as functions of spectrum envelope and fundamental frequency. *Journal of the Acoustical Society of America,* 1968, 43, 1, 87-101.

Smith, W.J. *The Behavior of Communicating*. Cambridge, Mass.: Harvard University Press, 1977.

Spearman, C. "General intelligence," objectively determined and measured. *American Journal of Psychology*, 1904, 15, 201-293.

Spenner, B.F., Engebretson, A.M., Miller, J.D., & Cox, J.R. Random-access programmable recorder of complex sounds (RAP): A digital instrument for auditory research. *Journal of the acoustical society of America,* 1974, 55, 427.

Steinschneider, M., Arezzo, J.C., & Vaughan, H.G.Jr. Tonotopic features of speech-evoked activity in primate auditory cortex. *Brain Research,* 1990, 519, 158-168.

Steinschneider, M., Schroeder, C.E., Arezzo, J.C., & Vaughan, H.G. Jr. Physiological mechanisms for categorical responses reflecting the voice onset time boundary in auditory cortex. *Soc. Neuroscience. Absts.* 1992, 18, 353.4, 841.

Strange, W. Evolving theories of vowel perception. *Journal of the Acoustical Society of America*, 1989, May, 85(5), 2081-2087.

Streeter, L.A. Language perception of 2-month-old infants shows effects of both innate mechanisms and experience. *Nature (London)*, 1976, 259, 38-41.

Suga, N. Auditory neuroethology and speech processing: complex-sound processing by combination-sensitive neurons. In *Auditory function - neurobiological bases of hearing*, G.M. Edelman, W.E. Gall, & W.M. Cowan (Eds.). New York: John Wiley & Sons, 1988.

Suga, N. The extent to which biosonar information is represented in the bat auditory cortex. In *Dynamic aspects of neocortical function*, G.M. Edelman, W.E. Gall, & W.M. Cowan (Eds.). New York: John Wiley & Sons, 1984.

Swoboda, P.J., Morse, P.A., & Leavitt, L. Continuous vowel discrimination in normal and at risk infants. *Child Development*, 1976, 47, 459-465.

Syrdal, A.K., and Gopal., H.S. A perceptual model of vowel recognition based on the auditory representation of American English vowels. *Journal of the Acoustical Society of America*, 1986, April, 79(4), 1086-1100.

Tartter, V. *Language Processes*. New York: Holt, Rinehart & Winston, 1986.

Terrace, H.S., Petitto, L.A., Sanders, R.J., & Bever, T.G. Can an ape create a sentence? *Science*, 1979, 206, 891-902.

Thompson, M., Porter, B., O'Bryan, J., Heffner, H.E., and Heffner, R.S. A syringe-pump food-paste dispenser. *Behavior Research Methods, Instruments, & Computers*, 1990, 22(50), 449-450.

Trehub, S.E. Infants' sensitivity to vowel and tonal contrasts. *Developmental psychology*, 1973, 9, 91-96.

Trehub, S.E., Endman, M.W., & Thorpe, L.A. Infants' perception of timbre: classification of complex tones by spectral structure. *Journal of Experimental Child Psychology*, 1990, 49, 300-313.

Van Veen, T.M. & Houtgast, T. On the perception of spectral modulations. In *Hearing - phsyiological bases and psychophysics*, R. Klinke & R. Hartment (Eds.). Berlin: Springer-Verlag, 1983.

Van Veen, T.M., & Houtgast, T. Spectral sharpness and vowel dissimilarity. *Journal of the Acoustical Society of America*, 1985, Feb., 77(2), 628-634.

Vranic, S., Versnel, H., & Shamma, S.A. Single and double spectral peaks: psychoacoustical and physiological results. *Association for Research in Otolaryngology Abstracts*, 1993, 188, p.47.

Waters, R.S., & Wilson, W.A. Speech perception by rhesus monkeys: the voicing distinction in synthesized labial and velar stop consonants. *Perception and psychophysics*, 1976, 19, 4, 285-9.

Watson, B.U. Some relationships between intelligence and auditory discrimination. *Journal of Speech and Hearing Research,* 1991, June, 34, 621-627.

Weinberger, N.M. & Diamond, D.M. Dynamic modulation of the auditory system by associative learning. In *Auditory function - neurobiological bases of hearing*, G.M.Edelman, W.E. Gall, & W.M. Cowan (Eds.). New York: John Wiley & Sons, 1988.

Wier, C.C., Jesteadt, W., & Green, D.M. Frequency discrimination as a function of frequency and sensation level. Journal of the Acoustical Society of America, 1977, Jan., 61, 1, 178-184.

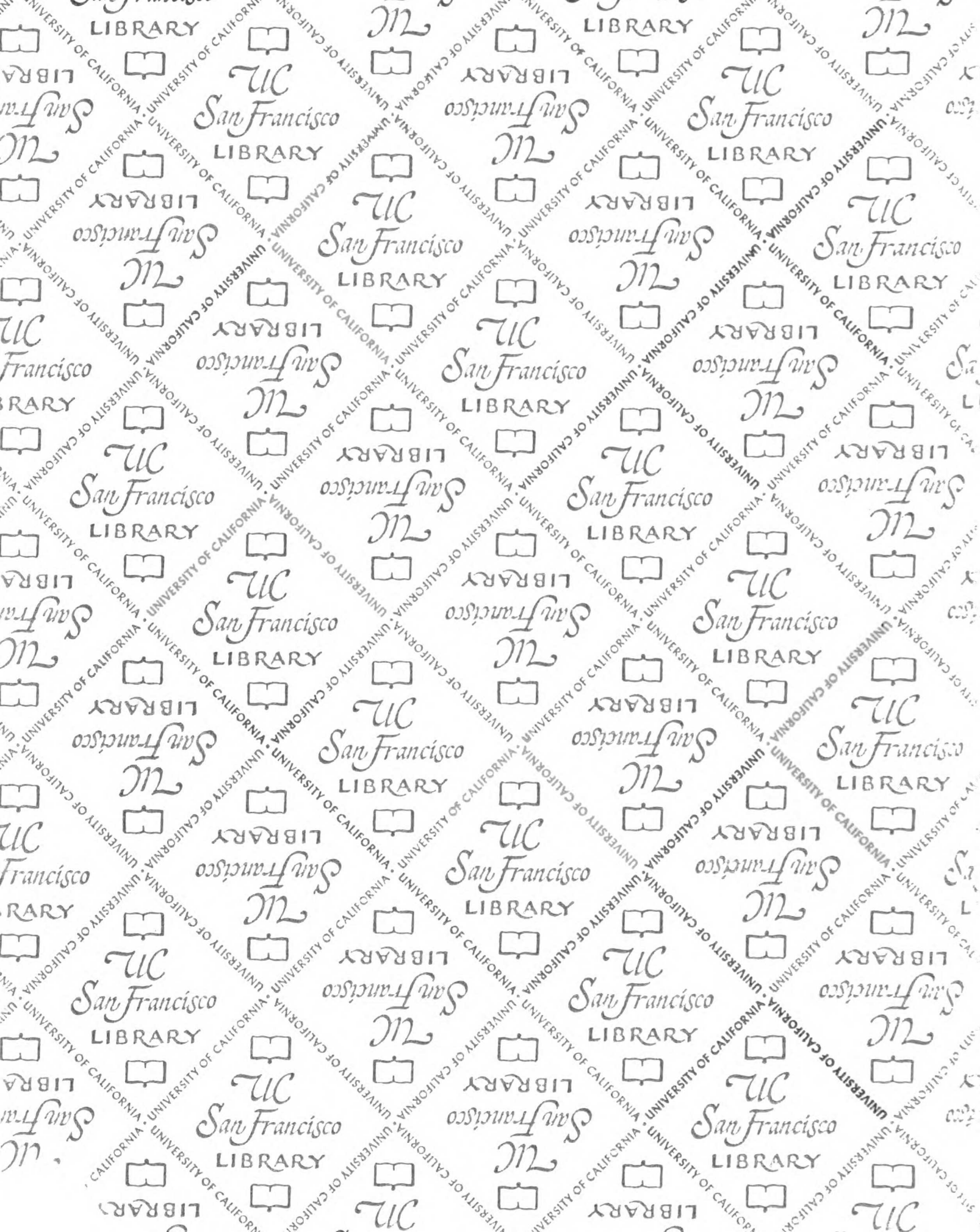
Winer, J.A., Diamond, I.T., & Raczkowski, D. Subdivisions of the auditory cortex of the cat: the retrograde transport of horseradish peroxidase to the medial geniculate body and posterior thalamic nuclei. *Journal of comparative neurology*, 1977, 176, 387-417.

Wollberg,A., & Newman, J.D. Auditory cortex of squirrel monkey: response patterns of single cells to species-specific vocalizations. *Science*, 1972, 175, 14 Jan., 212-4.

Woolsey, C.N., & Walzl, E.M. Topical projection of nerve fibers from local regions of the cochlea to the cerebral cortex of the cat. *Bulletin of Johns Hopkins Hospital*, 1942, 71, 315-44.

Zwicker, E. Subdivision of the audible frequency range into critical bands (Frequenzgruppen). *Journal of the Acoustical Society of America*, 1961, 60, 248.

Zwicker, E., Flottorp, G., & Stevens, S.S. Critical bandwidth in loudness summation. *Journal of the Acoustical Society of America*, 1957, 29, 548-557.


UC
San Francisco
LIBRARY
UNIVERSITY OF CALIFORNIA · UNIVERSITY OF CALIFORNIA