

# Neural Correlates of the Pitch of Complex Tones. I. Pitch and Pitch Salience

PETER A. CARIANI AND BERTRAND DELGUTTE

*Department of Otolaryngology, Harvard Medical School, and Eaton-Peabody Laboratory for Auditory Physiology, Massachusetts Eye and Ear Infirmary, Boston, Massachusetts 02114*

## SUMMARY AND CONCLUSIONS

1. The temporal discharge patterns of auditory nerve fibers in Dial-anesthetized cats were studied in response to periodic complex acoustic waveforms that evoke pitches at their fundamental frequencies. Single-formant vowels, amplitude-modulated (AM) and quasi-frequency-modulated tones, AM noise, click trains, and other complex tones were utilized. Distributions of intervals between successive spikes ("1st-order intervals") and between both successive and nonsuccessive spikes ("all-order intervals") were computed from spike trains. Intervals from many fibers were pooled to estimate interspike interval distributions for the entire auditory nerve. Properties of these "pooled interspike interval distributions," such as the positions of interval peaks and their relative heights, were examined for correspondence to the psychophysical data on pitch frequency and pitch salience.

2. For a diverse set of complex stimuli and levels, the most frequent all-order interspike interval present in the pooled distribution corresponded to the pitch heard in psychophysical experiments. Pitch estimates based on pooled interval distributions (30–85 fibers, 100 stimulus presentations per fiber) were highly accurate (within 1%) for harmonic stimuli that produce strong pitches at 60 dB SPL.

3. Although the most frequent intervals in pooled all-order interval distributions were very stable with respect to sound intensity level (40, 60, and 80 dB total SPL), this was not necessarily the case for first-order interval distributions. Because the low pitches of complex tones are largely invariant with respect to level, pitches estimated from all-order interval distributions correspond better to perception.

4. Spectrally diverse stimuli that evoke similar low pitches produce pooled interval distributions with similar most-frequent intervals. This suggests that the pitch equivalence of these different stimuli could result from central auditory processing mechanisms that analyze interspike interval patterns.

5. Complex stimuli that evoke strong or "salient" pitches produce pooled interval distributions with high peak-to-mean ratios. Those stimuli that evoke weak pitches produce pooled interval distributions with low peak-to-mean ratios.

6. Pooled interspike interval distributions for stimuli consisting of low-frequency components generally resembled the short-time autocorrelation function of stimulus waveforms. Pooled interval distributions for stimuli consisting of high-frequency components resembled the short-time autocorrelation function of the waveform envelope.

7. Interval distributions in populations of neurons constitute a general, distributed means of encoding, transmitting, and representing information. Existence of a central processor capable of analyzing these interval patterns could provide a unified explanation for many different aspects of pitch perception.

veys information concerning voicing, prosody, and speaker identity. In music, sequences of pitches comprise melodies, whereas their relationships determine harmonies. Common pitches facilitate the grouping of multiple sounds into auditory objects, whereas differences in pitch enhance their separation. The study of the physiological mechanisms subserving pitch therefore permits general insights into the neural basis of auditory information processing. Pitch perception has been studied extensively through psychophysical methods, and the nature of the neural mechanisms underlying pitch has been vigorously debated for over a century. As a consequence there exists a large bodies of highly developed psychophysical observation and theory for pitch with which observed patterns of neural activity can be compared.

With very few exceptions, periodic complex acoustic stimuli evoke low pitches associated with their fundamental frequencies ( $F_0$ ). These low pitches have been variously labeled "periodicity pitch," "the pitch of complex tones," "virtual pitch," "repetition pitch," "the pitch of the missing fundamental," "residue pitch," "synthetic pitch," or "musical pitch" (de Boer 1976; Evans 1978; Moore 1989; Nordmark 1978; Plomp 1967, 1976; Small 1970; Warren 1982). These low pitches associated with multicomponent spectral patterns can be contrasted with "place," "spectral," "analytical," or "pure tone" pitches that are associated with individual frequency components. Various psychophysical evidence suggests that place pitches and periodicity pitches may be produced by different neural mechanisms. Periodicity pitches can be reliably heard when there is little or no energy present at  $F_0$  (i.e., a missing fundamental) or when the frequency region of the fundamental is masked with noise (Licklider 1954). This means that a particular frequency region (place) need not be differentially excited in order to evoke a corresponding pitch. Conversely, a given frequency region can "give rise to sensations of widely different pitch" depending on the temporal pattern of stimulation (Ritsma 1962a). It appears that periodicity pitch is not limited to humans, but is also perceived by many other animals: monkeys (Tomlinson and Schwartz 1988), cats (Hefner and Whitfield 1976), birds (Cynx 1986), and fish (Fay 1972; Fay et al. 1983).

Auditory physiologists have long appreciated the ability of auditory nerve fibers (ANFs) to convey information concerning the frequency content of the stimulus through both spatial and temporal patterns of discharge (Evans 1978; Kiang et al. 1965; Rose et al. 1967; Wever 1949). Early on it was observed that neural discharge periodicities corresponding to the low pitches of complex tones are present in

## INTRODUCTION

Pitch is a fundamental auditory quality important for the perception of speech and music. In speech, voice pitch con-

ANFs, and most subsequent physiological studies of the pitch of complex tones have focused on temporal pattern information. In the auditory nerve, Rose and coworkers (Rose 1980; Rose et al. 1969) showed that pairs of harmonically related tones give rise to interspike intervals corresponding to the period of their common fundamental. They noted that these intervals are always at least as common as others in single-fiber interval distributions, and suggested that populations of fibers could convey this information more centrally. With the use of both harmonic and inharmonic, amplitude-modulated (AM) tones, Evans (1978, 1983) and Javel (1980) found interspike interval peaks corresponding to the perceived pitches. Temporal discharge patterns related to pitch have also been found for two-tone complexes (Greenberg and Rhode 1987), synthetic speech sounds (Delgutte 1980; Miller and Sachs 1984; Palmer et al. 1986), and rippled noise (Fay et al. 1983; ten Kate and van Bakkum 1988).

Models for pitch perception have coevolved with physiological observations. For over a century there has been an ongoing, shifting debate over the respective roles of place and temporal representations in the perception of pitch and other auditory forms (von Békésy 1963; de Boer 1976; Borring 1942; Evans 1978; Lyon and Shamma 1995; Nordmark 1970, 1978; Small 1970). Rate-place neural models use spatial discharge rate patterns in tonotopically ordered neural maps to represent the stimulus power spectrum. Spectrally based pattern recognition mechanisms for pitch then analyze this "central spectrum" to detect patterns of excitation that are indicative of sets of related harmonics, e.g., Goldstein (1973); Terhardt (1973). Temporal-place models represent the fine temporal structure of neural discharges of each tonotopic place. Information concerning discharge synchrony between neighboring neurons (Young and Sachs 1979) and interspike intervals within single neurons (Goldstein and Sruлович 1977; Houtsma and Smurzynski 1990; Sruлович and Goldstein 1983) can be combined with place information to form a sharpened central spectrum that is then analyzed with the use of pattern recognition mechanisms. Purely temporal models combine interspike intervals from single auditory neurons in all frequency regions to produce "population interval distributions," i.e., global distributions of intervals in neural populations. Purely temporal models can be based on either first-order intervals, i.e., between successive spikes (Moore 1989; van Noorden 1982), or all-order intervals, i.e., between successive and nonsuccessive spikes (de Cheveigné 1986; Licklider 1951, 1956, 1959; Meddis and Hewitt 1991a,b). All of these models embody a "predominant interval" hypothesis, that the pitch heard corresponds to the most frequent interval in a population interval distribution. A subsidiary hypothesis is that the strength of this pitch corresponds to the ratio between pitch-related intervals and the total number of intervals present. Similar temporal strategies for estimating pitch that use cochlear models and auto-correlation functions (Lyon and Shamma 1995; Slaney and Lyon 1993) have recently been implemented in silicon (Lazaro and Mead 1989; Moore 1990). All of these purely temporal models successfully predict pitch for a wide range of complex stimuli, reinforcing the plausibility that the predominant interval hypothesis might hold at the level of the auditory nerve.

In the present study we assess how well features of the

population all-order interspike interval distribution of the auditory nerve correspond with pitches perceived by human listeners. This study is the first attempt to test the predominant interval hypothesis directly, with the use of physiological data obtained from Dial-anesthetized cats. Pooled interspike interval distributions constructed from recorded responses of large numbers of single ANFs served as estimates of the distribution of intervals in the entire auditory nerve. Pitch predictions computed from these estimated population interval distributions were then compared with human pitch judgments to identify correspondences between pitch percepts and patterns of neural activity.

To carry out this project, complex stimuli that have played key roles in psychophysical studies of pitch were chosen. These stimuli included single-formant vowels, click trains, AM tones, quasi-frequency-modulated tones, AM broadband noise, Shepard-Risset tones, and equal-amplitude harmonic complex tones. The stimulus set covered a wide range of pitch phenomena: the missing fundamental, voice pitch, pitch equivalence between different stimuli, pitch of unresolved harmonics, pitch shift of inharmonic tones, phase invariance, pitch circularity, "rate pitch," and the dominance region for pitch. The extensive use of these kinds of stimuli in psychophysical experiments permitted detailed comparison between neural responses and patterns of pitch judgments. Stimuli continuously varying in pitch-related parameters were used to observe neural correlates of entire ranges of pitch.

In this paper we present the basic techniques that were used to estimate pitch and pitch salience from pooled interspike interval distributions and the basic findings concerning the pitch and pitch salience of harmonic stimuli and AM noise. A companion paper presents results for stimuli that produce more complex patterns of pitch judgments.

## METHODS

### *Recording techniques*

Surgical and recording methods were as described by Kiang et al. (1965). Healthy cats were maintained in a state of surgical anesthesia throughout each experiment by an initial administration of Dial (75 mg/kg) in urethan, with supplementary doses given as needed. A posterior craniectomy was performed, and the bullae and middle ear cavities were opened on both sides. The cerebellum was retracted medially to expose the auditory nerve, and glass micropipette electrodes filled with 2 M KCl were placed under visual observation. During the course of the experiment the animal was given dexamethasone to reduce brain swelling and Ringer saline to prevent dehydration. The condition of the cat's hearing was assessed by monitoring both pure tone thresholds for single ANFs and click-evoked thresholds of compound action potentials measured at the round window.

Recordings were carried out in a temperature-controlled, electrically shielded and acoustically isolated chamber. Spike arrival times were measured with 1  $\mu$ s accuracy. Spontaneous discharge rates (SRs) were measured over 20 s. Threshold tuning curves for discharge rate [50-ms tone bursts (Kiang et al. 1970)] were used to measure the characteristic frequency (CF) of each unit. Units exhibiting abnormally high rate thresholds, above the 95% confidence interval for normal cat ANFs (Lieberman 1982; Liberman and Kiang 1978), were excluded from analysis. The rate thresholds and CFs of the 507 fibers from 13 cats that were included for analysis are shown in Fig. 1.

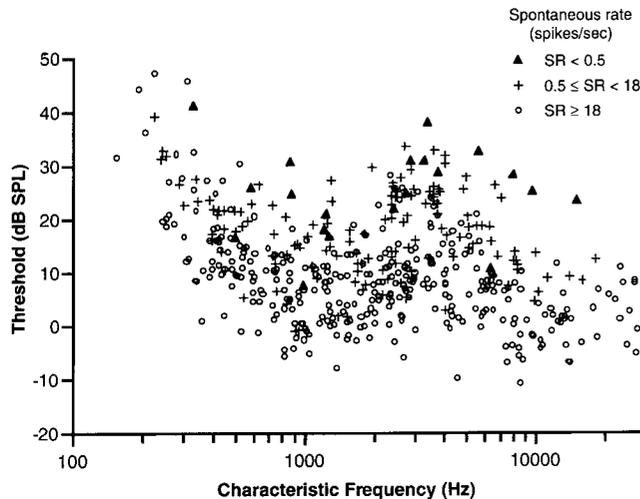


FIG. 1. Characteristic frequencies (CFs), discharge rate thresholds, and spontaneous rates (SRs) for the 507 auditory nerve fibers (ANFs) studied.

### Stimulus generation and delivery

Stimuli discussed in this paper are summarized in Table 1. All stimuli were synthesized numerically. To efficiently study neural responses for many different pitches, most stimuli were continuously varied in a parameter associated with pitch [e.g.,  $F_0$  or modulation frequency ( $F_m$ )]. The ranges of these stimulus variables are shown in *third column* of the table. An example of such a time-varying stimulus is the single-formant vowel shown in Fig. 2. The stimulus has an  $F_0$  that varies sinusoidally between 80 and 160 Hz over its 530-ms duration (Fig. 2B). Two segments of the stimulus waveform and their power spectra, at the beginning and middle of the stimulus, are shown in Fig. 2A.

The  $F_0$  for harmonic complexes was sinusoidally varied over 1 octave. For AM tones  $F_m$  was varied over a 1-octave range, while the carrier frequency ( $F_c$ ) was held constant. All stimuli were presented continuously.

Stimuli were generated by a 16-bit, digital-to-analog converter (MassComp DA04H). All signals were sampled at 100 kHz, low-pass filtered to reduce aliasing (cutoff 44 kHz), and delivered using a Beyer DT48 driver via closed, calibrated acoustic assembly. Total harmonic distortion for the sound delivery system was near  $-70$  dB re fundamental measured for 1-, 3-, and 5-kHz pure tones at 110 dB SPL. Distortion produced at the  $F_0$ s of two-tone com-

plexes (e.g., 3rd and 4th harmonics of 170 Hz) presented at  $\sim 60$  dB SPL was  $\geq 50$  dB below the primaries. To control sound pressure waveform present at the eardrum, broadband stimuli were digitally filtered to compensate for the transfer characteristics of the acoustic assembly. Thus a given stimulus always had the same power spectrum at the eardrum for each ear studied. Unless otherwise noted, each stimulus was continuously presented to a given fiber 100 times at 60-dB total SPL.

### Individual fiber autocorrelation histograms and autocorrelograms

Time intervals between spikes have been studied extensively in the spontaneous and stimulus-driven discharges of many different kinds of neurons (Perkell and Bullock 1968). Most analyses of interspike intervals have examined distributions of "first-order interspike intervals," which are intervals between successive spikes (Rodieck et al. 1962). Intervals between both successive and non-successive spikes, or "all-order intervals" can also be counted to form "all-order interval histograms" (Møller 1970; Rodieck 1967). Such histograms are also called "autocorrelation histograms" because of their formal equivalence to the autocorrelation function of spike trains (Licklider 1951). Because of their relation to renewal processes, these histograms are also called "renewal density functions," "expectation density functions," "intensity functions" and "postfiring interval distributions" (Møller 1970). Unless otherwise qualified, all mention here of interspike intervals refers to all-order intervals.

Most of the stimuli employed have time-varying periodicities that can produce rapid changes in interspike interval patterns. Visualization and analysis of changing interval distributions consequently requires a representation of spike train data that combines peristimulus time (PST) and interspike interval information. For similar reasons, running distributions of first-order intervals had been used by other workers to investigate interval coding in the visual system (Chung et al. 1970), discharge regularity in the cochlear nucleus (Bourk 1976), and ANF responses to frequency-modulated tones (Sinex and Geisler 1981). Because the running interval representations used here consist of all-order intervals, they are called autocorrelograms. These are related to analog "correlatograms" that were used for early autocorrelation analyses of speech (Lange 1967).

The construction of an autocorrelogram from a spike train is schematized in Fig. 3. Figure 3A shows two fundamental periods of a single-formant vowel and Fig. 3B shows a hypothetical response spike train consisting of four spikes (labeled a-d). First-

TABLE 1. Stimulus parameters

Stimulus	Variable	Range, Hz	Duration, ms	Level, dB SPL	Remarks
Pure tone	$F_0$	160 (Fixed)	100	60	Single component
Vowel, single formant	$F_0$	80–160 160–320	501	40, 60, 80	$F_1 = 640$ Hz Formant bandwidth = 50 Hz
AM tone	$F_m$	80–160 160–320	480	60	$F_c = 640$ Hz, $m = 1.0$
AM noise	$F_m$	80–160 160–320	530	60	Broadband noise
Click trains	$F_0$	160 (Fixed) 80–160 160–320	937 530.4	60 80	Uniform-polarity 40- $\mu$ s clicks Uniform-polarity 40- $\mu$ s clicks Alternating-polarity 40- $\mu$ s clicks
Harmonic complexes	$F_0$	80–160 160–320 240–480	498	60	Harmonics 3–5
	$F_0$	96–192 192–384 288–576	498	60	Harmonics 6–12

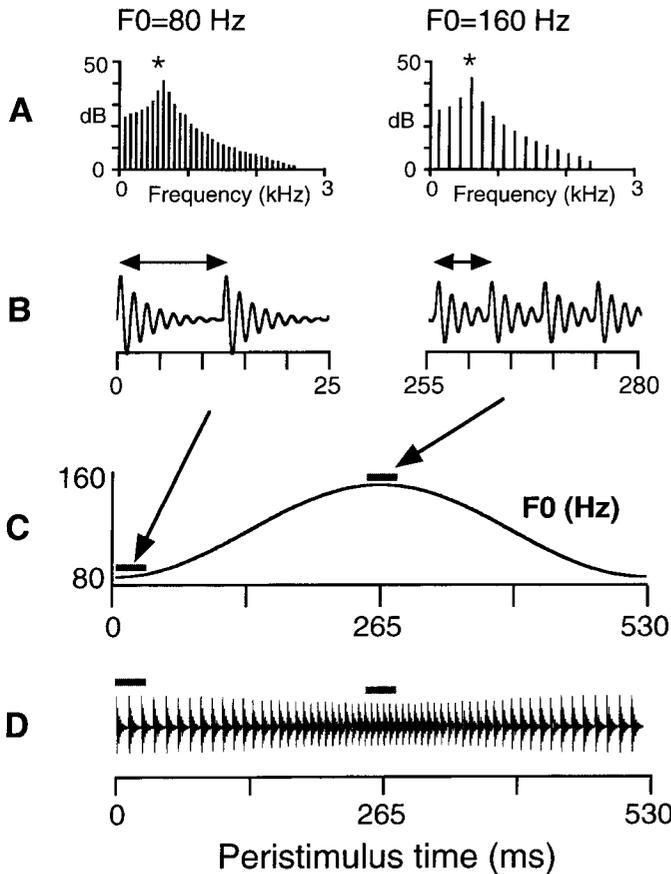


FIG. 2. Variable-fundamental stimulus. Single-formant vowel, fundamental frequency ( $F_0$ ) = 80–160 Hz, formant frequency ( $F_1$ ) = 640 Hz, duration = 530 ms. *A*: stimulus power spectra for 2 stimulus time segments, when  $F_0$  = 80 Hz (*left*) and when  $F_0$  = 160 Hz (*right*). Asterisks:  $F_1$  = 640 Hz. *B*: corresponding waveforms for the 2 segments. Arrows: fundamental period  $1/F_0$ . *C*: sinusoidal variation of the  $F_0$  over the course of the stimulus. *D*: complete waveform.

order intervals in this spike train are a-b, b-c, and c-d, whereas all-order intervals also include a-c, a-d, and b-d. For each spike, a list of all interspike intervals ending at that spike is compiled (e.g., intervals d-c, d-b, and d-a all end with spike d). The autocorrelogram (Fig. 3C) is a two-dimensional histogram whose bins contain the number of interspike intervals of a given length (vertical axis) ending at a particular time relative to the stimulus onset (horizontal axis). Thus horizontal bins represent PST, whereas vertical bins represent interspike interval. Autocorrelograms for three ANFs in response to a single-formant vowel are shown in Fig. 6, A–C. In these single-fiber autocorrelograms, each dot corresponds to one interspike interval.

*Pooled autocorrelation histograms and autocorrelograms*

The interval distribution of a population of neurons can be estimated by summing the interval distributions of many single units randomly sampled from the population. All-order interval (autocorrelation) histograms for single ANFs were summed together to form pooled histograms. Because pooled interval distributions obtained from cats are to be compared primarily with human psychophysical data, and the CF distributions of the two species are somewhat different (Fig. 4, *inset*), pooled distributions were weighted so as to approximate the estimated human CF distributions.

The same weighting procedures were used for autocorrelation histograms and autocorrelograms (Fig. 4). First, interval distributions for single fibers (e.g., Figs. 5E and 6, A–C) were grouped

by CF and summed to form octave band distributions. The octave bands had cutoffs at 0.36, 0.71, 1.4, 2.9, 5.7, and 11.4 kHz. Each octave band distribution was then normalized by the number of stimulus presentations represented in the octave band. The octave band distribution was weighted according to the proportion of similar CFs in the human distribution (Fig. 4, *inset*). This human CF distribution was estimated by scaling the distribution of cat ANF CFs (Liberman 1982; Melcher 1993) downward in frequency by a factor of 2.8 (Greenwood 1990). For example, the estimated proportion of human ANFs with CFs between 0.357 and 0.714 kHz was made equal to the estimated proportion of cat ANFs with CFs between 1 and 2 kHz (10.9%). The weighted octave band distributions were then added together to form the pooled interval distribution, which then served as an estimate of the population interval distribution of the human auditory nerve. For visual display purposes, pooled autocorrelograms were thresholded at the number of counts/bin, usually near the histogram mean, that best allowed them to be visualized and compared.

*Algorithm for pitch estimation*

A simple algorithm based on the predominant interval hypothesis was developed for estimating pitch from pooled autocorrelation histograms. For steady-state stimuli, pooled autocorrelation histograms were compiled over the entire duration of the stimulus. For time-varying stimuli, pooled autocorrelation histograms were constructed from intervals that ended in consecutive, nonoverlapping 20-ms PST segments. Pooled autocorrelation histograms with 10- $\mu$ s binwidths were then smoothed by taking a 300- $\mu$ s moving average. The position of the highest interval peak was then taken as the estimate for the pitch period. Because the stimuli are locally periodic and the neural discharges are largely stimulus-locked, autocorrelation histograms invariably show periodic patterns of major and minor maxima that are related to the fundamental period

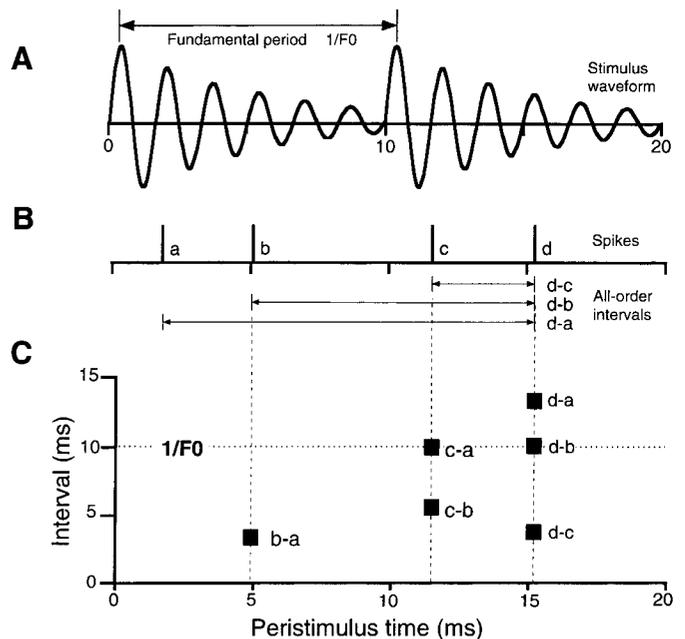


FIG. 3. Autocorrelogram construction. *A*: 2 periods from the waveform of a single-formant vowel,  $F_0$  = 100 Hz,  $F_1$  = 40 Hz. *B*: hypothetical ANF spike train showing 4 spikes labeled a-d. Double arrows: all-order interspike intervals that end with spike d. *C*: autocorrelogram computed from all of the intervals in the spike train. The autocorrelogram is a 2-dimensional histogram in which each bin holds the number of intervals of a given length (vertical axis) that end at a given peristimulus time (PST) (horizontal axis). Autocorrelogram binwidths were typically 700 ms in PST and 30 ms in time interval.

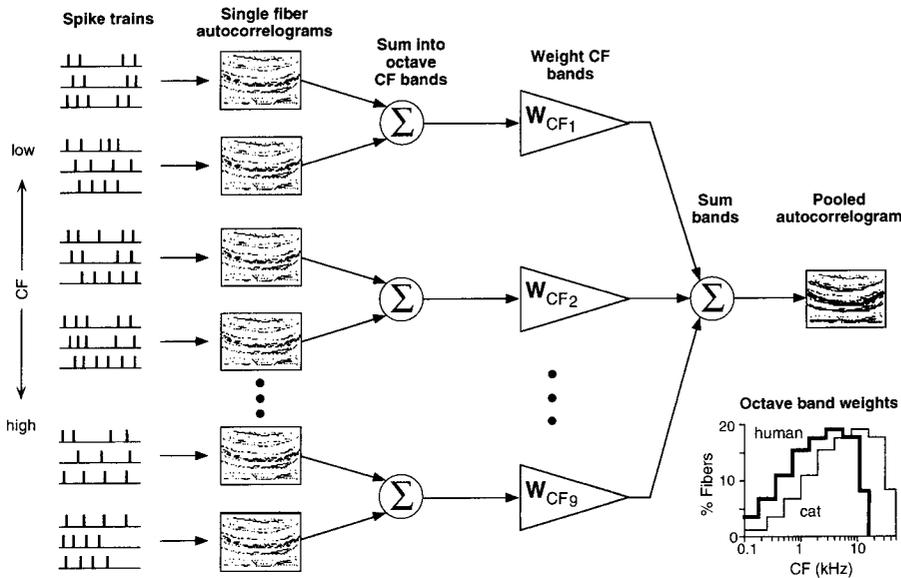


FIG. 4. Construction of pooled autocorrelograms. Autocorrelograms are 1st constructed directly from response spike trains of each fiber. Autocorrelograms for fibers in each octave CF band are normalized by number of stimulus presentations and summed together to form octave band autocorrelograms. These are then weighted according to the respective proportion of fibers in a human CF distribution and summed to form a pooled autocorrelogram. *Inset*: estimated cat and human CF distributions showing respective octave bands.

(as in Figs. 5F and 11). When these multiple, harmonically related peaks were present, the position of the major peak representing the shortest interspike interval was taken as the estimate of the pitch period.

#### Algorithm for pitch salience estimation

An interval-based signal-to-noise measure was developed for estimating pitch salience. This is the ratio of the number of intervals

in a particular interval bin of interest (signal) to the mean number of intervals per bin (background). ‘‘Fiber saliences’’ were peak-to-background ratios computed from smoothed autocorrelation histograms of individual fibers. Estimates for the saliences of pitches produced by various stimuli (Table 2) were peak-to-background ratios computed from smoothed, pooled autocorrelation histograms. Because periodic patterns of maxima and minima exist in these histograms, and their spacing changes with  $F_0$ , taking the mean number of bin counts over a fixed window (e.g., 0–15 ms)

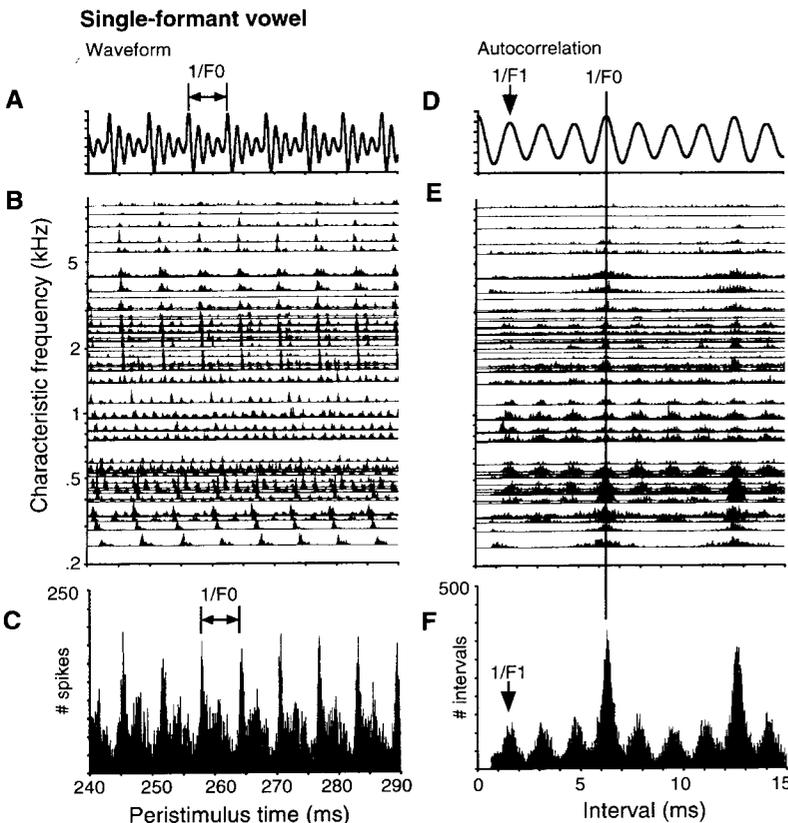


FIG. 5. ANF responses for a single-formant vowel,  $F_0 = 160$  Hz,  $F_1 = 640$  Hz. *A*: stimulus waveform. Interval indicates fundamental period  $1/F_0$ . *B*: PST histograms for 52 fibers, arranged by CF. *C*: pooled PST histogram for the 52 fibers. *D*: stimulus short-time autocorrelation function. *E*: single-fiber autocorrelation histograms, arranged by CF. *F*: pooled autocorrelation histogram. Vertical line in *D* and *F*: fundamental period  $1/F_0$ .

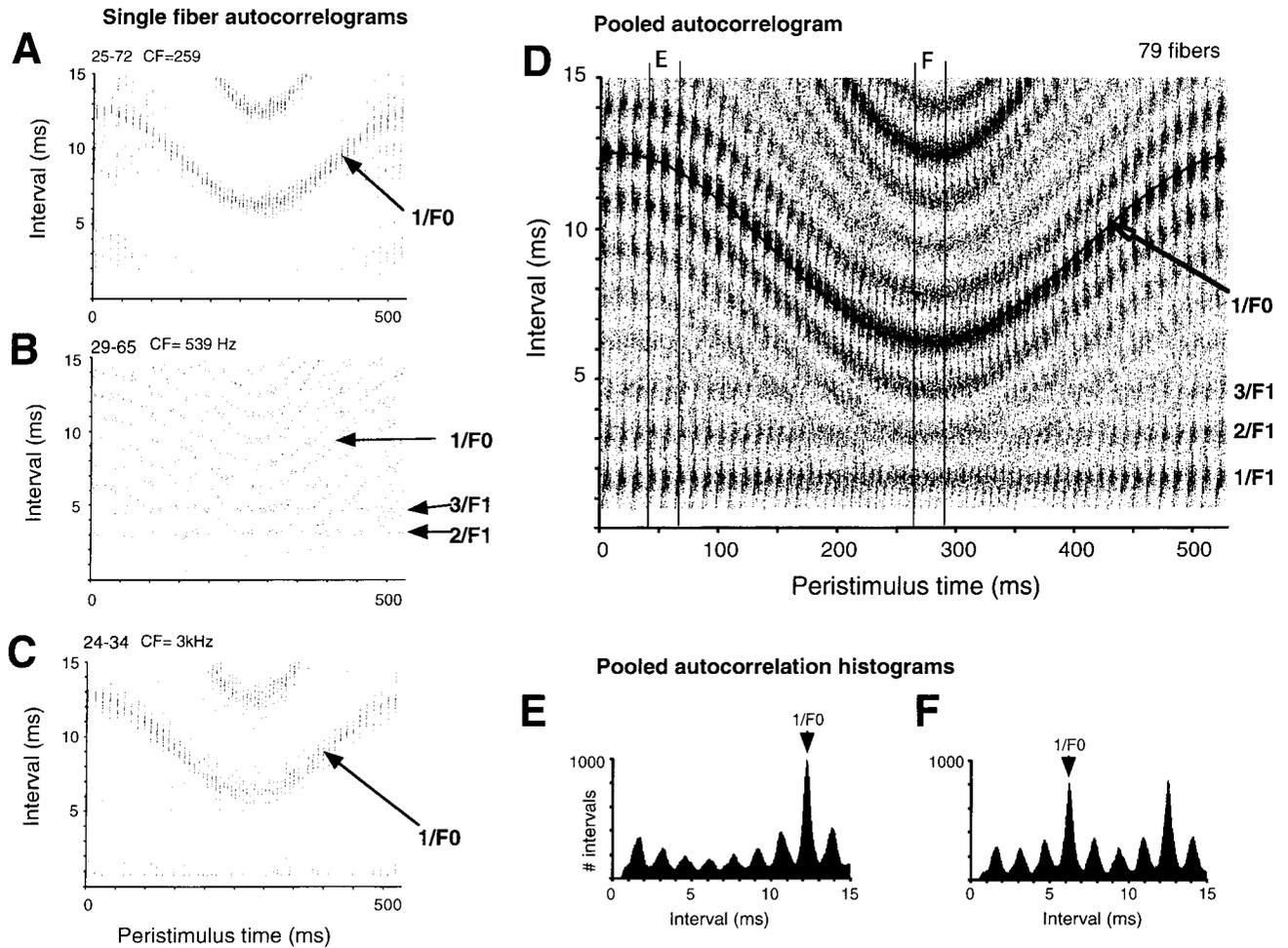


FIG. 6. Autocorrelograms and autocorrelation histograms in response to a variable- $F_0$  single-formant vowel,  $F_0 = 80$ – $160$  Hz. A–C: single-fiber autocorrelograms from 3 different CF regions. D: pooled autocorrelogram for a single-formant stimulus with  $F_0 = 80$ – $160$  Hz and  $F_1 = 640$  Hz. PST binwidth =  $714 \mu\text{s}$ , interval binwidth =  $30 \mu\text{s}$ . Solid line: fundamental period  $1/F_0$ . E and F: smoothed, pooled autocorrelation histograms from peristimulus regions labeled E and F in the pooled autocorrelogram (D).

can introduce “edge effects” into computation of the background. “Dead time” effects are similarly introduced if the computation includes all intervals from 0 to  $1/F_0$  ms. To minimize these effects, background level was computed over a window  $1/F_0$  ms wide centered on the interval of interest (signal). For example, peak-to-background ratios for 160-Hz pitches produced by stimuli with  $F_0 = 160$  Hz were computed by dividing the number of intervals

at 6.25 ms (1 bin) by the mean number of intervals per bin for the interval range of 3.125–9.375 ms.

Both the pitch estimation and salience measures were selected in part because they do not depend on absolute numbers of spikes, a dependency that would confound cross-stimulus comparisons. To facilitate direct, visual comparisons of estimated saliences, some pooled autocorrelation histograms (e.g., Fig. 11) were normalized

TABLE 2. Physiologically estimated pitches and pitch saliences

Stimulus	Level, dB SPL	Length, ms	Number of Fibers	Perceived Pitch, Hz	Estimated Pitch, Hz	Error, %	Estimated Salience
Pure tone, 160 Hz	60	125	85	160	161.0	0.64	1.7
Single formant vowel	40	50	34	160	158.7	−0.8	2.5
$F_0$ : 160 Hz	60	50	79	160	158.7	−0.8	3.1
$F_1$ : 640 Hz	80	50	50	160	159.5	−0.3	2.6
AM tone, $F_m$ : 160 Hz, $F_c$ : 640 Hz	60	100	60	160	160.3	0.2	2.7
Click train, unipolar, $F_0$ : 160 Hz	80	50	54	160	158.7	−0.8	2.8
Harmonics, 6–12 of 160 Hz	60	50	49	160	159.5	−0.3	2.7
AM tone, $F_m$ : 160 Hz, $F_c$ : 6400 Hz	60	100	54	160	166.7	4.2	1.1
AM noise, $F_m$ : 160 Hz	60	100	53	160	158.0	−1.3	1.3

For abbreviations, see Table 1.

by dividing each bin count by the average number of bin counts taken over the whole histogram.

## RESULTS

### *Pitch of single-formant vowels*

Single-formant synthetic vowels were utilized both because of their spectral simplicity and because they permit investigation into the neural correlates of voice pitch. Vowels are created when trains of volume velocity (glottal) pulses are passed through resonant (oropharyngeal) cavities. The resulting signals are harmonic complexes with most of their energy near the resonant frequencies (formants). Although human vowels generally have multiple formants, synthetic single-formant vowels have only one formant, and therefore have a fairly narrowband spectrum (Fig. 2A). Despite relatively little energy at their fundamentals, vowels evoke strong, low "voice pitches" at  $F_0$  that are very stable over a wide range of SPLs.

Two single-formant vowels were constructed, one with  $F_0$  varying between 80 and 160 Hz, in the range of a male voice, (Fig. 2) and the other with  $F_0$  varying between 160 and 320 Hz, in the range of a female voice. The formant frequency ( $F_1$ ) for both single-formant vowels was 640 Hz and the bandwidth was 50 Hz.

Neural responses for a single-formant vowel when  $F_0 \sim 160$  Hz are shown in Fig. 5. The stimulus waveform (Fig. 5A) is shown above the set of single-fiber PST histograms, arranged by CF (Fig. 5B). This PST "neurogram" (Kiang and Moxon 1972; Kiang et al. 1979) shows the stimulus-locked character of the responses. Widespread synchrony across CFs exists in high-CF regions, whereas systematic, CF-dependent latency shifts due to cochlear delays disrupt such synchronies in low-CF regions. Summing the single-fiber PST histograms produces a pooled PST histogram (Fig. 5C). Despite smearing due to the systematic latency shifts, periodicities corresponding to the fundamental and to the formant can be seen in the spacings of PST peaks.

The stimulus autocorrelation function (Fig. 5D) exhibits major maxima at the fundamental period  $1/F_0$  (vertical line) and its multiples. Minor maxima reflect the formant period  $1/F_1$  and its multiples. An autocorrelation neurogram (Fig. 5E) shows all-order interval distributions for each fiber in the sample [see other similar and related representations (Evans 1986; Lyon and Shamma 1995; Meddis and Hewitt 1991b; Patterson et al. 1995)]. Almost all of the fibers in the neurogram show intervals related to the fundamental ( $1/F_0 = 6.25$  ms). Fibers with CFs between 0.4 and 3.0 kHz additionally show intervals related to the formant ( $1/F_1 = 1.5625$  ms). Weighting and summing the single-fiber autocorrelation histograms produces a pooled autocorrelation histogram (Fig. 5F). Positions of major and minor peaks in the pooled interval distribution correspond to those in the stimulus autocorrelation function. The first major peak in the histogram, i.e., the most frequent interval in the distribution, lies very close to the fundamental period  $1/F_0$ , which is the period of the pitch that is heard. The second major peak at 12.5 ms represents intervals spanning two fundamental periods. Interval multiples such as these (e.g.,  $1/F_0$ ,  $2/F_0$ ,  $3/F_0$ , ...) are reflections of stimulus periodicities and the autocorrelation-like properties of interval representations.

Minor peaks are associated with the formant period  $1/F_1$  and its multiples.

Running interval distributions in response to the entire variable- $F_0$  single-formant vowel ( $F_0 = 80$ –160 Hz) are shown in Fig. 6. Single-fiber autocorrelograms for three fibers with different CFs (Fig. 6, A–C) all show many intervals that follow the fundamental as it changes. The lowest-CF fiber (Fig. 6A) has a CF well below  $F_1$ , whereas the highest-CF fiber (Fig. 6C) has a CF well above it. Both of these autocorrelograms show mostly intervals near the fundamental period. Upper interval bands correspond to twice the fundamental period. The middle-CF fiber (Fig. 6B) has a CF near  $F_1$ , and its autocorrelogram shows many more interval bands. One set of straight bands corresponds to multiples of the fixed formant period ( $2/F_1$ ,  $3/F_1$ ), whereas another set of undulating bands corresponds to the changing fundamental period ( $1/F_0$ ,  $2/F_0$ ). Other undulating bands represent intervals that are combinations of fundamental and formant periods (e.g.,  $1/F_0 \pm 1/F_1$ ,  $1/F_0 \pm 2/F_1$ , etc.). For this fiber, interval bands associated with the formant ( $2/F_1$ ,  $3/F_1$ ) are approximately as dense as those associated with the fundamental ( $1/F_0$ ,  $2/F_0$ ). For comparison, the autocorrelation histograms of Fig. 5E correspond to PSTs 240–290 in the autocorrelograms of Fig. 6.

To summarize the single-unit responses shown in Figs. 5 and 6, almost all fibers showed intervals related to the fundamental, and these intervals were generally at least as frequent as any other interval. Intervals for fibers with CFs away from the formant region were primarily related to the fundamental. In addition to  $F_0$ -related intervals, fibers with CFs near the formant region showed comparable numbers of various formant-related intervals. This is consistent with the findings of previous studies (Delgutte 1980; Delgutte and Kiang 1984; Miller and Sachs 1984; Palmer et al. 1986) in which responses of fibers with CFs close to vowel formants discharged throughout the vowel period, showing relatively little modulation at  $F_0$ . In these studies fibers with CFs far removed from formants discharged predominantly only over a fraction of the vowel period, generally showing more modulation at  $F_0$ . Thus the present results, based on autocorrelation histograms and autocorrelograms, are consistent with earlier ones based on PST and period histograms.

The pooled, weighted autocorrelogram for 79 fibers is shown in Fig. 6D. The (voice) pitches heard follow the changing fundamental period  $1/F_0$  (solid undulating line). The densest interval band follows the fundamental period quite closely throughout its entire octave range. Two cross sections of the autocorrelogram, labeled E and F, indicate 20-ms PST segments from which two pooled autocorrelation histograms (Fig. 6, E and F) were constructed. Major peaks in these autocorrelation histograms thus correspond to the densest bands in the autocorrelogram for those segments. In both autocorrelation histograms, the first major interval peak closely coincides with the fundamental period,  $1/F_0$ . As in Fig. 5F, minor peaks in the autocorrelation histograms correspond to formant-related periodicities. As in the autocorrelogram for the middle-CF fiber (Fig. 6B), the pooled autocorrelogram shows sets of minor interval bands related to the formant period and formant-fundamental combinations, their main difference being in the greater proportion of  $F_0$ -related intervals in the pooled case. Analogous patterns of intervals related to fundamentals, formants, and their in-

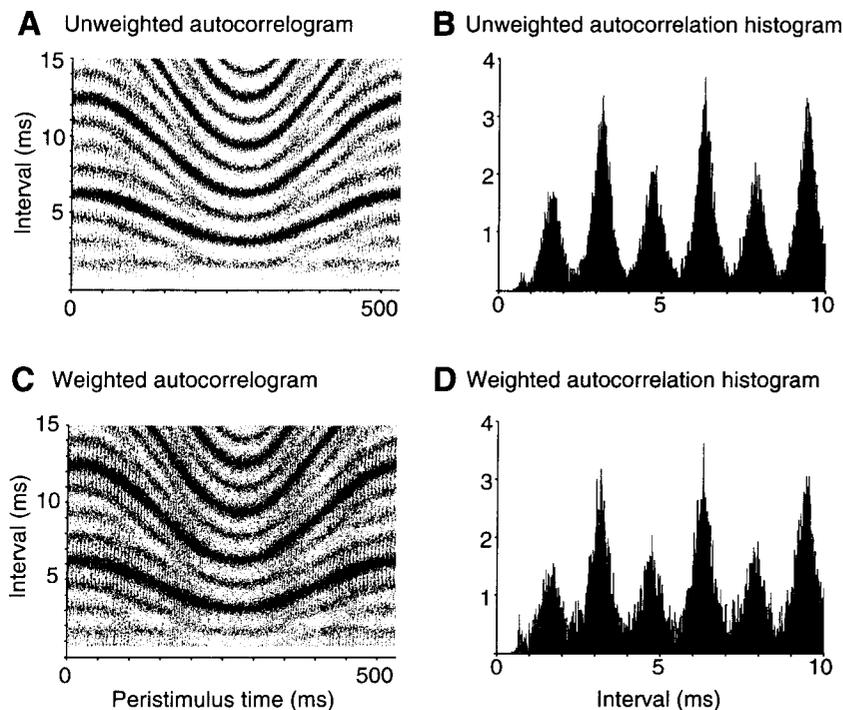


FIG. 7. Effects of explicit weighting by CF on pooled interval distributions. Stimulus: variable- $F_0$  single-formant vowel,  $F_0 = 160\text{--}320$  Hz,  $F_1 = 640$  Hz. *A* and *C*: unweighted and CF-weighted autocorrelograms. *B* and *D*: unweighted and CF-weighted pooled autocorrelation histograms compiled from peristimulus interval 240–290 ms ( $F_0 = 320$  Hz). *A–D* are all computed from the same set of 64 fibers. Pooled histograms were normalized to the mean histogram bin count and left unsmoothed. Autocorrelograms were thresholded at 1.25 times the mean bin count.

interactions are seen in pooled interval distributions for multiple-formant synthetic vowels (Delgutte 1996; Lyon and Shamma 1995), where these patterns can be utilized to discriminate single vowels (Palmer 1992), to identify concurrent vowel pairs (Cariani and Delgutte 1993, 1994), and to provide explicit neural representations of  $F_1$  (Hirahara et al. 1996).

More quantitative estimates for pitch and pitch salience are presented in Table 2. Each estimate is based on 100 stimulus presentations per fiber, with responses taken from stimulus segments of specified lengths. The estimated pitch for the single-formant vowel presented at 60 dB SPL when  $F_0 = 160$  Hz (PSTs 240–290 ms) was 158.7, a deviation of 0.8%.

#### *Robustness of pitch estimates with respect to sampling variability*

The estimation of pitch from pooled interval distributions was found to be very robust with respect to which fibers were sampled. Each stimulus of Table 1 produced a characteristic pooled interval distribution that remained quite stable once data from >15–20 fibers (100 stimulus presentations per fiber) were included. Subsequent exclusion or inclusion of single fibers, whatever their CF or SR, generally altered the shape of this distribution very little.

For harmonic stimuli, the procedure of weighting by CF band also had little effect on interval distributions. Figure 7 shows both unweighted and weighted pooled autocorrelograms and autocorrelation histograms for the variable- $F_0$  single-formant vowel ( $F_0 = 160\text{--}320$  Hz). Weighted distributions (Fig. 7, *C* and *D*) were sometimes slightly noisier than their unweighted counterparts (Fig. 7, *A* and *B*) because of underrepresentation of particular CFs in fiber samples (e.g., high CFs for a low-frequency stimulus or vice versa).

Noisier pooled interval distributions are produced when responses averaged across fewer fibers are weighted more heavily to compensate for their fewer numbers. This difference notwithstanding, the shapes and peak-to-background ratios of the weighted and unweighted distributions in Fig. 7 are very similar. Pitches estimated from weighted distributions were also very similar to those previously estimated from unweighted distributions (Cariani and Delgutte 1992; Delgutte and Cariani 1992). Although explicit weighting by CF band generally matters less for harmonic stimuli, it is potentially more important for inharmonic stimuli (e.g., as in the companion paper), especially where multiple competing periodicities are present in different frequency bands.

The stability of pooled interval distributions with respect to sampling variation gives some indication of the robustness of a population interval code for pitch. Because  $F_0$ -related intervals are often present across many CF regions, selection of fibers with particular CFs is generally not critical. Because the interval information produced by each fiber is relatively precise (Goldstein and Srulovicz 1977; Siebert 1970), highly accurate estimates of  $F_0$  are possible with the use of small numbers of randomly chosen fibers.

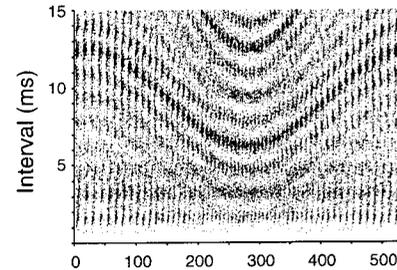
#### *Pitch invariance with stimulus level*

Pitch judgments for harmonic complex tones are nearly invariant over a wide range of stimulus levels, changing <1–2% as levels are raised from 40 to 80 dB SPL (Plomp 1976; Zwicker and Fastl 1990). The perceived strengths of the low pitches of complex tones are generally weak at low levels, rise rapidly at moderate levels, and plateau for higher levels (Bilsen and Ritsma 1970; Yost and Hill 1978).

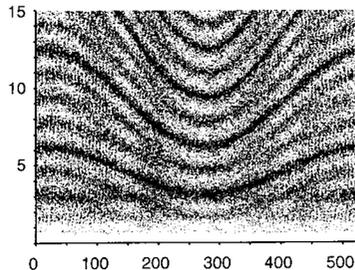
Pooled autocorrelograms for two single-formant vowels ( $F_0 = 80\text{--}160$  Hz;  $F_0 = 160\text{--}320$  Hz) at three stimulus levels (40, 60, and 80 dB SPL) are shown in Fig. 8. The

## Single Formant Vowels

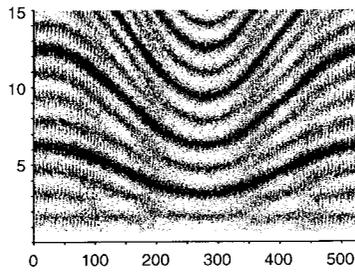
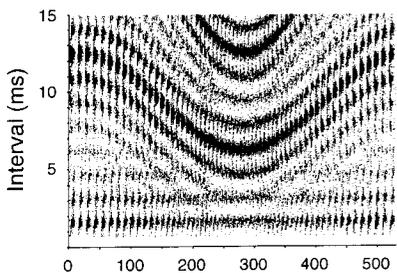
40 dB SPL  $F_0$ : 80-160 Hz



$F_0$ : 160-320 Hz



60 dB SPL



80 dB SPL

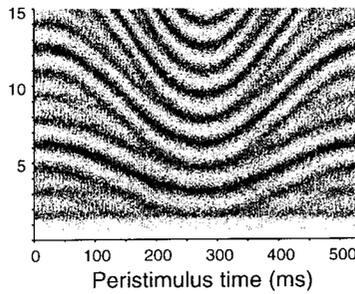
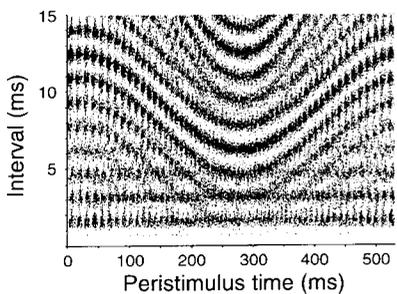


FIG. 8. Pooled response autocorrelograms for 2 variable- $F_0$  single-formant vowels presented at 40, 60, and 80 dB SPL.  $F_1 = 640$  Hz. *Left*:  $F_0 = 80-160$  Hz;  $N = 26, 79,$  and  $50$  fibers, respectively. *Right*:  $F_0 = 160-320$  Hz;  $N = 29, 70,$  and  $36$  fibers, respectively.

most frequent interval was always very close to the perceived pitch period over the entire range of levels (40–80 dB SPL) and fundamentals (80–320 Hz). Precise positions of interval peaks and their relative heights are best seen in pooled autocorrelation histograms for the same stimulus (Fig. 9, A–C). More quantitative estimates are presented in Table 2 for  $F_0 = 160$  Hz, where the difference between physiologically estimated pitches and the fundamental ranged from 0.3 to 0.8%. All of these measures confirm that pitches estimated from pooled interval distributions maintain a high degree of accuracy over a very wide dynamic range (40 dB).

Interval peak-to-background ratios, which were used as estimates of pitch salience, remained high from 40 to 80 dB SPL. This is qualitatively seen in the densities of  $F_0$ -related interval bands in autocorrelograms relative to their background (Fig. 8), and in peak-to-background ratios of pooled autocorrelation histograms (Fig. 9). Quantitatively, estimated saliences for  $F_0 = 160$  Hz were 2.5, 3.1, and 2.6 for 40, 60, and 80 dB SPL, respectively (Table 2). These high estimated saliences are in broad agreement with the strong pitches that are heard at these levels for similar harmonic complexes.

For single-formant vowels, salience estimates can be described in terms of three sets of intervals: those near the

fundamental period  $1/F_0$ , those related to components near  $F_1$ , and those produced by spontaneous activity. For low  $F_0$ s, the closely spaced components in the formant region produce fused interval peaks at  $1/F_1$  and its multiples. In the autocorrelation histograms of Fig. 9, intervals near  $1/F_0$  form the highest peak, other formant-related intervals form the smaller side-peaks, and spontaneous activity forms the baseline. At 40 dB SPL (Fig. 9A), salience is primarily the ratio of  $F_0$ -related intervals to spontaneous activity. As levels increase to 60 and 80 dB SPL (Fig. 9, B and C), the proportion of formant-related intervals increases, and salience is mainly determined by the ratio of  $F_0$ -related intervals to formant-related ones. The proportional increase in formant-related intervals causes the estimated salience for this stimulus to decline slightly at higher levels.

### First-order intervals versus all-order intervals

Some temporal models for pitch have utilized first-order interspike intervals (i.e., those between successive spikes) rather than all-order intervals (i.e., those between both successive and nonsuccessive spikes). Pooled distributions of all-order intervals (Fig. 9, A–C) were compared with those of first-order intervals (Fig. 9, D–F). For each of three

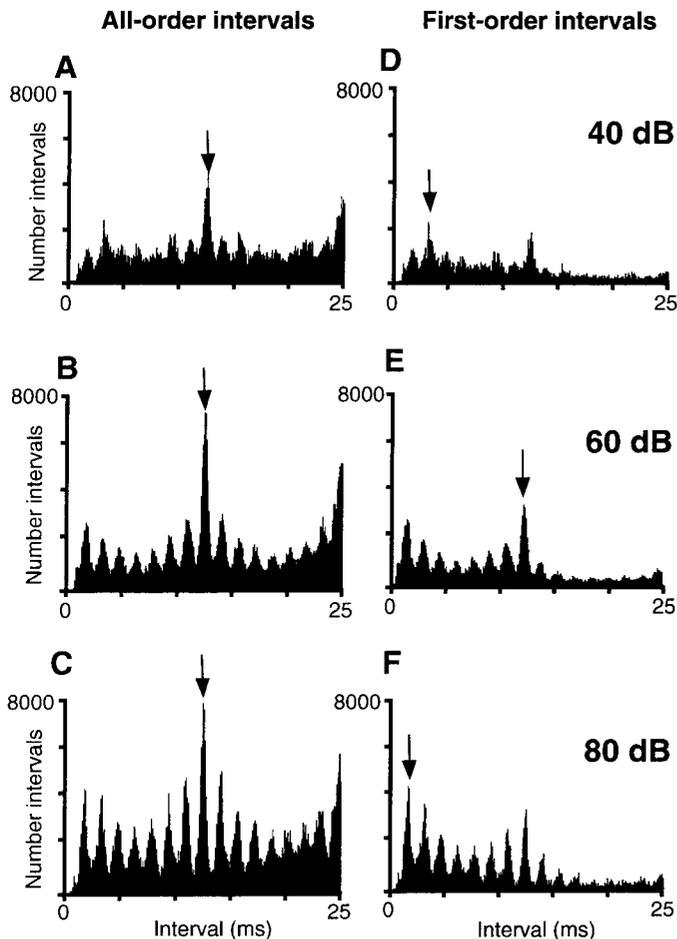


FIG. 9. Distributions of all-order vs. 1st-order intervals as a function of level. Arrows: pitch estimates. *A–C*: weighted pooled distributions of all-order intervals for a single-formant vowel,  $F_0 = 80$  Hz,  $F_1 = 640$  Hz, presented at 40, 60, and 80 dB SPL. *D–F*: weighted pooled distributions of 1st-order intervals for the same stimulus, levels, and set of fibers. *A* and *D*: 34 fibers; *B* and *E*: 60 fibers; *C* and *F*: 50 fibers.

stimulus levels, the same set of fibers was used to compute both interval distributions. For all levels the most frequent all-order interval is always at the fundamental  $1/F_0$ . In contrast, the most frequent first-order interval is dependent on stimulus level, moving from twice the formant period (3.1 ms) at 40 dB SPL to the fundamental period (12.5 ms) at 60 dB SPL, then back to the formant period (1.6 ms) at 80 dB SPL. Although positions of major peaks in the two distributions are the same, the relative heights of peaks in first-order interval distributions change with level, thereby altering pitch estimates. Thus, for this stimulus, pitch estimates based on all-order intervals are much more stable with respect to level than those based on first-order intervals. As a consequence, all further analyses here are based on all-order interspike intervals.

#### Pitch equivalence of spectrally diverse stimuli

Complex stimuli with very different power spectra can evoke the same low pitch. Such perceptual equivalences provide clues for which aspects of neural response are responsible for particular percepts. Neural responses to stimuli that evoke the same quality (pitch) can therefore be exam-

ined for commonalities and differences. Common aspects of neural response that covary with the percept (pitch) make plausible candidates for neural codes underlying that percept.

Waveforms, power spectra, and short-time autocorrelation functions for six stimuli that evoke the same pitch at 160 Hz are shown in Fig. 10. These stimuli are (*A*) a 160-Hz pure tone, (*B*) an AM tone with a low-frequency carrier ( $F_c = 640$  Hz,  $F_m = 160$  Hz), (*C*) a complex tone consisting of harmonics 6–12 of 160 Hz, (*D*) an AM tone with a high-frequency carrier ( $F_c = 6,400$  Hz,  $F_m = 160$  Hz), (*E*) a click train whose  $F_0$  is 160 Hz, and (*F*) an AM broadband noise ( $F_m = 160$  Hz, cutoff 20 kHz). The pure tone was included in the stimulus set because it is involved in many operational definitions of pitch, one definition being “that quality of a stimulus which can be matched to the frequency of a pure tone” (Burns and Feth 1983; Moore 1989). The AM tones and the equal-magnitude harmonic complex (Fig. 10, *B*, *C*, and *E*) have missing fundamentals, i.e., they have no spectral energy at 160 Hz. Their regions of greatest spectral energy are located in medium- to high-frequency regions, well above the fundamental.

Some of the stimuli, such as the pure tone and the AM tones, have their energy distributed within a relatively narrow spectral region, whereas others, such as the AM noise and the click train, have much more widely distributed energy. Except for the AM noise, all of the stimuli are harmonic. The five harmonic stimuli share the same  $F_0$ , the greatest common divisor of the frequencies of their components. The waveforms of harmonic stimuli show periodicities related to the fundamental, and their short-time autocorrelation functions show maxima at the fundamental period. Because noise is aperiodic, AM noise, strictly speaking, has no well-defined fundamental. Although its long-time spectrum is flat, the short-time spectrum of AM noise does have some harmonic structure. Its waveform envelope has a periodicity at the  $F_m$  and its short-time autocorrelation function has a maximum at the modulation period  $1/F_m$ . The autocorrelation functions of all six pitch-equivalent stimuli have common maxima at lags corresponding to their common pitch ( $1/160$  Hz = 6.25 ms).

Pooled all-order interval distributions for the six stimuli are shown in Fig. 11. Weighted, pooled autocorrelation histograms were smoothed and then normalized by dividing the number of intervals in each bin by the mean number of intervals per bin. In all six distributions the highest peak always lies close to the pitch period (arrows). The pitch estimates in Table 2 confirm this qualitative impression, with most estimates deviating  $<1\%$  from perceived pitches at either  $F_0$  or  $F_m$ . Those stimuli that produced larger deviations, the high- $F_c$  AM tone (4.2%) and the AM noise (1.3%), also have more diffuse pitches. In general, the forms of pooled interval distributions (Fig. 11) roughly resembled those of their respective stimulus autocorrelation functions (Fig. 10).

#### Neural correlates of pitch salience

Stimuli evoking the same pitch can differ in their pitch strength or salience (Fastl and Stoll 1979; Hermes 1993; Zwicker and Fastl 1990). Although all of the stimuli in Fig. 10 evoke the same low pitch, some of the pitches produced are much weaker and less definite than others. Specifically,

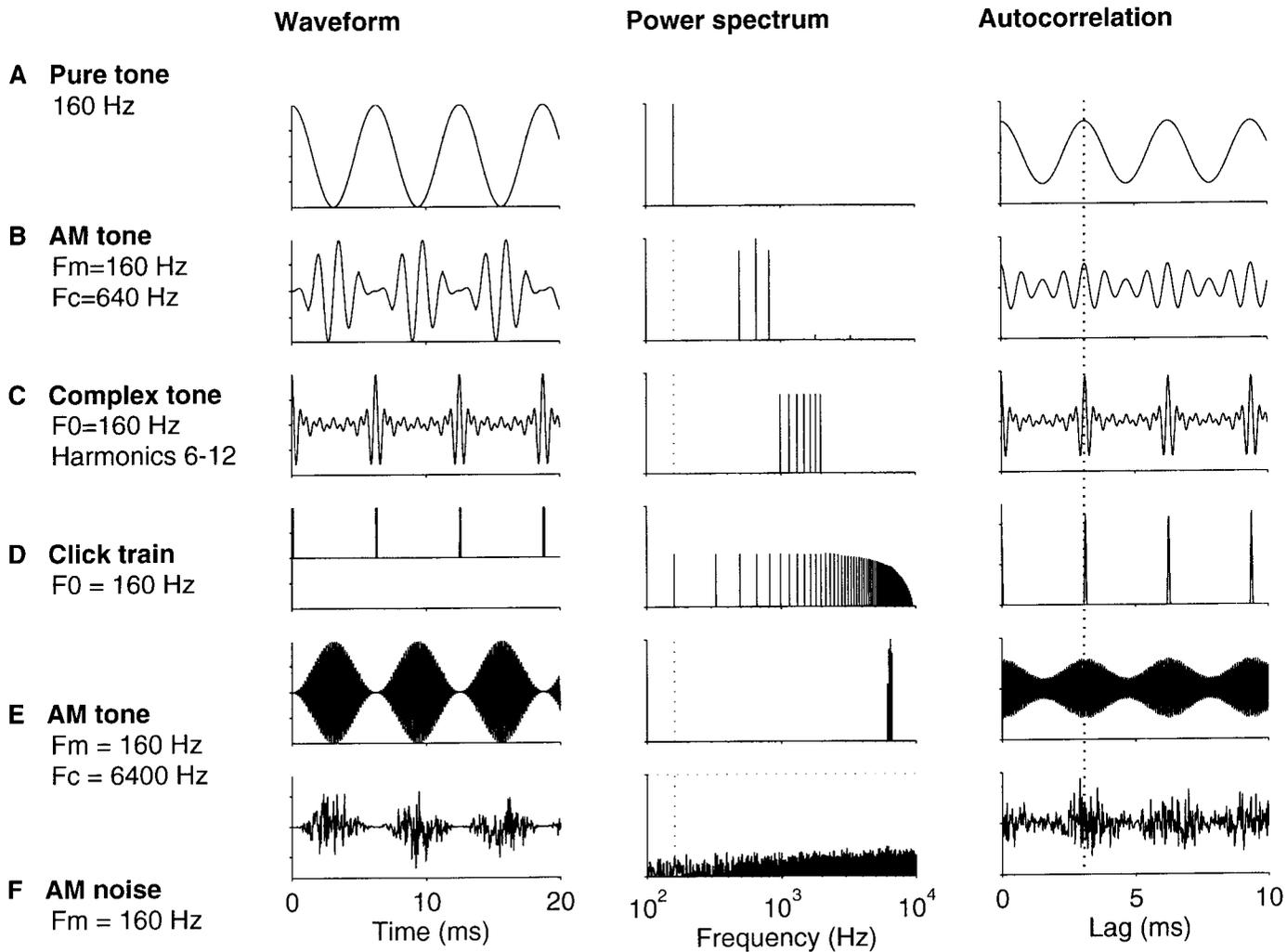


FIG. 10. A–F: waveforms, long-term power spectra, and short-term autocorrelation functions for 6 stimuli that evoke a pitch at 160 Hz. Magnitude range for power spectra is 50 dB. Dotted lines in power spectra: position of missing  $F_0$  for harmonic stimuli (B–D), and modulation frequency ( $F_m$ ) for amplitude-modulated (AM) noise (F). Dotted lines in autocorrelation functions:  $1/F_0$  (A–E) or  $1/F_m$  (F). Power spectra and autocorrelations were computed over 1.25 s for AM noise (F).

the top four stimuli (Fig. 10, A–D) evoke relatively strong pitches (Fastl and Stoll 1979), whereas the AM noise evokes a much weaker pitch (Burns and Viemeister 1976; Miller and Taylor 1948). The AM tone with the high-frequency carrier (Fig. 10E) lies just outside the region for which definite pitch matches can be made by human listeners (Fastl and Stoll 1979; Ritsma 1962a,b).

Examining the pooled autocorrelation histograms of Fig. 11, the peak-to-background ratios for the AM high- $F_c$  tone and the AM noise stimuli are noticeably smaller than those for the other four stimuli. This impression is confirmed in the computed salience estimates of Table 2. SDs were generally a few percent of the estimate. Stimuli that evoke strong pitches have estimated salience values ranging from 1.7 to 3.1, whereas the AM noise that produces a weak but discernible pitch range had an estimated salience of 1.3. The AM tone with the high-frequency carrier, with its ill-defined and barely perceptible pitch (Fastl and Stoll 1979), had a salience of 1.1. This would suggest that a peak-to-background ratio of roughly 1.2 or more is needed before a definite pitch can be detected. In addition to differences in peak-to-

background ratios, the interval peaks for these two stimuli are also broader and less well defined than those in the other histograms. There is thus a qualitative correspondence between the physiological salience measure and human rankings of pitch strength: those stimuli that evoke strong pitches produce relatively higher interval peaks, whereas those that evoke weak pitches produce flatter, more uniform interval distributions.

Despite this broad agreement between psychophysically observed saliences and their physiologically estimated counterparts, a discrepancy does exist between psychophysical rankings of pitch strength and those obtained from pooled interval distributions. In psychophysical studies, human listeners almost always judge a pure tone to have a stronger pitch than harmonic complexes with a missing fundamental (Fastl and Stoll 1979; Zwicker and Fastl 1990). However, examining the physiologically estimated saliences of Table 2, the pure tone had a considerably lower salience (1.7, Table 2) than most of the complex tones: the single-formant vowel (3.1), the low- $F_c$  AM tone (2.7), harmonics 6–12 (2.7), and the click train (2.6).

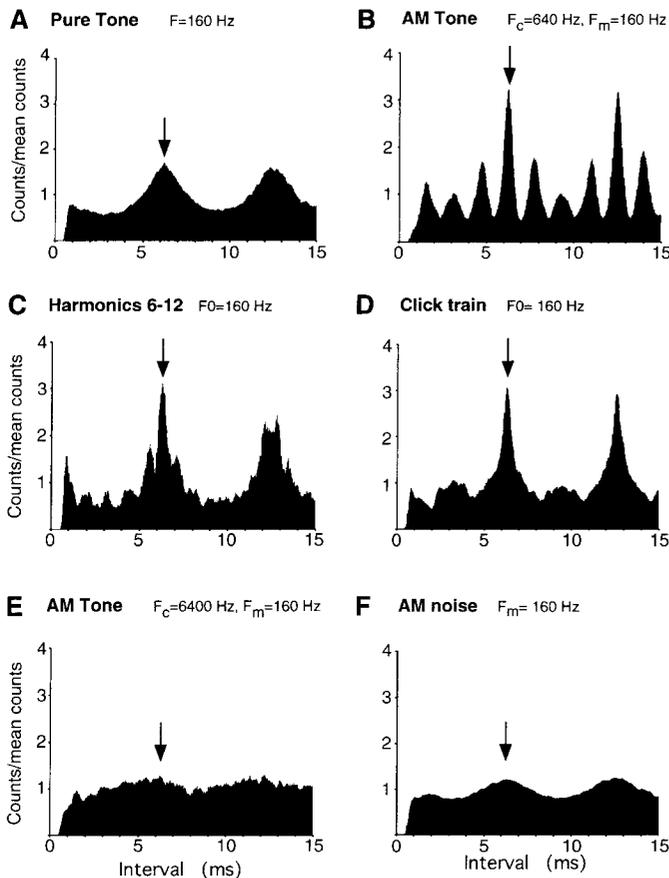


FIG. 11. Pooled autocorrelation histograms for 6 stimuli that evoke a pitch at 160 Hz. Histograms have been normalized to the mean number of counts per bin and smoothed. Arrows: position of the fundamental period  $1/F_0$  and/or modulation period  $1/F_m$  (6.25 ms). *A*: pure tone, 160 Hz. *B*: AM tone with a low-frequency carrier [carrier frequency ( $F_c$ ) = 640 Hz,  $F_m$  = 160 Hz]. *C*: harmonic complex (harmonics 6–12 of 160 Hz). *D*: AM tone with a high-frequency carrier ( $F_c$  = 6,400 Hz,  $F_m$  = 160 Hz). *E*: unipolar click train ( $F_0$  = 160 Hz). *F*: AM broadband noise with  $F_m$  = 160 Hz. All stimuli presented at 60 dB total SPL. Numbers of fibers for histograms *A*–*F*, respectively: 85, 54, 49, 56, 53, 50. Stimuli in *A*–*D* produce strong pitches, whereas stimuli in *E* and *F* produce weak ones.

The six stimuli of Fig. 10 have different power spectra, and consequently excite different regions of the auditory nerve. One means of assessing the contributions of different CF regions to physiologically estimated saliences is to examine peak-to-background ratios in interval histograms of single fibers. These ratios, or fiber saliences, are plotted against fiber CFs for each of the six stimuli in Fig. 12. Narrowband stimuli (pure tone, harmonics 6–12, AM tones) produced patterns of fiber saliences that broadly depended on fiber CF and stimulus spectrum. For example, fiber saliences for the 160-Hz pure tone (Fig. 12*A*) are generally highest for very low CFs, decreasing to unity as CFs increase beyond 1–2 kHz. For the low- $F_c$  AM tone (Fig. 12*B*) and harmonics 6–12 (Fig. 12*C*), the highest fiber saliences in each SR class are found for CFs bordering stimulus frequency regions containing intense components. Lower fiber saliences are found for CFs within those frequency regions, where ratios of  $F_0$ -related to component-related intervals decline as fibers are driven into saturation (as in fibers with CFs near vowel formants, Fig. 6*B*). Whether a similar notched pattern of highest fiber saliences exists for the high- $F_c$  AM tone (Fig.

12*E*) is less clear. In contrast to narrowband stimuli, broadband stimuli (click train, AM noise) produced elevated fiber saliences over wider ranges of CFs.

For any stimulus, low- and medium-SR fibers generally produce most of both the lowest and the highest fiber saliences. Fiber saliences of zero are usually the result of very low discharge rates, when no intervals at  $1/F_0$  may be produced. The highest fiber saliences tend to be produced by low- and medium-SR fibers, because a small number of  $F_0$ -related intervals can dramatically boost fiber saliences when there is little background activity (“spontaneous” discharges). Much of the scatter in distributions of fiber saliences (e.g., Fig. 12, *B*–*D*) is thus due to interfiber differences in rates of spontaneous and/or driven discharge. Pooling of single-fiber interval distributions averaged out many of these factors to produce much less variable estimates of pitch salience.

## DISCUSSION

### Neural correlates of pitch frequency

For a wide variety of stimuli with  $F_0$ s ranging from 80 to 500 Hz (Table 1), the most frequent interspike interval in

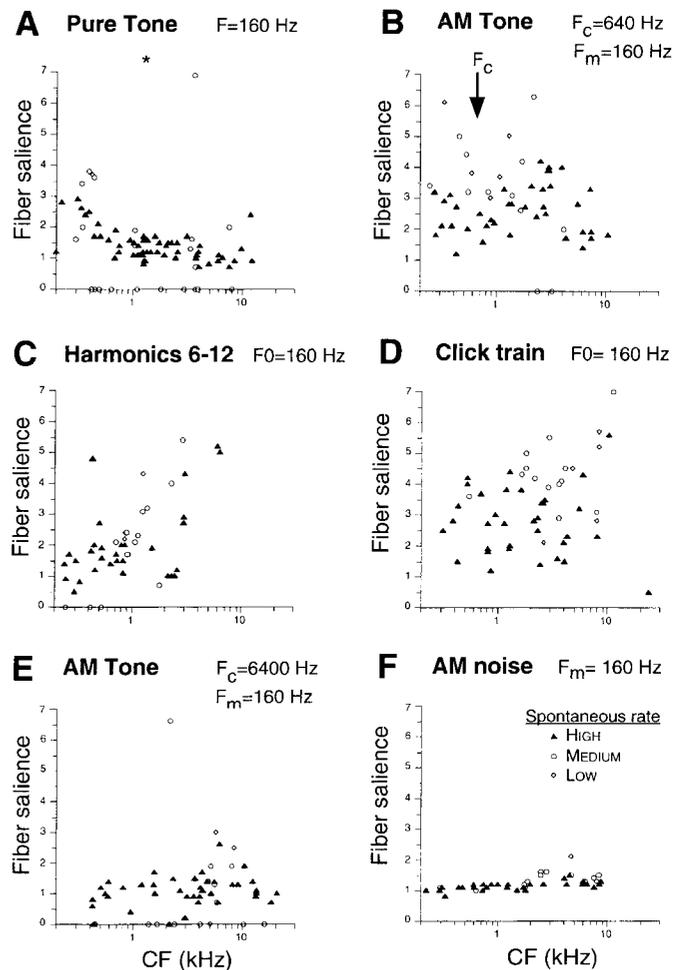


FIG. 12. Fiber saliences for 6 stimuli that evoke a pitch at 160 Hz. Fiber salience is the ratio of intervals per bin at 6.25 ms ( $1/160$  Hz) to mean intervals per bin in a smoothed single-fiber autocorrelation histogram. Asterisk in *A*: medium-spontaneous discharge rate (SR) fiber with a fiber salience of 9.5.

the auditory nerve corresponded closely to the low pitch that would be heard by a human listener (Table 2). The one exception to this rule that was found, involving alternating-polarity click trains, is discussed in the companion paper.

The results reported here are consistent with other published reports of the temporal discharges of ANFs in response to complex harmonic stimuli, which used interspike interval distributions (Evans 1978, 1983, 1986; Greenberg 1986; Horst et al. 1986, 1992; Javel 1980; Javel et al. 1988; Kim et al. 1990; Rose 1980; Voigt et al. 1982), period histograms (Delgutte 1980; Deng and Geisler 1987; Miller and Sachs 1984; Palmer et al. 1986; Young and Sachs 1979), and PST histograms (Feng et al. 1991; Secker-Walker and Searle 1990). Data from these studies indicate that interspike intervals and/or discharge periodicities related to the fundamental are present in the responses of many single ANFs. The present study is the first to explicitly combine interval distributions from many fibers to form an estimate of the population interval distribution of the auditory nerve, and to investigate correspondences between this population distribution and human pitch judgments. This study also employed a much more extensive set of stimuli than has been used previously.

Like human pitch judgments, interval-based pitch estimates were largely unaffected by stimulus level. In the case of single-formant vowels, the positions of interval peaks in a 2-octave range of stimulus  $F_0$ s (80–320 Hz) remained stable over a 40-dB range of levels (Figs. 8 and 9). These results are consistent with single-fiber interval distributions in response to click trains seen by Evans (1983). If interval information is used by central processors for the estimation of pitch, then relative invariance of pitch with respect to level is a direct consequence of the level-invariant nature of interspike interval distributions.

Although some level-dependent pitch shifts are observed psychophysically, these shifts are generally small (<1%) for harmonic complexes with many components (Zwicker and Fastl 1990). Thus systematic level-dependent pitch shifts would be expected to be minimal with the single-formant vowels used here, and none were seen. Small shifts in pooled interval distributions cannot be ruled out entirely, because means  $\pm$  SE for these pitch estimates were also on the order of 1%. Level-dependent pitch shifts could also be produced by combining interval information with that from other aspects of neural activity that do shift with level, such as rate-place patterns or relative spike latencies.

#### *Pitch of unresolved harmonics*

Spectral pattern models explain stimulus equivalence in terms of common harmonic relations in a central spectrum. A central pitch processor analyzes patterns of neural excitation in central auditory frequency maps in order to make pitch discriminations. For harmonic spectral patterns to be distinguished, the precise locations of spectral peaks must first be estimated. Harmonic complexes with higher-frequency partials, such as the harmonics 6–12 for fundamentals >180 Hz and the high- $F_c$  AM tone, contain only closely spaced harmonics that are not resolved psychophysically (Plomp 1964; Plomp and Mimpen 1968). Rate-place profiles of ANFs do not resolve these higher-frequency harmonics, and as a consequence, pure rate-place models are hard pressed to explain pitches evoked by these stimuli.

Interspike intervals related to pitches of unresolved harmonics were observed in individual ANFs for harmonics 6–12 for fundamentals >180 Hz and for the high- $F_c$  AM tone (see also Figs. 9 and 10 in the companion paper). Harmonics 6–12 of 160 Hz (Fig. 10C) are a borderline case, because the 6th harmonic would be barely discernible (Plomp 1964; Plomp and Mimpen 1968). As with harmonic stimuli consisting of resolved harmonics, the most frequent intervals in pooled interval distributions correspond to the low pitches that are heard. Pitches of stimuli consisting completely of unresolved harmonics are generally weak and diffuse, and the pooled interval distributions for these stimuli correspondingly show broad, shallow peaks that generate less precise pitch estimates.

A conceptual model (Fig. 13) is helpful in explaining these factors in the context of interval-based theories of pitch (Meddis and Hewitt 1991b; Moore 1989; van Noorden 1982). Figure 13A illustrates the case for sets of low-frequency (“resolved”) harmonics. The stimulus waveform is an AM tone consisting of harmonics 3, 4, and 5 ( $F_c - F_m$ ,  $F_c$ ,  $F_c + F_m$ ,  $F_0 = F_m$ ). The waveform is assumed to be filtered by the cochlea to produce completely resolved components in each frequency channel. Transduction at hair cells and synaptic transmission are represented by half-wave rectifications and low-pass filtering. The autocorrelation of the output at each idealized ANF approximates the all-order interval distribution for that fiber. Thus intervals corresponding to the period of the one resolved component and its subharmonics are produced in each of the channels. Because each frequency component is a harmonic of  $F_0$ , the fundamental period is an integer multiple of each component’s period. Although intervals at the fundamental period are no more frequent in any one channel than those associated with individual components, all resolved channels have intervals at the common fundamental period. Thus, when channel interval distributions are summed together, these intervals at the fundamental are most frequent in the pooled distribution. The resulting pooled autocorrelation function approximates the stimulus autocorrelation function (compare with Fig. 10B), which in turn closely resembles the pooled interval distribution that is observed physiologically (Fig. 11B). These similarities are obtained because the summed autocorrelations of sinusoidal components equal the autocorrelation of the sum of the components [i.e., the stimulus as a whole (Licklider 1951)].

Figure 13B illustrates the case for frequency channels whose center frequencies are well removed from the frequencies present in the stimulus. In this case there is no differential filtering of individual components, so that all components are attenuated roughly equally, leaving individual harmonics unresolved by the filter. The resulting waveform is rectified and low-pass filtered. Here each channel interval distribution will resemble the autocorrelation of the original waveform, with the most frequent interval being at the fundamental. If the waveform is attenuated sufficiently, only the interval peak at the fundamental will be produced. For the single-formant vowel (Figs. 5E and 6, A and C), interval distributions for fibers with CFs far removed from the formant show this pattern.

Figure 13C illustrates the case for a stimulus with high-frequency components, an AM tone similar to that of Fig. 10E. Because the components are spaced very closely rela-

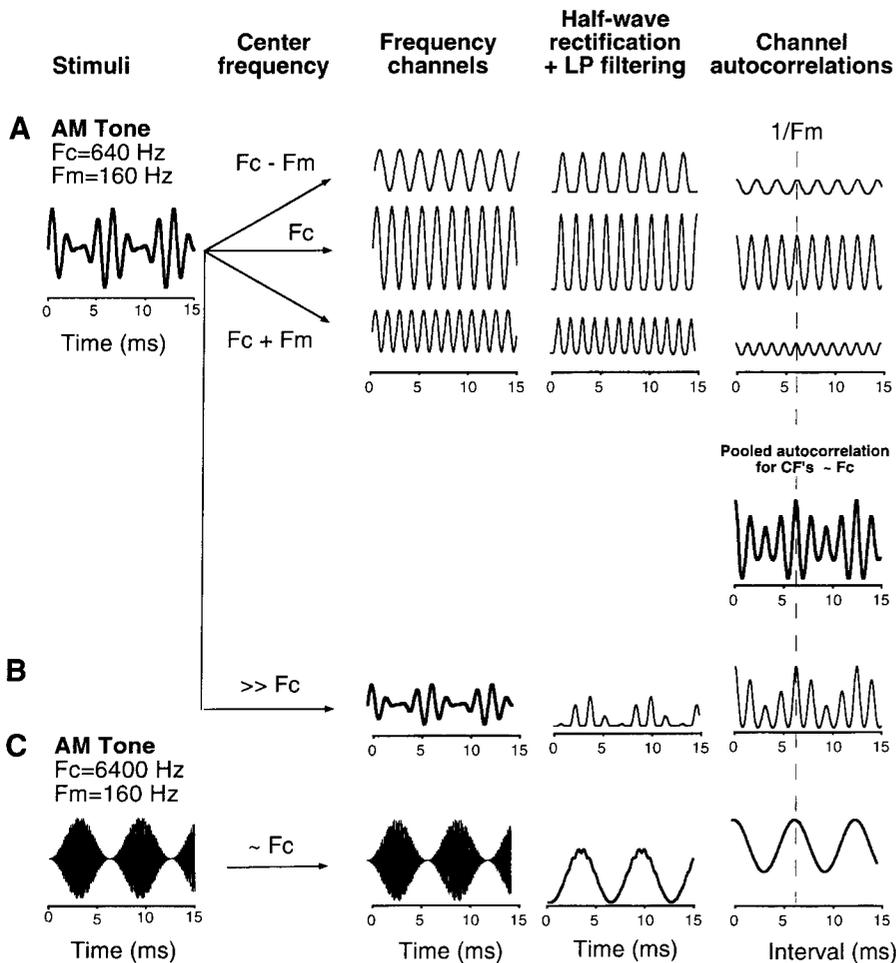


FIG. 13. Conceptual model for the generation of interspike intervals at the fundamental. Model stages (analogous physiological process), *left to right*: stimulus waveform (AM tones), filtered waveform (cochlear filtering), half-wave rectification (transduction) + low-pass (LP) filtering (synaptic transmission), short-term autocorrelation function (ANF all-order interspike interval distributions). *A and B*: harmonic AM tone with low-frequency carrier ( $F_c = 640$  Hz,  $F_m = F_0 = 160$  Hz). *A*: low-frequency harmonics resolved by narrowly tuned low-frequency band-pass filters produce channel autocorrelation maxima related to 1 frequency component. Pooling of low-frequency channel autocorrelations produces major pooled autocorrelation peak at the fundamental period ( $1/F_0 = 1/F_m = 6.25$  ms). *B*: higher-frequency channels with broader filters that do not resolve individual harmonics produce channel autocorrelations resembling the pooled autocorrelation of resolved harmonics (*A*). *C*: harmonic AM tone with high-frequency carrier ( $F_c = 6400$  Hz,  $F_m = 160$  Hz). As in *B*, broader filters of high-frequency channels do not resolve the individual harmonics, but rectification/low-pass filtering stage smears out high-frequency periodicities, leaving the waveform envelope that produces broad peaks at  $1/F_m$  in channel autocorrelation functions.

tive to the bandwidth of high-CF cochlear filters, there is no differential attenuation. Here rectification and low-pass filtering remove high-frequency components, leaving only the residue of envelope modulation. Thus for low frequencies the fine structure of the waveform is preserved, whereas for high frequencies only the envelope remains. The autocorrelation of this envelope has a broad peak at the fundamental period  $1/F_m$ .

Any combination of these processes produces pooled autocorrelation functions that have a maximum at the fundamental period, so that the distinction between resolved and "unresolved" harmonics is not an essential one for interval-based theories of pitch (Meddis and Hewitt 1991b; Moore 1989). Similarly, although different frequency channels can produce different channel autocorrelation functions, for harmonic stimuli the most frequent interval in the pooled distribution invariably corresponds to the fundamental. This potential difference between individual channel autocorrelations and the pooled autocorrelation underscores the importance of summing across all channels. Examination of the autocorrelation function in a particular frequency region, especially for inharmonic stimuli and/or analysis of single-unit responses in isolation, can therefore yield pitch estimates that are at variance with human pitch judgments. Models for pitch therefore need to take CF distributions into account (e.g., through peripheral weighting functions) in or-

der to obtain accurate and robust predictions (Yost and Hill 1978, 1979).

#### "Nonspectral" pitch

Neural discharge patterns associated with AM noise are of interest because these stimuli pose difficulties for pitch models that are based on analysis of long-term power spectra (Burns and Viemeister 1976; Moore and Glasberg 1986). Because AM noise has a flat, long-term power spectrum (Fig. 10F), there is no harmonic structure to be recognized if medium or long rate-integration periods ( $>20$  ms) are assumed for rate-place profiles. AM noise, nevertheless, produces a weak pitch near  $F_m$  that is capable of conveying a melody (Burns and Viemeister 1976, 1981). Whether the short-time spectrum of AM noise contains the requisite information for this nonspectral pitch has been a matter of some debate (Houtsma et al. 1980; Pierce et al. 1977).

In terms of temporal models, it has been pointed out that short-term temporal correlations near  $1/F_m$  are created through amplitude modulation, that these correlations are strengthened by half-wave rectification, and that therefore neural discharge periodicities at  $1/F_m$  should be produced that are amenable to analysis by temporal autocorrelation (Licklider 1951, 1959; Patterson et al. 1978). Human psychophysical data concerning the relation of modulation de-

tection and pitch discrimination thresholds for AM noise also appear to be consistent with a temporal autocorrelation analysis (Patterson et al. 1978).

Interspike intervals related to the modulation period  $1/F_m$  are indeed present in significant numbers of ANFs. Consequently, the most frequent intervals in the pooled distribution (Fig. 11F) are those near the modulation period  $1/F_m$ . In contrast to the difficulties of estimating such pitches from spectral profiles, the pitches of AM noise stimuli are simply explained in terms of interspike intervals. The relative weakness of these pitches is explained by the relatively lower proportion of intervals near  $1/F_m$  that are present in the auditory nerve.

### *Pitch equivalence*

Pitch equivalence, the production of similar pitches by complex tones of different spectral composition, is a fundamental feature of pitch perception. Perceptual equivalence of different stimuli permits the identification of commonalities in neural responses that may be involved in the production of the percept. The property of neural population responses that was similar for all of the pitch-equivalent stimuli studied was the duration of the most common interspike interval present. This is strong evidence in favor of an interspike interval code as a possible "neural code for pitch." If such a code is indeed used, perceptual equivalence would be a direct consequence of both the properties of interval codes and the action of a central processor that analyzes interval patterns.

Interspike interval codes are closely related to autocorrelation operations, because both essentially represent distributions of times between threshold crossings of an input signal (Licklider 1951, 1959). Pooled interval distributions were found to be very similar in form to the short-term autocorrelations of their respective stimuli, once transduction (rectification) and synaptic transmission (low-pass filtering) are taken into account (Figs. 10 and 11). Thus pooled interval distributions are autocorrelation-like neural representations of the stimulus, and analysis of these distributions effects an analysis of the stimulus autocorrelation function. Stimuli that produce similar periodicity pitches may do so because of basic properties inherent in autocorrelations, specifically in the harmonic structure that is inherent in sets of time intervals. Pitch equivalence would then be a direct consequence of a particular neural coding strategy rather than the result of complex spectral pattern equivalence classes that are either preestablished (Cohen et al. 1994; Dui fhuis et al. 1982; Gerson and Goldstein 1978; Terhardt et al. 1982b) or organized through experience (Bharucha 1991; Terhardt 1973).

Although diverse stimuli may have similar pitches, they can still differ in timbre, i.e., that quality of a sound that is neither loudness, pitch, duration, or location (Moore 1989). Timbre is a multidimensional quality influenced by spectral shape, amplitude dynamics, phase spectrum, and spectrotemporal structure. Those aspects of timbre related to periodicities  $<5$  kHz have neural correlates in temporal patterns of discharge (Lyon and Shamma 1995). For example, synthetic vowels can be discriminated purely on the basis of patterns of short intervals ( $<5$  ms) in pooled interval distributions (Cariani and Delgutte 1993, 1994; Palmer 1992). Subjectively,

the stimuli of Fig. 10 can be distinguished by their timbres, as can their corresponding patterns of minor peaks in pooled interval distributions (Fig. 11). On the other hand, pure tones, high- $F_c$  AM tones, and AM noise have strikingly different timbres, whereas the forms of their pooled interval distributions (Fig. 11, A, E, and F) bear considerable similarity. Clearly, shapes of pooled interval distributions alone cannot provide a full account for even these relatively simple timbre differences. Part of the difficulty is that pooled interval patterns produced by noise and high-frequency tones are not distinguishable from those produced by spontaneous activity. Thus more elaborate representations that incorporate both spatial and temporal structure (Lyon and Shamma 1995; Patterson and Akeroyd 1995; Patterson et al. 1995) are probably needed to explain the neural basis of timbre.

### *Neural correlates of pitch salience*

Pitch saliences estimated from peak-to-background ratios in pooled interval distributions generally corresponded well with rankings of pitch saliences in psychophysical studies (Fastl and Stoll 1979; Zwicker and Fastl 1990). Similar autocorrelation-based measures appear to account for the strengths of pitches produced by rippled noise (Shofner and Yost 1995; Yost et al. 1996). Where strong pitches would be expected from the psychophysics, high peak-to-background ratios were seen in interval distributions. Those stimuli that evoke weak or barely discernible pitches (AM noise, high- $F_c$  AM tones) had physiological salience measures of 1.1–1.3, whereas those that evoke strong pitches all had saliences  $>1.5$ . The physiological salience measure appeared to be consistent across stimuli, because stimuli with similar power spectra produced comparable salience estimates. For example, estimated saliences (Table 2) for the single-formant vowel (3.1) and the low- $F_c$  AM tone (2.7) were similar, as well as those estimated for harmonics 6–12 (2.7) and the click train (2.6).

In many respects the peak-to-background measure used here to estimate salience is analogous to others that have been used in autocorrelation-based models for pitch (Wightman 1973b; Yost and Hill 1978, 1979). Its principal shortcoming was its underestimation of the pure tone pitch relative to those of complex tones. Human listeners perceive pure tones to have pitches at least as strong as those produced by complex tones with the same  $F_0$ , yet the physiologically estimated salience of the 160-Hz pure tone (1.7, Table 2) was significantly lower than that of comparable complex tones (2.7–3.1). Possible reasons for this discrepancy could involve 1) underestimation of the contributions of low-CF fibers in the human auditory nerve, 2) dependence of pitch salience on the masses of interval peaks, rather than their peak-to-background ratios (interval peaks for pure tones are broader than for complex tones), 3) the presence of central matched filters favoring transmission of interspike intervals near  $1/CF$  to higher auditory stations, as is assumed in some auditory models (Srulovicz and Goldstein 1983), or 4) the action of a central pitch processor that combines both place and interval information.

Interval-based pitch representations provide possible explanations for why different stimuli produce stronger or weaker pitches. For example, why does an AM tone with a low-frequency carrier produce a much stronger pitch than

one with a high-frequency carrier? Peak-to-background ratios in pooled interval distributions are the consequence of how many fibers respond (spread of excitation), how strongly each fiber responds (discharge rate), and the proportion of pitch-related intervals that each fiber produces (single-fiber salience). Within each CF region, numbers of responding fibers and their discharge rates determine the magnitude of each region's contribution to the pooled interval distribution. Asymmetries in cochlear filters mean that stimuli with lower-frequency components produce a larger spread of excitation than higher-frequency ones. Thus the low- $F_c$  AM tone excites more fibers than its high- $F_c$  counterpart. Interval production is also affected by the frequencies of the components involved. Intervals related to pitch ( $1/F_m$ ) are produced by discharges that follow either the fine structure of the waveform or its envelope. Because phase-locking to fine structure declines as component frequencies increase, this mechanism produces most  $F_0$ -related intervals for sets of low-frequency components. In contrast, locking of discharges to the envelope produces most  $F_0$ -related intervals for sets of high-frequency components. Phase-locking to low-frequency harmonics produces proportionately more  $F_0$ -related intervals (higher fiber saliences) than phase-locking to the waveform envelope (compare magnitudes of fiber saliences in Fig. 11, *B* and *E*). All three factors (spread of excitation, discharge rates, fiber saliences) combine to produce dramatically higher peak-to-background ratios in pooled interval distributions for sets of lower-frequency harmonics (compare in Table 2 estimated saliences of the low- $F_c$  AM, 2.7, vs. its higher- $F_c$  counterpart, 1.1). Thus an interval-based theory of pitch predicts that, all other factors being equal, low-frequency harmonics should give rise to stronger pitches than higher-frequency ones. This is consistent with the existence of a low-frequency "dominance region" for pitch, whose neural correlates are investigated in the companion paper.

#### *Implications for temporal models for pitch*

The autocorrelation-based analysis utilized in this study is similar to models for pitch that combine the temporal response patterns of all frequency channels to compute a global temporal representation (Ghitza 1992; Licklider 1951, 1959; Lyon and Shamma 1995; Meddis and Hewitt 1991a,b; Moore 1989; Patterson 1987; Slaney and Lyon 1991, 1993; Yost et al. 1996). The results reported here are broadly compatible with those models. Many similarities and interrelations between the models are discussed elsewhere (Lyon and Shamma 1995; Meddis and Hewitt 1991a,b; Yost et al. 1996).

The operations used by Meddis and Hewitt (1991a,b) and Slaney and Lyon (1991, 1993) in their simulations parallel the data-processing operations used here. Autocorrelations of single-channel responses are summed to form pooled autocorrelation-like responses, from which pitch is then estimated. For each of their 128 simulated ANFs, Meddis and Hewitt used a multistage model of the auditory periphery to compute PST histograms (running discharge probabilities) from which single-fiber autocorrelation functions were then computed. Slaney and Lyon computed channel autocorrelation functions directly from a cochlear model. PST histograms and channel autocorrelation functions were then

displayed as "cochleograms" that are similar to the autocorrelation neurogram of Fig. 5*E*. All single-channel autocorrelation functions were added together to form "summary autocorrelation functions," wherein a pitch extractor finds the major maximum representing the shortest time interval, i.e., the highest pitch.

These models and the present study differed mainly in their respective use of simulated versus physiologically observed spike trains. Real spike trains reflect all of the underlying physiological processes at work, and therefore also incorporate (unknown or partially understood) factors that may not be represented in the models (e.g., certain cochlear nonlinearities, refractoriness). In the present study we also compiled interval distributions directly from spike trains rather than from simulated discharge probabilities or cochlear response functions. Temporal fine structure can be smeared out by such averaging of responses prior to counting of intervals. Although direct computation of intervals is not so critical in the auditory nerve, where responses are rigidly stimulus-locked, it becomes more important in auditory stations where response latencies can vary from presentation to presentation.

Some other temporal models for pitch use only first-order interspike intervals (Ghitza 1992; Moore 1989; van Noorden 1982). However, unlike pitch estimates based on all-order intervals, those based on first-order intervals can show large discontinuities with changes in level (Fig. 9). Because the low pitches that are heard are very stable over a wide range of levels, neural representations based on all-order intervals yield pitch estimates that are more in accord with patterns of human pitch judgments.

#### *Implications for place models of pitch*

In contrast to temporal models, place models for pitch use a central representation of the power spectrum of the stimulus. The central spectrum is formed by computing the relative amount of neural excitation at different places in a tonotopic map. The simplest rate-place models assume that the set of average discharge rates in a tonotopic map constitute the neural encoding of the central spectrum. Temporal-place processing schemes incorporate temporal discharge patterns into the computation of a central spectrum (Delgutte and Kiang 1984; Goldstein and Sruлович 1977; Miller and Sachs 1984; Seneff 1985; Sruлович and Goldstein 1983). Because complex tones have multiple peaks in their magnitude spectra, and their low pitch is related to the harmonic relations between these peaks, a pattern recognition process is needed to extract the  $F_0$ . This task of recognizing harmonic patterns in spectral excitation profiles can be realized via Fourier transform operations or autocorrelations on the spectra (Wightman 1973b; Yost and Hill 1979), analysis of subharmonic patterns (Terhardt et al. 1982a,b), optimal fitting to central templates (Sruлович and Goldstein 1983), or neural networks (Bharucha 1991; Cohen et al. 1994).

Because the same information is present in the autocorrelation function of a signal as exists in its power spectrum, differences between theories of auditory form discrimination based on spectral-pattern and temporal autocorrelation lie not in the mathematics, but in the means by which pitch mechanisms are realized neurally (Licklider 1951; Wightman 1973a,b). For every frequency domain operation there exists a corresponding one in the time domain.

Given the capabilities and limitations of auditory neurons, an analysis of periodicity pitch may be simpler to carry out in the time domain. A neural temporal processing strategy can use the harmonic structure inherent in time intervals and interspike interval representations to implement a harmonic analysis. For example, the fundamental of a flat harmonic complex (e.g., Fig. 10C) can be deduced by computing all of the subharmonics of stimulus components and finding the frequency common to the greatest number of subharmonics (Terhardt 1979; Terhardt et al. 1982b). Finding the most frequent interval in a population distribution implements an analogous temporal operation. Here each stimulus component generates shorter intervals that represent its own period, as well as longer ones that represent periods of subharmonics. When intervals produced by all stimulus components are added together, relative numbers of intervals roughly reflect relative numbers of components sharing the same subharmonic periods. Thus, in this example, the most frequent interval present corresponds to the subharmonic common to the greatest number of stimulus components. Rather than inferring absolute frequencies of stimulus harmonics and subharmonics from population rate profiles (Cohen et al. 1994), a temporally based central processor must instead determine which interspike interval is most prevalent in a population of auditory neurons (Licklider 1951). Which pitch estimation strategy is ultimately simpler to realize neurally depends on the particular neural processing capacities (connectivities, tunings, time delays, integration/coincidence windows) that are available at a given auditory station.

In comparison with rate-place profiles, interval-based representations are much more stable with respect to changes in stimulus level, particularly at higher levels (e.g., Fig. 9, A–C). Above 60 dB SPL, the discharge rates of a majority of ANFs are saturated, so that the abilities of population rate-place profiles to resolve individual harmonics may be degraded considerably as levels increase. Arguably, place-based representations that utilize interneural synchrony and lateral inhibition (Lyon and Shamma 1995) coupled with central compensatory mechanisms for “selective listening” and efferent control (Delgutte 1995) could potentially preserve spectral contrast at high levels. On the other hand, many critical questions also remain for purely temporal theories of pitch perception: whether the requisite temporal information is available at higher auditory stations, what forms such information might take, and what kinds of neural architectures would be needed to carry out an interval-based analysis.

### Dual nature of auditory representations

The interval statistics of populations of neurons constitute a general, distributed means of encoding, transmitting, and representing information (Cariani 1995; Perkell and Bullock 1968). Because all-order interspike interval distributions form autocorrelation-like time domain representations of the stimulus, they complement place-based, frequency domain representations. The distinguishability of high-frequency pure tones in the absence of temporal cues argues that central auditory mechanisms exist for using purely place-based information. The distinguishability of low pitches in the absence of place cues (unresolved harmonics, single-channel electrical stimulation) similarly argues that central auditory

mechanisms exist for using purely temporal information. The existence of multiple pitch mechanisms is therefore not at issue. What is not yet understood is how time- and frequency domain representations might be gracefully combined in central auditory stations (Licklider 1951, 1959) in a manner consistent with known neuroanatomy and neurophysiology. The close and pervasive correspondences between periodicity pitch and the pooled interval distribution of the auditory nerve suggest that mechanisms for processing temporal information may very well play a major role in producing the low pitches of complex tones. Exactly what the nature of that role is will remain unclear until the responses of central auditory neurons to complex sounds are better understood.

We acknowledge invaluable comments and criticisms that we received from N. Kiang, C. Brown, J. Melcher, C. Liberman, and M. Tramo, as well as all the support that was given by everyone at the Eaton Peabody Laboratory.

This work was supported by National Institute on Deafness and Other Communication Disorders Grants DC-00019, DC-00006, DC-02356, and DC-02258.

Address for reprint requests: P. Cariani, Eaton Peabody Laboratory, Massachusetts Eye and Ear Infirmary, 243 Charles St., Boston, MA 02114.

Received 10 October 1995; accepted in final form 6 May 1996.

### REFERENCES

- VON BEKESY, G. Hearing theories and complex sounds. *J. Acoust. Soc. Am.* 35: 588–601, 1963.
- BHARUCHA, J. J. Pitch, harmony and neural nets: a psychological perspective. In: *Connectionism and Music*, edited by P. Todd and G. Loy. Cambridge, MA: MIT Press, 1991, p. 84–99.
- BILSEN, R. A. AND RITSMA, R. J. Some parameters influencing the perceptibility of pitch. *J. Acoust. Soc. Am.* 47: 469–475, 1970.
- DE BOER, E. On the “residue” and auditory pitch perception. In: *Handbook of Sensory Physiology*, edited by W. D. Keidel and W. D. Neff. Berlin: Springer-Verlag, 1976, p. 479–583.
- BORING, E. G. *Sensation and Perception in the History of Experimental Psychology*. New York: Appleton-Century-Crofts, 1942.
- BOURK, T. R. *Electrical Responses of Neural Units in the Anteroventral Cochlear Nucleus of the Cat* (PhD thesis). Cambridge, MA: MIT, 1976.
- BURNS, E. M. AND FETH, L. L. Pitch of sinusoids and complex tones above 10 kHz. In: *Hearing—Physiological Bases and Psychophysics*, edited by R. Klinke and R. Hartmann. Berlin: Springer-Verlag, 1983, p. 327–333.
- BURNS, E. M. AND VIEMEISTER, N. F. Nonspectral pitch. *J. Acoust. Soc. Am.* 60: 863–869, 1976.
- BURNS, E. M. AND VIEMEISTER, N. F. Played-again SAM: further observations on the pitch of amplitude-modulated noise. *J. Acoust. Soc. Am.* 70: 1655–1660, 1981.
- CARIANI, P. As if time really mattered: temporal strategies for neural coding of sensory information. *Commun. Cognit. Artif. Intell.* 12: 161–229, 1995.
- CARIANI, P. AND DELGUTTE, B. Interspike interval distributions of auditory nerve fibers in response to variable-pitch complex stimuli. *Assoc. Res. Otolaryngol. Abstr.* 1992, p. 36.
- CARIANI, P. AND DELGUTTE, B. Interspike interval distributions of auditory nerve fibers in response to concurrent vowels with same and different fundamental frequencies. *Assoc. Res. Otolaryngol. Abstr.* 1993, p. 94.
- CARIANI, P. AND DELGUTTE, B. Transient changes in neural discharge patterns may enhance separation of concurrent vowels with different fundamental frequencies (Abstract). *J. Acoust. Soc. Am.* 95: 2842, 1994.
- DE CHEVEIGNÉ, A. A pitch perception model. *ICASSP Tokyo 86*: 897–900, 1986.
- CHUNG, S. H., RAYMOND, S. A., AND LETTVIN, J. Y. Multiple meaning in single visual units. *Brain Behav. Evol.* 3: 72–101, 1970.
- COHEN, M. A., GROSSBERG, S., AND WYSE, L. L. A spectral network model of pitch perception. *J. Acoust. Soc. Am.* 98: 862–879, 1994.
- CYNX, J. Periodicity pitch in a species of songbird, the European starling (*Sturnus vulgaris*). *Assoc. Res. Otolaryngol. Abstr.* 138, 1986.
- DELGUTTE, B. Representation of speech-like sounds in the discharge patterns of auditory-nerve fibers. *J. Acoust. Soc. Am.* 68: 843–857, 1980.

- DELGUTTE, B. Physiological models for basic auditory percepts. In: *Auditory Computation*, edited by H. Hawkins, T. McMullin, A. N. Popper, and R. R. Fay. New York: Springer-Verlag, 1995, p. 157–220.
- DELGUTTE, B. Auditory neural processing of speech. In: *The Handbook of the Phonetic Sciences*, edited by W. J. Hardcastle and J. Laver. Oxford, UK: Blackwell, in press.
- DELGUTTE, B. AND CARIANI, P. Coding of the pitch of harmonic and inharmonic complex tones in the interspike intervals of auditory nerve fibers. In: *The Processing of Speech*, edited by M. E. H. Schouten. Berlin: Mouton-DeGruyter, 1992, p. 37–45.
- DELGUTTE, B. AND KIANG, N. Y. S. Speech coding in the auditory nerve. I. Vowel-like sounds. *J. Acoust. Soc. Am.* 75: 866–878, 1984.
- DENG, L. AND GEISLER, S. Responses of auditory nerve fibers to multiple tone complexes. *J. Acoust. Soc. Am.* 82: 1989–2000, 1987.
- DUIFHUIS, H., WILLEMS, L. F., AND SLUYTER, R. J. Measurement of pitch in speech: an implementation of Goldstein's theory of pitch perception. *J. Acoust. Soc. Am.* 71: 1568–1580, 1982.
- EVANS, E. F. Place and time coding of frequency in the peripheral auditory system: some physiological pros and cons. *Audiology* 17: 369–420, 1978.
- EVANS, E. F. Pitch and cochlear nerve fibre temporal discharge patterns. In: *Hearing: Physiological Bases and Psychophysics*, edited by R. Klinke and R. Hartmann. Berlin: Springer-Verlag, 1983, p. 140–146.
- EVANS, E. F. Cochlear nerve fibre temporal discharge patterns, cochlear frequency selectivity and the dominant region for pitch. In: *Auditory Frequency Selectivity*, edited by B. C. J. Moore and R. D. Patterson. New York: Plenum, 1986, p. 253–264.
- FASTL, H. AND STOLL, G. Scaling of pitch strength. *Hear. Res.* 1: 293–301, 1979.
- FAY, R. R. Perception of amplitude-modulated auditory signals by the goldfish. *J. Acoust. Soc. Am.* 53: 660–666, 1972.
- FAY, R. R., YOST, W. A., AND COOMBS, S. Psychophysics and neurophysiology of repetition noise processing in a vertebrate auditory system. *Hear. Res.* 12: 31–55, 1983.
- FENG, A. S., HALL, J. C., AND SIDDIQUE, S. Coding of temporal parameters of complex sounds by frog auditory nerve fibers. *J. Neurophysiol.* 65: 424–445, 1991.
- GERSON, A. AND GOLDSTEIN, J. L. Evidence for a general template in central optimal processing for pitch of complex tones. *J. Acoust. Soc. Am.* 63: 498–510, 1978.
- GHITZA, O. Auditory nerve representation as a basis for speech processing. In: *Advances in Speech Signal Processing*, edited by S. Furui and M. M. Sondhi. New York: Dekker, 1992, p. 453–485.
- GOLDSTEIN, J. L. An optimum processor theory for the central formation of the pitch of complex tones. *J. Acoust. Soc. Am.* 54: 1496–1516, 1973.
- GOLDSTEIN, J. L. AND SRULOVICZ, P. Auditory-nerve spike intervals as an adequate basis for aural frequency measurement. In: *Psychophysics and Physiology of Hearing*, edited by E. F. Evans and J. P. Wilson. London: Academic, 1977.
- GREENBERG, S. Possible role of low and medium spontaneous rate cochlear nerve fibers in the encoding of waveform periodicity. In: *Auditory Frequency Selectivity*, edited by B. C. J. Moore and R. D. Patterson. New York: Plenum, 1986, p. 241–251.
- GREENBERG, S. AND RHODE, W. S. Periodicity coding in cochlear nerve and ventral cochlear nucleus. In: *Auditory Processing of Complex Sounds*, edited by W. A. Yost and C. S. Watson. Hillsdale, NJ: Erlbaum, 1987, p. 225–236.
- GREENWOOD, D. D. A cochlear frequency-position function for several species—29 years later. *J. Acoust. Soc. Am.* 87: 2592–2604, 1990.
- HEFNER, H. AND WHITFIELD, I. C. Perception of the missing fundamental in cats. *J. Acoust. Soc. Am.* 59: 915–919, 1976.
- HERMES, D. J. Pitch analysis. In: *Visual Representations of Speech Signals*, edited by S. Beet and M. Crawford. Chichester, UK: Wiley, 1993, p. 3–25.
- HIRAHARA, T., CARIANI, P., AND DELGUTTE, B. Representation of low-frequency vowel formants in the auditory nerve. *Assoc. Res. Otolaryngol. Abstr.* 80, 1996.
- HORST, J. W., JAVEL, E., AND FARLEY, G. R. Coding of spectral fine structure in the auditory nerve. I. Fourier analysis of period and interspike interval histograms. *J. Acoust. Soc. Am.* 79: 398–416, 1986.
- HORST, J. W., JAVEL, E., AND FARLEY, G. R. Coding of fundamental frequency in auditory nerve fibers: effects of signal level and phase spectrum. In: *The Processing of Speech*, edited by M. E. H. Schouten. Berlin: Mouton-DeGruyter, 1992, p. 29–36.
- HOUTSMA, A. J. M. AND SMURZYNSKI, J. Pitch identification and discrimination for complex tones with many harmonics. *J. Acoust. Soc. Am.* 87: 304–310, 1990.
- HOUTSMA, A. J. M., WICKE, R. W., AND ORDUBADI, A. Pitch of amplitude modulated low-pass noise and predictions by temporal and spectral theories. *J. Acoust. Soc. Am.* 67: 1312–1322, 1980.
- JAVEL, E. Coding of AM tones in the chinchilla auditory nerve: implications for the pitch of complex tones. *J. Acoust. Soc. Am.* 68: 133–146, 1980.
- JAVEL, E., MCGEE, J. A., HORST, W., AND FARLEY, G. R. Temporal mechanisms in auditory stimulus coding. In: *Auditory Function: Neurobiological Bases of Hearing*, edited by G. M. Edelman, W. E. Gall, and W. M. Cowan. New York: Wiley, 1988, p. 515–558.
- TEN KATE, J. AND VAN BEKKUM, M. F. Synchrony-dependent autocorrelation in eight-nerve-fiber response to rippled noise. *J. Acoust. Soc. Am.* 84: 2092–2102, 1988.
- KIANG, N. Y. S., EDDINGTON, D. K., AND DELGUTTE, B. Fundamental considerations in designing auditory implants. *Acta Otolaryngol.* 87: 294–218, 1979.
- KIANG, N. Y. S. AND MOXON, E. C. Physiological considerations in artificial stimulation of the inner ear. *Ann. Otol. Rhinol. Laryngol.* 81: 714–731, 1972.
- KIANG, N. Y. S., MOXON, E. C., AND LEVINE, R. A. Auditory-nerve activity in cats with normal and abnormal cochleas. In: *Sensorineural Hearing Loss*, edited by G. E. W. Wolstenholme and J. Knight. London: Churchill, 1970, p. 241–273.
- KIANG, N. Y. S., WATANABE, T., THOMAS, E. C., AND CLARK, L. F. *Discharge Patterns of Single Fibers in the Cat's Auditory Nerve*. Cambridge, MA: MIT Press, 1965.
- KIM, D. O., SIRIANNI, J. G., AND CHANG, S. O. Responses of DCN-PVCN neurons and auditory nerve fibers in unanesthetized decerebrate cats to AM and pure tones: analysis with autocorrelation/power-spectrum. *Hear. Res.* 45: 95–113, 1990.
- LANGE, F. H. *Correlation Techniques*. Princeton, NJ: Van Nostrand, 1967.
- LAZZARO, J. AND MEAD, C. Silicon modeling of pitch perception. *Proc. Natl. Acad. Sci. USA* 86: 9597–9601, 1989.
- LIBERMAN, C. E. AND KIANG, N. Y. S. Acoustic trauma in cats. *Acta Otolaryngol. Suppl.* 358: 1–63, 1978.
- LIBERMAN, M. C. The cochlear frequency map for the cat: labeling auditory-nerve fibers of known characteristic frequency. *J. Acoust. Soc. Am.* 72: 1441–1449, 1982.
- LICKLIDER, J. C. R. A duplex theory of pitch perception. *Experientia* VII: 128–134, 1951.
- LICKLIDER, J. C. R. "Periodicity" pitch and "place" pitch (Abstract). *J. Acoust. Soc. Am.* 26: 945, 1954.
- LICKLIDER, J. C. R. Auditory frequency analysis. In: *Information Theory*, edited by C. Cherry. London: Butterworths, 1956, p. 253–268.
- LICKLIDER, J. C. R. Three auditory theories. In: *Psychology: A Study of a Science. Study I. Conceptual and Systematic*, edited by S. Koch. New York: McGraw-Hill, 1959, p. 41–144.
- LYON, R. AND SHAMMA, S. Auditory representations of timbre and pitch. In: *Auditory Computation*, edited by H. Hawkins, T. McMullin, A. N. Popper, and R. R. Fay. New York: Springer-Verlag, 1995, p. 517.
- MEDDIS, R. AND HEWITT, M. J. Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I. Pitch identification. *J. Acoust. Soc. Am.* 89: 2866–2882, 1991a.
- MEDDIS, R. AND HEWITT, M. J. Virtual pitch and phase sensitivity of a computer model of the auditory periphery. II. Phase sensitivity. *J. Acoust. Soc. Am.* 89: 2883–2894, 1991b.
- MELCHER, J. R. *The Cellular Generators of the Brainstem Auditory Evoked Potential* (PhD thesis). Cambridge, MA: MIT, 1993.
- MILLER, G. A. AND TAYLOR, W. G. The perception of repeated bursts of noise. *J. Acoust. Soc. Am.* 20: 171–182, 1948.
- MILLER, M. I. AND SACHS, M. B. Representation of voice pitch in discharge patterns of auditory-nerve fibers. *Hear. Res.* 14: 257–279, 1984.
- MÖLLER, A. R. The use of correlation analysis in processing neuroelectric data. In: *Progress in Brain Research, Computers and Brains*, edited by J. P. Schade and J. Smith. Amsterdam: Elsevier, 1970, vol. 33, p. 87–99.
- MOORE, B. C. J. *Introduction to the Psychology of Hearing* (3rd ed.). London: Academic, 1989.
- MOORE, B. C. J. Pitch perception in silicon. *Nature Lond.* 343: 114–115, 1990.
- MOORE, B. C. J. AND GLASBERG, B. R. The role of frequency selectivity in the perception of loudness, pitch and time. In: *Frequency Selectivity in Hearing*, edited by B. C. J. Moore. London: Academic, 1986, p. 251–308.

- VAN NOORDEN, L. Two channel pitch perception. In: *Music, Mind and Brain*, edited by M. Clynes. New York: Plenum, 1982, p. 251–269.
- NORDMARK, J. O. Time and frequency analysis. In: *Foundations of Modern Auditory Theory*, edited by J. V. Tobias. New York: Academic, 1970, p. 57–83.
- NORDMARK, J. O. Frequency and periodicity analysis. In: *Handbook of Perception*, edited by E. C. Carterette and M. P. Friedman. New York: Academic, 1978, p. 243–282.
- PALMER, A. R. Segregation of the responses to paired vowels in the auditory nerve of the guinea pig using autocorrelation. In: *The Auditory Processing of Speech*, edited by M. E. H. Schouten. Berlin: Mouton de Gruyter, 1992, p. 115–124.
- PALMER, A. R., WINTER, I. M., AND DARWIN, C. J. The representation of steady-state vowel sounds in the temporal discharge patterns of the guinea pig cochlear nerve and primarylike cochlear nucleus neurons. *J. Acoust. Soc. Am.* 79: 100–113, 1986.
- PATTERSON, R. D. A pulse-ribbon model of peripheral auditory processing. In: *Auditory Processing of Complex Sounds*, edited by W. A. Yost and C. S. Watson. New Jersey: Erlbaum, 1987, p. 167–179.
- PATTERSON, R. D. AND AKEROYD, M. A. Time-interval patterns and sound quality. In: *Advances in Hearing Research*, edited by G. A. Manley, G. M. Klump, C. Köppl, H. Fastl, and H. Oeckinghaus. Singapore: World Scientific, 1995, p. 545–556.
- PATTERSON, R. D., ALLERHAND, M. H., AND GIGUERE, C. Time-domain modeling of peripheral auditory processing: a modular architecture and a software platform. *J. Acoust. Soc. Am.* 98: 1890–1894, 1995.
- PATTERSON, R. D., JOHNSON-DAVIES, D., AND MILROY, R. Amplitude-modulated noise: the detection of modulation versus the detection of modulation rate. *J. Acoust. Soc. Am.* 63: 1904–1911, 1978.
- PERKELL, D. H. AND BULLOCK, T. H. Neural Coding. Neurosciences Research Program Bulletin, Vol. 6, No. 3. Neurosciences Research Program, 1968.
- PIERCE, J. R., LIPES, R., AND CHEETHAM, C. Uncertainty concerning the direct use of time information in hearing: place clues in white-spectra stimuli. *J. Acoust. Soc. Am.* 61: 1609–1621, 1977.
- PLOMP, R. The ear as a frequency analyzer. *J. Acoust. Soc. Am.* 36: 1628–1636, 1964.
- PLOMP, R. Pitch of complex tones. *J. Acoust. Soc. Am.* 41: 1526–1533, 1967.
- PLOMP, R. *Aspects of Tone Sensation*. London: Academic, 1976.
- PLOMP, R. AND MIMPEN, A. M. The ear as a frequency analyzer. II. *J. Acoust. Soc. Am.* 43: 764–767, 1968.
- RITSMA, R. J. Existence region of the tonal residue. I. *J. Acoust. Soc. Am.* 34: 1224–1229, 1962a.
- RITSMA, R. J. Existence region of the tonal residue. II. *J. Acoust. Soc. Am.* 35: 1241–1245, 1962b.
- RODIECK, R., KIANG, N. Y. S., AND GERSTEIN, G. L. Some quantitative methods for the study of spontaneous activity of single neurons. *Biophys. J.* 2: 351–368, 1962.
- RODIECK, R. W. Maintained activity of cat retinal ganglion cells. *J. Neurophysiol.* 30: 1043–1071, 1967.
- ROSE, J. E. Neural correlates of some psychoacoustical experiences. In: *Neural Mechanisms of Behavior*, edited by D. McFadden. New York: Springer-Verlag, 1980, p. 1–33.
- ROSE, J. E., BRUGGE, J. R., ANDERSON, D. J., AND HIND, J. E. Phase-locked response to low-frequency tones in single auditory nerve fibers of the squirrel monkey. *J. Neurophysiol.* 30: 769–793, 1967.
- ROSE, J. E., BRUGGE, J. F., ANDERSON, D. J., AND HIND, J. E. Some possible neural correlates of combination tones. *J. Neurophysiol.* 32: 769–793, 1969.
- SECKER-WALKER, H. E. AND SEARLE, C. L. Time-domain analysis of auditory-nerve-fiber firing rates. *J. Acoust. Soc. Am.* 88: 1427–1436, 1990.
- SENEFF, S. *Pitch and Spectral Analysis of Speech Based on an Auditory Synchrony Model* (PhD thesis). Cambridge, MA: MIT, 1985.
- SHOFNER, W. P. AND YOST, W. A. Repetition pitch: auditory processing of rippled noise in the chinchilla. In: *Advances in Hearing Research*, edited by G. A. Manley, G. M. Klump, C. Köppl, H. Fastl, and H. Oeckinghaus. Singapore: World Scientific, 1995, p. 477–487.
- SIEBERT, W. M. Frequency discrimination in the auditory system: place or periodicity mechanisms? *Proc. IEEE* 58: 723–730, 1970.
- SINEX, D. G. AND GEISLER, D. Auditory nerve fiber responses to frequency-modulated tones. *Hear. Res.* 4: 127–148, 1981.
- SLANEY, M. AND LYON, R. F. Apple Hearing Demo Reel. Videotape and report. Cupertino, CA: Apple Computer, 1991.
- SLANEY, M. AND LYON, R. F. On the importance of time—a temporal representation of sound. In: *Visual Representations of Speech Signals*, edited by M. Cooke, S. Beet, and M. Crawford. New York: Wiley, 1993, p. 95–118.
- SMALL, A. M. Periodicity pitch. In: *Foundations of Modern Auditory Theory*, edited by J. V. Tobias. New York: Academic, 1970, p. 3–54.
- SRULOVICZ, P. AND GOLDSTEIN, J. L. A central spectrum model: a synthesis of auditory-nerve timing and place cues in monaural communication of frequency spectrum. *J. Acoust. Soc. Am.* 73: 1266–1276, 1983.
- TERHARDT, E. Pitch, consonance, and harmony. *J. Acoust. Soc. Am.* 55: 1061–1069, 1973.
- TERHARDT, E. Calculating virtual pitch. *Hear. Res.* 1: 155–182, 1979.
- TERHARDT, E., STOLL, G., AND SEEWANN, M. Pitch of complex signals according to virtual-pitch theory: test, examples, and predictions. *J. Acoust. Soc. Am.* 71: 671–678, 1982a.
- TERHARDT, E., STOLL, G., AND SEEWANN, M. Algorithm for extraction of pitch and pitch salience from complex tonal signals. *J. Acoust. Soc. Am.* 71: 679–688, 1982b.
- TOMLINSON, R. W. W. AND SCHWARTZ, D. W. F. Perception of the missing fundamental in nonhuman primates. *J. Acoust. Soc. Am.* 84: 560–565, 1988.
- VOIGT, H. F., SACHS, M. B., AND YOUNG, E. D. Representation of whispered vowels in discharge patterns of auditory-nerve fibers. *Hear. Res.* 8: 49–58, 1982.
- WARREN, R. M. *Auditory Perception: A New Synthesis*. New York: Pergamon, 1982.
- WEVER, E. G. *Theory of Hearing*. New York: Wiley, 1949.
- WIGHTMAN, F. L. Pitch and stimulus fine structure. *J. Acoust. Soc. Am.* 54: 397–406, 1973a.
- WIGHTMAN, F. L. The pattern-transformation model of pitch. *J. Acoust. Soc. Am.* 54: 407–416, 1973b.
- YOST, W. A. AND HILL, R. Strength of the pitches associated with ripple noise. *J. Acoust. Soc. Am.* 64: 485–492, 1978.
- YOST, W. A. AND HILL, R. Models of the pitch and pitch strength of rippled noise. *J. Acoust. Soc. Am.* 66: 400–410, 1979.
- YOST, W. A., PATTERSON, R., AND SHEFT, S. A time domain description for the pitch strength of rippled noise. *J. Acoust. Soc. Am.* 99: 1066–1078, 1996.
- YOUNG, E. D. AND SACHS, M. B. Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory nerve fibers. *J. Acoust. Soc. Am.* 66: 1381–1403, 1979.
- ZWICKER, E. AND FASTL, H. *Psychoacoustics: Facts and Models*. Berlin: Springer-Verlag, 1990.

# Neural Correlates of the Pitch of Complex Tones. II. Pitch Shift, Pitch Ambiguity, Phase Invariance, Pitch Circularity, Rate Pitch, and the Dominance Region for Pitch

PETER A. CARIANI AND BERTRAND DELGUTTE

*Department of Otolaryngology, Harvard Medical School, and Eaton-Peabody Laboratory for Auditory Physiology, Massachusetts Eye and Ear Infirmary, Boston, Massachusetts 02114*

## SUMMARY AND CONCLUSIONS

1. The neural correlates of low pitches produced by complex tones were studied by analyzing temporal discharge patterns of auditory nerve fibers in Dial-anesthetized cats. In the previous paper it was observed that, for harmonic stimuli, the most frequent interspike interval present in the population of auditory nerve fibers always corresponded to the perceived pitch (predominant interval hypothesis). The fraction of these most frequent intervals relative to the total number of intervals qualitatively corresponded to strength (salience) of the low pitches that are heard.

2. This paper addresses the neural correlates of stimuli that produce more complex patterns of pitch judgments, such as shifts in pitch and multiple pitches. Correlates of pitch shift and pitch ambiguity were investigated with the use of harmonic and inharmonic amplitude-modulated (AM) tones varying either in carrier frequency or modulation frequency. Pitches estimated from the pooled interval distributions showed shifts corresponding to "the first effect of pitch shift" (de Boer's rule) that is observed psychophysically. Pooled interval distributions in response to inharmonic stimulus segments showed multiple maxima corresponding to the multiple pitches heard by human listeners (pitch ambiguity).

3. AM and quasi-frequency-modulated tones with low carrier frequencies produce very similar patterns of pitch judgments, despite great differences in their phase spectra and waveform envelopes. Pitches estimated from pooled interval distributions were remarkably similar for the two kinds of stimuli, consistent with the psychophysically observed phase invariance of pitches produced by sets of low-frequency components.

4. Trains of clicks having uniform and alternating polarities were used to investigate the relation between pitches associated with periodicity and those associated with click rate. For bipolar click trains, where periodicity and rate coincide, physiologically estimated pitches closely follow the fundamental period. This corresponds to the pitch at the fundamental frequency ( $F_0$ ) that is heard. For alternating click trains, where rate and periodicity do not coincide, physiologically estimated pitches always closely followed the fundamental period. Although these pitch estimates corresponded to periodicity pitches that are heard for  $F_0s > 150$  Hz, they did not correspond to the rate pitches that are heard for  $F_0s < 150$  Hz. The predominant interval hypothesis thus failed to predict rate pitch.

5. When alternating-polarity click trains are high-pass filtered, rate pitches are strengthened and can also be heard at  $F_0s > 150$  Hz. Pitches for high-pass-filtered alternating click trains were estimated from pooled responses of fibers with characteristic frequencies (CFs)  $> 2$  kHz. Roughly equal numbers of intervals at  $1/\text{rate}$  and  $1/F_0$  were found for all  $F_0s$  studied, from 80 to 160 Hz, producing pitch estimates consistent with the rate pitches that are heard after high-pass filtering. The existence region for rate pitch also coincided with the presence of clear periodicities related to the click

rate in pooled peristimulus time histograms. These periodicities were strongest for ensembles of fibers with CFs  $> 2$  kHz, where there is widespread synchrony of discharges across many fibers.

6. The "dominance region for pitch" was studied with the use of two harmonic complexes consisting of harmonics 3–5 of one  $F_0$  and harmonics 6–12 of another fundamental 20% higher in frequency. When the complexes were presented individually, pitch estimates were always close to the fundamental of the complex. When the complexes were presented concurrently, pitch estimates always followed the fundamental of harmonics 3–5 for  $F_0s$  of 150–480 Hz. For  $F_0s$  of 125–150 Hz, pitch estimates followed one or the other fundamental, and for  $F_0s < 125$  Hz, pitch estimates followed the fundamental of harmonics 6–12. The results are generally consistent with the ranges of component frequencies (500–1,000 Hz) that are found psychophysically to be most important for low pitch.

7. Taken as a whole, the physiological data presented here provide strong evidence that interspike interval information plays an important role in the perception of the low pitch of complex tones. The predominant interval hypothesis for pitch yields surprisingly robust, comprehensive, and unified explanations for a very wide range of pitch phenomena: the missing fundamental, pitch invariance with respect to level, pitch equivalence of spectrally diverse stimuli, the pitch of unresolved harmonics, the pitch of AM noise, pitch salience, pitch shift of inharmonic AM tones, pitch ambiguity, phase insensitivity of pitch, and the dominance region for pitch. Its main weaknesses are its failure to account for the rate pitches of alternating click trains and its underestimation of the salience of low-frequency tones.

## INTRODUCTION

The "predominant interval" hypothesis for the low pitch of complex tones postulates that the pitch heard corresponds to the most frequent interval present in the population of auditory nerve fibers at any given time. Explicit models for pitch based on this hypothesis are able to readily account for a wide diversity of pitch phenomena (Licklider 1951; Meddis and Hewitt 1991a,b; Moore 1989; van Noorden 1982).

In the companion paper in this series, we tested the predominant interval hypothesis by examining pooled interspike interval distributions of ensembles of auditory nerve fibers. For many different harmonic complex tones, the most frequent interval in pooled interval distributions invariably corresponded to the period of the pitch that is heard. Peak-to-background ratios in pooled distributions qualitatively corresponded to the strength of the pitch that is heard. Those stimuli that evoke strong, salient pitches produced pooled

TABLE 1. *Stimulus parameters*

Stimulus	Variable	Range, Hz	Duration, ms	Level, dB SPL	Remarks
AM tone	$F_m$	80–160 160–320 240–480	480	60	$F_c = 640$ Hz, $m = 1.0$
QFM tone	$F_c$	500–700	480	60	$F_m = 125$ Hz, $m = 1.0$
	$F_m$	80–160 160–320 240–480	480	60	$F_c = 640$ Hz, $m = 1.0$
	$F_c$	500–700	480	60	$F_m = 125$ Hz, $m = 1.0$
Click trains	$F_0$	80–160 160–320	530.4	80	Uniform-polarity 40- $\mu$ s clicks
		80–160			Alternating-polarity 40- $\mu$ s clicks
Harmonic complexes	$F_0$	160–320	498	60 per component	Harmonics 3–5
		80–160			
		240–480			
	$F_0$	96–192 192–384 288–576	498	60 per component	Harmonics 6–12
Shepard-Risset	$F_0$	4.7–9.4	5,600	60	Octave harmonics (octaves 1–12)

AM, amplitude-modulated;  $F_m$ , modulation frequency;  $F_c$ , carrier frequency;  $m$ , modulation index; QFM, quasi-frequency-modulated;  $F_0$ , fundamental frequency.

distributions with high peak-to-background ratios, whereas those that produce weak pitches produced pooled distributions with low ratios. These psychoneural correspondences suggest that information subserving pitch perception could be represented in the interspike interval statistics of populations of auditory neurons.

In the present paper, we investigate the neural correlates of more complex pitch phenomena, using stimuli that produce multiple, competing pitches rather than one definite pitch. The resulting patterns of pitch judgments were studied by psychophysicists to probe auditory mechanisms underlying pitch perception: the pitch shifts of inharmonic complex tones (de Boer 1956, 1976; Schouten et al. 1962), pitch ambiguity, pitch circularity (de Boer 1956, 1976; Schouten et al. 1962; Shepard 1964), “rate pitch” (de Boer 1956, 1976; Flanagan and Gutman 1960a,b; Schouten et al. 1962; Shepard 1964), and the dominance region for pitch (de Boer 1956, 1976; Flanagan and Gutman 1960a,b; Plomp 1967; Ritsma 1967; Schouten et al. 1962; Shepard 1964). Stimuli having aperiodic waveforms and/or multiple, competing periodicities were used to study these effects: harmonic and inharmonic amplitude-modulated (AM) and quasi-frequency-modulated (QFM) tones, Shepard-Risset frequency-modulated tones, trains of alternating-polarity clicks, and pairs of harmonic complexes with competing fundamentals. The complex patterns of pitch judgments that these stimuli produce can be compared with pitch estimates from pooled interspike interval distributions of the auditory nerve to provide still stronger tests of the predominant interval hypothesis.

## METHODS

The experimental and analytical methods used here are described in detail in the companion paper. Responses of single auditory nerve fibers were recorded in Dial-anesthetized cats with the use of glass micropipette electrodes. Stimuli were numerically generated (sampling rate 100 kHz, 16-bit resolution) and presented via a calibrated, closed acoustic assembly.

Basic stimulus parameters are given in Table 1. Most stimuli

were sinusoidally varied in a parameter associated with pitch [e.g., fundamental frequency ( $F_0$ ), modulation frequency ( $F_m$ ), or carrier frequency ( $F_c$ )], with one cycle of variation occurring over the stimulus duration ( $\sim 500$  ms). AM and QFM stimuli were varied in either  $F_m$  or  $F_c$ . For AM and QFM tones, the modulation index was always 1.0. For the Shepard-Risset stimulus,  $F_0$  was linearly swept upward over 5.6 s. All stimuli were presented continuously, 40 times per fiber for the Shepard-Risset tones, 100 times for all others.

Distributions of all-order interspike intervals were analyzed by means of autocorrelation histograms and autocorrelograms (see companion paper). An autocorrelation histogram represents the distribution of all-order intervals, whereas an autocorrelogram represents the running distribution of all-order intervals. Thus the “cross section” of an autocorrelogram at a particular peristimulus time (PST) is an autocorrelation histogram (as in Fig. 3). Interval distributions from 30–90 single fibers were grouped into octave characteristic frequency (CF) bands, weighted to approximate a human distribution of CFs, and summed together to form pooled autocorrelation histograms and autocorrelograms that then served as estimates of the population interval distribution of the human auditory nerve.

Pitches were estimated from smoothed pooled autocorrelation histograms (300-ms rectangular window). The position of the highest interval peak was taken as the estimate of the pitch period. Except for the Shepard-Risset tones, when multiple, harmonically related major peaks of comparable height were present (e.g.,  $1/F_0$ ,  $2/F_0$ ,  $3/F_0$ ), the position of the peak representing the shortest interval was chosen as the pitch estimate. Pitch saliences were estimated from peak-to-background ratios in smoothed, pooled autocorrelation histograms (see companion paper). Fiber saliences are peak-to-background ratios in smoothed autocorrelation histograms of single fibers.

## RESULTS

### *Pitch shift and pitch ambiguity of inharmonic complex tones*

AM tones with variable  $F_c$ s (500–750 Hz) and fixed  $F_m$ s (125 or 250 Hz) were used to investigate the neural correlates of pitch shift.  $F_c$  varied sinusoidally from 500 to 750

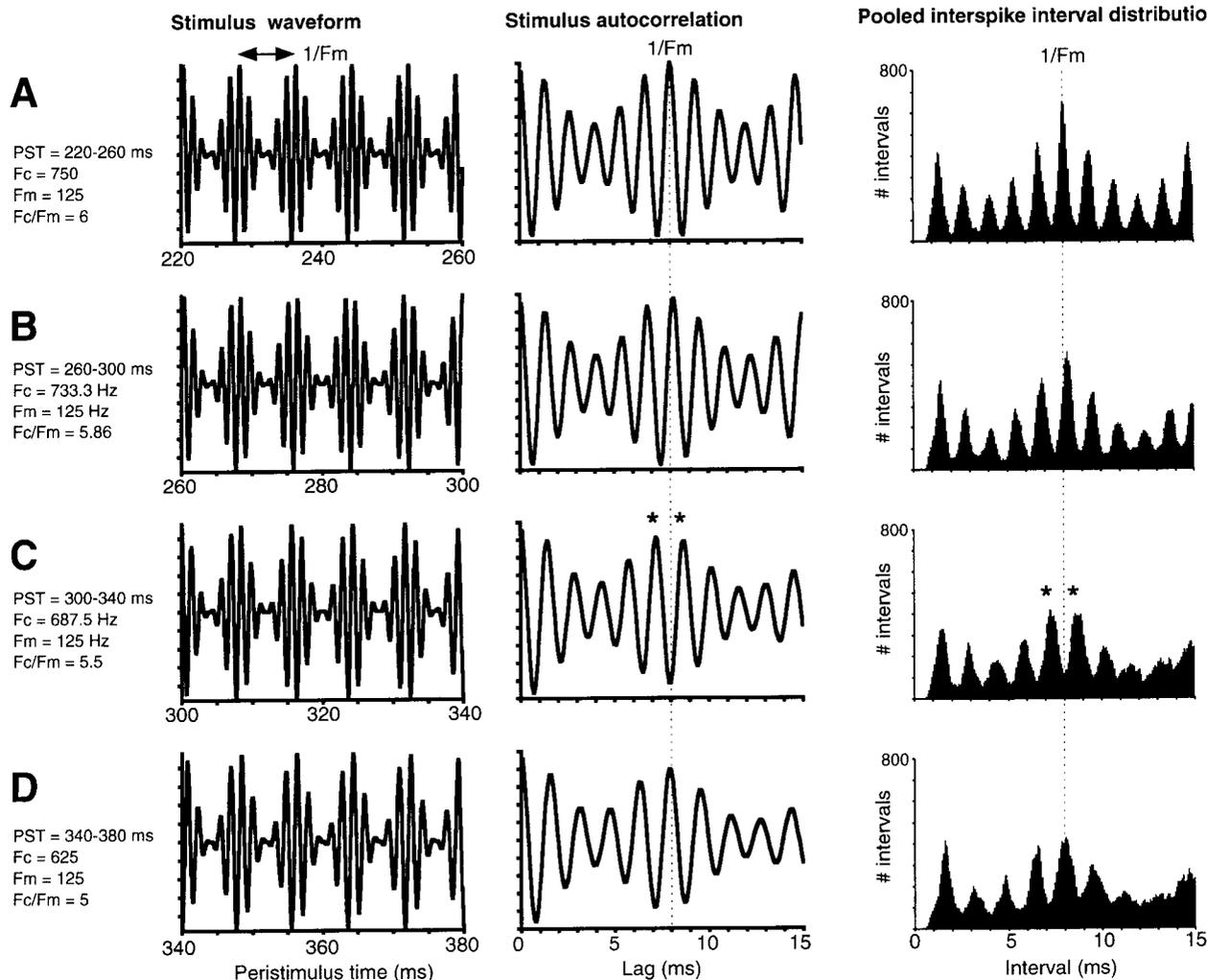


FIG. 1. Interspike interval correlates of the pitch shift of inharmonic complex tones. *Right*: stimulus waveforms, amplitude-modulated (AM) tones with fixed modulation frequency ( $F_m = 125$  Hz) and different carrier frequencies ( $F_c = 625$ – $750$  Hz). *A* and *D* are harmonic, whereas *B* and *C* are inharmonic. Peristimulus times (PSTs) listed correspond to those of Fig. 2A. *Middle*: stimulus short-term autocorrelation functions. Vertical dotted lines: modulation period  $1/F_m$ . *Left*: pooled all-order interval (autocorrelation) histograms for 31 auditory nerve fibers. Asterisks: multiple, ambiguous pitches that would be heard.

to 500 Hz over the 480-ms duration of the stimulus. Waveforms for four consecutive segments of this stimulus are shown in Fig. 1, *left*. The interval in the waveform indicated by the arrows is the modulation period  $1/F_m$  (8 ms). Autocorrelation functions for each waveform segment are shown in Fig. 1, *middle*. Here the vertical dotted line indicates the modulation period  $1/F_m$ .

When  $F_c/F_m$  is an integer (Fig. 1, *A* and *D*), the stimulus is harmonic and the waveform is periodic. Here  $F_0$  is equal to  $F_m$ , and the waveform has a period of  $1/F_m$ . For such harmonic AM tones, human listeners usually hear a clear, unambiguous pitch at  $F_m$  ( $F_m = F_0$ ).

When  $F_c/F_m$  is not an integer (Fig. 1, *B* and *C*), the stimulus is inharmonic and aperiodic. Although the envelope of the waveform still has a periodicity of  $F_m$ , the fine structure of the waveform contains periodicities that deviate slightly from  $F_m$ . For these inharmonic AM tones, human listeners hear pitches at fine-structure periodicities rather than at  $F_m$ . This deviation from  $F_m$  is the “first

effect of pitch shift,” and is described by de Boer’s rule (de Boer 1956, 1976). Stated in slightly different form, it is  $p = F_c/n$ , where  $p$  is the frequency of the pitch heard and  $n$  is an integer near  $F_c/F_m$ . The effect can also be described in terms of stimulus autocorrelation functions, where “pseudoperiods” correspond to maxima near  $1/F_m$  (de Boer 1956). Accompanying the shift in pitch is pitch ambiguity, wherein human listeners hear one of two or more pitches in the vicinity of, but not at,  $F_m$  (de Boer 1956, 1976; Schouten et al. 1962; Wiersinga-Post and Duifhuis 1995). Pitch ambiguity reaches a maximum when  $F_c/F_m$  is halfway between two integers, and the pitch alternatives for this condition are indicated by asterisks in Fig. 1C.

Pooled all-order interspike interval distributions (pooled autocorrelation histograms) for each of the stimulus segments are shown in Fig. 1, *right*. Most striking is the resemblance of each interval distribution to the autocorrelation function of the corresponding stimulus segment. Where a

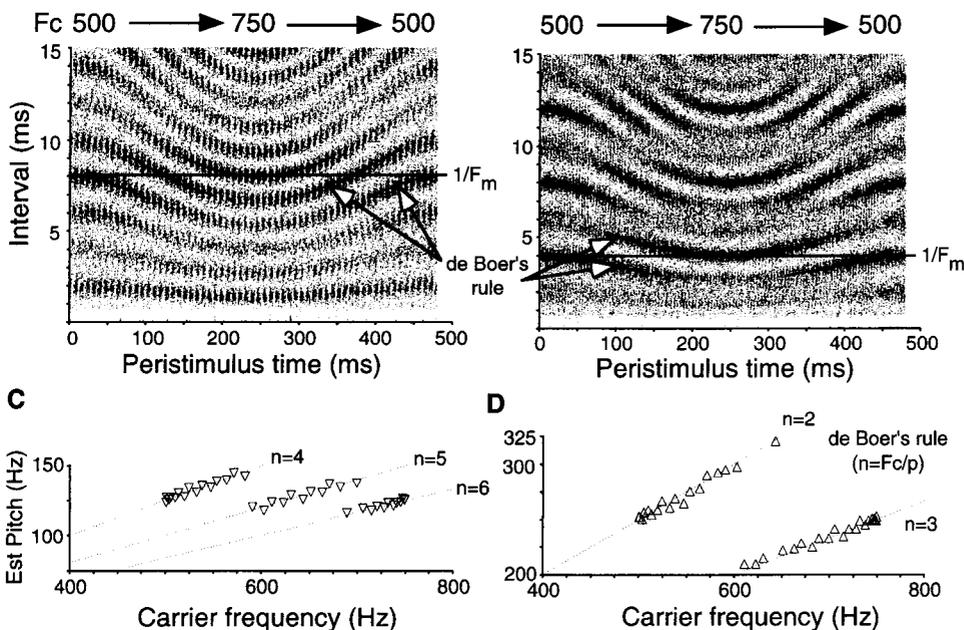
Variable carrier AM Tones  $F_c = 500\text{--}750\text{ Hz}$ A  $F_m = 125\text{ Hz}$ B  $F_m = 250\text{ Hz}$ 

FIG. 2. Neural correlates for pitch shift for variable-carrier AM tones.  $F_c = 500\text{--}700\text{ Hz}$ ,  $F_m = 125$  or  $250\text{ Hz}$ . *A* and *B*: pooled response autocorrelograms (47 and 53 fibers, respectively). Modulation period  $1/F_m$  (straight line) and pitch(es) heard (de Boer's rule, curved lines) are indicated. *C* and *D*: physiologically estimated pitches as a function of  $F_c$ . Diagonal lines: de Boer's rule.

clear, unambiguous pitch is heard at  $F_m$ , in the harmonic segments shown in Fig. 1, *A* and *D*, the highest interval peaks are located at  $1/F_m$ . Where a small pitch shift relative to  $F_m$  would be expected, as in Fig. 1*B*, the interval peak also shows a small deviation from  $1/F_m$ . Where two ambiguous pitches would be expected, as in the inharmonic segment shown in Fig. 1*C*, two roughly equal interval peaks appear in the histogram.

Running pooled interspike interval distributions (pooled autocorrelograms) in response to two variable-carrier AM tones ( $F_c = 500\text{--}750\text{ Hz}$ ,  $F_m = 125\text{ Hz}$  or  $250\text{ Hz}$ ) are shown in Fig. 2. The modulation period  $1/F_m$  is indicated by horizontal, straight lines, whereas curved black lines indicate the pitch(es) that would be heard (de Boer's rule). The two sets of lines converge when the stimulus is harmonic and one pitch is heard. The two sets diverge when the stimulus is inharmonic and one of several pitches can be heard. For inharmonic AM tones, the densest interval bands in the autocorrelograms always coincide with the pitches that are heard (de Boer's rule) rather than the modulation period  $1/F_m$ .

Pitches estimated from the autocorrelograms showed similar patterns of pitch shift and pitch ambiguity. Pitches estimated from successive 20-ms peristimulus segments are plotted against their respective  $F_c$ s in Fig. 2, *C* and *D*. Predictions of de Boer's rule are indicated by diagonal lines. All pitch estimates fall near those lines, in close agreement with de Boer's rule. In inharmonic regions where pitch ambiguity is expected (e.g.,  $F_c \sim 575\text{ Hz}$  and  $F_c \sim 675\text{ Hz}$  in Fig. 2*C*;  $F_c \sim 625\text{ Hz}$  in Fig. 2*D*), the pitch estimation algorithm finds multiple pitches.

The pitch shift experiment was also conducted using AM tones in which the modulator,  $F_m$ , was varied while the carrier,  $F_c$ , was kept constant (Fig. 3). Here a different

progression of pitches is heard than for the variable-carrier AM tones of Figs. 1 and 2. Instead of continuously shifting pitch values, ambiguities between several discrete pitches wax and wane as probabilities of making particular pitch judgments shift among them. The discrete pitches predicted by de Boer's rule are indicated by sets of straight horizontal lines in the pooled autocorrelograms (Fig. 3, *C* and *D*), whereas the continuous, undulating lines show modulation periods  $1/F_m$ . The stimulus segment pictured in Fig. 3*E* is a harmonic region ( $F_m = 320\text{ Hz}$ ,  $F_c/F_m = 2.0$ ) where one pitch at  $1/F_m$  is heard. The pooled autocorrelation histogram for the segment (Fig. 3*E*) has a maximum very close ( $+0.5\%$ ) to the modulation period. The stimulus segment pictured in Fig. 3*G* is an inharmonic region ( $F_m = 256\text{ Hz}$ ,  $F_c/F_m = 2.5$ ) where maximal pitch ambiguity is expected. Here two interval bands surrounding  $1/F_m$  appear to be of roughly equal density. These bands correspond to the two peaks that are indicated by arrows in the corresponding pooled autocorrelation histogram (Fig. 3*G*). These peaks both lie near, but not at,  $1/F_m$  (dashed line), and correspond to  $213\text{ Hz}$  ( $F_c/3$ ) and  $320\text{ Hz}$  ( $F_c/2$ ).

In regions of maximal pitch ambiguity, when  $F_c/F_m = n \pm 0.5$ , AM tones have a true  $F_0$  an octave below the  $F_m$ . Such "octave drops" are often reported by human listeners if they are allowed to choose pitch matches near the true fundamental (Gerson and Goldstein 1978). The partials present for the stimulus segment shown in Fig. 3*G* ( $F_c - F_m = 384\text{ Hz}$ ,  $F_c = 640\text{ Hz}$ , and  $F_c + F_m = 896\text{ Hz}$ ) are odd harmonics 3, 5, and 7 of a fundamental at  $128\text{ Hz}$ , which is an octave below  $F_m$  ( $256\text{ Hz}$ ). The octave drop pitch can be heard at this true fundamental if listeners consider pitches far from  $1/F_m$ , and correspondingly, the highest interval peak lies at this true fundamental.

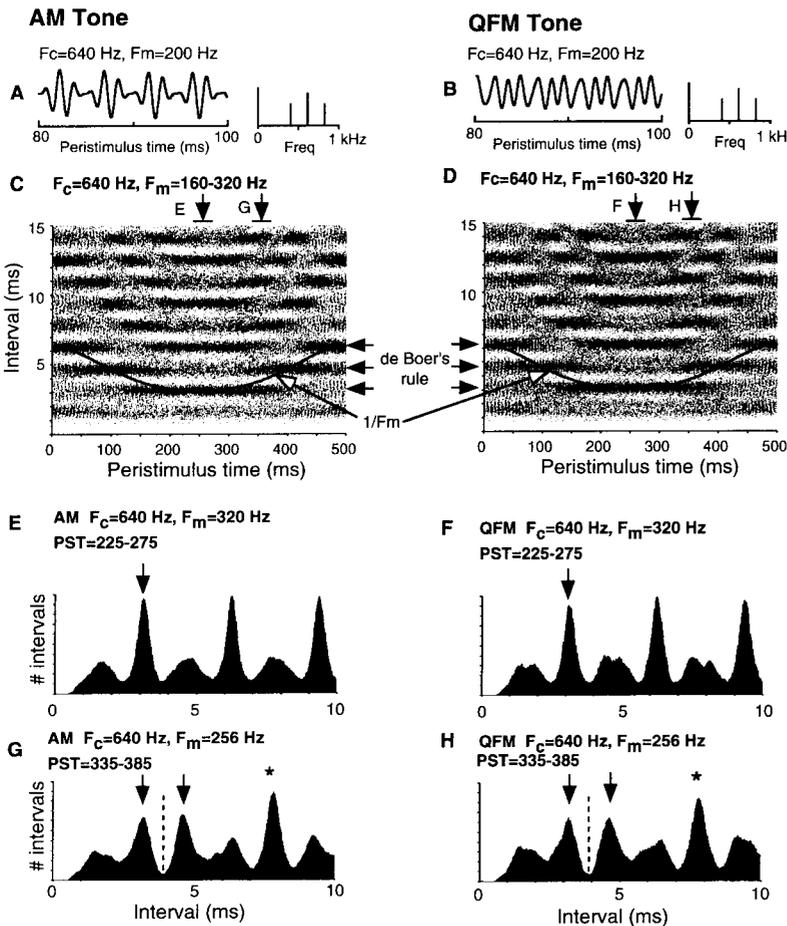


FIG. 3. Neural responses to variable-modulator AM and quasi-frequency-modulated (QFM) tones. *A* and *B*: waveforms and power spectra for AM and QFM tones when  $F_m = 200$  Hz. *C* and *D*: pooled response autocorrelograms. AM tone: 52 fibers; QFM tone: 44 fibers. Pitches heard correspond to de Boer's rule (horizontal lines indicated by triple arrows). *E*–*H*: pooled interval histograms for 50-ms stimulus segments. *E* and *F*: harmonic case ( $F_c/F_m = 2$ ) producing 1 clear pitch at  $1/F_m$  (arrow). *G* and *H*: inharmonic case ( $F_c/F_m = 2.5$ ) producing ambiguous pitches (dual arrows) near, but not at,  $1/F_m$ . Asterisks: lower "octave drop" pitch at  $1/\text{fundamental frequency } (F_0)$ . Dashed lines:  $1/F_m$ .

### Insensitivity of pitch to changes in phase spectrum and waveform envelope

To investigate the sensitivity of population interval distributions to changes in phase spectra, the pooled neural responses of AM tones were compared to those of their QFM counterparts (Fig. 3). QFM tones had the same carrier and variable  $F_m$ s as the variable-modulator AM tones ( $F_c = 640$  Hz,  $F_m = 160\text{--}320$  Hz). These AM and QFM tones thus differ only in their phase spectra, specifically in the phase of their  $F_c$ . Although their waveform envelopes are quite different (Fig. 3, *A* vs. *B*), these stimuli are subjectively very hard to tell apart.

Neural responses for the QFM tones (Fig. 3, *right*) are shown next to their AM counterparts (Fig. 3, *left*). The pooled autocorrelogram for the QFM tone (Fig. 3*D*) is strikingly similar to its AM counterpart (Fig. 3*C*) in both inharmonic and harmonic regions. Likewise, the pooled autocorrelation histograms for corresponding stimulus segments resembled those of their AM counterparts (i.e., Fig. 3, *E* and *F*, *G* and *H*). Interval correlates of pitch ambiguity and octave pitch drop were thus also seen for QFM stimuli. The only discernible difference between pooled interval histograms for AM and QFM lies in subtle splits in side peaks for the harmonic QFM case (Fig. 3*F*). The similarity of pooled interval distributions is consistent with the psychophysical observation that low pitches produced by complex

tones with low-frequency harmonics are highly invariant with respect to phase manipulations.

### Pitch circularity

Harmonic stimuli with many components spaced at octave intervals can also evoke a multiplicity of pitches ("octave confusions"). Shepard constructed a series of such ambiguous stimuli to illustrate the relativity of pitch intervals (Shepard 1964). Harmonics at octaves 1–11 comprise the stimulus. The waveform itself (Fig. 4*A*, *top*) has a fractal form that is invariant with respect to changes in time scale (Schroeder 1986). Shepard attenuated frequency components according to a fixed, log-cosine spectral envelope (Fig. 4*A*), so that  $F_0$  and all of the octave harmonics could then be slowly swept upward with minimal changes in timbre. Shepard used discrete versions of this timbre-invariant stimulus to demonstrate that pitch has a locally transitive but globally circular structure analogous to the visual staircase illusions of Penrose and Penrose (1958) and M. C. Escher. Subjectively, one hears a continuously rising pitch that at some ill-defined point seamlessly drops down an octave, only to continue rising again.

Pitch circularity was investigated with the use of a continuous Shepard-Risset tone (Risset 1971; Shepard 1964) in which an entire octave cycle of  $F_0$ s was continuously traversed over a 5.6-s period. The pooled response autocorrela-

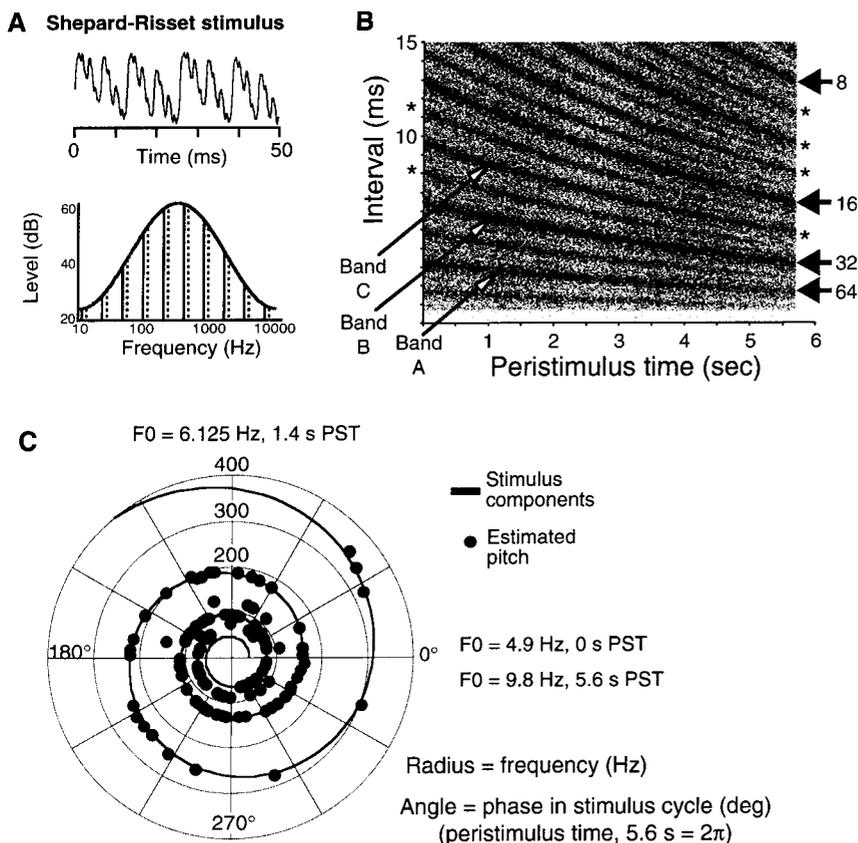


FIG. 4. Neural correlates of pitch circularity. *A*: waveform and power spectrum for the Shepard-Risset stimulus. All (octave) components are swept upward in frequency and attenuated according to the spectral envelope shown, with  $F_0 = 4.9$ –9.8 Hz. *B*: pooled response autocorrelogram (45 fibers). Harmonic numbers: interval bands associated with each octave harmonic. Asterisks: interval bands associated with component subharmonics. *C*: physiologically estimated pitches plotted on polar coordinates. PST increases as angle increases (counterclockwise). Frequency increases as radius increases.

gram (Fig. 4*B*) shows many interval bands of comparable density. The densest bands (←) correspond to stimulus components, whereas less dense ones (asterisks) correspond to component subharmonics. Following the interval bands marked “A” and “B” from the beginning of the stimulus to its end, the two bands begin as roughly equal, with band A gradually becoming less dense than band B as the stimulus ends. If the autocorrelogram is projected onto a cylinder with interval length as the vertical axis, and PST as its circumference, the sets of bands form spirals like those of a barber pole. These physiological results are comparable with those produced by autocorrelation models (Slaney and Lyon 1993).

Pitches estimated for each portion of the stimulus (Fig. 4*C*) show similar progressions. A polar coordinate system is useful for visualizing the spiral structure and multiplicity of pitch estimates. Because the stimulus traverses an octave cycle of  $F_0$ s, and then repeats, angle (0–360°) indicates where in the 5.6-s stimulus cycle a particular PST lies. Frequency is represented by radius. All frequency components present at a given time are therefore represented by a set of points having the same angle. Because  $F_0$  is continually increasing, the points sweep out a spiral figure over time (—).

When pitch estimates for each stimulus time are plotted on these polar coordinates, most estimates fall on the octave spiral (i.e., at one of the octave components). A few estimates in 50- to 100-Hz range (inner circle) not on the octave spiral line trace out a spiral path corresponding to subharmonics of higher-frequency components (band C in the auto-

correlogram). Following the octave spiral counterclockwise (outward), it can be seen that there are fewer and fewer pitch estimates as frequencies increase above 200 Hz (outer 2 rings). The probability of “dropping down” an octave in pitch therefore increases as one proceeds around the spiral (as time goes on). If a probabilistic model for pitch based on relative heights of interval peaks is combined with a “proximity principle” for maintaining pitch continuity in time (so that one only rarely jumps from 1 turn of the spiral to another), then the sequence of (mostly rising) pitch estimates resembles those heard by human listeners (Shepard 1964).

#### *Flanagan and Gutman rate pitch*

Three click trains with variable  $F_0$ s were used to study the neural correlates of rate pitch. Waveforms and power spectra for corresponding segments of the three trains are shown in Fig. 5. Two of the click trains (labeled A and B) consisted of pulses having the same polarity (condensation clicks), whereas the third train (labeled C) consisted of pulses alternating in polarity (condensation alternating with rarefaction). Train A and train C share the same fundamental, whereas trains B and C share the same click rate, interclick interval, and harmonic spacing. Trains A and B consist of consecutive harmonics, whereas train C consists of odd harmonics. Relations between the time patterns and spectra of the three trains were maintained as their fundamentals were sinusoidally varied over 530 ms. The fundamental of click train A varied from 80 to 160 Hz, that of

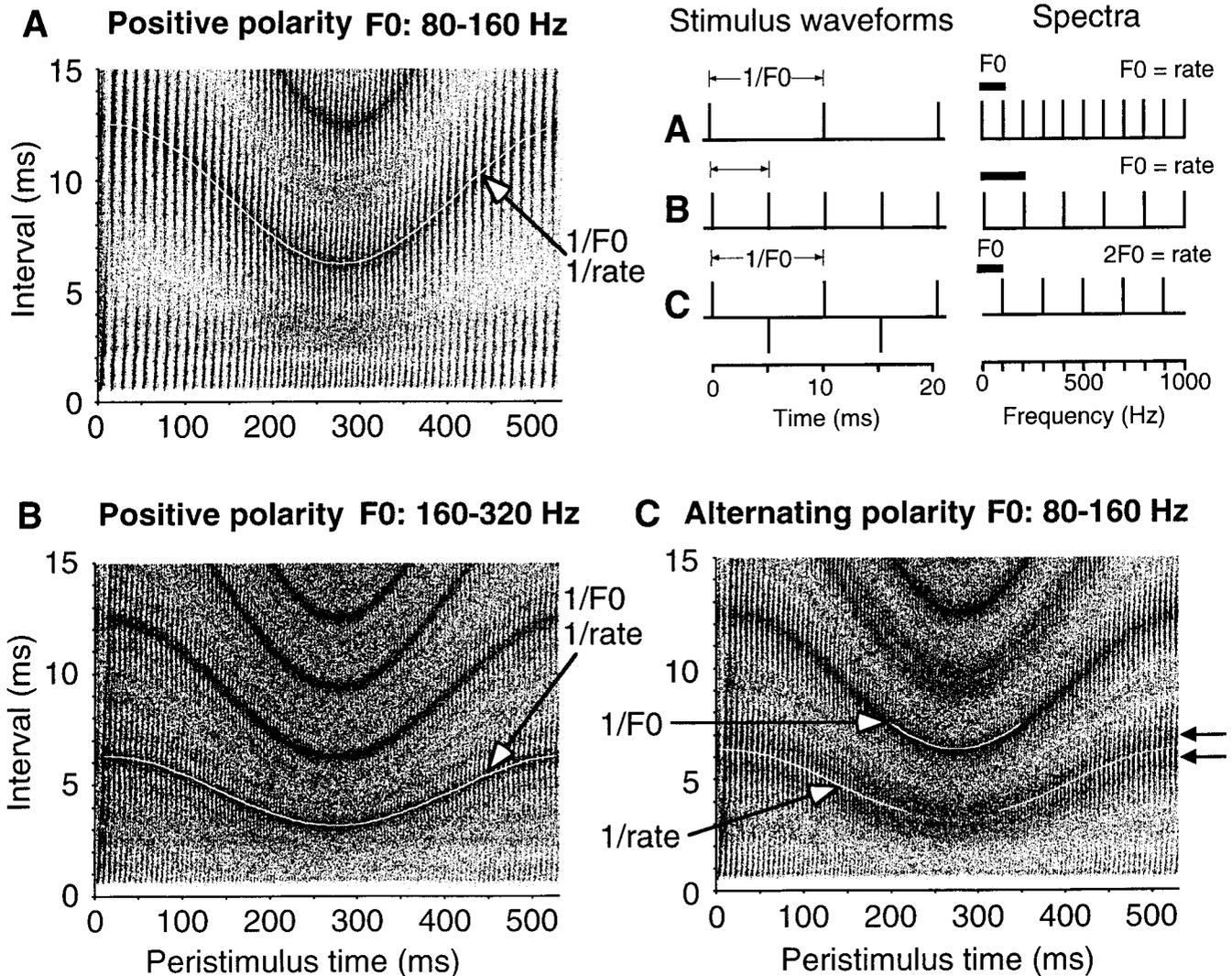


FIG. 5. Pooled interval distributions for click trains with variable fundamentals. *Top right*: representative stimulus waveforms and power spectra for click trains labeled A–C. A–C: pooled response autocorrelograms. A: unipolar click train with  $F_0 = 80$ –160 Hz. B: unipolar click train with  $F_0 = 160$ –320 Hz. C: alternating-polarity click train with  $F_0 = 80$ –160 Hz. White curved lines: periods of pitches heard ( $F_0$  and/or rate pitch).

train B varied from 160 to 320 Hz, and that of train C varied from 80 to 160 Hz.

Unipolar click trains produce clear unambiguous pitches at their fundamentals, which equal their click rates. Alternating click trains can produce rate pitches associated with interclick interval or a periodicity pitch associated with the period of the alternating click pattern, the rate pitch being an octave higher than the periodicity pitch. Whether a rate pitch or a periodicity pitch is heard depends on  $F_0$  (or click rate) (Flanagan and Gutman 1960a,b, 1964).  $F_0$ s < 150 Hz (click rates < 300 Hz) tend to produce rate pitches, whereas  $F_0$ s above 150 Hz (click rates > 300 Hz) tend to produce periodicity pitches at the fundamental.

Pooled autocorrelograms for the two unipolar click trains are shown in Fig. 5, A and B. In both cases, the densest interval bands closely follow the fundamental period  $1/F_0$  and the interclick interval, corresponding to the pitches that are heard in each case (white curved lines). The pooled autocorrelogram for the alternating train is shown in Fig.

5C. PSTs when rate and/or periodicity pitches could be heard are indicated by two sets of white curved lines. Here the densest interval band always follows the fundamental period  $1/F_0$  throughout the stimulus duration. Although the most frequent interval corresponds to the pitch that is heard for  $F_0$ s > 150 Hz (top curved line), the correspondence does not hold for  $F_0$ s < 150 Hz (bottom curved lines). In place of a clear interval band at the rate pitch, there are two diffuse bands.

These patterns are seen more clearly in Fig. 6, where autocorrelation histograms for  $F_0 = 80$  Hz (right) and  $F_0 = 160$  Hz (left) are presented. Histograms for 50 single fibers arranged by CF are shown in the neurograms of Fig. 6, A and D. The corresponding pooled autocorrelation histograms, which are CF weighted but left unsmoothed, are shown in Fig. 6, B and E. Asterisks indicate the periods of pitches that would be heard. For both  $F_0 = 80$  Hz and  $F_0 = 160$  Hz, the highest peak in the pooled interval distributions is clearly at  $1/F_0$ . For  $F_0 = 80$  Hz there are also two promi-

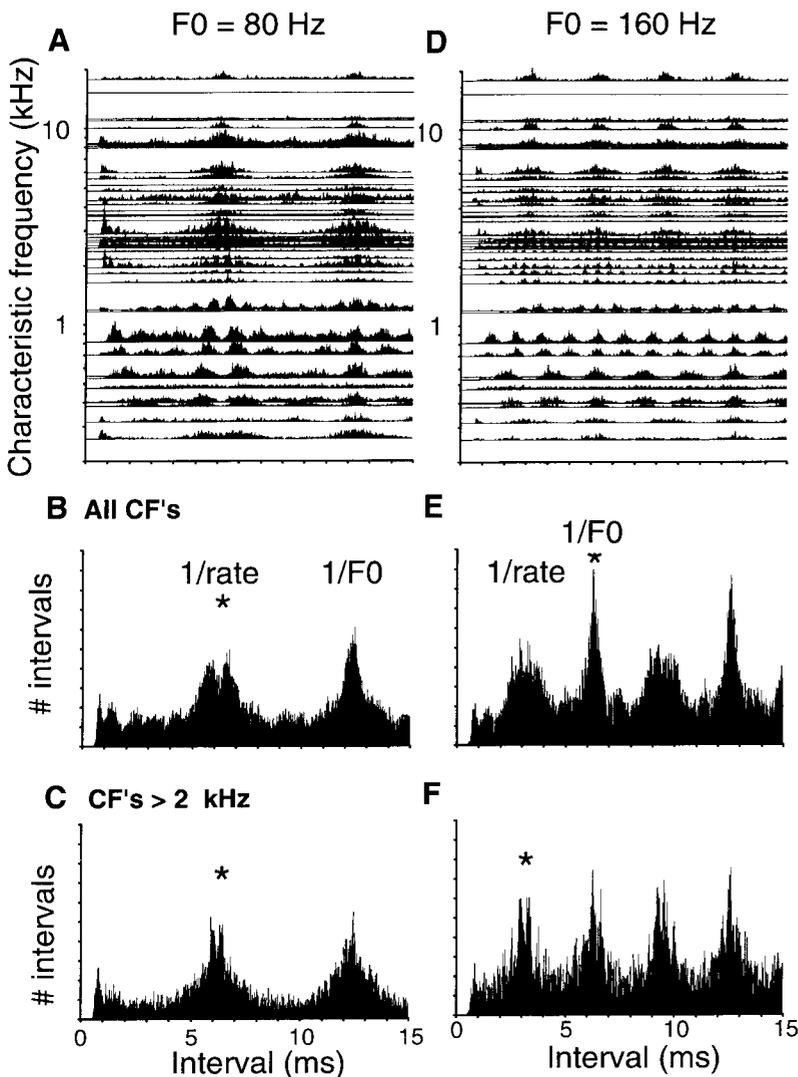


FIG. 6. Single-fiber and pooled autocorrelation histograms in response to alternating-polarity click trains. *A* and *D*: single-fiber autocorrelation histograms for 50 fibers arranged by characteristic frequency (CF), when  $F_0 = 80$  Hz and  $F_0 = 160$  Hz. *B* and *E*: pooled autocorrelation histograms. *C* and *F*: pooled autocorrelation histograms for 30 fibers with CFs  $> 2$  kHz. Asterisks: periods of pitches heard.

nent peaks near the interclick interval ( $1/\text{click rate}$ ), but neither of these peaks is as high as that at  $1/F_0$ . Thus, because the most frequent interval in the pooled distribution does not follow the rate pitch that is heard, there is a clear discrepancy between human pitch judgments and the predominant interval hypothesis for pitches produced by alternating-polarity click trains with low  $F_0$ s.

The reason that most frequent interval does not follow the rate pitch for  $F_0 < 150$  Hz is that interval peaks for  $1/\text{rate}$  are split in the pooled autocorrelation histograms. These split peaks correspond to the two diffuse interval bands in the autocorrelation neurogram (Fig. 5C). The origin of these two interval bands is evident in the autocorrelation neurogram of Fig. 6A. Whereas fibers with CFs  $> 2$  kHz show one interval peak near the interclick period ( $1/\text{rate}$ ), those with lower CFs show two peaks that diverge with decreasing CF. These pairs of interval peaks are created by different spike latencies ( $\pm \frac{1}{2}CF$ ) for clicks of positive and negative polarity. When a fiber responds to clicks of opposite polarity, intervals related to the interclick interval  $\pm$  the CF-dependent latency are generated (Kiang et al. 1965). Consequently, when the responses of many fibers with different CFs are summed, a

pair of interval peaks is created in the pooled autocorrelation histogram, with neither peak being as high as the interval peak at the fundamental period.

When alternating-polarity click trains are high-pass filtered, rate pitches are strengthened and their existence region is extended upward to  $F_0 \geq 400$  Hz (Flanagan and Gutman 1964). Because low-CF fibers respond poorly to higher-frequency stimuli, the population interval distribution for these high-pass-filtered click trains can be estimated by summing the responses of fibers with CFs above the cutoff frequency. From the neurograms of Fig. 6, *A* and *D*, it can be seen that high-CF fibers produce comparable interval peaks at  $1/\text{click rate}$  and at  $1/F_0$ . When only fibers with CFs  $> 2$  kHz are used to construct pooled autocorrelation histograms (Fig. 6, *C* and *F*), for  $F_0$ s of both 80 and 160 Hz, interval peaks at  $1/\text{click rate}$  become roughly equal in height to those at  $1/F_0$ . Thus the predominant interval hypothesis does appear to account for rate pitch in the high-pass-filtered case.

Other features of neural responses also correspond with rate pitch. Examining the autocorrelation of Fig. 5C, prominent vertical striations can be seen for PSTs 0–200 ms and 330–530 ms, when fundamentals are  $< 150$  Hz. These verti-

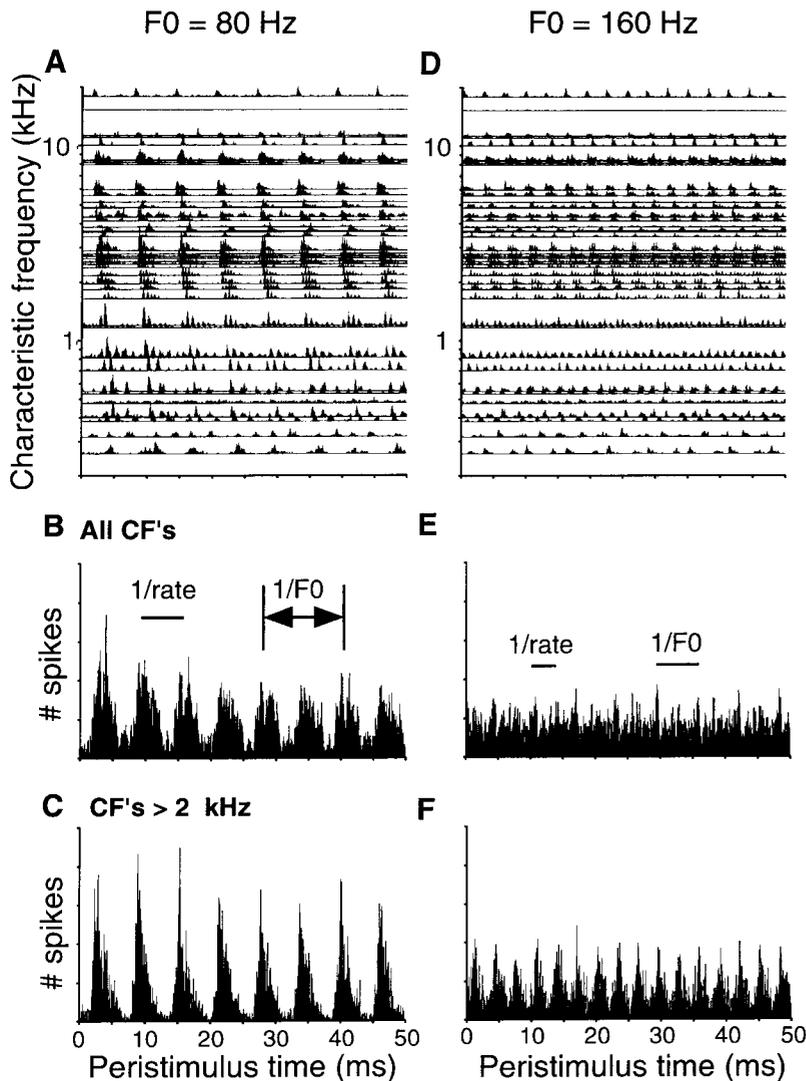


FIG. 7. Single-fiber and pooled PST histograms in response to alternating-polarity click trains. *A* and *D*: single-fiber PST histograms for 50 fibers arranged by CF, when  $F_0 = 80$  Hz and  $F_0 = 160$  Hz. *B* and *E*: pooled PST histograms. *C* and *F*: pooled PST histograms for 30 fibers with CFs  $> 2$  kHz.

cal striations indicate a high degree of interneural synchrony across the auditory nerve. This can be seen very clearly in the neurogram and pooled PST histograms of Fig. 7. For  $F_0 = 80$  Hz, when the rate pitch is heard, many fibers across the auditory nerve are discharging in rough synchrony (Fig. 7*A*), thereby producing regular, distinct peaks in the pooled PST histogram (Fig. 7*B*) that are separated by the interclick interval (or  $1/\text{click rate}$ ). In contrast, when  $F_0 = 160$  Hz, and the rate pitch is not heard, there are no clear peaks in the pooled PST histogram (Fig. 7*E*). This apparent loss of interneural synchrony on a population level is not due to changes in single-fiber discharge patterns, but to the smearing out of peaks and troughs in population PST histograms as CF-dependent latency differences reach significant fractions of interclick intervals. When pooled PST histograms of fibers with CFs  $> 2$  kHz are examined (Fig. 7, *C* and *F*), clear peaks separated by the interclick interval are apparent for both  $F_0 = 80$  Hz (Fig. 7*C*) and  $F_0 = 160$  Hz (Fig. 7*F*). Thus, when a rate pitch is heard, there exists widespread interfiber synchrony across the entire population that gives rise to clear periodicities in population discharge

rates. When rate pitches are absent, no clear, periodic patterns of population discharge rates are apparent.

#### *Dominance region for pitch*

Two sets of harmonic complexes were used to investigate the dominance region for pitch. One complex consisted of harmonics 3–5 of one fundamental ( $F_{0\ 3-5}$ ), whereas the other consisted of harmonics 6–12 of another fundamental ( $F_{0\ 6-12}$ ) that was always 20% higher in frequency. This  $F_0$  separation, also employed by Meddis and Hewitt (1991a), was made greater than those used in comparable psychophysical studies (Plomp 1967, 1976), to ensure that interval peaks could be clearly identified and separated. All frequency components had the same magnitude (60 dB SPL) and were in cosine phase. The waveforms and magnitude spectra of the individual complexes and their combination are shown in Fig. 8. Three sets of waveforms were constructed to cover the ranges of 80–160 Hz, 160–320 Hz, and 240–480 Hz. For the pitch dominance experiment, the two complexes were presented together. Thus for each range

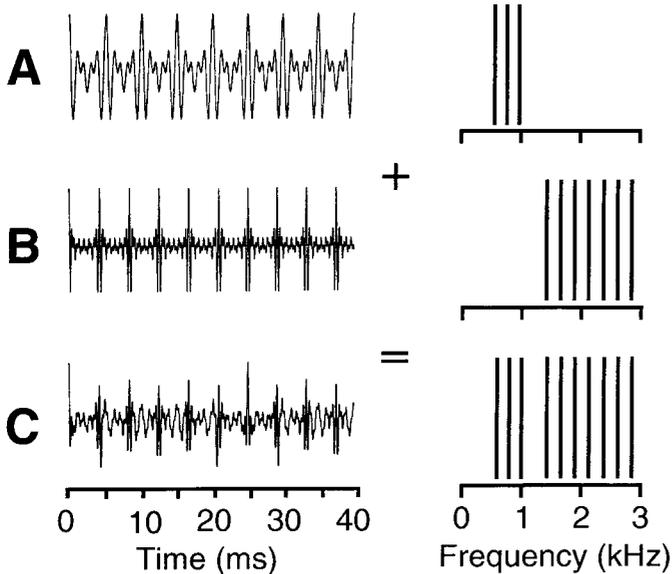


FIG. 8. Stimuli used to study the neural correlates of the dominance region for pitch. *Left*: waveforms. *Right*: power spectra. *A*: harmonics 3–5. *B*: harmonics 6–12 of a fundamental 20% higher in frequency. *C*: harmonics 3–5 and harmonics 6–12 presented concurrently.

of  $F_{0s}$ , there were three stimulus conditions: harmonics 3–5 presented alone, harmonics 6–12 presented alone, and both complexes presented together. From psychophysical studies (Moore and Glasberg 1985; Plomp 1967, 1976; Ritsma 1967), human listeners would be expected to hear the  $F_{0\ 3-5}$  over  $F_{0\ 6-12}$  throughout the range of fundamentals from 100 to 500 Hz.

Pooled autocorrelation histograms for  $F_{0s}$  near 80, 160, 240, 320, and 480 Hz are shown in Fig. 9. The two vertical dotted lines running through each histogram indicate the fundamental periods of harmonics 3–5 ( $1/F_{0\ 3-5}$ ) and harmonics 6–12 ( $1/F_{0\ 6-12}$ ). The *top row of histograms* shows the pooled interval distributions for harmonics 3–5 presented alone, and in each case the highest interval peak is very close to the fundamental period of the complex ( $1/F_{0\ 3-5}$ ). The *middle row of histograms* shows a similar result for harmonics 6–12. The *bottom row of histograms* shows the pooled distributions for the two complexes presented together. In the *leftmost histogram*, when  $F_{0\ 3-5} = 80$  Hz, the highest interval peak lies near the fundamental period of harmonics 6–12, whereas for all other  $F_{0s}$  shown (160, 240, 320, and 480 Hz) the highest peak lies near that of harmonics 3–5. For all  $F_{0s}$  the pooled distribution for the complexes presented together resembles the sum of the distributions of the complexes presented individually. Thus the complex producing the highest salience (peak-to-background ratio) when presented alone is the one that produces the highest salience (the dominant pitch) when the complexes are presented together. In all three cases, as  $F_0$  increases, peak-to-background ratios decrease, paralleling the decline in pitch salience that is observed when complex tones have higher harmonic numbers.

Algorithms for pitch and salience estimation were used to provide a more quantitative description of the results. Pitch estimates for the individual and combined complexes are plotted as a function of  $F_0$  in Fig. 10, *A* and *B*. The solid

lines indicate the  $F_{0s}$  of the two complexes. Over the entire range of  $F_{0s}$ , the estimated pitch of each complex presented alone (Fig. 10*A*) closely followed its corresponding fundamental (the corresponding solid line). For harmonics 6–12, as the  $F_0$  increased beyond 300 Hz and component frequencies increased beyond 2 kHz, pitch estimates showed wider deviations from the fundamental. For these  $F_{0s}$ , individual harmonics would not be psychophysically resolved.

When the two complexes were presented together (Fig. 10*B*), estimated pitches followed either one or the other fundamental, depending on the  $F_0$ . Pitch estimates for a given PST are plotted against  $F_{0\ 3-5}$  at that time. For  $F_{0s} > 160$  Hz, estimated pitches consistently followed  $F_{0\ 3-5}$ . For  $F_{0s}$  from 120 to 160 Hz, estimates included roughly equal numbers of pitches near each fundamental. For  $F_{0s} < 120$  Hz, the estimated pitches consistently followed  $F_{0\ 6-12}$ .

Estimated saliences for the harmonic complexes are plotted as a function of  $F_0$  in Fig. 10, *C* and *D*. Estimated salience curves for the fundamentals of each harmonic complex presented alone are shown in Fig. 10*C*. The salience curve for harmonics 6–12 has been shifted leftward (downward in frequency) by 20% to facilitate direct comparison with the salience curves in the concurrent case (Fig. 10*D*). The salience for harmonics 3–5 reaches a maximum when  $F_0$  is between 150 and 200 Hz, when the frequency components of the complex are between 450 and 1,000 Hz. If the maximum of the salience curve for harmonics 6–12 is taken to be in the vicinity of 100 Hz, then the maximum saliences for both complexes would be obtained when the frequency components are in the 600- to 1,000-Hz range. Thus salience estimates appear to depend more on the frequencies of stimulus components than on  $F_0$  or harmonic number.

Reinforcing this interpretation, estimated salience curves for the individual complexes undergo roughly parallel declines as the  $F_0$  increases, so that the salience of harmonics 3–5 when  $F_0 \approx 400$  Hz is roughly similar to that of harmonics 6–12 when  $F_0 \approx 200$ . The salience curves of the two complexes cross when the  $F_0$  ( $F_{0\ 3-5}$ ) is 160 Hz. Thus for  $F_{0s} < 160$  Hz, the estimated salience of harmonics 6–12 is greater than that of harmonics 3–5, whereas at higher  $F_{0s}$  the situation is reversed.

Estimated salience curves for the two fundamentals when both complexes are presented concurrently are shown in Fig. 10*D*. Estimated saliences for the two fundamentals are computed from the same pooled autocorrelation histograms with the use of the same background interval density value for both cases. Estimated saliences in the concurrent case have consistently lower values than those for separate presentation, indicating competition between the different complexes for production of intervals related to their fundamentals ("interval suppression"). Despite this difference, the forms of the salience curves and the  $F_{0s}$  at which they cross are very similar for the separate and competing conditions. Thus the fundamental that is most salient when presented in isolation is the one that dominates when the two are presented together.

Although all of the saliences discussed thus far have been estimates of interspike interval peak-to-background ratios for the entire auditory nerve, peak-to-background ratios were also computed from interval distributions of single fibers.

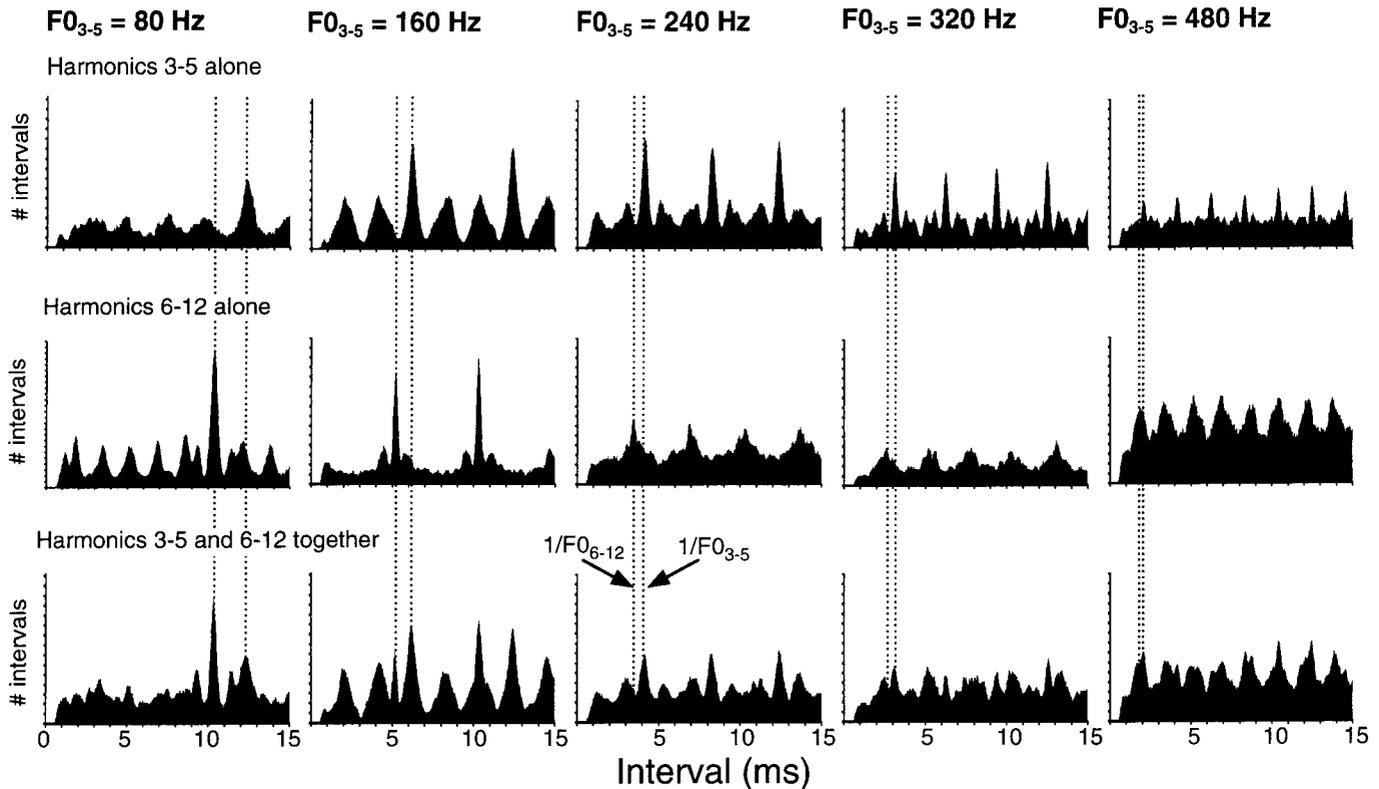


FIG. 9. Neural correlates of the dominance region for pitch.  $F_{0s}$  indicated are for harmonics 3–5.  $F_{0s}$  for harmonics 6–12 were always 20% higher. *Top row*: pooled interval histograms for harmonics 3–5 presented alone (47, 83, 49, 83, and 49 fibers, respectively). *Middle row*: pooled interval histograms for harmonics 3–5 presented alone (43, 70, 44, 70, and 44 fibers, respectively). *Bottom row*: pooled interval histograms for harmonics 3–5 and harmonics 6–12 presented together (60, 79, 72, 79, and 72 fibers, respectively). Vertical dotted lines: intervals corresponding to the respective fundamentals, harmonics 3–5 of one fundamental ( $F_{0\ 3-5}$ ) and harmonics 6–12 of another fundamental ( $F_{0\ 6-12}$ ) that was always 20% higher in frequency, with  $1/F_{0\ 6-12}$  always on the left and  $1/F_{0\ 3-5}$  always on the right. Ordinal axis range for histograms is 0–150 spikes per bin, except for the leftmost middle plot, where it is 0–250 spikes per bin.

These fiber saliences are useful for assessing relative contributions of different CF regions to overall salience estimates. Fiber saliences for the harmonic complexes presented both alone and concurrently are shown in Fig. 11. To facilitate comparisons between separate and concurrent presentations, fiber saliences were computed from the same histogram (i.e., the same peristimulus segment) when  $F_{0\ 3-5} = 150$  Hz and  $F_{0\ 6-12} = 180$  Hz. Fiber saliences for harmonics 3–5 presented alone were roughly equal for all CF regions (Fig. 11A, ●). Because stimulus components ranged from 450 to 750 Hz, there were very few fibers that did not respond to the stimulus (i.e., fiber saliences  $\leq 0.1$ ). In comparison, the fiber saliences for harmonics 6–12 presented alone (○) are noticeably smaller for CFs  $< 600$  Hz. Because stimulus components ranged from 1,080 to 2,160 Hz, this complex lies outside the response area of many low-CF fibers. Fiber saliences for each of the two fundamentals for the concurrent case are shown in Fig. 11B. Fiber saliences for  $F_{0\ 3-5}$  (●) drop off dramatically for CFs  $> 800$  Hz, whereas fiber saliences for  $F_{0\ 6-12}$  begin to rapidly increase. The frequency boundaries of the two complexes (450–750 Hz vs. 1,080–2,160 Hz) are consistent with the observed sharp crossover between 700 and 1,000 Hz. For higher  $F_{0s}$ , similar patterns of fiber saliences were seen, but with correspondingly higher crossover regions situated between the frequency boundaries

of the two complexes. Thus, for the concurrent case, different CF regions contribute intervals predominantly related to one or the other fundamental, giving rise to different temporal representations in different parts of the auditory nerve. Which CF regions contribute the most intervals can be determined from the stimulus spectrum.

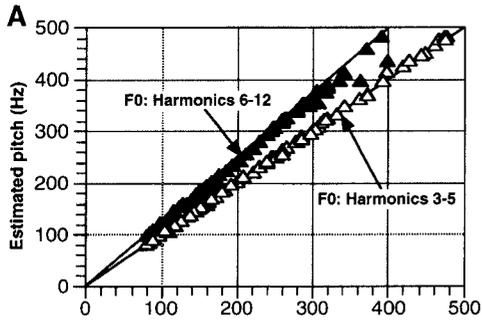
## DISCUSSION

### *Predominant interval hypothesis for the low pitch of complex tones*

The data presented in the preceding paper demonstrate that simple neural correlates of the low pitch of harmonic complexes exist in pooled interval distributions of the auditory nerve. Data presented here further demonstrate that neural correlates of more complex patterns of pitch judgments are also directly observable in pooled interval distributions of the auditory nerve.

The basic assumptions of the predominant interval hypothesis are that 1) the low pitch of complex tones corresponds to the most frequent interval in the population interval distribution of the auditory nerve and 2) the strength of this low pitch is related to the ratio of intervals associated with the pitch to the mean number of intervals present in

## Harmonics 3-5 and 6-12 presented separately



## Harmonics 3-5 and 6-12 presented concurrently

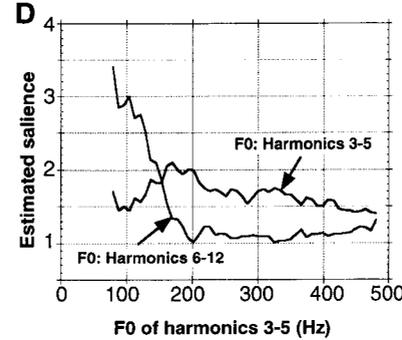
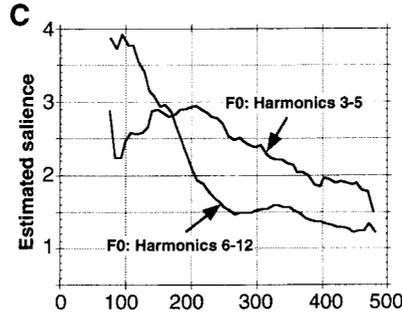
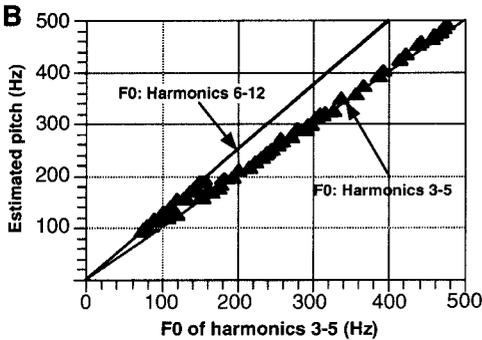


FIG. 10. Physiologically estimated pitches and pitch saliences for the dominance region experiment. All  $F_0$ s are plotted in terms of corresponding  $F_{0\ 3-5}$  (i.e.,  $F_{0\ 6-12}$  is always 20% higher than plotted). *A* and *B*: pitch estimates as a function of  $F_0$ ,  $F_{0\ 3-5}$ . Estimates were made from consecutive 20-ms peristimulus response segments. *C* and *D*: estimated salience as a function of  $F_0$ ,  $F_{0\ 3-5}$ . Curves are least-squares fits of 9th-order polynomials to estimates made from 600 overlapping 20-ms peristimulus response segments.

the population distribution. Taken together with the basic assumptions, the data presented here provide explanations for pitch shift, pitch ambiguity, pitch circularity, the phase insensitivity of pitch, the pitch of unresolved harmonics, and the dominance region for pitch. The only pitch phenomenon studied that is not readily explicable in these terms is the rate pitch of alternating-polarity click trains with low fundamentals.

## Pitch shift

Psychophysical experiments using inharmonic AM tones were originally carried out to probe the nature of the auditory pattern-recognition mechanisms responsible for the pitch of complex tones (de Boer 1976; Schouten 1940a,b; Schouten et al. 1962; Warren 1982). If auditory pitch mechanisms used local frequency spacings between spectral peaks ('dif-

ference frequencies'), or waveform envelopes to make pitch discriminations, then pitch should remain constant as all frequency components are shifted by a constant amount. When frequency components were all shifted, rather than hearing a constant pitch at the difference frequency  $F_m$ , listeners heard a small shift in pitch away from  $F_m$ . These experiments decisively ruled out 1) simple spectral pattern analyzers that detected absolute spacings between neighboring frequency components and 2) simple temporal pattern analyzers that detected periodicities of the waveform envelope. To account for these shifts, a spectral pattern processor must carry out an harmonic analysis of spectral patterns rather than simply recognizing constant frequency spacings, or alternately a temporal pattern processor must analyze the temporal fine structure of the stimulus rather than its envelope (de Boer 1956, 1976).

Studies have previously shown that the fine structure of

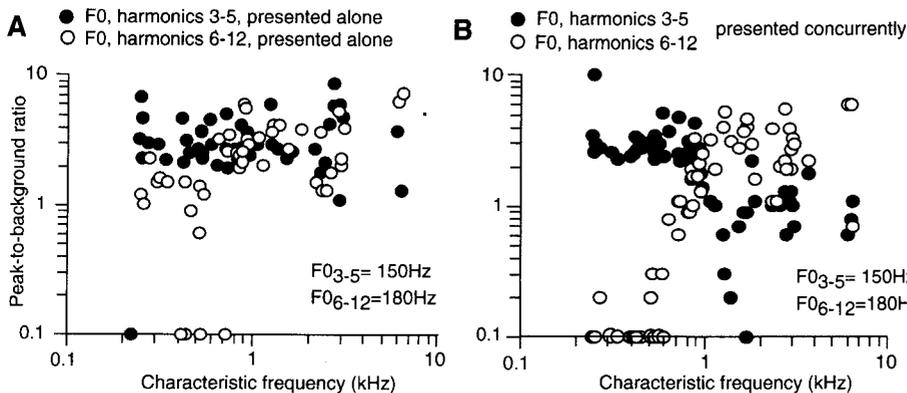


FIG. 11. Fiber saliences as a function of CF in the dominance region experiment.  $F_{0\ 3-5} = 150$  Hz,  $F_{0\ 6-12} = 180$  Hz. *A*: fiber saliences (peak-to-background ratios) for  $F_{0\ 3-5}$  and  $F_{0\ 6-12}$ , each presented individually. *B*: fiber saliences for  $F_{0\ 3-5}$  and  $F_{0\ 6-12}$  presented concurrently.

inharmonic complex tones is reflected in the temporal discharge patterns of single auditory nerve fibers (Evans 1978; Javel 1980). The physiological results presented here show that the first effect of pitch shift (de Boer's rule) can be directly explained in terms of an autocorrelation-like population-based predominant interval model for pitch. The more subtle "second effect of pitch shift," a small but systematic deviation from de Boer's rule, was not observed. This effect is thought to be caused by cochlear distortion products, "combination tones" (Smooenburg 1970), that would be expected to be small given the relatively large component spacings for our stimuli. Although there exist explanations for inharmonic pitch shifts in terms of central mechanisms that perform optimal matching to harmonic spectral templates (Gerson and Goldstein 1978; Goldstein 1973; Terhardt et al. 1982a,b), the physiological data here indicate that these pitch shifts also have direct correlates in the population interspike interval distribution of the auditory nerve.

#### *Pitch ambiguity, pitch circularity, and phase insensitivity*

Pitch ambiguity was found to correspond to the appearance of multiple major peaks in pooled interval distributions. Where several discrete pitches would be heard for an inharmonic AM stimulus, multiple independent peaks of comparable height were seen (Fig. 2). Where stimulus parameters are shifted to positions where octave drops in pitch are expected (e.g.,  $F_c/F_m = n + 0.5$ ), corresponding patterns of peaks in pooled interspike interval distributions were observed (Fig. 2G). These sorts of patterns have been seen previously in single-fiber responses to inharmonic stimuli: AM tones (Javel 1980), nonuniform click trains (Evans 1978, 1983, 1986), and two-tone complexes (Greenberg and Rhode 1987).

The shifting relative heights of the various peaks in pooled interval distributions roughly parallel the shifting probabilities of particular pitch alternatives. Where the probability of hearing one pitch was greater than that of hearing another alternative pitch, its corresponding interval peak was higher than the other's. Thus the probability that a given pitch is heard among several possibilities appeared to correspond to the height of its associated interval peaks, relative to those associated with the other pitches.

When this probabilistic pitch allocation strategy is applied to the Shepard-Risset stimulus, and assumptions concerning local temporal contiguity of pitch judgments are incorporated, a reasonable account of pitch circularity is given. The pitch allocation principle also suggests possible interval-based explanations for several aspects of musical pitch: the various pitches that can be heard for a chord, the relative strengths of these pitches, and the degree to which they fuse to form a unified percept (Boring 1942; DeWitt and Crowder 1987; Parncutt 1989).

A considerable body of psychophysical work has investigated the role of phase spectrum and waveform envelope in pitch perception. Because changes in phase spectra can result in large changes in waveform envelope without concomitant changes in pitch, invariance of pitch frequency with respect to phase was used to falsify simple temporal models that measured intervals between peaks in the unfiltered waveform (Moore 1989; Wightman 1973a,b). A useful tactic has been

to change phase spectra without altering magnitude spectra (Small 1970), thereby allowing spectral pattern cues (e.g., the difference frequencies) to be separated from temporal waveform cues (de Boer 1976; Schouten 1940b). Although low pitches of complex tones can be very sensitive to subtle shifts in magnitude spectra, as in "pitch shift" experiments, they are largely insensitive to changes in the phase spectra. Particular phase configurations can influence salience or distinctness of low pitches, especially for stimuli composed of higher-frequency components (Bilsen 1973; Licklider 1955; Nordmark 1978; Small 1970). However, these phase manipulations generally have little or no effect on the perceived frequencies of low pitches (Bilsen 1973; Meddis and Hewitt 1991b; Patterson 1973; Plomp 1976). These observations notwithstanding, some exceptions, involving inharmonic complex tones and phase-dependent magnitudes of distortion products, have prompted some debate over the issue (Green 1976). Other, high pitches can also be produced by particular phase spectra, but these do not alter the frequencies of the low pitches that are heard (Kubovy and Jordan 1979).

Although their power spectra are identical, waveforms of AM and QFM tones with low-frequency carriers are very different (Fig. 3, A and B). Despite their dissimilar waveforms, pitches produced by AM and QFM tones with low-frequency carriers (or low harmonic numbers) are largely indistinguishable (de Boer 1956; Zwicker 1962). The AM and QFM tones used in this study produce very similar pitches, despite different phase spectra and great dissimilarity in the temporal envelopes of their waveforms. Such stimuli also produced almost indistinguishable pooled interval histograms (Fig. 3). These results are broadly consistent with those obtained from the interval-based model of Meddis and Hewitt (1991a,b). For sets of low-frequency harmonics, the Meddis and Hewitt model produces pitch estimates that are highly independent of phase. The model can also account for some of the subtle differences in pitches and timbres of AM and QFM tones with higher-frequency carriers, higher harmonic numbers, and greater modulation indexes that have been debated in the literature (Ritsma and Engel 1964; Wightman 1973a). Although AM and QFM tones with some of these properties can produce changes in first-order interval distributions for single fibers when  $F_c = CF$  (Horst et al. 1992), many of these differences would not be expected to be present in pooled distributions of all-order intervals. In summary, for AM and QFM tones with low-frequency harmonics, the predominant interval appears to be affected little by changes in phase-spectra or the shape of the waveform envelope. This is consistent with the highly phase-invariant character of both simulated population interval distributions and human pitch judgments.

#### *Flanagan-Gutman rate pitch*

Click trains of uniform polarity produce strong pitches at their  $F_0$ s. For these click trains  $F_0$ , click rate, and difference frequency are the same. For click trains of alternating polarity, however, click rate and difference frequency are twice the fundamental. For fundamentals  $<150$  Hz, alternating click trains produce rate pitches at the click rate, whereas higher  $F_0$ s produce periodicity pitches at the fundamental (Flanagan and Gutman 1960a,b, 1964). Similar rate pitches

can also be produced by other alternating-polarity stimuli (Pierce 1991; Warren 1982) and by (Seebeck's) click trains with alternating interclick intervals (Evans 1978, 1983, 1986).

The phenomenon of rate pitch produced the only clear contradiction of the predominant interval hypothesis that was found. Observed pooled interval distributions always showed the most frequent interval to be at the fundamental, and consequently no interval correlates of rate pitch were initially seen. This discrepancy was reduced somewhat when correspondences between pooled interval distributions and rate pitches were found for high-pass-filtered click trains. When alternating click trains are passed through a high-pass filter, the rate pitch is strengthened and can be heard at higher fundamentals (Flanagan and Gutman 1964). Because few low-CF fibers would be expected to respond in the high-pass condition, distributions using only higher-CF fibers can be used to estimate the population response. When only intervals from higher-CF fibers are considered, then there are roughly as many intervals at  $1/\text{rate}$  as there are at  $1/F_0$ . Because the first interval maximum is chosen for the pitch, the predominant interval model does correctly predict the rate pitches heard for the high-pass condition.

Another, possibly stronger neural correlate of rate pitch is the presence of periodic patterns of discharge on a population level. For  $F_0$ s  $< 150$  Hz, pooled PST histograms (Fig. 7) showed clear peaks in population discharge rates for every click in the train, as well as clear troughs during each interclick interval. The time between these peaks corresponds to the interclick period ( $1/\text{click rate}$ ). When fundamentals increase to  $\geq 150$  Hz, CF-dependent latency differences become a large proportion of the interclick period, and population rate peaks are smeared out. When only high-CF fibers are included in the pooled PST histograms, clear peaks and troughs are present for all  $F_0$ s studied (80–160 Hz), consistent with the extension of the existence region of rate pitches by high-pass filtering. High-CF fibers have smaller CF-dependent latency-differences among themselves, and thus they discharge more synchronously as a group than their lower-frequency counterparts. Thus widespread discharge synchrony between fibers can lead to the generation of temporal discharge patterns on a population level. The "roughness" of some sounds (Plomp 1976; Terhardt 1973) might be explicable in such terms (Tramo et al. 1992). Such patterns could strengthen rate pitch relative to periodicity pitch if central analyzers count intervals between fibers (cross-correlation) as well as those within fibers (autocorrelation). Such a process might be consistent with the existence of two pitch mechanisms, one that produces a rough, buzzing, "rattle pitch" or rate pitch (Plomp 1976), and one that produces a smoother periodicity pitch. If both intraneural and interneural intervals are counted, the resulting predominant interval can coincide either with the most common interneural interval (rate pitch) or with the most common intraneural interval (periodicity pitch). Pitch would then be the result of a global correlation operation, possibly implemented via neural temporal cross-correlation architectures, e.g., (Braitenberg 1961; Cariani 1995a; Loeb et al. 1983).

#### *Dominance region for pitch*

Some early temporal theories of pitch (Schouten 1940b; Schouten et al. 1962) assumed that the low pitch of a har-

monic complex tone was generated by interaction of harmonics not resolved by peripheral auditory filters, the "residue" of an incomplete aural analysis. Because harmonic spacing remains constant while peripheral filter bandwidths increase with center frequency, residue pitch was thought to be generated mainly by sets of high-frequency components. To test this hypothesis, a number of different psychophysical experiments were carried out to determine the relative importance of low- and high-frequency harmonics for low pitch. When two harmonic complexes with slightly different fundamentals were presented together, human listeners almost invariably heard the pitch associated with the lower-frequency harmonics (Plomp 1967). This pitch is heard presumably because its constituent components carry greater weight in the central neural representation of pitch. Other experiments using different stimuli and psychophysical tasks (Bilsen 1973; Bilsen and Ritsma 1970; Moore and Glasberg 1985; Plomp 1976; Ritsma 1967; Yost 1982) confirmed the presence of a frequency region (the dominance region), roughly between 500 and 1,000 Hz, that disproportionately contributes to low pitch. All of the experiments suggested that resolved, lower-frequency harmonics have a much greater effect on the pitch of complex tones than unresolved, higher-frequency ones, and this led to a resurgence of spectral pattern theories for pitch (de Boer 1976).

The physiological data demonstrate that there are correlates of the dominance of low-frequency stimulus components in the population interspike interval distribution of the auditory nerve. For fundamentals from 160 to 500 Hz, intervals related to the fundamental of the lower-frequency harmonics clearly dominated in pooled distributions. As  $F_0$ s and thus component frequencies increase, the saliences of the associated low pitches decline, and this was seen directly in saliences estimated from the physiological data. For fundamentals  $< 125$  Hz, intervals related to the fundamental of the higher-frequency harmonics clearly dominated. This puts the crossover frequency for "interval dominance" roughly between 600 and 800 Hz, which is in the vicinity of crossover frequencies determined psychophysically. These results are generally in accord with those obtained by computer simulation (Meddis and Hewitt 1991b).

Predominant interval models thus provide an alternate, temporal interpretation of the dominance region for pitch. Here, the dominance region is a direct consequence of the differential abilities of auditory nerve fibers to phase-lock to stimulus components of lower and higher frequencies. The relative numbers of intervals produced by a given harmonic complex is thus a function of the aggregate numbers of discharges evoked by the stimulus and the degree to which the discharges phase-lock to stimulus components. Both aggregate numbers of discharges and degree of phase-locking depend on the distribution of fiber CFs and their positions relative to those of the stimulus harmonics present. In contrast to earlier temporal models based on unresolved harmonics, it is the lower-frequency components in the 500- to 1,000-Hz range that produce the most associated intervals. Thus, from the perspective of a population interval model for pitch, it is immaterial whether the frequency components are psychophysically resolved or not, as long as intervals related to the fundamental are produced (see discussion of Fig. 12 in the companion paper). Interval-based models of

pitch therefore have the advantage of explaining the pitches generated by both resolved and unresolved harmonics in terms of the same mechanism.

### *Summary of pitch results*

Taking the physiological data in its entirety, predominant interval models for pitch yield surprisingly satisfactory, comprehensive, and economical explanations for the pitch of complex tones. Many pitch phenomena are simply explained with the few model assumptions outlined above: the missing fundamental, pitch invariance with respect to level, pitch equivalence of spectrally diverse stimuli, pitch salience, pitch shift, phase insensitivity, and the dominance region for pitch. With the addition of a probabilistic pitch allocation rule, the population interval model can also explain the multiple pitches heard for some complex tones (pitch ambiguities). With assumptions concerning local temporal congruity of pitch judgments, the model can also give a reasonable account of pitch circularity.

Interval-based analysis could also be extended to form models for musical consonance (Tramo et al. 1992). In terms of pooled interval distributions, musical consonance is associated with fewer competing periodicities, whereas dissonance is associated with many competing periodicities (cf. Patterson 1986). Musical intervals with low integer ratios (e.g., 3:2, 4:3, 5:4) produce simpler autocorrelation patterns (fewer competing pitches) than do those with higher integer ratios (e.g., 12:11, 45:32), and these differences are directly reflected in pooled interval distributions.

As it stands, the predominant interval hypothesis has two apparent weaknesses. In the companion paper it was reported that the physiologically estimated salience of the 160-Hz pure tone was lower than expected relative to the estimated saliences of several complex tones. Likewise, in the present paper the dominance of harmonics 3–5 over harmonics 6–12 extended downward only to fundamentals as low as 160 Hz. Although the lower-frequency boundary of the dominance region is less well delineated in the psychophysical literature than the upper one, it was expected that harmonics 3–5 would dominate for fundamentals at least down to 100 Hz. Both of these results could be due to insufficient weight given to very low-CF responses when human population interval distributions are estimated from cat data. The second shortcoming of the predominant interval model is its inability to explain the Flanagan and Gutman click rate pitch. Here much more weight would need to be given to high-CF responses.

### *Central processing of temporal information for pitch*

The strong correspondence between pitch judgments and the population interval distribution of the auditory nerve begs the question of the mechanism by which the central auditory system computes pitch. Perhaps the most convincing evidence for a central temporal analysis comes from electrical stimulation of humans via single-channel cochlear implants, where only very minimal “place” information is thought to be present. Psychophysical studies indicate that temporal periodicities up to several hundred hertz can be distinguished (Eddington et al. 1978; Shannon 1992). How-

ever, the pitches heard are generally weak, and are very poorly defined above 400 or 500 Hz, where just-noticeable frequency differences can be  $\geq 100$  Hz. These pitches are affected by both temporal pattern and place of stimulation in the cochlea (Eddington et al. 1978). These findings have decidedly mixed implications for temporal models for pitch. On one hand, they constitute an existence proof that purely temporal information is utilized by the central auditory system in forming pitch percepts. On the other, it is known that electrical stimulation produces more precise phase-locking in auditory nerve fibers than does acoustic stimulation (Dynes and Delgutte 1992), so that electrically evoked interspike interval distributions would be expected to produce much finer pitch discriminations than acoustical stimulation, in marked contrast to what is observed. Given the enormous amount of timing information presumed to be available, it can be argued that the purely temporal processing mechanisms normally play only a weak role in pitch perception, participating only in the discrimination of pitches below 300 Hz (Shannon 1992). This strong line of reasoning notwithstanding, it remains to be verified whether peripheral interspike interval distributions in response to complex electrical stimuli are indeed qualitatively similar to their acoustically produced counterparts (Javel et al. 1988; Javel 1988). Recent cochlear compound action potential data from human cochlear implant users (Wilson et al. 1994) suggest that the electrically stimulated auditory nerve may not be capable of conveying interval information for frequencies much above 400 Hz. This might then explain why cochlear implants fail to produce pitch discriminations above that limit.

Is enough timing information available at central auditory stations to support a temporally based mechanism for pitch? There are several ways by which interspike interval information might be conveyed and utilized by central auditory stations. One possibility is that temporal discharge patterns in the auditory periphery are transformed to discharge rate codes more centrally (Eggermont 1993; Epping and Eggermont 1986; Langner 1992; Schreiner and Langner 1988). Central auditory units whose discharge rate and/or stimulus-locked synchronization of discharges is maximal for particular stimulus periodicities have been found in many central auditory stations (Epping and Eggermont 1986; Langner 1992; Langner and Schreiner 1988; Schreiner et al. 1983). However, central representations for pitch based on these “modulation detectors” have several major difficulties: 1) modulation tuning is coarse and degrades at high stimulus levels (Rees and Møller 1987) and in noise (Rees and Palmer 1989), 2) the range of periodicities represented decreases dramatically at more central auditory stations, such that the range of best  $F_m$ s at the cortex is far below the range of periodicity pitches heard (Eggermont 1993, 1994), 3) sharply tuned cortical “periodicity detectors” have yet to be found (Lyon and Shamma 1995; Schwartz and Tomlinson 1990; Sheich 1991), and 4) analysis of envelope modulation is not how the central processor computes periodicity pitch, as was ruled out by experiments demonstrating the pitch shift of inharmonic AM tones (de Boer 1956; Schouten et al. 1962). Whatever the central mechanism for pitch, it must operate on either the fine structure of the waveform or the harmonic structure of its spectrum.

Some peripheral interspike interval information may be

preserved in higher centers. In the cochlear nucleus, pitch-related interspike intervals (Boerger 1974; Cariani 1995b,c; Kim and Leonard 1988; Kim et al. 1990; Rhode 1995; Shofner 1991) and discharge periodicities (Blackburn and Sachs 1990; Frisina et al. 1990; Glatke 1969; Greenberg and Rhode 1987; Palmer et al. 1986; Wang and Sachs 1994) are observed in many types of neurons. In the auditory midbrain and cortex, discharge periodicities up to 1 kHz have been observed in single- and multiple-unit responses to periodic stimuli (Langner 1992; Licklider 1959; de Ribaupierre et al. 1972; Rouiller et al. 1981), although periodicities above a few hundred hertz are rare in cortical units. Even evoked potentials, which detect only those temporal discharge patterns that are widely synchronized over neural populations, show periodicities up to 2 kHz at the level of the midbrain (Greenberg 1980) and up to several hundred hertz at the level of the cortex (Goldstein and Kiang 1958; Kiang and Goldstein 1959; Mäkelä et al. 1990; Steinschneider et al. 1980). A central interval code for periodicity pitch is thus not completely out of the question.

Finally, synchronous temporal codes in the periphery may be transformed into asynchronous temporal codes more centrally. These could include patterns of long intervals at fundamental subharmonics, interspike intervals jittered with respect to the stimulus and each other, intervals that are interleaved with other spike patterns, and interval patterns that are gated in time. These kinds of patterns might be expected from the high degree of divergence and convergence found in the auditory pathway, but none of the methods commonly utilized in the central auditory system to analyze spike trains would detect these transformations. Period histograms of their derived measures (synchronization and modulation indexes), for example, detect only stimulus-locked time patterns, and will miss more subtle, asynchronous ones. When more powerful analytical methods have been used to study neural coding in other parts of the brain, complex temporal patterns have been found (Abeles et al. 1993; Chung et al. 1970; Emmers 1981; Lestienne and Strehler 1987; Mountcastle 1993; Richmond et al. 1989; Villa 1992; Wasserman 1992). Until better methods are used to study responses to complex stimuli in more central auditory stations, complex temporal pattern codes cannot be ruled out.

If temporal information is preserved and/or transformed in the ascending auditory pathway, how this information might be utilized by central auditory processors to discriminate pitch remains a key question. Neural architectures that compute temporal autocorrelations (Chung et al. 1970; Licklider 1951, 1959; Reichardt 1961) and temporal cross-correlations (Braitenberg 1961, 1967; Jeffress 1948) have been proposed in the past, and such architectures may provide starting points for new neurally based temporal processing strategies for pitch. Other possibilities lie in the development of artificial, adaptive time-delay neural networks (Cariani 1995a; Chappelier and Grumbach 1994; Hopfield 1995; MacKay 1962; Mozer 1993; Pratt 1990; Tank and Hopfield 1987; Torras i Genis 1985), which in turn may inspire entirely new classes of physiologically based models for the perception of auditory forms.

We acknowledge invaluable comments and criticisms that we received from N. Kiang, C. Brown, M. Tramo, and J. Melcher, and the support that

was given by everyone at the Eaton Peabody Laboratory. We also thank R. Meddis, A. de Cheveigné, and an anonymous reviewer for many helpful suggestions.

This work was supported by National Institute on Deafness and Other Communication Disorders Grants DC-00019, DC-00006, DC-02356, and DC-02258.

Address for reprint requests: P. Cariani, Eaton Peabody Laboratory, Massachusetts Eye & Ear Infirmary, 243 Charles St., Boston, MA 02114.

Received 10 October 1995; accepted in final form 6 May 1996.

## REFERENCES

- ABELES, M., BERGMAN, H., MARGALIT, E., AND VAADIA, E. Spatiotemporal firing patterns in the frontal cortex of behaving monkeys. *J. Neurophysiol.* 70: 1629–1638, 1993.
- BILSEN, F. A. On the influence of the number and phase of harmonics on the perceptibility of the pitch of complex signals. *Acustica* 28: 60–65, 1973.
- BILSEN, R. A. AND RITSMA, R. J. Some parameters influencing the perceptibility of pitch. *J. Acoust. Soc. Am.* 47: 469–475, 1970.
- BLACKBURN, C. C. AND SACHS, M. B. The representation of the steady-state vowel sound /e/ in the discharge patterns of cat anteroventral cochlear nucleus neurons. *J. Neurophysiol.* 63: 1191–1212, 1990.
- DE BOER, E. *On the "Residue" in Hearing* (PhD thesis). Amsterdam: Univ. of Amsterdam, 1956.
- DE BOER, E. On the "residue" and auditory pitch perception. In: *Handbook of Sensory Physiology*, edited by W. D. Keidel and W. D. Neff. Berlin: Springer-Verlag, 1976, p. 479–583.
- BOERGER, G. Coding of repetition noise in the cochlear nucleus of the cat. In: *Facts and Models in Hearing*, edited by E. Zwicker and E. Terhardt. New York: Springer-Verlag, 1974, p. 206–215.
- BORING, E. G. *Sensation and Perception in the History of Experimental Psychology*. New York: Appleton-Century-Crofts, 1942.
- BRAITENBERG, V. Functional interpretation of cerebellar histology. *Nature Lond.* 190: 539–540, 1961.
- BRAITENBERG, V. Is the cerebellar cortex a biological clock in the millisecond range? *Prog. Brain Res.* 25: 334–346, 1967.
- CARIANI, P. As if time really mattered: temporal strategies for neural coding of sensory information. *Commun. Cognit. Artif. Intell.* 12: 161–229, 1995a.
- CARIANI, P. Physiological correlates of periodicity pitch in the cochlear nucleus. *Assoc. Res. Otolaryn. Abstr.* 128, 1995b.
- CARIANI, P. A. Temporal coding of periodicity pitch in interspike interval distributions of cochlear nucleus populations (Abstract). *Ann. Biomed. Eng.* 23: S81, 1995c.
- CHAPPELIER, J. C. AND GRUMBACH, A. Time in neural networks. *SIGART Bull.* 5: 3–11, 1994.
- CHUNG, S. H., RAYMOND, S. A., AND LETTVIN, J. Y. Multiple meaning in single visual units. *Brain Behav. Evol.* 3: 72–101, 1970.
- DEWITT, L. A. AND CROWDER, R. G. Tonal fusion of consonant musical intervals: the oomph in Stumpf. *Percept. Psychophys.* 41: 73–84, 1987.
- DYNES, S. B. C. AND DELGUTTE, B. Phase-locking of auditory-nerve discharges to sinusoidal stimulation of the cochlea. *Hear. Res.* 58: 19–90, 1992.
- EDDINGTON, D. K., DOBELLE, W. H., BRACKMAN, D. E., MLADEJOVSKY, M. G., AND PARKIN, J. Place and periodicity pitch by stimulation of multiple scala tympani electrodes in deaf volunteers. *Trans. Am. Soc. Artif. Intern. Organs* 24: 1–5, 1978.
- EGGERMONT, J. J. Functional aspects of synchrony and correlation in the auditory nervous system. *Concepts Neurosci.* 4: 105–129, 1993.
- EGGERMONT, J. J. Temporal modulation transfer functions for AM and FM stimuli in cat auditory cortex. Effects of carrier type, modulating waveform and intensity. *Hear. Res.* 74: 51–66, 1994.
- EMMERS, R. *Pain: A Spike-Interval Coded Message in the Brain*. New York: Raven, 1981.
- EPPING, W. J. M. AND EGGERMONT, J. J. Sensitivity of neurons in the auditory midbrain of the grassfrog to temporal characteristics of sound. I. Stimulation with acoustic clicks. *Hear. Res.* 24: 37–54, 1986.
- EVANS, E. F. Place and time coding of frequency in the peripheral auditory system: some physiological pros and cons. *Audiology* 17: 369–420, 1978.
- EVANS, E. F. Pitch and cochlear nerve fibre temporal discharge patterns. In: *Hearing: Physiological Bases and Psychophysics*, edited by R. Klinke and R. Hartmann. Berlin: Springer-Verlag, 1983, p. 140–146.

- EVANS, E. F. Cochlear nerve fibre temporal discharge patterns, cochlear frequency selectivity and the dominant region for pitch. In: *Auditory Frequency Selectivity*, edited by B. C. J. Moore and R. D. Patterson. New York: Plenum, 1986, p. 253–264.
- FLANAGAN, J. L. AND GUTMAN, N. On the pitch of periodic pulses. *J. Acoust. Soc. Am.* 32: 1308–1319, 1960a.
- FLANAGAN, J. L. AND GUTMAN, N. Pitch of periodic pulses without fundamental component. *J. Acoust. Soc. Am.* 32: 1319–1328, 1960b.
- FLANAGAN, J. L. AND GUTMAN, N. Pitch of high-pass-filtered pulse trains. *J. Acoust. Soc. Am.* 36: 757–765, 1964.
- FRISINA, R. D., SMITH, R. L., AND CHAMBERLAIN, S. C. Encoding of amplitude modulation in the gerbil cochlear nucleus. I. A hierarchy of enhancement. *Hear. Res.* 44: 99–122, 1990.
- GERSON, A. AND GOLDSTEIN, J. L. Evidence for a general template in central optimal processing for pitch of complex tones. *J. Acoust. Soc. Am.* 63: 498–510, 1978.
- GLATTKE, T. J. Unit responses of the cat cochlear nucleus to amplitude-modulated stimuli. *J. Acoust. Soc. Am.* 45: 419–425, 1969.
- GOLDSTEIN, J. L. An optimum processor theory for the central formation of the pitch of complex tones. *J. Acoust. Soc. Am.* 54: 1496–1516, 1973.
- GOLDSTEIN, M. H., JR. AND KIANG, N. Y. S. Synchrony of neural activity in electric responses evoked by transient acoustic stimuli. *J. Acoust. Soc. Am.* 30: 107–114, 1958.
- GREEN, D. M. *An Introduction to Hearing*. Hillsdale, NJ: Erlbaum, 1976.
- GREENBERG, S. *Neural Temporal Coding of Pitch and Vowel Quality: Human Frequency-Following Response Studies of Complex Signals*. Los Angeles, CA: UCLA Working Papers in Phonetics #52, 1980.
- GREENBERG, S. AND RHODE, W. S. Periodicity coding in cochlear nerve and ventral cochlear nucleus. In: *Auditory Processing of Complex Sounds*, edited by W. A. Yost and C. S. Watson. Hillsdale, NJ: Erlbaum, 1987, p. 225–236.
- HOPFIELD, J. J. Pattern recognition computation using action potential timing for stimulus representation. *Nature Wash. DC* 376: 33–36, 1995.
- HORST, J. W., JAVEL, E., AND FARLEY, G. R. Coding of fundamental frequency in auditory nerve fibers: effects of signal level and phase spectrum. In: *The Processing of Speech*, edited by M. E. H. Schouten. Berlin: Mouton-DeGruyter, 1992, p. 29–36.
- JAVEL, E. Acoustic and electrical encoding of temporal information. In: *Cochlear Implants: Models of the Electrically Stimulated Ear*, edited by J. M. Miller and F. A. Spelman. New York: Springer-Verlag, 1988, p. 246–295.
- JAVEL, E. Coding of AM tones in the chinchilla auditory nerve: implications for the pitch of complex tones. *J. Acoust. Soc. Am.* 68: 133–146, 1980.
- JAVEL, E., MCGEE, J. A., HORST, W., AND FARLEY, G. R. Temporal mechanisms in auditory stimulus coding. In: *Auditory Function: Neurobiological Bases of Hearing*, edited by G. M. Edelman, W. E. Gall, and W. M. Cowan. New York: Wiley, 1988, p. 515–558.
- JEFFRESS, L. A. A place theory of sound localization. *J. Comp. Physiol. Psychol.* 41: 35–39, 1948.
- KIANG, N. Y. S. AND GOLDSTEIN, M. H., JR. Tonotopic organization of the cat auditory cortex for some complex stimuli. *J. Acoust. Soc. Am.* 31: 786–790, 1959.
- KIANG, N. Y. S., WATANABE, T., THOMAS, E. C., AND CLARK, L. F. *Discharge Patterns of Single Fibers in the Cat's Auditory Nerve*. Cambridge: MIT Press, 1965.
- KIM, D. O. AND LEONARD, G. Pitch-period following response of cat cochlear nucleus neurons to speech sounds. In: *Basic Issues in Hearing*, edited by H. Duijhuys, J. W. Horst, and H. P. Wit. London: Academic, 1988, p. 252–260.
- KIM, D. O., SIRIANNI, J. G., AND CHANG, S. O. Responses of DCN-PVCN neurons and auditory nerve fibers in unanesthetized decerebrate cats to AM and pure tones: analysis with autocorrelation/power-spectrum. *Hear. Res.* 45: 95–113, 1990.
- KUBOVY, M. AND JORDAN, R. Tone-segregation by phase: on the phase sensitivity of the single ear. *J. Acoust. Soc. Am.* 66: 100–106, 1979.
- LANGNER, G. Periodicity coding in the auditory system. *Hear. Res.* 60: 115–142, 1992.
- LANGNER, G. AND SCHREINER, C. E. Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanisms. *J. Neurophysiol.* 60: 1799–1822, 1988.
- LESTIENNE, R. AND STREHLER, B. L. Time structure and stimulus dependence of precise replicating patterns present in monkey cortical neuronal spike trains. *Brain Res.* 43: 214–238, 1987.
- LICKLIDER, J. C. R. A duplex theory of pitch perception. *Experientia* VII: 128–134, 1951.
- LICKLIDER, J. C. R. Influence of phase coherence upon the pitch of complex sounds (Abstract). *J. Acoust. Soc. Am.* 27: 996, 1955.
- LICKLIDER, J. C. R. Three auditory theories. In: *Psychology: A Study of a Science. Study I. Conceptual and Systematic*, edited by S. Koch. New York: McGraw-Hill, 1959, p. 41–144.
- LOEB, G. E., WHITE, M. W., AND MERZENICH, M. M. Spatial cross-correlation. A proposed model for acoustic pitch perception. *Biol. Cybern.* 47: 149–163, 1983.
- LYON, R. AND SHAMMA, S. Auditory representations of timbre and pitch. In: *Auditory Computation*, edited by H. Hawkins, T. McMullin, A. N. Popper, and R. Fay. New York: Springer-Verlag, 1995, p. 517.
- MACKAY, D. M. Self-organization in the time domain. In: *Self-Organizing Systems 1962*, edited by M. C. Yovitts, G. T. Jacobi, and G. D. Goldstein. Washington, DC: Spartan, 1962, p. 37–48.
- MÄKELÄ, J. P., KARMOSS, G., MOLNAR, M., CSEPE, V., AND WINKLER, I. Steady-state responses from cat auditory cortex. *Hear. Res.* 45: 41–50, 1990.
- MEDDIS, R. AND HEWITT, M. J. Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I. Pitch identification. *J. Acoust. Soc. Am.* 89: 2866–2882, 1991a.
- MEDDIS, R. AND HEWITT, M. J. Virtual pitch and phase sensitivity of a computer model of the auditory periphery. II. Phase sensitivity. *J. Acoust. Soc. Am.* 89: 2883–2894, 1991b.
- MOORE, B. C. J. *Introduction to the Psychology of Hearing* (3rd ed.). London: Academic, 1989.
- MOORE, C. J. AND GLASBERG, B. R. Relative dominance of individual partials in determining the pitch of complex tones. *J. Acoust. Soc. Am.* 77: 1853–1860, 1985.
- MOUNTCASTLE, V. Temporal order determinants in a somatosensory frequency discrimination: sequential order coding. *Ann. NY Acad. Sci.* 682: 151–170, 1993.
- MOZER, M. C. Neural net architectures for temporal sequence processing. In: *Predicting the Future and Understanding the Past*, edited by A. Weigend and N. Gershenfeld. Redwood City, CA: Addison-Wesley, 1993, p. 243–264.
- VAN NOORDEN, L. Two channel pitch perception. In: *Music, Mind and Brain*, edited by M. Clynes. New York: Plenum, 1982, p. 251–269.
- NORDMARK, J. O. Frequency and periodicity analysis. In: *Handbook of Perception*, edited by E. C. Carterette and M. P. Friedman. New York: Academic, 1978, p. 243–282.
- PALMER, A. R., WINTER, I. M., AND DARWIN, C. J. The representation of steady-state vowel sounds in the temporal discharge patterns of the guinea pig cochlear nerve and primarylike cochlear nucleus neurons. *J. Acoust. Soc. Am.* 79: 100–113, 1986.
- PARNCUTT, R. *Harmony: A Psychoacoustical Approach*. Berlin: Springer-Verlag, 1989.
- PATTERSON, R. D. The effects of relative phase and the number of components on residue pitch. *J. Acoust. Soc. Am.* 53: 1565–1572, 1973.
- PATTERSON, R. D. Spiral detection of periodicity and the spiral form of musical scales. *Psychol. Music* 14: 44–61, 1986.
- PENROSE, L. S. AND PENROSE, R. Impossible objects: a special type of visual illusion. *Br. J. Psychol.* 49: 31–33, 1958.
- PIERCE, J. R. Periodicity and pitch perception. *J. Acoust. Soc. Am.* 90: 1889–1893, 1991.
- PLOMP, R. Pitch of complex tones. *J. Acoust. Soc. Am.* 41: 1526–1533, 1967.
- PLOMP, R. *Aspects of Tone Sensation*. London: Academic, 1976.
- PRATT, G. *Pulse Computation* (PhD thesis). Cambridge, MA: MIT, 1990.
- REES, A. AND MØLLER, A. R. Stimulus properties influencing the responses of inferior colliculus neurons to amplitude-modulated sounds. *Hear. Res.* 27: 129–143, 1987.
- REES, A. AND PALMER, A. R. Neuronal responses to amplitude-modulated and pure-tone stimuli in the guinea pig inferior colliculus, and their modification by broadband noise. *J. Acoust. Soc. Am.* 85: 1978–1994, 1989.
- REICHARDT, W. Autocorrelation, a principle for the evaluation of sensory information by the central nervous system. In: *Sensory Communication*, edited by W. A. Rosenblith. New York: MIT, 1961, p. 303–317.
- RHODE, W. S. Interspike intervals as correlates of periodicity pitch in cat cochlear nucleus. *J. Acoust. Soc. Am.* 97: 2414–2429, 1995.
- DE RIBAUPIERRE, F., GOLDSTEIN JR., M. H., AND YENI-KOMSHIAN, G. Cortical coding of repetitive acoustic pulses. *Brain Res.* 48: 205–225, 1972.

- RICHMOND, B. J., OPTICAN, L. M., AND GAWNE, T. J. Neurons use multiple messages encoded in temporally modulated spike trains to represent pictures. In: *Seeing Contour and Colour*, edited by J. J. Kulikowski and C. M. Dickenson. New York: Pergamon, 1989, p. 705–713.
- RISSET, J. C. Paradoxes de hauteur: le concept de hauteur sonore n'est pas les meme pour tout le monde. In: *7th International Congress on Acoustics*. Budapest, 1971.
- RITSMA, R. J. Frequencies dominant in the perception of the pitch of complex sounds. *J. Acoust. Soc. Am.* 42: 191–198, 1967.
- RITSMA, R. J. AND ENGEL, F. L. Pitch of frequency-modulated signals. *J. Acoust. Soc. Am.* 36: 1637–1644, 1964.
- ROULLER, E., DE RIBAUPIERRE, Y., TOROS-MOREL, A., AND DE RIBAUPIERRE, F. Neural coding of repetitive sounds in the medial geniculate body of cat. *Hear. Res.* 5: 181–100, 1981.
- SCHOUTEN, J. F. The perception of pitch. *Phillips Tech. Rev.* 5: 286–294, 1940a.
- SCHOUTEN, J. F. The residue, a new concept in subjective sound. *Proc. K. Ned. Akad. Wet.* 43: 356–365, 1940b.
- SCHOUTEN, J. F., RITSMA, R. J., AND CARDOZO, B. L. Pitch of the residue. *J. Acoust. Soc. Am.* 34: 1418–1424, 1962.
- SCHREINER, C., URBAS, J. V., AND MEHRGARDT, S. Temporal resolution of amplitude modulation and complex signals in the auditory cortex of the cat. In: *Hearing—Physiological Bases and Psychophysics*, edited by R. Klinke and R. Hartmann. Berlin: Springer-Verlag, 1983, p. 169–175.
- SCHREINER, C. E. AND LANGNER, G. Coding of temporal patterns in the central auditory system. In: *Auditory Function: Neurobiological Bases of Hearing*, edited by G. M. Edelman, W. E. Gall, and W. M. Cowan. New York: Wiley, 1988, p. 337–362.
- SCHROEDER, M. R. S. Auditory paradox based on fractal waveform. *J. Acoust. Soc. Am.* 79: 186–189, 1986.
- SCHWARTZ, D. W. F. AND TOMLINSON, R. W. W. Spectral response patterns of auditory cortex neurons to harmonic complex tones in alert monkey (*Macaca mulatta*). *J. Neurophysiol.* 64: 282–298, 1990.
- SHANNON, R. V. Temporal modulation transfer functions in patients with cochlear implants. *J. Acoust. Soc. Am.* 91: 2156–2164, 1992.
- SHEICH, H. Auditory cortex: comparative aspects of maps and plasticity. *Curr. Opin. Neurobiol.* 1: 236–247, 1991.
- SHEPARD, R. N. Circularity in judgments of relative pitch. *J. Acoust. Soc. Am.* 36: 2346–2353, 1964.
- SHOFNER, W. P. Temporal representation of rippled noise in the anteroventral cochlear nucleus of the chinchilla. *J. Acoust. Soc. Am.* 90: 2450–2466, 1991.
- SLANEY, M. AND LYON, R. F. On the importance of time—a temporal representation of sound. In: *Visual Representations of Speech Signals*, edited by M. Cooke, S. Beet, and M. Crawford. New York: Wiley, 1993, p. 95–118.
- SMALL, A. M. Periodicity pitch. In: *Foundations of Modern Auditory Theory*, edited by J. V. Tobias. New York: Academic, 1970, p. 3–54.
- SMOORENBURG, G. F. Pitch perception of two-frequency stimuli. *J. Acoust. Soc. Am.* 48: 924–942, 1970.
- STEINSCHNEIDER, M., AREZZO, J., AND VAUGHAN, H. G., JR. Phase locked cortical responses to a human speech sound and low frequency tones in the monkey. *Brain Res.* 198: 75–84, 1980.
- TANK, D. W. AND HOPFIELD, J. J. Neural computation by concentrating information in time. *Proc. Natl. Acad. Sci. USA* 84: 1896–1900, 1987.
- TERHARDT, E. Pitch, consonance, and harmony. *J. Acoust. Soc. Am.* 55: 1061–1069, 1973.
- TERHARDT, E., STOLL, G., AND SEEWANN, M. Pitch of complex signals according to virtual-pitch theory: test, examples, and predictions. *J. Acoust. Soc. Am.* 71: 671–678, 1982a.
- TERHARDT, E., STOLL, G., AND SEEWANN, M. Algorithm for extraction of pitch and pitch salience from complex tonal signals. *J. Acoust. Soc. Am.* 71: 679–688, 1982b.
- TORRAS I GENIS, C. *Temporal-Pattern Learning in Neural Models*. Berlin: Springer-Verlag, 1985.
- TRAMO, M. J., CARIANI, P., AND DELGUTTE, B. Representation of tonal consonance and dissonance in the temporal firing patterns of auditory nerve fibers. *Soc. Neurosci. Abstr.* 18: 382, 1992.
- VILLA, A. E. P. Temporal aspects of information processing in the central nervous system. *Annales du groupe CARNAC (Swiss Federal Institute of Technology EPFL)* 5: 15–42, 1992.
- WANG, X. AND SACHS, M. B. Neural encoding of single-formant stimuli in the cat. II. Responses of anteroventral cochlear nucleus units. *J. Neurophysiol.* 71: 59–78, 1994.
- WARREN, R. M. *Auditory Perception: A New Synthesis*. New York: Pergamon, 1982.
- WASSERMAN, G. S. Isomorphism, task dependence, and the multiple meaning theory of neural coding. *Biol. Signals* 1: 117–142, 1992.
- WIERSINGA-POST, E. AND DUIFHUIS, H. Probability distributions of the pitch of harmonic and inharmonic three tone complexes: Symmetric mistuning. In: *Advances in Hearing Research. Proceedings of the 10th International Symposium on Hearing, June 26–July 1, 1994, Irsee, Bavaria*, edited by G. A. Manley, G. M. Klump, C. Köppl, H. Fastl, and H. Oeckinghaus. Singapore: World Scientific, 1995, p. 488–497.
- WIGHTMAN, F. L. Pitch and stimulus fine structure. *J. Acoust. Soc. Am.* 54: 397–406, 1973a.
- WIGHTMAN, F. L. The pattern-transformation model of pitch. *J. Acoust. Soc. Am.* 54: 407–416, 1973b.
- WILSON, B. S., FINLEY, C. C., ZERBI, M., AND LAWSON, D. T. *Speech Processors for Auditory Prostheses*. Quarterly Progress Report No. 7. Center for Auditory Prosthesis Research, Research Triangle Institute, 1994.
- YOST, W. A. The dominance region and ripple noise pitch: a test of the peripheral weighting model. *J. Acoust. Soc. Am.* 72: 416–425, 1982.
- ZWICKER, E. Direct comparisons between the sensations produced by frequency modulation and amplitude modulation. *J. Acoust. Soc. Am.* 34: 1425–1430, 1962.

# Neural representation of pitch through temporal autocorrelation

Peter Cariani<sup>1,2</sup>, Mark Tramo<sup>1,3</sup>, and Bertrand Delgutte<sup>1,2</sup>

<sup>1</sup>Eaton Peabody Laboratory, Massachusetts Eye & Ear Infirmary  
<sup>2</sup>Department of Otology and Laryngology, Harvard Medical School  
<sup>3</sup>Department of Neurobiology, Harvard Medical School  
243 Charles St., Boston MA 02114 USA  
tel (617) 573-4243, FAX (617) 720-4408  
email peter@epl.meei.harvard.edu

## Abstract

An enormous wealth of acoustic information is present in the temporal firing patterns of auditory neurons. Distributions of interspike intervals across neural populations in the auditory nerve and brainstem form autocorrelation-like stimulus representations that closely predict the low pitches of complex tones. Many diverse aspects of auditory perception are readily explained in terms of central analyses of these interval-based representations. To the extent that neural discharges are stimulus-locked in a given sensory system, distributions of all-order interspike intervals provide a neural representation of the stimulus autocorrelation function. These time-domain representations provide an alternative means for the nervous system to perform Fourier analysis.

## The neural coding problem

The neural coding problem – how populations of neurons represent and convey information through trains of spikes – is fundamental to our understanding how sensory systems function [1,2,29]. Although a great deal is known about neural response properties at many levels of the auditory system, we presently have only a very rudimentary understanding of how auditory forms are actually represented by the central auditory system.

There are fundamentally two basic ideas about how sensory information can be encoded in patterns of neural discharge: coding by spatial patterns of neural excitation vs. coding by temporal patterns in spike trains. These alternative coding strategies could be called, respectively, “coding-by-channel” and “coding-by-time structure.” Place-based or “labeled line” codes depend upon which particular neurons respond (i.e. which channels are activated). Strategies for pattern-recognition based on spatially-organized sensory maps and on specifically-tuned “feature detectors” both stem from this basic idea of coding by channel. Temporal codes, on the other hand, depend upon how neurons respond: the form of their response, rather than through which neural channels the message arrives, carries the message. Temporal codes depend upon either temporal patterns between spikes in a spike train or on the time-of-arrival of spikes relative to some reference event. Historically, coding-by-channel ideas developed from Mueller’s “specific nerve energies” and Helmholtz’s later resonance-place theory of auditory representation. Temporal coding ideas, on the other hand, were articulated through Rutherford’s “telephone” theory of hearing, Troland’s temporal-modulation representations for pitch and color, Wever’s volley principle, and Licklider’s duplex theory of hearing [1,22,23,40]. While the channel-coding idea has given rise to the highly developed connectionist networks of today, a theory of adaptive timing networks based on temporally-coded signals remains to be elaborated.

## Six stimuli that produce a low pitch at 160 Hz

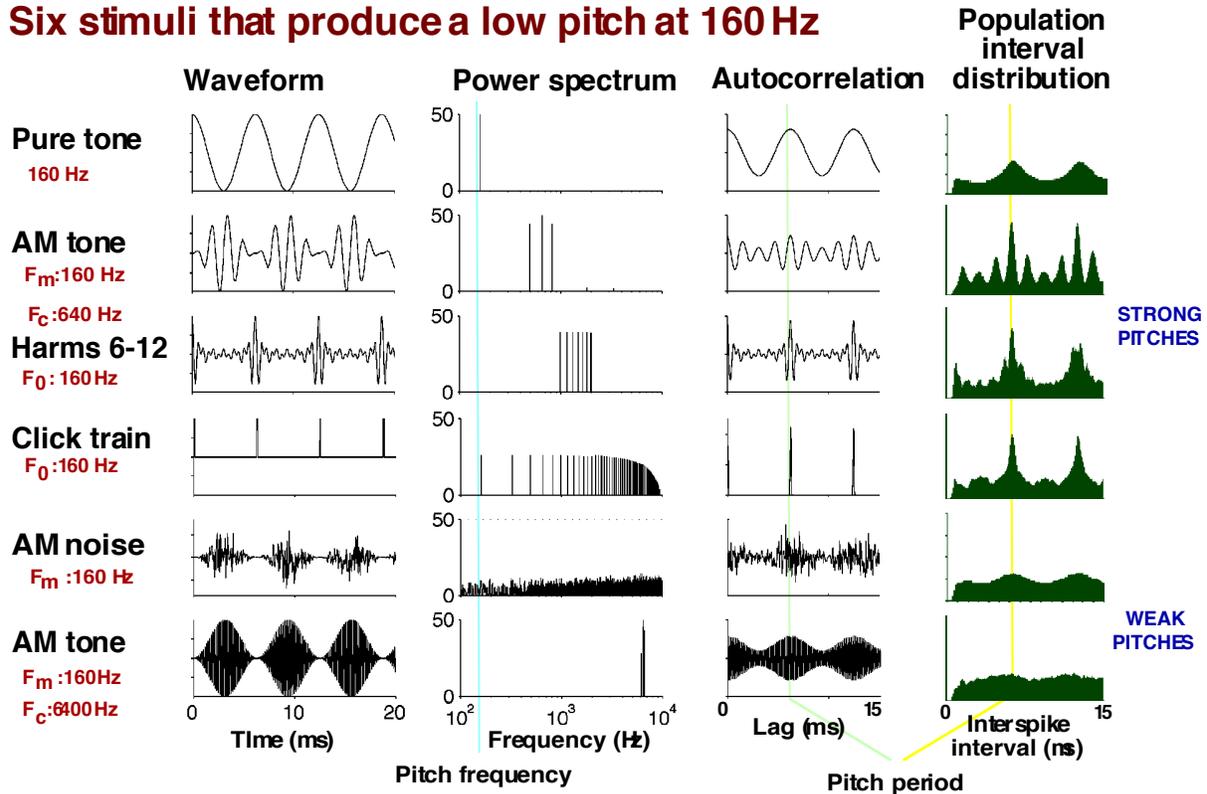


Figure 1. Neural responses to four stimuli evoking a pitch at 160 Hz but differing in pitch salience. Left to right: Stimulus waveform, power spectrum, short term autocorrelation function, and population-interval distribution for each stimulus. Population-interval distributions are constructed by summing together the all-order interspike interval distributions of many auditory nerve fibers having a wide range of characteristic frequencies (A-F: 85, 54, 49, 56, 53, 50 fibers). Histograms have been normalized to the mean number of counts/bin. Arrows indicate the position of the fundamental period (6.25 ms). A. Pure tone, 160 Hz. B. AM tone with a low frequency carrier ( $F_c=640$  Hz,  $F_m=160$  Hz). C. Harmonic complex (harmonics 6-12, of 160 Hz). D. Unipolar click train ( $F_0=160$  Hz.) E. AM tone with a high frequency carrier ( $F_c=6400$  Hz,  $F_m=160$  Hz). F. AM broadband noise with  $F_m=160$  Hz. All stimuli presented at 60 dB total SPL.

### Temporal coding of periodicity pitch

Pitch has played a pivotal role in many of the general debates about neural coding [1,12]. The mechanisms underlying the low pitches of complex tones (“periodicity pitches”) have been discussed and debated for over 150 years. Throughout this history auditory physiologists and theoreticians alike have simultaneously appreciated the great abundance of information about stimulus periodicities that temporal discharge patterns of auditory neurons carry, as well as the orderly, spatial organization of the cochlea by frequency. While this general, channel-based, “place principle” has dominated thinking about neural coding in most other sensory modalities, in audition there has always been a strong case for temporal coding of pitch. The pendulum of scientific opinion has swung back and forth between spectral pattern and temporal theories [12]. Temporal autocorrelation models for pitch held sway in the 1950’s [22,23], but with the discovery of the “dominance frequency region for pitch” in the early 1960’s, spectral pattern models regained support. Over the last decade the temporal autocorrelation models have been revived and extended [24,25,26,35,39].

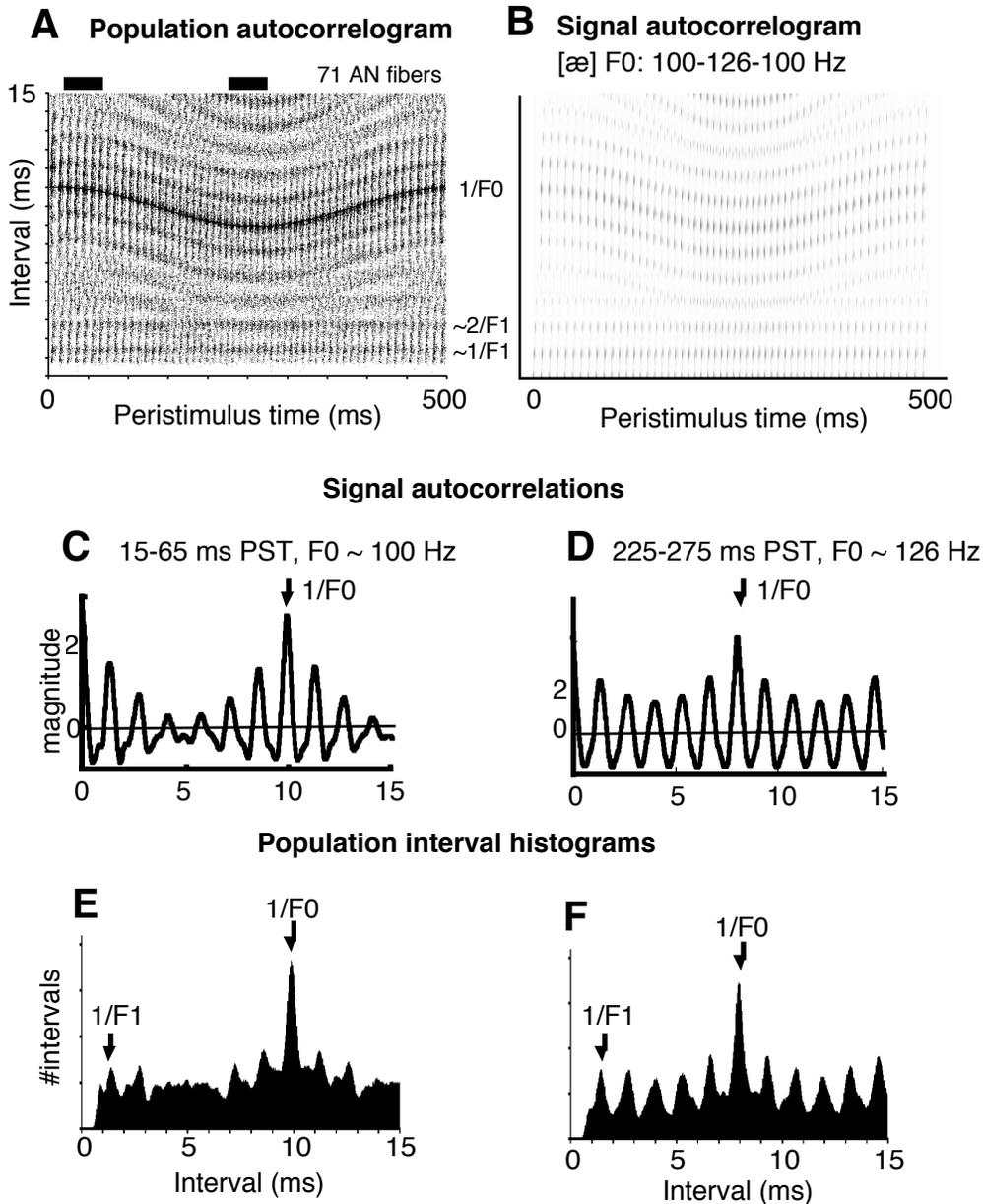


Figure 2. Autocorrelation-like representations of the vowel [æ]. Stimulus: Five-formant synthetic vowel [æ], as in "had". Formants are at 750, 1450, 2450, 3350, 3850 Hz, and the fundamental frequency  $F_0$  (heard as the voice pitch) sinusoidally varies between 100-126 Hz. A. Neural response. Running population-interval distribution (autocorrelogram) of 71 auditory nerve fibers of widely distributed characteristic frequencies. The stimulus was presented to each fiber 100 times at 60 dB SPL and running all-order interval distributions were weighted according to CF and summed. The result is an estimate of the distribution of all-order intervals across the entire auditory nerve. Each dot represents the occurrence of 10 or more intervals of a given length (y ms, range 0-15 ms) ending at a given peristimulus time (x ms, range 0-500 ms). Thin line indicates the fundamental period,  $1/F_0$ , the voice pitch that would be heard, as a function of peristimulus time. B. Stimulus autocorrelogram,  $SAC(\tau, t) = S(t)S(t-\tau)$ , computed at 10 kHz sampling rate and thresholded. C, D. Short-time autocorrelation functions for two stimulus segments indicated in A (bars). The highest peak at the fundamental period  $1/F_0$  corresponds to the voice pitch. E, F. Population interval histograms for the same segments.

In physiological studies at the level of the auditory nerve [7,8], we have found robust and pervasive correspondences between interspike interval statistics of populations of auditory nerve fibers and the pitches produced by a wide array of complex tones. In these studies we recorded the responses of many single auditory nerve fibers in Dial-anesthetized cats to complex stimuli that produce low, periodicity pitches in humans. We compiled all-order interspike interval distributions (i.e. counting intervals between both successive and nonsuccessive spikes) for each fiber. We weighted and summed the intervals from fibers according to their characteristic frequencies (CFs) in order to estimate what the all-order interval distribution would be for the entire population of auditory nerve fibers in a human listener. The result is a population-interval distribution, the distribution of intervals thought to be present in the entire auditory nerve. These kinds of interval-based representations constitute a possible means by which the auditory system might represent the structure of sounds. Such temporal neural representations complement those channel-based representations that are based on spatial patterns of discharge in auditory frequency maps.

With very few exceptions, we found that the most common all-order interval present in the population corresponds to the pitch that is heard. This can be seen in Figure 1, which shows the waveforms, power spectra, short-term autocorrelation functions for five stimuli that produce definite pitches at 160 Hz. A sixth stimulus (D) lies just outside the classical existence region for periodicity pitch, and produces a very weak, ill-defined pitch. Several of these stimuli (B, C, E) have “missing fundamentals” at 160 Hz. The population-interval distributions for these stimuli at the level of the auditory nerve are shown in the rightmost panels. In all cases, the positions of major interval peaks correspond to the period of the pitch that is heard (i.e. the fundamental period for harmonic complexes or the modulation period for AM noise) and its multiples. We found this rule to hold at low (40 dB SPL), moderate (60 dB) and high (80 dB) stimulus levels, and at all signal-to-noise ratios where the pitch could be heard. This suggests that all-order interval codes provide extremely robust representations of pitch, that, like the pitch percept itself, are not greatly distorted or degraded by high levels or background noise.

Our second major finding was that the relative proportion of pitch-related intervals amongst all intervals qualitatively corresponded to pitch strength. In Figure 1, stimuli A-D evoke strong periodicity pitches, whereas stimuli E and F evoke much weaker pitches. Correspondingly, in their respective population-interval distributions, the peak-to-background ratios of the major, pitch-related interval peaks are much higher for those stimuli (A-D) that produce strong periodicity pitches.

These findings taken together with the rest of our data suggest that many diverse aspects of pitch can be directly explained in terms of population-interval distributions at the level of the auditory nerve: the pitch of the “missing fundamental”, pitch equivalence of stimuli with very different power spectra, pitch shifts and pitch ambiguities produced by inharmonic AM tones, the relative phase- and level-invariance of periodicity pitches, pitches produced by unresolved harmonics and by AM noise, and the dominance (frequency) region for pitch. From studies of neural responses in the auditory brainstem [3,4,19,31], it appears that population-interval distributions can serve as representations for periodicity pitch in the central auditory system, although the extent to which pitch-related timing information exists at the level of primary auditory cortex is still unclear.

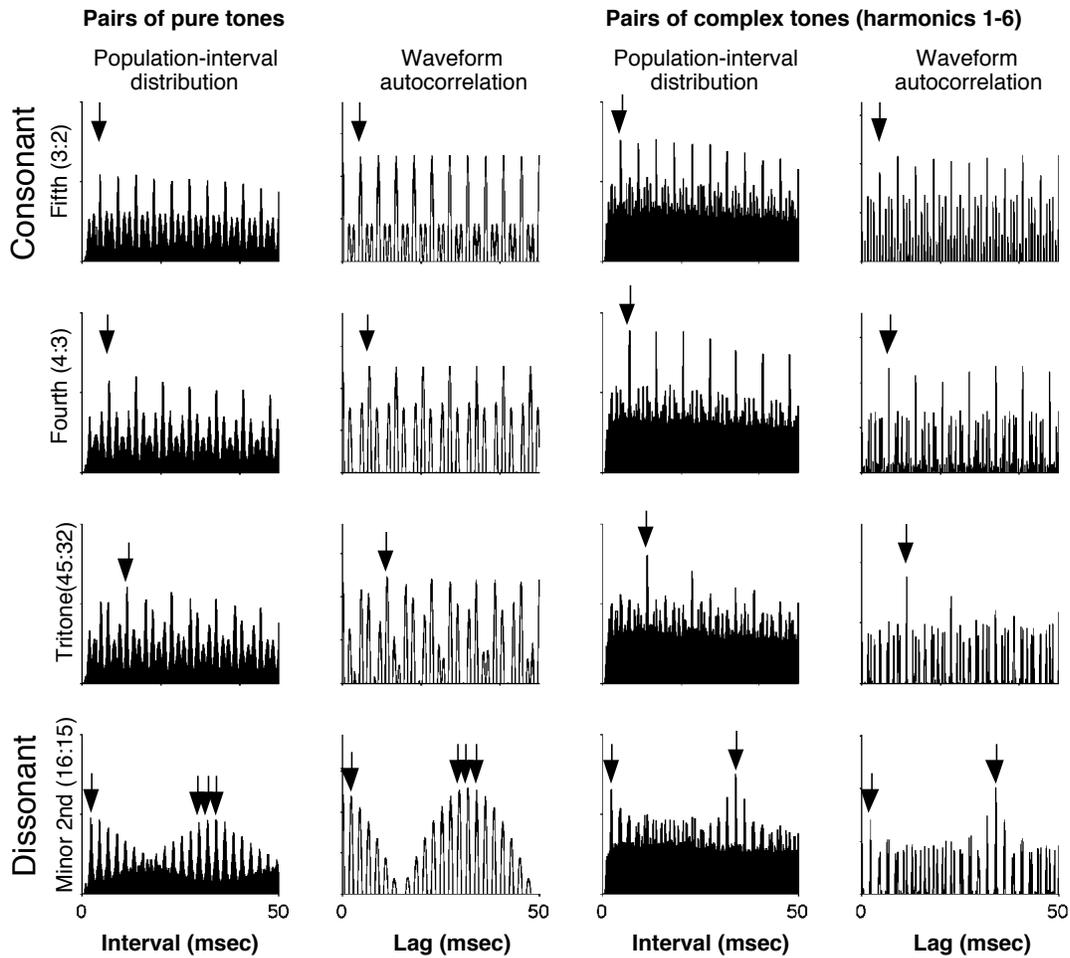


Figure 3. Population-interval neural representations and corresponding stimulus autocorrelation functions for five musical intervals. Stimuli were pairs of pure tones or pairs of complex tones consisting of harmonics 1-6 (equal amplitudes), with the lower fundamental frequency = 440 Hz. For complex tones, the fifth (3:2) and the fourth (4:3) are generally regarded as consonant intervals with the tritone (45:32) and minor second (16:15) being dissonant ones. For pure tones, the rank ordering follows frequency separation, with the tritone being slightly more consonant than the fourth. All 200 msec stimuli were presented at 60 dB SPL, 100 times to each auditory nerve fiber (ANF) through a closed, calibrated acoustic assembly. Population-interval distributions were compiled by summing together the all-order interspike interval distributions of 50-100 ANF's distributed across a wide range of characteristic frequencies (CF's). Positive portions of stimulus autocorrelation functions are shown. Arrows indicate positions of major peaks in population-interval and autocorrelation plots that may correspond to low pitches.

## Temporal autocorrelation

In our investigations of the representation of pitch at the level of the auditory nerve we also observed that the forms of population-interval distributions resembled their respective stimulus autocorrelation functions. In this paper we will discuss this observation more deeply, outlining in what sense these population-interval distributions might constitute autocorrelation-like representations of the stimulus. We will briefly explore what this might mean in terms of auditory representations and neural information processing.

Correlational analysis [20,41] was a much more widely used for the analysis of complex signals in the decades before the “rediscovery” of the fast Fourier transform in the 1960’s and the general availability of digital signal processing. Delay lines were common in the computing machinery of the day, and analog autocorrelators existed for real-time analysis of time-series inputs [20]. Temporal autocorrelation and cross-correlation architectures were proposed for the representation and the separation of auditory forms [9,23,36]. In other sensory modalities, temporal correlation mechanisms successfully explained many aspects of motion detection in the fly visual system [30] and flutter-vibration frequency discrimination in the somatosensory system [27]. Interspike intervals and temporal discharge patterns have also been proposed as means of multiplexing different kinds of visual information in neural pathways [10,32].

Temporal correlation functions entail the multiplication of a time-series signal by another signal at different relative time delays ( $\tau$ ). Cross-correlation entails the multiplication of two different signals, while autocorrelation entails the multiplication of a signal by itself. Because the output of the multiplication operation is a joint property of both signals, correlation functions reflect patterns of joint occurrences of events in time.

Correlation functions are intimately related to Fourier transforms. Temporal correlations are expressed as functions of time delays (time domain), while Fourier transforms and power spectra are expressed as functions of frequency (frequency domain). Power spectra can be obtained by computing the Fourier transform of the autocorrelation function. Because this operation is reversible (invertible), the autocorrelation function of a signal contains exactly the same information about a stimulus as its power spectrum. Compared to the original waveform, the autocorrelation function and the power spectrum retain periodicity information while discarding phase information.

Currently the power spectrum and the spectrogram are most commonly used representations of sound. Correspondingly, the auditory system is most often conceptualized of in terms of spatially-distributed spectral representations, where particular sub-populations of auditory neurons are tuned to particular frequency ranges. The profile of average discharge rates across these tonotopic auditory frequency maps in effect provides a neural representation of the stimulus power spectrum. These rate-place representations are channel-based: in their simplest forms, each neuron represents the frequency to which it is maximally tuned, the discharge rate encoding in some fashion the amount of stimulus energy in that frequency neighborhood.

Neural time domain representations of the stimulus waveform consist of times of arrivals in the spike trains themselves. Time domain representations of the stimulus autocorrelation function consist of distributions of interspike intervals, the times between spikes. The autocorrelation function is simply a time-series multiplied by itself as a function of relative time delay:  $f(\tau) = \sum S(t) S(t-\tau)$ , summed over all times  $t$ . Typically we compute the (retrospective) autocorrelation function only for short, positive time lags that are in the periodicity pitch range (e.g. 0-15 ms), thereby reducing the amount of computation required. Autocorrelation functions are intimately related to interspike interval histograms. If a spike train is described in terms of a time-series of 0’s (no spike

in a given time-bin) and 1's (spike in a given time-bin), then the retrospective autocorrelation function of the spike train is the same as a histogram of the time intervals between all possible pairs of spikes in the train, i.e. between both successive and non-successive spikes.

### **Temporal autocorrelation and all-order interspike intervals**

We found through observation that the delay-positions of major and minor peaks in population-interval distributions closely mirror those of the stimulus autocorrelation function. This can be readily appreciated by comparing in Figure 1 the population-interval distributions with their respective short-time stimulus autocorrelation functions. This correspondence holds to the extent that there is phase-locking, i.e. for periodicities up to a few kHz. While all of the major and minor peaks in autocorrelation of the AM tone with the low-frequency carrier (B) are replicated in its population-interval response, the corresponding fine structure of the AM tone with the high-frequency carrier (E) is completely absent: only a very shallow remnant of the envelope remains.

In another series of experiments, we observed the responses of auditory nerve fibers to five synthetic vowels [5,6]. These vowels had fundamental frequencies (F0s) that varied sinusoidally between 100-126 Hz. Stimulus autocorrelation functions and population-interval distributions for the vowel [ae] are shown in Figure 2. The population autocorrelogram, or running population-interval distribution, and the signal autocorrelogram are shown in panels A and B. The voice pitch of such a vowel is heard at the fundamental period. The densest interval band in the population autocorrelogram (A) closely followed the fundamental period and voice pitch period throughout the entire range of fundamental frequencies presented. There is strong correspondence between the structure of the population and signal autocorrelograms in panels A and B.

In panels C-F the signal short-time autocorrelation functions and population interval histograms are shown for two vowel segments, when  $F_0 = 100$  Hz and when  $F_0 = 126$  Hz. Major peaks at 10 ms and at 8 ms correspond to respectively to the fundamental periods (and voice pitches) or the two segments. Each vowel has a characteristic autocorrelation function, and we found that in general the population-interval distributions resembled their respective stimulus autocorrelation functions. Essentially each vowel's characteristic format structure sets up characteristic autocorrelation and interval patterns. This is consistent with previous physiological observations from both single ANFs and ensembles of ANFs [14,28]. We found that vowels could be discriminated on the basis of population-interval distributions consisting of short intervals (0-5 ms), i.e. on the basis of temporal information alone. This suggests that the timbres of stationary sounds with low- and medium-frequency components may be explicable in purely temporal terms [24]. In other experiments we have found that changes in population-interval distributions covary with vowel-class boundaries [18].

We have also used stimuli that have varying qualities of musical consonance or dissonance [38]. These stimuli consisted of either pairs of pure tones or pairs of complex tones (harmonics 1-6). The (fundamental) frequency of the first tone was always 440 Hz. The second frequency was separated from the first by various musical intervals (frequency ratios): fifth (3:2), fourth (4:3), tritone (45:32), or minor second (16:15). The population-interval distributions and the positive portions of their respective stimulus autocorrelation functions are shown in Figure 3. We are currently analyzing the neural responses to these stimuli using a variety of representations, including population-interval distributions, to explore possible neural correlates of musical consonance, roughness, fusion, and distributions of perceived pitch(es). In music theory, which primarily deals with relations between the complex tones produced by musical instruments, the musical fifth and fourth are generally considered to be consonant intervals, whereas the tritone and

the minor second are generally considered as dissonant. Qualitatively, the more consonant stimuli produce population-interval distributions with simpler, repetitive interval patterns, whereas the dissonant stimuli produce more complex, less repetitive ones. What is most striking in this context, however, is the similarity between the population-interval distributions and their respective stimulus autocorrelation functions. For virtually all sets of stimulus autocorrelation peaks there exist corresponding sets of interval peaks. The one exception is for the pure tone fourth (4:3), where there is an extra set of small peaks at half the fundamental period and its multiples (this corresponds to the distortion product  $2f_1 - f_2$ , which, for the 4:3 ratio of the fourth, equals  $2F_0$ ).

In retrospect, the reasons that population-interval distributions should resemble stimulus autocorrelation functions are fairly straightforward. They depend mainly on the phase-locked nature of auditory nerve fiber discharges. Each stimulus component produces discharges that are phase-locked discharges to itself, predominantly, but not limited to, those auditory nerve fibers whose characteristic frequencies (CF) are nearby. In doing so, intervals at subharmonics, integer multiples of the component's period, are produced. If the stimulus is a harmonic complex, then all stimulus components have a common subharmonic at the fundamental. When all of the intervals corresponding to all of the subharmonics are summed together in a population interval distribution, the most common intervals are invariably at the fundamental period and its multiples (i.e. the fundamental frequency and its subharmonics). This is the time-domain equivalent of Terhardt's frequency-based method of subharmonic coincidence [37]. If interspike intervals are the means by which the auditory system represents pitch, then central auditory analyzers interpret the interval-pattern associated with the fundamental frequency (even if it is "missing") as a low pitch. Thus, the perception of periodicity pitch could well be a direct consequence of the basic neural codes that the auditory system uses coupled with the phase-locked, stimulus-driven character of its neural discharges. If this is the case, pitch judgments are well-described by temporal autocorrelation models [11,12] precisely because the neural representations that subserve those judgments are themselves autocorrelation-like.

The similarity extends beyond the patterns of major peaks that are associated with periodicity pitch. Minor interval peaks are produced by other combinations of intervals (subharmonics of stimulus components). These patterns repeat at each fundamental period in the autocorrelation function (note the repeating patterns for the consonant stimuli and responses of Figure 3 and in the autocorrelations of Figure 2 in panels D and F). Different vowels with different formant regions (and different timbres) give rise to different repeating patterns of minor peaks that are nested within the  $F_0$ -related major peaks. It is not surprising then, that the delay-positions of interval peaks should mirror those of the stimulus, since each stimulus component produces intervals related to it.

In some cases (Figure 3), the relative heights of peaks are similar in stimulus autocorrelations and population-interval distributions, whereas in others (Figs. 1B and 2), the relative heights are noticeably different. These similarities and differences may stem from the degree to which cochlear filtering and spike initiation are linear processes. Autocorrelations of individual frequency components summed together equal the autocorrelation of the whole [22]. For an array of contiguous band-pass linear filters of uniform bandwidth, the sum of the channel autocorrelations is proportional to the autocorrelation of the unfiltered stimulus. To the extent that the production of intervals is the result of a linear process, then summing the intervals should yield relative amplitudes that mirror the stimulus autocorrelation function. To the extent that nonlinearities are created by cochlear filtering, hair cell transduction (half-wave rectification), synaptic transmission, nonuniform distributions of characteristic frequencies of auditory nerve fibers or their rate-level

functions (threshold and saturation effects), the respective heights of interval peaks will diverge from their counterparts in the stimulus autocorrelation function.

The functional effects of nonlinearities in the auditory system depend critically on the nature of the neural representations, i.e. what aspects of the neural signal are actually used by the auditory system to subserve a given auditory percept. Interval-based representations of stimulus periodicities appear to be relatively resilient to the introduction of many of the above-mentioned nonlinearities. Nonlinear changes in discharge rates with level do not distort the time intervals that correspond to a particular frequency component – they merely cause relatively fewer or more intervals associated with that component to be produced. Combination tones created by nonlinear distortion produce sets of related intervals that either augment those associated with stimulus components or add entirely new sets of intervals to the distribution [16], as was seen for the pure tone fourth in Figure 3. In population-interval distributions the delay positions of the interval peaks themselves are generally unchanged, only the relative heights of peaks are altered. Thus, under a population-interval code, information about the frequency of a stimulus component is generally not degraded, while the information concerning the relative intensity of that component may be considerably distorted by nonlinearities. Population-interval distributions therefore appear to be more faithful in their representation of the frequencies of stimulus components that are present than they are in their representation of the relative intensities of those components. In many ways this behavior parallels our auditory perceptions. Subtle changes in the relative levels of stimulus components generally affect the quality of sounds far less than comparable changes in component frequencies. For example, for pure tones at 1 kHz and moderate levels, the difference limens for intensity, expressed in terms of Weber fractions  $(I + \Delta I)/I$ , are some 40 times those for frequency,  $(f + \Delta f)/f$  [33]. Interestingly, given the discharge properties of auditory nerve fibers, autocorrelation analysis is by far the decision strategy that most closely approaches the performance of the ideal pure tone frequency discriminator [13,17,34]. Independent of whether the central auditory system utilizes such temporal coding strategies to represent auditory forms, receptor arrays capable of phase-locking coupled with temporal autocorrelation analysis offer extremely powerful and robust strategies for discriminating sounds, strategies that we have only barely begun to incorporate into devices for processing audio signals [15,21,24,35].

## Conclusions

The potential implications of autocorrelation-like representations in the auditory system are many. Neural codes based on interspike intervals allow the stimulus power spectrum to be represented and analyzed in the time domain. To the extent that there is phase-locking of neural discharges to stimulus components, such interval-based codes can form the basis of stimulus representations that complement spectral, tonotopically-based rate-place ones. Periodicity pitch along with many other aspects of auditory perception may thus be direct consequences of the kinds of temporally-based neural representations that the auditory system employs for the analysis of sounds. Major questions remain for how such temporal information might be utilized by the central auditory system to give rise to some of the qualities of sound that we hear. We need to better understand the extent to which the timing information that we observe in the auditory nerve is available at higher auditory centers as well as the neural computational strategies by which this information might be effectively used.

## Acknowledgments

This work was supported by Grant DC03054 from the National Institute for Deafness and Communications Disorders (NIDCD), of the National Institutes of Health (NIH).

## References

1. Boring, E.G. 1942, *Sensation and Perception in the History of Experimental Psychology*, Appleton-Century-Crofts, New York.
2. Cariani, P., 1995, As if time really mattered: temporal strategies for neural coding of sensory information, *Communication and Cognition - Artificial Intelligence (CC-AI)* 12(1-2): 161-229. Reprinted in: *Origins: Brain and Self-Organization*, (K. Pribram, ed.), Lawrence Erlbaum, Hillsdale, NJ, 1995.
3. Cariani, P., 1995, Physiological correlates of periodicity pitch in the cochlear nucleus, *Assoc. Res. Otolaryn. Abstr.* : 128.
4. Cariani, P., 1997, Population-interval representations of pitch in the auditory brainstem, *Assoc. Res. Otolaryn. Abstr.* :
5. Cariani, P. and B. Delgutte, 1993, Interspike interval distributions of auditory nerve fibers in response to concurrent vowels with same and different fundamental frequencies, *Assoc. Res. Otolaryngology. Abs.* : 373.
6. Cariani, P. and B. Delgutte, 1994, Transient changes in neural discharge patterns may enhance separation of concurrent vowels with different fundamental frequencies [Abstract], *J. Acoust. Soc. Am.* 95(5(2)): 2842.
7. Cariani, P.A. and B. Delgutte, 1996, Neural correlates of the pitch of complex tones. I. Pitch and pitch salience., *J. Neurophysiol.* 76(3): 1698-1716.
8. Cariani, P.A. and B. Delgutte, 1996, Neural correlates of the pitch of complex tones. II. Pitch shift, pitch ambiguity, phase-invariance, pitch circularity, and the dominance region for pitch, *J. Neurophysiol.* 76(3): 1717-1734.
9. Cherry, C., 1961, Two ears – but one world, in: *Sensory Communication*, (W. A. Rosenblith, ed.), MIT Press/John Wiley, New York.
10. Chung, S.H., S.A. Raymond and J.Y. Lettvin, 1970, Multiple meaning in single visual units, *Brain Behav Evol* 3: 72-101.
11. de Boer, E. *On the "residue" in hearing*. Ph.D., University of Amsterdam (1956)
12. de Boer, E., 1976, On the "residue" and auditory pitch perception, in: *Handbook of Sensory Physiology*, (W. D. Keidel and W. D. Neff, ed.), Springer Verlag, Berlin.
13. Delgutte, B., 1995, Physiological models for basic auditory percepts, in: *Auditory Computation*, (H. Hawkins, T. McMullin, A. N. Popper and R. R. Fay, ed.), Springer Verlag, New York.
14. Delgutte, B. and N.Y.S. Kiang, 1984, Speech coding in the auditory nerve: I. Vowel-like sounds, *J. Acoust. Soc. Am.* 75(3): 866-878.
15. Ghitza, O., 1992, Auditory nerve representation as a basis for speech processing, in: *Advances in Speech Signal Processing*, (S. Furui and M. M. Sondhi, ed.), Marcel Dekker, New York.
16. Goldstein, J.L. and N.Y.S. Kiang, 1968, Neural correlates of the aural combination tone  $2f_1-f_2$ , *IEEE Proc* 56: 981-992.
17. Goldstein, J.L. and P. Sruлович, 1977, Auditory-nerve spike intervals as an adequate basis for aural frequency measurement, in: *Psychophysics and Physiology of Hearing*, (E. F. Evans and J. P. Wilson, ed.), Academic Press, London.
18. Hirahara, T., P. Cariani and B. Delgutte, 1996, Representation of low-frequency vowel formants in the auditory nerve, in: *Proceedings, ESCA Research Workshop on The Auditory Basis of Speech Perception, Keele University, United Kingdom, July 15 - 19, 1996*, ed.),
19. Kim, D.O. and G. Leonard, 1988, Pitch-period following response of cat cochlear nucleus neurons to speech sounds, in: *Basic Issues in Hearing*, (H. Duifhuis, J. W. Horst and H. P. Wit, ed.), Academic Press, London.
20. Lange, F.H. 1967, *Correlation Techniques*, Van Nostrand, Princeton.
21. Lazzaro, J. and C. Mead, 1989, Silicon modeling of pitch perception, *Proc. Nat. Acad. Sci. U.S.A.* 86: 9597-9601.
22. Licklider, J.C.R., 1951, A duplex theory of pitch perception, *Experientia* VII(4): 128-134.
23. Licklider, J.C.R., 1959, Three auditory theories, in: *Psychology: A Study of a Science. Study I. Conceptual and Systematic*, (S. Koch, ed.), McGraw-Hill, New York.

24. Lyon, R. and S. Shamma, 1995, Auditory representations of timbre and pitch, in: *Auditory Computation*, (H. Hawkins, T. McMullin, A. N. Popper and R. R. Fay, ed.), Springer Verlag, New York.
25. Meddis, R. and M.J. Hewitt, 1991, Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I. Pitch identification, *J. Acoust. Soc. Am.* 89(6): 2866-2882.
26. Moore, B.C.J. 1989, *Introduction to the Psychology of Hearing*, Academic Press, London.
27. Mountcastle, V., 1993, Temporal order determinants in a somatosthetic frequency discrimination: sequential order coding, *Annals New York Acad. Sci.* 682: 151-170.
28. Palmer, A.R., 1988, The representation of concurrent vowels in the temporal discharge patterns of auditory nerve fibers, in: *Basic Issues in Hearing*, (H. Duifhuis, J. W. Horst and H. P. Wit, ed.), Academic Press, London.
29. Perkell, D.H. and T.H. Bullock, 1968, Neural Coding, *Neurosciences Research Program Bulletin* 6(3): 221-348.
30. Reichardt, W., 1961, Autocorrelation, a principle for the evaluation of sensory information by the central nervous system, in: *Sensory Communication*, (W. A. Rosenblith, ed.), MIT Press/John Wiley, New York.
31. Rhode, W.S., 1995, Interspike intervals as correlates of periodicity pitch in cat cochlear nucleus, *J. Acoust. Soc. Am.* 97(4): 2414-2429.
32. Richmond, B.J., L.M. Optican and T.J. Gawne, 1989, Neurons use multiple messages encoded in temporally modulated spike trains to represent pictures, in: *Seeing Contour and Colour*, (J. J. Kulikowski and C. M. Dickenson, ed.), Pergamon Press, New York.
33. Siebert, W.M., 1965, Some implications of the stochastic behavior of primary auditory neurons, *Kybernetik* 2: 206-215.
34. Siebert, W.M., 1970, Frequency discrimination in the auditory system: place or periodicity mechanisms?, *Proc IEEE* 58: 723-730.
35. Slaney, M. and R.F. Lyon, 1993, On the importance of time - a temporal representation of sound, in: *Visual Representations of Speech Signals*, (M. Cooke, S. Beet and M. Crawford, ed.), John Wiley, New York.
36. Stevens, K.N., 1950, Autocorrelation analysis of speech sounds, *J. Acoust. Soc. Am.* 22(6):769-71.
37. Terhardt, E., G. Stoll and M. Seewann, 1982, Algorithm for extraction of pitch and pitch salience from complex tonal signals, *J. Acoust. Soc. Am.* 71(3): 679-688.
38. Tramo, M.J., P. Cariani and B. Delgutte, 1992, Representation of tonal consonance and dissonance in the temporal firing patterns of auditory nerve fibers, *Soc. Neurosci. Abstr.* 18: 382.
39. van Noorden, L., 1982, Two channel pitch perception, in: *Music, Mind and Brain*, (M. Clynes, ed.), Plenum, New York.
40. Wever, E.G. 1949, *Theory of Hearing*, Wiley, New York.
41. Wiener, N. 1961, *Cybernetics: or Control and Communication in the Animal and in the Machine*, MIT Press, Cambridge, MA.