

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 μ 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.

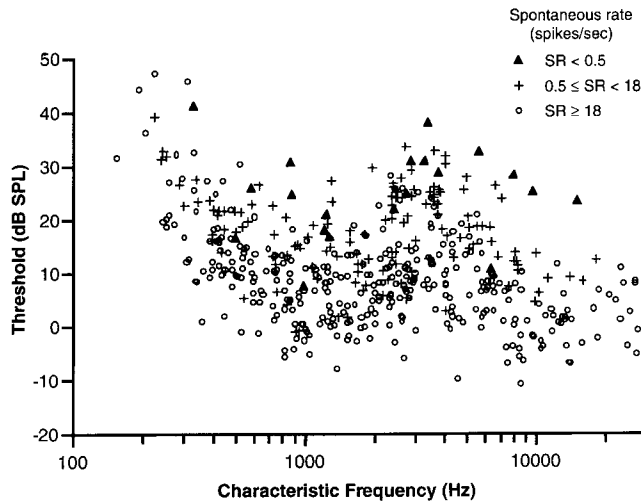


FIG. 1. Characteristic frequencies (CFs), discharge rate thresholds, and spontaneous rates (SRs) for the 507 auditory nerve fibers (ANFs) studied.

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 ~ 60 dB SPL was ≥ 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- μ s clicks Uniform-polarity 40- μ s clicks Alternating-polarity 40- μ 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

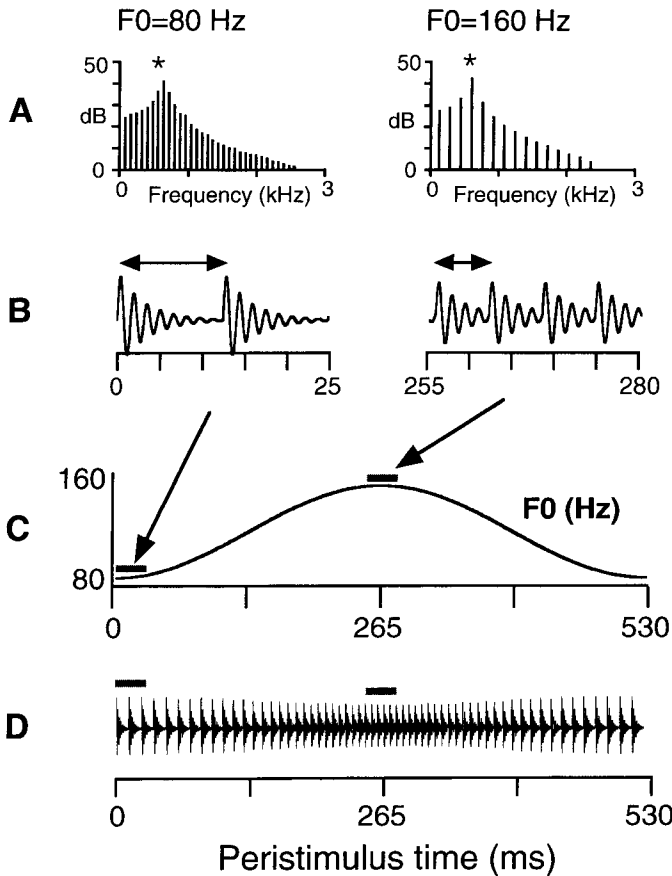


FIG. 2. Variable-fundamental stimulus. Single-formant vowel, fundamental frequency (F_0) = 80–160 Hz, formant frequency (F_1) = 640 Hz, duration = 530 ms. *A*: stimulus power spectra for 2 stimulus time segments, when F_0 = 80 Hz (*left*) and when F_0 = 160 Hz (*right*). Asterisks: F_1 = 640 Hz. *B*: 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 (Lieberman 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- μ s binwidths were then smoothed by taking a 300- μ 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

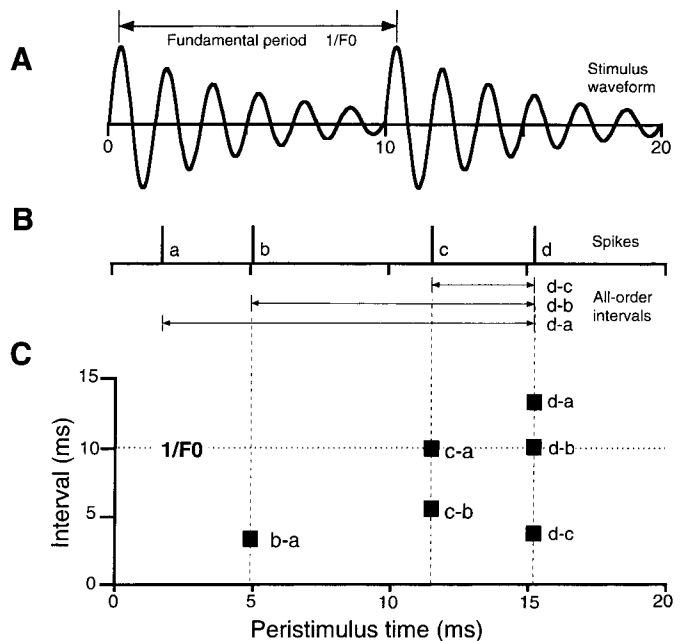


FIG. 3. Autocorrelogram construction. *A*: 2 periods from the waveform of a single-formant vowel, F_0 = 100 Hz, F_1 = 40 Hz. *B*: hypothetical ANF spike train showing 4 spikes labeled a-d. Double arrows: all-order interspike intervals that end with spike d. *C*: autocorrelogram computed from all of the 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.

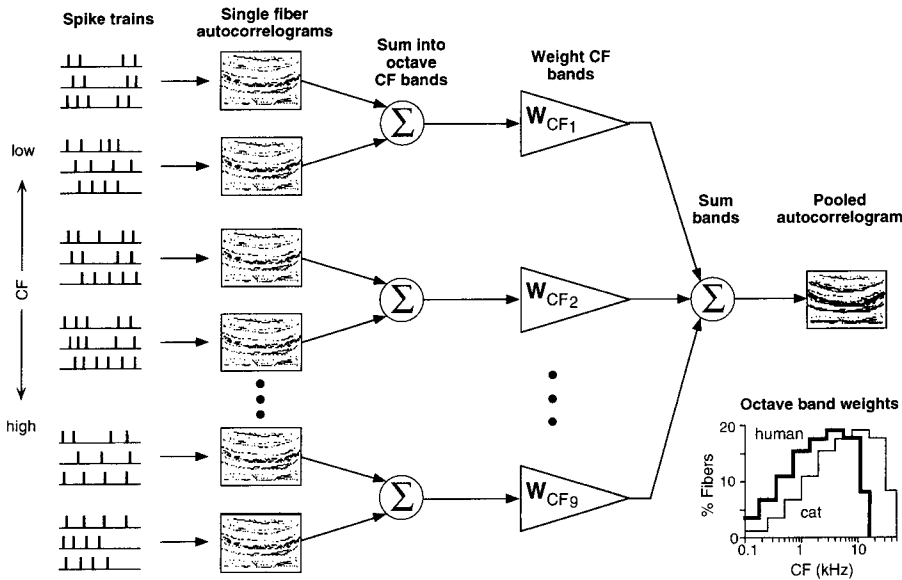


FIG. 4. Construction of pooled autocorrelograms. Autocorrelograms are 1st constructed directly from response spike trains of each fiber. Autocorrelograms for fibers in each octave CF band are normalized by number of stimulus presentations and summed together to form octave band autocorrelograms. These 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)

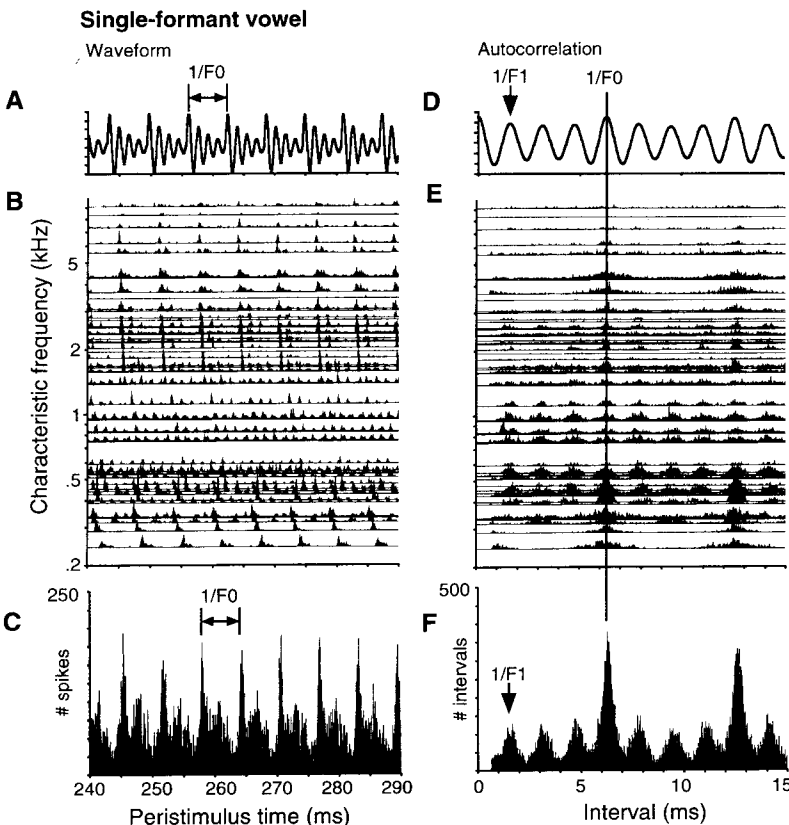


FIG. 5. ANF responses for a single-formant vowel, $F_0 = 160$ Hz, $F_1 = 640$ Hz. *A*: stimulus waveform. Interval indicates fundamental period $1/F_0$. *B*: PST histograms for 52 fibers, arranged by CF. *C*: pooled PST histogram for the 52 fibers. *D*: stimulus short-time autocorrelation function. *E*: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

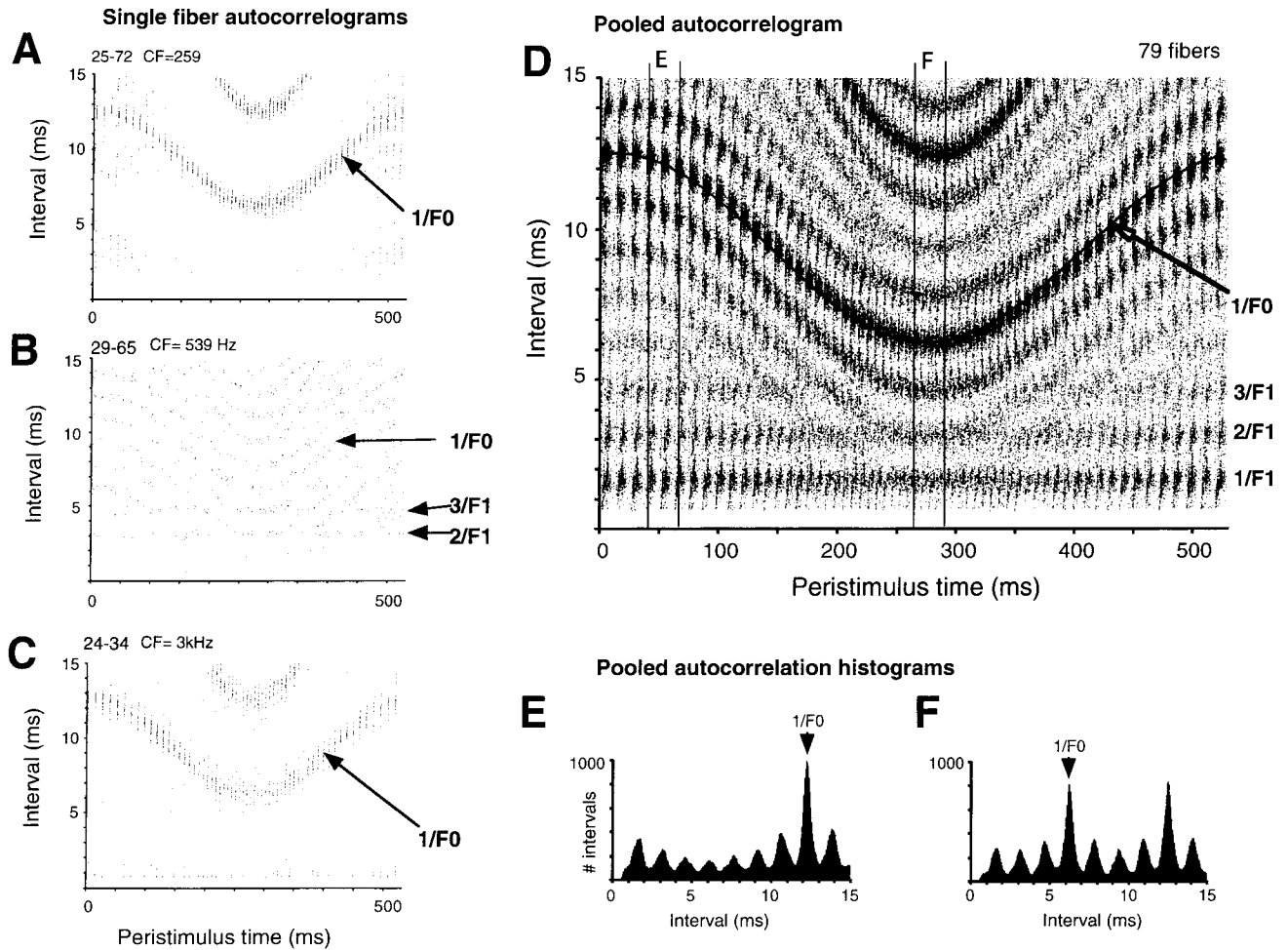


FIG. 6. Autocorrelograms and autocorrelation histograms in response to a variable- F_0 single-formant vowel, $F_0 = 80$ – 160 Hz. A–C: single-fiber autocorrelograms from 3 different CF regions. D: pooled autocorrelogram for a single-formant stimulus with $F_0 = 80$ – 160 Hz and $F_1 = 640$ Hz. PST binwidth = $714 \mu\text{s}$, 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-

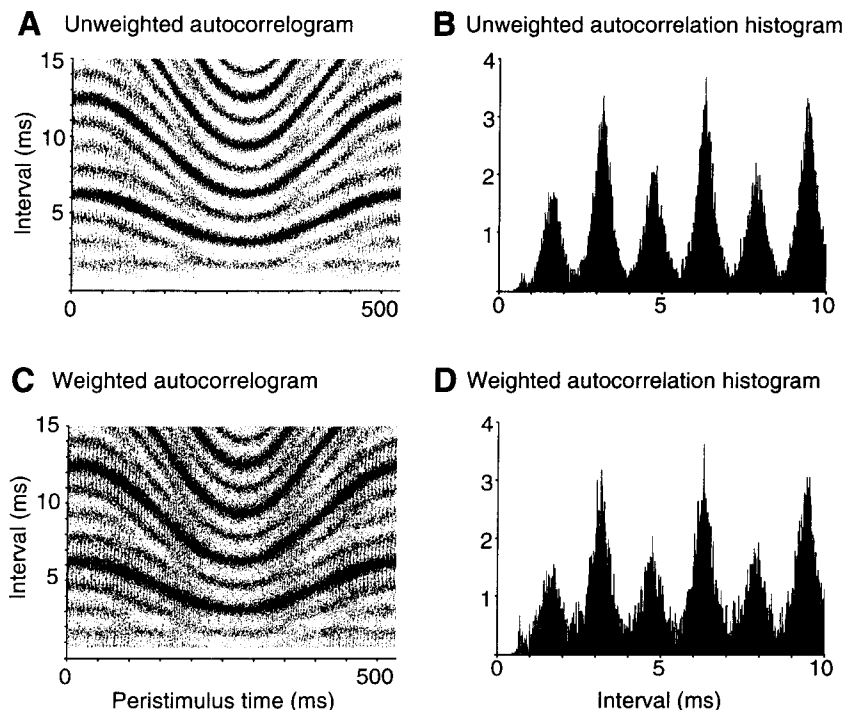


FIG. 7. Effects of explicit weighting by CF on pooled interval distributions. Stimulus: variable- F_0 single-formant vowel, $F_0 = 160\text{--}320$ Hz, $F_1 = 640$ Hz. *A* and *C*: unweighted and CF-weighted autocorrelograms. *B* and *D*: unweighted and CF-weighted pooled autocorrelation histograms compiled 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.

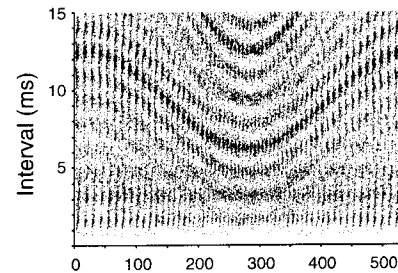
Pitch invariance with stimulus level

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).

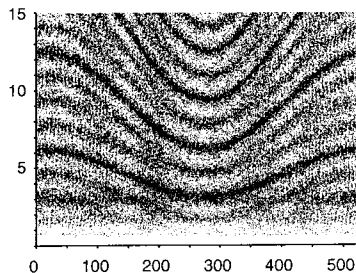
Pooled autocorrelograms for two single-formant vowels ($F_0 = 80\text{--}160$ Hz; $F_0 = 160\text{--}320$ Hz) at three stimulus levels (40, 60, and 80 dB SPL) are shown in Fig. 8. The

Single Formant Vowels

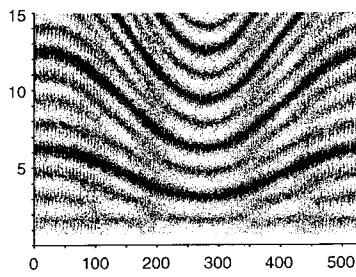
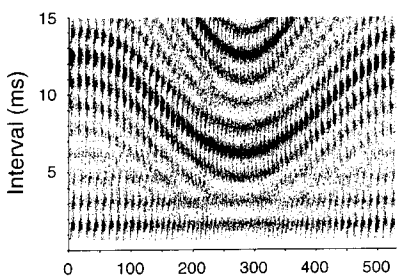
40 dB SPL F_0 : 80-160 Hz



F_0 : 160-320 Hz



60 dB SPL



80 dB SPL

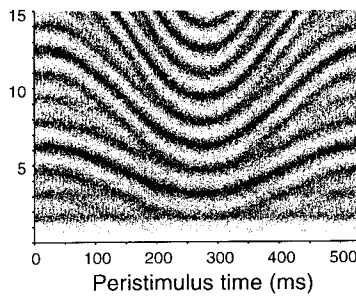
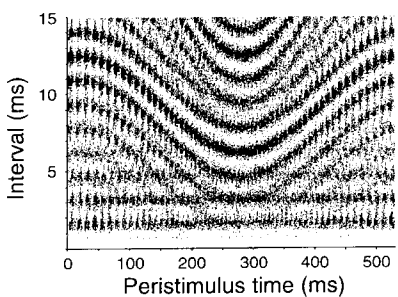


FIG. 8. Pooled response autocorrelograms for 2 variable- F_0 single-formant vowels presented at 40, 60, and 80 dB SPL. $F_1 = 640$ Hz. *Left*: $F_0 = 80-160$ Hz; $N = 26, 79,$ and 50 fibers, respectively. *Right*: $F_0 = 160-320$ Hz; $N = 29, 70,$ and 36 fibers, respectively.

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

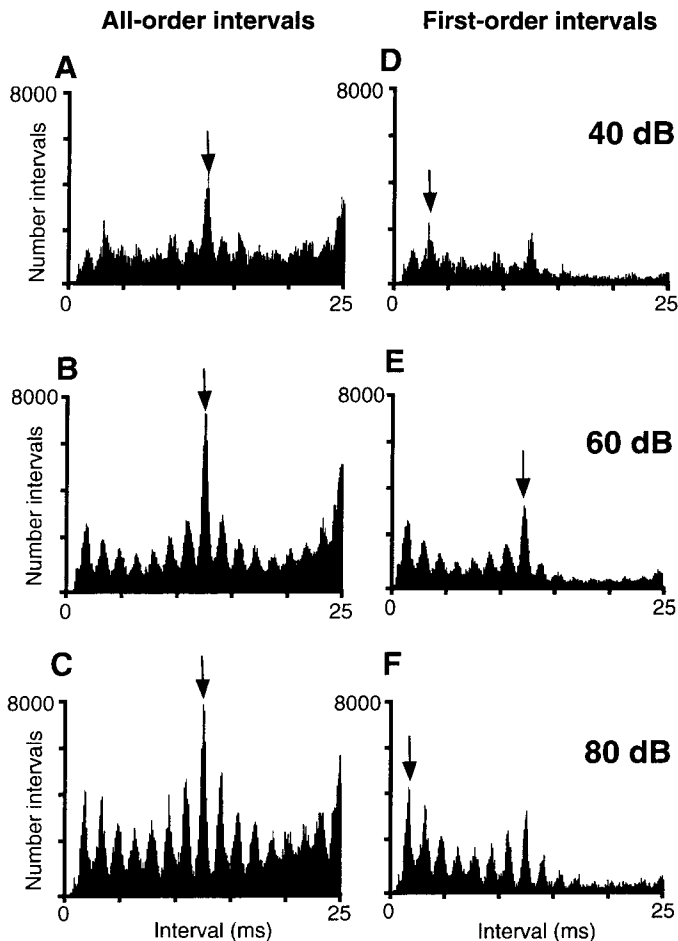


FIG. 9. Distributions of all-order vs. 1st-order intervals as a function of level. Arrows: pitch estimates. *A–C*: weighted pooled distributions of all-order intervals for a single-formant vowel, $F_0 = 80$ Hz, $F_1 = 640$ Hz, presented at 40, 60, and 80 dB SPL. *D–F*: weighted pooled distributions of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

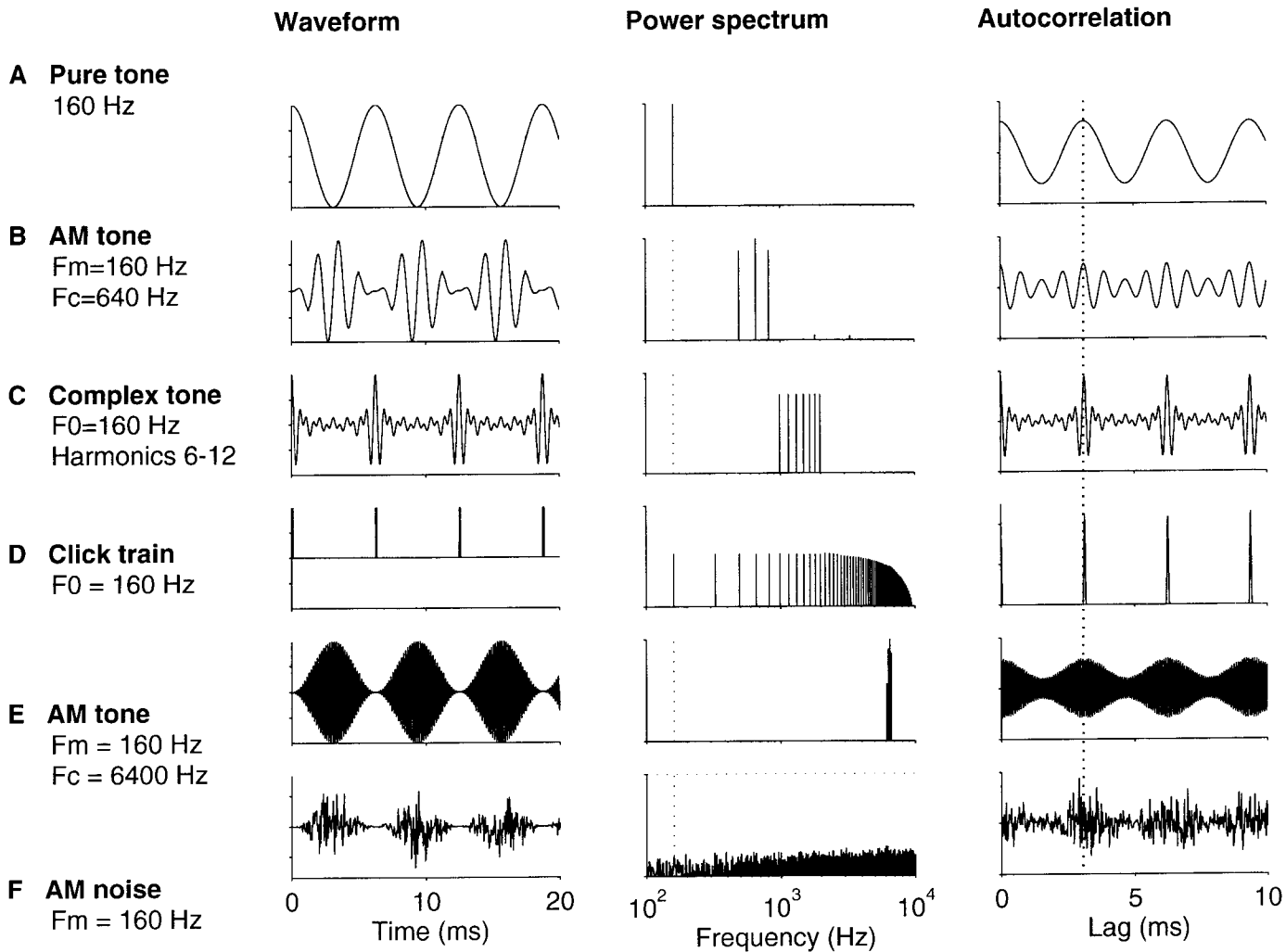


FIG. 10. A–F: waveforms, long-term power spectra, and short-term autocorrelation functions for 6 stimuli that evoke a pitch at 160 Hz. Magnitude range for power spectra is 50 dB. Dotted lines in power spectra: position of missing F_0 for harmonic stimuli (B–D), and modulation frequency (F_m) for amplitude-modulated (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).

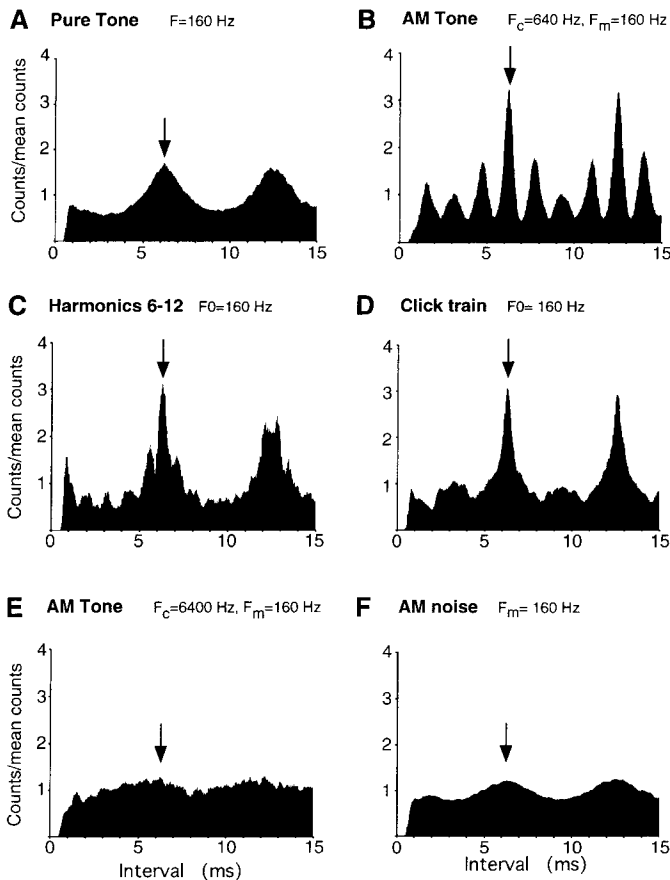


FIG. 11. Pooled autocorrelation histograms for 6 stimuli that evoke a pitch at 160 Hz. Histograms have been normalized to the mean number of counts per bin and smoothed. Arrows: position of the fundamental period $1/F_0$ and/or modulation period $1/F_m$ (6.25 ms). *A*: pure tone, 160 Hz. *B*: AM tone 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

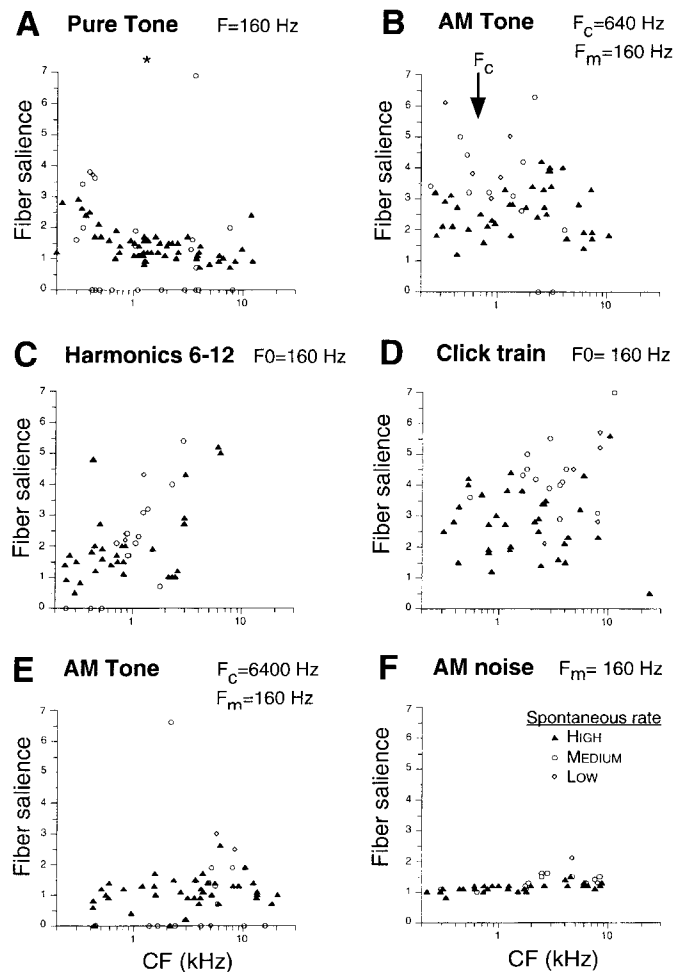


FIG. 12. Fiber saliences for 6 stimuli that evoke a pitch at 160 Hz. Fiber salience is the ratio of intervals per bin in a smoothed single-fiber autocorrelation histogram to mean intervals per bin in a smoothed single-fiber autocorrelation histogram. Asterisk in *A*: medium-spontaneous discharge rate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

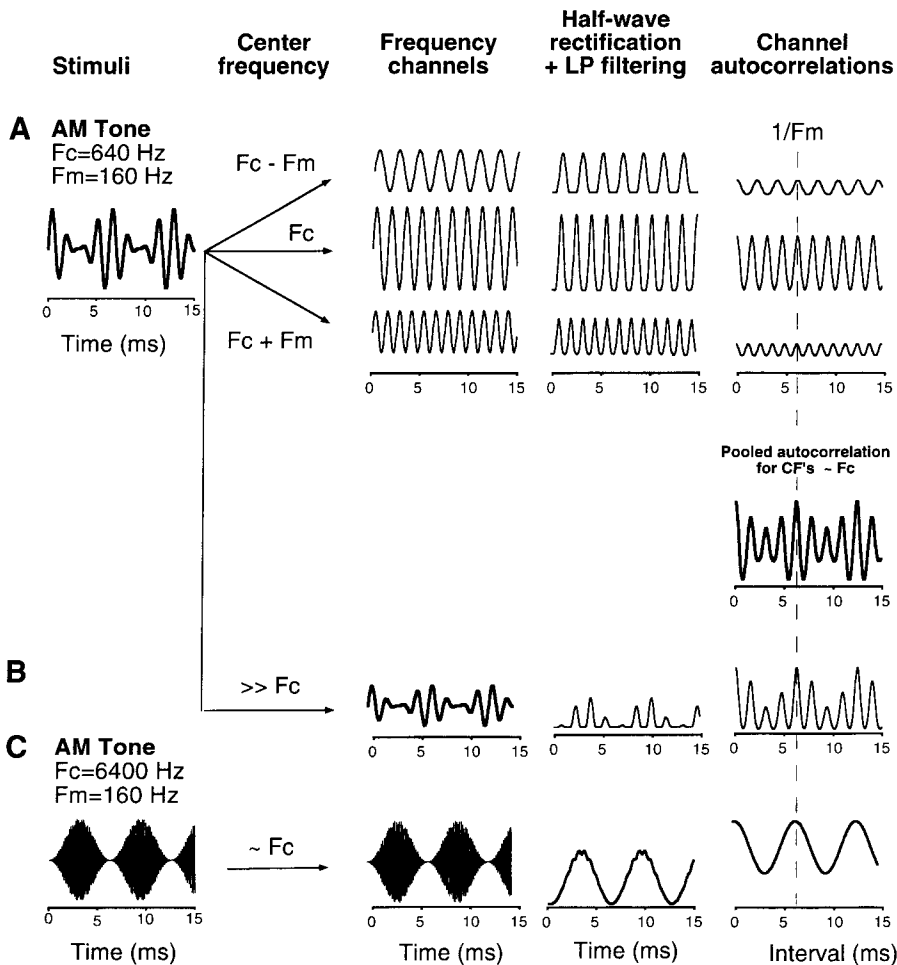


FIG. 13. Conceptual model for the generation of interspike intervals at the fundamental. Model stages (analogous physiological process), *left to right*: stimulus waveform (AM tones), filtered waveform (cochlear filtering), half-wave rectification (transduction) + low-pass (LP) filtering (synaptic transmission), 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; Perrell 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. Köppel, 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. Köppel, 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.