

# A possible neurophysiological basis of the octave enlargement effect

Martin F. McKinney<sup>a)</sup>

Speech and Hearing Sciences Program, Harvard University–Massachusetts Institute of Technology, Division of Health Sciences and Technology and Eaton-Peabody Laboratory, Massachusetts Eye and Ear Infirmary, 243 Charles Street, Boston, Massachusetts 02114

Bertrand Delgutte

Eaton-Peabody Laboratory, Massachusetts Eye and Ear Infirmary, 243 Charles Street, Boston, Massachusetts 02114 and Research Laboratory of Electronics, Massachusetts Institute of Technology and Speech and Hearing Sciences Program, Harvard University–Massachusetts Institute of Technology, Division of Health Sciences and Technology

(Received 24 August 1998; revised 14 July 1999; accepted 23 July 1999)

Although the physical octave is defined as a simple ratio of 2:1, listeners prefer slightly greater octave ratios. Ohgushi [J. Acoust. Soc. Am. **73**, 1694–1700 (1983)] suggested that a temporal model for octave matching would predict this *octave enlargement* effect because, in response to pure tones, auditory-nerve interspike intervals are slightly larger than the stimulus period. In an effort to test Ohgushi's hypothesis, auditory-nerve single-unit responses to pure-tone stimuli were collected from Dial-anesthetized cats. It was found that although interspike interval distributions show clear phase-locking to the stimulus, intervals systematically deviate from integer multiples of the stimulus period. Due to refractory effects, intervals smaller than 5 msec are slightly larger than the stimulus period and deviate most for small intervals. On the other hand, first-order intervals are smaller than the stimulus period for stimulus frequencies less than 500 Hz. It is shown that this deviation is the combined effect of phase-locking and multiple spikes within one stimulus period. A model for octave matching was implemented which compares frequency estimates of two tones based on their interspike interval distributions. The model quantitatively predicts the octave enlargement effect. These results are consistent with the idea that musical pitch is derived from auditory-nerve interspike interval distributions. © 1999 Acoustical Society of America. [S0001-4966(99)05111-5]

PACS numbers: 43.64.Pg, 43.66.Ba, 43.66.Hg [RDF]

## INTRODUCTION

The octave is the basis of most known tonal systems throughout the world (Dowling and Harwood, 1986).<sup>1</sup> Pitches that are an octave apart are deemed equivalent to some degree and can serve the same musical function within certain tonal contexts. The prevalence of the octave as the fundamental building block of tonal systems suggests that there may be a physiological basis for octave equivalence.

A physical octave is defined as a frequency ratio of 2:1. It is known, however, that listeners prefer octave ratios slightly greater than 2:1 (Ward, 1954; Walliser, 1969; Terhardt, 1971; Sundberg and Lindqvist, 1973). In a typical procedure to measure this *octave enlargement* effect, a subject listens to a lower standard tone alternating with an adjustable higher tone and is instructed to adjust the frequency of the higher tone until it sounds one octave above the lower tone. Results of three such experiments are shown in Fig. 1. The size of the preferred or subjective octave is close to 2:1 at low frequencies but increases with frequency and exceeds the physical octave by almost 3% at 2 kHz. It is difficult for listeners to make octave judgments for tones above about 2 kHz. This corresponds to an upper limit in musical pitch of

about 4–5 kHz (Ward, 1954; Attneave and Olson, 1971). There is considerable variability in the octave enlargement effect across listeners but it is nonetheless a statistically significant effect in all the reported studies. The effect is also seen in a wide variety of stimulus conditions and in subjects with various musical backgrounds. It is seen when the two tones are presented simultaneously (Ward, 1954; Demany and Semal, 1990) and under the method of constant stimuli (Dobbins and Cuddy, 1982). The studies shown in Fig. 1 were all performed using pure-tone stimuli but Sundberg and Lindqvist (1973) reported the effect with complex tones as well as pure tones. Ward (1954) reported the presence of the effect in listeners without musical training and in listeners with musical training as well as in possessors of absolute pitch. Dowling and Harwood (1986, p. 103) reported the effect in a number of musical cultures.

The presence of the octave enlargement effect under a wide range of subject and stimulus conditions suggests that the effect may have a general physiological basis. Ohgushi (1983) proposed an octave matching scheme based on a temporal model for pitch that predicts the octave enlargement effect. In an earlier study, he noticed that, in response to pure tones, auditory-nerve interspike intervals are slightly longer than integer multiples of the stimulus period (Ohgushi, 1978). He then showed, using a temporal model for octave

<sup>a)</sup>Author to whom correspondence should be addressed.

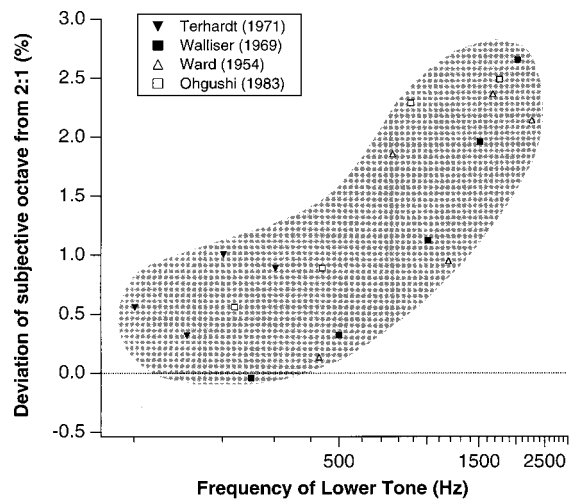


FIG. 1. Psychoacoustic measures of the octave enlargement. Adapted from Fig. 4 in Sundberg and Lindqvist (1973) and Fig. 9 in Ohgushi (1983). The subjective octave, obtained from octave matching experiments, is plotted as a deviation from the physical octave versus the frequency of the lower tone in the octave pair. The subjective octave is larger than the physical octave and the deviation grows with frequency.

matching, that these deviations lead to a prediction of the octave enlargement effect (Ohgushi, 1983).

Upon review of Ohgushi's (1983) model for octave matching, Hartmann (1993) pointed out an arbitrary factor of two. This scaling factor, which is not based on any physiological process, allows a model listener to theoretically set it, and thus the octave interval, to any value. Hartmann suggested a variation of the model that would not rely on such a scaling factor. He also suggested that if the model operated on all-order interspike intervals instead of first-order interspike intervals, it may better predict the psychoacoustic data.

The work presented here was motivated by the hypotheses presented by Ohgushi and Hartmann. Neither one of them could reliably test their predictions because the existing physiological data consisted of only a small number of coarse-resolution interspike-interval distributions. It was therefore difficult to measure the modes of the distributions, i.e., characterize the intervals, with high precision. Special methods were used in this study to ensure high-precision interval analyses so that predictions of temporal models for octave matching could be reliably evaluated. We combined spike data across fibers to form pooled interspike interval histograms which have been shown to reflect a wide variety of pitch phenomena (Cariani and Delgutte, 1996a, 1996b). In addition to characterizing interspike intervals, we have developed and evaluated models for octave matching, based on Ohgushi's and Hartmann's ideas, which operate on pooled interspike interval histograms.

## I. METHOD

The methods used in this study differ from typical auditory-nerve (AN) studies in that specific efforts were taken to ensure accurate estimation of interspike intervals (ISIs): Unusually long recordings were made to ensure the inclusion of a high number of spikes in each record; very fine

binwidths ( $1 \mu\text{sec}$ ) were used when generating ISI histograms in order to accurately estimate the modes.

## A. Experiment

Data were recorded from auditory-nerve fibers in six healthy, adult cats. Cat preparation and recording techniques were standard for our laboratory (Kiang *et al.*, 1965; Cariani and Delgutte, 1996a).

In each experiment, the cat was Dial-anesthetized with an initial dose of 75 mg per kg of body weight and subsequent doses of 7.5 mg per kg of body weight. A craniectomy was performed and the middle-ear and bulla cavities were opened to access the round window. The cerebellum was retracted to expose the AN. Injections of dexamethasone (0.26 mg/kg of body weight/day), to reduce brain swelling, and Ringer's saline (50 ml/day), to prevent dehydration, were given throughout the experiment.

The cat was placed on a vibration isolation table in an electrically shielded, temperature-controlled ( $38^\circ\text{C}$ ) chamber. The AN compound action potential (CAP) in response to click stimuli was monitored with a metal electrode placed near the round window. The cat's hearing was assessed by monitoring the CAP threshold and single-unit thresholds.

Sound was delivered to the cat's ear through a closed acoustic assembly driven by a (Beyerdynamic DT 48A) headphone. The acoustic assembly was calibrated with respect to the voltage delivered to the headphone, allowing for accurate control over the sound-pressure level at the tympanic membrane. Stimuli were generated by a 16-bit, Concurrent (DA04H) digital-to-analog converter using a sampling rate of 100 kHz. The total harmonic distortion for pure tones between 110 and 3000 Hz was less than  $-55 \text{ dB}$  re fundamental when measured at a stimulus level of 95 dB SPL.

AN action potentials (spikes) were recorded with glass micropipette electrodes filled with 2 M KCl. The electrodes were visually placed on the nerve and then mechanically stepped through the nerve using a micropositioner (Kopf 650). The electrode signal was band pass filtered and fed into a spike detector. The times of spike peaks were recorded with  $1\text{-}\mu\text{s}$  precision.

Nerve fibers were sought using a click (near 55 dB SPL) as a search stimulus. Upon contact with a fiber, a threshold tuning curve was generated using the Moxon (Kiang *et al.*, 1970) algorithm with a criterion of 0. The spontaneous rate of the fiber was then measured by counting the number of spikes over a 20-sec period. Units with a characteristic-frequency (CF) threshold more than two standard deviations away from the mean threshold for normal AN fibers (as found by Liberman and Kiang, 1978) were not included in the analysis.

An estimate of the number of false triggers in the spike record was derived from examination of the ISIs. Because the absolute refractory period of AN fibers prohibits ISIs smaller than about 0.5 msec (Gaumond *et al.*, 1982, 1983), intervals shorter than 0.5 msec were assumed to be false triggers. Spike records containing more than 0.1% of these short intervals were not included in the analysis.

The experimental data were recorded using pure-tone

stimuli at frequencies of 110, 220, 440, 880, 1500, 1760, and 3000 Hz and at levels of 5, 10, 15, 20, 25, 40, and 60 dB re threshold. The stimulus was presented once per second (400 msec on, 600 msec off, 2.5 msec rise and fall times) for 180 sec or until 20 000 spikes had been recorded, whichever came first. In order to avoid the possible complex effects of onset transients and adaptation, spikes that occurred during the first 20 msec following the onset of each stimulus and during the stimulus off-time were excluded. Recordings containing fewer than 5000 spikes were not included in the analysis. This unusually high requirement on the minimum number of spikes in the record ensures a reliable estimate of the ISI distribution.

## B. Analysis

Auditory-nerve responses to low-frequency stimuli tend to occur at a specific phase with respect to the stimulus (Rose *et al.*, 1967; Kiang *et al.*, 1965). Thus ISI distributions display modes at intervals corresponding, roughly, to integer multiples of the stimulus period. The main goal of the analysis in this study was to accurately estimate modes of AN ISI distributions in order to quantitatively verify Ohgushi's (1978) observation that the intervals deviate from the stimulus period.

There were three main steps to the analysis of the ISI distributions. First, a histogram of the intervals was produced. Second, the mean interval of each mode in the histogram was estimated by fitting, in the maximum likelihood sense, a Gaussian mixture density to the histogram. Third, deviation of the interval modes from stimulus periods was characterized.

### 1. Histogram generation

The first step in the analysis was to generate histograms of the ISIs. The histogram binwidths were 2  $\mu$ sec for frequencies less than 300 Hz and 1  $\mu$ sec for frequencies above 300 Hz. Both first-order and all-order ISI histograms were computed.

### 2. Mode estimation

The second step in the analysis was to estimate the modes of the interspike interval distribution. A maximum likelihood (ML) estimation approach was implemented in which the interval distributions were modeled as a mixture of Gaussian densities with each mode in the distribution corresponding to a single density. This *mixture density* was fit (in the ML sense) to the interval histograms and the means of the individual Gaussian densities were taken as the estimated modes in the histogram. Two forms of mixture density were used, one for estimating individual modes in the interval distributions and another for estimating the fundamental mode (i.e., stimulus period) in the distributions (and subsequently the stimulus frequency). In the first case, the individual Gaussian densities in the mixture had mutually independent means and variances. In the second case, they were assumed to have harmonically related means and a common variance.

Because obtaining the ML estimates of the parameters is not analytically straightforward, we used the expectation-maximization (EM) algorithm, an iterative technique which

converges to the ML estimate (Redner and Walker, 1984; Moon, 1996). Mathematical details of our implementation are included in the Appendix.

### 3. Mode offset

The third step in the analysis was to calculate the mode offset (MO), the difference between the mode estimate (ME) and the corresponding multiple of the stimulus period,

$$MO_n = ME_n - \frac{n}{f}, \quad (1)$$

where  $f$  is the frequency of the stimulus and  $n$  is the mode number (e.g., mode 1 contains intervals that are roughly one stimulus period in length and mode 2 contains intervals that are roughly 2 stimulus periods in length). Figure 3(e) illustrates the above calculation for  $MO_1$ .

In an effort to represent the total AN population response to the stimuli, pooled histograms were generated by summing all of the individual ISI histograms for a specific stimulus frequency. Mode estimates of the pooled histograms were calculated as well.

## II. RESULTS

From six experiments, a total of 399 spike records from 164 fibers were obtained that met our requirements in terms of the minimum number of spikes, normal thresholds, and small number of false triggers. The majority (79%) of the records were from high spontaneous rate fibers. CFs ranged from 150 to 17 000 Hz.

Figure 2(a) shows a schematic representation of a stimulus wave form and a hypothetical spike record. ISIs are roughly integer multiples of the stimulus period. First-order intervals are those between consecutive spikes, second-order intervals are those between every other spike, etc.

Figure 2(b) shows a histogram of first-order ISIs from a single-unit recording generated with an 880-Hz tone stimulus. The modal distribution of intervals clearly reflects the synchronization of the spike train to the stimulus and the position of the modes provides information about the stimulus frequency (Rose *et al.*, 1967).

Figure 2(c) displays first-, second-, and third-order histograms based on the same spike record as in (b). As one would expect, first-order intervals are, on average, shorter than second- and third-order intervals and thus fall into earlier modes. There is, however, a great deal of overlap in the distributions of the intervals of different orders and the intervals of a particular order are not confined to a single mode in the histogram.

The histogram shown in Fig. 2(d) contains ISIs of all orders, and is thus termed the all-order ISI histogram. This histogram is sometimes referred to as the autocorrelation or autocoincidence histogram (Perkel *et al.*, 1967; Rodieck, 1967; Ruggero, 1973; Evans, 1983).

An important difference between first-order and all-order ISI histograms is their general shape: the size of the modes in the first-order ISI histogram tends to decrease as the mode number increases; the size of the modes in the all-order ISI histogram is relatively constant. In other words,

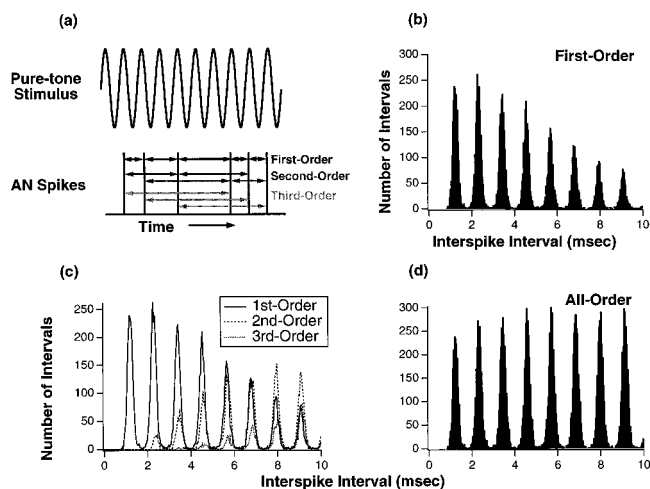


FIG. 2. Histogram generation. (a) is a schematized representation of a pure-tone stimulus and corresponding spike record from the auditory nerve. The order of the interspike interval is based on the number of spikes included in the interval: first-order intervals are those between consecutive spikes; second-order intervals are those between every other spike; third-order intervals are those between every third spike. (b) and (c) are histograms of the various types of interspike intervals. (d) is an interval histogram containing intervals of all orders, thus termed an all-order histogram. All of the histograms were generated from the same spike record. The stimulus was an 880-Hz pure tone at 84 dB SPL. The auditory-nerve fiber from which the recording was made had the following properties: CF: 2609 Hz; SR: 29 spikes/sec. The histograms have a binwidth of 40  $\mu$ sec and the following number of total intervals: first-order: 10 305; second-order: 5696; third-order: 1643; all-order: 17 874.

when one examines very long ISIs, few are first-order intervals. The all-order ISI histogram does not reflect the decaying trend because higher-order intervals are included and “fill in” the modes at long intervals.

In addition to the 399 spike records included in the analysis, 28 spike records that met our data requirements were excluded from the analysis because their histograms displayed peak-splitting. At moderate to high levels of pure-tone low-frequency stimulation, AN ISI histograms can exhibit two or sometimes three peaks per stimulus cycle instead of the normal one (Kiang and Moxon, 1972; Kiang, 1980; Liberman and Kiang, 1984; Kiang, 1990; Ruggero *et al.*, 1996). Most of the fibers from which we recorded did not exhibit this behavior within our stimulus-level range, but those records that did were excluded to simplify the analysis. In our data, peak splitting occurred primarily at stimulus frequencies below 440 Hz.

### A. All-order interspike intervals

Figure 3(a)–(d) are all-order ISI histograms from one AN fiber for four different stimulus frequencies. Figure 3(e) is a magnification of the histogram in (a) with the mode estimates indicated by  $\times$ 's above each mode. As previously reported by Ohgushi (1978, 1983), the short intervals (early modes) are slightly longer than stimulus periods. This deviation is presumed to be at least partially due to the refractory period of the auditory-nerve fiber (Ohgushi, 1978). The mode offset for the first mode is labeled in the figure.

Mode offsets from the histograms in Fig. 3(a)–(d) are plotted in (f) as a function of ISI length. The mode offset decreases monotonically as the ISI increases [Fig. 3(f)] and

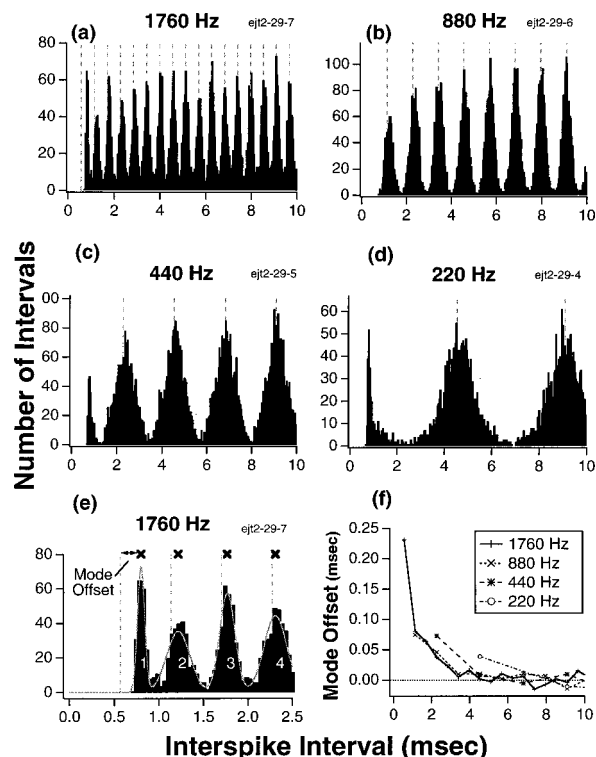


FIG. 3. Histogram mode offset. (a)–(d) are all-order ISI histograms of specified frequency with 40- $\mu$ sec binwidths. Vertical dashed lines mark integer multiples of the stimulus period. (e) is a magnification of the first four modes of (a). The gray curve outlining the histogram is the ML estimate of the Gaussian mixture density corresponding to the histogram. The  $\times$ 's above the modes in (e) mark the ML estimate of the mode (the ML means of the individual Gaussian pdfs in the mixture density), obtained from Eq. (A7) operating on a histogram with 1- $\mu$ sec binwidths. The mode offset is the deviation of the mode estimate from the corresponding integer multiple of the stimulus period. Each histogram was generated from a separate spike record but each spike record was obtained from the same auditory-nerve fiber. Fiber characteristics: CF=2602 Hz; SR=66 spikes/sec. The stimulus levels were all 10 dB re threshold, corresponding to the following levels for each spike record (in dB SPL): (a) 27; (b) 45; (c) 62; (d) 70. (f) displays the mode offsets from the histograms in (a)–(d). Mode offsets are primarily a decreasing function of interval, although, at corresponding intervals, lower frequency stimuli yield slightly larger mode offsets.

for intervals greater than about 5 msec, mode offsets are insignificant. To a first approximation, the mode offsets depend primarily on ISI and not stimulus frequency. However, at any particular ISI  $\leq 5$  msec, lower frequency stimuli generally yield slightly larger mode offsets.

Figure 4(a)–(c) show how mode offsets vary with fiber CF, spontaneous rate (SR), and discharge rate (DR) for all-order histograms of 220 and 1760 Hz. The DR is typically a compressed function of stimulus level ranging from SR to saturation rate. The mode offsets in all-order ISI histograms do not obviously depend on the fiber CF, SR, or DR. Because of this, we decided to pool the ISI data (across fibers and stimulus levels) and use pooled histograms for testing the model presented in the next section. Because pooled histograms contain many more intervals than single-fiber histograms, they more accurately represent the underlying interval probability distributions. Figure 4(d) shows mode offsets grouped by the cat from which they were measured. There is a small, but statistically significant (see caption) variation across cats for the 1760-Hz data. Despite this trend, we de-

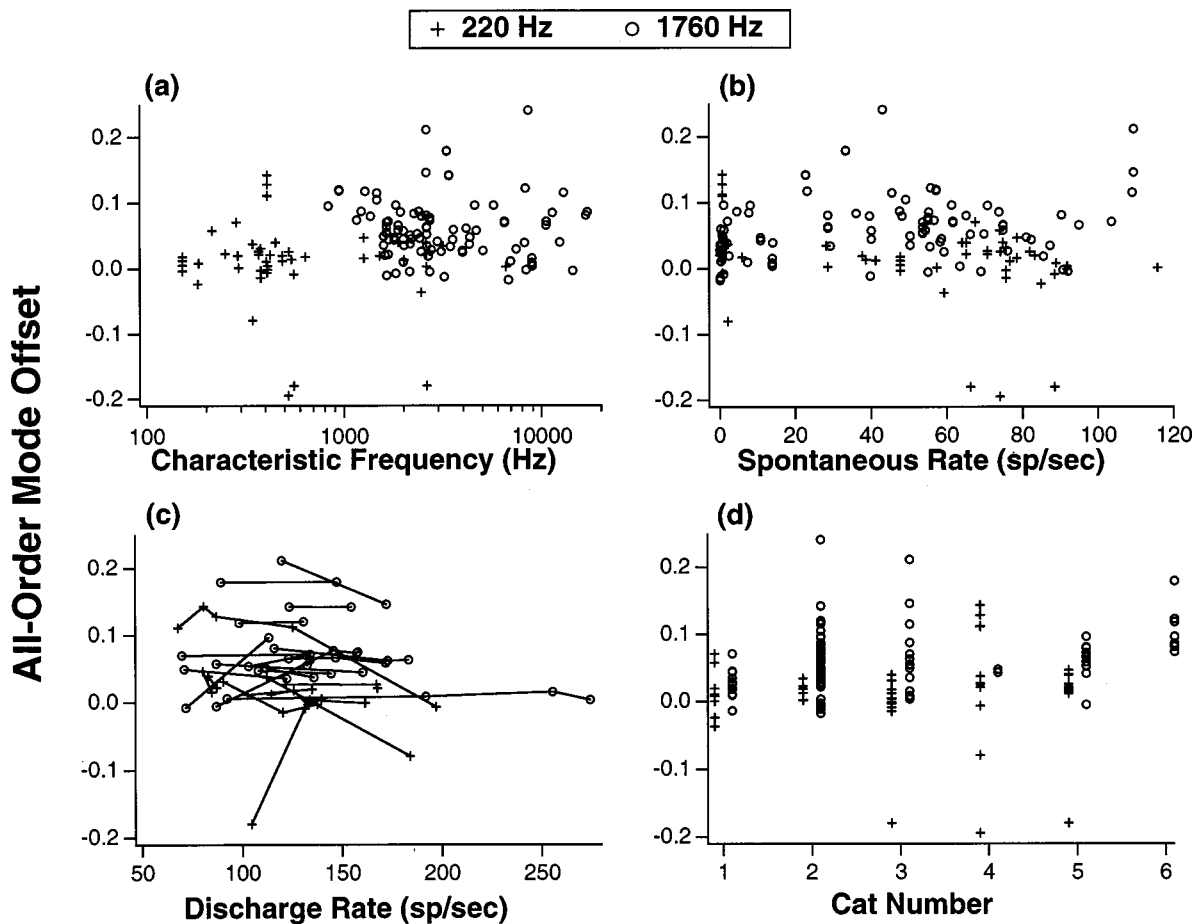


FIG. 4. Variation of mode offset across spontaneous rate, characteristic frequency, discharge rate, and cat. +’s mark the first mode offset of every individual 220-Hz data record and ○’s mark the second mode offset of every individual 1760-Hz data record. These frequencies and mode numbers were chosen as typical representatives of our low- and high-frequency data. In (c), lines connect mode offsets (plotted against DR) that were derived from the same fiber. (a), (b), and (c) show that there is no obvious dependence of mode offset on CF, SR, or DR. (d) shows how mode offset depends on the cat from which it was measured. One-way ANOVA on the data in (d), using the cat number as the individual factor, yielded the following  $p$ -values:  $p=0.003$  for 1760 Hz and  $p=0.139$  for 220 Hz. Although there were significant differences in mode offsets across cats, our decision to pool data across cats did not affect our general conclusions.

cided to also pool data across cats. Conclusions based on the analysis of data from individual cats were not different from those based on pooled data.

Figure 5 shows pooled histograms for six stimulus frequencies. The pooled histograms are much smoother than the single-fiber histograms due to the large number of intervals they contain. Mode offsets are clearly visible at intervals less than about 5 msec. Modes in the 110 and 220 Hz histograms show no offset because even the earliest modes occur at intervals greater than or near 5 msec.

Figure 6(a)–(e) show mode offsets as a function of interval length for pooled histograms as well as for single-fiber histograms. Figure 6(f) shows just the pooled histogram mode offsets for five stimulus frequencies. Although there is some variation across fibers in the size of mode offsets, the characteristics seen in the single-fiber data are evident in the pooled data: the mode offset is a monotonically decreasing function of ISI; mode offsets for intervals greater than about 5 msec are insignificant; and for a given ISI, lower frequency stimuli yield slightly larger mode offsets. Thus these characteristics seem to be general phenomena and not just particular to one type of auditory-nerve fiber or stimulus intensity.

## B. First-order interspike intervals

The general shape of first-order histograms changes with fiber discharge rate while the shape of all-order histograms remains relatively constant (Cariani and Delgutte, 1996a). Figure 7 shows interval histograms from one AN fiber for a 220-Hz tone at three stimulus levels. As the SPL, and therefore discharge rate, increases, the average first-order interval gets shorter and the relative sizes of the histogram modes reflect this change: the later modes get smaller and the early modes get larger. In contrast, as the discharge rate increases, higher-order intervals fill in the modes that get depleted of first-order intervals so that the general shape of all-order histograms remains unchanged.

The main difference between mode offsets of first-order and all-order intervals is the presence of negative mode offsets for low stimulus frequencies in the first-order data. A negative mode offset means that a particular ISI mode is shorter than the corresponding stimulus-period multiple. This is illustrated in the first-order low-frequency histograms in Fig. 8(a) and (b): the modes occur slightly to the left of the stimulus period lines. Mode offset data for low-frequency

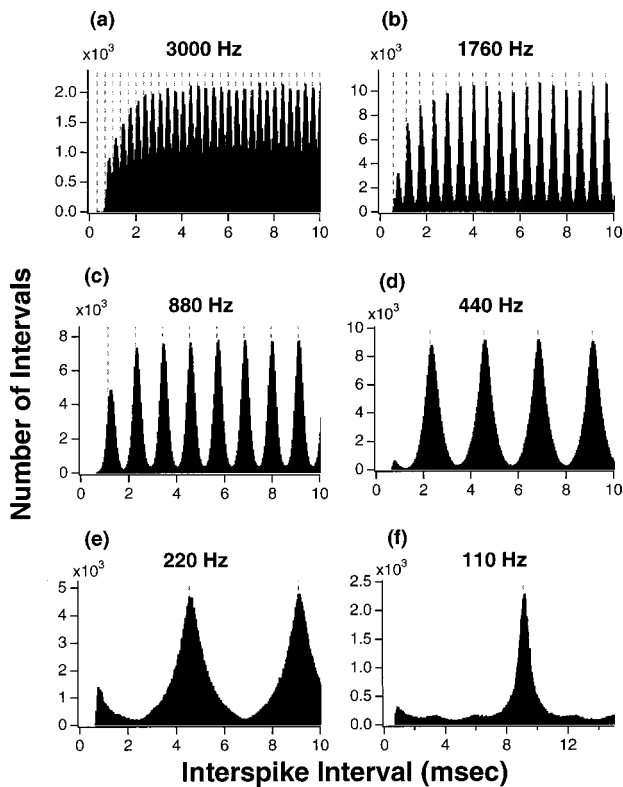


FIG. 5. Pooled histograms. (a)–(f) are pooled all-order ISI histograms of specified frequency. The histograms have the same format as Fig. 3(a)–(d). The intervals are pooled from the following number of fibers: (a) 26; (b) 75; (c) 58; (d) 47; (e) 33; (f) 10. Positive mode offsets are visible for intervals  $\leq 5$  msec (i.e., modes at intervals smaller than 5 msec are shifted slightly to the right of their corresponding stimulus-period multiple). Note that the scale of the abscissa in (f) is different than the other panels.

first-order ISI histograms are shown in panels (c)–(f). These mode offsets show a greater variability at low stimulus frequencies than those from all-order ISIs.

The negative mode offsets in low-frequency first-order ISI histograms are due to the presence of intervals in Mode zero (0). As Fig. 9(a) and (b) illustrate, an interval falls into Mode 0 if two spikes occur within the same half-period of the stimulus. Mode 0 is bounded on the left by the absolute refractory period and on the right by half the stimulus period. Due to the refractory period of the AN fiber, only low-frequency stimuli ( $\leq 500$  Hz) produce ISI histograms that contain a Mode 0. When an interval occurs in Mode 0, the preceding and following first-order intervals tend to be smaller than if just a single spike had occurred in that half-period. The relationship between consecutive intervals can be seen by examining a joint ISI histogram (Rodieck *et al.*, 1962), as shown in Fig. 9(c). The joint ISI histogram is a two-dimensional histogram which plots the ISI size against that of the previous ISI. It is displayed here as a gray scale image in which gray level indicates the number of interval pairs in a small square bin. As is the case for one-dimensional ISI histograms, the modal distribution of intervals is clearly evident in this plot: the intervals tend to cluster near integer multiples of the stimulus period. We will use the notation  $\text{Mode}(n,m)$  to refer to the mode in which the previous interval is in Mode  $n$  and the current interval is in Mode  $m$ . Examination of  $\text{Mode}(0,1)$ ,  $\text{Mode}(0,2)$ , and

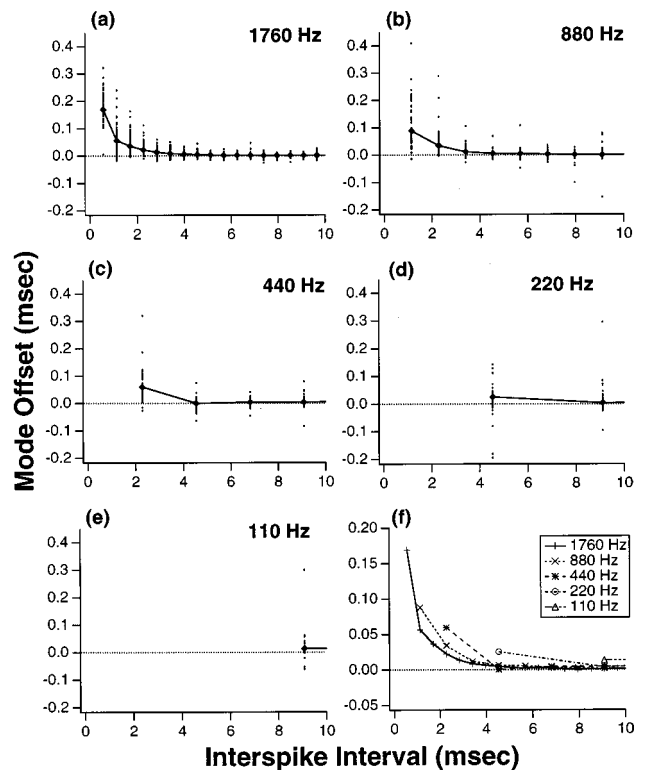


FIG. 6. Mode offsets of all-order ISI histograms. (a)–(e) display the mode offsets of pooled and individual histograms of specified frequency. Lines connect the mode offsets of pooled histograms and dots mark the mode offsets of individual histograms. (f) shows the pooled-histogram mode offsets for most of the experimental stimulus frequencies. Mode offsets in pooled histograms show the same trend with interval as those in individual histograms: mode offset is primarily a monotonically decreasing function of interval, although lower-frequency stimuli yield slightly larger mode offsets at corresponding intervals. Note that the scale of the ordinate in (f) is different than the other panels.

$\text{Mode}(0,3)$  shows that if the previous interval lies in Mode 0, the current interval tends to be shorter than the corresponding stimulus-period multiple. Also, examination of  $\text{Mode}(1,0)$ ,  $\text{Mode}(2,0)$ , and  $\text{Mode}(3,0)$  shows a similar dependency on mode 0 for the current interval. Thus in a first-order ISI histogram, the presence of intervals in Mode 0 effectively biases the other modes toward smaller values.

Offsets of higher modes in all-order ISI histograms are not affected by the presence of intervals in Mode 0 because these histograms include higher-order intervals. For every first-order interval that is shortened by the presence of an interval in Mode 0, there is a second-order interval (which includes the one in Mode 0) that is lengthened. This can be seen schematically in Fig. 9(a). The lengthened second-order interval falls into the same mode as the shortened first-order interval and counteracts its effect on the mode offset.

The effect of Mode 0 on first-order ISI histograms can be quantified by selecting only those intervals that do not precede or follow intervals in Mode 0. This conditioning was performed on all of the 220-Hz stimulated spike records and then histograms of the conditioned intervals were generated. The distribution of the mode estimates for Mode 1 in these conditioned interval distributions is plotted in Fig. 9(d) along with similar unconditioned distributions from first-order and all-order ISI histograms. The alignment of the mode-estimate

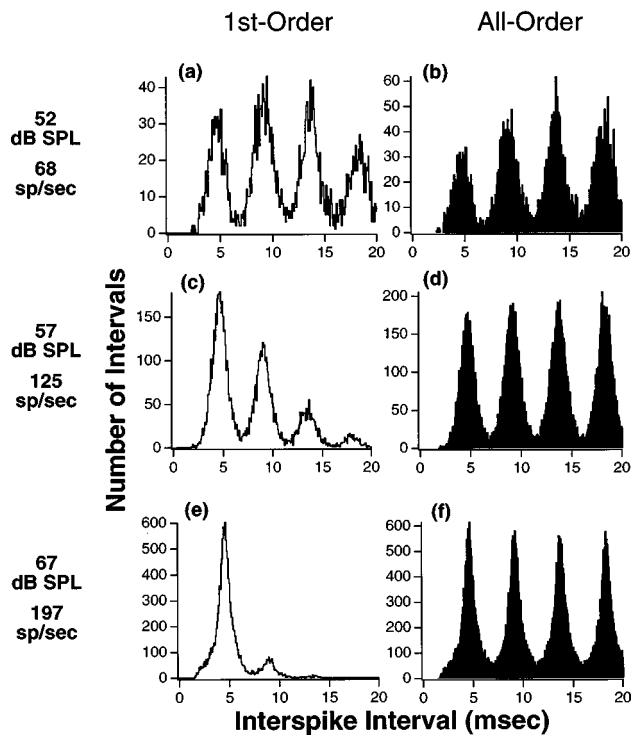


FIG. 7. A series of 220-Hz ISI histograms (from the same auditory-nerve fiber) over a range of discharge rates. Stimulus level and fiber discharge rate are indicated to the left of the plots. First-order ISI histograms are plotted in (a), (c), and (e). All-order ISI histograms are plotted in (b), (d), and (f). Fiber characteristics: CF: 409 Hz; SR: 0.7 spikes/sec; threshold at 220 Hz: 47 dB SPL. The histogram binwidths are 80  $\mu$ sec.

distributions for all-order and conditioned first-order ISIs indicates that the presence of intervals in Mode 0 accounts for nearly all of the difference between mode estimates in all-order and first-order ISI histograms.

The negative correlation between consecutive intervals is a characteristic of our data that is not well documented in the literature. The joint ISI histogram in Fig. 9(c) shows a clear dependence between the previous and current first-order ISI. All of the modes are oval with the long axis going diagonally from the top left to the bottom right of the figure. This means that if the previous interval was shorter than average, the current interval will tend to be longer than average and vice versa. This is a consequence of phase-locking: every interval longer than the stimulus period must be compensated for by a shorter interval if the spikes are to remain phase-locked.

### III. MODEL

Our primary objective in formulating a (central) model for octave matching is to evaluate how physiological constraints in the auditory periphery, i.e., deviations in AN ISIs, affect the central processor. This is best accomplished with simple models that have few, if any, free parameters, so that the effect of the peripheral physiological behavior is not clouded. With this in mind, we developed a temporal model for pure-tone octave matching based on Ohgushi's (1983) model.

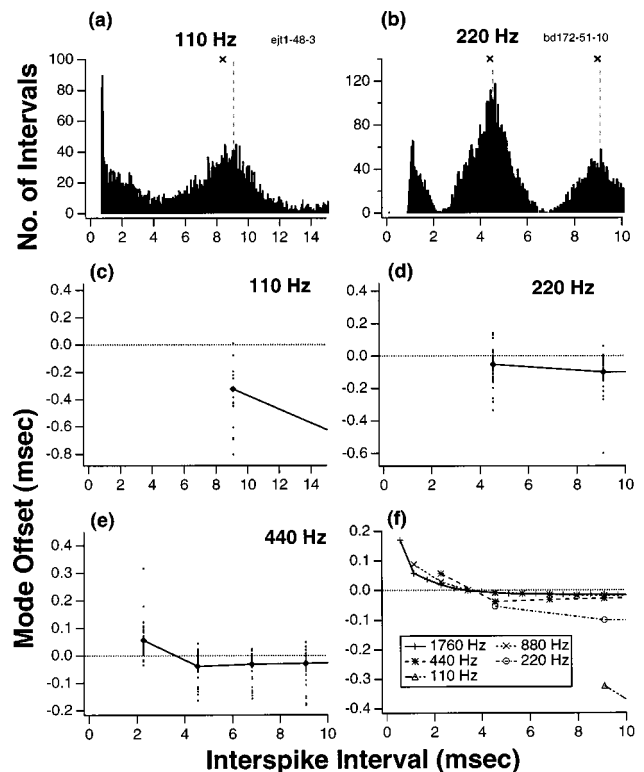


FIG. 8. Low-frequency first-order ISI histograms display negative mode offsets. (a) and (b) show first-order ISI histograms in the same format as those in Fig. 3. The  $\times$ 's above the modes mark the ML estimate of the mode. (c), (d), and (e) show mode offsets from first-order ISI histograms in the same format as Fig. 6. (f) shows the pooled-histogram mode offsets for most of the stimulus frequencies. Below  $\sim$ 500 Hz, first-order ISI histograms display large negative mode offsets (i.e., modes are shifted slightly to the left of their corresponding stimulus-period multiple), in contrast to the insignificant mode offsets present in low-frequency all-order ISI histograms. Note that the scales of the axes vary across panels.

### A. Model for estimating pure-tone frequency

The basic assumption of the model is that perceived pitch is equal to a biased estimate of the stimulus frequency derived from AN ISIs. The bias in the frequency estimate comes from the mode offsets in the ISI histograms. Frequency estimates were derived from interval histograms using the EM algorithm [Eqs. (A1) and (A2)], assuming a mixture density of Gaussians with harmonically related means [Eq. (A9)],

$$\hat{f} = \frac{1}{\hat{\mu}_{\text{ML}}(N_{\text{max}})}, \quad (2)$$

where  $\hat{f}$  is the estimate of stimulus frequency  $f$ ,  $\hat{\mu}_{\text{ML}}$  is the ML estimate of the fundamental mean in the mixture density [ $\mu^+$  in Eq. (A10)], and  $N_{\text{max}}$  is the number of modes included in the calculation [ $M$  in Eqs. (A10) and (A11)]. If the modes occur exactly at integer multiples of the stimulus period,  $\hat{\mu}_{\text{ML}}$  will equal the stimulus period and the frequency estimate will be equal to the stimulus frequency.

Estimates for each stimulus frequency were calculated using pooled ISI histograms and their deviations from the stimulus frequency were derived as follows:

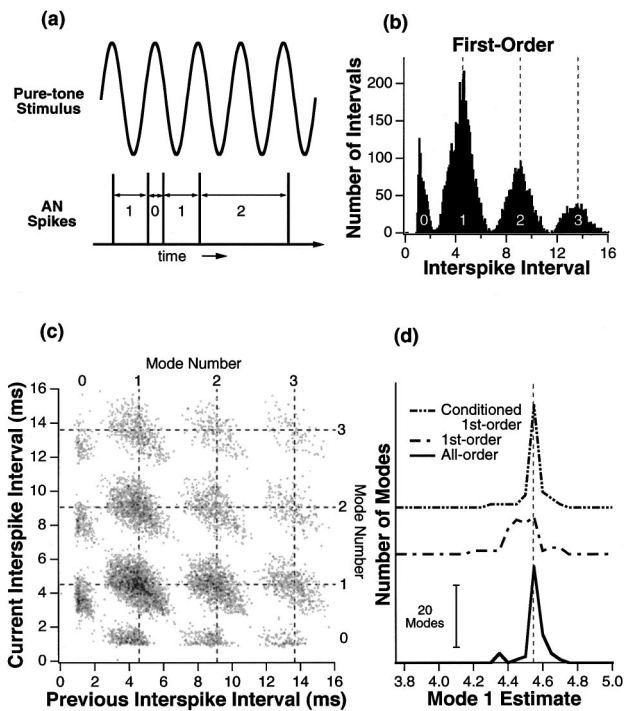


FIG. 9. Intervals in Mode 0. (a) is a schematic representation of a stimulus and a corresponding spike record. The number between the spikes in (a) indicates the mode of the histogram, shown in (b), to which the first-order interval belongs. (b) is a first-order ISI histogram and (c) is a joint first-order ISI histogram of the same auditory-nerve spike record. The joint ISI histogram is a two-dimensional histogram which plots ISI size against that of the previous ISI. It is displayed as a gray-scale image in which gray level indicates the number of interval pairs in a small square bin. The dashed lines in (b) and (c) mark integer multiples of the stimulus period. The stimulus was a 220-Hz pure tone at 61 dB SPL. The fiber characteristics are: CF: 379 Hz; SR: 76 spikes/sec; threshold at 220 Hz: 36 dB SPL. The histogram binwidth is 80  $\mu$ sec in (b) and 64  $\mu$ sec for both dimensions in (c). (d) shows the distribution of the estimates of Mode 1 in 50 individual 220 Hz histograms for all-order, first-order, and conditioned first-order ISIs. The condition in the third case is that the intervals do not follow or precede an interval in Mode 0. The traces are vertically offset for clarity and the vertical bar in the lower left denotes 20 modes. The binwidth of the mode estimate distribution is 50  $\mu$ sec. The vertical dashed line marks the stimulus period. Negative mode offsets in low-frequency first-order ISI histograms are due to shortened intervals caused by intervals in Mode 0 (i.e., two spikes within the same half-period of the stimulus).

$$\hat{f}_{\text{DEV}} = 100 \cdot \frac{\hat{f} - f}{f}, \quad (3)$$

where  $\hat{f}_{\text{DEV}}$  is the percent deviation of the frequency estimate and  $\hat{f}$  is the frequency estimate.

$\hat{f}_{\text{DEV}}$  is plotted versus stimulus frequency in Fig. 10 for three values of  $N_{\text{max}}$ . For both all-order intervals and first-order intervals,  $\hat{f}_{\text{DEV}}$  is a decreasing function of stimulus frequency.<sup>2</sup> This trend is a direct result of the dependence of mode offset on interval size. As the stimulus frequency increases, the stimulus period decreases and the offset for any given mode number increases. This results in a larger estimate of the fundamental period,  $\hat{\mu}_{\text{ML}}$ , and hence, a decrease in the frequency estimate. For all-order intervals,  $\hat{f}_{\text{DEV}}$  is always negative because mode offsets are always positive. On the other hand, first-order ISI intervals yield positive

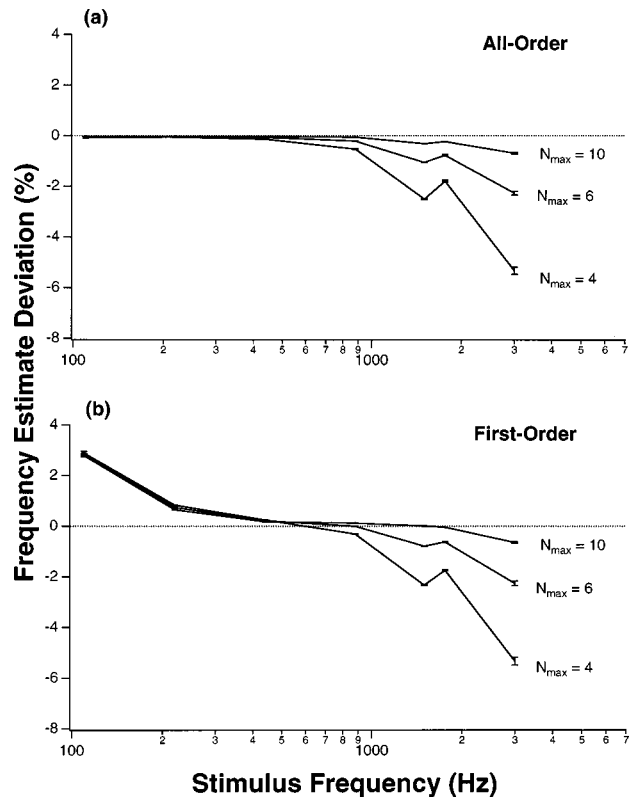


FIG. 10. Frequency estimate deviation ( $\hat{f}_{\text{DEV}}$ ) vs frequency. (a) displays  $\hat{f}_{\text{DEV}}$  calculated from pooled all-order ISI histograms for the values of  $N_{\text{max}}$  shown next to each trace. (b) displays  $\hat{f}_{\text{DEV}}$  calculated from pooled first-order ISI histograms. For frequencies  $\geq 500$  Hz,  $\hat{f}_{\text{DEV}}$  is a decreasing function of both frequency and  $N_{\text{max}}$ . Error bars show an estimate of the standard error of  $\hat{f}_{\text{DEV}}$ . The estimate was calculated using the bootstrap technique (Efron and Tibshirani, 1993): 50 simulations of the frequency estimate were calculated [Eq. (2)] in which pooled histograms were generated by randomly choosing (with replacement) spike records of individual stimulus presentations. The standard deviation of the frequency estimates from these simulations is an estimate of the standard error of the mean.

$\hat{f}_{\text{DEV}}$ 's for low stimulus frequencies because the histograms contain negative mode offsets.

Figure 10 also shows that the free parameter  $N_{\text{max}}$  greatly influences the frequency estimate at high frequencies. For low values of  $N_{\text{max}}$ , the frequency estimate has a relatively large bias from the mode offsets of the lower modes. Since the mode offset is minimal in the higher modes, the frequency estimate becomes less biased as  $N_{\text{max}}$  increases. On the other hand,  $N_{\text{max}}$  has little effect on the estimates at low frequencies because either the mode offsets are consistently small for all modes (all-order ISIs), or the higher modes contain few intervals and thus little weight in the calculation of  $\hat{f}$  (first-order ISIs).

## B. Model for octave matching

The model operates on two sets of pooled ISI histograms to predict the size of the pitch interval separating their respective stimuli. The pitch interval prediction is obtained by comparing the frequency estimate [Eq. (2)] of a low-frequency tone,  $f_1$ , with the frequency estimate of a high-frequency tone,  $f_2$ . The model predicts that  $f_1$  and  $f_2$  are separated by a subjective octave when,

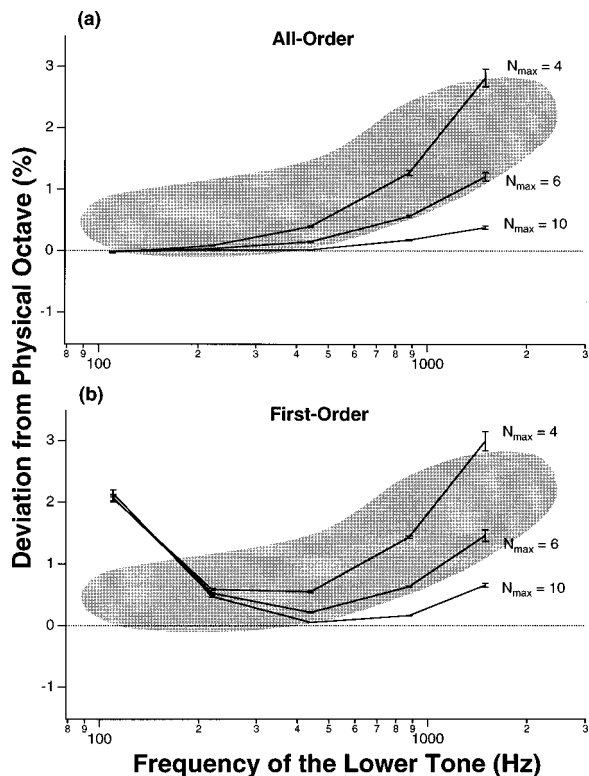


FIG. 11. Model predictions of the octave enlargement effect. The model predictions are based on pooled histograms for each stimulus frequency. Error bars show the estimated standard error of the subjective octave prediction and were calculated in a similar manner to those in Fig. 10. (a) shows the model predictions for all-order ISIs and several values of  $N_{max}$ . (b) shows the same for first-order ISIs. Although low-frequency data are not well predicted by the model, the predictions based on all-order intervals are within the range of the psychoacoustic data with  $N_{max} \approx 4-6$ .

$$\hat{f}_2 = 2 \cdot \hat{f}_1. \quad (4)$$

The model algorithm can be interpreted graphically as attempting to align the modes of the scaled (by two)  $f_1$  histogram with the modes of the  $f_2$  histogram. An octave is predicted when the modes are best aligned.

The deviation of the model prediction (i.e., “subjective octave”) from the physical octave,  $\Delta_{SO}$ , is:

$$\Delta_{SO} = 100 \cdot \frac{2 \cdot \hat{f}_1 - \hat{f}_2}{f_1}, \quad (5)$$

for  $f_1$  and  $f_2$  separated by a physical octave.

Model predictions are shown in Fig. 11 for several values of  $N_{max}$ . As in the frequency estimate (Fig. 10), variation in  $N_{max}$  results in large changes in model predictions at high frequencies. As  $N_{max}$  increases, more modes with little or no offset are included in the frequency estimates and the resulting deviation of the subjective octave decreases.

When all-order ISIs are used as model input, the model predicts an octave enlargement in general agreement with the psychoacoustic data [Fig. 11(a)] at most frequencies for  $N_{max} \approx 4-6$ . At low frequencies, the model underestimates the psychoacoustic octave enlargement for all values of  $N_{max}$ , but its predictions are still within the range of the

psychoacoustic data. At 1500 Hz, the model predicts the range of psychoacoustic data simply by varying  $N_{max}$  from 4 to 6.

When operating on first-order ISIs, the model, with  $N_{max} = 4$ , predicts an octave enlargement in general agreement with the psychoacoustic data, except at 100 Hz, where the model predicts a much larger deviation [Fig. 11(b)]. In addition, the model predicts a decrease in deviation as frequency increases (at low frequencies) but the psychoacoustic data show the opposite trend. The model’s predicted octave enlargement at low frequencies is due to the negative mode offsets in the first-order ISI histograms. The frequency estimates of these low-frequency tones are higher than the true frequency [see Fig. 10(b)] and when they are matched to estimates of (upper) tone frequencies that produce little or no negative mode offsets, an octave enlargement is predicted.

In summary, the model, operating on first- or all-order ISIs with  $N_{max} \approx 4-6$ , predicts the octave enlargement effect at mid- to high frequencies. At low frequencies, the model underestimates the effect when operating on all-order ISIs and overestimates it when operating on first-order ISIs.

## IV. DISCUSSION

### A. Auditory nerve physiology

We have shown that, in response to low-frequency pure tones, AN ISIs deviate systematically from integer multiples of the stimulus period. When quantitatively expressed as mode offsets in ISI histograms [Eq. (1)], the deviations are positive for ISIs less than 5 msec and decrease with increasing ISI until they become insignificant for ISIs greater than 5 msec. In addition, first-order intervals show negative mode offsets for stimulus frequencies less than 500 Hz. These robust phenomena exist for all CFs and SRs and over a wide range of stimulus levels. Our quantitative characterization of these physiological properties provides a solid basis to study how they can effect any temporally based estimate of the stimulus frequency.

Our data and analyses suggest that positive and negative mode offsets in ISI histograms arise from fundamentally different mechanisms. We showed in Fig. 9 that negative mode offsets, seen in first-order ISI distributions for low-frequency stimuli, are due to the occurrence of multiple spikes within the same half-period. In order to maintain phase-locking between the stimulus and AN response, the intervals before and after these multiple spikes tend to be slightly shorter, on average, than multiples of the stimulus period. Positive mode offsets, on the other hand, have been attributed to the refractory properties of the neurons (Ohgushi, 1983, 1978) and, specifically, to a reduction in conduction velocity during the relative refractory period (de Cheveigné, 1985). While these ideas are reasonable, it is important to note that the delays causing the offsets could arise at any point from the basilar membrane to the AN fiber.

A physiological characteristic that we saw in our data but ignored in the analysis is peak splitting. This phenomenon causes two or more modes of intervals to be present within a single stimulus period of an ISI histogram instead of the usual one mode per stimulus period. The multiple modes

are the result of the AN response going through a change in phase (as much as  $180^\circ$ ) relative to the stimulus as the stimulus level is increased (Kiang and Moxon, 1972; Johnson, 1980; Kiang, 1980, 1990).

At first sight, peak splitting would seem to wreak havoc on temporal models for pitch. At stimulus intensities where peak splitting occurs, a model operating on the intervals would estimate multiple frequencies, depending on the degree of phase shift. However, because the stimulus intensity at which peak splitting occurs depends on both fiber CF and stimulus frequency, only a small fraction of AN fibers will exhibit peak splitting at the same stimulus intensity. So, in a temporal model for pitch that operates on intervals pooled from fibers across many CFs, peak splitting most likely has a small effect on the pooled interval distribution, leaving the overall frequency estimate relatively unchanged.

## B. Temporal models for octave matching and pitch perception

Our model for octave matching makes pitch-interval judgments based on frequency estimates of two tones. Each frequency estimate is computed from a pooled AN ISI histogram by fitting it with a Gaussian mixture density with harmonically related means. An octave is predicted when the frequency estimate of one tone is twice that of another tone. The model predicts the octave enlargement effect except at very low frequencies, where it slightly underestimates the effect when operating on all-order ISIs and overestimates the effect when operating on first-order ISIs.

### 1. Comparison with Ohgushi's model

Our model is similar to Ohgushi's (1983) model for octave matching. The basic elements of the models are the same although there are three primary differences in his implementation: he uses first-order ISIs only; his frequency estimates were based on just the first two modes of the histogram while we used a variable number ( $N_{\max}$ ) of modes; and he calculates frequency estimates from the modes with weights obtained by fitting the model predictions to the psychoacoustic data and adjusting two free variables. These differences lead to different predictions at low frequencies when operating on first-order ISIs. Ohgushi's model predictions are consistent with the psychoacoustic data on the octave enlargement for all frequencies while our model has difficulties at very low frequencies ( $<200$  Hz). It should be pointed out that with two free parameters, Ohgushi had more flexibility with which to fit the data.

In addition, Ohgushi operated on rather coarse (100- $\mu$ sec binwidth) single-fiber ISI histograms from only four AN fibers, published by Rose *et al.* (1967, 1968), while our model predictions were based on fine-resolution pooled histograms which represent a large number of fibers and spikes. Analysis of our data using Ohgushi's method yields results similar to his.

### 2. Interpretation of $N_{\max}$

The one free parameter in our model is  $N_{\max}$ , the number of modes over which the frequency estimate is calculated.  $N_{\max}$  can greatly affect the frequency estimate and re-

sulting octave interval prediction. Rather than treating it as an arbitrary free parameter, it would be nice to give  $N_{\max}$  a physiological or psychoacoustic interpretation. If one assumes that pure-tone pitch is based on the interspike interval distribution of AN spikes,  $N_{\max}$  could be related to the minimum tone duration required to elicit a pitch.

A number of psychoacoustic studies have investigated the effect of tone duration (for very short tones) on pitch (Doughty and Garner, 1947, 1948; Pollack, 1967) and the ability to recognize musical melodies (Patterson *et al.*, 1983). A general result from these studies is that, for tones below about 1000 Hz, a minimum number of cycles ( $6 \pm 3$ ) is required to elicit a stable pitch or to achieve maximum performance in melody recognition. On the other hand, above 1000 Hz, a minimum tone duration ( $\sim 10$  msec) is required to elicit a stable pitch (Gulick *et al.*, 1989). If  $N_{\max}$  is taken as the number of cycles required to elicit a pitch for low frequencies, our empirically derived range for  $N_{\max}$  ( $\sim 4-6$ ) is consistent with this result. It should be noted that our analyses do not include the first 20 msec of the AN response. Verification of such a relationship between  $N_{\max}$  and minimum duration for pitch would require a study which carefully addresses the effects of adaptation and ringing of the cochlear filter for short-duration tones. Nevertheless, our results suggest that there may be a link between the two "integration times."

Another consideration related to  $N_{\max}$  is that the overall neural delay required to perform octave matches for low-frequency tones may be physiologically implausible. For example, with  $N_{\max}=5$ , the total delay required to obtain a frequency estimate for a 60-Hz tone is 83 msec. There is however, evidence for the existence of a lower limit to musical pitch around 90 Hz (Biasutti, 1997), which reduces the maximum required neural delay in our model to about 55 msec.

### 3. An alternative model for octave matching

We developed and implemented a second model for octave matching following a suggestion by Hartmann (1993). Noting that the scaling factor of two in Ohgushi's (1983) model for octave matching is arbitrary, Hartmann suggested that a more physiologically grounded model is one that attempts to correlate the ISIs without first scaling those from the low-frequency tone. The comparison is then made between two tones using only the intervals from the even modes in the ISI histogram for the high-frequency tone. This model can be interpreted graphically as attempting to align the modes of the  $f_1$  histogram with the even modes of the  $f_2$  histogram.

Despite the appeal of Hartmann's suggestion, we found that this model fails to predict the octave enlargement phenomenon and instead predicts a slight octave contraction. The cause of this prediction can be seen by examining the mode offsets at the same interval size in Fig. 6(f). In an octave comparison between two tones separated by a physical octave, the second mode in the ISI histogram for the high-frequency tone has a smaller mode offset than the first mode of the lower frequency tone. This causes the sub-octave estimate of the higher tone to be slightly higher than

the frequency estimate of the lower tone. In order to achieve a subjective octave match, the higher frequency tone needs to be slightly lower in frequency than the physical octave above the lower tone. This results in a predicted octave contraction rather than an octave enlargement.

#### 4. Temporal models for pitch

Our model for octave matching is similar to existing models for frequency discrimination (Siebert, 1970; Goldstein and Srulovicz, 1977) in that they are based on the idea that pitch is a frequency estimate of a pure-tone stimulus based on temporal discharge patterns. Both Siebert (1970) and Goldstein and Srulovicz (1977) represent AN activity with nonhomogeneous Poisson processes. Siebert's main objective was to investigate the limitations in frequency discrimination of an optimal processor operating on spike times of modeled AN activity. He discovered that there is enough temporal information in the all-order intervals from a small number of auditory-nerve fibers to account for the psychoacoustic data on frequency discrimination. However, the slope of the predicted frequency discrimination limen versus stimulus duration far exceeded psychoacoustic performance. Goldstein and Srulovicz showed that a similar model operating on only first-order ISIs better predicts the dependence of the psychoacoustic frequency difference limen on stimulus duration.

The essential difference between these models and ours is that the optimal processor models give unbiased (ML) estimates of the stimulus frequency. The octave matching model relies on biased frequency estimates which result from the assumption that modes of the ISI distribution are harmonically related to the stimulus period. These biases were lacking in the Siebert and Goldstein models because refractor effects were not included in the Poisson processes.

An important distinction within the class of temporal models for pitch is between those that operate on first-order ISIs and those that operate on all-order ISIs. All-order intervals can be obtained from a spike train using delay lines and coincidence detectors as proposed by Licklider (1951). Analysis of first-order intervals, on the other hand, requires an extra stage of processing to eliminate the higher-order intervals. This makes a model based on first-order ISIs less appealing, physiologically, than one that operates on all-order intervals. A further advantage for a model operating on all-order intervals may be the fact that all-order interval distributions tend to be more stable across stimulus level than first-order interval distributions, as shown in Fig. 7.

We have seen in this study, as have Goldstein and Srulovicz (1977), that model predictions based on one or the other type of ISI can yield different results. Goldstein and Srulovicz show that in the context of frequency discrimination, operating on first-order ISIs results in a better fit to the psychoacoustic data than operating on all-order ISIs. Also, psychophysical experiments attempting to distinguish between the two kinds of ISI-based pitch models have favored first-order ISIs (Kaernbach and Demany, 1996). Kaernbach and Demany used random click train stimuli with specified first- and higher-order interclick distributions and found that discrimination between those stimuli and randomly distrib-

uted clicks was better for regular first-order interclick intervals. Results of our study do not strongly favor either first- or all-order intervals. Model predictions based on first-order ISIs overestimate the subjective octave at low frequencies and those based on all-order ISIs slightly underestimate the subjective octave. The trend with frequency, however, of those predictions based on all-order ISIs is more consistent with the psychoacoustic data. Nevertheless, we can not rule out models based on intermediate combinations of the two types of intervals or other more complex models, such as Ohgushi's, which predict the octave enlargement based on first-order ISIs. Also, it is conceivable that different physiological cues may be responsible for discriminating frequency than for matching octaves or for performing other tasks involving musical pitch.

Our model for octave matching is also analogous to the optimum processor introduced by Goldstein (1973). He uses a template of Gaussian density functions spaced harmonically along the spectral axis to fit, in the ML sense, the excitation pattern produced by a complex tone. Although his implementation operates on spectral excitation, there is nothing inherent to the model that precludes its operation on interval distributions. Our model is similar to his in that it fits harmonic templates to noisy and possibly inharmonic data. In Goldstein's case, the inharmonicity only arises if the stimulus contains inharmonically related partials. In our case, the inharmonicity is always present and comes from mode offsets in ISI distributions.

Although we have concentrated solely on temporal models, we should not forget that there exist alternative schemes for octave matching, namely rate/place models. Terhardt's (1971, 1974) model for virtual pitch theoretically predicts the octave enlargement effect and is discussed in that light by Hartmann (1993). Terhardt suggests that through pervasive listening to natural tone complexes we develop memory templates of tonotopic excitation patterns and that we make octave judgments based on the places of maximum excitation in these memory templates. He further postulates that these templates are stretched, i.e., the places of maximum excitation corresponding to the harmonics in the tone complex are shifted (upwards in frequency) due to masking effects caused by the presence of the lower harmonics. Thus the subjective octave, based on these stretched templates, is slightly larger than the physical octave. There is some evidence that lower-frequency masking stimuli can lower the CF of an AN fiber (Kiang and Moxon, 1974; Delgutte, 1990) but the effect of masking depends on the overall stimulus level and on the relative levels of the signal and masker. It is not known whether these effects are quantitatively adequate to validate Terhardt's theory.

#### V. CONCLUSION

We have shown that, in response to low-frequency pure tones, AN ISIs less than 5 msec are systematically larger than integer multiples of the stimulus period and, for frequencies less than 500 Hz, first-order ISIs are smaller than integer multiples of the stimulus period. These deviations result in biased estimates of frequency and can lead directly to a prediction of the octave enlargement effect by temporal-

based models. Thus computational models for pitch may have to incorporate detailed physiological properties of the auditory periphery, such as refractoriness, in order to predict effects such as octave enlargement.

Correlating psychoacoustic behavior in the context of pitch effects with physiological responses to the same set of stimulus conditions can lead to valuable insights into the neurophysiological basis of pitch. Here, we have examined models for octave matching operating on two forms of ISIs and, although no model is completely satisfactory, one of them, operating on all-order intervals, comes close to predicting the octave enlargement effect over its entire frequency range. This result is consistent with the notion that musical pitch is based on a temporal code.

## ACKNOWLEDGMENTS

The authors would like to thank Drs. P. A. Cariani, J. J. Guinan, M. C. Liberman, and three anonymous reviewers for comments on previous revisions of this manuscript. Dr. L. D. Braida suggested the use of ML estimation for analyzing interval histograms and provided helpful guidance on the EM algorithm. This work was supported by Grant Nos. R01 DC02258 and T32 DC00038 from the NIDCD, NIH.

## APPENDIX: THE EM ALGORITHM

In order to find the ML estimates of parameters in the Gaussian mixture densities described in Eqs. (A3), (A4), and (A9), we used the iterative EM algorithm (Redner and Walker, 1984; Moon, 1996). This Appendix briefly describes the EM algorithm and shows the mathematical details of our implementation.

The general idea of the EM algorithm is as follows: Ideally, one would like to obtain ML estimates for parameters,  $\Phi$ , of a pdf,  $f(\mathbf{y}|\Phi)$ , over the complete sample space,  $\mathbf{Y}$ . At hand, however, is an incomplete data sample,  $\mathbf{x}$ , which is insufficient to compute and maximize the log-likelihood function over  $\mathbf{Y}$ . In our case, the vector  $\mathbf{x} = \{x_k : k = 1, N\}$  is the interspike interval distribution where  $x_k$  is a single interval and  $N$  is the number of intervals. The data sample is incomplete because the component density in the mixture from which a particular interval arises is not known. A complete data sample,  $y_k = (x_k, i_k)$ , would consist of the interspike interval,  $x_k$ , and an indicator,  $i_k$ , of the component density from which  $x_k$  originated. So, instead of maximizing the log-likelihood over  $\mathbf{Y}$ , the EM algorithm maximizes the expectation of  $\log(f(\mathbf{y}))$  given the data,  $\mathbf{x}$ , and the current parameter estimates,  $\Phi'$ . The two-step EM algorithm is,

E-step: Determine:  $Q(\Phi|\Phi') = E(\log(f(\mathbf{y}|\Phi))|\mathbf{x}, \Phi')$ . (A1)

M-step: Choose:  $\Phi^+ \in \arg \max_{\Phi} Q(\Phi|\Phi')$ . (A2)

With each iteration, the next parameter estimates,  $\Phi^+$ , replace the current parameter estimates,  $\Phi'$ , until convergence or until the difference between sequential sets of parameters is less than some designated  $\epsilon$ . Our implementation of the EM algorithm follows directly from equations developed in Redner and Walker (1984), so we refer the reader to their

paper for details on the preliminary derivations and focus here on details pertinent to our implementation.

## 1. Gaussians with independent means and variances

To characterize the individual modes of ISI histograms, we modeled each interval distribution as a mixture of  $M$ -weighted, univariate Gaussian probability density functions (PDF) with independent means and variances,

$$p(x|\Phi) = \sum_{i=1}^M \alpha_i p_i(x|\phi_i), \quad (\text{A3})$$

where  $x$  is a single interval in the distribution,  $\Phi = (\alpha_1, \dots, \alpha_M, \phi_1, \dots, \phi_M)$ ,  $\alpha_i$  is a non-negative weighting,  $\sum_{i=1}^M \alpha_i = 1$ , and  $p_i$  is a univariate Gaussian pdf with parameters  $\phi_i = (\mu_i, \sigma_i)$ ,

$$p_i(x|\phi_i) = \frac{1}{\sqrt{2\pi}\sigma_i} e^{-(x-\mu_i)^2/2\sigma_i^2}. \quad (\text{A4})$$

For a mixture of Gaussian densities in the form of Eqs. (A3) and (A4), Redner and Walker (1984) derive  $Q(\Phi|\Phi')$  in their Eq. (4.1),

$$Q(\Phi|\Phi') = \sum_{i=1}^M \left[ \sum_{k=1}^N \frac{\alpha'_i p_i(x_k|\phi'_i)}{p(x_k|\Phi')} \right] \log \alpha_i + \sum_{i=1}^M \sum_{k=1}^N \log p_i(x_k|\phi_i) \frac{\alpha'_i p_i(x_k|\phi'_i)}{p(x_k|\Phi')}, \quad (\text{A5})$$

where  $N$  is the number of data samples (number of intervals in the histogram), and the other variables are as defined in Eq. (A3). Note that maximization of  $Q(\Phi|\Phi')$  with respect to the weights,  $\alpha_i$ , is independent of the parameters,  $\phi_i$ , of the individual densities. Maximizing  $Q(\Phi|\Phi')$  with respect to the individual parameters leads to the following relations, which are special cases of Eqs. (4.5), (4.8), and (4.9) in Redner and Walker (1984),

$$\alpha_i^+ = \frac{\alpha'_i}{N} \sum_{k=1}^N \frac{p_i(x_k|\phi'_i)}{p(x_k|\Phi')}, \quad (\text{A6})$$

$$\mu_i^+ = \left\{ \sum_{k=1}^N x_k \frac{\alpha'_i p_i(x_k|\phi'_i)}{p(x_k|\Phi')} \right\} / \left\{ \sum_{k=1}^N \frac{\alpha'_i p_i(x_k|\phi'_i)}{p(x_k|\Phi')} \right\}, \quad (\text{A7})$$

$$\sigma_i^{+2} = \frac{\left\{ \sum_{k=1}^N (x_k - \mu_i^+)^2 \frac{\alpha'_i p_i(x_k|\phi'_i)}{p(x_k|\Phi')} \right\}}{\left\{ \sum_{k=1}^N \frac{\alpha'_i p_i(x_k|\phi'_i)}{p(x_k|\Phi')} \right\}}, \quad (\text{A8})$$

where  $\alpha_i^+$ ,  $\mu_i^+$ , and  $\sigma_i^{+2}$  are the parameter values used in the subsequent iteration of the algorithm. In this form of mixture density, the weights, means, and variances of the individual densities in the mixture are mutually independent. This form was used to characterize the individual modes in the interval histograms. The ML estimate of  $\mu_i$  was used as an estimate of the  $i$ th mode.

## 2. Gaussians with harmonically related means and a common variance

To estimate the fundamental mode, i.e., stimulus period, of ISI histograms, we modeled their distribution as a Gaussian mixture density with harmonically related means and a common variance,

$$p_i(x|\phi_i) = \frac{1}{\sqrt{2\pi}\sigma} e^{(x-i\cdot\mu)^2/2\sigma^2}. \quad (\text{A9})$$

The ML estimate of  $\mu$  was used as an estimate of the stimulus period.

A different set of iteration equations result when considering the mixture density described by Eqs. (A3) and (A9). In this case, maximizing  $Q(\Phi|\Phi')$  with respect to the individual parameters leads to the following iteration equations, similar to Eqs. (A7) and (A8),

$$\mu^+ = \frac{\left\{ \sum_{i=1}^M \sum_{k=1}^N x_k \cdot i \frac{\alpha'_i p_i(x_k|\phi'_i)}{p(x_k|\Phi')} \right\}}{\left\{ \sum_{i=1}^M \sum_{k=1}^N i^2 \frac{\alpha'_i p_i(x_k|\phi'_i)}{p(x_k|\Phi')} \right\}}, \quad (\text{A10})$$

$$\sigma^{+2} = \frac{\left\{ \sum_{i=1}^M \sum_{k=1}^N (x_k - i \cdot \mu^+)^2 \frac{\alpha'_i p_i(x_k|\phi'_i)}{p(x_k|\Phi')} \right\}}{\left\{ \sum_{i=1}^M \sum_{k=1}^N \frac{\alpha'_i p_i(x_k|\phi'_i)}{p(x_k|\Phi')} \right\}}. \quad (\text{A11})$$

The weights of the individual densities,  $\alpha'_i$ , are the same as in Eq. (A6).

<sup>1</sup>Dowling and Harwood (1986) report (on p. 93) only one known tonal system, from an aboriginal culture in Australia, that is not based on the octave.

<sup>2</sup>The slight deviation from monotonicity near 1500 Hz is due to differences in mode offsets across cats and uneven sampling across cats. The data at 1500 and 3000 Hz are primarily from two cats which showed relatively large mode offsets in their AN responses [cats 5 and 6 in Fig. 4(d)].

Attneave, F., and Olson, R. (1971). "Pitch as a medium: A new approach to psychophysical scaling," *Am. J. Psychol.* **84**, 147–166.

Biasutti, M. (1997). "Sharp low- and high-frequency limits on musical chord recognition," *Hear. Res.* **105**, 77–84.

Cariani, P. A., and Delgutte, B. (1996a). "Neural correlates of the pitch of complex tones. I. Pitch and pitch salience," *J. Neurophysiol.* **76**, 1698–1716.

Cariani, P. A., and Delgutte, B. (1996b). "Neural correlates of the pitch of complex tones. II. Pitch shift, pitch ambiguity, phase invariance, pitch circularity, rate pitch, and the dominance region for pitch," *J. Neurophysiol.* **76**, 1717–1734.

de Cheveigné, A. (1985). "A nerve fiber discharge model for the study of pitch," in *Transactions of the Committee on Speech Research/Hearing Research* (The Acoustical Society of Japan, Tokyo), S85-37, pp. 279–286.

Delgutte, B. (1990). "Physiological mechanisms of psychophysical masking: Observations from auditory-nerve fibers," *J. Acoust. Soc. Am.* **87**, 791–809.

Demany, L., and Semal, C. (1990). "Harmonic and melodic octave templates," *J. Acoust. Soc. Am.* **88**, 2126–2135.

Dobbins, P. A., and Cuddy, L. L. (1982). "Octave discrimination: An experimental confirmation of the 'stretched' subjective octave," *J. Acoust. Soc. Am.* **72**, 411–415.

Doughty, J., and Garner, W. (1947). "Pitch characteristics of short tones. I. Two kinds of pitch threshold," *J. Exp. Psychol.* **37**, 351–365.

Doughty, J., and Garner, W. (1948). "Pitch characteristics of short tones. II. Pitch as a function of tonal duration," *J. Exp. Psychol.* **38**, 478–494.

Dowling, W. J., and Harwood, D. L. (1986). *Music Cognition* (Academic, San Diego), Series in Cognition and Perception.

Efron, B., and Tibshirani, R. (1993). *An Introduction to the Bootstrap* (Chapman & Hall, New York), Monographs on Statistics and Applied Probability.

Evans, E. (1983). "Pitch and cochlear nerve fibre temporal discharge patterns," in *Hearing: Physiological Bases and Psychophysics*, edited by R. Klinke and R. Hartmann (Springer Verlag, Berlin), pp. 140–146.

Gaumont, R., Kim, D., and Molnar, C. (1983). "Response of cochlear nerve fibers to brief acoustic stimuli: Role of discharge-history effects," *J. Acoust. Soc. Am.* **74**, 1392–1398.

Gaumont, R., Molnar, C., and Kim, D. (1982). "Stimulus and recovery dependence of cat cochlear nerve fiber spike discharge probability," *J. Neurophysiol.* **48**, 856–873.

Goldstein, J. (1973). "An optimum processor theory for the central formation of the pitch of complex tones," *J. Acoust. Soc. Am.* **54**, 1496–1516.

Goldstein, J., and Srulovicz, P. (1977). "Auditory-nerve spike intervals as an adequate basis for aural frequency measurement," in *Psychophysics and Physiology of Hearing*, edited by E. Evans and J. Wilson (Academic, London), pp. 337–346.

Gulick, W., Gescheider, G., and Frisina, R. (1989). *Hearing: Physiological Acoustics, Neural Coding and Psychoacoustics* (Oxford University Press, New York).

Hartmann, W. (1993). "On the origin of the enlarged melodic octave," *J. Acoust. Soc. Am.* **93**, 3400–3409.

Johnson, D. H. (1980). "The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones," *J. Acoust. Soc. Am.* **68**, 1115–1122.

Kaernbach, C., and Demany, L. (1998). "Psychophysical evidence against the autocorrelation theory of auditory temporal processing," *J. Acoust. Soc. Am.* **104**, 2298–2306.

Kiang, N. (1980). "Peripheral neural processing of auditory information," in *Handbook of Physiology*, edited by I. Darian-Smith (American Physiological Society, Bethesda, MD).

Kiang, N. (1990). "Curious oddments of auditory-nerve studies," *Hear. Res.* **49**, 1–16.

Kiang, N., and Moxon, E. (1972). "Physiological considerations in artificial stimulation of the inner ear," *Ann. Otol. Rhinol. Laryngol.* **81**, 714–730.

Kiang, N., and Moxon, E. (1974). "Tails of tuning curves of auditory-nerve fibers," *J. Acoust. Soc. Am.* **55**, 620–630.

Kiang, N., Moxon, E., and Levine, R. (1970). "Auditory-nerve activity in cats with normal and abnormal cochleas," in *Sensorineural Hearing Loss*, edited by G. Wolstenholme and J. Knight (J. & A. Churchill, London), pp. 241–273.

Kiang, N., Watanabe, T., Thomas, E., and Clark, L. (1965). *Discharge Patterns of Single Fibers in the Cat's Auditory Nerve* (The MIT Press, Cambridge, MA).

Lieberman, M., and Kiang, N. (1978). "Acoustic trauma in cats: Cochlear pathology and auditory-nerve activity," *Acta Oto-Laryngol. Suppl.* **358**, 1–63.

Lieberman, M. C., and Kiang, N. Y. (1984). "Single-neuron labeling and chronic cochlear pathology. IV. Stereocilia damage and alterations in rate- and phase-level functions," *Hear. Res.* **16**, 75–90.

Licklider, L. (1951). "A duplex theory of pitch perception," *Experientia* **7**, 128–134.

Moon, T. (1996). "The expectation-maximization algorithm," *IEEE Signal Process. Mag.* **13**(6), 47–60.

Ohgushi, K. (1978). "On the role of spatial and temporal cues in the perception of the pitch of complex tones," *J. Acoust. Soc. Am.* **64**, 764–771.

Ohgushi, K. (1983). "The origin of tonality and a possible explanation of the octave enlargement phenomenon," *J. Acoust. Soc. Am.* **73**, 1694–1700.

Patterson, R., Peters, R., and Milroy, R. (1983). "Threshold duration for melodic pitch," in *Hearing: Physiological Bases and Psychophysics*, edited by R. Klinke and R. Hartmann (Springer-Verlag, Berlin), pp. 321–325.

Perkel, D., Gerstein, G., and Moore, G. (1967). "Neuronal spike trains and stochastic point processes. I. The single spike train," *Biophys. J.* **7**, 391–418.

Pollack, I. (1967). "Number of pulses required for minimal pitch," *J. Acoust. Soc. Am.* **42**, 895.

Redner, R., and Walker, H. (1984). "Mixture densities, maximum likelihood and the EM algorithm," *SIAM Rev.* **26**, 195–239.

- Rodieck, R. (1967). "Maintained activity of cat retinal ganglion cells," *J. Neurophysiol.* **30**, 1043–1071.
- Rodieck, R., Kiang, N., and Gerstein, G. (1962). "Some quantitative methods for the study of spontaneous activity of single neurons," *Biophys. J.* **2**, 351–368.
- Rose, J. E., Brugge, J. F., Anderson, D. J., and Hind, J. E. (1967). "Phase-locked response to low-frequency tones in single auditory nerve fibers of the squirrel monkey," *J. Neurophysiol.* **30**, 769–793.
- Rose, J. E., Brugge, J. F., Anderson, D. J., and Hind, J. E. (1968). "Patterns of activity in single auditory nerve fibers of the squirrel monkey," in *Hearing Mechanisms in Vertebrates*, edited by A. V. S. de Reuck and J. Knight (Churchill, London), pp. 144–168.
- Ruggero, M. (1973). "Response to noise of auditory nerve fibers in the squirrel monkey," *J. Neurophysiol.* **36**, 569–587.
- Ruggero, M. A., Rich, N. C., Shivapuja, B. G., and Temchin, A. N. (1996). "Auditory-nerve responses to low-frequency tones: Intensity dependence," *Aud. Neurosci.* **2**, 159–185.
- Siebert, W. M. (1970). "Frequency discrimination in the auditory system: Place or periodicity mechanisms?," *Proc. IEEE* **58**, 723–730.
- Sundberg, J., and Lindqvist, J. (1973). "Musical octaves and pitch," *J. Acoust. Soc. Am.* **54**, 922–929.
- Terhardt, E. (1971). "Die tonhöhe harmonischer klänge und das oktavierintervall," *Acustica* **24**, 126–136.
- Terhardt, E. (1974). "Pitch, consonance, and harmony," *J. Acoust. Soc. Am.* **55**, 1061–1069.
- Walliser, V. (1969). "Über die spreizung von empfundenen intervallen gegenüber mathematisch harmonischen intervallen bei sinustönen," *Frequenz* **23**, 139–143.
- Ward, W. (1954). "Subjective musical pitch," *J. Acoust. Soc. Am.* **26**, 369–380.