

Pitch Representations in the Auditory Nerve: Two Concurrent Complex Tones

by

Erik Larsen

M.Sc. Applied Physics
Technische Universiteit Delft, 1998

SUBMITTED TO
THE DEPARTMENT OF ELECTRICAL ENGINEERING AND COMPUTER SCIENCE
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE IN
ELECTRICAL ENGINEERING AND COMPUTER SCIENCE
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

SEPTEMBER 2008

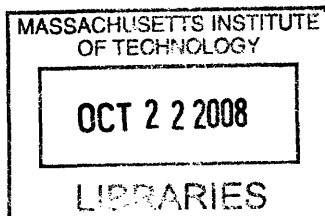
© 2008 Erik Larsen. All rights reserved.

The author hereby grants to MIT permission to reproduce
and to distribute publicly paper and electronic
copies of this thesis document in whole or in part
in any medium now known or hereafter created.

Signature of author: _____
Harvard-MIT Division of Health Sciences and Technology
August 19, 2008

Certified by: _____
Bertrand Delgutte
Senior Research Scientist, Research Laboratory of Electronics
Thesis Supervisor

Accepted by: _____
Terry P. Orlando
Professor of Electrical Engineering and Computer Science
Chair, Department Committee on Graduate Students



ARCHIVES

Pitch Representations in the Auditory Nerve: Two Concurrent Complex Tones

by

Erik Larsen

Submitted to the Department of Electrical Engineering and Computer Science
On September, 2008 in partial fulfillment of the
Requirements for the Degree of Master of Science in
Electrical Engineering and Computer Science

ABSTRACT

Pitch differences between concurrent sounds are important cues used in auditory scene analysis and also play a major role in music perception. To investigate the neural codes underlying these perceptual abilities, we recorded from single fibers in the cat auditory nerve in response to two concurrent harmonic complex tones with missing fundamentals and equal-amplitude harmonics. We investigated the efficacy of rate-place and interspike-interval codes to represent both pitches of the two tones, which had fundamental frequency (F_0) ratios of 15/14 or 11/9. We relied on the principle of scaling invariance in cochlear mechanics to infer the spatiotemporal response patterns to a given stimulus from a series of measurements made in a single fiber as a function of F_0 . Templates created by a peripheral auditory model were used to estimate the F_0 s of double complex tones from the inferred distribution of firing rate along the tonotopic axis. This rate-place representation was accurate for F_0 s above about 900 Hz. Surprisingly, rate-based F_0 estimates were accurate even when the two-tone mixture contained no resolved harmonics, so long as some harmonics were resolved prior to mixing. We also extended methods used previously for single complex tones to estimate the F_0 s of concurrent complex tones from interspike-interval distributions pooled over the tonotopic axis. The interval-based representation was accurate for F_0 s below about 900 Hz, where the two-tone mixture contained no resolved harmonics. Together, the rate-place and interval-based representations allow accurate pitch perception for concurrent sounds over the entire range of human voice and cat vocalizations.

Thesis Supervisor: Bertrand Delgutte

Title: Senior Research Scientist, Research Laboratory of Electronics

TABLE OF CONTENTS

| | | |
|-----------------|--|------------------|
| <u>A</u> | <u>INTRODUCTION</u> | <u>4</u> |
| A.1 | PITCH DISCRIMINATION AND IDENTIFICATION FOR CONCURRENT COMPLEX TONES | 4 |
| A.2 | ROLE OF F ₀ DIFFERENCES IN THE IDENTIFICATION OF CONCURRENT VOWELS..... | 5 |
| A.3 | NEURAL REPRESENTATIONS OF PITCH FOR SINGLE AND CONCURRENT COMPLEX TONES | 5 |
| A.4 | UTILIZATION OF SCALING INVARIANCE IN COCHLEAR MECHANICS | 7 |
| <u>B</u> | <u>MATERIALS AND METHODS</u> | <u>9</u> |
| B.1 | ANIMAL PREPARATION AND NEURAL RECORDING | 9 |
| B.2 | STIMULI..... | 9 |
| B.3 | ELECTROPHYSIOLOGICAL PROCEDURES..... | 10 |
| B.4 | DATA ANALYSIS..... | 10 |
| B.4.I | RATE-BASED ANALYSIS | 11 |
| B.4.II | INTERSPIKE-INTERVAL ANALYSIS | 12 |
| <u>C</u> | <u>RESULTS</u> | <u>14</u> |
| C.1 | RATE-BASED REPRESENTATION OF F ₀ FOR DOUBLE COMPLEX TONES | 14 |
| C.2 | PITCH ESTIMATION FROM RATE RESPONSES WORKS BEST FOR HIGHER CFs..... | 16 |
| C.3 | INTERSPIKE-INTERVAL ANALYSIS | 21 |
| C.4 | PITCH ESTIMATION FROM INTERSPIKE INTERVALS WORKS BEST FOR LOWER CFs | 21 |
| C.5 | EFFECT OF SPONTANEOUS RATE ON PITCH ESTIMATION FROM INTERSPIKE INTERVALS | 25 |
| <u>D</u> | <u>DISCUSSION.....</u> | <u>28</u> |
| D.1 | SCALING INVARIANCE IN COCHLEAR MECHANICS..... | 28 |
| D.1.I | DEVIATIONS FROM SCALING INVARIANCE IN COCHLEAR MECHANICS | 29 |
| D.1.II | EFFECT ON PHASE LOCKING..... | 29 |
| D.1.III | EFFECTS ON ADAPTATION AND REFRACTORINESS..... | 30 |
| D.2 | RATE-PLACE REPRESENTATION OF PITCH..... | 31 |
| D.3 | PITCH REPRESENTATION IN INTERSPIKE INTERVALS | 32 |
| D.4 | RELATION TO PSYCHOPHYSICS | 34 |
| <u>E</u> | <u>CONCLUSIONS</u> | <u>37</u> |
| <u>F</u> | <u>ACKNOWLEDGEMENTS.....</u> | <u>38</u> |
| <u>G</u> | <u>LIST OF REFERENCES.....</u> | <u>39</u> |

A Introduction

In everyday listening situations, multiple sound sources are usually present. For example, various talkers may be speaking at the same time, or different musical instruments may be playing together. To understand speech and recognize auditory objects in these situations, it is necessary to segregate sound sources from one another. Many natural sounds such as speech, animal vocalizations and the sounds of most musical instruments contain harmonic complex tones, where all the frequency components are multiples of a common fundamental frequency (F0) that gives rise to a strong pitch percept. For such harmonic sounds, a pitch difference is an important cue underlying the segregation ability (Scheffers 1983; Bregman 1990; Darwin and Carlyon 1995), particularly in adverse signal-to-noise ratios. The ability to use pitch differences to segregate sound sources is severely degraded in the hearing impaired and wearers of cochlear implants (Moore and Carlyon 2005; Deeks and Carlyon 2004; Qin and Oxenham 2005; Stickney et al. 2007; Carlyon et al. 2007; Summers and Leek 1998; Rossi-Katz and Arehart 2005). Thus, a better understanding of pitch processing for simultaneous complex tones may shed light on neural mechanisms of auditory scene analysis and lead to improved assistive devices for the deaf and hearing impaired. Yet, surprisingly few psychophysical studies (Beerends and Houtsma 1989; Carlyon 1996, 1997; Assmann and Paschall 1998; Micheyl et al. 2006) and even fewer neurophysiological studies (Tramo et al. 2000, 2001; and for concurrent vowels Palmer 1990, 1992; Keilson et al. 1997) have directly addressed the identification and discrimination of the F0s of concurrent complex tones. In this paper, we quantitatively characterize the representation of the F0s of two concurrent complex tones in both the average firing rates and the temporal discharge patterns of auditory-nerve fibers in anesthetized cat.

A.1 Pitch discrimination and identification for concurrent complex tones

An important factor in pitch perception for harmonic complex tones is the ability to hear out (“resolve”) individual harmonics. In general, tones containing resolved harmonics evoke stronger pitches and have better F0 discrimination thresholds than tones consisting entirely of unresolved harmonics (Plomp 1967; Houtsma and Smurzynski 1990; Carlyon and Shackleton 1994; Bernstein and Oxenham 2003). Psychophysical studies of F0 identification and discrimination for concurrent complex tones have also stressed the role of harmonic resolvability. Beerends and Houtsma (1989) found that musically-trained listeners could accurately identify both pitches of two concurrent complex tones, each consisting of two components, as long as at least one component of each tone was resolved. Carlyon (1996) found that F0 discrimination for a target harmonic complex tone containing resolved harmonics was not severely impaired by the presence of a masker complex tone whose components occupied the same restricted frequency region as the target. In contrast, for targets consisting of unresolved harmonics, listeners heard a single “crackling” sound rather than two tones with clear pitches, and appeared to base their judgments on complex, irregular envelope cues formed by the superposition of the target and masker waveforms. Carlyon (1996) concluded that identification of individual pitches in a tone mixture is only possible when the tones have resolved harmonics.

Micheyl et al. (2006) measured the threshold target-to-masker ratio (TMR) for discriminating the F0 of a target complex tone in the presence of a complex tone masker occupying the same frequency region as the target (1200 – 3600 Hz). Discrimination performance improved (threshold TMR decreased) when the target’s harmonics were better resolved by increasing

target F0. At the lowest F0 (100 Hz), where the target consisted entirely of unresolved harmonics, the threshold TMR for F0 discrimination was always above 0 dB, suggesting that the target had to dominate the percept for listeners to do the task, and that the masker's F0 could not be heard separately at threshold, consistent with Carlyon's (1996) results. For higher F0s (200 and 400 Hz), where some or all of the target's harmonics were resolved, threshold TMRs were typically below 0 dB, suggesting the listeners could hear out the F0s of both target and masker. However, simulations with an auditory filter model suggested that, even if the target by itself contained resolved harmonics, these harmonics were rarely resolved after mixing with the masker, suggesting that harmonic resolvability *in the tone mixture* may not be necessary for both F0s to be heard.

Taken together, these studies suggest that, while peripheral resolvability is an important factor in pitch identification and discrimination for concurrent complex tones, there are still questions about its exact role. The present study was designed to include stimulus conditions with both resolved and unresolved harmonics in order to assess the role of resolvability in the neural coding of concurrent F0s.

A.2 Role of F0 differences in the identification of concurrent vowels

Many studies on the perceptual efficacy of pitch differences for segregating sound sources have focused on a relatively simple task, the identification of two concurrent, synthetic vowels (e.g. Scheffers 1983; Assmann and Summerfield 1989; Culling and Darwin 1993; de Cheveigné 1997a,b, 1999a). These studies have shown that identification performance improves with increasing difference in F0 between the two vowels, although performance is already well above chance when both vowels have the same F0. Most models for this phenomenon predict that the performance improvement is dependent upon the identification of at least one of the two pitches of the concurrent vowels (Meddis and Hewitt 1992; de Cheveigné 1997c), and some models require the identification of both pitches (Scheffers 1983; Assmann and Summerfield 1990). However, Assmann and Paschall (1998) found that listeners can reliably match both pitches of a concurrent vowel to that of a harmonic complex tone when the F0 separation is 4 semitones or greater, but that they appear to hear a single pitch at smaller separations. Most of the improvement in identification performance with concurrent vowels occurs for F0 separations below 1 semitone, in a range where Assmann and Paschall's listeners seem to hear only one pitch intermediate between the vowels' two F0s. For small F0 separations, the waveforms of concurrent vowels contain cues to vowel identity (such as beats between neighboring harmonics) that do not require an explicit identification of either F0 (Culling and Darwin 1994; Assmann and Summerfield 1994; but see de Cheveigné 1999b for a contrasting view). Thus, concurrent vowel identification may rely on different strategies depending on the size of the F0 difference between the vowels. One goal of the present study was to evaluate whether a neural correlate of this difference is found at the level of the auditory nerve.

A.3 Neural representations of pitch for single and concurrent complex tones

Studies of the coding of harmonic complex tones in the auditory nerve (AN) and cochlear nucleus (CN) have shown that pitch cues are available in both the temporal discharge patterns and the spatial distribution of activity along the tonotopic axis. Most studies have focused on temporal pitch cues, particularly those available in interspike interval distributions (Javel 1980; Evans 1983; Palmer 1990; Shofner 1991; Palmer and Winter 1993; Rhode 1995; Cariani and

Delgutte 1996a,b; Winter et al. 2003). These cues are closely related to the autocorrelation model of pitch (Licklider 1951; Meddis and Hewitt 1991) because the all-order interspike interval distribution is formally equivalent to the autocorrelation of the spike train. This interval-based pitch representation works with both resolved and unresolved harmonics (Meddis and Hewitt 1991; Cariani and Delgutte 1996a; Carlyon 1998; Cedolin and Delgutte 2005).

Fewer studies have focused on place cues to pitch, perhaps because such cues are not found in experimental animals such as cat and guinea pig when using F0s in the range of human voice (100-200 Hz). However, Cedolin and Delgutte (2005), using F0s above 400-500 Hz appropriate for cat vocalizations, found that the spatial profiles of average firing rates of AN fibers along the tonotopic axis have peaks at the locations of resolved harmonics for low and moderate stimulus levels. In principle, these rate-place cues to pitch could be extracted by a central harmonic template mechanism (Goldstein 1973; Wightman 1973; Shamma and Klein 2000) to obtain precise estimates of the stimulus F0. The place cues can also be combined with temporal cues to give various spatio-temporal representations of pitch (Loeb et al. 1983; Shamma 1985; Cedolin and Delgutte 2007; de Cheveigné and Pressnitzer 2006). The present study is a direct extension of the work of Cedolin and Delgutte (2005) to concurrent complex tones, and looks at both rate-place and interval-based representations of pitch over a wider range of F0s than in previous studies.

Only a few studies have directly examined the representation of the F0s of concurrent complex tones in the AN and CN (Palmer 1990, 1992; Tramo et al. 2000; Keilson et al. 1997; Sinex 2008). Using two concurrent vowels with F0s of 100 and 125 Hz, Palmer found that the temporal discharge patterns of AN fibers contained sufficient information to identify both F0s. In particular, each of the two F0s appeared to be individually represented in the pooled interspike interval distribution (obtained by summing interval distributions over the entire sample of AN fibers). Using the F0 of the dominant vowel estimated from the pooled distribution, Palmer (1992) successfully implemented the Meddis and Hewitt (1992) vowel segregation model using his AN data as input. This work suggests that the F0s of two concurrent vowels can be estimated from purely temporal information, while the vowel identities can be determined from a combination of place and temporal information once the dominant F0 is known.

Tramo et al. (2000) measured responses of AN fibers to pairs of concurrent complex tones consisting of 6 equal-amplitude harmonics. The lower F0 was always 440 Hz, and the F0 ratios were chosen to form musical intervals varying in consonance: minor second ($16/15$, ~1 semitone), perfect fourth ($4/3$, ~5 semitones), tritone ($45/32$, ~6 semitones), and perfect fifth ($3/2$, ~7 semitones). For all musical intervals, the pooled interspike interval distribution showed peaks at the periods of both F0s and their multiples. In addition, for musically consonant intervals (fourth and fifth), there was a pronounced peak at the fundamental period of the two-tone complex, consistent with the perception of a low pitch at that frequency¹ (Terhardt 1974). These results and those of Palmer (1990) suggest that interspike interval distributions contain detailed information about the pitches produced by concurrent complex tones with F0s in the range of speech and music.

Keilson et al. (1997) measured single-unit responses to two concurrent vowels in the cat ventral cochlear nucleus, using F0 separations of 11%, 14%, and 27%. They proposed a “periodicity-tagged” spectral representation in which a unit’s average firing rate in response to a

¹ Peaks at the fundamental period of the mixture were also present for the dissonant musical intervals (minor second and tritone) but they occurred for very long interspike intervals (> 30 ms) and were therefore unlikely to be associated with pitch percepts (Pressnitzer et al. 2001).

double vowel is partly assigned to each vowel in proportion to the synchrony to the F0 of each vowel. The periodicity-tagged representation was most effective in representing both vowel spectra in chopper units, and also worked to some extent in primary-like units. This scheme has the advantage of not requiring precise phase locking to the harmonics of the vowels; such phase locking to the fine time structure becomes increasingly rare as one ascends the auditory pathway. However, this study did not directly address how the F0s of the two vowels are estimated from the neural data since the analysis assumed the F0s were known *a priori*. Moreover, periodicity tagging requires the neuron responses to be temporally modulated at the F0s of the vowels, which can only occur with unresolved harmonics. Thus, this scheme is not likely to work with resolved harmonics, which appear to be necessary for precise F0 identification and discrimination with concurrent complex tones (Carlyon 1996; Micheyl et al. 2006).

The present study systematically investigates the effect of F0 range and F0 differences on the ability of AN discharges to represent both F0s of two concurrent complex tones. Unlike previous studies, we use stimulus conditions with both resolved and unresolved harmonics, and examine both rate-place and interval-based representations of pitch over a wide range of F0s. With both representations, we derive quantitative estimates of pitch that can be compared with each other and with psychophysical data. We use tones with equal-amplitude harmonics instead of vowels in order to give equal weight to all spectral regions and to facilitate the use of the scaling invariance principle (see below). We use two different F0 separations (about 1 and 4 semitones) to approximate the conditions when pitch matches by the listeners of Assmann and Paschall (1998) were unimodal and bimodal, respectively. A preliminary report of this work has been presented (Larsen et al. 2005), and a full report has been published (Larsen et al. 2008).

A.4 Utilization of scaling invariance in cochlear mechanics

The most direct way to study the neural representation of pitch would be to measure the response to a given stimulus as a function of both time and cochlear place, which maps to characteristic frequency (CF). Since a fine and regular sampling of the CF axis with a resolution of less than a semitone is hard to achieve in neurophysiology, we relied instead on the principle of scaling invariance in cochlear mechanics (Zweig 1976) to infer the spatio-temporal response pattern from measurements made at a single CF. Scaling invariance means that the response to a tone with frequency f at the cochlear location tuned to CF is only dependent on the ratio f/CF . It implies that the response to a *single* F0 over a *range* of CFs can be inferred from the response to a *range* of F0s at a *single* CF, if time t and frequency $F0$ are represented in dimensionless units: $t \times F0$ (cycles) and $CF/F0$ ("neural harmonic number"). Similar ideas have been used by other investigators without explicitly invoking the principle of scaling invariance (Pickles 1984; Young et al. 1992; Keilson et al. 1997; May et al. 1998; May 2003; Heinz 2005). Figure 1 illustrates scaling invariance using a model based on a bank of gammatone auditory filters (Patterson et al. 1995) with bandwidths typical for the cat cochlea (Carney and Yin 1988), followed by half-wave rectification. The left panel shows the model spatio-temporal response pattern for a harmonic complex tone with an F0 of 1 kHz. This pattern is very similar to that shown on the right, obtained by plotting the response of one model filter ($CF = 3.5$ kHz) to a series of tones with varying F0, chosen to yield the same $CF/F0$ values as on the left. The "rate-place profiles" obtained by averaging the spatio-temporal patterns over time, are also similar for the two methods. Although the model used in Fig. 1 is highly simplified and does not include many of the cochlear nonlinearities, similar results are obtained with a more sophisticated model (Zhang et al. 2001), as shown in Fig. 2 of Cedolin and Delgutte (2007).

Scaling invariance is a good approximation when applied to a local region of the cochlea, but does not hold over wide cochlear spans (Shera and Guinan 2003; van der Heijden and Joris 2006). Since F_0 was varied over a limited range in our experiments (about 2 octaves), deviations from scaling invariance may not present a major problem, as Fig. 1 suggests. This and other issues related to scaling invariance are addressed in the Discussion.

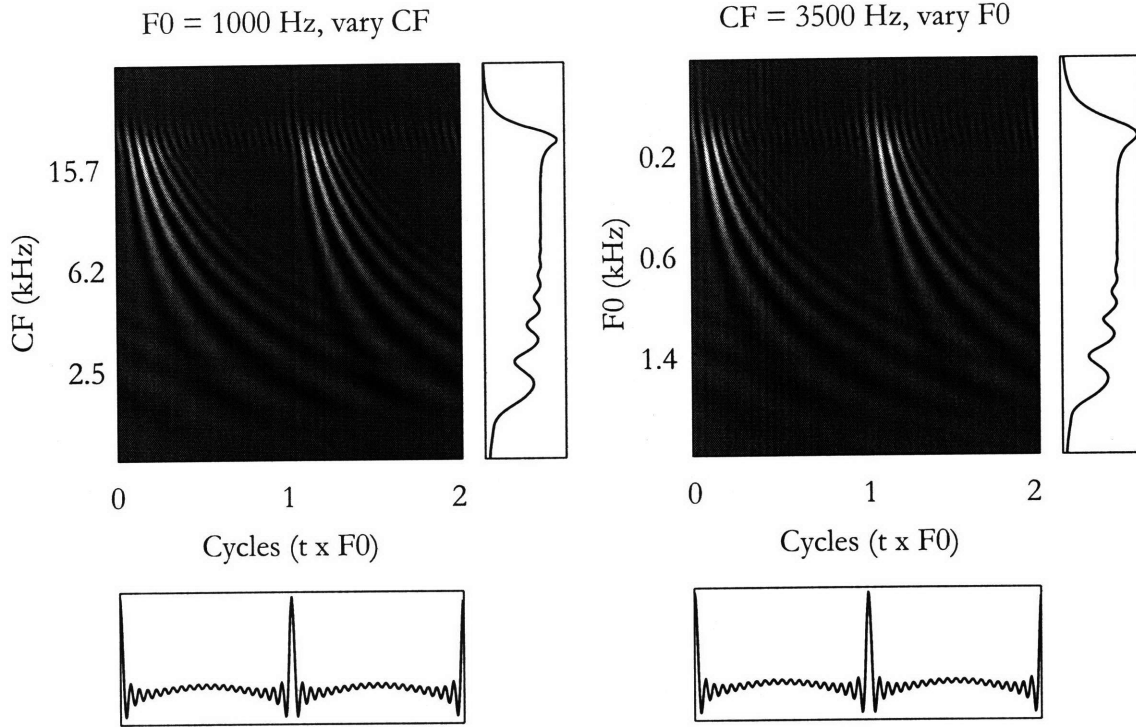


Figure 1. Illustration of scaling invariance in cochlear mechanics using a peripheral auditory model. Left: Model response of an array of auditory nerve fibers with different characteristic frequencies (CF) to a harmonic complex tone with equal-amplitude harmonics ($F_0 = 1000$ Hz). Right: Model response for one AN fiber ($CF = 3500$ Hz) as a function of the F_0 of a harmonic complex tone. F_0 values were chosen to obtain the same set of normalized frequencies CF/F_0 as on the left, so that responses in the two panels should be identical if scaling invariance holds. For both panels, gray scale represents the response amplitude, and the time scale is normalized to units of stimulus cycle ($t \times F_0$). Lower panels show the stimulus waveform on the same normalized scale. Small panels on the right of each main panel show the average model firing rate as a function of CF/F_0 , obtained by summing the spatio-temporal response patterns over one stimulus cycle.

B Materials and Methods

B.1 Animal preparation and neural recording

Single-unit responses were obtained from the auditory nerve (AN) of five female cats, aged 4-8 months. The surgical and experimental procedures were as described in Kiang et al. (1965) and Cariani and Delgutte (1996a) and were approved by the Animal Care Committees of both the Massachusetts Eye & Ear Infirmary and MIT. Briefly, animals were anesthetized with dial-in-urethane, 75 mg/kg I.P. initially, and boosters were given as needed to maintain an areflexive state. Dexamethasone (0.25 mg/kg, I.M.) was given every three hours to reduce edema and Ringer solution (50 ml/d, I.V.) was given to prevent dehydration. The AN was exposed via a posterior craniotomy and medial retraction of the cerebellum. The bulla was opened to enable measurement of gross cochlear potentials at the round window, and the middle-ear cavity was vented. General physiological state was assessed by monitoring heart rate, respiratory rate, exhaled CO₂ concentration, and rectal temperature, which was maintained at 37°C by a thermostat-controlled heating pad.

Cochlear function and stability were assessed by monitoring both pure-tone thresholds of high spontaneous-rate fibers and compound action potential (CAP) threshold to clicks measured with a silver-wire electrode placed on the bone near the round window. A significant increase (5 dB or more) in either CAP threshold or single-unit thresholds would cause termination of the experiment.

Single-unit activity was measured with glass micropipettes filled with 2M KCl. The electrode signal was amplified, band-pass filtered (0.3 – 3 kHz), and fed to a custom spike detector. Spikes were timed at 1 μ s resolution, and only recordings with good signal-to-noise ratio were used.

After each experiment, fiber thresholds were plotted as a function of CF and compared to thresholds from animals raised in a sound-proof chamber (Lieberman 1978). Data from CF regions with a high proportion of abnormally high fiber thresholds were excluded.

B.2 Stimuli

All complex tones consisted of equal-amplitude harmonics (numbers 2-20, i.e. excluding the fundamental) in cosine phase. “Double complex tones” consisted of two complex tones with different F0s. The ratio of F0s in the double complex tone was either 15/14 (~7%, slightly larger than one semitone) or 11/9 (~22%, slightly less than four semitones). These particular ratios were chosen as a compromise between two competing goals: minimizing the overlap between harmonics of the two tones that arises with ratios of small integers, while still having a mixture waveform with a well defined period so as to facilitate data analysis. Levels of complex tones are expressed as dB SPL per component.

Stimuli were generated by a 16-bit D/A converter (National Instruments NIDAC 6052e) using a sampling rate of 100 kHz. They were delivered to the tympanic membrane via a calibrated closed acoustic system comprising an electrodynamic loudspeaker (Realistic 40-1377) and a probe-tube microphone. The frequency response of the acoustic system was measured between 0 and 35 kHz and used to design digital inverse filters that equalized the sound pressure (magnitude and phase) at the tympanic membrane for all acoustic stimuli.

B.3 Electrophysiological procedures

Clicks (100 μ s, 10/s) at 55 dB SPL were used as search stimuli. Upon contacting a fiber, a frequency tuning curve was measured with an automated tracking algorithm (Kiang and Moxon 1974) to determine the CF and threshold at CF. Spontaneous discharge rate (SR) was measured over 20 s. A rate-level function was measured for a 500-ms single complex tone with an F0 such that the fifth harmonic would be near the fiber CF. Tone level was varied from approximately –10 to 60 dB re threshold at CF in 10 dB steps. This measurement was used to determine the level that produced approximately half the maximum driven rate; this was typically 15-35 dB above threshold. Subsequently, this stimulus level was used to measure responses to both single and double complex tones as a function of F0. The corresponding absolute levels ranged between 15 and 85 dB SPL per component, although most (about 3/4) were between 30 and 60 dB SPL.

For each fiber, the F0 range of single complex tones was selected in relation to the CF such that the “neural harmonic number” CF/F0 varied from approximately 1.5 to approximately 5.5 in steps of 1/8, creating 33 F0 values in total². This fine sampling of F0 causes successive low-order harmonics (2 through 5, which are most important for determining pitch for missing-fundamental stimuli) to slowly traverse the auditory filter centered at the CF, leading to a regular modulation in firing rate as a function of F0 if these harmonics are resolved. For double complex tones, the lower F0 was varied over the same range as for single complex tones, while the higher F0 was varied proportionately to keep the frequency ratio at either 15/14 or 11/9. Each of the 33 F0 steps lasted 520 ms, including a 20-ms transition interval during which the waveform for one F0 gradually decayed while overlapping with the gradual build up of the waveform for the subsequent F0. Responses were typically collected over 20 repetitions of the 17.15-s stimulus (33 steps \times 520 ms) with no interruption, for a total duration of nearly 6 min.

Depending on contact time with the fiber, we were able to measure responses as a function of F0 for a single complex tone, a double complex tone with F0 ratios of either 15/14 or 11/9, or all three stimuli. Most fibers were studied with 2-3 of these stimuli. Because the measurement order was randomized for each fiber, in some cases responses are only available for a single complex tones or 1 or 2 double complex tones. For a small number of fibers, we were able to measure responses to single and double complex tones at more than one stimulus level; however, these data were too limited to warrant a detailed analysis of the effect of level on the F0 representations.

B.4 Data Analysis

We developed quantitative methods for estimating the F0s of double complex tones from both the rate responses and the interspike interval distributions of auditory-nerve fibers. These methods are generalizations of those used by Cedolin and Delgutte (2005) to assess the representation of single complex tones.

A consequence of applying the principle of cochlear scaling invariance is that the term ‘F0’ will be used in two distinct ways. First, we use stimuli with varying F0 to “probe” the spatio-temporal response pattern using data from a single fiber, and thus use the term *probe F0* (F0_p). If

² To expedite data collection, complete sets of single and double complex tone stimuli were pre-synthesized for a limited number of CFs spaced one half-octave apart. After measuring a fiber’s CF, the stimulus set synthesized for the nearest CF was selected for study. For this reason, the range of neural harmonic numbers can deviate by as much as $\pm 1/4$ octave from the nominal 1.5 to 5.5. This is apparent, for example, on the horizontal axes of Fig. 2, and on the vertical axis of Fig. 5B.

scaling invariance holds, the observed response pattern is the same as would be obtained by measuring the response of an array of “virtual fibers” to a single complex tone as a function of cochlear place or CF (cf. Fig. 1). We call the F0 of this hypothetical complex tone the “effective F0”. The effective F0 and the CFs of the virtual fibers are constrained in that their ratios $CF_{\text{virt}}/F0_{\text{eff}}$ must match the neural harmonic numbers $CF/F0_p$ used in probing the single-fiber response. In practice, we define the effective F0 to be geometric mean of the set of probe F0s used to study a given fiber, i.e. approximately $CF/3.3$. This choice ensures that the CFs of the virtual fibers are geometrically centered at the CF of the actual fiber from which responses were measured, thereby minimizing the effects of deviations from scaling invariance. Once the virtual CFs are defined, we quantitatively assess how well the effective F0s of double complex tones (assumed to be unknown) can be estimated from the measured rate responses to double tones. We independently estimate how well the effective F0s of both single and double tones can be estimated from interspike interval distributions.

The first step in the analysis was to select the spikes occurring during the 500-ms steady-state portion of the complex tones for each probe F0, excluding the 20-ms transition intervals over which waveforms for subsequent F0 values overlap. For double complex tones, the analysis interval was further constrained to span an integer number of periods of the two-tone mixture to avoid possible biases resulting from the varying phase relationships between the two probe F0s over each cycle of the complex. This fundamental period corresponds to 9 times the period of the lower probe F0 (11 times the period of the higher F0) for the 11/9 ratio, and 14 times the period of the lower probe F0 (15 times the period of the higher F0) for the 15/14 F0 ratio.

B.4.i Rate-based analysis

The rate-place analysis is based on the idea that the average firing rate of an AN fiber should vary systematically as resolved partials of a single or double complex tone move across the fiber’s response area when probe F0 is varied; the rate should show a maximum when a partial coincides with the CF and a minimum when the CF falls between two resolved partials (Cedolin and Delgutte 2005). The locations of these maxima and minima give information about the effective F0s of double complex tones. Specifically, we used a three-step process for quantitatively estimating the two effective F0s of double complex tones from the rate responses of each fiber. In the first step, the parameters of a phenomenological model for the rate responses of AN fibers are fit to the response to a single complex tone as a function of probe F0. In the second step, scaling invariance is used to convert the single-fiber model fit in Step 1 into a model for an array of virtual fibers with varying CFs. In the third step, we find the two effective F0s of a double complex tone with equal-amplitude harmonics which, when input to the virtual fiber array model, give the best approximation to the measured responses to a double complex tone as a function of probe F0. Note that this method requires measurements of both single and double complex tone responses, which were not available for every fiber.

The phenomenological model of rate responses of a single fiber (Cedolin and Delgutte 2005) consists of three cascaded stages: (1) a rounded exponential filter (Patterson and Nimmo-Smith 1980) representing peripheral frequency selectivity; (2) computation of the root mean square (r.m.s.) amplitude over time at the filter output, and (3) a saturating nonlinearity representing the dependence of rate on level (Sachs and Abbas 1974). The model has 5 free parameters which are fit to the single-tone rate response as a function of probe F0: (i) the filter center frequency, (ii) the filter bandwidth, (iii) the spontaneous discharge rate, (iv) the maximum driven rate, and (v) the sound level at which driven rate is 50% of maximum. The filter center frequency estimated

by this fitting procedure is called BF_{CT} ('best frequency' in response to a complex tone) to distinguish it from the CF measured from pure-tone tuning curves.

In Step 2 of the estimation procedure, scaling invariance is used to convert the single-fiber model from in Step 1 into a model for the rate response of an array of virtual fibers with varying CFs. Specifically, each probe F0 is mapped into the CF of one virtual fiber using the equation:

$$\{CF_{virt}\} = F0_{eff} \times \{nh\}, \quad \text{with } \{nh\} = BF_{CT}/\{F0_p\} \quad (1)$$

$F0_{eff}$ is the effective F0 (the geometric mean of the probe F0s), $\{nh\}$ is the vector of neural harmonic numbers (varying from approximately 1.5 to 5.5), and $\{F0_p\}$ is the vector of 33 probe F0s of the single complex tones used in Step 1. With this convention, the CF_{virt} of the virtual fibers are approximately geometrically centered at BF_{CT} and encompass harmonics 2 through 5 of a single complex tone at the effective F0. All the model parameters are determined from the fit in Step 1, except that CF_{virt} varies as in Eq. (1), and the filter bandwidths vary proportionately to CF_{virt} in order to enforce scaling invariance. Thus specified, the model can predict the rate response of the virtual fiber array to any sum of sinusoids, including double complex tones with arbitrary F0s.

In Step 3 of the estimation procedure, the two effective F0s of a double complex tone input to the model are adjusted to best predict the measured rate responses to a set of double complex tones with varying probe F0s. The best matching input F0s are the estimated effective F0s of the double complex tone. Note that the effective F0s are assumed to be unknown in order to quantitatively assess how well they can be estimated from the neural data, assuming the virtual CFs specified in Eq. (1).

A Levenberg-Marquardt iterative least-squares optimization routine implemented in MATLAB (Mathworks, Framingham MA) was used both to fit model parameters to the single-tone response (Step 1), and to find the effective F0s that give the best match between model predictions and measured rate responses to double complex tones (Step 3). In order to reduce the possibility of finding a local minimum of the residuals rather than the true minimum, five randomized sets of starting values (typically differing by $\pm 20\%$) were used for the fitted parameters, and the best resulting fit was retained. Standard deviations of the effective F0 estimates (Fig. 4) were computed based on the r.m.s. residuals and the Jacobian at the solution vector (Press et al. 1992).

B.4.ii Interspike-interval analysis

Our method for estimating the two F0s of double complex tones from the temporal discharge patterns of AN fibers is a direct extension of methods used previously to estimate the F0 of single complex tones from interspike interval distributions (Palmer 1990; Cariani and Delgutte 1996a,b; Cedolin and Delgutte 2005). The main difference is that, using scaling invariance, the present method gives effective F0 estimates from the response of a *single* fiber measured as a function of probe F0, while previous methods estimated F0 from the response of a population of fibers to a *single* stimulus. The method consists of two steps (Fig. 5): (1) computation of a "pseudo-pooled" interspike interval distribution from the responses of a single fiber as a function of probe F0, (2) estimation of the effective F0 by fitting periodic templates to the pseudo-pooled interval distribution.

We first compute an all-order interspike interval distribution for every probe F0 in a series of single- or double complex tones. To implement scaling invariance, the interspike intervals are

computed on a normalized time scale ($t \times F_0$) by always using 45 bins in each stimulus cycle (in case of double tones, this is the period of the tone with the lower F_0), meaning the bin width is inversely proportional to probe F_0 . The time-normalized interspike interval distributions are then summed across all probe F_0 s to form pseudo-pooled interval distributions. These are not true pooled distribution since pooling normally refers to summation across fibers for a single stimulus, while we sum across stimuli (across probe F_0 s) for a single fiber. Pooling the scaled interval distributions allows a single estimate of the effective F_0 to be obtained from responses to 33 different probe F_0 s.

To estimate the effective F_0 from pseudo-pooled interval distributions, we used periodic templates that select intervals at a given period and its multiples. For each template, *contrast* is defined as the ratio of the mean number of intervals in the template bins to the mean number of intervals per bin in the entire histogram (Cariani and Delgutte 1996a,b). A contrast value of 1 implies no temporal structure at the template F_0 , while larger contrast values imply that the fiber preferentially fires at that interval. Contrast has been shown to correlate with psychophysical pitch strength for a wide variety of stimuli (Cariani and Delgutte 1996a,b). Contrast values are computed for a range of template F_0 s (from 0.29 to 3.5 times the effective F_0), and effective F_0 s are estimated based on maxima in contrast. For a single complex tone, the estimated F_0 is simply the template F_0 which maximizes contrast. For a double complex tone, the two template F_0 s with the highest contrasts are selected, with the constraint that the F_0 of the second estimate cannot be a multiple or sub-multiple of the F_0 giving the largest contrast.

To make this method more robust, the pseudo-pooled ISID was weighted with an exponentially-decaying function which de-emphasizes long interspike intervals corresponding to low effective F_0 s. This weighting implements the idea that the existence of a lower F_0 limit to pitch perception (Pressnitzer et al. 2001) implies that the auditory system is unable to use very long intervals in forming pitch percepts. In practice, the weighting reduces the template contrast at subharmonic frequencies of the effective F_0 , thereby preventing F_0 matches to these subharmonics. A decay constant equal to 0.75 times the period of the lower F_0 was found empirically to give a good compromise between reducing subharmonic errors and decreasing template contrast at the effective F_0 too much (which could lead to harmonic errors).

C Results

Results are based on recordings from 107 AN fibers in five cats. Fifty of these fibers (47%) had high spontaneous discharge rates (>18 spikes/s, Liberman 1978), 43 (40%) had medium-spontaneous rate ($0.5 < \text{SR} < 18/\text{s}$), and 14 (13%) had low spontaneous rates ($< 0.5/\text{s}$). The CF distribution was fairly uniform on a logarithmic scale between 1 and 14 kHz, but the sampling was somewhat less dense below 1 kHz down to 200 Hz. Single- and double complex tones were typically presented at 15-35 dB above the fiber's pure tone threshold at CF, about half-way into the fiber's dynamic range as determined by the rate-level function for a single complex tone.

C.1 Rate-based representation of F0 for double complex tones

Figure 2 illustrates the procedure for estimating the effective F0s of double complex tones from rate responses to both single and double complex tones using an example for a medium spontaneous-rate fiber (CF = 7000 Hz). Figure 2A shows the rate response to a single complex tone as a function of probe F0 (filled circles) together with the fitted response of the peripheral auditory model (solid trace). Consistent with previous results for higher-CF fibers at moderate stimulus levels (Cedolin and Delgutte 2005), the rate response of this fiber shows peaks at integer values of the “neural harmonic number” CF/F0. These peaks occur when a resolved harmonic coincides with the fiber CF. The pattern of harmonically-related peaks allows the fiber's best frequency (BF_{CT}) to be precisely estimated from the rate response to the single complex tone. The model fit for this fiber gave a BF_{CT} of 7049 Hz, very close to the CF measured from the pure-tone tuning curve (0.8% difference).

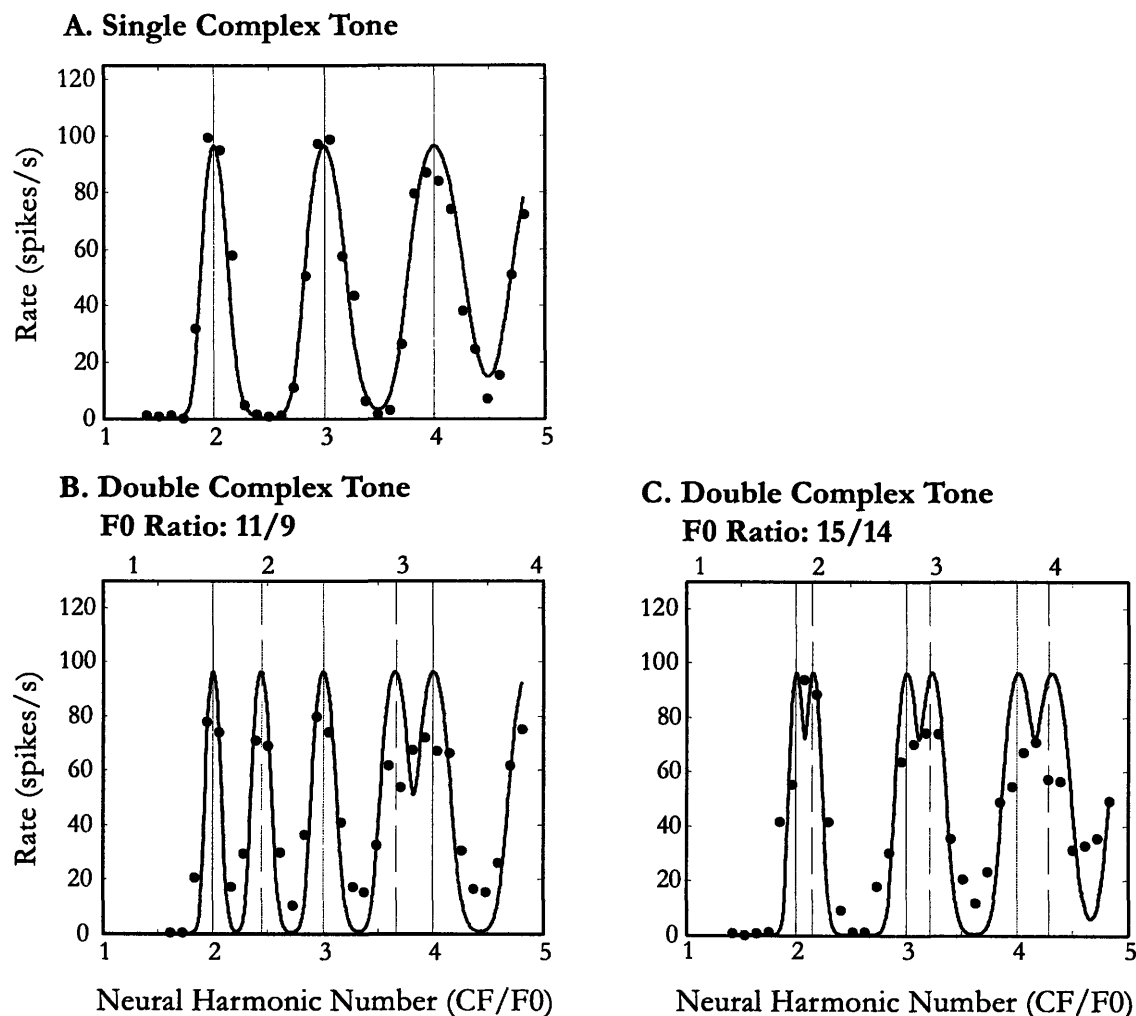


Figure 2. Template-matching procedure used to estimate both F0s of a double complex tone from the rate response of an AN fiber ($CF = 7000$ Hz). A: Rate response to a single complex tone as a function of F0. The abscissa represents the neural harmonic number $CF/F0$. The measured data (dots) were used to fit the response of an auditory nerve model (solid line). Vertical lines show the F0s for which Harmonics 2, 3, and 4 coincide with the CF. B and C: Measured rate responses (dots) and model predictions (solid line) for concurrent complex tones in which the F0s were varied proportionately to keep the F0 ratio at 11/9 (B), and 15/14 (C). Model parameters were fixed as in A, and the F0s of a double tone input to the model were adjusted to best predict the data, thereby giving quantitative estimates of the F0s. Lower and upper horizontal axes show the ratio $CF/F0$ for the lower and the higher tone, respectively, while solid and dashed vertical lines represent Harmonics 2, 3 and 4 of the lower and higher tone, respectively.

Figure 2B shows the measured rate response and model predictions for a double complex tone with an F0 ratio of 11/9. The vertical lines show the positions of the harmonics of both tones (lower tone: solid lines, higher tone: dashed lines), where we would expect maxima in the rate response if these harmonics were resolved in the two-tone mixture. Indeed, the rate response shows peaks at Harmonics 2 and 3 of the lower tone and Harmonic 2 of the higher tone. In contrast, Harmonic 4 of the lower tone and Harmonic 3 of the higher tone are poorly separated, even though these harmonics were well resolved before mixing (Fig. 2A). The predicted model response captures the main peaks and troughs in the response fairly well, although it tends to overestimate the peak amplitudes, perhaps because the model does not

explicitly include adaptation. Adaptation may be stronger for double complex tones than for single complex tones because, with the double tone, the fiber more frequently receives strong stimulation from a component close to the CF. Note that this is a prediction, not a fit, since the model parameters were fixed to the values derived from Fig. 2A. However, the F0s at the input to the model were adjusted so as to obtain the best prediction. The effective F0s estimated in this way have errors of 0.10% and -0.34%, for the lower and higher tone, respectively. These estimates are remarkably accurate considering that they are based on data from a single fiber, with 20 stimulus repetitions for each probe F0.

Figure 2C shows measured responses and model predictions for a double complex tone with an F0 ratio of 15/14. In this case, the rate response shows three broad peaks encompassing Harmonics 2, 3 and 4 of both tones, but there is no dip in between equal-numbered harmonics of the two tones due to the close spacing of these harmonics relative to cochlear bandwidth. One cue to the presence of two complex tones is that each peak in the rate response becomes broader with increasing harmonic number because the separation between same-numbered harmonics of the two tones increases. Moreover, the peaks are broader than corresponding peaks in the single-tone response of Fig. 2A. Again, the model prediction captures the main peaks and troughs in the response, with a tendency to overestimate the peak amplitudes for the higher harmonics. In this case, the effective F0 estimates have errors of 0.35% and 0.68% for the lower and higher tone, respectively. These estimates are quite accurate despite the lack of a peak in the rate response at any individual harmonic. This result challenges the conventional assumption that a tone mixture must have resolved partials for a spectral pitch code to be effective. Even though peripheral frequency resolution is not good enough to separate same-numbered harmonics from the two tones in the mixture, our template-matching procedure does give accurate estimates of the underlying F0s.

C.2 Pitch estimation from rate responses works best for higher CFs

Rate responses to single complex tones were measured for 74 fibers, giving a total of 85 responses. From 55 of these fibers we also measured responses to double complex tones. In total we obtained 112 double complex tone responses (about two per fiber). A t-test revealed no significant difference ($p=0.87$) in mean pitch estimation performance between high (>18.5 spikes/s) and low/medium (<18.5 spikes/s) spontaneous rate (SR) groups, so we analyze data for these two groups of fibers together. The lack of an SR effect on rate-based pitch estimation is most likely explained by our choice of stimulus levels about half-way into each fiber's dynamic range, ensuring a strong response to the complex tone partials yet avoiding saturation for all SR groups.

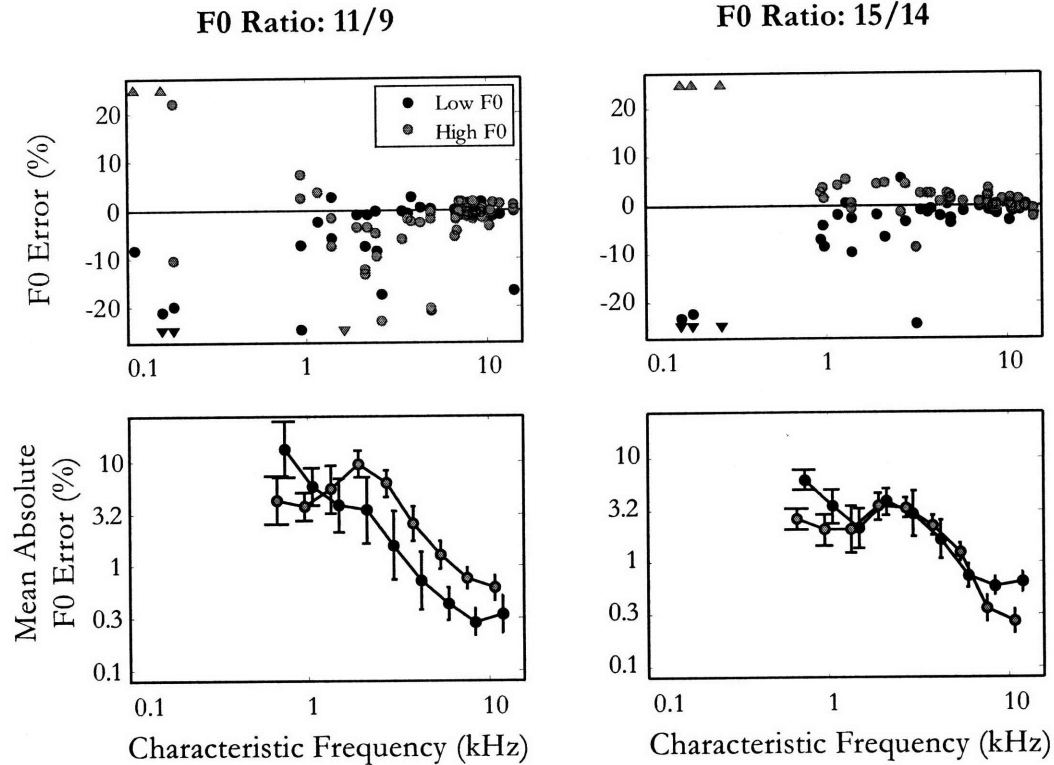


Figure 3. Percent F0 estimation errors from rate responses to double complex tone as a function of CF for the population of AN fibers. Left and right panels show results for F0 ratios of 11/9 and 15/14, respectively. Upper panels show errors for individual AN fibers; lower panels show moving window averages of the log-transformed absolute errors, using one-octave wide CF bins with 50% overlap. Estimation errors for the low (dark) and high (light) tone in each double complex tone are shown separately. Error bars in the bottom represent ± 1 standard error of the mean. In the top panels, triangles show data that lie out of the vertical range.

Figure 3 shows percent errors of F0 estimation for double complex tones with F0 ratios of 11/9 (left) and 15/14 (right) for these 112 measurements. The upper panels show estimation errors for individual fibers, while the lower panels show moving-window averages of the log-transformed absolute estimation errors. The horizontal axis in these figures is the CF obtained from the pure-tone tuning curve, not BF_{CT} .

The absolute estimation errors in Fig. 3 are relatively large ($>2\%$) for CF below 2-3 kHz, and decrease with increasing CF to fall below 1% above 5 kHz. This improvement in estimation performance is gradual and does not show a sharp transition, as was also found by Cedolin and Delgutte (2005) for single complex tones. It is consistent with the gradual improvement in *relative* frequency selectivity of AN fibers (as measured by the quality factor Q) with increasing CF (Kiang et al. 1965; Liberman et al. 1978). The data were processed with a three-way analysis of variance (ANOVA) with F0 ratio (11/9 and 15/14), CF range (0.5-2 kHz, 2-4 kHz, 4-8 kHz, 8-16 kHz), and tone height (low and high) as factors. There was a significant main effect of CF, as expected ($F(3, 188)=29.96$, $p<0.001$), but no effect of F0 ratio and tone height. However, there was a significant interaction between F0 ratio and tone height ($F(1, 188)=4.89$, $p=0.028$), indicating that that mean absolute errors for the higher tone are larger than errors for the lower tone for the 11/9 F0 ratio, while they are similar for the 15/14 F0 ratio (Tukey-Kramer post-hoc analysis).

There is a tendency, strongest with the 15/14 F0 ratio, for the low F0 to be underestimated, and for the high F0 to be overestimated (from the upper panels in Fig. 3). This bias is most likely the result of assigning the two effective F0 estimates as either “low” or “high” so as to minimize the combined error with respect to both F0s. This makes it relatively unlikely that an estimated effective F0 which is greater than the actual low F0 will be categorized as “low”, unless it lies in a narrow region halfway between the actual low and high F0s. Effectively, “low F0s” are biased to be underestimates, and “high F0s” are biased to be overestimates. These biases are more pronounced for closely spaced F0 pairs.

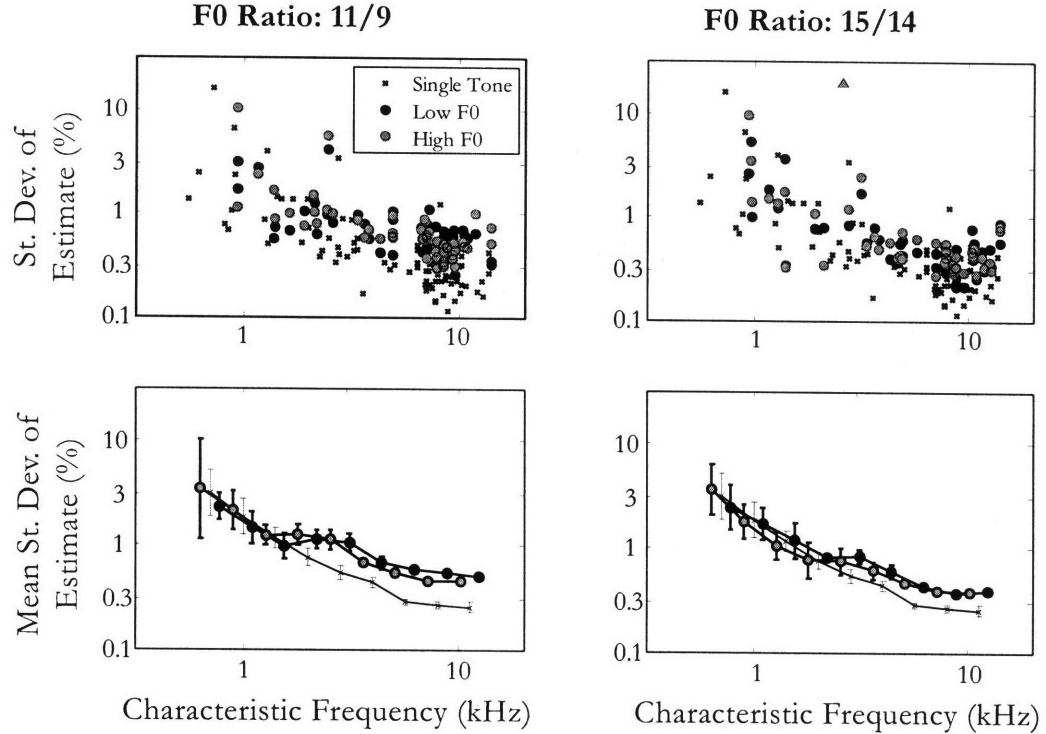


Figure 4. Precision (standard deviation) of the rate-based F0 estimates for double tones and of the BF_{CT} estimate from single-tone responses as a function of CF for the AN fiber population. Left and right panels show double-tone results for F0 ratios of 11/9 and 15/14, respectively; the single-tone BF_{CT} results are shown on both sides. Upper panels show errors for individual AN fibers; lower panels show moving window averages of the log-transformed absolute errors, using one-octave wide CF bins with 50% overlap. Results for the low (dark symbols) and high (light symbols) tone in each double complex, as well as for BF_{CT} estimation (crosses) from single complex tones are shown separately. Error bars represent ± 1 standard error of the mean.

To quantify the precision of F0 estimates, we used the standard deviations of the fitted F0 parameters calculated from the residuals (see *Data Analysis*). Figure 4 shows the standard deviation of the F0 estimates for double complex tones as a function of CF. The left and right panels show results for the 11/9 and 15/14 F0 ratio, respectively. The upper and lower panels show data for individual fibers and moving averages of the log-transformed data, respectively. Also included in Fig. 4 are the standard deviations of the BF_{CT} estimates from single-tone responses. Since the CF_{virt} of the virtual fiber array are directly proportional to BF_{CT} (Equation 1), the reliability of F0 estimation for double complex tones ultimately depends on the precision of the BF_{CT} estimate in Step 1 of the estimation procedure.

The standard deviations of the F0 and BF_{CT} estimates decrease gradually with increasing CF, again consistent with the improvement in relative cochlear frequency selectivity. For CFs above 2 kHz, the BF_{CT} estimates are more precise than the F0 estimates, but the two estimates are comparable at low CFs. A two-way ANOVA on the single- and double-complex tone data for the 11/9 F0 ratio, using CF (0.5-2 kHz, 2-4 kHz, 4-8 kHz, and 8-16 kHz) and tone type (single, low-F0 of double, high-F0 of double) as factors revealed significant main effects of both factors (CF: $F_{(3, 188)}=50.82$, $p<0.001$; tone type: $F_{(2, 188)}=17.44$; $p<0.001$). The precision was better (standard deviation lower) for higher CFs, as well as for single complex tones compared to either the lower or higher tone of the double complex, which were statistically equivalent (Tukey-Kramer). These differences between single and double tones were significant only for CFs above 4 kHz, which resulted in a CF \times tone type interaction ($F_{(6, 188)}=2.52$, $p=0.023$). The same analysis using the data for the 15/14 F0 ratio gave significant main effects of CF and tone type, but no interaction (CF: $F_{(3, 175)}=63.61$, $p<0.001$; tone type: $F_{(2, 175)}=6.74$, $p<0.002$).

Because the reliability of F0 estimation for double complex tones depends on the BF_{CT} parameter fitted to the rate responses to single complex tones, it is of interest to compare BF_{CT} to the CF obtained from pure-tone tuning curves (no figure). Since most measurements were made at low stimulus levels, the two parameters might be expected to be close, although in a nonlinear system they do not have to be identical. As expected, fibers with higher CFs tended to have smaller differences between BF_{CT} and CF: the median absolute differences were 9.3%, 3.2%, and 2.8% in the CF ranges below 2 kHz, between 2 and 5 kHz, and above 5 kHz, respectively. The small differences between tuning curve CF and BF_{CT} above 2 kHz suggests that the procedure for fitting the model to the rate responses is reliable. The larger discrepancies for low-CF fibers are consistent with previous results with single complex tones (Cedolin and Delgutte 2005). In this range, the relative sharpness of cochlear tuning (expressed as Q) is too poor to resolve the harmonics, leading to difficulties in fitting the model.

In summary, both F0s of a double complex tone can be accurately and reliably estimated from rate responses of AN fibers for CFs above 2-3 kHz, where the relative frequency resolution of the cochlea is the best. For CFs below 2-3 kHz, F0 estimation for double tones appears to be limited by the ability to fit the peripheral model to the single complex tone responses in Step 1 of the estimation procedure because (1) BF_{CT} and pure-tone CF could differ appreciably in this CF range, and (2) the precision of the F0 estimates for double complex tones was comparable to that of the BF_{CT} estimate (Fig. 4). Unexpectedly, for CFs above 2-3 kHz, the F0 estimation procedure was equally effective for both F0 ratios, even though the two-tone mixture contained resolved partials for the 11/9 ratio but not for the 15/14 ratio (Fig. 2). This suggests that our template matching procedure can make use of information contained in the shapes and widths of the peaks and valleys of the rate profile as well as their location along the tonotopic axis.

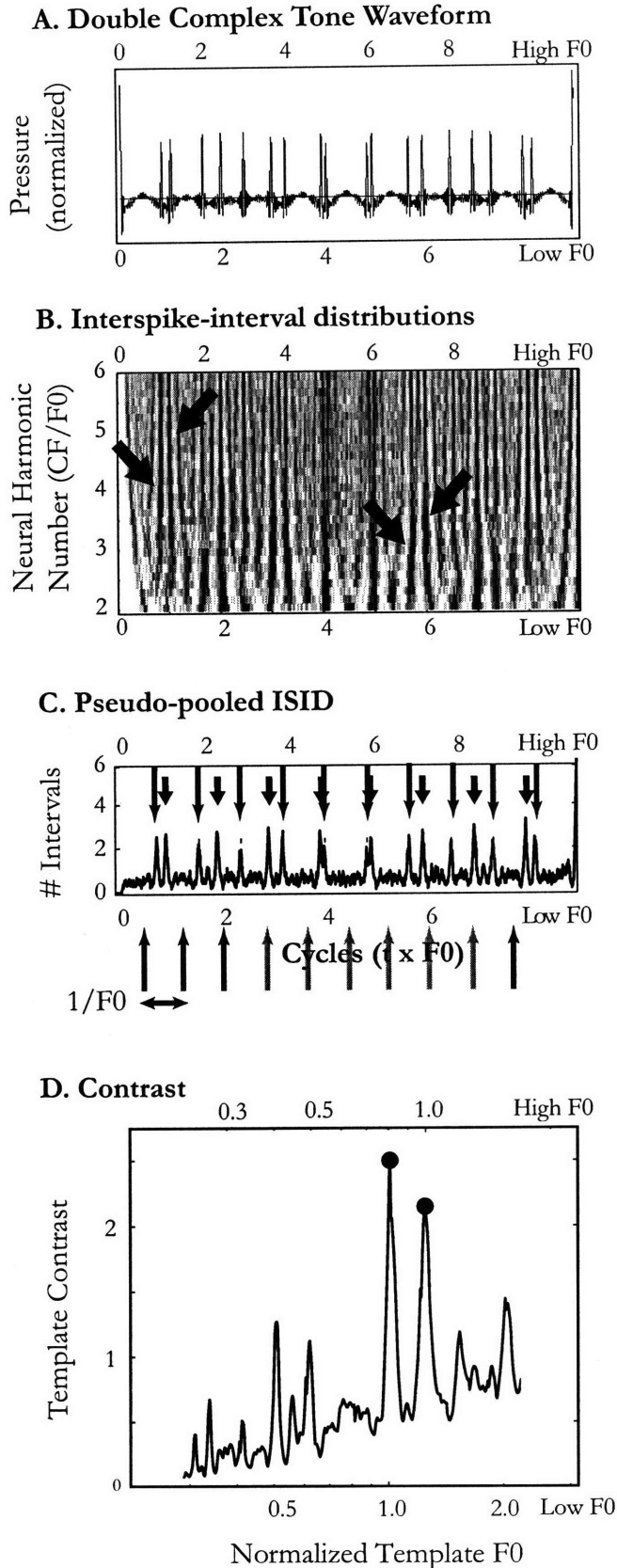


Figure 5. Method for estimating the F0s of double complex tones from the interspike-interval distributions (ISIDs) of an auditory-nerve fiber (CF = 816 Hz). A: One period of the waveform of double complex tone with F0 ratio of 11/9. B: ISIDs measured in response to the double complex tone as a function of probe F0. Gray scale represents the number of intervals in each time bin. ISIDs are plotted on normalized time scales in units of number of cycles of either tone (lower scale on panel: low F0 tone, upper scale on panel: high F0 tone). This scaling leads to vertical ridges in the ISID at the periods of the two complex tones and their multiples (block arrows). C: Pseudo-pooled ISIDs obtained by summing the time-normalized ISIDs over all probe F0s. Wide and thin downward arrows show the periods of the lower and the upper tone, respectively. Upward arrows at the bottom point to the time bins at which a periodic template with period $1/F0$ tallies interval counts from the pseudo-pooled ISID. These tallies are then normalized by the mean number of intervals per bin in the pseudo-pooled ISID to obtain the *template contrast*, and the operation is repeated for a wide range of template F0s. D: Template contrast as a function of normalized template F0 for the pseudo-pooled ISID in C. The template F0 is normalized to the F0 of the lower tone at the bottom of the panel, and to the F0 of the higher tone at the top of the panel. The F0 estimates for the double tone are the locations of the two largest peaks in the contrast function.

C.3 Interspike-interval analysis

Figure 5 illustrates our method for estimating the effective F0s of double complex tones using interspike interval distributions from a medium spontaneous-rate fiber (CF=816 Hz, SR=2.6 spikes/s, threshold=37 dB SPL). Panel A shows one period of the double complex tone waveform (F0 ratio: 11/9), which contains 9 and 11 periods of the lower and higher tones, respectively. Since all the components are of equal amplitude and in cosine phase, the waveform is mathematically equivalent to its autocorrelation. Panel B shows all-order interspike-interval distributions (ISIDs) in response to this stimulus as a function of probe F0. The ISIs are plotted in units of normalized time (cycles of the lower tone), and the vertical scale is the neural harmonic number $CF/F0_{\text{low}}$, which varied from approximately 1.5 to 5.5; $F0_{\text{high}}$ varied proportionately to maintain the F0 ratio at 11/9. For single complex tones (not shown) with F0s within the range of phase locking, ISIDs of responding AN fibers show peaks at the period of F0 and its multiples, and these peaks are reinforced in the pooled ISID obtained by summing single-fiber ISIDs over a wide range of CFs (Cariani and Delgutte 1996a,b). In Fig. 5, the ISIDs for double tones show clear vertical ridges at the normalized periods of both F0s and their multiples (block arrows in the figure). The time-normalized ISIDs were summed across the vertical axis (probe F0) to yield a “pseudo-pooled” ISID (panel C). The pseudo-pooled ISID also displays strong peaks at the periods of both complex tones and their multiples. Because ISIs are plotted in units of normalized time (cycles), ISID peaks occur at the same locations along the horizontal axis for all probe F0s and are therefore reinforced in the pseudo-pooled ISID.

Effective F0s were estimated from the pseudo-pooled ISID using a periodic template (or sieve), which takes the mean interval count in histogram bins at integer multiples of the template period (in units of cycles, rather than absolute time). This value is then divided by the mean of all histogram bins to obtain a template *contrast*. By repeating this procedure for a wide range of template periods, a contrast function as in Fig. 5D is obtained. The horizontal axis is in units of normalized template F0 so that, a peak corresponding to the lower F0 will always be near 1, while the peak corresponding to the higher F0 will be near the F0 ratio (11/9 or 15/14). In this case, the two largest peaks in the contrast function do occur near 1 and 11/9. The effective F0s estimated from the peak template contrasts are very accurate, with errors of 0.53% and 0.36% for the lower and higher F0, respectively.

While in this case the two largest peaks in the template contrast function occurred at the two F0s present in the double-tone stimulus, in some cases the second largest peak occurred at a harmonic or subharmonic of the F0 of the largest peak rather than at the second F0. Our algorithm therefore uses the second largest peak that is not harmonically related to the largest peak to estimate the second F0. This constraint is psychophysically reasonable, as two tones with harmonically-related F0s would likely be heard as a single tone. On the other hand, a method that directly cancels the first tone from the input spike train (de Cheveigné 1993) could in principle avoid this constraint.

C.4 Pitch estimation from interspike intervals works best for lower CFs

We analyzed temporal responses to 85 single complex tones (from 73 fibers) and 155 double complex tones (from 89 fibers). Unlike the rate-based F0 estimation method, the interval-based method does not require an auxiliary measurement of single-tone responses to obtain F0 estimates for double complex tones, and also provides an independent F0 estimate for single complex tones when responses are available. We first present results on estimation errors and

contrast together for all spontaneous-rate groups, and then analyze the effect of spontaneous rate (SR) on contrast.

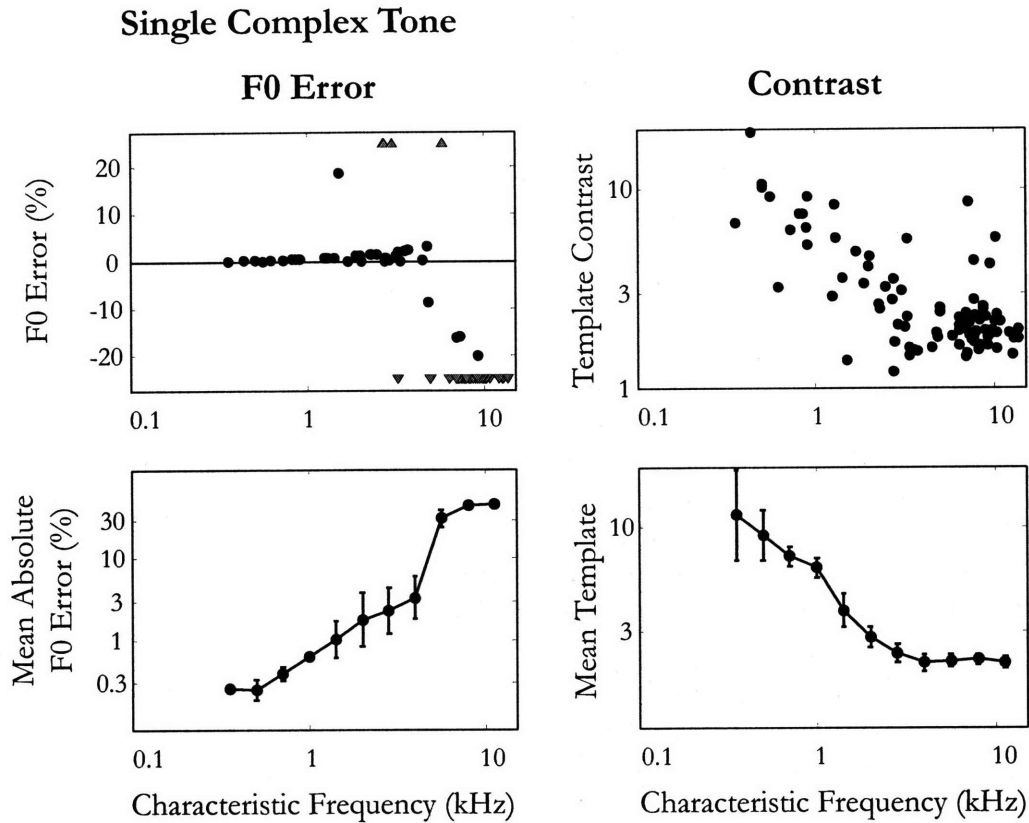


Figure 6. Interval-based F0 estimation errors (left) and maximum template contrasts (right) for single complex tones as a function of CF for the AN fiber population. Upper panels show errors for individual AN fibers; lower panels show moving window averages of the log-transformed data, using one-octave wide CF bins with 50% overlap. Error bars represent ± 1 standard error of the mean. In the top left panel, triangles show data that lie out of the vertical range.

Figure 6 shows percent F0 estimation errors (left) and the associated contrast values (right) obtained from responses to single complex tones. The upper and lower panels show data for individual fibers and moving averages of the data, respectively. The mean F0 errors increase gradually from about 0.3% at CFs below 500 Hz to about 3% around 4 kHz; over the same CF range, contrast declines gradually from about 10 to about 2. For CFs above 4 kHz, F0 estimation errors increase faster, while contrast remains at a floor level of about 2. These patterns of F0 errors and contrast values obtained from pseudo-pooled ISIDs are very similar to those obtained with true pooled ISIDs by Cedolin and Delgutte (2005), thereby lending support to the use of scaling invariance.

Figure 7 shows the percent F0 estimation errors for double complex tones with 11/9 and 15/14 F0 ratios in left and right panels, respectively. The moving average data shown in the lower panels also include the single-tone data for comparison purposes. As in the single-tone case, the mean absolute F0 errors generally increase with CF, gradually at low CFs but more steeply above 2-3 kHz. Compared to the single-tone data, the double-tone data show more variability. The data for the 11/9 F0 ratio and the single-tone data were together submitted to a

two-way ANOVA with CF (half-octave bins from 500 Hz to 16 kHz, except 125-500 Hz and 1-2 kHz due to sparser sampling these CF regions) and tone type (single, low-F0 of double, high-F0 of double) as factors. The main effects of both CF ($F(9, 258)=35.36, p<0.001$) and tone type ($F(2, 258)=5.84, p=0.0034$) were significant, while their interaction was not. The errors for the higher F0 of the double complex tone were significantly higher than either errors from the lower F0 or those from the single complex tone; the last two were not significantly different from each other. Thus, estimation of the lower F0 of a double tone is just as accurate as for complex tones presented in isolation, and breaks down at about the same CF (4 kHz), while the higher F0 of the double tone is estimated less accurately. The same analysis applied to the data for the 15/14 F0 ratio gave the same results as far as significance of the main effects (CF: $F(9, 220)=35.71, p<0.001$; tone type: $F(2, 220)=3.67, p=0.021$). However, in this case both constituents of the double complex tone had statistically equivalent estimation errors that were significantly higher than F0 errors for the single complex tone. Finally, a three-way ANOVA using the double-tone data for both F0 ratios (but excluding the single-tone data) revealed significant main effects of CF ($F(9,270)=42.33, p<0.001$) and tone height (high vs. low F0, $F(1,270)=4.31, p=0.039$), but no effect of F0 ratio. Putting these analyses together, there is a tendency for the higher F0 of a double complex tone to be estimated less accurately from ISIDs than the lower F0, and a weaker tendency for the lower F0 of a double complex tone to be estimated less accurately than the F0 of a single complex tone. However, the F0 separation between the two tones is not a major factor.

Figure 8 shows contrast values for the same conditions as for the F0 estimation errors in Fig. 7, and using a similar organization of the panels. Contrast decreases gradually with increasing CF for all stimuli, and the single-tone contrast is clearly greater than the double-tone contrasts, particularly at lower CFs (<3 kHz). A two-way ANOVA on the single-tone data and the double-tone data for the 11/9 F0 ratio reveals significant main effects of CF ($F(2, 258)=74.16, p<0.001$) and tone type ($F(9, 258)=18.17, p<0.001$), as well as a significant interaction ($F(18, 258)=5.23, p<0.001$). A Tukey-Kramer post-hoc analysis showed that contrast for the single complex tone is significantly higher than contrast for either of the constituents of the double complex tone, which have statistically equivalent contrast values. The interaction indicates that the difference between single-tone and double-tone contrasts is greater at low CFs than at high CFs. The same main effects and interactions were found when the data for the 15/14 F0 ratio were analyzed, with all p -values also <0.001. Finally, a three-way ANOVA including double tone data for both F0 ratios (but excluding single complex tones) showed significant main effects of CF and tone height (both $p<0.001$), but no effect of F0 ratio and no interaction.

In summary, F0s of both single and double complex tones can be accurately estimated (errors < 3%) from pseudo-pooled interspike distributions for CFs below 2-4 kHz. Since the range of probe F0s in our stimuli is proportional to the CF, this means that F0 estimation based on interspike intervals works best for lower F0s. Our correlate of pitch strength, the template contrast, is at least two times smaller for double complex tones than for a single complex tone with the same F0, reflecting the competition for phase locked spikes (a.k.a. “synchrony suppression” – Greenwood 1986) between the two F0s of a double complex tone. This pronounced decrease in contrast does not translate into a proportional increase in F0 estimation errors for double complex tones compared to single complex tones, and in many cases the error for the lower F0 of a double tone was comparable to the single-tone error (Fig. 7). As was the case for F0 estimation from rate responses, the ratio of constituent F0s in a double complex tone appears to have little effect on accuracy of F0 estimation and contrast.

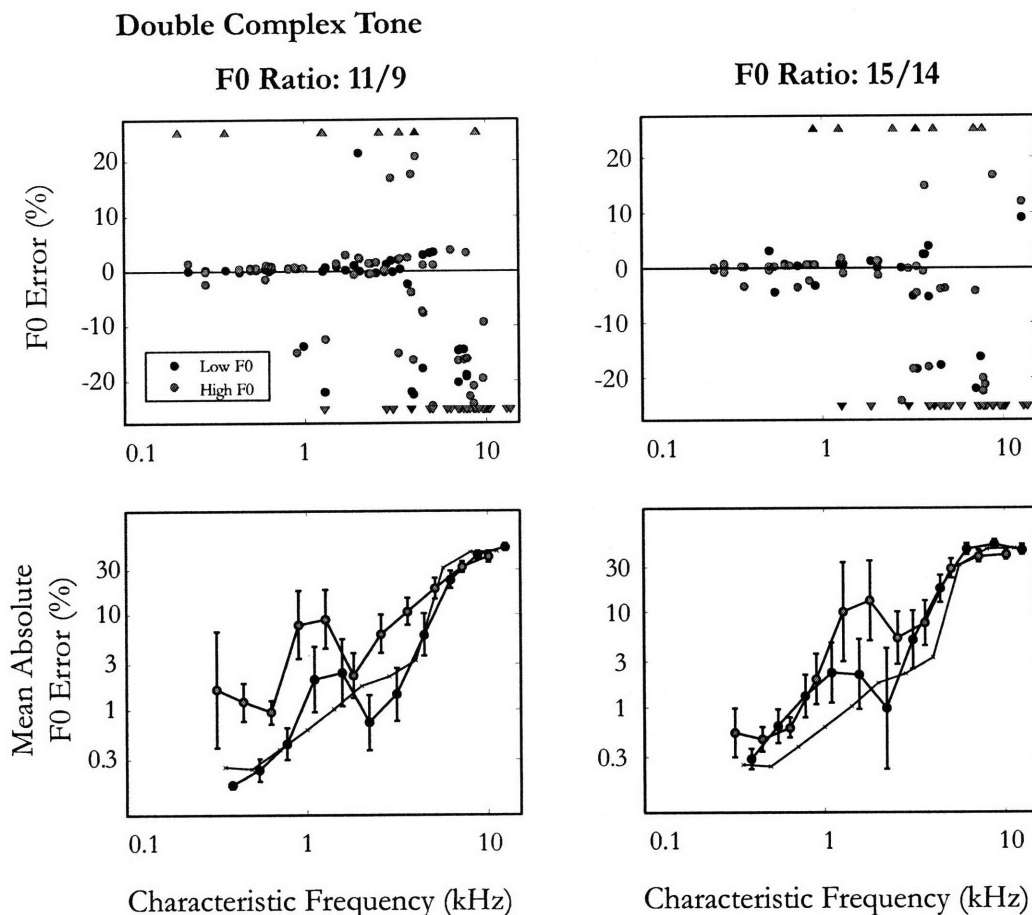


Figure 7. Percent F0 estimation errors from pseudo-pooled ISIDs for double complex tones as a function of CF for the AN fiber population. Left and right panels show results for F0 ratios of 11/9 and 15/14, respectively. Upper panels show errors for individual AN fibers; lower panels show moving window averages of the log-transformed absolute errors, using one-octave wide CF bins with 50% overlap. For reference, the lower panels also reproduce the mean F0 estimation errors for single complex tones (crosses) from Fig. 6. Error bars represent ± 1 standard error of the mean. In the top panels, triangles show data that lie out of the vertical range.

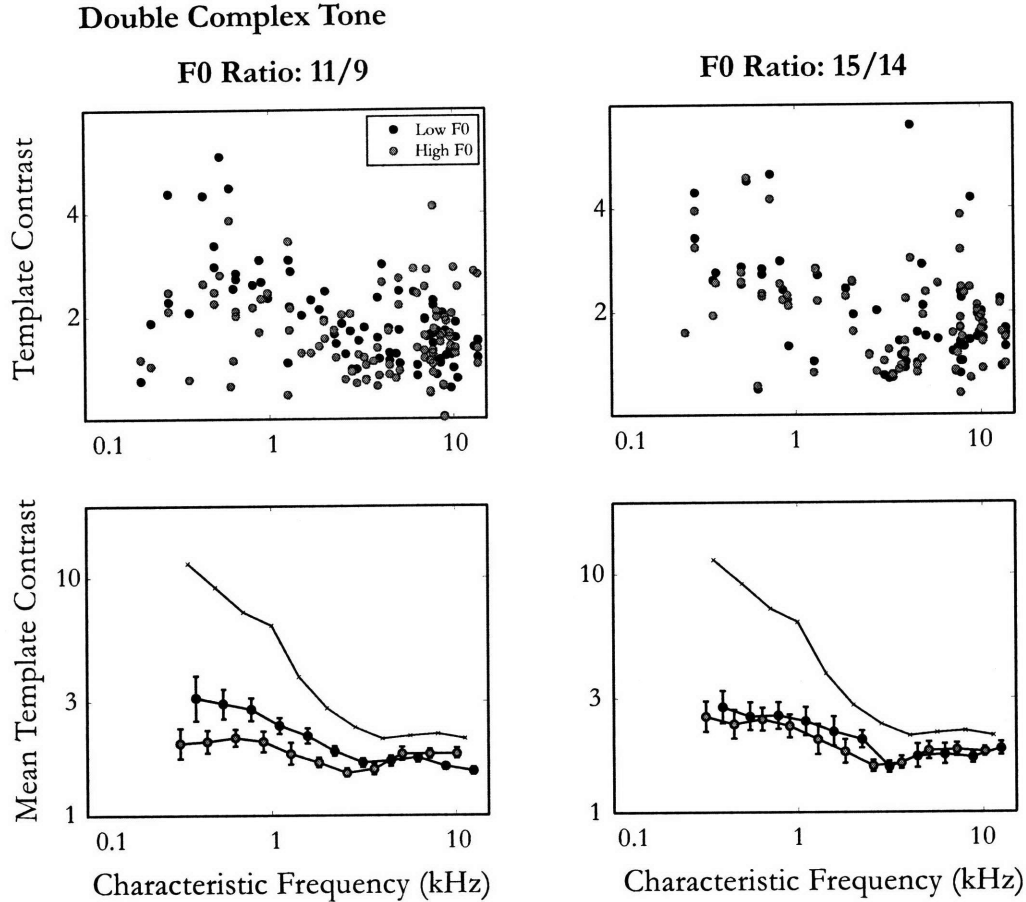


Figure 8. Neural pitch strength (template contrast), from pseudo-pooled ISIDs for double complex tones as a function of CF for the AN fiber population. Left and right panels show results for F0 ratios of 11/9 and 15/14, respectively. Upper panels show errors for individual AN fibers; lower panels show moving window averages of the log-transformed absolute errors, using one-octave wide CF bins with 50% overlap. For reference, the lower panels also reproduce the template contrasts for single complex tones (crosses) from Fig. 6. Error bars represent ± 1 standard error of the mean.

C.5 Effect of spontaneous rate on pitch estimation from interspike intervals

Fig. 9A and 9B show interspike interval distributions (ISIDs) in response to a single complex tone for two fibers with similar CFs (500-600 Hz) and similar pure-tone thresholds (10-15 dB SPL), but different spontaneous rates; the probe F0s were also identical in the two panels. The fiber shown in Fig. 9A had a high SR (66 spikes/s), while the fiber in Fig. 9B had a medium SR (3 spikes/s). For the medium-SR fiber, the vast majority of intervals occur at the stimulus period and its multiples, creating sharp vertical ridges in the ISID that are further reinforced in the pseudo-pooled ISID below. In contrast, the high-SR fiber shows a relatively large fraction of ISIs that do not correspond to the periodicity in the stimulus, leading to smaller period peaks in the pseudo-pooled ISID relative to the medium-SR fiber, and more background activity between the peaks. These smaller peaks result in lower contrast at the effective F0 for the high-SR fiber vs. the medium-SR fiber. This difference may be explained by considering that, with low/medium SR fibers, essentially all spikes are phase locked to the stimulus, whereas in high-SR fibers, there are “spontaneous spikes” that occur randomly in time, as well as phase-locked

spikes. These spontaneous spikes lower contrast by increasing the background rate in the denominator of the expression for contrast.

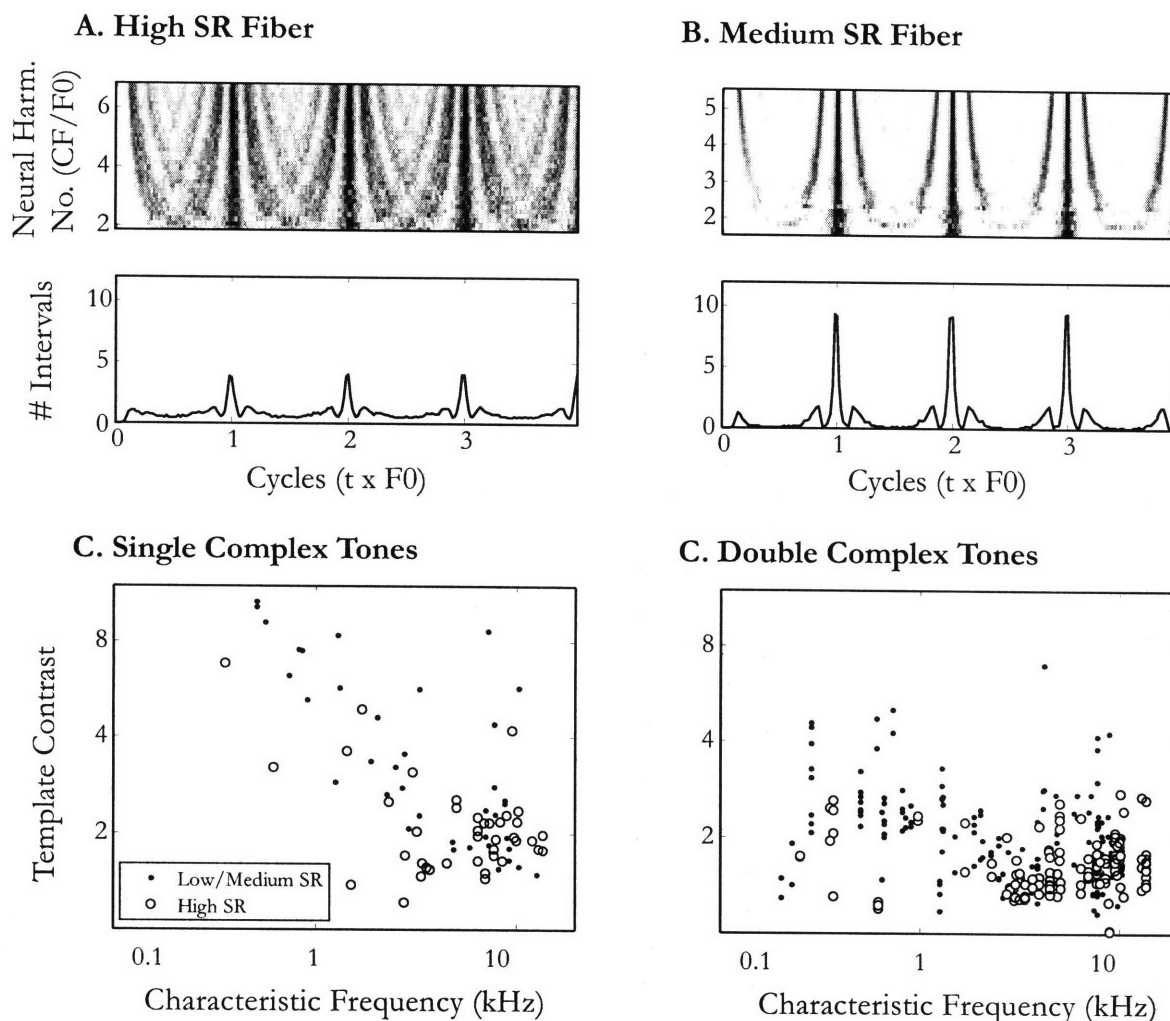


Figure 9. Effect of spontaneous rate on neural template contrast (a correlate of pitch strength). A-B: Interspike interval distribution as a function of probe F0 (top) and pseudo-pooled interval distribution (bottom) in response to a single complex tone for a high-SR fiber (A, 66 spikes/s) and a medium-SR fiber (B, 3 spikes/s). Both fibers have similar thresholds (10-15 dB SPL) and CF (500-600 Hz). C-D: Template contrast at the estimated F0 as a function of CF for the AN fiber population in response to single (C) and double complex tones (D). The data shown in D include contrasts values for both F0 ratios and both the lower and higher F0s of each double tone.

Fig. 9C and 9D show template contrast at the estimated F0 as a function of CF for single complex tones and double complex tones, respectively. High-SR (> 18.5 spikes/s) and low/medium SR fibers are shown by different symbols. For double complex tones, data for both F0 ratios and both the lower and higher F0 are included. On average, contrast for low/medium-SR fibers is higher than that for high-SR fibers, and this observation is confirmed by a two-way ANOVAs with CF (bins: 0.25-1 kHz, 1-2.5 kHz, 2.5-5 kHz) and SR group as factors. For single complex tones, the effects of both CF ($F(2,39)=19.5$, $p<0.001$) and SR ($F(1,39)=11.24$, $p=0.002$) were significant, but their interaction was not ($F(2,39)=0.19$, $p=0.83$). The same effects were

found in a separate ANOVA for double complex tones (CF: $F(3,169)=8.51$, $p<0.001$; SR: $F(1,169)=14.9$, $p<0.001$; CF \times SR: $F(3,169)=2.27$, $p=0.082$).

Although SR has a significant effect on contrast (a correlate of pitch strength), it does not seem to have an effect of the accuracy of F0 estimation from interspike interval distributions. Specifically, a two way ANOVA on pitch estimation errors for single tones revealed a significant main effect of CF ($F(2, 39)=3.62$, $p=0.038$), but not of SR ($F(1,39)=0.05$, $p=0.82$), and no significant interaction. Results were similar in a separate ANOVA for double complex tones (CF: $F(3,169)=13.7$, $p<0.001$, SR: $F(1,169)=0.02$, $p=0.88$, CF \times SR: $F(3,169)=0.27$, $p=0.84$). The lack of an SR effect on the accuracy of F0 estimation, despite an effect on contrast, suggests that accurate F0 estimation may be less dependent on the heights of the period peaks in the pooled ISID than on their widths, which are similar for both SR groups (compare Fig. 9A and 9B).

D Discussion

We measured responses of auditory nerve (AN) fibers to pairs of concurrent harmonic complex tones, and quantitatively assessed the representation of the two F0s in both rate responses and interspike interval distributions. We relied on scaling invariance in cochlear mechanics to infer spatiotemporal response patterns to an “effective” stimulus from a series of measurements made in a single fiber as a function of “probe” F0. A template matching procedure, in which the templates were synthesized by a peripheral auditory model, was used to estimate the effective F0s of double complex tones from pseudo rate-place profiles. This rate-place representation was accurate (mean estimation errors <2-3%) for fibers with CFs above about 3 kHz. Periodic templates were used to estimate the effective F0s of single and double complex tones from pseudo-pooled interspike interval distributions. This temporal representation was accurate for CFs below about 3 kHz.

While we are reporting the range of effectiveness of the two pitch representations in terms of CF because F0 estimates were obtained for each fiber, it is useful to convert these ranges into units of F0 in order to compare our results with those of other psychophysical and physiological studies. To do so, we note that the range of probe F0s was chosen for each fiber so that the neural harmonic number $CF/F0$ would vary from 1.5 to 5.5, with a geometric mean of 3.3. This means that the mean F0 of the lower tone in our stimulus set (the effective F0) was approximately equal to the fiber CF divided by 3.3. Using this scaling factor, the rate-place representation is expected to be effective for F0s above about 900 Hz ($3000/3.3$), while the interval-based representation should be effective for F0s below 900 Hz. Together, the two representations cover a wide range of F0s, including the 500-1000 Hz range most important for cat vocalizations and also the range of human voice (80-400 Hz).

These F0 limits should not be interpreted too literally because they depend on arbitrary criteria such as the 2-3% error ceiling for accurate estimation, the range of probe F0s chosen for each fiber, and also on the signal-to-noise ratio of the recordings, which in turn depends on the number of stimulus presentations and the duration of each stimulus. Also, we report average performance across fibers, while the brain may use the best-performing fibers instead. Thus, measures of accuracy of our F0 estimates are only relative, and most useful for comparing between stimulus conditions (e.g. single vs. double tone, or low vs. high F0) and between putative neural codes (e.g. rate-place vs. intervals).

D.1 Scaling invariance in cochlear mechanics

In a perfectly scaling invariant cochlea, the response to frequency f at the cochlear location tuned to CF is only dependent on the ratio f/CF (Zweig 1976). Therefore, the magnitude and phase of the response to a pure tone of frequency f at the location tuned to CF_0 are equal to the magnitude and the phase of the response to frequency f_0 at the cochlear location tuned to βCF_0 , where $\beta = f_0/f$. This means that the waveforms of the two responses are the same except for a scaling in time by a factor $1/\beta$. By varying f (and therefore β) and measuring responses at a fixed location tuned to CF_0 , one can therefore infer the response to a stimulus of fixed frequency f_0 as a function of cochlear location. The same reasoning applies to a harmonic complex tone in which all the components are multiples of a given F0, and to a double complex tone with a given F0 ratio.

There are two issues with the scaling invariance assumption. First, actual cochleae are only approximately scaling invariant, the approximation being fairly good locally but not over the

whole length of the cochlea (Shera and Guinan 2003; van der Heijden et al. 2005). Specifically, while scaling invariance requires filters with constant quality factor Q (the ratio of CF to bandwidth), Q is known to increase with increasing CF (Kiang et al. 1965; Liberman 1978). This deviation from scaling invariance is of primary concern for the rate-place representation. The second issue is that the time constants of cochlear processing in hair cells and their afferent synapses appear to be largely constant along the length of the cochlea, and therefore are not scaling invariant. These non-scalable parameters include the upper frequency limit of phase locking, neural refractory periods, and adaptation time constants. Yet by analyzing AN response patterns on a normalized time scale ($t \times F_0$), we effectively assume these time constants and cutoff frequencies do scale with CF. This issue is of primary concern for the temporal representation. We address the two issues in turn.

D.1.i Deviations from scaling invariance in cochlear mechanics

While scaling invariance assumes constant- Q tuning throughout the length of the cochlea, Q is actually an increasing function of CF. A power law with an exponent of 0.37 provides a good fit to AN data in cat measured with either pure or complex tones (Shera et al. 2002; Shera and Guinan 2003; Cedolin and Delgutte 2005); perfect scaling invariance would lead to an exponent of 0. Our stimuli were designed so that the ratio CF/F_0 varied from 1.5 to 5.5, a range of 1.87 octaves, which corresponds to about 25% of the length of the cochlea in cat for more basal locations (Liberman 1982). Using the geometric mean of this range (neural harmonic number 3.3) as a reference, a fiber with CF at Harmonic 1.5 would have a Q that is $(1.5/3.3)^{0.37}$, i.e. 25% smaller than that of the fiber with CF at harmonic 3.3 actually used to make the measurements. Similarly, a fiber with CF at Harmonic 5.5 would have a Q that is $(5.5/3.3)^{0.37}$, i.e. 21% greater than that of the fiber with CF at harmonic 3.3. Thus, deviations from scaling invariance lead to discrepancies in sharpness of tuning of ± 21 -25% over the range of probe F_0 s of our stimuli. Somewhat smaller deviations (± 9 -13%) are obtained if we use AN fiber bandwidths measured by reverse correlation with broadband noise (Carney and Yin 1988), rather than the power law. Although these deviations are not insignificant, they are not likely to greatly alter the effectiveness of F_0 estimation from rate responses. This conclusion is supported by the observation that the range of effective F_0 s over which we could reliably estimate BF_{CT} from responses to single complex tones corresponds well to the F_0 range previously obtained by measuring responses to a single complex tone as a function of CF for a sample of AN fibers (Cedolin and Delgutte 2005). Specifically, our BF_{CT} estimates were reliable (standard deviation $< 1\%$) for CFs above 1.5 kHz, which corresponds to F_0 s above 450 Hz, while Cedolin and Delgutte (2005) could reliably estimate pitch from rate-place profiles for F_0 s above 400-500 Hz.

In summary, the deviations in cochlear bandwidths from perfect scaling invariance are likely to have moderate effects on F_0 estimates from rate-place representations. Because pooled autocorrelations work equally well with resolved and unresolved harmonics, and are not very sensitive to the widths of the cochlear filters (Meddis and Hewitt 1991; Cariani and Delgutte 1996a; Carlyon 1998), the effect on interval-based representations of pitch are expected to be even smaller.

D.1.ii Effect on phase locking

The ability of AN fibers to phase lock to the fine structure of harmonic complex tones is known to have a major effect on the accuracy of pitch estimates based on interspike interval distributions (Meddis and Hewitt 1991; Cariani and Delgutte 1996a). The strength of phase

locking in AN fibers (as measured by the synchronization index, a.k.a. vector strength) drops rapidly with increasing stimulus frequency above 1 kHz, until phase locking is hard to detect above 4-5 kHz (Johnson 1980). Because this cutoff frequency does not depend strongly on CF, phase locking properties are not scaling invariant. As a result, the dependence of phase locking on normalized frequency CF/F_0 may differ substantially between a true cochlea and our approximation based on scaling invariance, particularly for tones having harmonics near 2-3 kHz, where synchrony drops rapidly.

To understand this effect, we first assume for simplicity that every harmonic over the range 2-5 is resolved, so that the response at each cochlear location resembles a sinusoid at the harmonic frequency closest to the CF; we call this the “dominant frequency”. For the response to a single or double complex tone as a function of cochlear place, the dominant frequency increases monotonically with CF, so that synchrony falls monotonically with normalized frequency CF/F_0 . In contrast, using measurements made at a fixed CF as a function of probe F_0 by assuming scaling invariance, the dominant frequency never deviates much from the CF, so that synchrony is nearly constant as a function of CF/F_0 . The net result is that the scaling invariance approximation acts like a highpass filter: low-order harmonics (relative to the mean neural harmonic number 3.3) are attenuated, while high-order harmonics are boosted. A similar reasoning applies to the case of unresolved harmonics if the dominant frequency is now the center of gravity of a group of harmonics that pass through the auditory filter.

To quantitatively assess this highpass filtering effect, we ran simulations using a 4th-order Butterworth lowpass filter with a 3-dB cutoff frequency of 2570 Hz to characterize the dependence of synchrony on frequency. This filter gave an excellent fit to the Johnson (1980) data describing synchrony as a function of pure-tone frequency. As expected, the magnitude of the highpass filtering effect due to the scaling invariance assumption depended strongly on CF, and reached a maximum of ± 10 dB near 3 kHz, the highest CF for which we could reliably estimate F_0 from ISIDs. Next, to simulate the effect of pooling the interspike interval distributions, we summed the squared synchronies to the dominant frequencies over the entire range of CF/F_0 (1.5 to 5.5) in our stimuli. (We squared the synchrony before summing, because the autocorrelation of a sinusoid is a sinusoid with an amplitude proportional to the square of the original signal’s amplitude.) For all CFs, the deviations in the pooled synchrony between the fixed-CF and the fixed- F_0 conditions never exceeded 1 dB. Thus, while the highpass filtering effect resulting from the scaling invariance assumption can have substantial effects on interspike interval distributions for individual probe F_0 s, the effect on the pseudo-pooled distribution obtained by summing across probe F_0 s is expected to be small.

This conclusion is supported by the observation that the upper frequency limit over which we could reliably estimate the F_0 of single complex tones from pseudo-pooled interval distributions corresponds well to the F_0 range previously found by Cedolin and Delgutte (2005) using true pooled distributions. Specifically, our F_0 estimates for single complex tones were reliable for CFs below 4 kHz (Fig. 6A), which corresponds to F_0 s below 1200 Hz, while Cedolin and Delgutte (2005) could estimate pitch from pooled distributions for F_0 s below 1300 Hz.

D.1.iii Effects on adaptation and refractoriness

Not only the upper frequency limit of phase locking, but also the time constants of short-term and rapid adaptation (Westerman and Smith 1984) and the neural refractory period are nearly independent of CF and therefore not scaling invariant. Our analysis discarded spikes during the initial 20 ms of each probe stimulus, when the effects of adaptation are strongest.

Also, rapid adaptation of firing rates was minimized in our stimulus paradigm because our probe stimuli had a long duration (520 ms), and consecutive probe F0s were presented in small increments without an intervening silence. Average firing rates were usually low ($< 70 - 140$ spikes/s, and always less than 200 spikes/s), so that effects of refractoriness were also minimized. The scaling of the refractory period is apparent in the interspike interval distributions of Fig. 5 and 9, where the duration of the time period devoid of intervals near the origin varies inversely with CF/F0. This scaling is not likely to have an effect on F0 estimation from pseudo-pooled interval distributions since the major peaks that contribute to template contrast occur for longer intervals. The effect of refractoriness on interval distributions could be avoided altogether by using shuffled autocorrelations (Louage et al. 2004), but we chose traditional autocorrelation histograms to facilitate comparison with previous work, especially that of Cedolin and Delgutte (2005). Overall, the lack of scaling invariance of adaptation and refractoriness is expected to have only small effects on F0 estimation from pooled interspike interval distributions for the stimulus conditions of our experiments.

D.2 Rate-place representation of pitch

We used a template-matching procedure to estimate both F0s of double complex tones from virtual rate-place patterns. As in previous work (Cedolin and Delgutte 2005), the templates were generated by a simple peripheral auditory model, making this a form of ‘analysis by synthesis’. The templates generated by the peripheral model had relatively wide peaks at harmonic frequencies, as determined by the frequency resolution of the peripheral model. These broad templates contrast with the narrow templates or sieves typically used in spectral models of pitch perception (Goldstein 1973; Duifhuis et al. 1982; Scheffers 1983; Terhardt 1974; but see Wightman 1973 for an exception). These wide peaks made the F0 estimation procedure sensitive to the widths and shapes of the peaks and valleys in the rate-place profiles as well as peak locations and were essential for achieving accurate F0 estimates for double tones with the 15/14 ratio (Fig. 2).

Using similar templates, Cedolin and Delgutte (2005) were able to reliably estimate the F0 of single complex tones from true rate-place profiles for F0s above 400-500 Hz. As we have pointed out, this fits well with the CF range over which we could estimate the BF_{CT} parameter of the peripheral model from single-tone responses using the scaling factor $CF \approx 3.3 F0$. However, the F0 range over which both F0s of double complex tones could be accurately estimated using rate-place information was more restricted, with a lower limit about an octave higher (900 Hz) than for single complex tones. This decrease in the F0 range of effectiveness of the rate-place representation likely reflects the more stringent requirements on harmonic resolvability for double complex tones. At lower F0s, cochlear filtering becomes too broad to clearly resolve individual harmonic frequencies, which results in weak modulation of firing rate across cochlear place. A double complex tone has twice as many harmonics in the same frequency range, which makes it more difficult for individual harmonics to be resolved in the mixture.

Remarkably, F0 estimation from pseudo rate-place profiles was equally effective for both F0 ratios, even though individual harmonics were only resolved in the mixture for the 11/9 ratio (Fig. 2B). For the 15/14 ratio, same-numbered harmonics of the two tones were not individually resolved, but formed pairs that were separated from other pairs (Fig. 2C). The model-generated templates were nevertheless able to estimate both F0s by making use of information about the widths and shapes of the peaks in the pseudo rate-place profiles. Thus, for our data, broad templates are more effective than narrow sieves that are only sensitive to peak locations.

Previous reports (e.g. Assmann and Summerfield 1990; Palmer 1990; de Cheveigné 1999a,b) have rejected spectral models for concurrent vowel identification because narrow harmonic sieves had trouble reliably estimating F0 when the mixture contained no resolved harmonics; these conclusions may need to be reexamined in light of the present results.

Although F0 estimation worked equally well for both F0 ratios, the higher tone was identified less accurately than the lower tone for the 11/9 ratio, but not for the 15/14 ratio, even though the harmonics of higher tone were better resolved for the 11/9 ratio. This unexpected result may be a consequence of how we chose the range of probe F0s for our stimuli. Specifically, the range was chosen for each fiber so that $CF/F0$ for the lower tone would range from 1.5 to 5.5. This range contains Harmonics 2-5 of the higher tone for the 15/14 ratio, but only Harmonics 2-4 for the 11/9 ratio. Thus, only 3 harmonics of the upper tone could be represented in pseudo rate-place profiles for the 11/9 ratio, vs. 4 harmonics for the 15/14 ratio, possibly explaining the decreased F0 estimation accuracy of the upper tone with the 11/9 ratio.

We found no effect of spontaneous rate (SR) on pitch estimation accuracy or precision from virtual rate-place profiles. We chose stimulus levels such that a complex tone with a fifth harmonic near the CF would yield firing rates halfway between spontaneous rate and maximum rate, ensuring that there would be a good amount of rate modulation with varying probe F0 (except at low CF, where tuning is broad), without saturating the fiber.

The physiological plausibility and generality of our F0 estimation procedure deserves comment. The idea of matching incoming sensory data to templates generated from internal models of sensory signal processing is widely accepted in studies of sensori-motor control (e.g. Merfeld et al. 1999; Todorov 2004; Guenther et al. 2006). Although we allowed both F0s of the double complex tone input to the model to vary freely, we did constrain the estimation algorithm by providing knowledge about the number of complex tones present (2) as well as their amplitude spectrum (equal-amplitude harmonics, missing fundamental). These constraints may be appropriate in the context of psychophysical experiments on F0 identification and discrimination for concurrent tones, where stimuli with equal-amplitude harmonics have typically been used (Beerends and Houtsma 1989; Carlyon 1996; Micheyl et al. 2006). Although here we assumed templates based on equal-amplitude harmonics, in general, the templates may incorporate all available *a priori* information about the stimulus and may also refine over time as the task is learned. For example, in concurrent-vowel experiments (e.g. Assmann and Paschall 1998), the templates might initially be generated assuming a spectrum envelope corresponding to an average vowel spectrum; as the task is learned, the templates might incorporate information about the spectral envelopes of specific vowels in the stimulus set. However, in real-life cocktail party situations, there will be more uncertainty about the stimulus than we have assumed in our estimation algorithm, so the F0 identification performance is likely to be overestimated. Psychophysical experiments that systematically manipulate stimulus uncertainty in F0 discrimination or identification tasks with concurrent complex tones would be helpful to assess the generality of the type of estimation algorithm dependent on a peripheral model we have proposed here.

D.3 Pitch representation in interspike intervals

Our results confirm and extend previous findings that both F0s of concurrent complex tones can be accurately estimated from pooled interspike interval distributions (Palmer 1990, 1992; de Cheveigné 1993; Tramo et al. 2000). Unlike the estimation method used with rate-place profiles, F0 estimation from interspike intervals did not require the use of a peripheral model, and was

directly based on the presence of modes at the periods of both F0s in the pooled interval distribution. This estimation method is a direct extension of that used previously for single complex tones (Cedolin and Delgutte 2005), which was itself a refinement of an earlier method based on the largest mode in the pooled distribution (Palmer 1990; Meddis and Hewitt 1991; Cariani and Delgutte 1996a,b). This single-mode method frequently leads to subharmonic errors when applied to neural data that are intrinsically noisy (Cedolin and Delgutte 2005). In the case of double tones, the single mode method has the additional disadvantage that a mode at the shortest period of one of the two F0s is sometimes absent (see Palmer 1990 for an example). Thus, the present results provide additional support for using periodic templates that look at series of harmonically related modes for F0 estimation rather than a single mode.

While previous physiological studies of concurrent complex tones were restricted to F0s in the range of human voice, we found that F0 estimation from interval distributions is effective up to about 900 Hz for double tones, and 1200 for single complex tones, consistent with an earlier study of single complex tones that did not rely on the scaling invariance assumption (Cedolin and Delgutte 2005). The decrease in estimation performance in the double-tone case is likely due to the decreased number of spikes that are phase locked to each tone of the double complex. Without significant changes in firing rate, each individual tone in a double complex tone can only generate half the number of phase-locked spikes in comparison to a single complex tone. This reduces the number of interspike-intervals that are associated with the periodicity of each tone, which in turn reduces the contrast of that tone's period with respect to background. This effect, which has been referred to as "synchrony suppression" (Greenwood 1986), is most clearly apparent in the template contrast, which was at least a factor of two smaller for double complex tones than for single complex tones (Fig. 8). The effect on accuracy of F0 estimation was less dramatic (Fig. 7), perhaps because accuracy depends more on the widths of the interval modes than on their heights, as long as the modes remain well above background. Apparently, the interval modes at F0 periods are essentially as sharp for double tones as for single complex tones.

We also found that low/medium SR fibers (<18.5 spikes/s) yielded higher template contrasts than high-SR fibers (>18.5 spikes/s), without a concomitant difference in mean F0 estimation errors. This was true for both single and double complex tones. This effect is consistent with the observation that synchrony to pure tones is somewhat higher for low/medium SR fibers than for high-SR fibers (Johnson 1980). Cariani and Delgutte (1996a) also reported that low/medium-SR fibers tended to produce larger template contrasts (which they called "pitch salience") than high-SR fibers for single complex tones having energy near the CF, but they did not quantify this dependence on SR. This difference in contrast can be explained by the greater fraction of non-phase-locked spikes in the high-SR fibers, which increases the denominator in the equation for contrast. However, the difference in contrast may have been exacerbated to some extent by the fact that we presented stimuli relative to fiber threshold, rather than at a fixed dB SPL. Since synchrony generally increases with level (Johnson 1980), presenting stimuli at a constant SPL rather than a constant dB re. threshold would tend to increase synchrony in high-SR fibers, which have lower threshold than low-SR fibers. However, this effect would only be significant at levels very close to threshold since synchrony saturates at lower levels than firing rate (Johnson 1980).

While F0 estimation worked equally well with both F0 ratios, template contrast for the higher F0 was smaller than the contrast for the lower F0 (Fig. 8). This effect was also observed to a lesser extent in F0 estimation errors (Fig. 7). A similar effect occurred for F0 estimation from

rate-place profiles, and we suggested that it might result from the higher tone containing a smaller number of harmonics (3 vs. 4) for the 11/9 ratio than for 15/14 ratio within the range of CF/F0 from 1.5 to 5.5. This explanation may also apply for the interval representation because the autocorrelation model predicts a stronger pitch with increasing number of harmonics (Meddis and Hewitt 1991; Cariani and Delgutte 1996b). An alternative explanation is that each harmonic of the higher tone for the 11/9 ratio is higher in frequency than the corresponding harmonic for the 15/14 ratio (because the lower F0 is the same for both ratios), and therefore produces weaker phase locking. However, this explanation is not supported by a quantitative analysis similar to that used in the discussion of scaling invariance. Specifically, we found that the pooled synchrony to the higher tone never differed by more than 1 dB between the two F0 ratios. We conclude that the difference in number of harmonics present is the most likely explanation for the lower contrast of the higher tone with the 11/9 ratio.

D.4 Relation to psychophysics

Psychophysical studies indicate that accurate pitch perception for both single and double complex tones depends strongly on the presence of resolved harmonics (Houtsma and Smurzynski 1990; Beerends and Houtsma 1989; Carlyon and Shackleton 1994; Carlyon 1996; Bernstein and Oxenham 2003, 2005; Micheyl et al. 2006). Generally, pitch discrimination and identification is poorer for stimuli that contain no resolved harmonics than for those that do. This is consistent with our results for the rate-place representation of pitch, which inherently requires resolved harmonics. Micheyl et al. (2006) introduced an important distinction between harmonics that are resolved in a constituent of a double complex tone presented by itself (before mixing) and harmonics that are resolved in the mixture of the two tones. They suggested that while resolved harmonics prior to mixing are necessary for accurate F0 discrimination, resolvability in the mixture may not be necessary (although it does help, see Micheyl et al. 2008). This is consistent with our finding that F0 estimation from rate-place information was equally accurate for both F0 ratios, even though the mixture typically contained resolved harmonics only for the 11/9 ratio (Fig. 2). We have argued that this ability results from the use of broad harmonic templates that are sensitive to the shapes and widths of peaks of activity associated with resolved harmonics rather than just the peak locations. At the same time, the F0 range over which rate-place estimation was effective was more restricted for double tones than for single complex tones, indicating that resolvability in the mixture is not wholly unimportant.

While the rate-place representation is consistent with the dependence of psychophysical results on harmonic resolvability, it poorly accounts for the range of F0s over which discrimination is accurate with double tones. Estimation from rate-place information was only effective for F0s above about 900 Hz, while psychophysical discrimination is accurate for F0s as low as 100 Hz so long as the tone target contains resolved harmonics. This discrepancy may partly result from species differences and would be alleviated if, as some biomechanical and psychophysical data suggest (Shera et al. 2002, 2007), cochlear frequency resolution is 2-3 times sharper in humans than in cat and other popular experimental animals (but see Ruggero and Temchin 2005 for a different opinion). Moreover, as we have pointed out, the 900 Hz lower F0 limit is a rough estimate dependent on our particular choice of stimuli and signal-to-noise ratio of our recordings. However, even if this F0 limit can be substantially lowered by taking into account species differences and using all the rate information available in the auditory nerve, F0 estimation from rate-place information remains a challenge for the lower F0s in the range of male voices. Even at higher F0s, estimation is constrained by the dynamic range problem (Sachs

and Young 1979; Colburn et al. 2003) which we avoided in the present study by using moderate stimulus levels. Our goal was to assess the pitch information available from resolved harmonics without confounding by the dynamic range issue.

Unlike the rate-place representation, F0 estimation from interspike interval distributions worked best for lower F0s (< 900 Hz). This temporal representation is limited by the phase-locking ability of AN fibers, which in turn is determined by the membrane time constant and the dynamics of basolateral Ca^{2+} and K^{+} channels in inner hair cells (Trussell 1999). Although there are some variations in phase locking ability across mammals (Palmer and Russell 1986), in the absence of direct physiological data, we shall assume that the frequency limit of phase locking in humans is similar to that in cat (Johnson 1980) and squirrel monkey (Rose et al. 1967). If so, the interspike interval representation is likely to be effective over the F0 range of male and female voices as well as the 100-400 Hz range used in previous human psychophysical studies of F0 identification and discrimination for double complex tones. However, accurate estimation from interspike intervals was achieved despite the lack of resolved harmonics in this F0 range (as assessed from the rate-place analysis for single complex tones), in contrast to psychophysical results pointing to the importance of resolved harmonics for accurate F0 discrimination in double complex tones (Beerends and Houtsma 1989; Carlyon 1996; Micheyl et al. 2006). Thus our results provide additional evidence that at least some of the temporal information present in AN interspike interval distributions may not be used at higher levels of the auditory pathway (Siebert 1970; Carlyon 1998; Kaernbach and Demany 1998; Oxenham et al. 2004; Cedolin and Delgutte 2005).

Various spatiotemporal models of pitch processing have been proposed in order to overcome difficulties with both purely spectral and purely temporal models (Shamma 1985; Bernstein and Oxenham 2005; de Cheveigné and Pressnitzer 2006; Cedolin and Delgutte 2007). These models depend on both precise phase locking and, at least implicitly, on harmonic resolvability. Our result that there is little or no overlap between the F0 ranges over which rate-place and interspike-interval representations are effective suggests that F0 discrimination and identification for double complex tones may prove to be a challenging task for some of these models.

For both rate-place and interspike-interval representations, our results do not indicate large differences in F0 estimation accuracy for the 15/14 vs. 11/9 F0 ratios. This finding contrasts with psychophysical results for concurrent vowels (Assmann and Summerfield 1990; Assmann and Paschall 1998), where listeners gave unimodal distribution of pitch matches for double vowels having a one semitone difference in F0 (which is close to our 15/14 ratio), while they gave bimodal distribution of matches for double vowels with a four semitones difference (close to our 11/9 ratio). Assmann and Paschall interpreted these results as indication that listeners perceived a single pitch with a one semitone difference but could hear two separate pitches with the 4 semitone difference. If this interpretation is correct, the failure of Assmann and Paschall's listeners to identify two pitches with F0 differences of one or two semitones may appear surprising given that musical competence routinely requires accurate identification of these intervals. However, musical intervals might be recognized without explicit identification of the individual pitches on which the interval is based (Burns and Campbell 1994). The listeners of Beerends and Houtsma (1989) were able to identify both pitches of double complex tones with F0 separations as low as two semitones, but the use of only 5 distinct F0 values and response choices in this study probably made the task much easier than the pitch masking task used by Assmann and Paschall (1998). A comparison of Assmann and Paschall's results with those of Micheyl et al. (2006) is made difficult by the fact that the complex tone stimuli of Micheyl et al.

contained no harmonics below 1200 Hz (and therefore no resolved harmonics for a 100 Hz F0), while the vowel stimuli of Assmann and Paschall had resolved harmonics in the first formant region. In addition, the prominent formant-related spectral peaks in the synthetic vowel stimuli used by Assmann and Paschall (1998) may have made the task of F0 identification more difficult by introducing possible confusions between pitch and timbre. Overall, psychophysical data on F0 identification for concurrent complex tones are too limited for an effect of F0 separation to be confidently assessed.

E Conclusions

We measured responses of auditory nerve fibers to pairs of concurrent harmonic complex tones, and quantitatively assessed the representation of the two F0s in both rate responses and interspike interval distributions.

1. We found that both F0s were accurately represented over a wide range of F0s. The interspike-interval representation was effective at low F0s (below 900 Hz), while the rate-place representations was effective at high F0s (above 900 Hz). If cochlear tuning is 2-3 sharper in humans than in cat (Shera et al. 2002) and auditory nerve phase locking is similar in the two species, then the lower limit for accurate rate-place representations of double complex tones would be 300-450 Hz, while the higher limit for accurate interspike-interval representations remain at 900 Hz. The two representations together would thus cover the entire human F0 range, with overlap in the region between about 400 and 900 Hz.
2. We used broad harmonic templates generated by a peripheral auditory model to estimate F0 from rate-place information. Consistent with psychophysical data (Micheyl et al. 2006), this estimation method was effective even when the two-tone mixture did not contain any resolved harmonic, so long as each constituent tone contained resolved harmonics prior to mixing. This property results from the sensitivity of the template matching method to the shapes and widths of peaks and valleys in rate-place profiles, rather than the locations of peaks alone. Previous conclusions that spectral models fail to account for F0 identification with concurrent complex tones, which were based on results with narrow harmonic templates or sieves, need to be reexamined in view of the present results.
3. Although the interspike interval representation supported accurate estimation of both F0s of double complex tones over the range of F0s most important for human voice, this accuracy was achieved despite the lack of resolved harmonics over much of this range (based on the rate-place analysis), in contrast to psychophysical results pointing to the importance of resolved harmonics for accurate pitch identification.
4. F0 estimation from both rate-place and interspike interval information was equally effective for F0 ratios of 15/14 (~1 semitone) and 11/9 (~4 semitones), in contrast to the psychophysical results of Assmann and Paschall (1998) with concurrent vowels, where listeners appeared to hear only one pitch for F0 separations below 4 semitones. Additional psychophysical data on F0 identification with double complex tones are needed to understand the reasons for this apparent discrepancy.

F Acknowledgements

I would like to thank my advisor, Bertrand Delgutte, for his guidance and support during this work. Although the experimental part was done in a relatively short time before my PhD thesis work started, the analysis and manuscript writing was done throughout my duration at the Eaton-Peabody lab with varying levels of intensity. Throughout this period I enjoyed working with and learning from Bertrand, and I feel fortunate to have had this additional opportunity beside my PhD thesis work.

Next I would like to give great thanks to Leonardo Cedolin, who was instrumental in getting me started with the experiments and data analysis, and helped with data collection also. Without him I could not have gotten the work done so fast, nor would it have been as enjoyable.

Lastly, I thank Connie Miller for expert surgery, and Christophe Micheyl for suggesting the possible implication of having resolved harmonics prior vs. after mixing of two harmonic complexes. NIH financial support made all of this possible.

G List of references

- Assmann PF and Paschall DD. Pitches of concurrent vowels. *J Acoust Soc Am* 103: 1150 – 1160, 1998.
- Beerends JG and Houtsma AJM. Pitch identification of simultaneous diotic and dichotic two-tone complexes. *J Acoust Soc Am* 86: 1835 – 1844, 1989.
- Bernstein JGW and Oxenham AJ. Pitch discrimination of diotic and dichotic tone complexes: Harmonic resolvability or harmonic number? *J Acoust Soc Am* 113: 3323 – 3334, 2003.
- Bernstein JGW and Oxenham AJ. An autocorrelation model with place dependence to account for the effect of harmonic number on fundamental frequency discrimination. *J Acoust Soc Am* 117: 3816 – 3831, 2005.
- Bregman AS. Auditory scene analysis: The perceptual organization of sound. Cambridge, MA: MIT Press, 1990.
- Burns EM and Campbell SL. Frequency and frequency-ratio resolution by possessors of absolute and relative pitch: Examples of categorical perception? *J Acoust Soc Am* 96: 2704 – 2719, 1994.
- Cariani PA and Delgutte B. Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. *J Neurophysiol* 76: 1698 – 1716, 1996a.
- Cariani PA and Delgutte B. Neural correlates of the pitch of complex tones. II. Pitch shift, pitch ambiguity, phase invariance, pitch circularity, and the dominance region for pitch. *J Neurophysiol* 76: 1716 – 1734, 1996b.
- Carlyon RP. Encoding the fundamental frequency of a complex tone in the presence of a spectrally overlapping masker. *J Acoust Soc Am* 99: 517 – 524, 1996.
- Carlyon RP. Comments on “A unitary model of pitch perception” [J. Acoust. Soc. Am. 102, 1811 – 1820 (1997)]. *J Acoust Soc Am* 104: 1118 – 1121, 1998.
- Carlyon RP, Long CJ, Deeks JM, and McKay CM. Concurrent sound segregation in electric and acoustic hearing. *J Assoc Res Otolaryngol* 8: 119 – 133, 2007.
- Carlyon RP and Shackleton TM. Comparing the fundamental frequencies of resolved and unresolved harmonics: Evidence for two pitch mechanisms? *J Acoust Soc Am* 95: 3541 – 3554, 1994.
- Carney LH and Yin TCT. Temporal coding of resonances by low-frequency auditory nerve fibers: Single-fiber responses and a population model. *J Neurophysiol* 60: 1653 – 1677, 1988.
- Cedolin L and Delgutte B. *Frequency selectivity of auditory-nerve fibers studied with band-reject noise* (Abstract #330). Association for Research in Otolaryngology, 2002.
- Cedolin L and Delgutte B. Pitch of complex tones: Rate-place and interspike interval representations in the auditory nerve. *J Neurophysiol* 94: 347 – 362, 2005.
- Cedolin L, and Delgutte B. Spatio-temporal representation of the pitch of complex tones in the auditory nerve. In: *Hearing – From Basic Research to Applications*, edited by Kollmeier B, Klump G, Hohmann V, Langemann U, Mauermann M, Uppenkamp S, Verhey J. New York, NY: Springer-Verlag, 2007, pp. 61 – 70.
- Colburn HS, Carney LH, and Heinz MG. Quantifying the information in auditory-nerve responses for level discrimination. *J Assoc Res Otolaryngol* 4: 294 – 311, 2003.
- Culling JF and Darwin CJ. Perceptual separation of simultaneous vowels: Within and across-formant grouping by F0. *J Acoust Soc Am* 93: 3454 – 3467, 1993.

- Culling JF Darwin CJ.** Perceptual and computational separation of simultaneous vowels: Cues arising from low-frequency beating. *J Acoust Soc Am* 95: 1559 – 1569, 1994.
- Darwin CJ and Carlyon RP.** Auditory grouping. In: *The Handbook of Perception and Cognition*, vol. 6, *Hearing*. Edited by Moore BCJ. London, UK: Academic Press, pp: 387-424, 1995.
- de Cheveigné A.** Separation of concurrent harmonic sounds: Fundamental frequency estimation and a time-domain cancellation model of auditory processing. *J Acoust Soc Am* 93: 3271 – 3290, 1993.
- de Cheveigné A.** Concurrent vowel identification. III. A neural model of harmonic interference cancellation. *J Acoust. Soc Am* 101: 2857 – 2865, 1997c.
- de Cheveigné A.** Vowel-specific effects in concurrent vowel identification. *J Acoust Soc Am* 106: 327 – 340, 1999a.
- de Cheveigné A.** Waveform interactions and the segregation of concurrent vowels. *J Acoust Soc Am* 106: 2959 – 2972, 1999b.
- de Cheveigné A, Kawahara H, Tsuzaki M, and Aikawa K.** Concurrent vowel identification. I. Effects of relative amplitude and F0 difference. *J Acoust Soc Am* 101: 2839 – 2847, 1997a.
- de Cheveigné A, McAdams S, and Marin CMH.** Concurrent vowel identification. II. Effects of phase, harmonicity, and task. *J Acoust Soc Am* 101: 2848 – 2856, 1997b.
- de Cheveigné A, Pressnitzer D.** The case of the missing delay lines: Synthetic delays obtained by cross-channel interaction. *J Acoust Soc Am* 119: 3908 – 3918, 2006.
- Deeks JM and Carlyon RP.** Simulations of cochlear implant hearing using filtered harmonic complexes: Implications for concurrent sound segregation. *J Acoust Soc Am* 115: 1736 – 1746, 2004.
- Duifhuis H, Willems LF, and Sluyter RJ.** Measurement of the pitch in speech: An implementation of Goldstein's theory of pitch perception. *J Acoust Soc Am* 71: 1568 – 1580, 1982.
- Evans EF.** Pitch and cochlear nerve fibre temporal discharge patterns. In: *Hearing: Physiological Bases and Psychophysics*, edited by Klinke R and Hartmann R. Berlin: Springer-Verlag, 1983.
- Goldstein JL.** An optimum processor theory for the central formation of the pitch of complex tones. *J Acoust Soc Am* 54: 1496 – 1516, 1973.
- Greenwood DD.** What is "Synchrony suppression"? *J Acoust Soc Am* 79: 1857 – 1972, 1986.
- Guenther F, Ghosh S, and Tourville J.** Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain Lang* 96: 280 – 301, 2006.
- Heinz M.** *Spectral coding based on cross-frequency coincidence detection of auditory-nerve responses* (Abstract #700). Association for Research in Otolaryngology, 2005.
- Houtsma AJM and Smurzynski J.** Pitch identification and discrimination for complex tones with many harmonics. *J Acoust Soc Am* 87: 304 – 310, 1990.
- 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.
- Johnson DH.** The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones. *J Acoust Soc Am* 68: 1115 – 1122, 1980.
- Kaernbach C and Demany L.** Psychophysical evidence against the autocorrelation theory of auditory temporal processing. *J Acoust Soc Am* 104: 2298 – 2306, 1998.

- Keilson SE, Richards VM, Wyman BT, and Young ED.** The representation of concurrent vowels in the cat anesthetized ventral cochlear nucleus: Evidence for a periodicity-tagged spectral representation. *J Acoust Soc Am* 102: 1056 – 1071, 1997.
- Kiang NYS and Moxon EC.** Tails of tuning curves of auditory nerve fibers. *J Acoust Soc Am* 55: 620 – 630, 1974.
- Kiang NYS, Watanabe T, Thomas EC, and Clark LF.** Discharge patterns of single fibers in the cat's auditory nerve. Cambridge, MA: MIT Press, 1965.
- Larsen E, Cedolin L, and Delgutte B.** *Coding of pitch in the auditory nerve: Two simultaneous complex tones* (Abstract #73). Association for Research in Otolaryngology, 2005.
- Larsen E, Cedolin L, and Delgutte B.** Pitch representations in the auditory nerve: Two concurrent complex tones. *J Neurophysiol* (in press), 2008.
- Liberman MC.** Auditory-nerve response from cats raised in a low-noise chamber. *J Acoust Soc Am* 63: 442 – 455, 1978.
- Liberman MC.** Single-neuron labeling in the cat auditory nerve. *Science* 216: 1239 – 1241, 1982.
- Licklider JCR.** A duplex theory of pitch perception. *Cell Mol Life Sci* 7: 128 – 134, 1951.
- Loeb GE, White MW, and Merzenich MM.** Spatial cross-correlation. *Biol Cybernetics* 47: 149 – 163, 1983.
- Louage DHG, van der Heijden M, and Joris PX.** Temporal properties of responses to broadband noise in the auditory nerve. *J Neurophysiol* 91: 2051 – 2065, 2004.
- May BJ.** Physiological and psychophysical assessment of the dynamic range of vowel representations in the auditory periphery. *Speech Comm* 41: 49 – 57, 2003.
- May BJ, Le Prell GS, and Sachs MB.** Vowel representations in the ventral cochlear nucleus of the cat: Effects of level, background noise, and behavioral state. *J Neurophysiol* 82: 152 – 163, 1998.
- Meddis R and Hewitt MJ.** Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I. Pitch identification. *J Acoust Soc Am* 89: 2866 – 2882, 1991.
- Meddis R and Hewitt MJ.** Modeling the identification of concurrent vowels with different fundamental frequencies. *J Acoust Soc Am* 91: 233 – 245, 1992.
- Merfeld DM, Zupan L, and Peterka RJ.** Humans use internal models to estimate gravity and linear acceleration. *Nature* 398: 615 – 618, 1999.
- Micheyl C, Bernstein JGW, and Oxenham AJ.** Detection and F0 discrimination of harmonic complex tones in the presence of competing tones or noise. *J Acoust Soc Am* 120: 1493 – 1505, 2006.
- Micheyl C, Keebler MV, and Oxenham AJ.** *The role of frequency selectivity in the perception of concurrent harmonic sounds* (Abstract #402). Association for Research in Otolaryngology, 2008.
- Moore BCJ and Carlyon RP.** Perception of pitch by people with cochlear hearing loss and by cochlear implant users. In: *Pitch: Neural Coding and Perception*, edited by Plack CJ, Oxenham AJ, Fay RR, and Popper AN. New York NY: Springer Science, 2005.
- Oxenham AJ, Bernstein JGW, and Penagos H.** Correct tonotopic representation is necessary for complex pitch perception. *Proc Nat Acad Sci* 101: 1421 – 1425, 2004.
- Palmer AR.** Representation of the spectra and fundamental frequencies of steady-state single and double-vowel sounds in the temporal discharge patterns of guinea pig cochlear-nerve fibers. *J Acoust Soc Am* 88: 1412 – 1426, 1990.

- Palmer AR.** Segregation of the responses to paired vowels in the auditory nerve of the guinea-pig using autocorrelation. In: *The Auditory Processing of Speech: From Sounds to Words*, edited by Schouten MEHS. Berlin: Mouton de Gruyter, 1992.
- Palmer AR and Russell IJ.** Phase-locking in the cochlear nerve of the guinea-pig and its relation to the receptor potential of inner hair cells. *Hear Res* 24: 1 – 15, 1986.
- Palmer AR and Winter IM.** Coding of the fundamental frequency of voiced speech sounds and harmonic complex tones in the ventral cochlear nucleus. In: *The Mammalian Cochlear Nuclei: Organization and Function*, edited by Merchan JM, Godfrey DA, and Mugnaini E. New York, NY: Plenum, 1993.
- Patterson RD, Allerhand M, and Giguère C.** Time-domain modeling of peripheral auditory processing: a modular architecture and a software platform. *J Acoust Soc Am* 98: 1890 – 1894, 1995.
- Patterson RD and Nimmo-Smith I.** Off-frequency listening and auditory-filter asymmetry. *J Acoust Soc Am* 67: 229 – 245, 1980.
- Pickles JO.** Frequency threshold tuning curves and simultaneous masking functions in single fibres of the guinea pig auditory nerve. *Hear Res* 14: 245 – 256, 1984.
- Press WH, Teukolsky SA, Vetterling WT, and Flannery BP.** Modeling of data. In: *Numerical Recipes in FORTRAN*. Cambridge, UK: Cambridge University Press, 1992, chapt. 15.
- Pressnitzer D, Patterson RD, and Krumbholz K.** The lower limit of melodic pitch. *J Acoust Soc Am* 109: 2074 – 2084, 2001.
- Qin M and Oxenham AJ.** Effects of envelope-vocoder processing on F0 discrimination and concurrent-vowel identification. *Ear & Hearing* 26: 451 – 460, 2005.
- Rhode WS.** Interspike intervals as a correlate of periodicity pitch in cat cochlear nucleus. *J Acoust Soc Am* 97: 2414 – 2429, 1995.
- Rossi-Katz JA and Arehart KH.** Effects of cochlear hearing loss on perceptual grouping cues in competing-vowel perception. *J Acoust Soc Am* 118: 2588 – 2598, 2005.
- Ruggero M and Temchin AN.** Unexceptional sharpness of frequency tuning in the human cochlea. *Proc Nat Acad Sci* 102: 18614 – 18619, 2005.
- Sachs MB and Abbas PJ.** Rate vs. level functions for auditory-nerve fibers in cat: tone-burst stimuli. *J Acoust Soc Am* 56: 1835 – 1847, 1974.
- Sachs MB and Young ED.** Encoding of steady-state vowels in the auditory nerve: Representation in terms of discharge rate. *J Acoust Soc Am* 66: 470 – 479, 1979.
- Scheffers MTM.** Sifting vowels: auditory pitch analysis and sound segregation (Doctoral Thesis). University of Groningen, The Netherlands, 1983.
- Shamma SA.** Speech processing in the auditory system. II. Lateral inhibition and the central processing of speech-evoked activity in the auditory nerve. *J Acoust Soc Am* 78: 1622 – 1632, 1985.
- Shamma SA and Klein D.** The case of the missing templates: How harmonic templates emerge in the early auditory system. *J Acoust Soc Am* 107: 2631 – 2644, 2000.
- Shera CA, Guinan Jr JJ, and Oxenham AJ.** Otoacoustic estimates of cochlear tuning: Validation in the chinchilla (Abstract #455). Association for Research in Otolaryngology, 2007.
- Shera CA and Guinan Jr JJ.** Stimulus-frequency-emission delay: A test of coherent reflection filtering and a window on cochlear tuning. *J Acoust Soc Am* 113: 2762 – 2772, 2003.
- Shera CA, Guinan Jr JJ, and Oxenham AJ.** Revised estimates of human cochlear tuning from otoacoustic and behavioral estimates *Proc Nat Acad Sci* 99: 3318 – 3323, 2002.

- Shofner WP.** Temporal representation of rippled noise in the anteroventral cochlear nucleus of the chinchilla. *J Acoust Soc Am* 90: 2450 – 2466, 1991.
- Siebert WM.** Frequency discrimination in the auditory system: Place or periodicity mechanisms? *Proc IEEE* 58: 723 – 730, 1970.
- Sinex DG.** Responses of cochlear nucleus neurons to harmonic and mistuned complex tones. *Hear Res* 238: 39 – 48, 2008.
- Stickney GS, Assman PF, Chang J, and Zeng F-G.** Effects of cochlear implant processing and fundamental frequency on the intelligibility of competing sentences. *J Acoust Soc Am* 122: 1069 – 1078, 2007.
- Summers V and Leek MR.** F0 processing and the separation of competing speech signals by listeners with normal hearing and with hearing loss. *J Speech Lang Hear Res* 41: 1294 – 1306, 1998.
- Terhardt E.** Pitch, consonance, and harmony. *J Acoust Soc Am* 55: 1061 – 1069, 1974.
- Todorov E.** Optimality principles in sensorimotor control. *Nat Neurosci* 7: 907 – 915, 2004.
- Tramo MJ, Cariani PA, McKinney MF, and Delgutte B.** *Neural coding of consonance and dissonance* (Abstract #5641). Association for Research Otolaryngology, 2000.
- Tramo MJ, Cariani PA, Delgutte B, and Braida LD.** Neurobiological foundations for the theory of harmony in Western tonal music. *Annals New York Acad Sci* 930: 92-116, 2001.
- Trussell LO.** Synaptic mechanisms for coding timing in auditory neurons. *Ann Rev Physiol* 61: 477 – 496, 1999.
- van der Heijden M and Joris PX.** Panoramic measurements of the apex of the cochlea. *J Neurosci* 26: 11462 – 11473, 2006.
- Wightman FL.** The pattern-transformation model of pitch. *J Acoust Soc Am* 54: 407 – 416, 1973.
- Winter IM, Palmer AR, Wiegand L, and Patterson RD.** Temporal coding of the pitch of complex sounds by presumed multipolar cells in the ventral cochlear nucleus. *Speech Comm* 41: 135 – 149, 2003.
- Young ED and Sachs MB.** 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.
- Young ED, Spirou GA, Rice JJ, and Voigt HF.** Neural organization and responses to complex stimuli in the dorsal cochlear nucleus. *Phil Trans: Biol Sci* 336: 407 – 413, 1992.
- Zweig G.** Basilar membrane motion. *Cold Spring Harb Symp Quant Biol* 40: 619 – 633, 1976.