Pitch Representations in the Auditory Nerve: Two Concurrent Complex Tones

Erik Larsen,1,2 Leonardo Cedolin,1,2 and Bertrand Delgutte1,2,3

1Eaton–Peabody Laboratory, Massachusetts Eye and Ear Infirmary, Boston; 2Speech and Hearing Bioscience and Technology Program, Harvard–Massachusetts Institute of Technology Division of Health Sciences and Technology; and 3Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts

Submitted 17 December 2008; accepted in final form 11 July 2008

Larsen E, Cedolin L, Delgutte B. Pitch representations in the auditory nerve: two concurrent complex tones. J Neurophysiol 100: 1301–1319, 2008. First published July 16, 2008; doi:10.1152/jn.01361.2007. 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 (F0) 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 F0. Templates created by a peripheral auditory model were used to estimate the F0s of double complex tones from the inferred distribution of firing rate along the tonotopic axis. This rate-place representation was accurate for F0s at least 900 Hz. Surprisingly, rate-based F0 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 F0s of concurrent complex tones from interspike-interval distributions pooled over the tonotopic axis. The interval-based representation was accurate for F0s at least 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.

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 (Bregman 1990; Darwin and Carlyon 1995; Scheffers 1983), 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 (Carlyon et al. 2007; Deeks and Carlyon 2004; Moore and Carlyon 2005; Qin and Oxenham 2005; Rossi-Katz and Archard 2005; Stickney et al. 2007; Summers and Leek 1998). 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 (Assmann and Paschall 1998; Beerends and Houtsma 1989; Carlyon 1996, 1997; Micheyl et al. 2006) and even fewer neurophysiological studies (Tramo et al. 2000, 2001; and for concurrent vowels Keilson et al. 1997; Palmer 1990, 1992) have directly addressed the identification and discrimination of the F0s of concurrent complex tones. Herein, 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.

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 (Bernstein and Oxenham 2003; Carlyon and Shackleton 1994; Houtsma and Smurzynski 1990; Plomp 1967). Psycho-physical 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 that of 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 possible only when the tones have resolved harmonics. Micheyli 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 that of the target (1,200–3,600 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 >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 results reported by Carlyon (1996). For higher F0s (200 and 400 Hz), where some or all of the target’s harmonics were resolved, threshold TMRs were typically <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, although 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 to assess the role of resolvability in the neural coding of concurrent F0s.

**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., Assmann and Summerfield 1989; Culling and Darwin 1993; de Cheveigne 1997a,b, 1999a; Scheffers 1983). 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 on the identification of at least one of the two pitches of the concurrent vowels (de Cheveigne 1997c; Meddis and Hewitt 1992) and some models require the identification of both pitches (Assmann and Summerfield 1990; Scheffers 1983). 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 at least four semitones, 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 one 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 (Assmann and Summerfield 1994; Culling and Darwin 1994; but see de Cheveigne 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.

**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 (ISIDs) (Cariani and Delgutte 1996a,b; Evans 1983; Javel 1980; Palmer 1990; Palmer and Winter 1993; Rhode 1995; Shofner 1991; 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 ISID is formally equivalent to the autocorrelation of the spike train. This interval-based pitch representation works with both resolved and unresolved harmonics (Cariani and Delgutte 1996a; Carlyon 1998; Cedolin and Delgutte 2005; Meddis and Hewitt 1991).

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 >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; Shamma and Klein 2000; Wightman 1973) to obtain precise estimates of the stimulus F0. The place cues can also be combined with temporal cues to give various spatiotemporal representations of pitch (Cedolin and Delgutte 2007; de Cheveigne and Pressnitzer 2006; Loeb et al. 1983; Shamma 1985). 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 that in previous studies.

Only a few studies have directly examined the representation of the F0s of concurrent complex tones in the AN and CN (Keilson et al. 1997; Palmer 1990, 1992; Sinex 2008; Tramo et al. 2000). 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 ISID (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, whereas 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 six 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 ISID 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 frequency\(^1\) (Terhardt 1974). These results and those of Palmer (1990) suggest that ISIDs 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 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 occur only 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; Michely 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 to give equal weight to all spectral regions and to facilitate the use of the scaling invariance principle (see following text). We use two different F0 separations (about one and four 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).

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 spatiotemporal 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 dependent only 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 (Heinz 2005; Keilson et al. 1997; May 2003; May et al. 1998; Pickles 1984; Young et al. 1992). 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 spatiotemporal 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 those on the left. The “rate-place profiles” obtained by averaging the spatiotemporal 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 Jr 2003; van der Heijden and Joris 2006). Since F0 was varied over a limited range in our experiments (~2 octaves), deviations from scaling invariance may not present a major problem, as Fig. 1 suggests. This and

\(\text{FIG. 1. Illustration of scaling invariance in cochlear mechanics using a peripheral auditory model. Left: model response of an array of auditory-nerve (AN) fibers with different characteristic frequencies (CFs) to a harmonic complex tone with equal-amplitude harmonics (F0 = 1,000 Hz). Right: model response for one AN fiber (CF = 3,500 Hz) as a function of the F0 of a harmonic complex tone. F0 values were chosen to obtain the same set of normalized frequencies CF/F0 as on the left, so that responses in the 2 panels should be identical if scaling invariance holds. For both panels, gray scale represents the response amplitude, and the timescale is normalized to units of stimulus cycle (t \times F0). Bottom panels show the stimulus waveform on the same normalized scale. Small panels on the right of each main panel show the average model firing rate as a function of CF/F0, obtained by summing the spatiotemporal response patterns over one stimulus cycle.}\)
other issues related to scaling invariance are addressed in the Discussion.

METHODS

Animal preparation and neural recording

Single-unit responses were obtained from the auditory nerve (AN) of five female cats, aged 4–8 mo. 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 and Ear Infirmary and MIT. Briefly, animals were anesthetized with diaz-in-urethane, 75 mg/kg initially administered intraperitoneally, and boosters were given as needed to maintain an areflexive state. Dexamethasone (0.25 mg/kg, administered intramuscularly) was given every 3 h to reduce edema and prevent dehydration. The AN was exposed via a posterior craniotomy and 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 thermostatically 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) in either CAP threshold or single-unit thresholds would cause termination of the experiment.

Single-unit activity was measured with glass micropipettes filled with 2 M 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-μs 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 with thresholds from animals raised in a sound-proof chamber (Liberman 1978). Data from CF regions with a high proportion of abnormally high fiber thresholds were excluded.

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 (NIDAC 6052e; National Instruments) using a sampling rate of 100 kHz. They were delivered to the tympanic membrane via a calibrated closed acoustic system consisting of 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.

Electrophysiological procedures

Clicks (100 μs, 10/s) at 55 dB SPL were used as search stimuli. On 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 three fourths) 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 5.5 in steps of 1/8, creating 33 F0 values in total. 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 that for single complex tones, whereas 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 buildup of the waveform for the subsequent F0. Responses were typically collected over 20 repetitions of the 17.15-s stimulus (33 steps × 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 two or three of these stimuli. Because the measurement order was randomized for each fiber, in some cases responses are available for only a single complex tone or one or two 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.

Data analysis

We developed quantitative methods for estimating the F0s of double complex tones from both the rate responses and the ISIDs of AN 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 spatiotemporal response pattern using data from a single fiber and thus use the term probe F0 (F0p). If 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 CFv/F0eff must match the neural harmonic numbers CF/F0p used in probing the single-fiber response. In practice, we define the effective F0 to be the geometric mean of the set of probe F0s used to study a given fiber, i.e.,

\[ F_0^{\text{effective}} = \sqrt[3]{F_0_1 \cdot F_0_2 \cdots F_0_n} \]

To expedite data collection, complete sets of single and double complex tone stimuli were presynthesized for a limited number of CFs spaced 0.5 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 ±0.25 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.
which were not available for every fiber. Tone as a function of probe \( F_0 \). Note that this method requires the best approximation to the measured responses to a double complex harmonics that, when input to the virtual fiber array model, give the two effective \( F_0 \)s of a double complex tone with equal-amplitude array of virtual fibers with varying CFs. In the third step, we find the used to convert the single-fiber model fit in step 1 into a model for an effective \( F_0 \)s of both single and double tones can be estimated from ISIDs.

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 \( F_0 \), excluding the 20-ms transition intervals over which waveforms for subsequent \( F_0 \) 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 \( F_0 \)s over each cycle of the complex. This fundamental period corresponds to 9-fold the period of the lower probe \( F_0 \) (11-fold the period of the higher \( F_0 \)) for the 11/9 ratio and 14-fold the period of the lower probe \( F_0 \) (15-fold the period of the higher \( F_0 \)) for the 15/14 \( F_0 \) ratio.

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 \( F_0 \) 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 \( F_0 \)s of double complex tones. Specifically, we used a three-step process for quantitatively estimating the two effective \( F_0 \)s 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 \( F_0 \). 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 \( F_0 \)s of a double complex tone with equal-amplitude harmonics that, 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 \( F_0 \). 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 five free parameters that are fit to the single-tone rate response as a function of probe \( F_0 \): \( i \) 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 the 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 that in step 1 into a model for the rate response of an array of virtual fibers with varying CFs. Specifically, each probe \( F_0 \) is mapped into the CF of one virtual fiber using the equation

\[
\{CF_{\text{virt}}\} = F_0_{\text{act}} \times [nh], \quad \text{with} \quad [nh] = \text{BF}_{\text{CT}}/\left[F_0_p\right]
\]

(1)

where \([nh] = \text{BF}_{\text{CT}}/\left[F_0_p\right] \); \( F_0_{\text{act}} \) is the effective \( F_0 \) (the geometric mean of the probe \( F_0 \)s); \([nh] \) is the vector of neural harmonic numbers (varying from \( \sim 1.5 \) to 5.5); and \( \{F_0_p\} \) is the vector of 33 probe \( F_0 \)s of the single complex tones used in step 1. With this convention, the \( CF_{\text{virt}} \) values 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 \( F_0 \). All the model parameters are determined from the fit in step 1, except that \( CF_{\text{virt}} \) varies as in Eq. 1, and the filter bandwidths vary proportionately to \( CF_{\text{virt}} \) 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 \( F_0 \)s.

In step 3 of the estimation procedure, the two effective \( F_0 \)s 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 \( F_0 \)s. The best matching input \( F_0 \)s are the estimated effective \( F_0 \)s of the double complex tone. Note that the effective \( F_0 \)s are assumed to be unknown to quantitatively assess how well they can be estimated from the neural data, assuming the virtual CFs specified in any particular model.

A Levenberg–Marquardt iterative least-squares optimization routine implemented in MatLab (The MathWorks, Framingham, MA) was used both to fit model parameters to the single-tone response (step 1) and to find the effective \( F_0 \)s that give the best match between model predictions and measured rate responses to double complex tones (step 3). 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. SDs of the effective \( F_0 \) estimates (Fig. 4) were computed based on the r.m.s. residuals and the Jacobian at the solution vector (Press et al. 1992).

Interspike-interval analysis

Our method for estimating the two \( F_0 \)s of double complex tones from the temporal discharge patterns of AN fibers is a direct extension of methods used previously to estimate the \( F_0 \) of single complex tones from ISIDs (Cariani and Delgutte 1996a,b; Cedolin and Delgutte 2005; Palmer 1990). The main difference is that, using scaling invariance, the present method gives effective \( F_0 \) estimates from the response of a single fiber measured as a function of probe \( F_0 \), whereas previous methods estimated \( F_0 \) from the response of a population of fibers to a single stimulus. The method consists of two steps (Fig. 5): 1) computation of a “pseudopooled” ISID from the responses of a single fiber as a function of probe \( F_0 \) and 2) estimation of the effective \( F_0 \) by fitting periodic templates to the pseudopooled interval distribution.

We first compute an all-order ISID for every probe \( F_0 \) in a series of single or double complex tones. To implement scaling invariance, the interspike intervals are computed on a normalized timescale (\( t \times F_0 \)) 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 ISIDs are then summed across all probe \( F_0 \)s to form pseudopooled interval distributions. These are not true pooled distributions since pooling normally refers to summation across fibers for a single stimulus, whereas 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 pseudopooled interval distributions, we used periodic templates that select intervals at a given period and its multiples. For each template, \( \text{contrast} \) 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 \( F_0 \), whereas larger contrast values imply that the fiber preferentially fires at that interval. Contrast has been shown to

\[ J \text{ Neurophysiol} \cdot \text{VOL} 100 \cdot \text{SEPTEMBER} 2008 \cdot \text{WWW.JN.ORG} \]
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 F0s (from 0.29- to 3.5-fold the effective F0) and effective F0s are estimated based on maxima in contrast. For a single complex tone, the estimated F0 is simply the template F0 that maximizes contrast. For a double complex tone, the two template F0s with the highest contrasts are selected, with the constraint that the F0 of the second estimate cannot be a multiple or submultiple of the F0 giving the largest contrast.

To make this method more robust, the pseudopooled ISID was weighted with an exponentially decaying function that deemphasizes long interspike intervals corresponding to low effective F0s. This weighting implements the idea that the existence of a lower F0 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 F0, thereby preventing F0 matches to these subharmonics. A decay constant equal to 0.75-fold the period of the lower F0 was found empirically to give a good compromise between reducing subharmonic errors and decreasing template contrast at the effective F0 too much (which could lead to harmonic errors).

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 < SR < 18/s), and 14 (13%) had low spontaneous rates (<0.5/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. Both single and double complex tones were typically presented at 15–35 dB above the fiber’s pure-tone threshold at CF, about halfway into the fiber’s dynamic range as determined by the rate-level function for a single complex tone.

**Figure 2.** Template-matching procedure used to estimate both F0s of a double complex tone from rate responses to both single and double complex tones using an example for a medium spontaneous-rate fiber (CF = 7,000 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 (BFCT) to be precisely estimated from the rate response to the single complex tone. The model fit for this fiber gave a BFCT of 7,049 Hz, very close to the CF measured from the pure-tone tuning curve (0.8% difference).

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 fairly well captures the main peaks and troughs in the response, 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 that 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 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 tones, 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 tones, 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.

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 halfway into each fiber’s dynamic range, ensuring a strong response to the complex tone partials yet avoiding saturation for all SR groups.

Figure 3 shows percentage errors of F0 estimation for double complex tones with F0 ratios of 11/9 (left) and 15/14 (right) for these 112 measurements. The top panels show estimation errors for individual fibers, whereas the bottom 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 BFCT.

The absolute estimation errors in Fig. 3 are relatively large (2% for CF < 2–3 kHz and decrease with increasing CF to fall < 1% at > 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 ANOVA with F0 ratio (11/9 and 15/14), CF range (0.5–2, 2–3 kHz and decrease with increasing CF), 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 greater than errors for the lower tone for the 11/9 F0 ratio, whereas 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 top panels in Fig. 3). This bias is most likely the result of assigning the two effective F0 estimates as...
either “low” or “high” to minimize the combined error with respect to both F0s. This makes it relatively unlikely that an estimated effective F0 that 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.

To quantify the precision of F0 estimates, we used the SDs of the fitted F0 parameters calculated from the residuals (see Data analysis). Figure 4 shows the SD 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 ratios, respectively. The top and bottom panels show data for individual fibers and moving averages of the log-transformed data, respectively. Also included in Fig. 4 are the SDs of the BFCT estimates for double complex tones ultimately depends on the precision of the BFCT estimate in step 1 of the estimation procedure.

The SDs of the F0 and BFCT estimates decrease gradually with increasing CF, again consistent with the improvement in relative cochlear frequency selectivity. For CFs >2 kHz, the BFCT estimates are more precise than the F0 estimates, although 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, 2–4, 4–8, 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 (SD lower) for higher CFs, as well as for single complex tones, compared with either the lower or higher tone of the double complex (statistically equivalent; Tukey–Kramer). These differences between single and double tones were significant only for CFs >4 kHz, which resulted in a CF × 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 BFCT parameter fitted to the rate responses to single complex tones, it is of interest to compare BFCT 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 BFCT and CF: the median absolute differences were 9.3, 3.2, and 2.8% in the CF ranges <2 kHz, between 2 and 5 kHz, and >5 kHz, respectively. The small differences between tuning curve CF and BFCT >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 >2–3 kHz, where the relative frequency resolution of the cochlea is the best. For CFs <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) BFCT 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 BFCT estimate (Fig. 4). Unexpectedly, for CFs >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 4](http://jn.physiology.org/ by 10.220.33.2 on May 24, 2017)
**Interspike-interval analysis**

Figure 5 illustrates our method for estimating the effective F0s of double complex tones using ISIDs from a medium spontaneous-rate fiber (CF = 816 Hz, SR = 2.6 spikes/s, threshold = 37 dB SPL). Figure 5A 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. Figure 5B shows all-order 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(low), which varied from about 1.5 to 5.5; F0(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 “pseudopooled” ISID (Fig. 5C). The pseudopooled 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 pseudopooled ISID.

Effective F0s were estimated from the pseudopooled 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, whereas 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.

Although in this case the two largest peaks in the template contrast function occurred at the two F0s present in the double-tones.
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, in that 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.

*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 shows percentage F0 estimation errors (*left*) and the associated contrast values (*right*) obtained from responses to single complex tones. The *top* and *bottom 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 <500 Hz to about 3% around 4 kHz; over the same CF range, contrast declines gradually from about 10 to about 2. For CFs >4 kHz, F0 estimation errors increase faster, whereas contrast remains at a floor level of about 2. These patterns of F0 errors and contrast values obtained from pseudopooled 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 percentage 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 *bottom 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 with 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 (0.5-octave bins from 500 Hz to 16 kHz, except 125–500 Hz and 1–2 kHz due to sparser sampling in 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, whereas their interaction was not. The errors for the higher F0 of the double complex tone were significantly higher than errors either from the lower F0 or 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), whereas 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 with respect to 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 \(F_{(9,220)} = 42.33, P < 0.001\) and tone height [high vs. low F0, \(F_{(1,220)} = 4.31, P = 0.039\)], but no effect of F0 ratio. In combining these analyses, we found 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.

**Fig. 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. *Top panels* show errors for individual AN fibers; *bottom panels* show moving window averages of the log-transformed data, using 1-octave-wide CF bins with 50% overlap. Error bars represent ±1SE. In the *top left panel*, triangles show data that lie out of the vertical range.
Figure 8 shows contrast values for the same conditions as for the F0 estimation errors in Fig. 7, 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 both the single-tone 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 the 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 that 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.

**FIG. 7.** Percentage F0 estimation errors from pseudopooled ISIDs for double complex tones as a function of CF for the AN fiber population. \textit{Left and right panels} show results for F0 ratios of 11/9 and 15/14, respectively. \textit{Top panels} show errors for individual AN fibers; \textit{bottom panels} show moving window averages of the log-transformed absolute errors, using 1-octave-wide CF bins with 50% overlap. For reference, the \textit{bottom panels} also reproduce the mean F0 estimation errors for single complex tones (crosses) from Fig. 6. Error bars represent ±1SE. In the \textit{top panels}, triangles show data that lie out of the vertical range.

**FIG. 8.** Neural pitch strength (template contrast), from pseudopooled ISIDs for double complex tones as a function of CF for the AN fiber population. \textit{Left and right panels} show results for F0 ratios of 11/9 and 15/14, respectively. \textit{Top panels} show errors for individual AN fibers; \textit{bottom panels} show moving window averages of the log-transformed absolute errors, using 1-octave-wide CF bins with 50% overlap. For reference, the \textit{bottom panels} also reproduce the template contrasts for single complex tones (crosses) from Fig. 6. Error bars represent ±1SE.
In summary, F0s of both single and double complex tones can be accurately estimated (errors <3%) from pseudopooled ISIDs for CFs <2–4 kHz. Since the range of probe F0s in our stimuli is proportional to the CF, this means that F0 estimation based on ISIs works best for lower F0s. Our correlate of pitch strength, the template contrast, is at least two times smaller for double complex tones than that for a single complex tone with the same F0, reflecting the competition for phase-locked spikes (also known as “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 with 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.

**Effect of spontaneous rate on pitch estimation from interspike intervals**

Figure 9, A and B shows 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), whereas 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 pseudopooled 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 pseudopooled 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 versus 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, C and D shows 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 F0s are included. On average, contrast for low-/medium-SR fibers is higher than that for high-SR fibers, an observation confirmed by a two-way ANOVAs with CF (bins: 0.25–1, 1–2.5, and 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, although 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 \[ CF: F(3,169) = 8.51, P < 0.001; SR: F(1,169) = 14.9, P < 0.001; CF × SR: F(3,169) = 2.27, P = 0.082 \].

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 ISIDs. 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 \[ CF: F(3,169) = 13.7, P < 0.001; SR: F(1,169) = 0.02, P = 0.88; CF × SR: F(3,169) = 0.27, P = 0.84 \]. 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. 9, A and B).

**FIG. 9.** Effect of spontaneous rate on neural template contrast (a correlate of pitch strength). A and B: ISID as a function of probe F0 (top) and pseudopooled interval distribution (bottom) in response to a single complex tone for a high-spontaneous discharge rate (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 and 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 contrast values for both F0 ratios and both the lower and higher F0s of each double tone.
**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 ISIDs. 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 ≥3 kHz. Periodic templates were used to estimate the effective F0s of single and double complex tones from pseudopooled ISIDs. This temporal representation was accurate for CFs ≤3 kHz.

Although 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 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 ≥900 Hz (3,000/3.3), whereas the interval-based representation should be effective for F0s <900 Hz. Together, the two representations cover a wide range of F0s, including the 500- to 1,000-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, whereas 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).

**Scaling invariance in cochlear mechanics**

In a perfectly scaling invariant cochlea, the response to frequency f at the cochlear location tuned to CF is dependent only 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 is equal to the magnitude and the phase of the response to frequency f0 at the cochlear location tuned to βCF, where β = f/CF. This means that the waveforms of the two responses are the same except for a scaling in time by a factor 1/β. By varying f (and therefore β) and measuring responses at a fixed location tuned to CF, one can therefore infer the response to a stimulus of fixed frequency f0 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 Jr 2003; van der Heijden et al. 2005). Specifically, whereas 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 thus 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 timescale (t × F0), 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.

**Deviations from scaling invariance in cochlear mechanics**

Although 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 (Cedolin and Delgutte 2005; Shera and Guinan Jr 2003; Shera et al. 2002); perfect scaling invariance would lead to an exponent of 0. Our stimuli were designed so that the ratio CF/F0 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 value 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 F0s 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 F0 estimation from rate responses. This conclusion is supported by the observation that the range of effective F0s over which we could reliably estimate BFCT from responses to single complex tones corresponds well to the F0 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 BFCT estimates were reliable (SD <1%) for CFs >1.5 kHz, which corresponds to F0s >450 Hz, whereas Cedolin and Delgutte (2005) could reliably estimate pitch from rate-place profiles for F0s >400–500 Hz.

In summary, the deviations in cochlear bandwidths from perfect scaling invariance are likely to have moderate effects.
on F0 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 (Cariani and Delgutte 1996a; Carlyon 1998; Meddis and Hewitt 1991), the effect on interval-based representations of pitch is expected to be even smaller.

**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 ISIDs (Cariani and Delgutte 1996a; Meddis and Hewitt 1991). The strength of phase locking in AN fibers (as measured by the synchronization index, also known as vector strength) drops rapidly with increasing stimulus frequency >1 kHz, until phase locking is hard to detect at >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/F0 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/F0. In contrast, using measurements made at a fixed CF as a function of probe F0 by assuming scaling invariance, the dominant frequency never deviates much from the CF, so that synchrony is nearly constant as a function of CF/F0. The net result is that the scaling invariance approximation acts like a high-pass filter: low-order harmonics (relative to the mean neural harmonic number 3.3) are attenuated, whereas 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 high-pass filtering effect, we ran simulations using a fourth-order Butterworth low-pass filter with a 3-dB cutoff frequency of 2,570 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 high-pass 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 F0 from ISIDs. Next, to simulate the effect of pooling the ISIDs, we summed the squared synchronies to the dominant frequencies over the entire range of CF/F0 (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-F0 conditions never exceeded 1 dB. Thus whereas the high-pass filtering effect resulting from the scaling invariance assumption can have substantial effects on ISIDs for individual probe F0s, the effect on the pseudopooled distribution obtained by summing across probe F0s is expected to be small.

This conclusion is supported by the observation that the upper frequency limit over which we could reliably estimate the F0 of single complex tones from pseudopooled interval distributions corresponds well to the F0 range previously found by Cedolin and Delgutte (2005) using true pooled distributions. Specifically, our F0 estimates for single complex tones were reliable for CFs <4 kHz (Fig. 6A), which corresponds to F0s <1,200 Hz, whereas Cedolin and Delgutte (2005) could estimate pitch from pooled distributions for F0s <1,300 Hz.

**Effects on adaptation and refractoriness**

Both the upper frequency limit of phase locking and 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. Moreover, 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 <200 spikes/s), so that effects of refractoriness were also minimized. The scaling of the refractory period is apparent in the ISIDs of Figs. 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 pseudopooled 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 ISIDs for the stimulus conditions of our experiments.

**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 (Duifhuis et al. 1982; Goldstein 1973; 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 > 400–500 Hz. As we have pointed out, this fits well with the CF range over which we could estimate the BF\textsubscript{CT} parameter of the peripheral model from single-tone responses using the scaling factor CF \approx 3.3F0. 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 that 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, resulting in weak modulation of firing rate across the 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 resolved in the mixture only 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 sensitive only to peak locations. Previous reports (e.g., Assmann and Summerfeld 1990; de Cheveigné 1999a,b; Palmer 1990) 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 three harmonics of the upper tone could be represented in pseudo rate-place profiles for the 11/9 ratio, versus four 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 sensorimotor control (e.g., Guenther et al. 2006; Merfeld et al. 1999; Todorov 2004). 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 (two) 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.

**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 ISIDs (de Cheveigné 1993; Palmer 1990, 1992; Tramo et al. 2000). Unlike the estimation method used with rate-place profiles, F0 estimation from ISIDs 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 (Cariani and Delgutte 1996a,b; Meddis and Hewitt 1991; Palmer 1990). 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.

Although 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 1,200 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
generate only half the number of phase-locked spikes com-
pared with a single complex tone. This reduces the number of
ISIs 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 “syn-
chrony suppression” (Greenwood 1986), is most clearly appar-
tent in the template contrast, which was at least a factor of 2
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 those for single complex tones.

We also found that low-/medium-SR fibers (<18.5 spikes/s)
yielded higher template contrasts than those of 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 that 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 those of 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 nonphase-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 with respect to threshold would tend to
increase synchrony in high-SR fibers, which have a lower
threshold than that of low-SR fibers. However, this effect
would be significant only at levels very close to threshold since
synchrony saturates at lower levels than firing rate (Johnson
1980).

Although 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 contain-
ing 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 (Cariani and Delgutte
1996b; Meddis and Hewitt 1991). 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. Specifi-
cally, we found that the pooled synchrony to the higher tone
never differed by >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.

Relation to psychophysics

Psychophysical studies indicate that accurate pitch percep-
tion for both single and double complex tones depends strongly
on the presence of resolved harmonics (Beerends and Houtsmma
1989; Bernstein and Oxenham 2003, 2005; Carlyon 1996;
Carlyon and Shackleton 1994; Houtsm and Smurzynski 1990;
Micheyl et al. 2006). Generally, pitch discrimination and
identification are 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 inher-
ently requires resolved harmonics. Micheyl et al. (2006) intro-
duced 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 although re-
solved harmonics prior to mixing are necessary for accurate F0
discrimination, resolvability in the mixture may not be neces-
sary (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 for
only 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 associ-
ated with resolved harmonics rather than just the peak loca-
tions. At the same time, the F0 range over which rate-place
estimation was effective was more restricted for double tones
than that for single complex tones, indicating that resolvability
in the mixture is not wholly unimportant.

Although the rate-place representation is consistent with the
dependence of psychophysical results on harmonic resolvabil-
ity, it poorly accounts for the range of F0s over which discrim-
ination is accurate with double tones. Estimation from rate-
place information was effective only for F0s \(\approx 900\) Hz,
whereas 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),
ocochlear frequency resolution is two- to threefold sharper in
humans than that in cat and other popular experimental animals
(but see Ruggiero 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 (Colburn et al. 2003; Sachs and
Young 1979), 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 ISIDs 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 ISI representation is likely to be effective over the F0 range of male and female voices as well as the 100- to 400-Hz range used in previous human psychophysical studies of F0 identification and discrimination for double complex tones. However, accurate estimation from ISIs 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 ISIDs may not be used at higher levels of the auditory pathway (Carlyon 1998; Cedolin and Delgutte 2005; Kaerbach and Demany 1998; Oxenham et al. 2004; Siebert 1970).

Various spatiotemporal models of pitch processing have been proposed to overcome difficulties with both purely spectral and purely temporal models (Bernstein and Oxenham 2005; Cedolin and Delgutte 2007; de Cheveigné and Pressnitzer 2006; Shamma 1985). 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 ISI 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 ISI representations, our results do not indicate large differences in F0 estimation accuracy for the 15/14 versus 11/9 F0 ratios. This finding contrasts with psychophysical results for concurrent vowels (Assmann and Paschall 1998; Assmann and Summerfield 1990), 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), whereas they gave bimodal distribution of matches for double vowels with a four-semitone difference (close to our 11/9 ratio). Assmann and Paschall interpreted these results as indicating that listeners perceived a single pitch with a one-semitone difference but heard two separate pitches with the four-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 five 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 <1,200 Hz (and therefore no resolved harmonics for a 100-Hz F0), whereas 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.

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 ISI representation was effective at low F0s (<900 Hz), whereas the rate-place representations was effective at high F0s (>900 Hz). If cochlear tuning is two- to threefold sharper in humans than that 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, whereas the higher limit for accurate ISI 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 ISI 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 ISI information was equally effective for F0 ratios of 15/14 (approximately one semitone) and 11/9 (approximately four 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 four semitones. Additional psychophysical data on F0 identification with double complex
tones are needed to understand the reasons for this possible discrepancy.

A C K N O W L E D G M E N T S

We thank C. Miller for expert surgery, C. Micheyl for suggesting the possible implication of having resolved harmonics prior versus after mixing of two harmonic complexes, and G. Wang for reading a preliminary version of the manuscript.

G R A N T S

This work was supported by National Institute on Deafness and Other Communication Disorders (NIDCD) Grants R01 DC-002258 and P30 DC-005209. E. Larsen was partly supported by NIDCD Training Grant T32 DC-00038. This article is based on a thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in the Department of Electrical Engineering and Computer Science at the Massachusetts Institute of Technology in September 2008.

R E F E R E N C E S


