Neural Representations of Pitch: Role of Peripheral Frequency Selectivity

ARCHIVES

by

Leonardo Cedolin

Laurea in Electrical Engineering,
Politecnico di Milano, 1999

SUBMITTED TO THE DIVISION OF HEALTH SCIENCES AND TECHNOLOGY
IN PARITAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN
HEALTH SCIENCES AND TECHNOLOGY

AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
FEBRUARY 2006

Signature of Author: _____
Division of Health Sciences and Technology
January 27, 2006

Certified by: _____
Bertrand Delgutte
Associate Professor of Otology and Laryngology
and Health Sciences and Technology, Harvard Medical School
Thesis Supervisor

Accepted by: _____
Martha L. Gray
Professor of Medical Engineering and Electrical Engineering
Director, Harvard-MIT Division of Health Sciences and Technology

Neural Representations of Pitch: Role of Peripheral Frequency Selectivity

By

Leonardo Cedolin

Laurea in Electrical Engineering, Politecnico di Milano, 1999

Submitted to the Division of Health Sciences and Technology on January 27, 2006
in partial fulfillment of the requirements for the degree Doctor of Philosophy in
Health Sciences and Technology

Abstract

Investigating the neural mechanisms underlying the perception of the pitch of harmonic complex tones is of great importance for many reasons. Changes in pitch convey melody in music, and the superposition of different pitches is the basis for harmony. Pitch has an important role in speech, where it carries prosodic features and information about speaker identity. Pitch plays a major role in auditory scene analysis: differences in pitch are a major cue for sound source segregation, while frequency components that share a common fundamental frequency (F0) tend to be grouped into a single auditory object.

In psychophysics, a positive correlation is commonly observed between the estimated "resolvability" of individual harmonics of complex tones, assumed to depend primarily on the frequency selectivity of the cochlea, and the strength of the corresponding pitch percepts. In this study, possible neural codes for the pitch of harmonic complex tones were investigated in the auditory nerve of anesthetized cats, with particular focus on their dependence on cochlear frequency selectivity, which was measured directly using both complex tones and band-reject noise.

A "rate-place" representation of pitch, based on cues to peripherally-resolved harmonics in profiles of average discharge rate along tonotopically-arranged neurons, was compared to a "temporal" representation, based on periodicity cues in the distributions of interspike intervals of the entire auditory nerve. Although both representations were viable in the range of F0s of cat vocalizations, neither was entirely satisfactory in accounting for psychophysical data. The rate-place representation degraded rapidly with increasing stimulus level and could not account for the upper limit of the perception of the pitch of missing-F0 in humans, while the interspike-interval representation could not predict the correlation between psychophysical pitch salience and peripheral harmonic resolvability.

Therefore, we tested an alternative, "spatio-temporal" representation of pitch, where cues to the resolved harmonics arise from the spatial pattern in the phase of the basilar membrane motion. The spatio-temporal representation was relatively stable with level and was consistent with an upper limit for the pitch of missing-F0, thus becoming the strongest candidate to explain several major human pitch perception phenomena.

Thesis Supervisor: Bertrand Delgutte
Title: Associate Professor of Otology and Laryngology and Health Sciences and Technology
Harvard Medical School

Acknowledgements

Questo lavoro è dedicato ai miei genitori, Carla and Luigi Cedolin.

Grazie, mamma e papà, per l'affetto, il sostegno morale e l'incitamento che mai mi avete fatto mancare nonostante la lontananza, e per essere sempre stati il mio modello di onestà, impegno e perseveranza.

I am deeply and truly grateful to my advisor Bertrand, for constantly steering me back onto the right path, and to Nelson, for giving me this life-changing opportunity. Thanks!

Among many others, the following people contributed to making my academic life and my adventure in Boston enjoyable and unforgettable. Thanks to all of you !

Andrew Oxenham, Chris Shera and Garrett Stanley, members of my thesis committee; Julie Greenberg and John Wyatt; the entire Speech and Hearing faculty; the HST staff at E-25; Peter Cariani; Sergio Cerutti e Emanuele Biondi; Dianna, Laura, Mike R., Connie, Leslie, Stephane, Kelly, Ish, Frank, Chris and Wen at EPL; Dr. Merchant, Dr. Halpin, Dr. Varvares and Dr. Nadol; Jocelyn, Erik, Suzuki and Waty ("coffee? Yeah, why not?"); my awesome lunch buddies Zach, Brad and Dave; Courtney, Chandran, and Anna ("is my music too loud? is the light too dim?"); Craig ("The strongest man alive"), Sasha, Ken ("I'll put anything on pizza"), Teresa; Keith and Darren ("don't talk about Fight Club"); Josh B. ("amici!!!!") and Jill; Hector, Isabel and Joey; the softball tribe and Amy at Shay's; Ashish and Minnan; the Mays, the Rizzos and the Stasons; Ozzie, Barclay and Finnigan; Jamie, Kate, Mark, Meri, Cis, George and Kari; Chris Nau, Greg and Ray; Chris B. ("I'll gamble"); "Orange" Jeroen, "Crazy" Hannah, Eric P., Juan-ita, Michi, Jeff "Gold Man"; my great HMS friends Kedar, Theo, Irina, Mindy and Sarah; Jean; my roommates (and their families): "hawaii" Chris E., "kimchi is gone" Kwang-Wook, "VV" Virgilio, Mia (and Cathy and the Irish gang); my amazing Italian buddies Ila, Ponchia, Zed, Claudia, Sbira, Drago e Pellegra; the 67a Dana gang: Paul, Remo ("whazzuuup") and Steak ("hey-hey"); Bruce ("nice hat dude"); my brothers Martino, Sridhar, Mike C. and Noop; Ali, my funny, smart, sweet, loving Bambina.

Table of Contents:

INTRODUCTION

Harmonic complex tones are the sum of single sinusoidal sounds ("pure tones") whose frequencies are all integer multiples of a common fundamental frequency (F0). Harmonic complex tones are an extraordinarily important class of sounds found in natural environments. For example, harmonic complex tones are produced by the vibrations of the vocal folds, source of "voiced" sounds in human speech. Sounds produced by a wide variety of musical instruments are also harmonic, as often are animal vocalizations.

Typically, individual frequency components of a harmonic complex tone are not perceived by human listeners as separate entities, but they are "grouped" together into a single auditory percept, commonly known as *pitch*. Pitch is defined as "*that attribute of auditory sensation in terms of which sounds may be ordered on a scale extending from low to high*" (ANSI 1973). Many classes of sounds can evoke a pitch sensation, including for example pure tones, complex tones and even some natural or appropriately manipulated types of noise. The pitch evoked by a harmonic complex tone with a fundamental frequency F0 typically coincides with the pitch evoked by a pure tone with frequency equal to F0. For F0s up to 1400 Hz (Moore 1973), this holds even when the component at the F0 is not physically present in the sound (Schouten 1940) or is masked with noise (Licklider 1954). This phenomenon is known as "pitch of the missing fundamental". Interestingly, a pitch at the F0 can still be heard even when a harmonic complex lacks a number of components, and in this case the strength of the perceived pitch depends strongly on *which* harmonics are missing (e.g. Houtsma and Smurzynski 1990).

Pitch plays many important roles in speech, where it provides the foundations for prosody and information used to segregating simultaneous speakers (Darwin and Hukin 2000). In tone languages such as Mandarin Chinese, pitch also carries lexical information. Changes in pitch convey melody in music, and the superposition of different pitches is the basis for harmony. Pitch plays a major role in auditory scene analysis: differences in pitch are a major cue for sound source segregation, while frequency components that share a common fundamental tend to be grouped into a single auditory object (Bregman 1990). Pitch may also be of great importance for animals to process conspecific vocalizations, often consisting of harmonic complex tones. Moreover, the pitch of the missing fundamental is known to be

perceived not only by humans, but also by cats (Heffner and Whitfield 1976), monkeys (Tomlinson and Schwartz 1988) and birds (Cynx and Shapiro 1986). These observations support the use of animal models (in our case the cat) to study neural representations of pitch.

The main goal of this thesis was to study possible neural mechanisms underlying the perception of the pitch of harmonic complex tones, with particular focus on seeking a neural representation of pitch that is consistent with major trends observed in human psychophysics experiments, as no current physiologically-based model can fully account for the whole body of psychophysical data available on pitch perception.

Despite having been investigated for over a century (Seebeck 1841; Ohm 1843), the exact nature of the neural coding of pitch is still debated. The peripheral auditory system can, in principle, generate two types of cues to the pitch of harmonic complex tones. On one hand, the harmonicity of the frequency spectrum (at F0) can be encoded thanks to the tonotopic mapping and frequency analysis performed by the cochlea. On the other hand, harmonic sounds with the same fundamental frequency F0 also share the same waveform periodicity (at T0 = 1/F0), reflected in regularities in the precise timings of neural action potentials thanks to phase locking.

The peripheral auditory system can be thought of as a bank of bandpass filters (commonly referred to as "auditory filters"), which represent the mechanical frequency analysis performed by the cochlea. When the frequency separation between two harmonics exceeds the bandwidth of the auditory filters, these harmonics result in local maxima (at locations tuned to the harmonics) and minima (at locations tuned to frequencies in between those of the harmonics) in the spatial pattern of basilar membrane motion, reflected in the resulting pattern of activation of the auditory nerve (AN). In this case, the harmonics are said to be "resolved" by the auditory periphery. On the other hand, when two or more harmonics fall within the pass-band of a single peripheral filter, they are said to be "unresolved". The bandwidths of the auditory filters are known to increase with their center frequency (e.g. Kiang et al. 1965; Shera et al. 2002), so that only low-order harmonics are resolved. "Spectral" or "place" models of pitch perception propose that the pitch of complex tones containing resolved harmonics may be extracted by matching the pattern of activity across a tonotopic neural map to internally stored harmonic templates (Cohen et al. 1994; Goldstein 1973; Terhardt 1974; Wightman 1973).

However, also harmonic complex tones consisting entirely of high-order, unresolved harmonics evoke a pitch percept. Unresolved harmonics do not produce local maxima in the spatial pattern of activation along the cochlea, thereby excluding a purely spectral, template-based, explanation for pitch extraction. On the other hand, unresolved harmonics can produce temporal cues to pitch because any combination of unresolved harmonics has a period equal to that of the complex tone (T0). These periodicities, reflected in neural phase locking, result in *interspike intervals* of individual AN fibers most frequently corresponding to T0 and its integer multiples. "Temporal" models of pitch perception propose that periodicity cues to pitch may be extracted by an autocorrelation-type mechanism (Licklider 1951; Meddis and Hewitt 1991; Moore 1990; Yost 1996), which is mathematically equivalent to an all-order interspike-interval distribution for neural spike trains. This model can in principle account also for the perception of the pitch of only resolved harmonics: since T0 is an integer multiple of the period of any of the harmonics, by "*pooling*" (summing) autocorrelation functions (or, equivalently, all-order interspike-interval distributions) across frequency channels one can extract the information about the common periodicity (Meddis and Hewitt 1991; Moore 1990).

It cannot be ruled out that the true neural code for pitch might be neither purely place-based nor purely temporal. On one hand, the co-existence of two separate representations (a place representation for resolved harmonics and a temporal representation for unresolved harmonics) has been hypothesized (Carlyon and Shackleton 1994). On the other, more than one plausible "spatio-temporal" mechanism have been proposed, by which a combination of place and temporal information might be used to identify individual components of a harmonic complex tone. For example, Srulovicz and Goldstein (1983) hypothesized that individual harmonic frequencies may be extracted by passing the interspike-interval distribution of each AN fiber through a filter matched to the period of the characteritisc frequency (CF) of that fiber. Others (Loeb et al. 1983; Shamma 1985a) have suggested that cues to spectral features of a sound could emerge when comparing the timings of spikes in AN fibers with neighboring CFs, due to changes in the velocity of propagation of the cochlear traveling wave.

In particular, the idea for an alternative, *spatio-temporal* representation of pitch stems from the observation that the fastest variation with cochlear place of the phase of basilar

membrane motion in response to a pure tone occurs at the cochlear location tuned to the tone frequency (Anderson 1971; Pfeiffer and Kim 1975; van der Heijden and Joris 2003). At frequencies below the limit of phase-locking, this rapid spatial change in phase of the basilar membrane motion results in a rapid variation of the latency of the response of auditory-nerve fibers with characteristic frequency, thus generating a cue to the frequency of the pure tone (Shamma 1985a). Our hypothesis is that, for a harmonic complex tone, rapid changes in phase may occur at each of the spatial locations tuned to a resolved harmonic. Hence, latency cues to the frequencies of each of the resolved harmonics of a complex tone would be generated by the wave of excitation produced in the cochlea by each cycle of the stimulus, as the wave travels from the base to the apex. In principle, these spatio-temporal cues to resolved harmonics can be extracted by a central neural mechanism sensitive to the relative timing of spikes in AN fibers innervating neighboring cochlear locations, effectively converting spatio-temporal cues into rate-place cues, which in turn can be combined to generate a pitch percept by a template-matching mechanism.

The concept of harmonic resolvability has been introduced as a possible explanation for a variety of phenomena observed in human studies of the perception of pitch. For example, human subjects with normal hearing perform much better in detecting small differences in the F0 of harmonic complexes in the presence of low-order harmonics than in their absence (Bernstein and Oxenham 2003b; Houtsma and Smurzynski 1990). This difference, in combination with the increase in bandwidth of auditory filters with their center frequency, is usually attributed to a neural mechanism more sensitive to resolved harmonics than to unresolved harmonics. Hearing impaired listeners with sensorineural hearing loss exhibit a decreased performance in detecting small F0-differences (e.g. Moore and Peters 1992; Bernstein and Oxenham 2006), at least partly as a result of their impaired cochlear frequency selectivity (Bernstein and Oxenham 2006), thought in turn to result in a decreased degree of harmonic resolvability. Moreover, the pitch evoked by exclusively high-order harmonics depends strongly on the phase relationships among the partials, while pitch evoked by low-order harmonics is phase-invariant (Carlyon and Shackleton 1994; Houtsma and Smurzynski 1990). Since a dependence on phase is expected only for harmonics interacting within single cochlear filters, these results can also be explained in terms of harmonic resolvability.

Human psychophysics experiments present the limitation that the sharpness of tuning of the mechanical frequency analysis performed by the auditory periphery, assumed to determine whether harmonics are resolved or not, cannot be estimated directly. Specifically, the extent to which psychophysical measures of peripheral tuning sharpness are influenced by more central stages of neural processing is unknown. In this thesis, the strengths and limitations of each of three possible neural representations of pitch mentioned above were investigated in the auditory nerve of anesthetized cats and discussed principally in terms of their dependence on harmonic resolvability in the species studied, which was directly quantified for single auditory-nerve fibers in Chapter 1.

Chapter 1 explores, at the level of the cat auditory nerve, the viability and effectiveness of a rate-place representation of the pitch of missing-fundamental harmonic complex tones, based on the harmonicity of their frequency spectrum, the tonotopic structure of the cochlea and harmonic resolvability. The results are compared to those obtained for a temporal representation of pitch, based on waveform periodicity and on neural phase-locking reflected in pooled all-order interspike-interval distributions. The fundamental frequencies used varied over a wide range (110 – 3520 Hz), extending over the entire F0-range of cat vocalizations [500 – 1000 Hz (Brown et al. 1978; Nicastro and Owren 2003; Shipley et al. 1991)] and beyond. The results of this study indicate that, although both rate-place and interspike-interval representations of pitch are viable in the cat AN over the entire F0-range of conspecific vocalizations, neither is entirely consistent with human psychophysical data. The rate-place representation predicts the greater strength and the phase-invariance of the pitch of resolved harmonics, but it degrades rapidly with stimulus level and it fails to account for the existence of an upper limit to the perception of the pitch of the missing fundamental in humans. On the other hand, a main limitation of the interspike-interval representation lies in the fact that it predicts greater strength for unresolved harmonics than for resolved harmonics, in sharp contrast with human psychophysics results.

Chapter 2 presents for the first time a physiological test of a spatio-temporal representation of the pitch of harmonic complex tones, based on the availability of temporal cues to resolved harmonics at the specific cochlear locations tuned to the harmonic frequencies. Effectiveness and strength of the spatio-temporal representation were first studied in single AN fibers, and then these results were re-interpreted as a function of

fundamental frequency using the *scaling invariance* principle (Zweig, 1976) in cochlear mechanics. The results of this study indicate that the spatio-temporal representation overcomes the limitations of both a strictly spectral and of a strictly temporal representation in accounting for trends observed in many studies of human psychophysics.

Chapter 3 tests the physiological basis of a method widely used in psychophysics to estimate peripheral frequency selectivity: the *notched-noise method* (Patterson 1976). Like other similar methods [for example *psychophysical tuning curves*, using pure tones (Moore 1978), or methods using *rippled noise* (Houtgast 1972)], the notched-noise method is based on the phenomenon of *masking*. Masking is the process by which one sound (the "masker") can interfere with the perception of another sound (the "signal"). The notched-noise method consists of adjusting the spectrum level of a band-reject ("notched") noise masker to a threshold at which a listener fails to detect a fixed-level pure-tone signal, as a function of the width of the rejection band and its placement with respect to the tone. The frequency-resolving power of the auditory periphery is modeled as a bank of "auditory filters", whose parameters are then estimated using the assumptions ("*power spectrum model of masking*", Fletcher 1940) that (1) when detecting a signal in the presence of a masking sound, the listener attends to the filter with the highest signal-to-noise power ratio at its output, (2) detection thresholds correspond to a constant signal-to-masker power ratio at the output of the filter, and (3) the filter is linear for the range of signal and masker levels used to define it. In our physiological experiment, cochlear auditory filters were fit to thresholds for pure tones in band-reject noise measured from single auditory nerve fibers using a paradigm adapted from the one used in psychophysical experiments. A neural population model was then developed to test the extent to which auditory filters estimated psychophysically match the underlying neural filters. The similarity observed between the frequency selectivity measured physiologically and the predicted psychophysical tuning supports the use of the notched-noise method in human psychophysics. Despite the highly nonlinear nature of the cochlea, estimates of neural frequency selectivity derived using the notched-noise method show an increase in tuning sharpness with CF similar to that obtained in previous studies using a variety of different stimuli and techniques and in broad agreement with that inferred from results related to harmonic resolvability in Chapter 1.

Chapter 1

Pitch of Complex Tones: Rate-Place and Interspike-Interval Representations in the Auditory Nerve

The work described in this chapter is published in the *Journal of Neurophysiology:*

Cedolin L and Delgutte B. Pitch of Complex Tones: Rate-Place and Interspike Interval Representations in the Auditory Nerve. *J Neurophysiol* 94: 347–362, 2005.

1.1 ABSTRACT

Harmonic complex tones elicit a pitch sensation at their fundamental frequency (F0), even when their spectrum contains no energy at F0, a phenomenon known as "pitch of the missing fundamental". The strength of this pitch percept depends upon the degree to which individual harmonics are spaced sufficiently apart to be "resolved" by the mechanical frequency analysis in the cochlea. We investigated the resolvability of harmonics of missing-fundamental complex tones in the auditory nerve (AN) of anesthetized cats at low and moderate stimulus levels, and compared the effectiveness of two representations of pitch over a much wider range of F0s (110-3520 Hz) than in previous studies. We found that individual harmonics are increasingly well resolved in rate responses of AN fibers as the characteristic frequency (CF) increases. We obtained rate-based estimates of pitch dependent upon harmonic resolvability by matching harmonic templates to profiles of average discharge rate against CF. These estimates were most accurate for F0s above 400-500 Hz, where harmonics were sufficiently resolved. We also derived pitch estimates from all-order interspike-interval distributions, pooled over our entire sample of fibers. Such interval-based pitch estimates, which are dependent upon phase-locking to the harmonics, were accurate for F0s below 1300 Hz, consistent with the upper limit of the pitch of the missing fundamental in humans. The two pitch representations are complementary with respect to the F0 range over which they are effective; however, neither is entirely satisfactory in accounting for human psychophysical data.

1.2 INTRODUCTION

A harmonic complex tone is a sound consisting of frequency components that are all integer multiples of a common fundamental (F0). The pitch elicited by a harmonic complex tone is normally very close to that of a pure tone at the fundamental frequency, even when the stimulus spectrum contains no energy at that frequency, a phenomenon known as "pitch of the missing fundamental".

Investigating the neural mechanisms underlying the perception of the pitch of harmonic complex tones is of great importance for a variety of reasons. Changes in pitch convey melody in music, and the superposition of different pitches is the basis for harmony. Pitch has an important role in speech, where it carries prosodic features and information about speaker identity. In tone languages such as Mandarin Chinese, pitch also cues lexical contrasts. Pitch plays a major role in auditory scene analysis: differences in pitch are a major cue for sound source segregation, while frequency components that share a common fundamental tend to be grouped into a single auditory object (Bregman 1990; Darwin and Carlyon 1995).

Pitch perception with missing-fundamental stimuli is not unique to humans, but also occurs in birds (Cynx and Shapiro 1986) and non-human mammals (Heffner and Whitfield 1976; Tomlinson and Schwartz 1988), making animal models suitable for studying neural representations of pitch. Pitch perception mechanisms in animals may play a role in processing conspecific vocalizations, which often contain harmonic complex tones.

The neural mechanisms underlying pitch perception of harmonic complex tones have been at the center of a debate among scientists for over a century (Seebeck 1841; Ohm 1843). This debate arises because the peripheral auditory system provides two types of cues to the pitch of complex tones: place cues dependent upon the frequency selectivity and tonotopic mapping of the cochlea, and temporal cues dependent on neural phase locking.

The peripheral auditory system can be thought of as containing a bank of bandpass filters representing the mechanical frequency analysis performed by the basilar membrane. When two partials of a complex tone are spaced sufficiently apart relative to the auditory filter bandwidths, each of them produces an individual local maximum in the spatial pattern of basilar membrane motion. In this case, the two harmonics are said to be "resolved" by the

auditory periphery. On the other hand, when two or more harmonics fall within the pass-band of a single peripheral filter, they are said to be "unresolved". Because the bandwidths of the auditory filters increase with their center frequency, only low-order harmonics are resolved. Based on psychophysical data, the first 6-10 harmonics are thought to be resolved in humans (Bernstein and Oxenham 2003b; Plomp 1964).

When a complex tone contains resolved harmonics, its pitch can be extracted by matching the pattern of activity across a tonotopic neural map to internally stored harmonic templates (Cohen et al. 1994; Goldstein 1973; Terhardt 1974; Wightman 1973). This type of model accounts for many pitch phenomena, including the pitch of the missing fundamental, the pitch shift associated with inharmonic complexes and the pitch ambiguity of complex tones comprising only a few harmonics. However, a key issue in these models is the exact nature of the neural representation upon which the hypothetical template-matching mechanism operates.

Pitch percepts can also be produced by complex tones consisting entirely of unresolved harmonics. In general, though, these pitches are weaker and more dependent on phase relationships among the partials than the pitch based on resolved harmonics (Bernstein and Oxenham 2003b; Carlyon and Shackleton 1994; Houtsma and Smurzynski 1990). With unresolved harmonics, there are no spectral cues to pitch and therefore harmonic template models are not applicable. On the other hand, unresolved harmonics produce direct temporal cues to pitch because the waveform of a combination of unresolved harmonics has a period equal to that of the complex tone. These periodicity cues, which are reflected in neural phase locking, can be extracted by an autocorrelation-type mechanism (Licklider 1951; Meddis and Hewitt 1991; Moore 1990; Yost 1996), which is mathematically equivalent to an all-order interspike interval distribution for neural spike trains. The autocorrelation model also works with resolved harmonics, since the period of the F0 is always an integer multiple of the period of any of the harmonics; this common period can be extracted by combining (e.g. summing) autocorrelation functions from frequency channels tuned to different resolved harmonics (Meddis and Hewitt 1991; Moore 1990).

Previous neurophysiological studies of the coding of the pitch of complex tones in the auditory nerve and cochlear nucleus have documented a robust temporal representation based on pooled interspike interval distributions obtained by summing the interval distributions

from neurons covering a wide range of characteristic frequencies (Cariani and Delgutte 1996a,b; Palmer 1990; Palmer and Winter 1993; Rhode 1995; Shofner 1991). This representation accounts for a wide variety of pitch phenomena, such as the pitch of the missing fundamental, the pitch shift of inharmonic tones, pitch ambiguity, the pitch equivalence of stimuli with similar periodicity, the relative phase invariance of pitch and, to some extent, the dominance of low-frequency harmonics in pitch. Despite its remarkable effectiveness, the autocorrelation model has difficulty in accounting for the greater pitch salience of stimuli containing resolved harmonics compared to stimuli consisting entirely of unresolved harmonics (Bernstein and Oxenham 2003a; Carlyon 1998; Carlyon and Shackleton 1994; Meddis and O'Mard 1997). This issue was not addressed in previous physiological studies because they did not have a means of assessing whether individual harmonics are resolved or not. Moreover, the upper F0 limit over which the interspike-interval representation of pitch is physiologically viable has not been determined. The existence of such a limit is expected due to the degradation in neural phase locking with increasing frequency (Johnson 1980).

In contrast to the wealth of data on the interspike-interval representation of pitch, possible rate-place cues to pitch that might be available when individual harmonics are resolved by the peripheral auditory system have rarely been investigated. The few studies that provide relevant information (Hirahara et al. 1996; Sachs and Young 1979; Shamma 1985a,b) show no evidence for rate-place cues to pitch, even at low stimulus levels where the limited dynamic range of individual neurons is not an issue. The reason for this failure could be that the stimuli used had low fundamental frequencies in the range of human voice (100-300 Hz) and therefore produced few, if any, resolved harmonics in typical experimental animals, which have a poorer cochlear frequency selectivity compared to humans (Shera et al. 2002). Rate-place cues to pitch might be available in animals for complex tones with higher F0s in the range of conspecific vocalizations, which corresponds to about 500-1000 Hz for cats (Brown et al. 1978; Nicastro and Owren 2003; Shipley et al. 1991). This hypothesis is consistent with a report that up to 13 harmonics of a complex tone could be resolved in the rate responses of high-CF units in the cat anteroventral cochlear nucleus (Smoorenburg and Linschoten 1977).

In the present study, we investigated the resolvability of harmonics of complex tones in the cat auditory nerve, and compared the effectiveness of rate-place and interval-based representations of pitch over a much wider range of fundamental frequencies (110-3520 Hz) than in previous studies. We found that the two representations are complementary with respect to the F0 range over which they are effective, but that neither representation is entirely satisfactory in accounting for human psychophysical data. Preliminary reports of our findings have been presented (Cedolin and Delgutte 2003, 2005a).

1.3 MATERIALS AND METHODS

Procedure

Methods for recording from auditory-nerve fibers in anesthetized cats were as described by Kiang et al. (1965) and Cariani and Delgutte (1996a). Cats were anesthetized with Dial in urethane (75 mg/kg), with supplementary doses given as needed to maintain an areflexic state. The posterior portion of the skull was removed and the cerebellum retracted to expose the auditory nerve. The tympanic bullae and the middle-ear cavities were opened to expose the round window. Throughout the experiment the cat was given injections of dexamethasone (0.26 mg/kg) to prevent brain swelling, and Ringer's solution (50 ml/day) to prevent dehydration.

The cat was placed on a vibration-isolated table in an electrically-shielded, temperature-controlled, sound-proof chamber. A silver electrode was positioned at the round window to record the compound action potential (CAP) in response to click stimuli, in order to assess the condition and stability of cochlear function.

Sound was delivered to the cat's ear through a closed acoustic assembly driven by an electrodynamic speaker (Realistic 40-1377). The acoustic system was calibrated to allow accurate control over the sound-pressure level at the tympanic membrane. Stimuli were generated by a 16-bit digital-to-analog converter (Concurrent DA04H) using sampling rates of 20 kHz or 50 kHz. Stimuli were digitally filtered to compensate for the transfer characteristics of the acoustic system.

Spikes were recorded with glass micropipettes filled with 2 M KCl. The electrode was inserted into the nerve and then mechanically advanced using a micropositioner (Kopf 650).

The electrode signal was bandpass filtered and fed to a custom spike detector. The times of spike peaks were recorded with 1-μs resolution and saved to disk for subsequent analysis.

A click stimulus at approximately 55 dB SPL was used to search for single units. Upon contact with a fiber, a frequency tuning curve was measured by an automatic tracking algorithm (Kiang et al. 1970) using 100-ms tone bursts, and the characteristic frequency (CF) was determined. The spontaneous firing rate (SR) of the fiber was measured over an interval of 20 s. The responses to complex-tone stimuli were then studied.

Complex-tone stimuli

Stimuli were harmonic complex tones whose fundamental frequency (F0) was stepped up and down over a two-octave range. The harmonics of each complex tone were all of equal amplitude, and the fundamental component was always missing. Depending on the fiber's CF, one of four pre-synthesized stimuli covering different F0 ranges was selected so that some of the harmonics would likely be resolved (Table 1). For example, for a fiber with a 1760 Hz CF, we typically used F0s ranging from 220 to 880 Hz so that the order of the harmonic closest to the CF would vary from 2 to 8. In each of the four stimuli, the harmonics were restricted to a fixed frequency region as F0 varied (Table I). For each fiber, the stimulus was selected so that the CF fell approximately at the center of the frequency region spanned by the harmonics. In some cases, data were collected from the same fiber in response to two different stimuli whose harmonics spanned overlapping frequency ranges.

Each of the 50 F0 steps (25 up, 25 down) lasted 200 ms, including a 20-ms transition period during which the waveform for one F0 gradually decayed while overlapping with the gradual build up of the waveform for the subsequent F0. Spikes recorded during these transition periods were not included in the analysis. Responses were typically collected over 20 repetitions of the 10-s stimulus (50 steps × 200 ms) with no interruption.

We used mostly low and moderate stimulus levels in order to minimize rate saturation, which would prevent us from accurately assessing harmonic resolvability by the cochlea. Specifically, the sound pressure level of each harmonic was initially set at 15-20 dB above the fiber's threshold for a pure tone at CF, and ranged from 10 to 70 dB SPL with a median of 25 dB SPL. Since our stimuli contain many harmonics, overall stimulus levels are about 5-

10 dB higher than the level of each harmonic, depending on F0. In some cases, responses were measured for two or more stimulus levels differing by 10-20 dB.

In order to compare neural responses to psychophysical data on the phase dependence of pitch, three versions of each stimulus were generated with different phase relationships among the harmonics: cosine phase, alternating (sine-cosine) phase, and negative Schroeder phase (Schroeder 1970). The three stimuli have the same power spectrum and autocorrelation function, but differ in their temporal fine structure and envelope: while the cosine-phase and alternating-phase stimuli have very "peaky" envelopes, the envelope of the Schroeder-phase stimulus is nearly flat (Figure 1.1). Moreover, the envelope periodicity is at F0 for the cosine-phase stimulus, but at 2×F0 for the alternating-phase stimulus. Alternating-phase stimuli have been widely used in previous studies of neural coding (Horst et al. 1990; Palmer and Winter 1992, 1993).

Average-rate analysis

For each step in the F0 sequence, spikes were counted over a 180-ms window extending over the stimulus duration, but excluding the transition period between F0 steps. Spikes counts from the two stimulus segments having the same F0 (from the ascending and descending parts of the F0 sequence) were added together because response to both directions were generally similar. The spike counts were converted to units of discharge rate (spikes/s), and then plotted either as a function of F0 for a given fiber, or as a function of fiber CF for a given F0 to form a "rate-place profile" (Sachs and Young 1979).

In order to assess the statistical reliability of these discharge rate estimates, "bootstrap" resampling (Efron and Tibshirani 1993) was performed on the data recorded from each fiber. One hundred resampled data sets were generated by drawing with replacement from the set of spike trains in response to each F0. Spike counts in the ascending and descending part of the F0 sequence were drawn independently from each other. Spike counts from each bootstrap data set were converted to discharge rate estimates as for the original data, and the standard deviation of these estimates used as an error bar for the mean discharge rate.

Simple phenomenological models were used to analyze average-rate responses to the complex-tone stimuli. Specifically, a single-fiber model was fit to responses of a given fiber as a function of stimulus F0 in order to quantify harmonic resolvability, while a population

model was used to estimate pitch from profiles of average discharge rate against CF for a given F0.

The single-fiber model (Figure 1.2) is a cascade of three stages. The linear bandpass filtering stage, representing cochlear frequency selectivity, is implemented by a symmetric rounded exponential function (Patterson 1976). The Sachs and Abbas (1974) model of rate-level functions is then used to derive the mean discharge rate r from the r.m.s. amplitude p at the ouput of the bandpass filter:

$$r = r_{sp} + r_{dmax} \frac{p^{\alpha}}{p^{\alpha} + p_{50}^{\alpha}} \tag{1}$$

In this expression, r_{sp} is the spontaneous rate, r_{dmax} is the maximum driven rate, and p_{50} is the value of p for which the driven rate reaches half of its maximum value. The exponent α was fixed at 1.77 to obtain a dynamic range of about 20 dB (Sachs and Abbas 1974). The single-fiber model has a total of 5 free parameters: the center frequency and bandwidth of the bandpass filter, r_{sp}, r_{dmax}, and p_{50}. This number is considerably smaller than the 25 F0 values for which responses were obtained in each fiber. The model was fit to the data by the least squares method using the Levenberg-Marquardt algorithm as implemented by Matlab's "lsqcurvefit" function.

The population model is an array of single-fiber models indexed on CF so as to predict the auditory-nerve rate response to any stimulus as a function of cochlear place. The population model has no free parameters; rather, it is used to find the stimulus parameters (F0 and SPL) most likely to have produced the measured rate-place profile, assuming that the spike counts are statistically-independent random variables with Poisson distributions whose expected values are given by the model response at each CF. The resulting maximum-likelihood F0 estimate gives a rate-based estimate of pitch that does not require *a priori* knowledge of the stimulus F0. This strategy effectively implements the concept of "harmonic template" used in pattern recognition models of pitch (Cohen et al. 1994; Goldstein 1973; Terhardt 1974; Wightman 1973): here, the template is the model response to a harmonic complex tone with equal amplitude harmonics. In practice, the maximum-likelihood F0 estimate was obtained by computing the model responses to complex tones covering a wide range of F0 in fine increments (0.1%) and finding the F0 value that maximizes the likelihood of the data.

While the population model has no free parameters, five fixed (i.e. stimulus-independent) parameters still need to be specified for each fiber in the modeled population. These parameters were selected so as to meet two separate requirements: (1) the model's normalized driven rate must vary smoothly with CF, (2) the model must completely specify the Poisson distribution of spike counts for each fiber so as to be able to apply the maximum-likelihood method. To meet these requirements, three of the population-model parameters were directly obtained from the corresponding parameters for the single-fiber model: the center frequency of the bandpass filter (effectively the CF), the spontaneous rate $r_{sp,}$, and the maximum driven rate r_{dmax}. The sensitivity parameter p_{50} in the population model was set to the median value of this parameter over our fiber sample. Finally, the bandwidth of the bandpass filter was derived from its center frequency by assuming a power law relationship between the two (Shera et al. 2002). The parameters of this power function were obtained by fitting a straight line in double logarithmic coordinates to a scatter plot of filter bandwidth against center frequency for our sample of fibers.

Interspike-interval analysis

As in previous studies of the neural coding of pitch (Cariani and Delgutte 1996a, b; Rhode 1995), we derived pitch estimates from pooled interspike interval distributions. The pooled interval distribution is the sum of the all-order interspike interval distributions for all the sampled auditory-nerve fibers, and is closely related to the summary autocorrelation in the Meddis and Hewitt (1991) model. The single-fiber interval distribution (bin width 0.1 ms) was computed for each F0 using spikes occurring in the same time window as used in the rate analysis.

To derive pitch estimates from pooled interval distributions, we used "periodic templates" that select intervals at a given period and its multiples. Specifically, we define the *contrast ratio* of a periodic template as the ratio of the weighted mean number of intervals for bins within the template to the weighted mean number of intervals per bin in the entire histogram. The estimated pitch period is the period of the template that maximizes the contrast ratio. In computing the contrast ratio, each interval is weighted by an exponentially decaying function of its length in order to give greater weight to short intervals. This weighting implements the idea that the lower F0 limit of pitch at about 30 Hz (Pressnitzer et

al. 2001) implies that the auditory system is unable to use very long intervals in forming pitch percepts. A 3.6-ms decay time constant was found empirically to minimize the number of octave and sub-octave errors in pitch estimation. The statistical reliability of the pitch estimates was assessed by generating 100 bootstrap replications of the pooled interval distribution (using the same resampling techniques as in the rate analysis), and computing a pitch estimate for each bootstrap replication.

1.4 RESULTS

Our results are based on 122 measurements of responses to harmonic complex tones recorded from 75 auditory-nerve fibers in two cats. Of these, 54 had high SR (> 18 spikes/s), 10 had low SR (< 0.5 spike/s), and 11 had medium SR. The CFs of the fibers ranged from 450 to 9200 Hz. We first describe the rate-responses of single fibers as a function of F0 to characterize harmonic resolvability. We then derive pitch estimates from both rate-place profiles and pooled interspike interval distributions, and quantify the accuracy and precision of these estimates as a function of F0.

Single-fiber cues to resolved harmonics

Figure 1.3 shows the average discharge rate as a function of complex-tone F0 (harmonics in cosine phase) for two auditory-nerve fibers with CFs of 952 Hz (panel A) and 4026 Hz (panel B), respectively. Data are plotted against the dimensionless ratio of fiber CF to stimulus F0, which we call *harmonic number* (lower horizontal axis). Because this ratio varies inversely with F0, F0 increases from right to left along the upper axis in these plots. The harmonic number takes an integer value when the CF coincides with one of the harmonics of the stimulus, while it is an odd integer multiple of 0.5 (2.5, 3.5, etc...) when the CF falls halfway between two harmonics. Thus, resolved harmonics should appear as peaks in firing rate for integer values of the harmonic number, with valleys in between. This prediction is verified for both fibers at lower values of the harmonic number (higher F0s), although, the oscillations are more pronounced and extend to higher harmonic numbers for the high-CF fiber than for the low-CF fiber. This observation is consistent with the higher

quality factor (Q = CF/Bandwidth) of high-CF fibers compared to low-CF fibers (Kiang et al. 1965; Liberman 1978).

In order to quantify the range of harmonics that can be resolved by each fiber, a simple peripheral auditory model was fit to the data (see Methods). For both fibers, the response of the best-fitting model (solid lines in Fig. 1.3) captures the oscillatory trend in the data. The rate of decay of these oscillations is determined by the bandwidth of the bandpass filter representing cochlear frequency selectivity in the model (Fig. 1.2). The harmonics of F0 are considered to be resolved so long as the oscillations in the fitted curve exceed two typical standard errors of the discharge rate obtained by bootstrapping (gray shading). The maximum resolved harmonic number N_{max} is 4.1 for the low-CF fiber, smaller than N_{max} for the high-CF fiber (6.3). The ratio CF/N_{max} gives $F0_{min}$, the lowest fundamental frequency for which harmonics are resolved in a fiber's rate response. In the examples of Fig. 1.3, $F0_{min}$ is 232 Hz for the low-CF fiber, and 639 Hz for the high-CF fiber.

Figure 1.4 shows how $F0_{min}$ varies with CF for our entire sample of fibers. To be included in this plot, the variance of the residuals after fitting the single-fiber model to the data had to be significantly smaller ($p < 0.05$, F-test) than the variance of the raw data so that N_{max} (and therefore $F0_{min}$) could be reliably estimated. Thirty five out of 122 measurements were thus excluded; 23 of these had CFs below 2000 Hz. On the other hand, the figure includes data from fibers (shown by triangles) for which $F0_{min}$ was bounded by the lowest F0 presented and was therefore overestimated. $F0_{min}$ increases systematically with CF, and the increase is well fit by a power function with an exponent of 0.63 (solid line). This increase is consistent with the increase in tuning curve bandwidths with CF (Kiang et al. 1965).

Rate responses of AN fibers to complex stimuli are known to depend strongly on stimulus level (Sachs and Young 1979). The representation of resolved harmonics in rate responses is expected to degrade as the firing rates become saturated. The increase in cochlear filter bandwidths with level may further degrade harmonic resolvability. We were able to reliably fit single-fiber models to the rate-F0 data for stimulus levels as high as 38 dB above the threshold at CF per component. In general, this limit increased with CF, from roughly 15 dB above threshold for CFs below 1 kHz to about 30 dB above threshold for CFs above 5 kHz. No obvious dependence of this limit on fiber's spontaneous rate was noticed.

To more directly address the level dependence of responses, we held 24 fibers long enough to record the responses to harmonic complex tones at two or more stimulus levels differing by 10-20 dB. In 23 of these 24 cases, the maximum resolved harmonic number N_{max} decreased with increasing level. One example is shown in Fig. 1.5 for a fiber with CF at 1983 Hz. For this fiber, N_{max} decreased from 7.1 at 20 dB SPL to 4.9 at 30 dB SPL.

The observed decrease in N_{max} with level could reflect either broadened cochlear tuning or rate saturation. To distinguish between these two hypotheses, two versions of the single-fiber model were compared when data were available at two stimulus levels. In one version, all the model parameters were constrained to be the same at both levels; in the other version, the bandwidth of the bandpass filter representing cochlear frequency selectivity was allowed to vary with level. The variable-bandwidth model is guaranteed to fit the data better (in a least squares sense) than the fixed-bandwidth model because it has an additional free parameter. However an F-test for the ratio of the variances of the residuals revealed no statistically significant difference between the two models at the 0.05 level for any of the 24 fibers, meaning than the additional free parameter of the variable-bandwidth model gave it only an insignificant advantage over the fixed-bandwidth model. This result suggests that rate saturation, which is present in both models, may be the main factor responsible for the decrease in N_{max} with stimulus level.

Pitch estimation from rate-place profiles

Having characterized the limits of harmonic resolvability in rate responses of auditory-nerve fibers, the next step is to determine how accurately pitch can be estimated from rate-place cues to resolved harmonics. For this purpose, we fit harmonic templates to profiles of average discharge rate against CF, and derive pitch estimates by the maximum likelihood method, assuming that the spike counts from each fiber are random variables with statistically-independent Poisson distributions. In our implementation, a harmonic template is the response of a peripheral auditory model to a complex tone with equal-amplitude harmonics. The estimated pitch is therefore the F0 of the complex tone most likely to have produced the observed response if the stimulus-response relationship were defined by the model.

Figure 1.6 shows the normalized driven discharge rate of AN fibers as a function of CF in response to two complex tones (harmonics in cosine phase) with F0s of 541.5 Hz (Panel A) and 1564.4 Hz (Panel C). The rate is normalized by subtracting the spontaneous rate and dividing by the maximum driven rate (Sachs and Young 1979), and these parameters are estimated by fitting the single-fiber model to the rate-F0 data. As for the single-fiber responses in Fig. 1.3 and 1.5, responses are plotted against the dimensionless harmonic number CF/F0, with the difference that F0 is now fixed while CF varies, instead of the opposite. Resolved harmonics should again result in peaks in firing rate at integer values of the harmonic number. Despite considerable scatter in the data, this prediction is verified for both F0s, although the oscillations are more pronounced for the higher F0. Many factors are likely to contribute to the scatter, including the threshold differences among fibers with the same CF (Liberman 1978), pooling data from two animals, intrinsic variability in neural responses, and inaccuracies in estimating the minimum and maximum discharge rates used in computing the normalized rate.

The solid lines in Fig. 1.6A and 1.6C show the normalized rate response of the population model to the complex tone whose F0 maximizes the likelihood, i.e. the best fitting harmonic template. Note that, while the figure shows the normalized model response, *unnormalized* rates (actually, spike counts) are used when applying the maximum likelihood method because only spike counts have the integer values required for a Poisson distribution. For both F0s, the model response shows local maxima near integer values of the harmonic number, indicating that the pitch estimates are very close to the stimulus F0s. This point is shown more precisely in Fig. 1.6B and 1.6D, which show the log likelihood of the model response as a function of template F0. Despite the very moderate number of data points in the rate-place profiles, for both F0s the likelihood shows a sharp maximum when the template F0 is very close to the stimulus F0. For the complex tone with F0 at 541.5 Hz, the estimated pitch is 554.2 Hz, about 2% above the actual F0. For the 1564.4 Hz tone, the estimated pitch is 1565.7 Hz, only 0.1% above the actual F0. The likelihood functions also show secondary maxima for template F0s that form ratios of small integers (e.g. 4/3, 3/4) with respect to the stimulus F0. However, because these secondary maxima are much lower than the absolute maximum, the estimated pitch is highly unambiguous, consistent with

psychophysical observations for complex tones containing many harmonics (Houtsma and Smurzynski 1990).

To assess the reliability of the maximum-likelihood pitch estimates, estimates were computed for 100 bootstrap resamplings of the data for each F0 (see Method). Figure 1.7A shows the median absolute estimation error of these bootstrap estimates as a function of F0 for complex tones with harmonics in cosine phase. With few exceptions, median pitch estimates only deviate by a few percent from the stimulus F0 above 500 Hz. Larger deviations are more common for lower F0s. The number and CF distribution of the fibers had to meet certain constraints for each F0 to be included in the figure because, to reliably estimate F0, the sampling of the CF axis has to be sufficiently dense to capture the harmonically-related oscillations in the rate-CF profiles. This is why Fig. 1.7 shows no estimates for F0s below 220 Hz and for a small subset of F0s (12 out of 56) above 220 Hz.

In order to quantify the salience of the rate-place cues to pitch as a function of F0, we used the *Fisher Information*, which is the expected value of the curvature of the log likelihood function, evaluated at its maximum. The expected value was approximated by averaging the likelihood function over 100 bootstrap replications of the rate-place data. A steep curvature means that the likelihood varies fast with template F0, and therefore that the F0 estimate is very reliable. The Fisher Information was normalized by the number of data points in the rate profile for each F0 to allow comparisons between data sets of different size. Fig. 1.7B shows that the Fisher Information increases monotonically with F0 up to about 1000 Hz and then remains essentially constant. Overall, pitch estimation from rate-place profiles works best for F0s above 400-500 Hz, although reliable estimates were obtained for F0s as low as 250 Hz.

Harmonic templates were fit to rate-place profiles obtained in response to complex tones with harmonics in alternating phase and in Schroeder phase as well as in cosine phase in order to test whether the pitch estimates depend on phase. Figure 1.8 shows an example for an F0 of 392 Hz. The numbers of data points differ somewhat for the three phase conditions because we could not always "hold" a unit sufficiently long to measure responses to all three conditions. Despite these sampling differences, the pitch estimates for the three phase conditions are similar to each other (Panels A-C) and similar to the pitch estimate obtained by combining data across all three phase conditions (Panel D).

We devised a statistical test for the effect of phase on pitch estimates. This test compares the likelihoods of the rate-place data given two different models. In one model, the estimated pitch is constrained to be the same for all three phase conditions by finding the F0 value that maximizes the likelihood of the combined data (Fig. 1.8D). In the other model, a maximum likelihood pitch estimate is obtained separately for each phase condition (Fig. 1.8A-C), and then the maximum log likelihoods are summed over the three conditions, based on the assumption that the 3 data sets are statistically-independent. If the rate-place patterns for the different phases differed appreciably, the maximum likelihood for the phase-dependent model should be higher than that for the phase-independent model because the phase-dependent model has the additional flexibility of fitting each data set separately. Contrary to this expectation, when the two models were fit to 1000 bootstrap replications of the rate-place data, the distributions of the maximum likelihoods for the two models did not significantly differ ($p = 0.178$), indicating that the additional free parameters of the phase-dependent model offer no significant advantage for this F0.

This test was performed for three different values of F0 (612 Hz, 670 Hz and 828 Hz) in addition to the 392 Hz case shown in Fig. 1.8[1]. In three of these four cases, the results were as in Fig. 1.8 in that the differences in maximum likelihoods for the two models did not reach statistical significance ($p < 0.05$). For 612 Hz, the comparison did reach significance ($p = 0.007$), but for this F0 the rate-place profiles for harmonics in alternating and Schroeder phase showed large gaps in the distribution of data points over harmonics numbers, making the reliability of the F0-estimates for these two phases questionable. When the actual pitch estimates for the different phase conditions were compared, there was no clear pattern to the results across F0s, i.e. the pitch estimate for any given phase condition could be the largest in one case and the smallest in another case. These results indicate that phase relationships among the partials of a complex tone do not seem to greatly influence the pitch estimated from rate-place profiles, consistent with psychophysical data on the phase invariance of pitch based on resolved harmonics (Houtsma and Smurzynski 1990).

[1] A programming error resulted in erroneous phase relationships among the harmonics for most F0s. We only performed the test for the four F0 values for which the phases were correct.

Pitch estimation from pooled interspike interval distributions

Pitch estimates were derived from pooled interspike interval distributions to compare the accuracy of these estimates with that of rate-place estimates for the same stimuli. Figure 1.9A-B show pooled all-order interspike interval distributions for two complex-tone stimuli with F0s of 320 and 880 Hz (harmonics in cosine phase). For both F0s, the pooled distributions show modes at the period of F0 and its integer multiples. However, these modes are less prominent at the higher F0 for which only the first few harmonics are located in the range of robust phase locking.

In previous work (Cariani and Delgutte 1996a,b; Palmer 1990; Palmer and Winter 1993), the pitch period was estimated from the location of the largest maximum (mode) in the pooled interval distribution. This simple method is also widely used in autocorrelation models of pitch (Meddis and Hewitt 1991; Yost 1982, 1996). However, when tested over a wide range of F0s, this method was found to yield severe pitch estimation errors for two reasons. First, for higher F0s as in Fig. 1.9B, the first interval mode near 1/F0 is always smaller than the modes at integer multiples of the fundamental period, due to the neural relative refractory period. Moreover, the location of the first mode is slightly but systematically delayed with respect to the period of F0 (Fig. 1.9B), an effect also attributed to refractoriness (McKinney and Delgutte 1999; Ohgushi 1983). In fact, a peak at the pitch period is altogether lacking if the period is shorter than the absolute refractory period, about 0.6 ms for auditory-nerve fibers (McKinney and Delgutte 1999). These difficulties at higher F0s might be overcome by using shuffled autocorrelograms (Louage et al. 2004) which, unlike conventional autocorrelograms, are not distorted by neural refractoriness. However, a more fundamental problem is that, at lower F0s, the modes at 1/F0 and its multiples all have approximately the same height (Fig. 1.9A) so that, due to the intrinsic variability in neural responses, some of the later modes will unavoidably be larger than the first mode in many cases, and therefore lead to erroneous pitch estimates at integer submultiples of F0.

We therefore modified our pitch estimation method to make use of all pitch-related modes in the pooled interval distribution rather than just the first one. Specifically, we used periodic templates that select intervals at a given period and its multiples, and determined the template F0 which maximizes the *contrast ratio,* a signal-to-noise ratio measure of the number of intervals within the template relative to the mean number of intervals per bin (see

Method). When computing the contrast ratio, short intervals were weighted more than long intervals according to an exponentially decaying weighting function of interval length. This weighting implements the psychophysical observation of a lower limit of pitch near 30 Hz (Pressnitzer et al. 2001) by preventing long intervals to contribute significantly to pitch. Fig. 1.9C-D show the template contrast ratio as a function of template F0 for the same two stimuli as on top. For both stimuli, the contrast ratio reaches an absolute maximum when the template F0 is very close to the stimulus F0, although the peak contrast ratio is larger for the lower F0. The contrast ratio also shows local maxima one octave above and below the stimulus F0. In Fig. 1.9C, these secondary maxima are small relative to the main peak at F0, but in Fig. 1.9D the maximum at F0/2 is almost as large as the one at F0. Despite the close call, F0 was correctly estimated in both cases of Fig. 1.9 and, overall, our pitch estimation algorithm produced essentially no octave or sub-octave errors over the entire range of F0 investigated (110-3520 Hz).

Figure 1.10 shows measures of the accuracy and strength of the interval-based pitch estimates as a function of F0 for harmonics in cosine phase. The accuracy measure is the median absolute value of the pitch estimation error over bootstrap replications of the pooled interval distributions. The estimates are highly accurate below 1300 Hz, where their medians are within 1-2% of the stimulus F0 (panel A). However, the interval-based estimates of pitch abruptly break down near 1300 Hz. While the existence of such an upper limit is consistent with the degradation in phase locking at high frequencies, the location of this limit at 1300 Hz is low compared to the 4-5 kHz upper limit of phase locking, a point to which we return in the Discussion.

Fig. 1.10B shows a measure of the strength of the estimated pitch, the contrast ratio of the best-fitting periodic template, as a function of F0. The contrast ratio is largest below 500 Hz, then decreases gradually with increasing F0, to reach essentially unity (meaning a flat interval distribution) at 1300 Hz. For F0s above 1300 Hz the modes in the pooled interval distribution essentially disappear into the noise floor. Thus, the strength of interval-based estimates of pitch is highest in the F0 range where rate-based pitch estimates are the least reliable due to the lack of strongly resolved harmonics. Conversely, rate-based estimates of pitch become increasingly strong in the range of F0s where the interval-based estimates break down.

For a few F0s, interval-based estimates of pitch were derived for complex tones with harmonics in alternating phase and in Schroeder phase as well as for harmonics in cosine phase. Figure 1.11 compares the pooled all-order interval distributions in the three phase conditions for two F0s, 130 Hz (left) and 612 Hz (right). Based on the rate-place results, the harmonics of the 130 Hz F0 are not resolved, while some of the harmonics of the 612 Hz F0 are resolved. This is because we obtained a reliable pitch estimate based on rate-place profiles at 612 Hz, but not at 130 Hz (Fig. 1.7).

For both F0s and all phase conditions, the distributions have modes at the period of the fundamental frequency and its integer multiples (Fig. 1.11 A-C, E-G). For the lower F0 (130 Hz), the pooled interval distribution for harmonics in alternating phase (panel B) also shows secondary peaks at half the period of F0 and its odd multiples (arrows), reflecting the periodicity of the stimulus envelope at 2×F0 (Fig 1.1). Such frequency doubling has previously been observed in auditory-nerve fiber responses to alternating-phase stimuli using both period histograms (Horst et al. 1990; Palmer and Winter 1992) and autocorrelograms (Palmer and Winter 1993). At 130 Hz, the pooled interval distribution for harmonics in negative Schroeder phase (panel C) shows pronounced modes at the period of F0 and its multiples and strongly resembles the interval distribution for harmonics in cosine phase (panel A), even though the waveforms of the two stimuli have markedly different envelopes (Fig 1.1).

The interval-based pitch estimates are nearly identical for all three phase conditions, but the maximum contrast ratio is substantially lower for harmonics in alternating phase than for harmonics in cosine or in Schroeder phase (Fig. 1.11D). In addition, for harmonics in alternating phase, the contrast ratio of the periodic template at the envelope frequency 2×F0 is almost as large as the contrast ratio at F0. In contrast, for the higher F0 (612 Hz), there are no obvious differences between phase conditions in the pooled all-order interval distributions (Fig. 1.11E-G). In particular, the secondary peaks at half the period of F0, which were found at 130 Hz for the alternating-phase stimulus, are no longer present at 612 Hz. Moreover, the maximum contrast ratios are essentially the same for all three phase conditions (Fig. 1.11H).

Overall, these results show that, while phase relationships among harmonics have little effect on the pitch values estimated from pooled interval distributions, which are always close to the stimulus F0, the salience of these estimates can be significantly affected by phase

when harmonics are unresolved. These results are consistent with psychophysical results showing a greater effect of phase on pitch and pitch salience for stimuli consisting of unresolved harmonics than for stimuli containing resolved harmonics (Houtsma and Smurzynski 1990; Shackleton and Carlyon 1994). However, these results fail to account for the observation that the dominant pitch is often heard at the envelope frequency 2×F0 for unresolved harmonics in alternating phase.

1.5 DISCUSSION

Harmonic resolvability in auditory-nerve fiber responses

We examined the response of cat auditory-nerve fibers to complex tones with a missing fundamental and equal-amplitude harmonics. We used low and moderate stimulus levels (15-20 dB above threshold) to minimize rate saturation which would prevent us from accurately assessing cochlear frequency selectivity and therefore harmonic resolvability from rate responses. In general, the average-rate of a single auditory-nerve fiber was stronger when its CF was near a low-order harmonic of a complex tone than when the CF fell halfway in between two harmonics (Fig. 1.3). This trend could be predicted using a phenomenological model of single-fiber rate responses incorporating a bandpass filter representing cochlear frequency selectivity (Fig. 1.2). The amplitude of the oscillations in the response of the best-fitting single-fiber model, relative to the typical variability in the data, gave an estimate of the lower F0 of complex tones whose harmonics are resolved at a given CF (Fig. 1.3). This limit, which we call $F0_{min}$, increases systematically with CF and this increase is well fit by a power function with an exponent of 0.63 (Fig. 1.4). That the exponent is less than 1 is consistent with the progressive sharpening of peripheral tuning with increasing CF when expressed as a Q factor, the ratio CF/Bandwidth. The exponent for Q would be 0.37, which closely matches the 0.37 exponent found by Shera et al. (2002) for the CF dependence of Q_{10} in pure-tone tuning curves from AN fibers in the cat.

Our definition of the lower limit of resolvability $F0_{min}$ is to some extent arbitrary because it depends on the variability in the average discharge rates, which in turn depends on the number of stimulus repetitions and the duration of the stimulus. Nevertheless, our results are

consistent with those of Wilson and Evans (1971) for auditory-nerve fibers in the guinea pig using ripple noise (a.k.a. comb-filtered noise), a stimulus with broad spectral maxima at harmonically-related frequencies. These authors found that the number of such maxima that can be resolved in the rate responses of single fibers (equivalent to our N_{max}) increases with CF from 2-3 at 200 Hz to about 10 at 10 kHz and above. Similarly, Smoorenburg and Linschoten (1977) report that the number of harmonics of a complex tone that are resolved in the rate responses of single units in the cat anteroventral cochlear nucleus (AVCN) increases from 2 at 250 Hz to 13 at 10 kHz. Despite the different metrics used to define resolvability, both studies are in good agreement with the data of Fig. 1.4 if we use the conversion $F0_{min} = CF/N_{max}$.

Consistent with a previous report for AVCN neurons (Smoorenburg and Linschoten 1977), we found that the ability of auditory-nerve fibers to resolve harmonics in their rate response degrades rapidly with increasing stimulus level (Fig. 1.5). This degradation could be due either to the broadening of cochlear tuning with increasing level, or to saturation of the average rate. Saturation seems to be the most likely explanation because a single-fiber model with level-dependent bandwidth did not fit the data significantly better than a model with fixed bandwidth. However, the level dependence of cochlear filter bandwidths might have a greater effect on responses to complex tones if level were varied over a wider range than the 10-20 dB used here (Cooper and Rhode 1997; Ruggero et al. 1997).

Rate-place representation of pitch

A major finding is that the pitch of complex tones could be reliably and accurately estimated from rate-place profiles for fundamental frequencies above 400-500 Hz by fitting a harmonic template to the data (Fig. 1.6 and 1.7A,B). The harmonic template was implemented as the response of a simple peripheral auditory model to a harmonic complex tone with equal-amplitude harmonics, and the estimated pitch was the F0 of the complex tone most likely to have produced the rate-place data assuming that the stimulus-response relationship is characterized by the model. Despite the non-uniform sampling of CFs and the moderate number of fibers sampled at each F0 (typically 20-40), these pitch estimates were accurate within a few percent.

Pitch estimation became increasingly less reliable for F0s below 400-500, with large estimation errors becoming increasingly common. Nevertheless, some reliable estimates could be obtained for F0s as low as 250 Hz. This result is consistent with the failure of previous studies to identify rate-place cues to pitch in auditory-nerve responses to harmonic complex tones with F0s below 300 Hz (Hirahara et al. 1996; Sachs and Young 1979; Shamma 1985a,b), although Hirahara et al. did find a weak representation of the first 2-3 harmonics in rate-place profiles for vowels with an F0 at 350 Hz.

In interpreting these results, it is important to keep in mind that the precision of the rate-based pitch estimates depends on many factors such as the number of fibers sampled, the CF distribution of the fibers, pooling of data from two animals, the number of stimulus repetitions, and the particular method for fitting harmonic templates. For example, since the lowest CF sampled was 450 Hz, the second harmonic and, in some cases, the third could not be represented in the rate-place profiles for F0s below 220 Hz, possibly explaining why we never obtained a reliable pitch estimate in that range. In fact, since our stimuli had missing fundamentals, we cannot rule out that the fundamental might always be resolved when it is present.

In one respect, our method may somewhat overestimate the accuracy of the rate-based pitch estimates because we only included data from measurements for which the rate response as a function of F0 oscillated sufficiently to be able to reliably fit a single-fiber model. This constraint was necessary because, for responses that do not oscillate, we could not reliably estimate the minimum and maximum discharge rates which are essential in fitting harmonic templates to the rate-place data. Thirty five of 122 responses were thus excluded. Because our design minimizes rate saturation, and because 23 or these 35 excluded responses were from fibers with CFs below 2 kHz, we infer that insufficient frequency selectivity for resolving harmonics rather than rate saturation was the primary reason for the lack of F0-related oscillations in these measurements.

A factor whose effect on pitch estimation performance is hard to evaluate is that the rate-place profiles included responses to stimuli presented at different sound levels. At first sight, pooling data across levels might seem to increase response variability and therefore decrease estimation performance. However, because the stimulus level was usually selected to be 15-20 dB above the threshold of each fiber so that responses would be robust without being

saturated, our procedure might actually have reduced the variability due to threshold differences among fibers. The rationale for this procedure is that an optimal central processor would focus on unsaturated fibers because these fibers are the most informative. Because level re. threshold rather than absolute level is the primary determinant of rate responses, we are effectively invoking a form of the "selective listening hypothesis" (Delgutte 1982; Delgutte 1987; Lai et al. 1994), according to which the central processor attends to low-threshold, high spontaneous rate fibers at low levels, and to high-threshold, low-spontaneous fibers at high levels.

Our harmonic template differs from those typically used in pattern recognition models of pitch in that it has very broad peaks at the harmonic frequencies. Most pattern recognition models (Duifhuis et al. 1982; Goldstein 1973; Terhardt 1974) use very narrow templates or "sieves", typically a few percent of each harmonic's frequency. One exception is the Wightman (1973) model, which effectively uses broad cosinusoidal templates by performing a Fourier transform operation on the spectrum. Our method also resembles the Wightman model, and differs from the other models in that it avoids an intermediate, error-prone stage which estimates the frequencies of the individual resolved harmonics; rather, a global template is fit to the entire rate-place profile. Broad templates are well adapted to the measured rate-place profiles because the dips between the harmonics are often sharper than the peaks at the harmonic frequencies (Figs. 6 and 8). On the other hand, the templates are the response of the peripheral model to complex tones with equal-amplitude harmonics, which exactly match the stimuli that were presented. It remains to be seen how well such templates would work when the spectral envelope of the stimulus is unknown, or when the amplitudes of the individual harmonics are roved from trial to trial, conditions which cause little degradation in psychophysical performance (Bernstein and Oxenham 2003a; Houtsma and Smurzynski 1990).

Given the uncertainties about the various factors discussed above may affect our pitch estimation procedure, a comparison of the pitch estimation performance with psychophysical data should focus on robust overall trends as a function of stimulus parameters rather than on absolute measures of performance. Both the precision of the pitch estimates (Fig. 1.7A) and their salience (as measured by the Fisher information; Fig. 1.7B), improve with increasing F0 as the harmonics of the complex become increasingly resolved. This result is in agreement

with psychophysical observations that both pitch strength and pitch discrimination performance improve as the degree of harmonic resolvability increases (Bernstein and Oxenham 2003b; Carlyon and Shackleton 1994; Houtsma and Smurzynski 1990; Plomp 1967; Ritsma 1967). However, the continued increase in Fisher information with F0 beyond 1000 Hz conflicts with the existence of an upper limit to the pitch of missing-fundamental stimuli, which occurs at about 1400 Hz in humans (Moore 1973b). This discrepancy between the rapid degradation in pitch discrimination at high frequencies and the lack of a concomitant degradation in cochlear frequency selectivity is a general problem for place models of pitch perception and frequency discrimination (Moore 1973a).

We also found that the relative phases of the resolved harmonics of a complex tone do not greatly influence rate-based estimates of pitch (Fig. 1.8). This result is consistent with expectations for a purely place representation of pitch, as well as with psychophysical results for stimuli containing resolved harmonics (Houtsma and Smurzynski 1990; Shackleton and Carlyon 1994; Wightman 1973).

The restriction of our data to low and moderate stimulus levels raises the question of whether the rate-place representation of pitch would remain robust at the higher stimulus levels typically used in speech communication or when listening to music. Previous studies have used signal detection theory to quantitatively assess the ability of rate-place information in the auditory nerve to account for behavioral performance in tasks such as intensity discrimination (Viemeister 1988; Delgutte 1987; Winslow and Sachs 1988; Winter and Palmer 1991; Colburn et al. 2003) and formant-frequency discrimination for vowels (Conley and Keilson 1995; May et al. 1996). These studies give a mixed message. On the one hand, the rate-place representation generally contains sufficient information to account for behavioral performance up to the highest sound levels tested. On the other hand, because the fraction of high-threshold fibers is small compared to low-threshold fibers, predicted performance of optimal processor models degrades markedly with increasing level, while psychophysical performance remains stable. Thus, while a rate-place representation cannot be ruled out, it fails to account for a major trend in the psychophysical data. Extending this type of analysis to pitch discrimination for harmonic complex tones is beyond the scope of this paper. Given the failure of the rate-place representation to account for the level dependence of performance in the other tasks, a more productive approach may be to explore

alternative spatio-temporal representations that would rely on harmonic resolvability like the rate-place representation, but would be more robust with respect to level variations by exploiting phase locking (Shamma, 1985a, Heinz et al. 2001). Preliminary tests of one such spatio-temporal representation are encouraging (Cedolin and Delgutte 2005b).

Interspike-interval representation of pitch

Our results confirm previous findings (Cariani and Delgutte 1996a,b; Palmer 1990; Palmer and Winter 1993), that fundamental frequencies of harmonic complex tones are precisely represented in pooled all-order interspike-interval distributions of the auditory nerve. These interval distributions have prominent modes at the period of F0 and its integer multiples (Fig. 1.9A, B). Pitch estimates derived using periodic templates that select intervals at a given period and its multiples were highly accurate (often within 1%) for F0s up to 1300 Hz (Fig. 1.10). The determination of this upper limit to the interval-based representation of pitch is a new finding. Moreover, the use of periodic templates for pitch estimation improves upon the traditional method of picking the largest mode in the interval distribution by greatly reducing sub-octave errors.

While the existence of an upper limit to the representation of pitch in interspike intervals is expected from the degradation in phase locking at high frequencies, the location of this limit at 1300 Hz is low compared to the usually quoted 4-5 kHz limit of phase locking in the auditory nerve (Johnson 1980; Rose et al. 1967). Of course, both the limit of pitch representation and the limit of phase locking depend to some extent on the signal-to-noise ratio of the data, which in turn depends on the duration of stimulus, the number of stimulus repetitions and, for pooled interval distributions, the number of sampled fibers. However the discrepancy between the two limits appears too large to be entirely accounted by differences in signal-to-noise ratio. Fortunately, the discrepancy can be largely reconciled by taking into account harmonic resolvability and the properties of our stimuli. For F0s near 1300 Hz, all the harmonics within the CF range of our data (450-9200 Hz) are well resolved (Fig. 1.4), so that information about pitch in pooled interval distributions must depend on phase locking to individual resolved harmonics rather than on phase locking to the envelope generated by interactions between harmonics within a cochlear filter. Moreover, because our stimuli have missing fundamentals, an unambiguous determination of pitch from the pooled distribution

requires phase locking to at least two resolved harmonics. As F0 increases above 1300 Hz, the third harmonic (3900 Hz) begins to exceed the upper limit of phase locking, leaving only ambiguous pitch information and therefore leading to severe estimation errors.

A major finding is that the range of F0s over which interval-based estimates of pitch are reliable roughly covers the entire human perceptual range of the pitch of missing-fundamental stimuli, which extends up to 1400 Hz for stimuli containing many harmonics (Moore 1973b). It is widely recognized that the upper limit of phase locking to pure tones matches the limit in listeners' ability to identify musical intervals (Semal and Demany 1990; Ward 1954). The present results extend this correspondence to complex tones with missing F0.

Our results predict that pitch based on pooled interval distributions is strongest for F0s below 400 Hz (Fig. 1.10B), a range for which the decreased effectiveness of pitch estimation based on rate-place information implies that individual harmonics are poorly resolved in the cat. Thus, the interval-based representation of pitch seems to have trouble predicting the greater salience of pitch based on resolved harmonics compared to that based on unresolved harmonics (Shackleton and Carlyon 1994; Bernstein and Oxenham 2003b). This conclusion based on physiological data supports similar conclusions previously reached for autocorrelation models of pitch perception (Bernstein and Oxenham 2003a; Carlyon 1998; Meddis and O'Mard 1997).

We found that neither the pitch values nor the pitch strength estimated from pooled interval distributions depend on the phase relationships among the harmonics at higher F0s where some harmonics are well resolved (Fig. 1.11E-G). This finding is consistent with psychophysical observations on the phase invariance of pitch and pitch salience for stimuli containing resolved harmonics (Carlyon and Shackleton 1994; Houtsma and Smurzynski 1990; Wightman 1973). However, pitch salience and, in some cases, pitch values do depend on phase relationships for stimuli consisting of unresolved harmonics (Houtsma and Smurzynski 1990; Lundeen and Small 1984; Ritsma and Engel 1964). In particular, for stimuli in alternating sine-cosine phase, the pitch often matches the envelope periodicity at 2×F0 rather than the fundamental (Lundeen and Small 1984). Consistent with previous studies (Horst et al. 1990; Palmer and Winter 1992; Palmer and Winter 1993), we found a correlate of this observation in the pooled interval distributions in that our interval-based

measure of pitch strength was almost as large at the envelope frequency 2×F0 as at the fundamental F0 for alternating-phase stimuli with unresolved harmonics (Fig 11D). Despite this frequency doubling, the pitch values estimated from interval distributions were always at F0 and never at 2×F0, in contrast to psychophysical judgments for unresolved harmonics in alternating phase (Lundeen and Small 1984). Thus, pitch estimation based on interspike intervals does not seem to be sufficiently sensitive to the relative phases of unresolved harmonics compared to psychophysical data. A similar conclusion has been reached for the autocorrelation model of pitch, and modifications to the model have been proposed in part to handle this difficulty (de Cheveigné 1998; Patterson and Holdsworth 1994).

Vocalizations and pitch perception

A widely held view is that pitch perception for harmonic complex tones is closely linked to the extraction of biologically relevant information from conspecific vocalizations including human speech. For example, Terhardt (1974) argues: "The virtual pitch cues can be generated only if a learning process previously has been performed. ... In that process, the correlations between the spectral-pitch cues of voiced speech sounds ... are recognized and stored. The knowledge about harmonic pitch relations which is acquired in this way is employed by the system in the generation of virtual pitch." While the role of a learning mechanism may be questioned given that the perception of missing-fundamental pitch appears to be already present in young infants (Clarkson and Clifton 1985; Montgomery and Clarkson 1997), a link between vocalization and pitch perception is supported by other arguments. Many vertebrate vocalizations contain harmonic complex tones such as the vowels of human speech. Because the fundamental component is rarely the most intense component in these vocalizations, it would often be masked in the presence of environmental noise, thereby creating a selective pressure for a missing-fundamental mechanism. The link between vocalization and pitch perception has recently been formalized using a model which predicts many psychophysical pitch phenomena from the probability distribution of human voiced speech sounds, without explicit reference to any specific pitch extraction mechanism (Schwartz and Purves 2004).

In cats, the fundamental frequency range most important for vocalizations lies at 500-1000 Hz (Brown et al. 1978; Nicastro and Owren 2003; Shipley et al. 1991). In this range,

pitch is robustly represented in both rate-place profiles and pooled interspike interval distributions. The difficulties encountered by the rate-place representation over the F0-region below 300 Hz, which is the most important for human voice, may reflect the poorer frequency selectivity of the cat cochlea compared to the human (Shera et al. 2002). If we assume that human cochlear filters are about three times as sharply tuned as cat filters, consistent with the Shera et al. data, the rate-place representation would hold in humans for F0s at least as low as 100 Hz, thereby encompassing most voiced speech sounds. Thus, in humans as well as in cats, the pitch of most conspecific vocalizations may be robustly represented in both rate-place profiles and interspike interval distributions. Such a dual representation may be advantageous in situations when either mechanism is degraded by either cochlear damage or central disorders of temporal processing because it would still allow impaired individual to extract pitch information.

Despite its appeal, the idea that the pitch of conspecific vocalizations has a dual representation in spatial and temporal codes is not likely to hold across vertebrate species. At one extreme is the mustached bat, where the F0s of vocalizations range from 8 to 30 kHz (Kanwal et al. 1994), virtually ruling out any temporal mechanism. Evidence for a perception of the pitch of missing fundamental stimuli at ultrasonic frequencies is available for one species of bats (Preisler and Schmidt 1998). At the other extreme is the bullfrog, where the fundamental frequency of vocalizations near 100 Hz appears to be coded in the phase locking of auditory-nerve fibers to the sound's envelope rather than by a place mechanism (Dear et al. 1993; Simmons et al. 1990). Although this species is sensitive to the fundamental frequency of complex tones (Capranica and Moffat 1975), it is not known whether it experiences a missing-fundamental phenomenon similar to that in humans. These examples suggest that a tight link between pitch and vocalization may be incompatible with the existence of a general pitch mechanism common to all vertebrate species. Either different species use separate pitch mechanisms to different degrees, or the primary function of the pitch mechanism is not to extract information from conspecific vocalizations, or both.

1.6 CONCLUSIONS

We compared the effectiveness of two possible representations of the pitch of harmonic complex tones in the responses of the population of auditory-nerve fibers at low and moderate stimulus levels: a rate-place representation based on resolved harmonics and a temporal representation based on pooled interspike-interval distributions. A major finding is that the rate-place representation was most effective for F0s above 400-500 Hz, consistent with previous reports of a lack of rate-place cues to pitch for lower F0s, and with the improvement in cochlear frequency selectivity with increasing frequency. The interspike-interval representation gave precise estimates of pitch for low F0s, but broke down near 1300 Hz. This upper limit is consistent with the psychophysical limit of the pitch of the missing fundamental for stimuli containing many harmonics, extending to missing-fundamental stimuli the correspondence between the frequency range of phase locking and that of musical pitch. Both rate-place and interspike-interval representations were effective in the F0 range of cat vocalizations, and a similar result may hold for human voice if we take into account the differences in cochlear frequency selectivity between the two species. Consistent with psychophysical data, neither of the two pitch representations was sensitive to the relative phases of the partials for stimuli containing resolved harmonics.

On the other hand, neither representation of pitch is entirely consistent with the psychophysical data. The rate-place representation fails to account for the upper limit of musical pitch and is known to degrade rapidly with increases in sound level and decreases in signal-to-noise ratio. The interval representation has trouble accounting for the greater salience of pitch based on resolved harmonics compared to pitch based on unresolved harmonics and appears to be insufficiently sensitive to phase for stimuli consisting of unresolved harmonics. These conclusions suggest a search for alternative neural codes for pitch that would combine some of the features of place and temporal codes in order to overcome the limitations of either code. One class of codes that may meet these requirements are spatio-temporal codes which depend on both harmonic resolvability and

phase locking (Loeb et al. 1983; Shamma and Klein 2000; Shamma 1985a; Heinz et al. 2001).

1.7 ACKNOWLEDGMENTS

The authors would like to thank Christophe Micheyl, Christopher Shera, and Joshua Bernstein for valuable comments on the manuscript and Andrew Oxenham for his advice on the design of the experiments. Thanks also to Connie Miller for her expert surgical preparation of the animals. This work was supported by NIH grants RO1 DC02258 and P30 DC05209.

CF Range (Hz)	F0 Range (Hz)	Fixed stimulus frequency region (Hz)
440-1760	110-440	440-1760
880-3520	220-880	880-3520
1760-7040	440-1760	1760-7040
3520-14080	880-3520	3520-14080

Table 1. Parameters of the four complex-tone stimuli with varying F0, and range of CFs for which each stimulus was used.

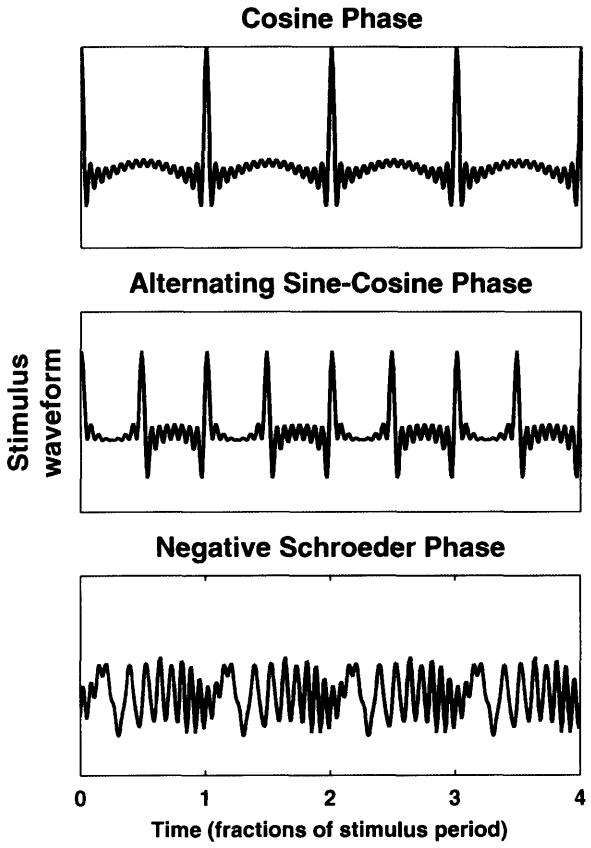


Figure 1.1: different phase relationships among the harmonics give rise to different stimulus waveforms. For harmonics in cosine phase (top), the waveform shows one peak per period of the F0. When the harmonics are in alternating phase (middle), the waveform peaks twice every period of the F0. A negative Schroeder phase relationship among the harmonics (bottom) minimizes the amplitude of the oscillations of the envelope of the waveform.

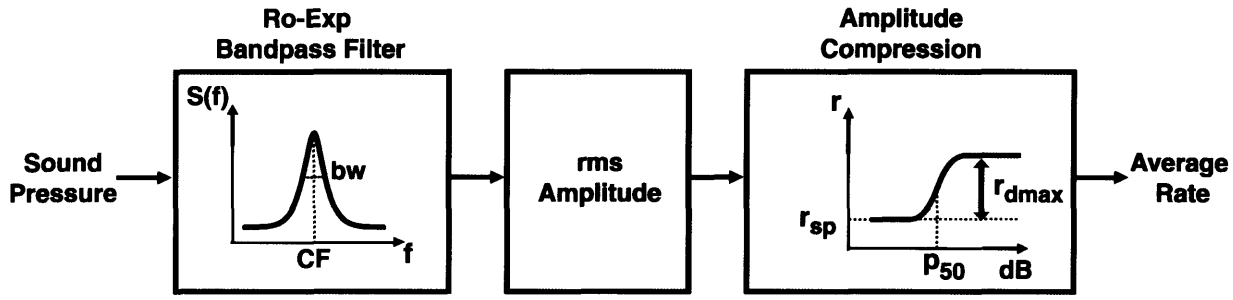


Figure 1.2. Single-fiber average-rate model. The first stage represents cochlear frequency selectivity, implemented by a symmetric rounded-exponential bandpass filter (Patterson 1977). The Sachs and Abbas (1974) model of rate-level functions is used to compute the mean discharge rate from the r.m.s. amplitude at the ouput of the bandpass filter.

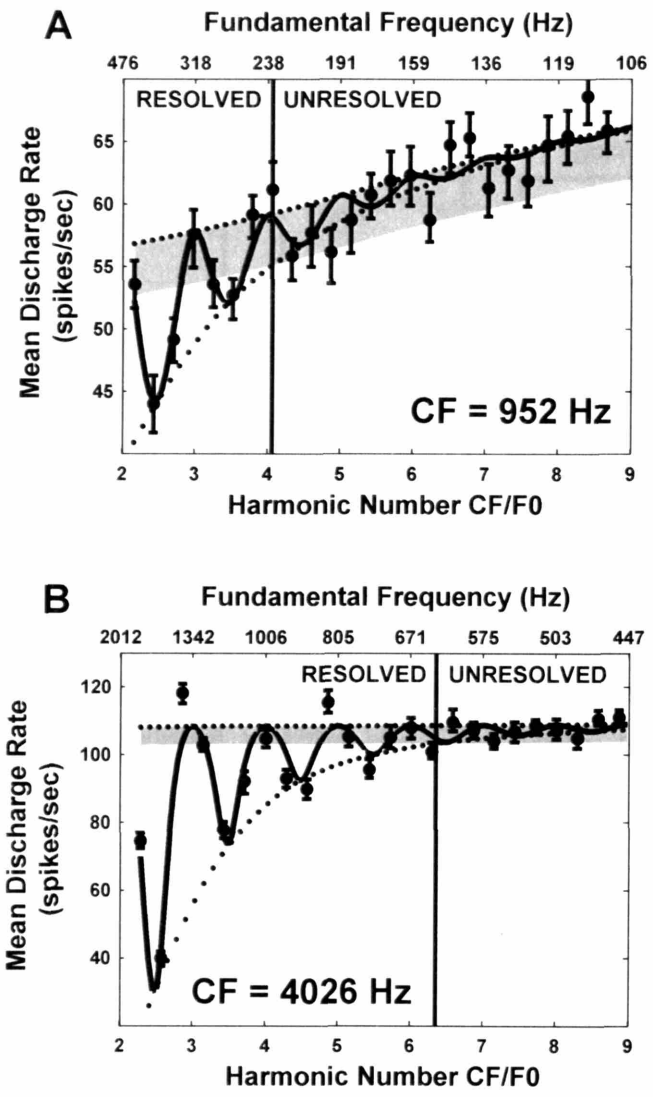


Figure 1.3. Average discharge rate against complex-tone F0 for two AN fibers from the same cat with CFs of 952 Hz (A) and 4026 Hz (B). Because the lower axis shows the harmonic number CF/F0, F0 increases from right to left along the upper axis. Filled circles with errorbars show the mean discharge rate ±1 standard deviation obtained by bootstrap resampling of the stimulus trials (see Method). The solid lines show the response of the best-fitting single-fiber model (Fig. 1.2). The upper and lower envelopes of the fitted curve are shown by dotted lines. The intersection of the lower envelope with two typical standard deviations from the upper envelope (gray shading) gives the maximum harmonic number N_{max} for which harmonics are resolved (vertical lines).

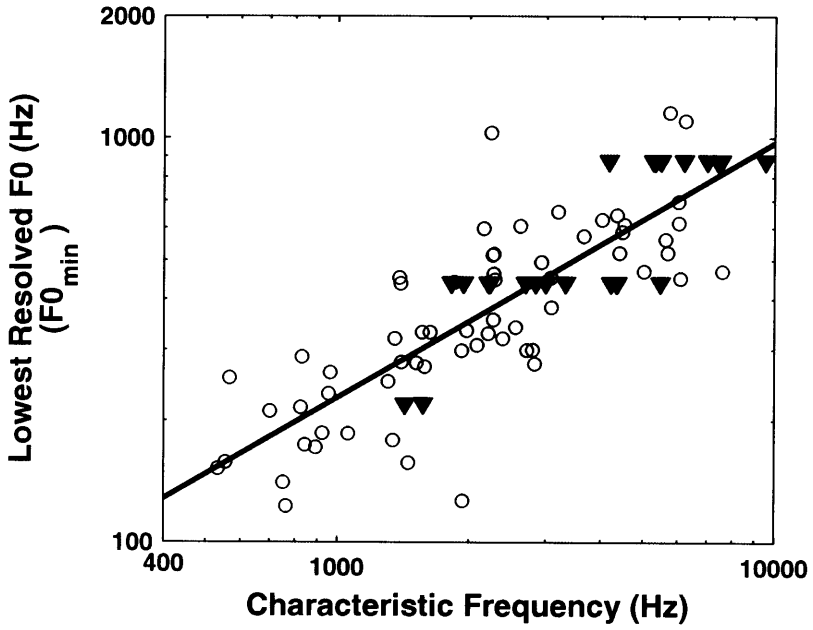


Figure 1.4. Lowest resolved F0 as a function of characteristic frequency. Each point shows data from one auditory-nerve fiber. Triangles show data point for which $F0_{min}$ was somewhat overestimated because harmonics were still resolved for the lowest F0 presented. The solid line shows the best-fitting straight line on double logarithmic coordinates (a power law).

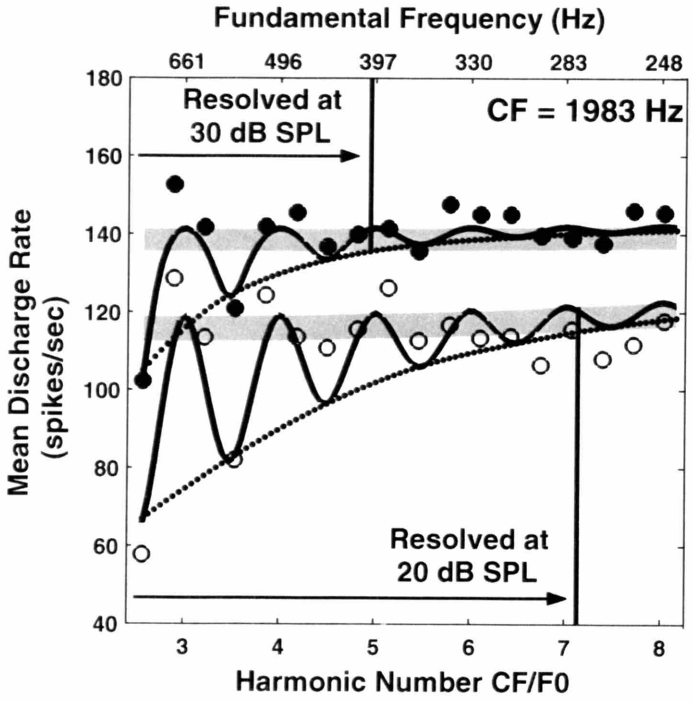


Figure 1.5. Effect of stimulus level on harmonic resolvability in rate responses of an auditory-nerve fiber (CF = 1983 Hz). Open and filled circles show mean discharge rate against F0 for complex tones at 20 and 30 dB SPL, respectively. Solid lines show response of the best fitting model when model parameters were constrained to be the same for both stimulus levels. Other features as in Fig. 1.3.

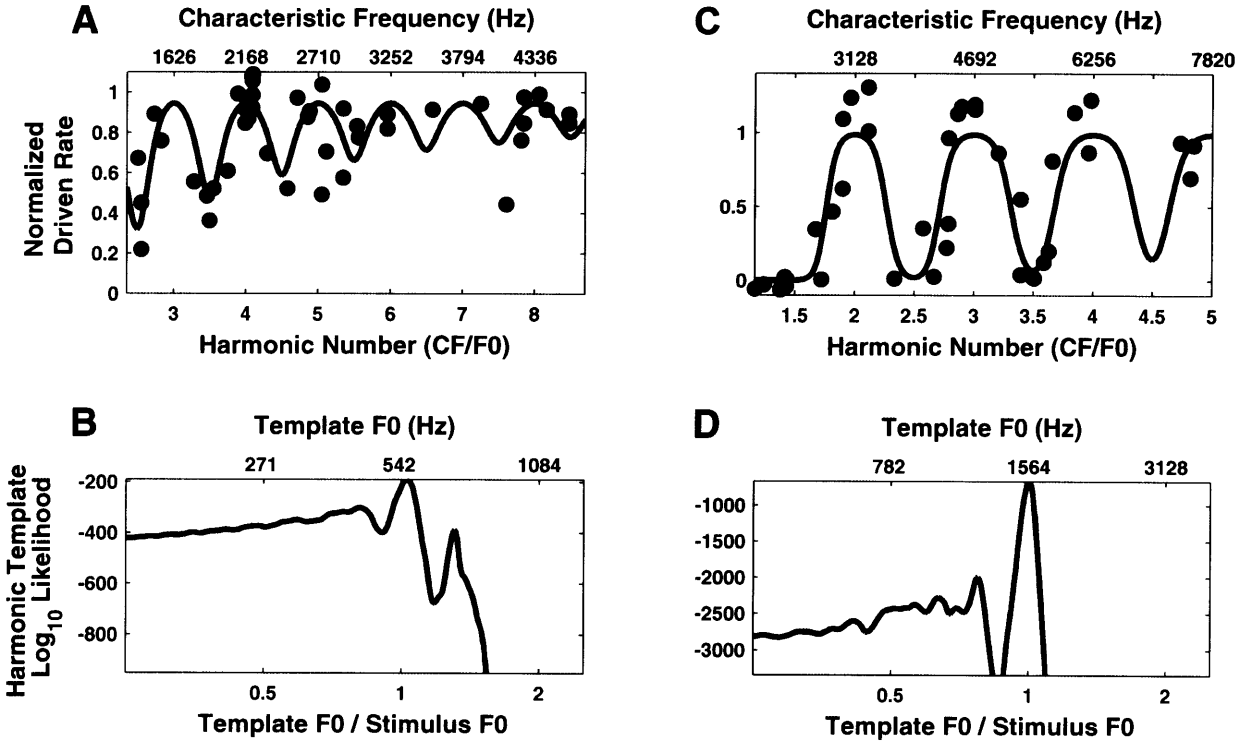


Figure 1.6. Maximum-likelihood pitch estimation from rate-place profiles using harmonic templates for two complex tones with F0s at 542 Hz (A, B) and 1564 Hz (C, D), respectively. A and C: Filled circles show normalized driven rate as a function of both CF (upper axis) and harmonic number CF/F0 (lower axis). Solid lines show the maximum-likelihood harmonic template, which is the response of a population model to a complex tone with equal-amplitude harmonics (see Method). B and D: Log likelihood of the harmonic template model in producing the data points as a function of template F0.

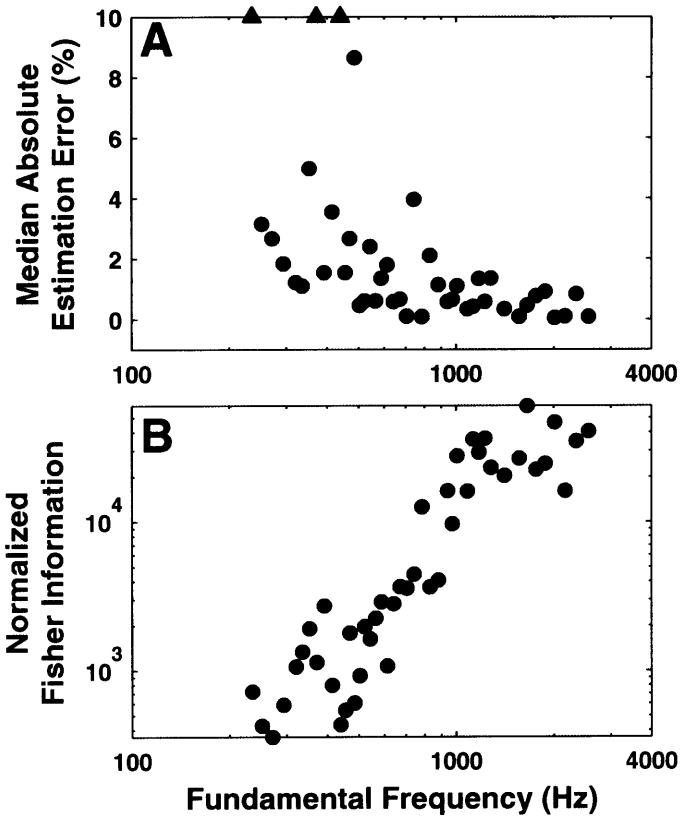


Figure 1.7. Pitch estimation based on rate-place profiles: median absolute estimation error (A) and normalized Fisher Information (B) as a function of F0. The median is obtained over 100 bootstrap resamplings of the data (see Method). Triangles indicate data points for which the median was out of the range defined by the vertical axes. The Fisher Information is a measure of pitch strength defined as the curvature of the log likelihood with respect to template F0 at the location of the maximum.

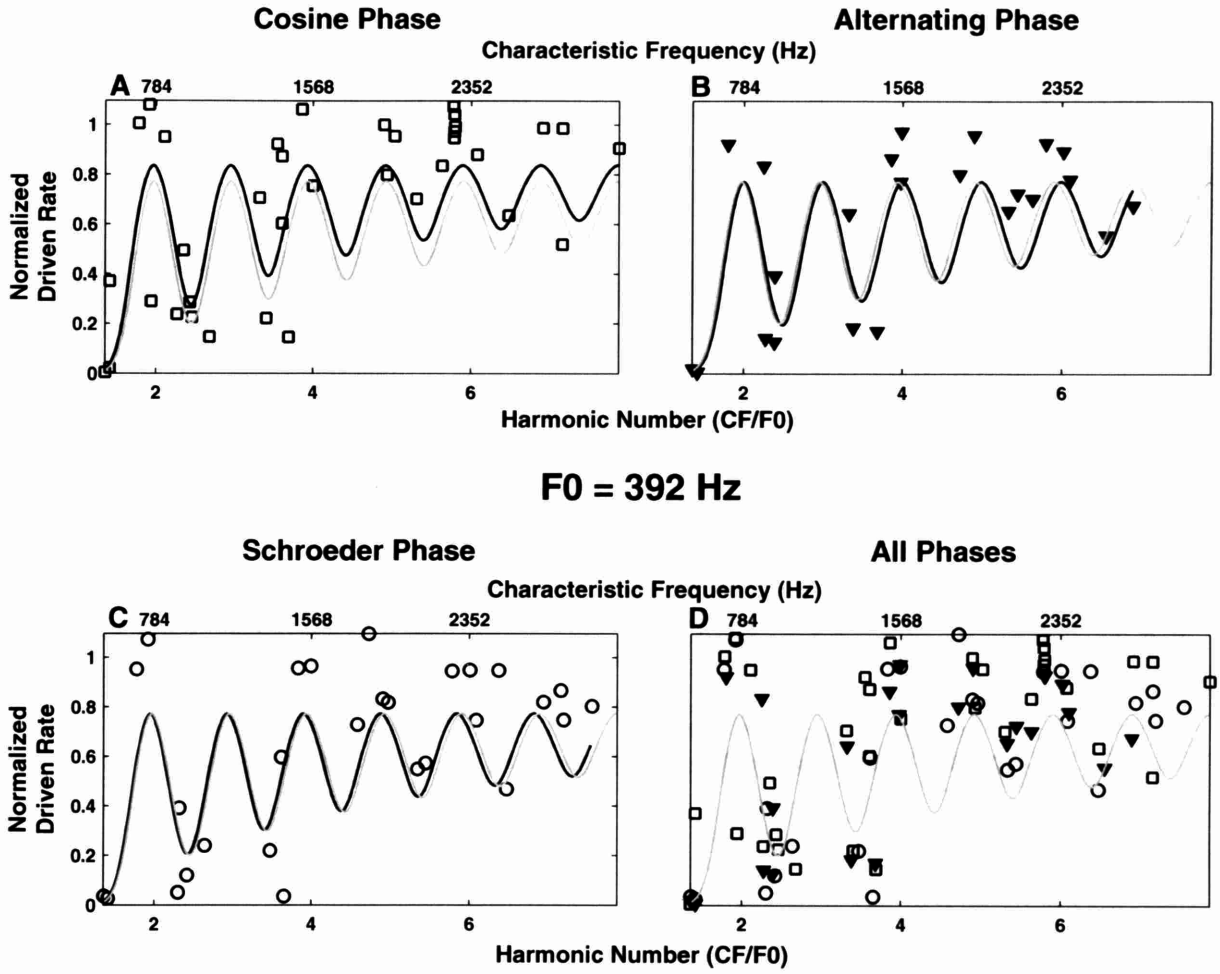


Figure 1.8. Effect of the relative phase of the harmonics on pitch estimation based on rate-place profiles. Stimulus F0 = 392 Hz. A,B and C: normalized driven rate (symbols) and maximum likelihood harmonic templates (black lines) for harmonics in cosine (squares), alternating (triangles) and Schroeder (circles) phase, respectively. D: normalized driven rate and maximum likelihood harmonic template (gray line) for data pooled across phases (symbols as in panels A-C). The maximum likelihood harmonic template for data pooled across phases is plotted in gray also in panels A-C.

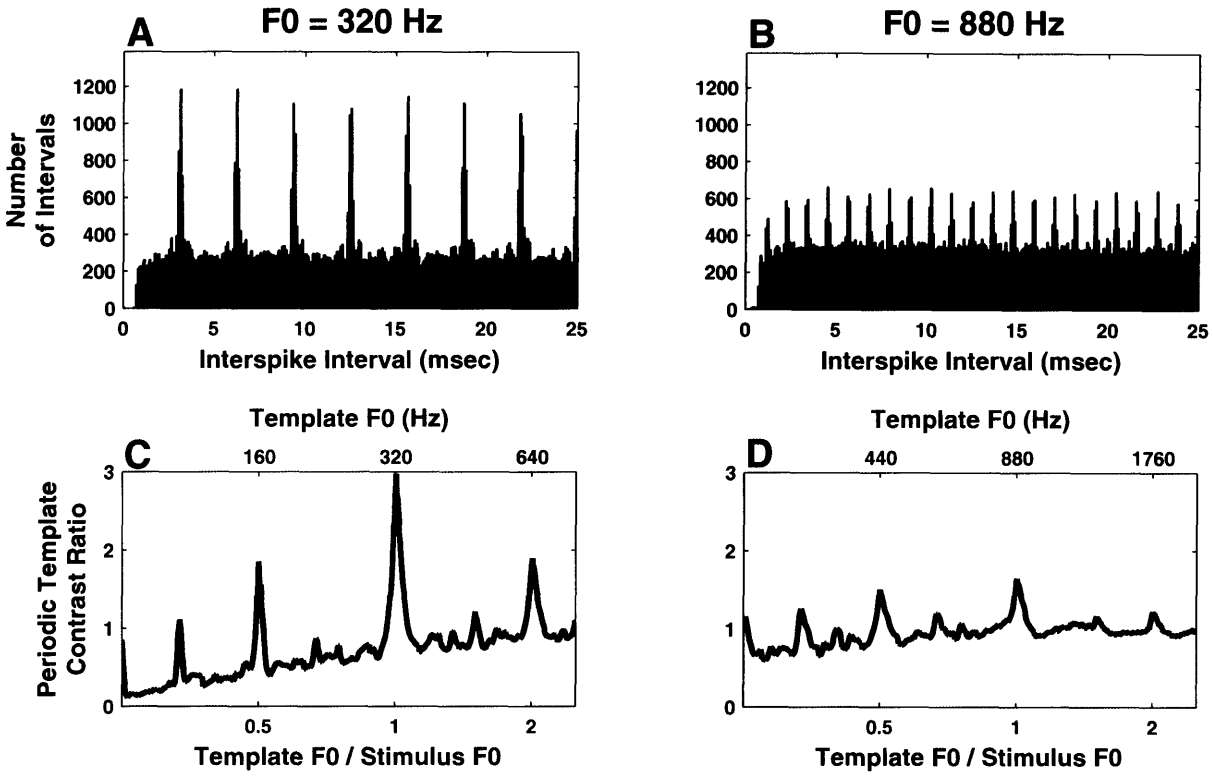


Figure 1.9. Pitch estimation based on pooled all-order interspike interval distributions for two F0s, 320 Hz (A,C) and 880 Hz (B,D). A, B: Pooled all-order interspike interval distributions with periodic templates maximizing the contrast ratio (vertical dotted lines). C, D: periodic template contrast ratio as a function of its F0.

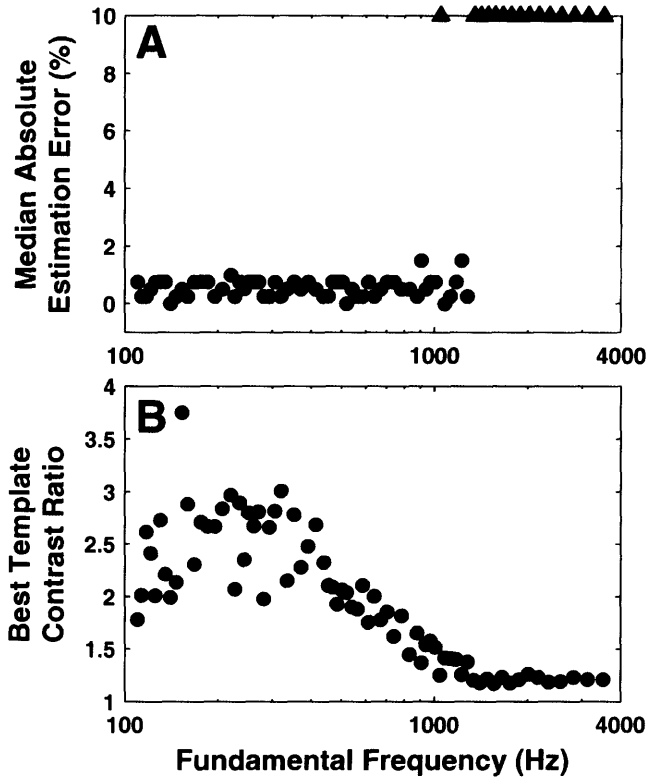


Figure 1.10. Pitch estimation based on pooled all-order interspike interval distributions: median absolute estimation error and best- template "contrast ratio" (see Methods) as a function of F0. Panel A: median (over 100 bootstrap resampling trials) pitch absolute estimation error, expressed as percentage of the true F0. Triangles: the pitch absolute estimation error exceeded 10% of the true F0. Panel B: contrast ratio of the best-matching periodic template.

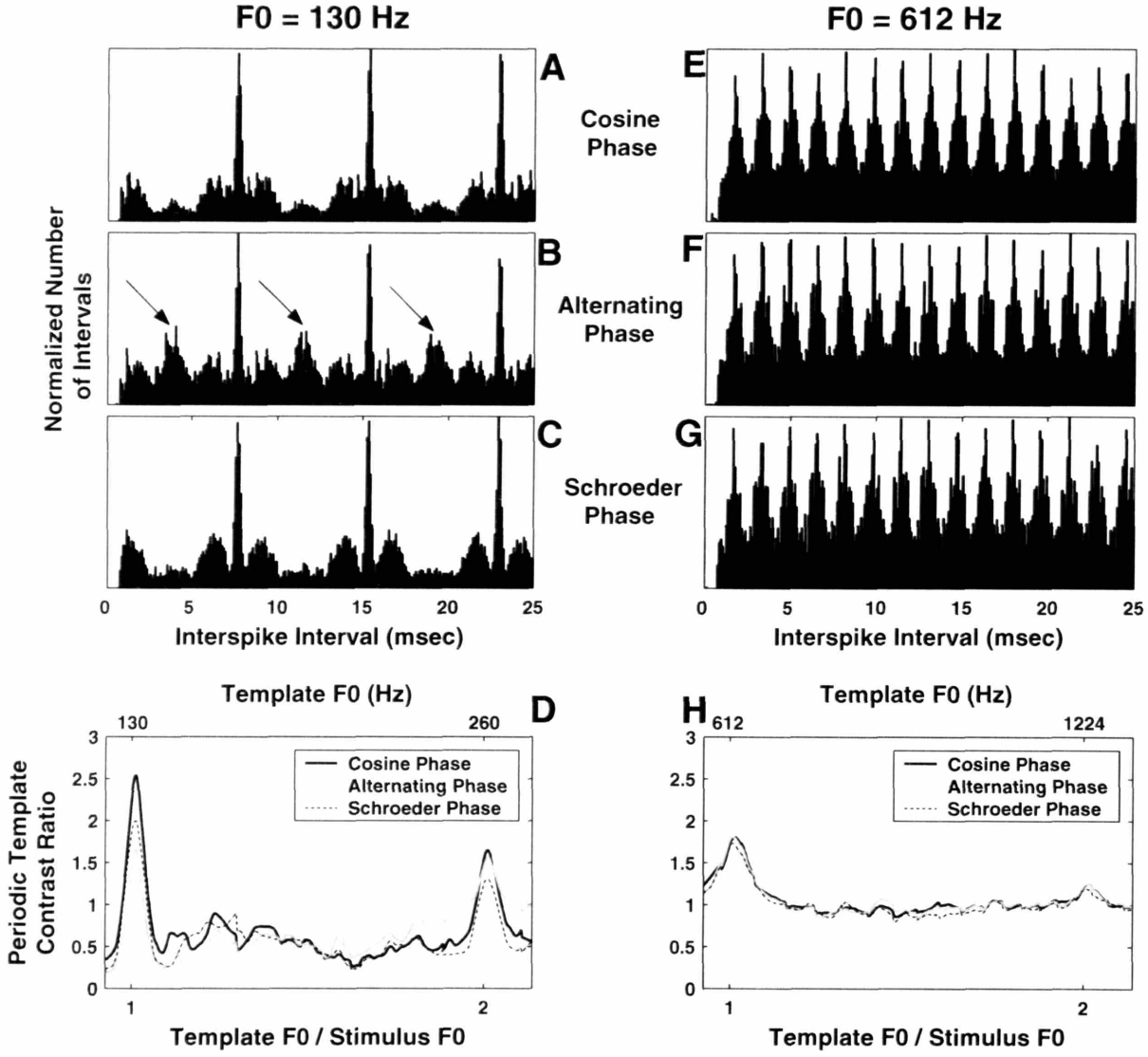


Figure 1.11. Effect of the relative phase of the harmonics on pitch estimation based on pooled all-order interval distributions for two fundamental frequencies (130 Hz in A-D, and 612 Hz in E-H). A-E: Pooled interval distributions for harmonics in cosine phase (A, E), alternating sine-cosine phase (B,F) and Schroeder phase (C,G). Arrows in B point to secondary local maxima in the interval distribution at half the period (i.e. twice the frequency) of the stimulus F0 and its odd multiples. D, H: Periodic template contrast ratio as a function of its F0 for the three phase conditions.

Chapter 2

Spatio-Temporal Representation of the Pitch of Complex Tones in the Auditory Nerve

2.1 INTRODUCTION

In Chapter 1 (Cedolin and Delgutte 2005c), we tested the effectiveness of two classic representations of the pitch of harmonic complex tones at the level of the auditory nerve in the cat: a rate-place representation based on resolved harmonics and a temporal representation based on pooled interspike-interval distributions. The rate-place representation was most effective for fundamental frequencies (F0s) above 400-500 Hz, while the interspike-interval representation gave precise estimates of pitch for low F0s, but broke down near 1300 Hz. Both rate-place and interspike-interval representations were effective in the F0 range of cat vocalizations, which corresponds to about 500-1000 Hz (Brown et al. 1978; Nicastro and Owren 2003; Shipley et al. 1991). Neither of the two pitch representations was sensitive to the relative phases of the partials for stimuli containing resolved harmonics, consistent with human psychophysical data (Houtsma and Smurzynski 1990; Shackleton and Carlyon 1994).

Neither representation of pitch was entirely consistent with available human psychophysical data. The strength of the rate-place representation was predicted to increase monotonically with F0, thus failing to predict the existence of an upper F0-limit for the pitch of missing-fundamental complex tones. The rate-place representation also degrades rapidly with increases in sound level and decreases in signal-to-noise ratio, in contrast to the relatively robust representation suggested by human psychophysics. The interval representation had trouble accounting for the greater salience of pitch based on resolved harmonics compared to pitch based on unresolved harmonics and appeared to be insufficiently sensitive to phase for stimuli consisting of unresolved harmonics.

The present study was aimed at investigating an alternative, "*spatio-temporal*" neural representation of pitch, which might combine the advantages and overcome the limitations of the rate-place and interval representations.

Spatio-temporal representation of pitch

The variation with cochlear place of the phase of basilar membrane motion in response to a pure tone is fastest in proximity of the cochlear place tuned to the tone frequency (Anderson 1971; Pfeiffer and Kim 1975; van der Heijden and Joris 2003). At frequencies

below the limit of phase-locking, this rapid spatial change in phase is reflected in the timing of auditory-nerve firings, thus generating a cue to the frequency of the pure tone (Shamma 1985a).

Our hypothesis is that rapid changes in phase should occur at each of the spatial locations tuned to a resolved harmonic. Hence, "*spatio-temporal*" cues to the frequencies of each of the resolved harmonics of a complex tone would be provided by the wave of excitation produced in the cochlea by each cycle of the stimulus, as the wave travels from the base to the apex. We illustrate this concept by showing (Figure 2.1) the response of a physiologically-realistic peripheral auditory model (Zhang et al., 2001) to a missing-fundamental harmonic complex tone with a 200 Hz F0. In this example, the bandwidths of the peripheral filters were adjusted to fit human psychophysical masking data (Glasberg and Moore 1990). Each cycle of the stimulus waveform creates a wave of excitation traveling from the base (high CFs) to the apex (low CFs) of the cochlea. The velocity of the traveling wave is not uniform, but varies such that the response latency changes more rapidly for CFs near a resolved harmonic of the complex tone than for CFs in between two harmonics.

In principle, these latency cues can be extracted by a central neural mechanism sensitive to the relative timing of spikes in AN fibers innervating neighboring cochlear locations, effectively converting spatio-temporal cues into rate-place cues. One possible mechanism is lateral inhibition, whose effect would likely be increments in discharge rate as inputs with neighboring CFs become less coincident (as would be the case when the phase change with CF is fast). Such mechanism would perform the equivalent of a derivative with respect to space of the AN spatio-temporal response pattern (Shamma, 1985b).

To simulate the extraction of these spatio-temporal cues by a lateral inhibition mechanism, we compute the spatial derivative of the response pattern (a point by point difference between adjacent rows in the image of Fig. 2.1), then integrate the absolute value of the derivative over time. This "*mean absolute spatial derivative*" shows local maxima at CFs corresponding to the frequencies of harmonics 2-6. The average discharge rate, on the other hand, is largely saturated at this stimulus level.

Thus, according to our model simulation, spatio-temporal pitch cues should be available in the spatio-temporal response pattern of AN fibers, and can, in principle, be decoded by a lateral inhibition mechanism. Moreover, spatio-temporal cues to resolved harmonics may

persist at stimulus levels at which the rate-place representation fails due to the saturation of individual fibers average-rate response.

Scaling Invariance

To accurately measure the entire spatio-temporal response pattern of the AN for a complex-tone stimulus composed of many harmonics would be extremely difficult, because it would require a very fine, regular and extensive sampling of CFs. We overcame this hurdle by instead applying the principle of local "*scaling invariance*" in cochlear mechanics (Zweig 1976). Scaling invariance means that, at any particular location x_0 on the basilar membrane whose CF is CF_0, the response H (magnitude and phase) of the basilar membrane to a given frequency f can be expressed as a function only of the ratio f/CF_0. That is to say, $H(f, x_0) = H(f/CF_0)$. Therefore, for a stimulus of frequency f_1, $H(f_1, x_0) = H(f_1/CF_0)$. If we call β the ratio between f_1 and f ($\frac{f_1}{f} = \beta$), then $H(f_1, x_0) = H(\beta \cdot f/CF_0) = H\left(\frac{f}{CF_0/\beta}\right)$.

The magnitude and phase of the response measured at x_0 to a frequency f_1 are therefore equal to the magnitude and the phase of the response to f at the cochlear location tuned to CF_0/β. This means that the waveforms of the two responses have the same shape and they are scaled in time by a factor β. By varying f_1 (and therefore β) and measuring responses at a fixed location tuned to CF_0, one can therefore infer the response to a stimulus of fixed frequency f at different cochlear locations.

Assuming that this scaling property holds for each of the frequency components of a complex sound, and that the locality of this principle is not violated (as would be the case if we were interested in single-fiber responses to frequencies very far away from CF), the spatio-temporal response pattern to a complex tone with a given F0 is equivalent to the response at a given CF to a complex tone with varying F0, if cochlear place and time are plotted in dimensionless units CF/F0 ("*harmonic number*") and t×F0 ("*normalized time*"), respectively.

This principle is very effectively visualized in Figure 2.2.

On the left is the spatio-temporal response pattern predicted by the Zhang et al. model, for AN fibers with CFs spanning the range between 750 and 2250 Hz, to a complex tone with fixed F0 at 500 Hz. The harmonic number varies from 1.5 to 4.5.

On the right is the response pattern predicted by the same model, for a single AN fiber with CF at 1500 Hz, to complex tones whose F0s encompass the range between 333 and 1000 Hz, such that the harmonic number varies over the same range (1.5 to 4.5) as in the figure on the left. Here, the time axis is expressed in dimensionless units T = t×F0 (number of cycles).

The response patterns for the two conditions (the stimulus F0 is fixed on the left, while the CF is fixed on the right) are nearly undistinguishable: they both show fast latency changes in correspondence of integer values of harmonic number, and relatively more constant latency for harmonic numbers that are integer multiples of 0.5 (2.5, 3.5, etc.). Average discharge rate and mean absolute spatial derivative, computed as for Fig. 2.1, also exhibit nearly identical features (peaks at integer values of harmonic number and valleys in between).

Thus, the results of these model simulations support the effectiveness of our strategy of relying on scaling invariance to infer the response of a population of AN fibers to a complex tone of a given F0 by measuring the response of a single AN fiber to complex tones with varying F0.

In this study, we tested the spatio-temporal representation of the pitch of complex tones by recording from single AN fibers in anesthetized cats. We found that this representation is viable for F0 in the range of cat vocalizations, and that it may overcome some of the limitations of the rate-place and of the interval representation in accounting for trends observed in human psychophysics experiments. Preliminary reports of our findings have been presented (Cedolin and Delgutte 2004, 2005a).

2.2 MATERIALS AND METHODS

Procedure

Methods for recording from auditory-nerve fibers in anesthetized cats were as described by Cedolin and Delgutte (2005b). Cats were anesthetized with Dial in urethane (75 mg/kg), with supplementary doses given as needed to maintain an areflexic state. The posterior portion of the skull was removed and the cerebellum retracted to expose the auditory nerve. The tympanic bullae and the middle-ear cavities were opened to expose the round window. Throughout the experiment the cat was given injections of dexamethasone (0.26 mg/kg) to prevent brain swelling, and Ringer's solution (50 ml/day) to prevent dehydration.

The cat was placed on a vibration-isolated table in an electrically-shielded, temperature-controlled, sound-proof chamber. A silver electrode was positioned at the round window to record the compound action potential (CAP) in response to click stimuli, in order to assess the condition and stability of cochlear function.

Sound was delivered to the cat's ear through a closed acoustic assembly driven by an electrodynamic speaker (Realistic 40-1377). The acoustic system was calibrated to allow accurate control over the sound-pressure level at the tympanic membrane. Stimuli were generated by a 16-bit digital-to-analog converter (NI 6052E) using sampling rates of 20 kHz or 50 kHz. Stimuli were digitally filtered to compensate for the transfer characteristics of the acoustic system.

Spikes were recorded with glass micropipettes filled with 2 M KCl. The electrode was inserted into the nerve and then mechanically advanced using a micropositioner (Kopf 650). The electrode signal was bandpass filtered and fed to a custom spike detector. The times of spike peaks were recorded with 1-μs resolution and saved to disk for subsequent analysis.

A click stimulus at approximately 55 dB SPL was used to search for single units. Upon contact with a fiber, a frequency tuning curve was measured by an automatic tracking algorithm (Kiang et al. 1970) using 50-ms tone bursts, presented at a rate of 10/s, and the characteristic frequency (CF) was determined. The spontaneous firing rate (SR) of the fiber was measured over an interval of 20 s. The responses to complex-tone stimuli were then studied.

Stimuli

Stimuli were harmonic complex tones with missing fundamentals. The fundamental frequency (F0) of a complex tone was stepped up and down over a range such that the ratio of fiber's CF to F0 ("harmonic number") typically varied from 1.5 to 4.5, in order to capture low-order harmonics, likely to be resolved. For each fiber, the values of F0 were chosen so as to regularly sample the range of harmonic numbers used. The typical harmonic-number step was 1/8. Therefore, the typical number of F0 for which responses were collected in each fiber was 25 ($8 \times (4.5 - 1.5) + 1$). Each complex tone was composed of harmonics 2 to 20, all of equal amplitude, and the fundamental frequency was always missing. Each of the F0 steps lasted 200 ms, including a 20-ms transition period during which the waveform for one F0 gradually decayed while overlapping with the gradual build up of the waveform for the subsequent F0. Responses were typically collected over 20 repetitions of the stimulus, with no interruption.

The sound pressure level of each harmonic was initially set at 10-15 dB above the fiber's threshold for a pure tone at CF. When possible, the stimulus level was then varied over a wide range (typically 20-30 dB) to investigate the stability of the spatio-temporal representation at levels at which the average-rate response is saturated.

In order to compare neural responses to psychophysical data on the phase dependence of pitch, three versions of each stimulus were generated with different phase relationships among the harmonics: cosine phase, alternating (sine-cosine) phase, and negative *Schroeder* phase (Schroeder 1970). The three stimuli have the same power spectrum and autocorrelation function, but differ in their temporal envelope (Figure 2.3). The cosine-phase stimulus has a very "peaky" envelope with periodicity at F0. The envelope of the alternating-phase stimulus is also very peaky, but its periodicity is at 2xF0, even though the period of the waveform remains equal to 1/F0. Finally, a Schroeder phase relationship among the harmonics produces a waveform whose envelope has minimum oscillations. While markedly dissimilar in their envelope, stimuli with harmonics in these three phase configurations produce similar pitches and pitch strengths in psychophysical experiments, as long as some of the harmonics are resolved (Houtsma and Smurzynski 1990; Carlyon and Shackleton 1994).

Analysis

In response to each complex tone, we constructed period histograms using spikes occurring in a 180-ms window extending over each F0 step, but excluding the transition period between F0 steps. The non-scaling conduction delay typical for each cochlear location (Carney and Yin 1988) was subtracted from the timing of each spike. Period histograms were visualized as a function of both time and F0, as in Fig. 2.2B. Consistent with scaling invariance, the time and F0 axes were expressed in dimensionless units T = t×F0 (number of cycles) and n = CF/F0 (harmonic number, or normalized place), respectively. As a consequence of this normalization of the time axis, the number of bins over which the period histograms were computed was the same (50) for all F0s. Two metrics were computed from the resulting spatio-temporal response pattern H(*n,T*): the average discharge rate ("R_{avg}"), obtained by integrating each period histogram over time and converting to spikes/s, and the "mean absolute spatial derivative" ("*MASD*"), obtained by differentiating the time-F0 pattern with respect to harmonic number, then taking the absolute value and integrating over time.

$$R_{avg}(n) = \int_0^1 H(n,T)dT\,; \quad MASD(n) = \int_0^1 \left|\frac{\partial H(n,T)}{\partial n}\right| dT$$

The spatial derivative represents the effect of a hypothetical central mechanism based on lateral inhibition that could, in principle, extract spatio-temporal pitch cues (Shamma 1985b).

Both R_{avg} and MASD were smoothed by linear convolution with a three-point triangular filter and then plotted as functions of harmonic number. Integer values of harmonic number occur when the F0 is such that a fiber's CF coincides with one of the harmonics of the stimulus, while the harmonic number is an odd integer multiple of 0.5 (2.5, 3.5, etc...) when the F0 is such that the CF falls halfway between two harmonics. Resolved harmonics are expected to result in peaks in either the R_{avg} or the MASD (or both) for integer values of the harmonic number, with valleys in between.

As in our previous study (Cedolin and Delgutte 2005c), we used "bootstrap" resampling (Efron and Tibshirani 1993) of the data recorded from each fiber, in order to evaluate the statistical properties of the estimates of R_{avg} and MASD. In particular, one hundred resampled data sets were generated by drawing with replacement from the set of spike trains,

typically recorded over 20 stimulus repetitions. Response intervals corresponding to the ascending and descending part of the F0 sequence were drawn independently from each other. Average discharge rate and mean absolute spatial derivative were computed from each bootstrap data set, and the standard deviations of these measures were used as error bars.

A simple mathematical function was fit to a fiber's response to the complex-tone stimuli. The function, which can capture the main features of both the R_{avg} and the MASD, is the following:

$$f(n) = A \cdot \cos(2\pi \cdot \delta \cdot n) \cdot e^{\frac{-n}{n_0}} + B \cdot e^{\frac{-n}{n_0}} + C \qquad (1)$$

where n is the harmonic number (CF/F0). A co-sinusoidal oscillating component at frequency δ, whose amplitude decays exponentially with harmonic number n (CF/F0), is added to a constant term C and to an exponential function of harmonic number. In the case where δ is equal to one, the function peaks exactly at integer values of harmonic number, effectively resulting in the "characteristic frequency" of the oscillating component (the product of δ and the pure-tone CF) to match the tuning-curve CF.

The model has 5 free parameters: the amplitude (A) and frequency (δ) of the oscillating component; the time constant (n_0) of the decay of the oscillation (fixed to be equal to the time constant of the baseline trend); the amplitude (B) of the baseline increase (or decrease); the value of the DC component (C). The model was fit independently to profiles of R_{avg} and MASD by the least squares method using the Levenberg-Marquardt algorithm as implemented by Matlab's "lsqcurvefit" function. The procedure was repeated for each of the one hundred profiles of R_{avg} and MASD derived from the bootstrap data sets to quantify the variability of parameter estimates.

Based on the values of the parameters of the best fitting curves, we defined metrics that were used to directly compare precision, accuracy and strength of the pitch cues provided by the rate-place and the spatio-temporal representation.

2.3 RESULTS

Our results are based on 173 measurements of responses to harmonic complex tones recorded from 94 auditory-nerve fibers in 6 cats. Of these, 76 had high SR (> 18 spikes/s), 3 had low SR (< 0.5 spike/s), and 15 had medium SR. The CFs of the fibers ranged from 300 Hz to 5 kHz.

Figure 2.4 shows the responses to complex tones with harmonics in cosine phase for three auditory-nerve fibers with CFs of 700 Hz (A), 2150 Hz (B) and 4300 Hz (C), respectively. The responses shown were measured at 10, 20 and 30 dB, respectively, above each fiber's threshold for a pure-tone at CF. Period histograms are displayed as a function of both normalized time (txF0, horizontal axis) and harmonic number (CF/F0, vertical axis) (see Methods) in the left panels. The stimulus waveform is also plotted against normalized time. The right panels show the average discharge rate (R_{avg}) and the mean absolute spatial derivative (MASD) as a function of harmonic number derived from the corresponding spatio-temporal response pattern (see Methods).

For the low-CF fiber, the latency of the response varies more or less uniformly with harmonic number, even at very low (10 dB) level above threshold. Strong cues to resolved harmonics are therefore not present in the spatio-temporal response pattern of this fiber. This is reflected in the absence of pronounced peaks in the MASD at integer harmonic numbers. The same observation is valid for the R_{avg}, which is nearly constant. This suggests that, at this CF, both the rate-place and the spatio-temporal representations do not provide strong evidence for resolved harmonics for the F0s in the range used.

The spatio-temporal pattern of the response of the fiber with CF at 2150 Hz does show variations in response latency with harmonic number, qualitatively similar to those predicted by the spatio-temporal model of Fig. 2.2. The phase varies rapidly at integer harmonic numbers, while it changes more slowly at harmonic numbers that are odd integer multiples of 0.5. As a result, the MASD shows local maxima at harmonic numbers equal to 2, 3 and 4, and local minima in between, thus providing evidence for spatio-temporal cues to pitch in the AN response. The R_{avg} also shows peaks at integer harmonic numbers, even though they appear less pronounced than those of the MASD.

As CF increases, the ability of single AN fibers to resolve harmonics of complex stimuli based on average rate is expected to improve, due to the progressive sharpening of the relative bandwidth of cochlear filters with respect to their center frequency (Kiang et al. 1965; Shera et al. 2002; Present study, Fig. 1.3 and 1.4). On the other hand, the effectiveness of the spatio-temporal representation may decrease at high CFs, due to the rapid worsening in phase-locking above 2-3 kHz (Johnson 1980) in the cat AN. Both these predictions are confirmed for a fiber with a CF of 4.3 kHz, in Fig. 2.4C. At this level (30 dB above threshold), up to 6 harmonics are resolved based on R_{avg}, while no strong latency cues are apparent from examining the spatio-temporal response pattern. Although small peaks in the MASD are present at the second and third harmonic number, they appear to reflect predominantly the variation in average rate with harmonic number, rather than differences in response latency.

As pointed out in the introduction, an appealing aspect of the spatio-temporal representation of pitch is that it is expected to work at levels at which the rate-place representation breaks down. To test this hypothesis, we compared the effectiveness of the two representations at different stimulus levels. Figure 2.5 shows data for one AN fiber with pure-tone CF at around 1920 Hz. Responses to complex tones with harmonics in cosine phase were recorded at 10, 25 and 40 dB, respectively, above the fiber's threshold at CF (25 dB SPL). At the low level (top), the second, the third, and arguably the fourth harmonic appear as distinct peaks in both the R_{avg} and the MASD. As the level of each harmonic is increased to 25 dB above threshold (middle), the R_{avg} begins to show signs of saturation as only the second harmonic appears to be resolved. In contrast, strong latency cues to the second, third and arguably fourth harmonic are still present in the spatio-temporal response pattern, resulting in corresponding prominent peaks in the MASD. At the highest level tested (65 dB SPL, bottom) the R_{avg} is completely saturated, while peaks at the second and third harmonic number are still easily detectable in the MASD. This example seems therefore to support our hypothesis that a spatio-temporal representation might still work at levels at which the effectiveness of a strictly rate-based representation is decreased.

Population results

To quantitatively compare the precision and the strength of the pitch cues provided by the rate-place and by the spatio-temporal representation, a mathematical function (see Methods) was fit independently to profiles of R_{avg} and MASD for each fiber in our sample. An example is shown in Figure 2.6, for the same fiber as in Fig. 2.4B. The tuning-curve CF of this fiber was 2150 Hz and the level of each of the cosine-phase harmonics was 20 dB above the threshold at CF. The best-fitting curves (solid lines) closely capture the oscillations of both the R_{avg} and the MASD. The "characteristic frequency" CF_{osc} of the oscillating component of the fitted curve (the product of the tuning-curve CF and the parameter δ of the best-fitting curve) is effectively an estimate of the characteristic frequency of the fiber based on its response to complex tones. When δ is equal to one, the function peaks exactly at integer values of harmonic number, effectively resulting in CF_{osc} to match the tuning-curve CF.

In the example of Fig. 2.6, the estimate of the CF provided by the MASD fit is 2263 ± 4 Hz, while the one based on the R_{avg} fit is 2210 ± 5 Hz. Both are slightly larger than the estimate based on the pure-tone tuning curve (2150 Hz).

Figure 2.7 shows the relationships between the estimates of CF derived from profiles of R_{avg} and MASD and from pure-tone tuning curves (TCs) for our entire sample of fibers. Differences between CF estimates, expressed as a percentage of the tuning-curve CF, are plotted against tuning-curve CF. Data are grouped by level relative to each fiber's threshold for a pure tone (columns). For a point to be included in this figure (and in all the following summary plots), the variance of the residuals after fitting the single-fiber model to the data had to be significantly smaller ($P < 0.05$, F-test) than the variance of the residuals for a model without the oscillating component (see Fig. 2.6), so that the parameter δ (and therefore the CF) could be reliably estimated. Excluded measurements are indicated by black crosses in Fig. 2.7.

Analysis of covariance was performed on the data displayed in each row of Fig. 2.7 to quantify whether the data depended on CF, stimulus level, or a combination of both. Red lines indicate the linear models with the minimum number of parameters (based on F-test on the variance of the residuals) that provided the best fits (on a log-frequency scale) across all

level groups. Results of this analysis are stated only if statistically significant at the 0.01 level (or better).

At low CFs, differences between MASD-based and TC-based CF estimates are typically positive and can be as large as 20-25% (Fig. 2.7, A-C). These deviations exhibit a similar decrease with CF in all three level ranges (red lines). The R_{avg}-based estimates of CF are also typically larger than the TC-based ones at low CFs (Fig. 2.7, D-F), and these differences also decrease significantly with CF at all levels (red lines). The best-fitting linear model has fixed slope (-13.92) across levels, but its intercept decreases significantly at the highest levels. For fibers with CFs above about 3 kHz, tuning curve CFs tend to become larger, on average, than R_{avg}-based CFs at very high levels (F).

MASD-based CF-estimates are generally larger than R_{avg}-based estimates at all levels (Fig. 2.7, G-I). The trend in the data showed no significant dependence on CF, but a significant effect of level (the mean difference going from 0.75% at low levels (G) up to 2.67% at the highest levels (I)). This result suggests that, at a particular cochlear place, the pure-tone frequency for which the magnitude of the basilar membrane response is maximum is lower than the frequency for which the rate of variation of the phase of the response is fastest. This is true even at very low levels, and the effect grows with stimulus level.

By virtue of the scaling invariance principle, and for both the R_{avg} and the MASD, the statistical properties of an estimate of a fiber's CF provided by the single-fiber model should be analogous to those of a hypothetical estimate of the pitch of the complex tone that would produce a spatio-temporal response pattern equivalent to the one observed (Fig. 2.2). In particular, precision and strength of an estimate of CF yielded by the single-fiber model can be considered as equivalent to the precision and the strength of pitch estimation, if the appropriate conversion from CF back to F0 is made (we will return to this point with more detail in the Discussion).

A measure of precision of the CF estimates obtained from best-fitting single-fiber models is shown in Figure 2.8 for profiles of R_{avg} (black) and MASD (red). The standard deviation of the estimates across bootstrap trials, expressed as a percentage of the tuning-curve CF, is plotted against CF for low (A), moderate (B) and high (C) stimulus levels relative to each fiber's threshold at CF. With few exceptions, estimates are highly precise for both representations, up to the highest levels tested, as their standard deviations are in most cases

within 2% of the tuning-curve CF. At about 3 kHz and above, the few reliable estimates of CF derived from model MASD appear to be more variable than those derived from model R_{avg}, suggesting that the degradation of phase-locking at higher frequencies limits latency cues and thus impairs the effectiveness of the spatio-temporal representation. In contrast, the rate representation continues to improve due to better harmonic resolvability.

The more pronounced the oscillations in the R_{avg} and in the MASD, the better individual harmonics are considered to be resolved. To quantify this observation, we use the area between the top and bottom envelope of the oscillating component of the fitted curve (light shadings in the examples of Fig. 2.6) as a measure of the strength of average-rate and spatio-temporal representations. Since the area within the oscillation is very different for profiles of R_{avg} compared to profiles of MASD, we use the ratio of the area within the oscillation to the typical standard deviation of the data points (dark shadings in Fig. 2.6) to directly compare the strengths of the two representations. We call this ratio the "*harmonic strength*" of the MASD and of the R_{avg}, respectively.

Harmonic strengths, computed for both the R_{avg} (black) and the MASD (red), are plotted against CF in Figure 2.9 for our entire sample of responses to complex tones with harmonics in cosine-phase. Below about 1 kHz, where individual harmonics are poorly resolved, the harmonic strengths of both R_{avg} and MASD are low at all levels. For CFs above 1 kHz, the harmonic strengths for both representations are highest at low levels (A), while they tend to progressively decrease at higher levels (B-C). This decrease may be due to broadening of peripheral filters, rate saturation (in the case of the R_{avg}), or both. However, for CFs above about 3 kHz, the harmonic strengths of the R_{avg} at the highest levels tested (C) often remain as large as at low and medium levels. For CFs above 3 kHz, the fraction of measurements for which a curve with an oscillating component gave only an insignificant improvement over a curve without an oscillation when fitting profiles of MASD, increased dramatically with level (from 8/26 at low and medium levels up to 10/16 at high levels), indicating that the spatio-temporal representation degrades rapidly with level in this frequency range.

To directly compare the strengths of the rate-based and of the spatio-temporal representations, we defined a metric that we call "*normalized strength difference*" as the difference between the harmonic strength of the MASD and that of the R_{avg}, divided by their sum. For example, if the harmonic strength of the MASD is three times the harmonic

strength of the R_{avg}, the corresponding normalized strength difference is equal to 0.5, while if the harmonic strength of the MASD is one-third of the harmonic strength of the R_{avg}, the normalized strength difference is -0.5. This metric assumes values between -1 and 1, and it is chosen because reciprocal ratios of harmonic strengths result in normalized strength differences with the same absolute value but opposite sign.

Normalized strength differences are shown against CF in Figure 2.10 for those measurements for which profiles of MASD and R_{avg} oscillated sufficiently to reliably fit a curve including an oscillating component. Two main trends are noticeable, for CFs below and above about 3 kHz, respectively.

For CFs between 1 and 3 kHz, the strengths of the two representations are generally comparable at low levels (A). In this range of CFs, as level increases the MASD tends to produce larger harmonic strength than the R_{avg}, resulting in normalized strength differences that are more often positive than negative at moderate levels (B). At high levels (C), the vast majority of normalized strength differences are positive, indicating that although both representations are less effective than they are at low levels (Fig. 2.9), the oscillations in the spatial derivative remain relatively more pronounced than those in the average rate. An analysis of covariance for fibers with CFs between 1 and 3 kHz (red lines) showed a decrease of normalized strength differences with CF at all levels (with same slope) and a significant increase of the intercept of the best-fitting line with level.

For CFs above about 3 kHz, normalized strength differences tend to assume large negative values at all levels, indicating that rate cues to resolved harmonics are significantly stronger than latency cues. An analysis of covariance for fibers with CF between 3 and 5 kHz (purple lines) showed a significant decrease of normalized strength differences with CF and no significant effect of level. This result is most likely caused by the worsening of phase-locking with the increasing frequency of the (resolved) harmonics to which high-CF fibers are tuned. This phenomenon has already been pointed out as a possible cause for the higher variability of MASD-based CF estimates at high frequencies relative to that of R_{avg} -based estimates (Fig. 2.9).

Phase effects

The results of psychophysical studies with complex tones containing resolved harmonics (Houtsma and Smurzynski 1990; Shackleton and Carlyon 1994) are generally not greatly affected by the relative phase relationship among the partials. To test whether the rate-place and spatio-temporal representations are consistent with this finding, we also measured responses to complex tones with harmonics in alternating phase and in Schroeder phase. Figure 2.11 shows an example for a fiber with a CF of 2520 Hz. The lowest F0 presented in our stimulus was therefore 560 Hz. In general, for fibers with CF around 2500 Hz, the lowest F0 for which harmonics are resolved in the average-rate response should be around 400-500 Hz (Present study, Fig. 1.4). Thus, we can assume that in our example, the stimulus contained well-resolved harmonics at all F0s. The spatio-temporal response patterns are very similar across phase conditions (Fig. 2.11, A-C), as are the strong pitch features present in both R_{avg} and MASD (Fig. 2.11, D,E).

A comparison of the strengths of the cues to resolved harmonics provided by the two representations in the three phase conditions is shown in Figure 2.12. The CFs of the AN fibers included in these plots varied from just above 1 kHz to 3 kHz. Harmonic strengths of the MASD for harmonics in cosine and alternating phase are very similar (A), and both are only slightly larger (non-significantly at the 0.01 level, paired t-test) than the harmonic strengths for harmonics in Schroeder phase (B-C). Similarly, we found no statistically-significant difference for harmonic strengths of the R_{avg} for harmonics in the three phase configurations (D-F), in agreement with our previous findings (Cedolin and Delgutte 2005c). These results suggest that the salience of pitch cues available in both the response latency and magnitude of single AN fibers is to a large extent independent of the phase relationship among resolved harmonics, consistent with observations from studies of human psychophysics.

2.4 DISCUSSION

We investigated the effectiveness of a spatio-temporal representation of the pitch of complex tones in the cat auditory nerve. This study was based upon the hypothesis that cues

to resolved harmonics might be generated by rapid changes in the latency of the basilar membrane response at cochlear locations tuned to each resolved harmonic frequency (Fig. 2.1). This code is "*spatio-temporal*" in that it requires a combination of resolvability of individual harmonics and effective phase-locking at their frequencies.

The most direct method to appropriately test this hypothesis would be to measure the response of an array of AN fibers with very finely and regularly spaced characteristic frequencies, spanning a range wide enough to encompass several harmonic frequencies. Because of the excessive difficulty in making this measurement, we used a simplified, yet nearly equivalent approach based on the principle of scaling invariance in cochlear mechanics (Zweig 1976). The scaling invariance principle allows us to infer the spatio-temporal response pattern of the AN to a complex tone with a fixed F0 from the response pattern of a single AN fiber, measured as a function of the F0 of a complex tone (Fig. 2.2).

Profiles of "mean absolute spatial derivative" (MASD) and "absolute discharge rate" (R_{avg}) were derived from single-fiber spatio-temporal discharge patterns. The MASD operates on a spatio-temporal discharge pattern to simulate the extraction of spatio-temporal cues to resolved harmonics by a putative lateral-inhibition mechanism, sensitive to the timings of its inputs, which are assumed to originate from neighboring cochlear locations. The R_{avg} is used to quantify the cues to pitch provided by a traditional rate-based representation. Simple mathematical functions were fit to profiles of both MASD and R_{avg} (Fig. 2.6), and the parameters of the best-fitting curves were used to compare the effectiveness and strength of the two representations.

Whenever possible, we varied the stimulus level over a wide range to test whether the spatio-temporal representation would still hold at levels at which the rate-based representation is no longer viable.

Spatio-temporal cues to resolved harmonics in single AN fiber responses

Our results show that strong spatio-temporal cues to resolved harmonics are available in the response of AN fibers whose CFs are high enough for harmonics to be sufficiently resolved (above around 1 kHz) but below the limit (around 3 kHz) above which phase-locking is significantly degraded (Fig. 2.4). For fibers with CF below 1 kHz, rate-based cues

to resolved harmonics are also generally weak, due to poor harmonic resolvability. At high CFs, on the other hand, the strength of the rate-place representation grows, due to the progressive sharpening of cochlear filters with their characteristic frequency (Kiang et al. 1965; Shera et al. 2002, Cedolin and Delgutte 2005c), while the spatio-temporal representation breaks down due to the rapid degradation of phase-locking with frequency (Johnson, 1980).

Estimates of CF based on profiles of MASD were generally slightly higher than those based on R_{avg} (Fig. 2.7, G-I). The differences between the two increased significantly with level. We will briefly discuss the implications of this finding in the next section. Both MASD– and R_{avg}–based estimates of CF were generally higher than CFs estimated from the pure-tone tuning curves (Fig. 2.7, A-F) at low levels, and this effect was larger at low CFs. One possibility is that the particular method (Kiang et al. 1970) we used to measure pure-tone tuning curves could have led to estimates of CF with a slightly negative bias. Thresholds were measured for tone frequencies in decreasing order, and each threshold was used as a starting level for the threshold-seeking algorithm at the next (slightly lower) frequency. For tone frequencies on the high-frequency side of the true CF, each starting condition is therefore likely to be a frequency-level combination falling in a highly excitatory area, while the opposite is true for tone frequencies lower than the true CF. Due to the stochastic nature of AN fibers responses, there is a bias towards overestimating thresholds on the high-frequency side of the true-CF and thus to underestimating the value of the CF. However, it is not clear whether this effect should be more pronounced for high-CF fibers because their tuning curves are steeper, on the high-frequency side, than for low-CF fibers, or vice versa, as we observed.

When the MASD- and R_{avg}-based estimates of CF were reliable, their standard deviations were generally within a few percent at all levels (Fig. 2.8), with larger deviations becoming more common for the MASD-based estimates at CFs above about 3 kHz.

For CFs between 1 and 3 kHz, the strengths of the spatio-temporal representation and of the rate-place representation were comparable at low levels (Fig. 2.9A, 10A), but as level increased the spatio-temporal representation was increasingly more effective than the rate representation (Fig. 2.10B,C). The rate-place representation degraded very rapidly with level, due to broadening of cochlear filters (Ruggero et al. 1992) and saturation of single AN

fibers responses (Cedolin and Delgutte 2005c). The spatio-temporal representation also shows signs of deterioration with level (Fig. 2.5), consistent with the finding that responses from fibers innervating neighboring cochlear locations become more coincident at high levels (Anderson 1971; van der Heijden and Joris 2003, Carney and Yin 1988).

The profiles of MASD and R_{avg} derived from spatio-temporal patterns of response to stimuli with harmonics in cosine, alternating and Schroeder phase were very similar for CFs in the range where the harmonics of the stimuli were likely to be resolved (Fig. 2.11). In this CF-range, the strength of both representations also did not vary significantly with phase (Fig. 2.12). Together, these findings suggest that the spatio-temporal representation produces results that are broadly consistent with the lack of phase effects observed in psychophysics experiments when the stimuli contain resolved harmonics.

Scaling invariance "unfolded"

So far, we have discussed the results of our experiment in terms of the characteristic frequencies of the single AN fibers whose spatio-temporal response patterns were measured as a function of complex-tone F0. However, it is interesting to "unfold" the scaling invariance principle and try to draw conclusions in terms of stimulus F0. Since, for the vast majority of our stimuli, we used F0s such that the harmonic number (the ratio of fiber's CF to stimulus F0) varied from 1.5 to 4.5, on the average the single-fiber spatio-temporal discharge pattern observed at a given CF most closely corresponds to the one that would be generated by a single complex tone with F0 roughly equal to CF/3 (Fig. 2.2). This conversion from CF to F0 is also supported by the observation that an unambiguous determination of the (missing-fundamental) stimulus F0 can only be possible if at least 2 of the harmonics are resolved. All our results can therefore be reinterpreted in terms of F0 by simply dividing the CF at which the recording was made by three.

As we have seen, the spatio-temporal representation works best for CFs in the range roughly between 1 and 3 kHz. This range would correspond to a range of F0 between 333 Hz and 1 kHz, covering the entire range of conspecific vocalizations in the cat.

For F0s between 333 Hz and 1 kHz, the rate-place and the spatio-temporal representation are equally effective at low levels (Fig. 2.10A). As stimulus level increases, although both representations degrade, the strength of the spatio-temporal representation exceeds that of the

rate representation (Fig. 2.10 B,C). For F0s below 333 Hz, both the rate-place and the spatio-temporal representation work poorly at all levels, indicating insufficient harmonic resolvability in the cat. This limit is broadly consistent with the finding of our previous study (Cedolin and Delgutte 2005c) that pitch estimation from rate-place profiles was problematic for F0s below 400-500 Hz (although some reliable estimates could be obtained for F0s as low as 250 Hz). For F0s above 1 kHz, as harmonics become increasingly resolved due to the progressive sharpening of cochlear filters relative to their center frequency, the rate representation becomes increasingly more effective (Fig. 2.9). Conversely, for missing-fundamental stimuli, phase-locking at the frequency of the harmonics worsens dramatically for F0s above 1 kHz. This results in the spatio-temporal representation becoming progressively weaker (Fig. 2.10) and less precise (Fig. 2.8) than the rate-representation.

It is interesting to attempt an extension of these cat findings to humans. Two important assumptions have to be made. First, that the lower limit for the viability of a spatio-temporal representation (around 333 Hz in cats) is determined by sharpness of tuning. Second, that phase-locking is primarily responsible for the decreased effectiveness of a spatio-temporal representation above an upper F0-limit (of about 1 kHz in cats). Although the result is a matter of current debate (Ruggero 2005), it has been argued (Shera et. al 2002, Oxenham and Shera, 2003) that the frequency tuning of the auditory periphery in humans might be up to 3 times shaper than in cats. There are no strong reasons to assume that the degree of neural phase-locking and its upper limit should be significantly different for humans than for cats. Based on these assumptions, the lower F0-limit for a spatio-temporal representation of the pitch of complex tones with a missing fundamental might translate from about 333 Hz in cats to about 111 Hz in humans, while the upper limit should remain the same, at about 1 kHz. This range encompasses most of the range of F0 of human voice (80-300 Hz), and the upper limit is consistent with the existence of an upper limit to the pitch of missing-fundamental stimuli [about 1400 Hz in humans (Moore 1973b)].

Another interesting result of our study was that the estimates of CF derived from profiles of spatial derivative were slightly, but systematically larger than those derived from profiles of average rate (Fig. 2.7, G-I), even at low stimulus levels. As stated in the results section, this suggests that, at a particular cochlear place, the pure-tone frequency for which the magnitude of the basilar membrane response is maximum is lower than the frequency for

which the rate of variation of the phase of the response is fastest. The magnitude of this effect increases significantly with stimulus level. Unfolding scaling invariance, these differences indicate that, for a pure-tone stimulus, the magnitude of the basilar membrane response might be maximum at a cochlear place increasingly basal with level to the cochlear place at which the phase varies most rapidly. Although, to our knowledge, this effect has not been systematically investigated for a wide range of CFs in studies of basilar membrane mechanics, there is evidence for a downwards shift of BF with level at frequencies above 1 kHz in the basilar membrane response of the guinea pig (de Boer and Nuttal 2000) and in the AN response of the cat (Greenberg et al. 1986) and of the squirrel monkey (Rose et al. 1971). A study by Shera (2001) suggests that these shifts with level do not originate because of local changes of the resonant frequency of the cochlear partition, but are consequences of the global increase of the driving pressure with stimulus level.

Possible mechanisms for the extraction of spatio-temporal pitch cues

So far we have discussed the availability of cues to resolved harmonics of complex tones in the spatio-temporal pattern of activity of tonotopically-arranged neurons at the level of the auditory nerve. These spatio-temporal cues to harmonic frequencies could theoretically be extracted by neurons that are (1) innervated by auditory nerve fibers with neighboring characteristic frequencies and (2) sensitive to slight differences in the timing of their inputs.

One possible mechanism is lateral inhibition, likely producing local maxima in discharge rate as inputs with neighboring CFs become less coincident (as would be the case when the phase change with CF is fast). An alternative decoding mechanism is cross-frequency coincidence detection (Carney 1994; Heinz et al. 2001) which would result in local minima at locations corresponding to the harmonics since those are the places where the rapid phase changes would cause inputs with neighboring CFs to be less coincident.

Carney (1990) showed that some neurons in the posterior antero-ventral cochlear nucleus (AVCN), most frequently with primarylike-with-notch, transient-chopper and onset response characteristics, are sensitive to the relative timing of their inputs. Precisely, the discharge rate of some of the cells in response to Huffman sequences (clicks filtered in such a way that their magnitude spectra remain flat while a phase variations of desired width is introduced at the neuron's CF) varied with the width of the phase transition. Although many cells

responded more vigorously to shallow phase transitions, indicating "preference" for more coincident inputs, suggesting coincidence detection as the mechanism extracting phase information, in some cases this trend was reversed, consistent with lateral inhibition. These finding suggest that coincidence detection and lateral inhibition might not be mutually exclusive in the AVCN. Interestingly, for some cells the preferred slope of the phase transition depended on stimulus level (typically being shallow at low levels and steep at high levels), indicating that coincidence detection and lateral inhibition might actually be implemented in the same cells at different stimulus levels.

2.5 CONCLUSIONS

We tested the effectiveness of a spatio-temporal representation of the pitch of complex tones in the cat auditory nerve at a wide range of stimulus level. We found that spatio-temporal cues to resolved harmonics are available in the cat AN for F0s in the range between 333 Hz and 1 kHz and can, in principle, be extracted by a lateral-inhibition mechanism. We argue that the lower F0-limit for the spatio-temporal representation is primarily determined by the limited frequency selectivity of auditory-nerve fibers at low CFs, while the upper limit is caused by the abrupt degradation of phase-locking for AN fibers with CFs above 3 kHz. The spatio-temporal representation is viable in the entire F0-range of cat vocalizations, and the same result may be extended to humans by taking into account their sharper cochlear frequency tuning. The spatio-temporal representation is consistent with the existence of an upper F0-limit to the perception of the pitch of complex tones with a missing F0, and its strength does not depend on the relative phase between the harmonics, when these are resolved. The spatio-temporal representation is thus consistent with trends observed in many studies of human psychophysics, and it is more effective than the classical rate-place representation at high stimulus level.

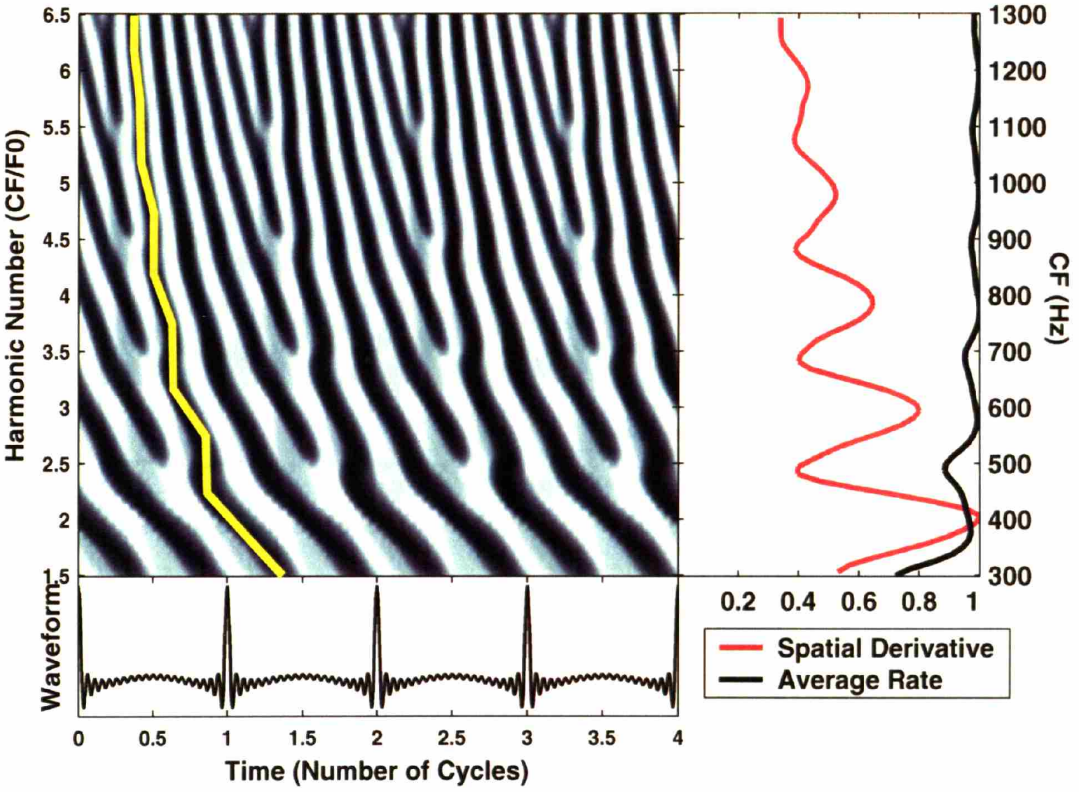


Figure 2.1: Spatio-temporal response pattern predicted by the Zhang et al. (2001) model to a complex tone with F0 of 200 Hz at 50 dB SPL. The response is displayed as a function of time and harmonic number (CF/F0). Fast variations in response latency with CF at integer harmonic numbers are highlighted in yellow. The resulting profiles of average rate (black) and mean absolute spatial derivative (red), normalized by their maximum values, are shown on the right.

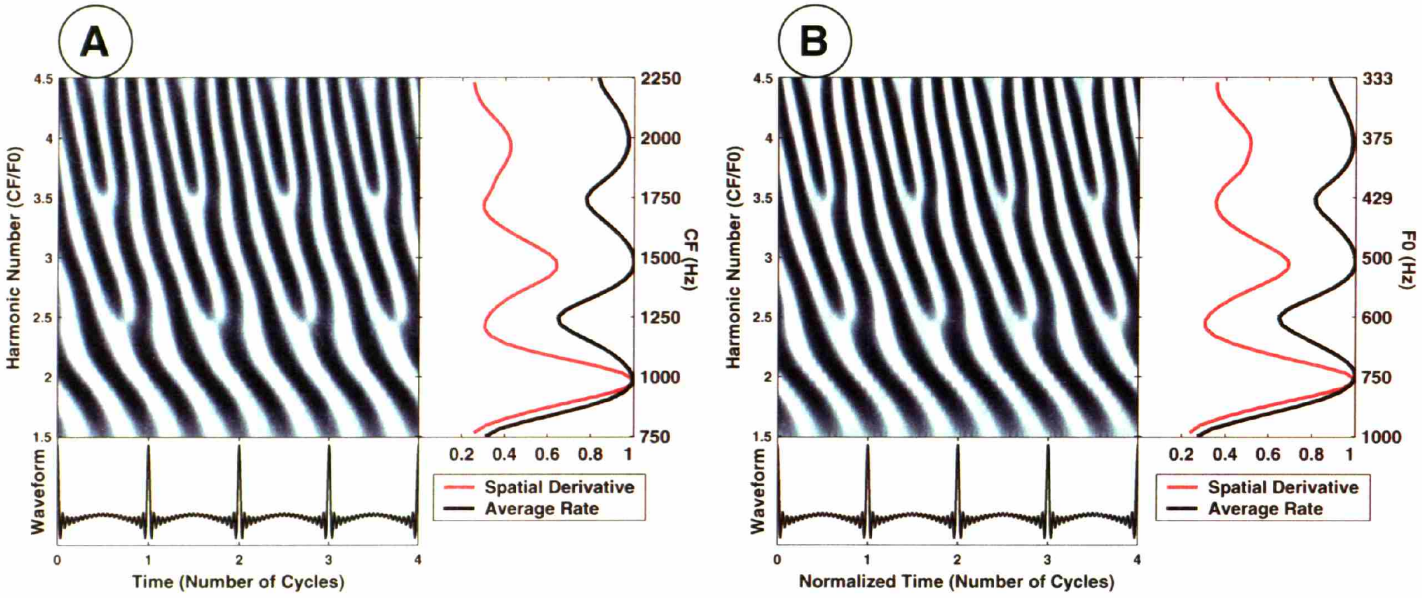


Figure 2.2: Scaling invariance principle. A: spatio-temporal response pattern predicted by the Zhang model for AN fibers with CFs from 750 to 2250 Hz to a complex tone with F0 of 500 Hz at 40 dB SPL, displayed as a function of harmonic number (CF/F0) and time. B: response pattern generated using the same model for one AN fiber with CF of 1500 Hz to complex tones at 40 dB SPL with F0s from 333 to 1000 Hz, displayed as a function of harmonic number and normalized time txF0. Solid lines show the profiles of normalized average discharge rate (black) and mean absolute spatial derivative (red) derived from the two response patterns.

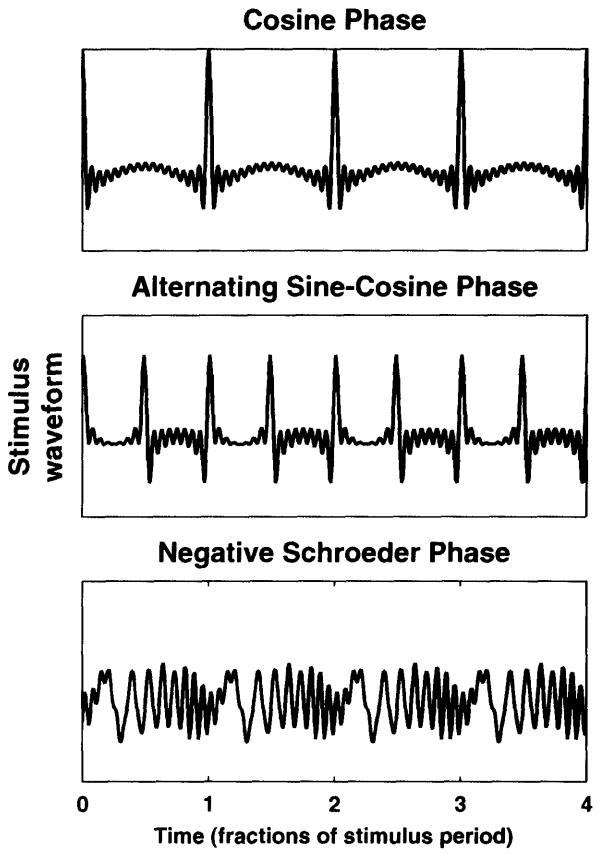
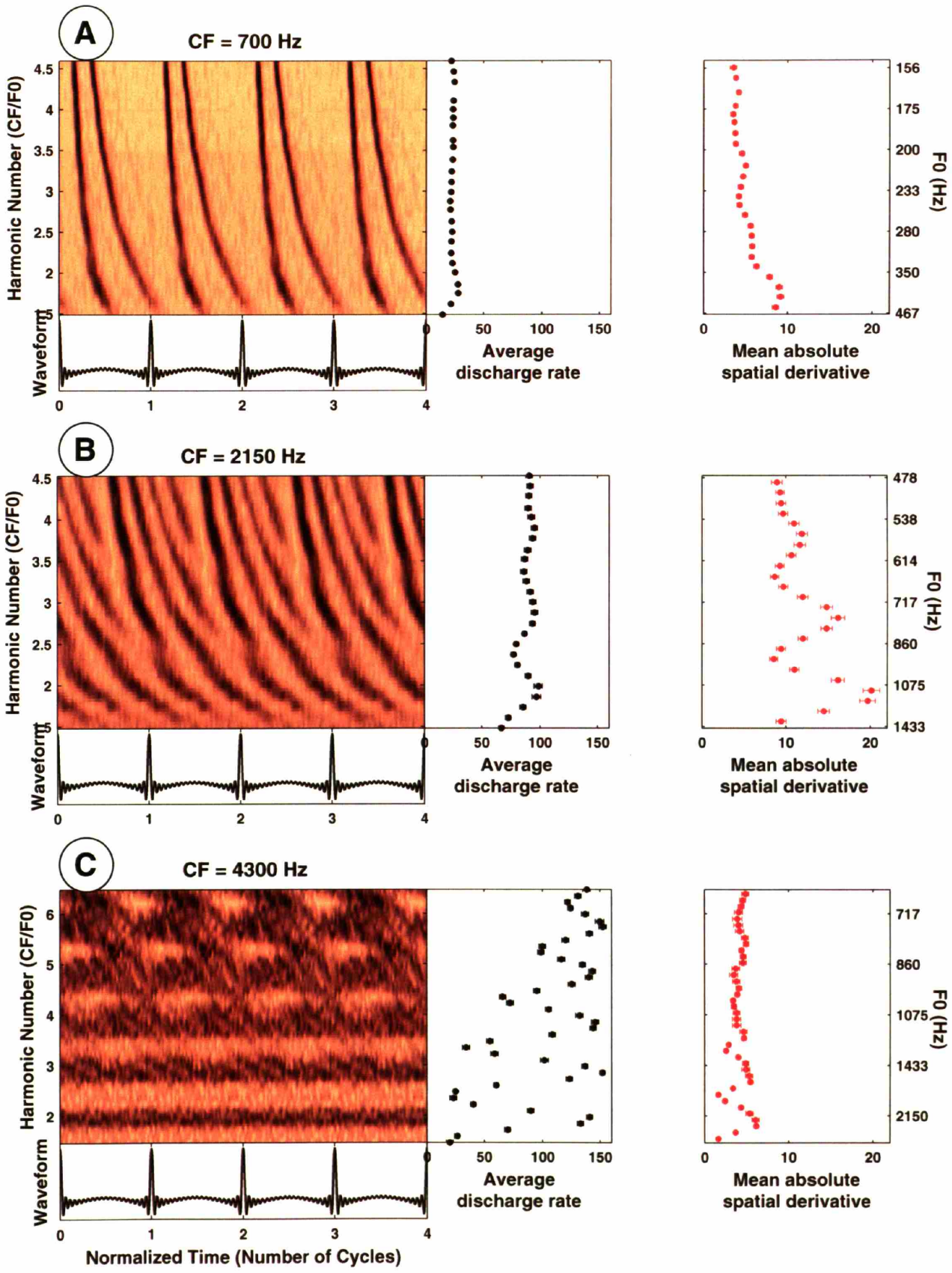


Figure 2.3: different phase relationships among the harmonics give rise to different stimulus waveforms. For harmonics in cosine phase (top), the waveform shows one peak per period of the F0. When the harmonics are in alternating phase (middle), the waveform peaks twice every period of the F0. A negative Schroeder phase relationship among the harmonics (bottom) minimizes the amplitude of the oscillations of the envelope of the waveform.

Figure 2.4. (Next page) Response patterns of three AN fibers with CFs of 700 Hz (A), 2150 Hz (B) and 4300 Hz (C), respectively. Harmonics in cosine phase. Left panels: spatio-temporal discharge pattern, displayed as a function of normalized time (horizontal axis) and harmonic number (vertical axis). Right panels: R_{avg} (black) and MASD (red) derived from the corresponding response patterns. Error bars correspond to ± one standard deviation obtained by bootstrap resampling of the stimulus trials (see Methods).


A
CF = 700 Hz
Harmonic Number (CF/F0)
Waveform
Average discharge rate
Mean absolute spatial derivative
F0 (Hz)
B
CF = 2150 Hz
C
CF = 4300 Hz
Normalized Time (Number of Cycles)

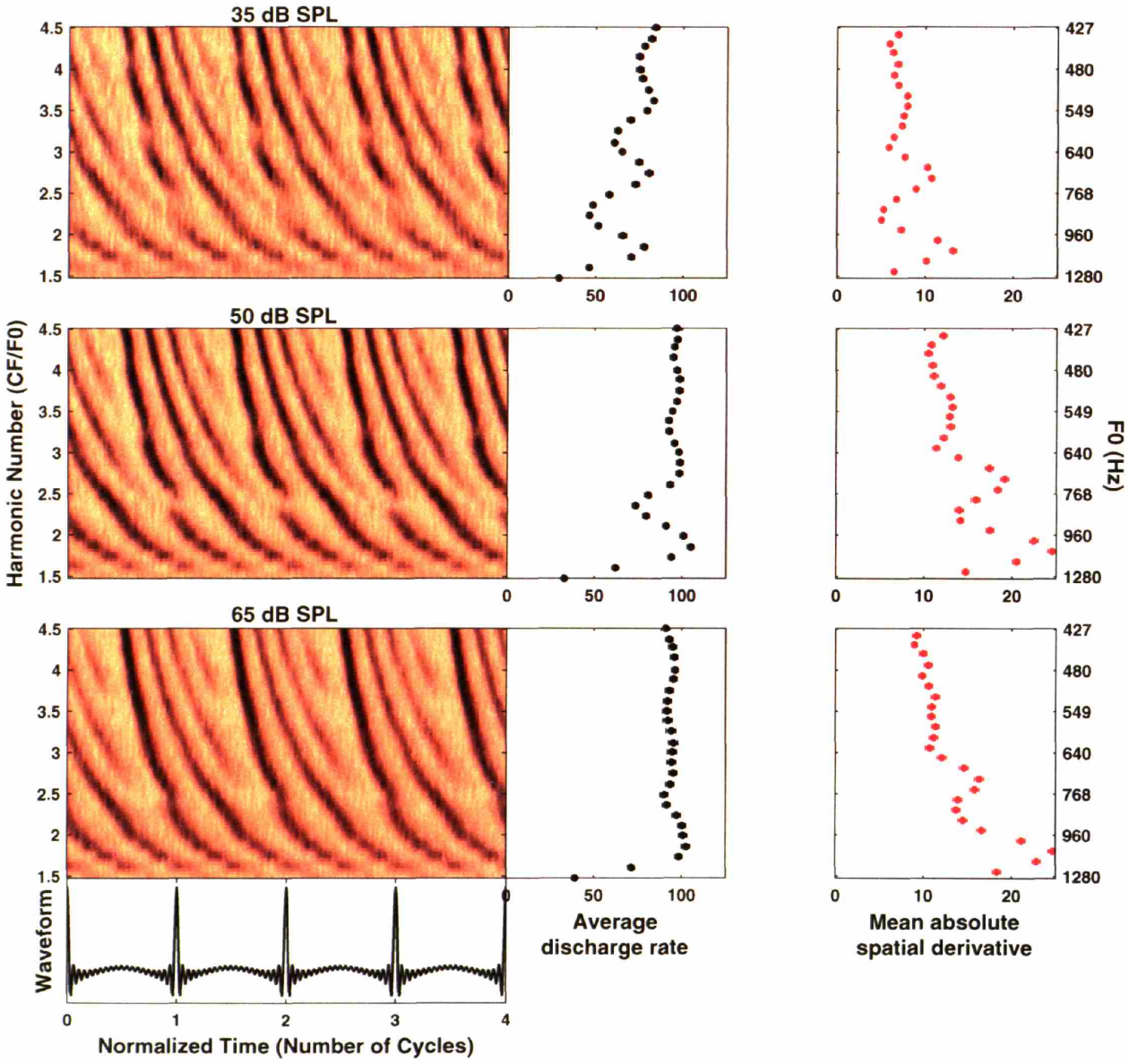


Figure 2.5. Effect of stimulus level on the response pattern of one AN fiber (CF = 1920 Hz). R_{avg} (black) and MASD (red) at 35 (top), 50 (middle) and 65 (bottom) dB SPL, respectively. The threshold for a pure tone at CF was 25 dB SPL. Features as in Fig. 2.4.

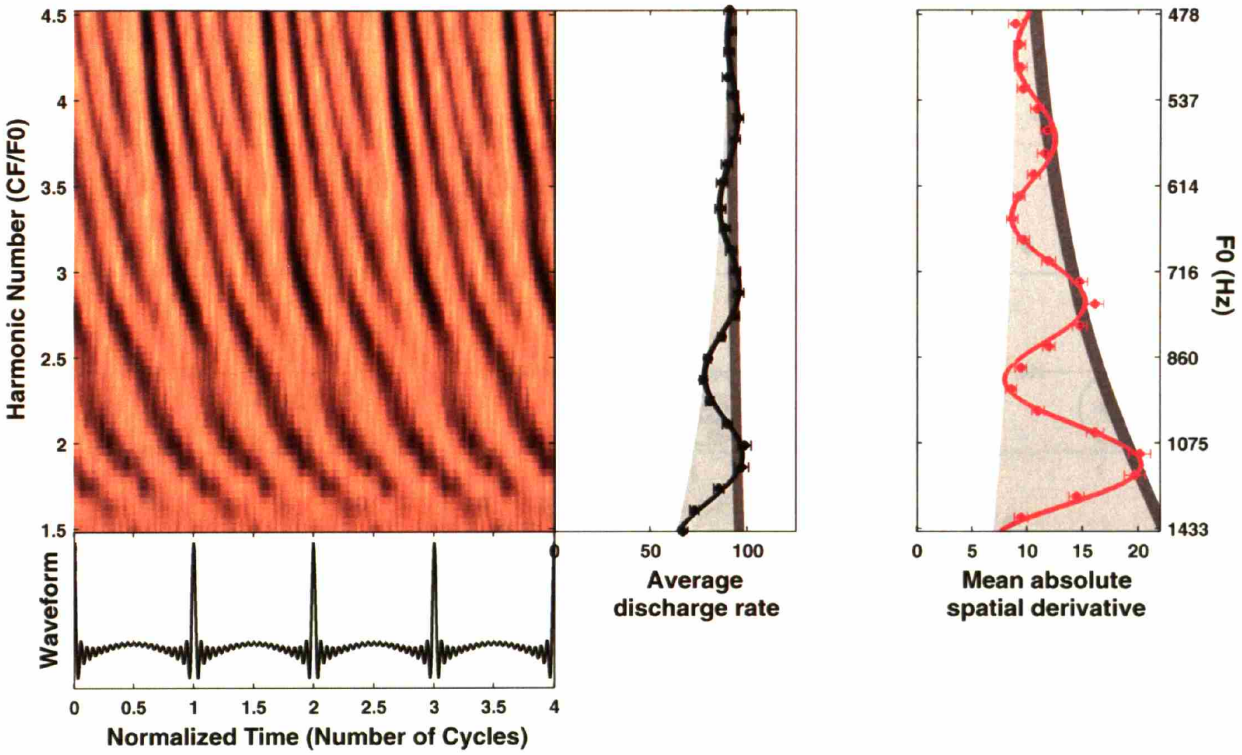


Figure 2.6. Solid lines show single-fiber models fit to profiles of R_{avg} (black) and MASD (red) derived from the spatio-temporal response pattern of a AN fiber with CF of 2150 Hz (same fiber as in Fig. 2.4B). Light shadings indicate the area between the top and bottom envelopes of each fitted curve. Dark shadings correspond to two typical standard deviations of the data points, derived from bootstrapping (see Methods). The ratio of these two quantities is used as an estimate of the strength of the pitch cues provided by each pitch representation.

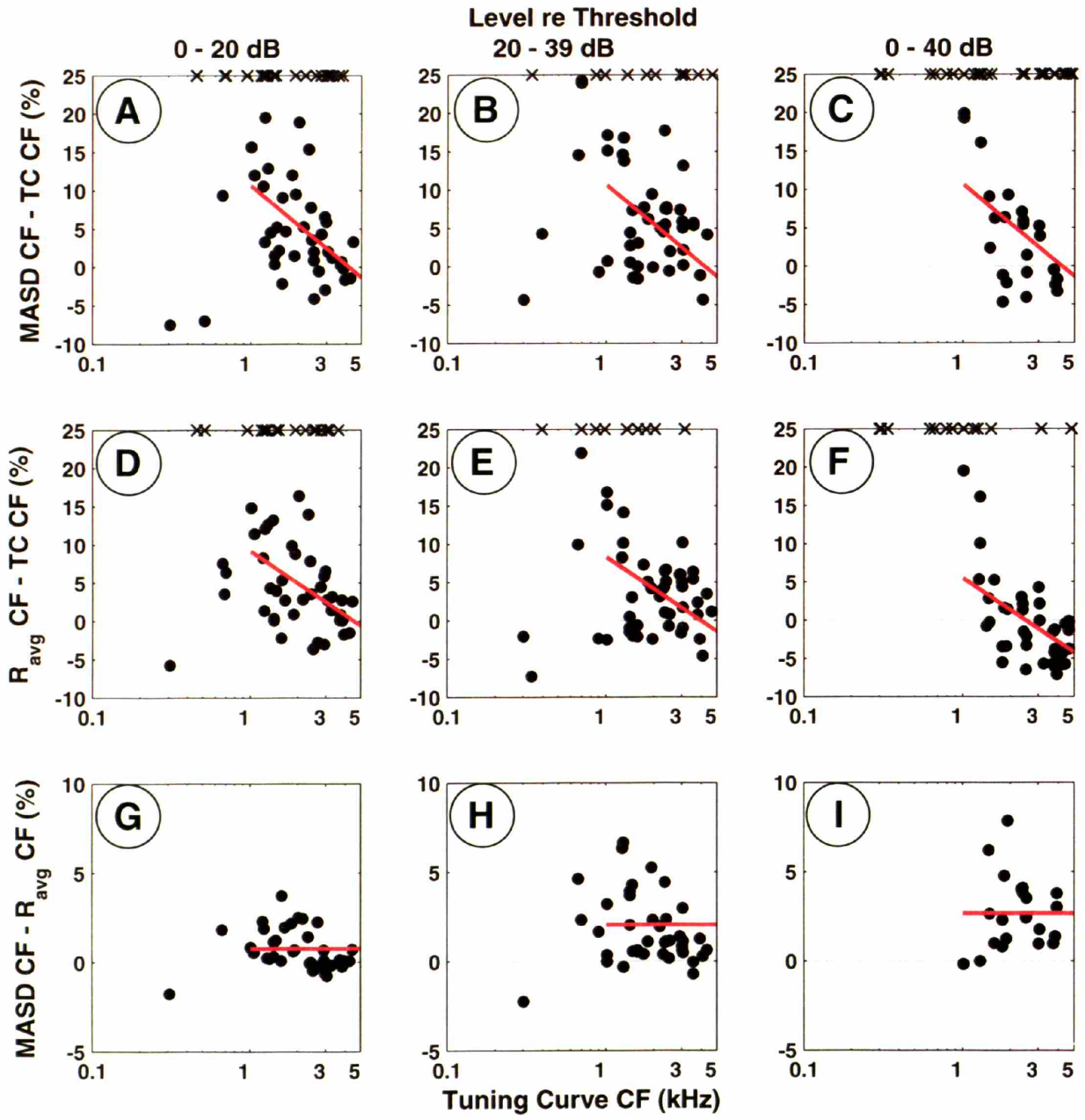


Figure 2.7. Relationships between estimates of CF derived from profiles of R_{avg}, MASD and from pure-tone tuning curves. Filled circles show differences between estimates, expressed as a percentage of the tuning-curve CF. Results are grouped by level relative to each fiber's threshold for a pure tone at CF (columns). Crosses indicate the measurements for which it was not possible to reliably estimate the CF from profiles of R_{avg} and MASD. Solid lines indicate best linear fits (significant at the 0.01 level or better) across all levels. Harmonics in cosine phase.

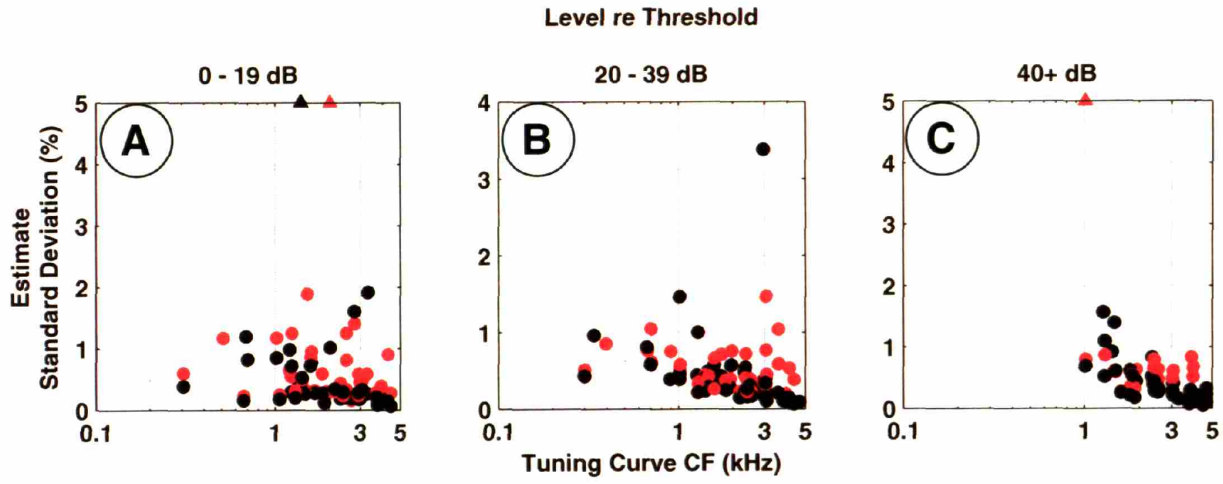


Figure 2.8. Precision of CF estimation based on profiles of R_{avg} (black) and MASD (red): filled circles show standard deviations of the MASD- and R_{avg}-based estimates of CF (expressed as percentage of the tuning curve CF). Results are grouped by level relative to each fiber's threshold at CF. Triangles indicate data points for which the SD was out of the range defined by the vertical axes. Harmonics in cosine phase.

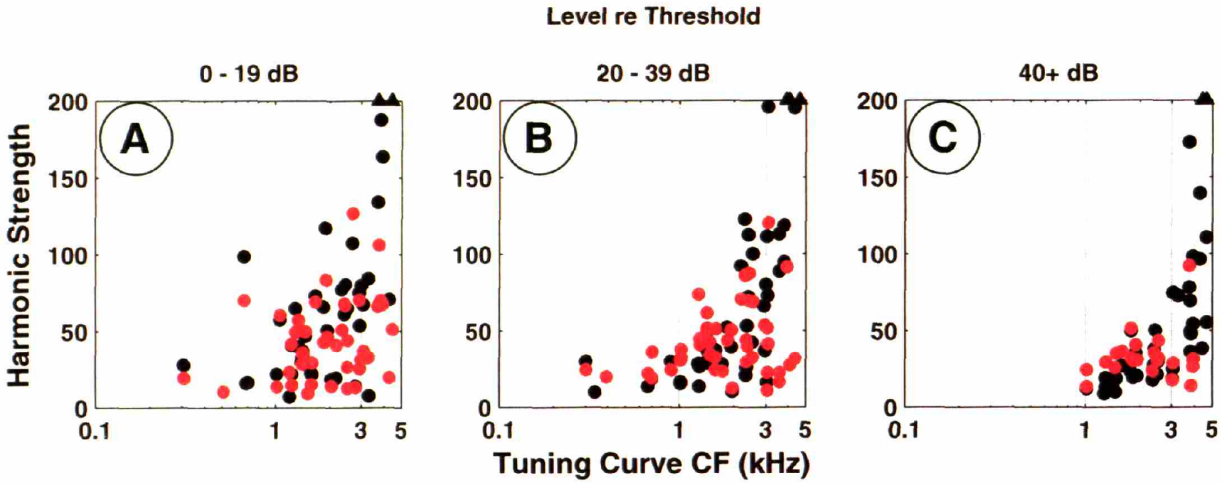


Figure 2.9. Strength of rate-place and spatio-temporal cues to resolved harmonics. Filled circles show harmonic strengths computed for best-fitting curves to profiles of R_{avg} (black) and MASD (red). Triangles indicate data points for which the harmonic strength was out of the range defined by the vertical axes. Results are grouped by level relative to each fiber's threshold for a pure tone at CF. Harmonics in cosine phase.

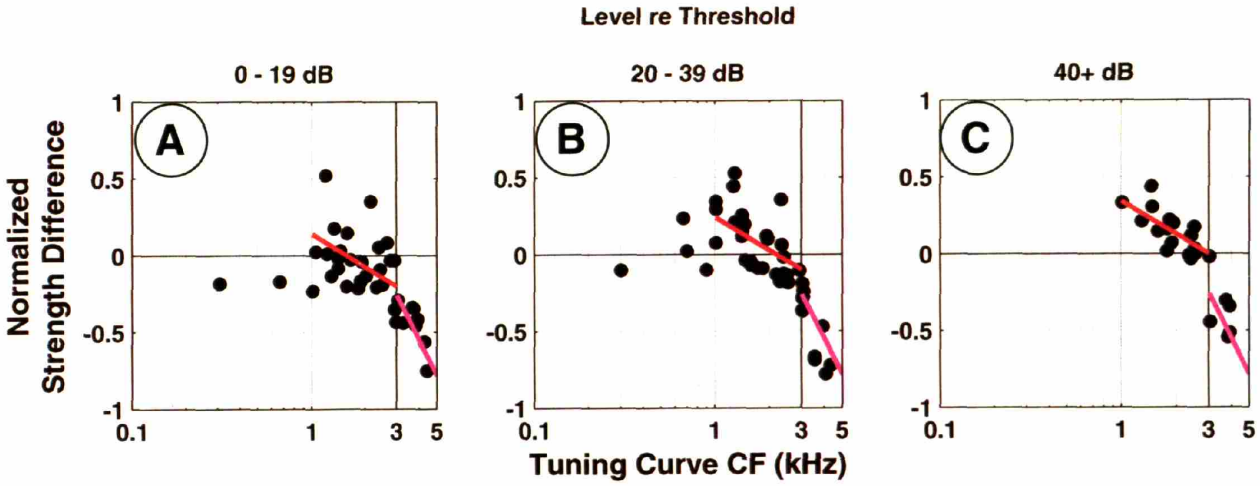


Figure 2.10. Comparison of the strengths of rate-based and spatio-temporal representations. Filled circles show normalized strength differences for the harmonic strengths shown in Fig 9. Positive values mean greater strength for the spatio-temporal representation, and vice versa. Results are grouped by level relative to each fiber's threshold for a pure tone at CF. Red lines indicate best linear fits (significant at the 0.01 level or better) across all levels in the 1-3 kHz CF-range. Purple lines indicate best linear fits (significant at the 0.01 level or better) across all levels in the 3-5 kHz CF-range. Harmonics in cosine phase.

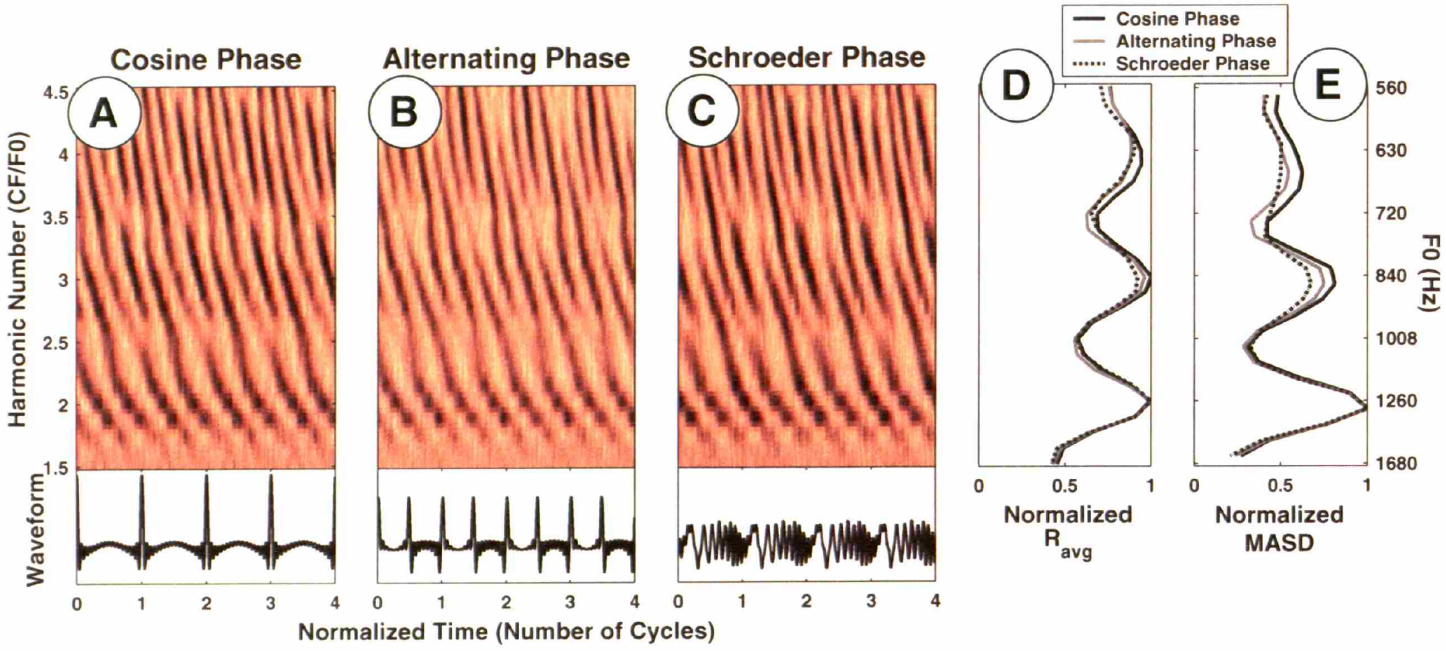


Figure 2.11. Effect of phase relationship among the harmonics on spatio-temporal response patterns and on rate-place and spatio-temporal representations. Response patterns are shown for one fiber with CF at 2520 Hz for stimuli with harmonics in cosine (A), alternating (sine-cosine) (B), and Schroeder (C) phase. The corresponding R_{avg}s and MASDs, normalized by the respective maxima, are plotted in panels D and E.

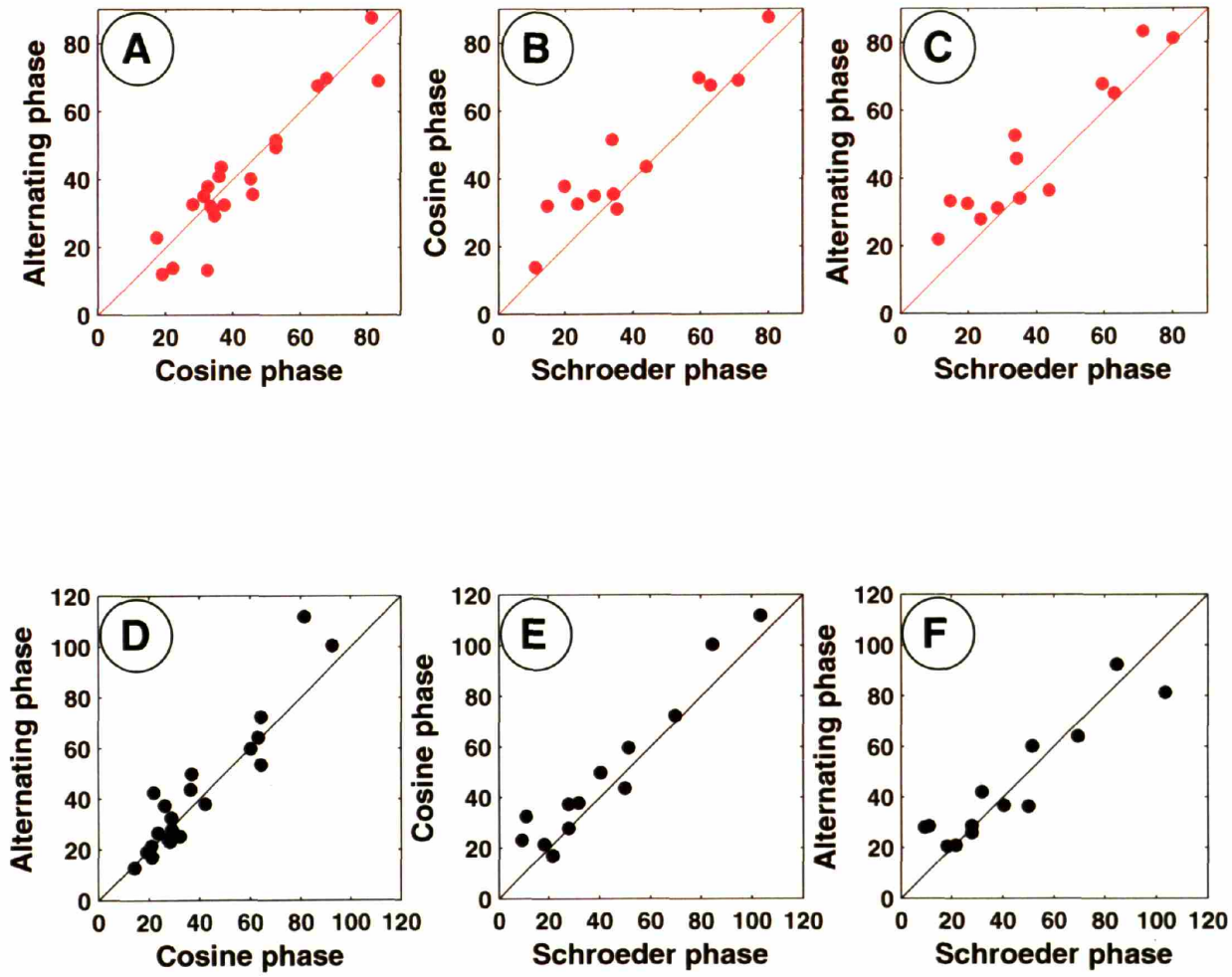


Figure 2.12. Effect of phase relationship among the harmonics on the strength of the spatio-temporal (A-C) and rate-place (D-F) cues to resolved harmonics. Filled circles show comparisons of harmonic strengths for profiles of MASD (red) and Ravg (black) derived from spatio-temporal patterns of response to complex tones with harmonics in cosine, alternating and Schroeder phase. CFs vary from 1 kHz to 3 kHz. Solid lines indicate equality.

Chapter 3

Frequency Selectivity of Auditory-Nerve Fibers Studied with Band-Reject Noise

3.1 INTRODUCTION

One of the most important functions of the cochlea is to mechanically separate sounds into their frequency components. Models of auditory processing, used to predict a wide variety of psychophysical phenomena, simulate the frequency analysis performed by the cochlea by means of a bank of overlapping band-pass filters, referred to as "*auditory filters*".

The most popular technique in psychophysics for estimating the parameters of human auditory filters is the "notched-noise method", introduced by Patterson (1976) and refined by others (for example Patterson et al. 1982; Glasberg and Moore 1990; Rosen et al. 1998). The notched-noise method consists of measuring the threshold level at which a band-reject noise impedes the detection of a fixed-level pure-tone signal, as a function of the width of the rejection band and its placement with respect to the tone. Since the ability of one sound (the "masker") to interfere with the perception of another sound (the "signal") is commonly known as "*masking*", such thresholds are commonly referred to as "*masked thresholds*". Auditory filters are then estimated under the assumptions ("*power spectrum model of masking*", Fletcher 1940) that (1) when detecting a signal in the presence of a masking sound, the listener attends to the filter with the highest signal-to-noise power ratio at its output, (2) detection thresholds correspond to a constant signal-to-masker power ratio at the output of the filter, and (3) the filter is linear for the range of signal and masker levels used to define it.

There is a general agreement about the fact that none of these assumptions is strictly true. For example, psychophysical studies have shown that auditory filters are non-linear in that their parameters depend on signal level (Moore & Glasberg 1987). Also, listeners can combine information from more than one auditory filter in order to enhance signal detection (Spiegel 1981; Buus et al. 1986).

These inconsistencies in the model do not mean that the basic concept of the auditory filter is wrong. On the contrary, this concept has proven to be very useful for studying a wide variety of psychophysical phenomena [for example loudness (Moore et al. 1997), "resolvability" of individual harmonics of complex tones (Bernstein 2006), certain aspects of hearing impairment (Patterson et al. 1982)] and is a fundamental stage of models at the basis

of many useful applications (for example music or speech coding, noise reduction, hearing aids).

Despite the importance of the notched-noise method, its physiological validity has never been directly tested. In this study, we defined and measured masked thresholds of auditory-nerve fibers in anesthetized cats using the same stimuli as in psychophysics, and compared neural auditory filters derived by the notched-noise method to traditional measures of frequency selectivity in auditory-nerve fibers such as pure-tone tuning curves.

A key question is whether the frequency selectivity observed in human psychophysics experiments is primarily determined by peripheral tuning. We approached this issue by developing a quantitative description of auditory filters parameters for individual fibers that allowed us to constrain a neural population model. The population model was used to generate predictions of psychophysical notched-noise masked thresholds (largely unavailable for the cat), which allowed us to test whether predicted psychophysical auditory filters match the corresponding neural filters in the cat.

3.2 MATERIALS AND METHODS

Procedure

Methods for recording from auditory-nerve fibers in anesthetized cats were as described by Kiang et al. (1965) and Cariani and Delgutte (1996a).

Healthy adult cats of either sex and free of middle-ear infection were used in these experiments. Cats were anesthetized with Dial in urethane (75 mg/kg), with supplementary doses given as needed to maintain an areflexic state. The posterior portion of the skull was removed and the cerebellum retracted to expose the auditory nerve. The ear canals were transected for insertion of closed acoustic systems. The tympanic bullae and the middle-ear cavities were opened to expose the round window. A silver electrode was positioned at the round window to record the compound action potential (CAP) in response to click stimuli, in order to assess the condition and stability of cochlear function. Throughout the experiment the cat is given injections of dexamethasone (0.26 mg/kg) to prevent brain swelling, and

Ringer's solution (50 ml/day) to prevent dehydration. All animals remained anesthetized at all times until death or euthanasia.

The cat was placed on a vibration-isolated table in an electrically-shielded, temperature-controlled, sound-proof chamber. Sound was delivered to the cat's ear through a closed acoustic assembly driven by an electrodynamic speaker (Realistic 40-1377). The acoustic system was calibrated to allow accurate control over the sound-pressure level at the tympanic membrane. Stimuli were generated by a 16-bit digital-to-analog converter (Concurrent DA04H) using sampling rates of 20 kHz, 50 kHz or 100 kHz, depending on CF. Stimuli were digitally filtered to compensate for the transfer characteristics of the acoustic system.

Spikes were recorded with glass micropipettes filled with 2 M KCl. The electrode was inserted into the auditory nerve and then mechanically advanced using a micropositioner (Kopf 650). The electrode signal was bandpass filtered and fed to a custom spike detector. The times of spike peaks were recorded with 1-μs resolution using custom circuits and saved to disk for subsequent analysis.

A click stimulus at approximately 55 dB SPL was used to search for single units. Upon contact with a fiber, a frequency tuning curve was measured by an automatic tracking algorithm (Kiang et al. 1970) using 50-ms tone bursts, presented at a rate of 10/s, and the characteristic frequency (CF) was determined. The spontaneous firing rate of the fiber was measured over an interval of 20s. The responses of the fiber to notched-noise stimuli were then studied.

Stimuli

Stimuli were pure tones in Gaussian band-reject noise, with rejection bands ("notches") placed both symmetrically and asymmetrically around the tone frequency. The signals were 100-msec tone bursts at 0.5, 0.75, 1, 1.4, 2, 2.8, 4, 6, 8, 12, 16 kHz (an half-octave spacing), whichever of these frequencies was closest to the CF of the fiber, estimated from the tuning curve (although a one-octave spacing was used in some early experiments). The fixed signal level (Rosen and Baker 1994) was chosen to be 10-15 dB above threshold at the tone frequency. The noises were band-reject noises with the rejection band placed either symmetrically at the signal frequency, or centered 10% above or 10% below the signal

frequency. These asymmetric notch placements are essential for deriving auditory filter shapes without assuming that the filters are symmetric. The width of the rejection band was varied from 0% to 200% of the signal frequency, in steps of 20%. The upper passband of the noise extended to 4 times the signal frequency. One interval, containing both the tone and the masker, alternated with one interval containing only the masker. The masker duration was 124 ms, with cosine-square rise-fall time of 12 ms, and the signal occupied the central 100 ms of the masker (Figure 3.1A). The Gaussian random noises were synthesized by first creating the desired spectrum in the frequency domain (real and imaginary part), and then inverse Fourier transforming and windowing to the desired length.

Experiment Paradigms

For each notch condition, an automated tracking algorithm (PEST, Taylor and Creelman 1967) was employed to determine the threshold noise level which just masks the rate response to the tone. Specifically, the number of spikes in response to signal in noise had to exceed the spike count for the noise alone on 75% (threshold criterion) of stimulus presentations for the signal to be detected. At the end of the experiment, a cumulative Gaussian distribution (error function) was fitted to the measured samples of the "neurometric" function (percent correct as a function of masker level) by the maximum-likelihood method (Hall, 1968) to obtain a more reliable threshold estimate (Figure 3.2). This "*simultaneous masking*" paradigm is comparable to those used in psychophysics.

One of the assumptions of the power spectrum model of masking is that masking occurs only if the masker contributes significant energy at the output of the auditory filter tuned to the signal frequency, thus "swamping" the response in the channel dedicated to the detection of the signal. This type of masking is commonly referred to as "*excitatory*" masking. However, there is evidence that a masker can "*suppress*" the response of an AN fiber to a tone signal even if it does not, by itself, elicit any excitatory response in that fiber (Sachs & Kiang 1968; Costalupes et al. 1984; Delgutte, 1990). Excitatory and suppressive masking are not mutually-exclusive phenomena and can occur together (Costalupes et al. 1984; Delgutte, 1990). Excitatory masking dominates when the signal and the masker are close in frequency, while suppressive masking is significant for masker frequencies far from the signal frequency, particularly at high masker levels (Delgutte 1990a).

To estimate the contribution of suppressive masking, when possible thresholds were also measured using a non-simultaneous, "*pseudo-masking*" paradigm, which differs from the simultaneous masking paradigm described above in that the tone alone alternates with the masker alone (Figure 3.1B). Masking per se is does not occur in this paradigm, because signal and masker are sufficiently separated in time. However, non-simultaneous threshold corresponds to the masker level that at which the response to the signal alone exceeds the response to the masker by a just-detectable increment, effectively representing the threshold that would be measured if masking were due exclusively to an excitatory mechanism. Therefore, the difference between simultaneous and non-simultaneous threshold is an estimate of the contribution of suppression to simultaneous masking (Delgutte 1990a).

Single-fiber analysis

Linear auditory-filter models were fit to single-fiber neural masked thresholds by assuming that, for each fiber, threshold corresponds to a constant signal-to-masker power ratio at the filter output for all notch widths (a "physiological adaptation" of the power spectrum model of masking). A Rounded Exponential ("Ro-Exp") function (Rosen et al. 1998; Patterson et al.1982), was used to express the power spectrum of the auditory filters. The general expression is the following:

$$|H(f)|^2 = (1-w)(1+p\frac{(f_c-f)}{f_c})e^{-p\frac{(f_c-f)}{f_c}} + w(1+t\frac{(f_c-f)}{f_c})e^{-t\frac{(f_c-f)}{f_c}} \quad \text{for } f < f_c$$

$$|H(f)|^2 = (1+q\frac{(f-f_c)}{f_c})e^{-q\frac{(f-f_c)}{f_c}} \quad \text{for } f > f_c$$

f_c is the center frequency of the filter, while p and q independently control the low- and high-frequency slopes, respectively. On the low frequency side of f_c, the "tail" of the auditory filter is approximated by a second, shallower rounded-exponential term whose weight and slope are controlled by w and t, respectively.

Fits provided by three different versions of this model (Figure 3.3) were compared: a symmetric model without tail ($w = 0$; $p = q$), characterized by 2 free parameters (f_c and p); an asymmetric model without tail ($w = 0$; $p \neq q$), characterized by 3 free parameters (f_c, p and q)

and a complete model (asymmetric with tail), characterized by 5 free parameters (f_c, p and q, w and t). The best-fitting model was selected by minimizing the variance of the residuals, which takes into account the number of degrees of freedom of each model.

"*Bootstrap*" resampling (Efron and Tibshirani 1993) was performed on the data recorded from each fiber, in order to evaluate the statistical properties of the results and derive error bars on the parameters of the fitted filters. In particular, for each notch width and noise level, two-hundred resampled data sets were generated by drawing with replacement from the set of original spike trains. Auditory filters were fit independently to sets of thresholds computed from each bootstrap data set.

Population model

A neural population model was developed to allow a rigorous comparison between simulated psychophysical and physiological auditory filters. Filter models for single fibers cannot be compared directly with psychophysical auditory filters because of the lack of cat behavioral masking data from experiments using notched noise. Predictions of behavioral masked thresholds were generated, based on the assumption that behavioral threshold corresponds to the best (highest) threshold in the entire simulated neural population (Delgutte, 1990).

"*Pseudo-psychophysical*" auditory filters were predicted using the following procedure:

1. A neural filter population was simulated, consisting of 250 logarithmically equally spaced center frequencies (CFs) over the range 500 Hz - 16 kHz. The parameters of the population filters at each CF were obtained by sampling uniformly within ± one standard deviation of the mean of the single-fiber model parameters for all fibers in an octave-wide band including the CF.
2. Neural masked thresholds for a given tone signal were computed as a function of noise condition for all the filters in the population. Threshold was assumed to correspond to a constant signal-to-noise ratio (SNR) at the output of each filter. For each filter, the SNR at threshold was set to the mean value for all fibers in the corresponding frequency band.
3. For each noise condition, the predicted population masked threshold was defined as the maximum masked threshold (meaning best performance) in the entire population.

While it is theoretically possible for an ideal processor to achieve better performance than that of any single neuron (Green 1958, Siebert 1968), this simple rule has the advantage of requiring no assumptions about correlation between the discharge patterns of different neurons.

4. A pseudo-psychophysical filter model was fit to the set of predicted population masked thresholds using the same method as for the neural data.

The entire procedure was repeated 100 times to obtain measures of the variability of the results.

3.3 RESULTS

Single-fiber results

Our results are based on responses to tones in notched noise recorded from 47 auditory-nerve fibers in 7 cats. Of these, 31 had high SR (> 18 spikes/s), 12 had medium SR (between 0.5 and 18 spikes/s) and 4 had low SR (< 0.5 spike/s). The CFs of the fibers ranged from 200 Hz to 23 kHz.

Figure 3.4 shows results for two auditory-nerve fibers with CFs of 1454 Hz (A) and 6775 Hz (B), respectively. The frequency of the pure tone was 1000 Hz for the low-CF fiber and 8 kHz for the high-CF fiber. The tones were presented at 12 and 10 dB, respectively, above threshold at the tone frequency. Masked thresholds (black) are plotted as a function of notch width for three different placements of the notch center frequency with respect to the tone frequency (see Methods). For the low-CF fiber (A), a Ro-Exp model without tail gave accurate predictions (orange) of masked thresholds. For the high-CF fiber (B), on the other hand, a model without a low-frequency tail fit poorly (orange), while a complete model including a low frequency tail (blue) was in this case necessary to capture the main trend in the dependence of masked thresholds on notch configuration.

Panels C and D of Figure 3.4 show comparisons of the best-fitting neural auditory-filters (orange and blue, as in A and B) to pure-tone tuning curves (black) measured from the same fibers. Note that these are not fits, but rather two independent measures of frequency selectivity obtained with very different stimuli. The asymmetrical Ro-Exp filter without tail

(orange) crudely approximates the pure-tone tuning curve for the low-CF fiber (C), while only a complete Ro-Exp model with tail (blue) can mimic the low-frequency portion of the tuning curve of a high-CF fiber (D).

Figure 3.5A shows the overall errors (r.m.s., the square root of the mean square error) of the best-fitting Ro-Exp models in predicting neural masked thresholds across all the fibers in our sample. R.m.s. errors were typically below 4 dB throughout the 0.2-23 kHz range of CFs, while no obvious trend can be detected in the values of the errors as a function of fibers CF. Figure 3.5B shows the fraction of the variance in the data that was accounted for by the best-fitting Ro-Exp model for each fiber in our sample (R^2, or "R-square"). R-squares were greater than 0.8 in 96% of the cases, indicating that linear filter models are applicable to predict masked thresholds in notched noise of AN fibers in the cat.

Different versions of the Ro-Exp model gave the best fits to neural masked thresholds depending on the CF range (Figure 3.6). For low-CF (< 1 kHz) fibers, as in the example of Fig. 3.4 (A,C), asymmetrical models without a low frequency tail (orange) gave the best fits in 12 of 15 cases. On the other hand, for high-CF (> 3 kHz) fibers, as in the example of Fig. 3.4 (B,D), a model with a low frequency tail (blue) was necessary to obtain the best fits in 13 of 15 cases. For CFs between 1 kHz and 3 kHz, any of the 3 models could give the best fit, depending on the fiber.

Figure 3.7 shows that the center frequencies of the best-fitting model neural filters are in very good agreement with the CFs estimated from the pure-tone tuning curves (correlation coefficient = 0.994, $p < 0.0001$). This is a sanity check of the validity of our approach: although auditory filters estimated with the notched-noise method are not constrained to have the same center frequency as the CF of tuning curves obtained with pure tones, a large discrepancy between the two would certainly raise some questions about the validity of the method.

A commonly-used measure of sharpness of tuning for both the model filters and pure-tone tuning curves is the Q_{ERB}, defined as the ratio of the center frequency (or CF) to the equivalent rectangular bandwidth (ERB) (the ERB of a filter is defined as the bandwidth of an ideal rectangular filter which passes the same amount of energy as the filter for white noise inputs and has the same maximum gain). Figure 3.8 shows Q_{ERB}s of model filters and tuning curves (black diamonds for the tuning curves and blue for the Ro-Exp filters) as a

function of CF for single-fibers. Solid lines show the median values of the single-fiber Q_{ERB}s across 6 non-overlapping octave-wide frequency bands centered at 0.5, 1, 2, 4, 8, and 16 kHz, respectively. Error bars span the interquartile ranges of the Q_{ERB}s across the same frequency bands. Q_{ERB}s of model filters (blue) and tuning curves (black) are very similar throughout the entire frequency range and they increase systematically with CF, consitent with previous results suggesting a progressive sharpening of the relative bandwidth of cochlear filters with respect to their center frequency (Kiang et al. 1965; Shera et al. 2002). Analysis of covariance was performed on the data of Fig. 3.8 to quantify whether there was a significant dependence on CF, group (tuning curves or model filters), or a combination of both. The result of this analysis showed that the Q_{ERB}s increased significantly ($p = 0$) with CF, and their values were fit by a straight line (red) on log-log coordinates whose slope (0.49 $\pm$ 0.07) and intercept were not significantly different for model filters than for tuning curves ($p = 0.63$).

Effect of suppression

In order to quantify the contribution of suppression to masked thresholds and to evaluate its effect on the parameters of auditory filters, we compared thresholds measured with the simultaneous-masking and a non-simultaneous, "*pseudo-masking*" paradigm (see Methods).

Figure 3.9 shows results for a fiber with a CF of 1100 Hz. Thresholds measured with the non-simultaneous paradigm (squares) are generally higher (meaning less masking) than those obtained with the simultaneous paradigm (diamonds). Differences between non-simultaneous and simultaneous thresholds are effectively estimates of the contribution of suppression to simultaneous masking (Delgutte 1990a). These differences correspond to the additional amount of noise power needed to just mask the signal in the non-simultaneous condition, designed to be equivalent to one in which masking were exclusively excitatory. Non-simultaneous thresholds also increase more rapidly with notch width than simultaneous thresholds, consistent with observations that suppressive masking is increasingly significant for masker frequencies farther from the signal frequency (Delgutte 1990a). Correspondingly, the auditory filter estimated with the non-simultaneous pseudo-masking paradigm is sharper than the one derived using the simultaneous masking paradigm.

This trend was observed for the majority of the fibers for which we were able to measure masked thresholds in the two conditions at the same level (Figure 3.10). In 13 of 21 cases (CFs ranging from 975 to 22834 Hz), the ERB of filters derived with the simultaneous masking paradigm were significantly ($p < 0.001$) wider than those of non-simultaneous filters (paired t-test on the sets of ERBs obtained over 100 bootstrap replications of the data for each fiber). Surprisingly, in 4 cases (CFs of 340, 4950, 6775 and 22854 Hz; all high spontaneous rate fibers), the opposite was true, i.e. the simultaneous filters was significantly narrower than the corresponding non-simultaneous ones. Finally, in 4 cases (CFs of 1450, 1520, 2052 and 4750 Hz, respectively; two high- and two medium- spontaneous rate fibers) the ERBs obtained using the two paradigms were not significantly different.

Population model and "pseudo-psychophysical" filters

Pseudo-psychophysical auditory filters were fit to sets of masked thresholds predicted from a population of neural auditory filters (see Methods), for tone signals ranging from 750 Hz to 14 kHz. The shape of each of the population filters was selected to be consistent with the trend observed for individual fibers (Figure 3.6): filters were asymmetrical without tail below 1 kHz, asymmetrical with low-frequency tail above 5 kHz, while filters between 1 kHz and 3 kHz were randomly selected among the three models. In order to evaluate the statistical properties of the results, one hundred populations of neural auditory filters were generated (see Methods), and one example is shown in Figure 3.11. As for the neural data, the same three versions of the Ro-Exp function were used to estimate the frequency response of the pseudo-psychophysical auditory filters.

Figure 3.12 shows two examples of the results of this simulation, for tone frequencies of 750 Hz (A) and 8000 Hz (B), respectively. In both cases the tone level was set at 20 dB SPL. Predictions of behavioral masked thresholds (black) are plotted as a function of notch width for three different placements of the notch center frequency with respect to the tone frequency. Purple solid lines show thresholds for the best fitting pseudo-psychophysical auditory-filters. At the lower frequency (A), the best-fitting model was asymmetrical without tail, while at the higher frequency (B), a model including a low frequency tail produced the best fits. Panels C and D show the transfer function of the best-fitting filters.

In general, pseudo-psychophysical filters produced excellent fits (1.3 - 4.8 dB r.m.s. error, R-squares all greater than 0.98) to behavioral masked thresholds predicted from the simulated neural populations throughout the 0.75 - 14 kHz frequency range. Similar to the trend observed for neural filters (Fig. 3.6), asymmetrical filters without tail gave the best fits for tone frequencies below 3 kHz, while filters with a low-frequency tail were necessary to fit predicted behavioral masked threshold at frequencies above 3 kHz. The center frequencies of the pseudo-psychophysical filters were very similar (median absolute deviation equal to 0.7%) to the frequencies of the corresponding tone stimuli.

Figure 3.13 shows how Q_{ERB}s of pseudo-psychophysical auditory filters (purple) compared to Q_{ERB}s of neural auditory filters (blue). For the pseudo-psychophysical filters, error bars indicate the interquartile ranges of the Q_{ERB}s across 100 simulated neural populations. Q_{ERB}s of the pseudo-psychophysical filters are in line with those of neural filters throughout the entire frequency range, the only small difference being at the lowest CFs, where pseudo-psychophysical filters appear to be slightly sharper (larger Q_{ERB}) than the corresponding neural filters. These findings are similar to existing data comparing behavioral and neural measures of frequency tuning using rippled and band-reject noise in the guinea pig (Evans et al. 1991). We are unaware of the existence of behavioral estimates of auditory filters using band-reject noise in the cat.

3.4 DISCUSSION

We measured masked thresholds of auditory-nerve (AN) fibers in anesthetized cats for pure tones in band-reject noise at low stimulus levels, and compared neural auditory filters derived by the notched-noise method to pure-tone tuning curves measured in the same fibers.

We found that linear Ro-Exp (rounded-exponential) filter models successfully predict the dependence of masked thresholds on notch configuration for fibers of all CF (Fig. 3.4 and 3.5). Different versions of the Ro-Exp model gave the best fits to thresholds depending on the CF: best-fitting filters were typically asymmetric for fibers with CFs below 1 kHz, while only models including a low-frequency "tail" (mimicking the low-frequency portion of the tuning curves) produced satisfactory fits for fibers with CFs above 3 kHz. For CFs between

1 kHz and 3 kHz, different forms of the model could produce the best fit, depending on the fiber.

The center frequencies of the model filters closely agreed with tuning-curve CFs (Fig. 3.7) despite the fact that the signal frequency did not exactly match the CF and could be as far as half an octave away from the CF). This result supports the robustness of the notched noise method, since signals away from the CF could generate suppressive effects on masker components at frequencies close to the CF (particularly for narrow notch widths) (Delgutte 1990a) that are not consistent with the assumptions of the power spectrum model of masking.

Measures of sharpness of tuning (Q_{ERB}s, the ratio of CF to Equivalent Rectangular Bandwidth) of tuning curves and neural filters were also in good agreement (Fig. 3.8). Q_{ERB}s of both neural filters and tuning curves increased systematically with CF, consistent with the progressive sharpening of peripheral tuning with increasing CF, and this increase was well fit by a power function of CF with an exponent of 0.49 ± 0.07. The 95% confidence interval for this parameter overlaps with both that for the 0.37 (± 0.10) exponent found by Shera et al. (2002) (for the CF dependence of Q_{10} in pure-tone tuning curves from AN fibers in the cat) and that for the 0.37 (± 0.09) exponent found by Cedolin and Delgutte (2005) (for the increase in resolvability of harmonics of complex tones with CF in the cat AN – see Chapter 1).

It is interesting to notice that, despite the highly nonlinear nature of the cochlea, there is broad agreement among values and trends observed in estimates of frequency selectivity (at least at low sound levels) of cat AN fibers obtained using very different stimuli and methods: pure-tone tuning curves (Kiang et al.1965; Shera et al. 2002), broadband-noise reverse-correlation techniques (de Boer and de Jongh 1978; Carney and Yin 1988), synchrony to pairs of tones (Greenberg et al.1986), rippled noise (Evans & Wilson 1973), complex tones (Cedolin and Delgutte 2005) and masking by notched-noise (present study).

One important aspect of peripheral frequency selectivity that was not addressed in our study is its level dependence. We typically were not able to "hold" fibers long enough to obtain data at different levels. However, even if holding time were not an issue, measuring thresholds at high levels (which would lead to an inaccurate assessment of cochlear frequency selectivity) would be hard because of rate saturation, especially for fibers with high-SR, due to their limited dynamic ranges. An alternative strategy to compare frequency

tuning at different levels would be looking at data obtained from low-threshold, high-SR fibers at low levels and from high-threshold, low-SR fibers at high levels (ideally in the same animal). The very small number of fibers with high threshold and low spontaneous rate in our sample prevented us from performing this comparison.

To quantify the contribution of suppression to masked thresholds and to evaluate its effect on the model filters parameters, we compared thresholds measured with the simultaneous-masking and a non-simultaneous, "*pseudo-masking*" paradigm, designed to be equivalent to one in which masking is exclusively excitatory. For a majority of fibers (Fig. 3.9), thresholds measured with the non-simultaneous paradigm were better than those obtained with the simultaneous paradigm, suggesting a contribution of suppression to simultaneous masking (Delgutte 1990a). Differences in threshold, and therefore the relative amount of suppressive masking, increased with notch width, consistent with previous results with pure tones (Delgutte 1990a) showing large suppression for masker frequencies farther from the signal frequency. In these cases, auditory filter estimated with the non-simultaneous pseudo-masking paradigm were somewhat sharper than those derived using the simultaneous masking paradigm (Fig. 3.10). This result suggests that frequency selectivity estimates derived in psychophysics using forward masking paradigms (e.g. Glattke 1967; Houtgast 1972; Moore 1978), designed to minimize the effect of suppression, may be more accurate than those derived using simultaneous masking paradigms. For some fibers, auditory filters bandwidths estimates with the two paradigms were not significantly different, indicating weak or no suppression, and in 4 of 21 cases, simultaneous filters were sharper than non-simultaneous filters. Interestingly, the only CF-range consistently showing an effect of suppression was the one between 600 Hz and 2 kHz (Fig. 3.10), in contrast with the findings of previous studies showing weak suppression for low-CF fibers compared to high-CF fibers in cat (Delgutte 1990b), chinchilla (Harris 1979) and gerbil (Schmiedt 1982).

An important issue is whether the frequency selectivity observed in human psychophysical experiments is primarily determined by peripheral tuning or whether it is sharpened by some more central neural mechanism. We approached this issue by using typical auditory filters parameters for individual fibers to simulate the performance of a neural population in the detection of pure tones in notched noise. Specifically, the population model was used to generate predictions of psychophysical masked thresholds in notched

noise based on the simplifying assumption that, for each noise condition, the predicted population masked threshold corresponds to the maximum masked threshold (meaning best performance) in the entire population. By fitting "pseudo-psychophysical" auditory filters to masked thresholds based on the population model, we were able to compare the predicted psychophysical tuning sharpness to that of the underlying neural filters in each frequency region. The results of this analysis suggest that the two are indeed similar at low levels, thus supporting the use of the notched noise method to measure tuning sharpness from psychophysical masking data.

3.5 CONCLUSIONS

We tested the physiological basis of the *notched-noise method,* widely used in psychophysics to estimate peripheral frequency selectivity. Neural auditory filters were fit to masked thresholds for pure tones in band-reject noise. Estimates of neural frequency selectivity derived using the notched-noise method show an increase in tuning sharpness with CF that is similar to that obtained in previous studies using a variety of different stimuli and techniques and is also in broad agreement with that inferred from results related to harmonic resolvability in Chapter 1. A neural population model was used to test the extent to which auditory filters estimated psychophysically match the underlying neural filters. The similarity observed between the frequency selectivity measured physiologically and the predicted psychophysical tuning sharpness supports the use of the notched-noise method in human psychophysics.

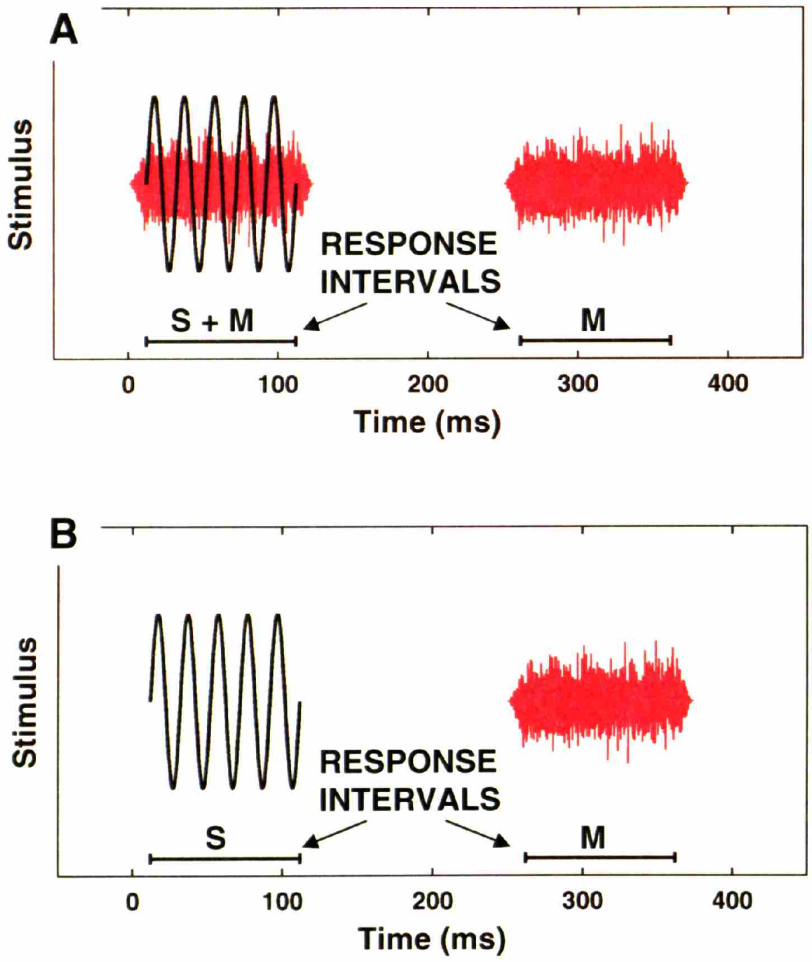


Figure 3.1: Simultaneous (A) and Non-Simultaneous (B) masking paradigms. In the simultaneous masking paradigm, and interval containing both a pure tone (black) and a Gaussian noise masker (red) alternated with one interval containing only the masker. In the non-simultaneous masking paradigm, designed to quantify the contribution of suppression, the first interval contained only the pure tone. The duration of each noise burst was 124 ms, with cosine-square rise-fall time of 12 ms, and the signal occupied the central 100 ms of the masker.

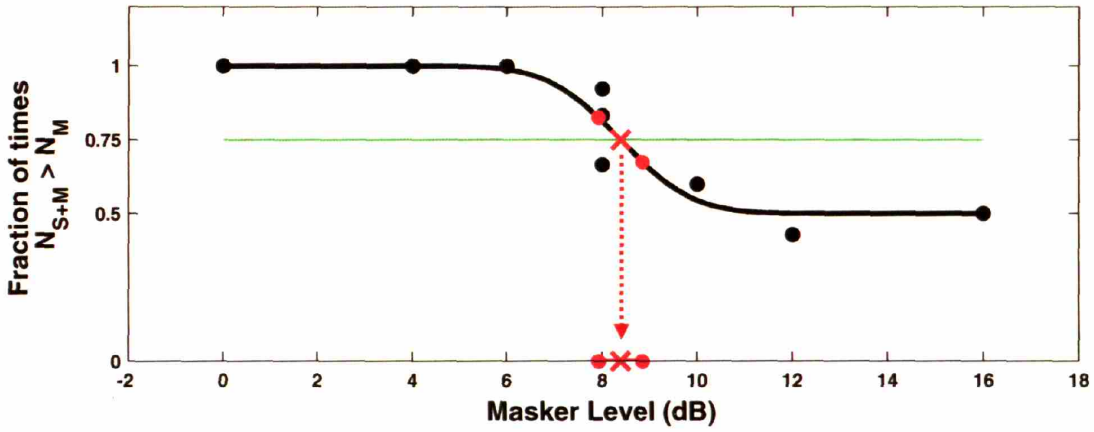


Figure 3.2. Threshold determination. During the experiment, a PEST tracking algorithm (Taylor and Creelman 1967) was employed to determine the threshold (red cross) for each notch condition. The number of spikes in response to signal in noise had to exceed the spike count for the noise alone on 75% (green) of stimulus presentations. At the end of the experiment, a cumulative Gaussian distribution (error function, black solid line) was fitted to the measured samples of the neurometric function by the maximum-likelihood method (Hall, 1968). An estimate of the threshold error was determined by calculating the mean error ME between the cumulative Gaussian and the measured samples of the neurometric function, and determining the two masker levels (red circles) at which the cumulative Gaussian assumed values equal to 0.75 ± ME.

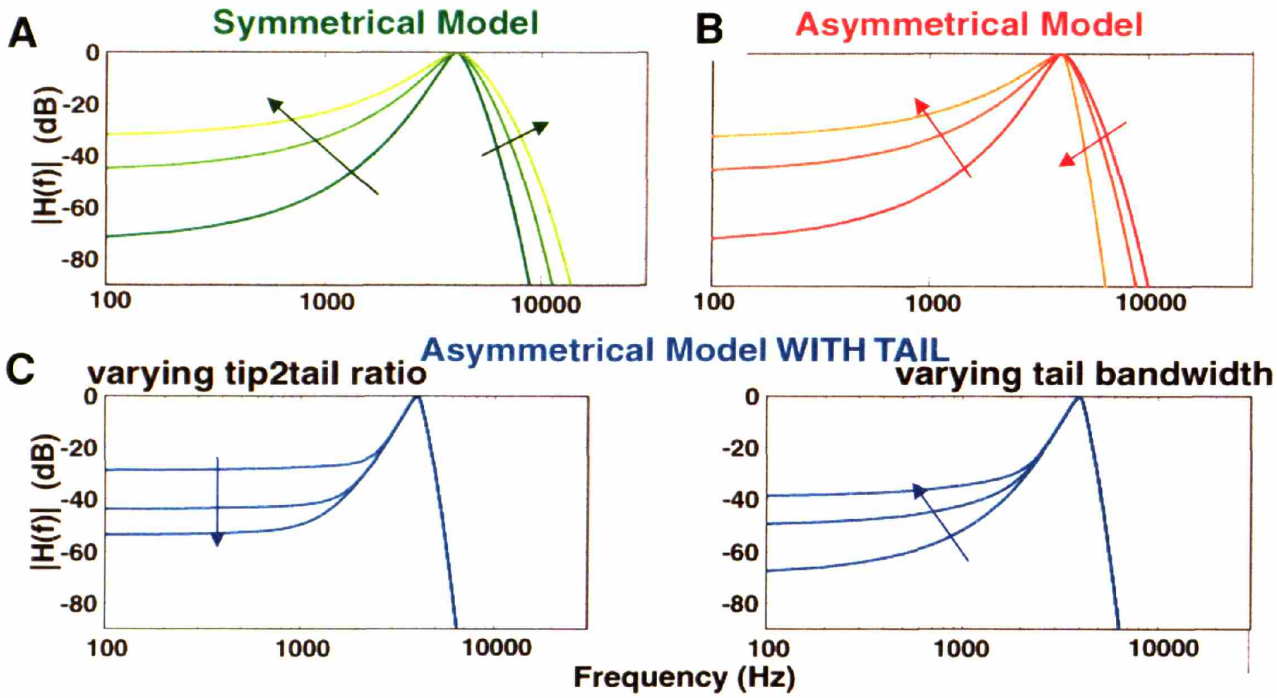
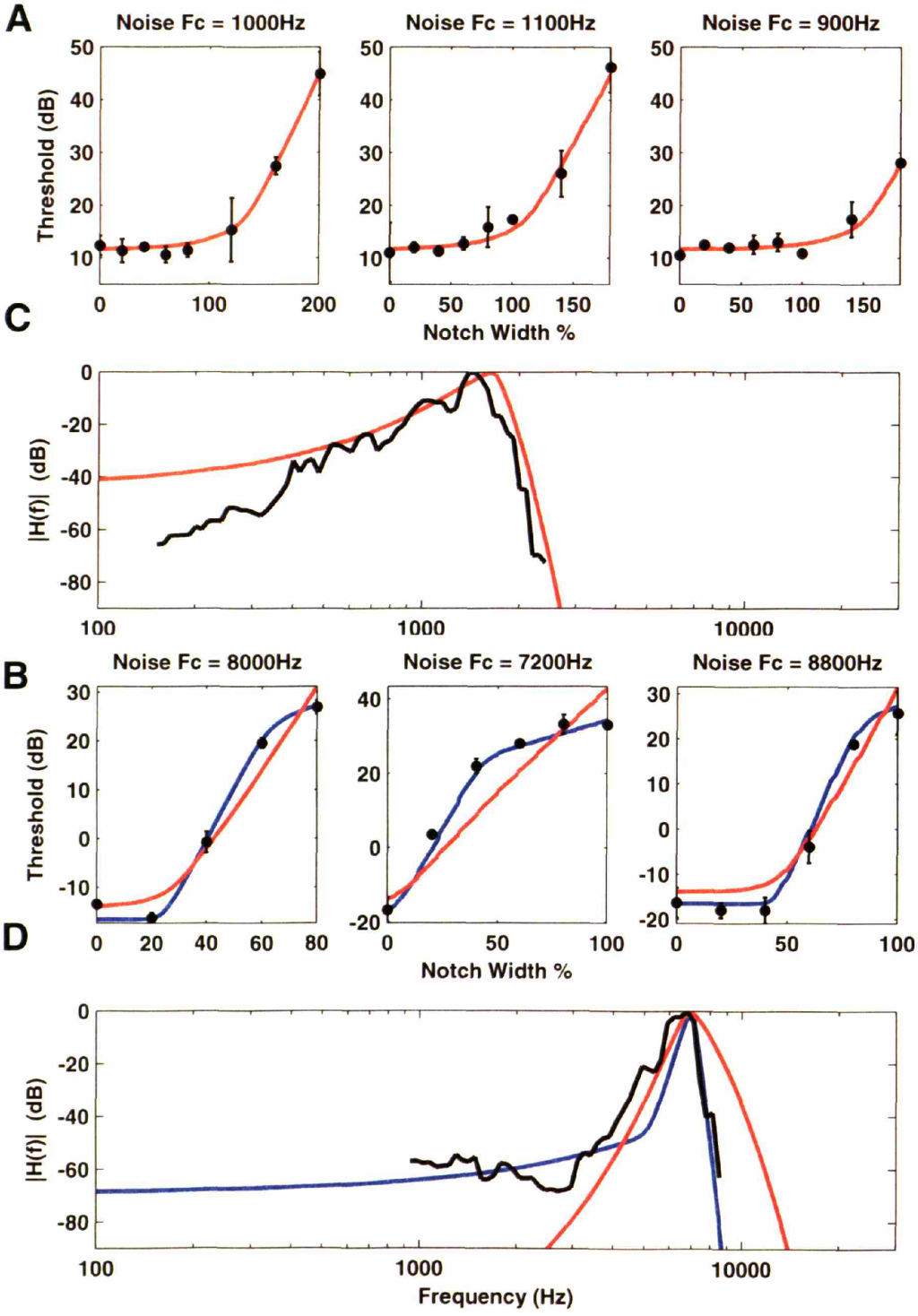


Figure 3.3. A: symmetric Ro-Exp model without low-frequency tail ($w = 0$; $p = q$). B: asymmetric Ro-Exp model without tail ($w = 0$; $p \neq q$). C: complete Ro-Exp model (asymmetric with low-frequency tail).

Figure 3.4 (next page): Masked thresholds (black circles) and the fits provided by the asymmetrical Ro-Exp models with (blue) and without (orange) low-frequency tail. A, C: fiber CF = 1454 Hz; tone frequency = 1 kHz; tone level = 12 dB above threshold at 1 kHz. B, D: fiber CF = 6775 Hz; tone frequency = 8 kHz; tone level = 10 dB above threshold at 8 kHz. A, B: thresholds and fits are plotted as a function of notch width (expressed as a percentage of the tone frequency) in three separate panels for notches placed symmetrically and asymmetrically (10% above and 10% below) with respect to the tone frequency. C, D: comparisons between the pure-tone tuning curves (black) and the best-fitting Ro-Exp models with (blue) and without (orange) low-frequency tail.


A
Noise Fc = 1000Hz
Noise Fc = 1100Hz
Noise Fc = 900Hz
Threshold (dB)
Notch Width %
C
|H(f)| (dB)
B
Noise Fc = 8000Hz
Noise Fc = 7200Hz
Noise Fc = 8800Hz
Threshold (dB)
Notch Width %
D
|H(f)| (dB)
Frequency (Hz)

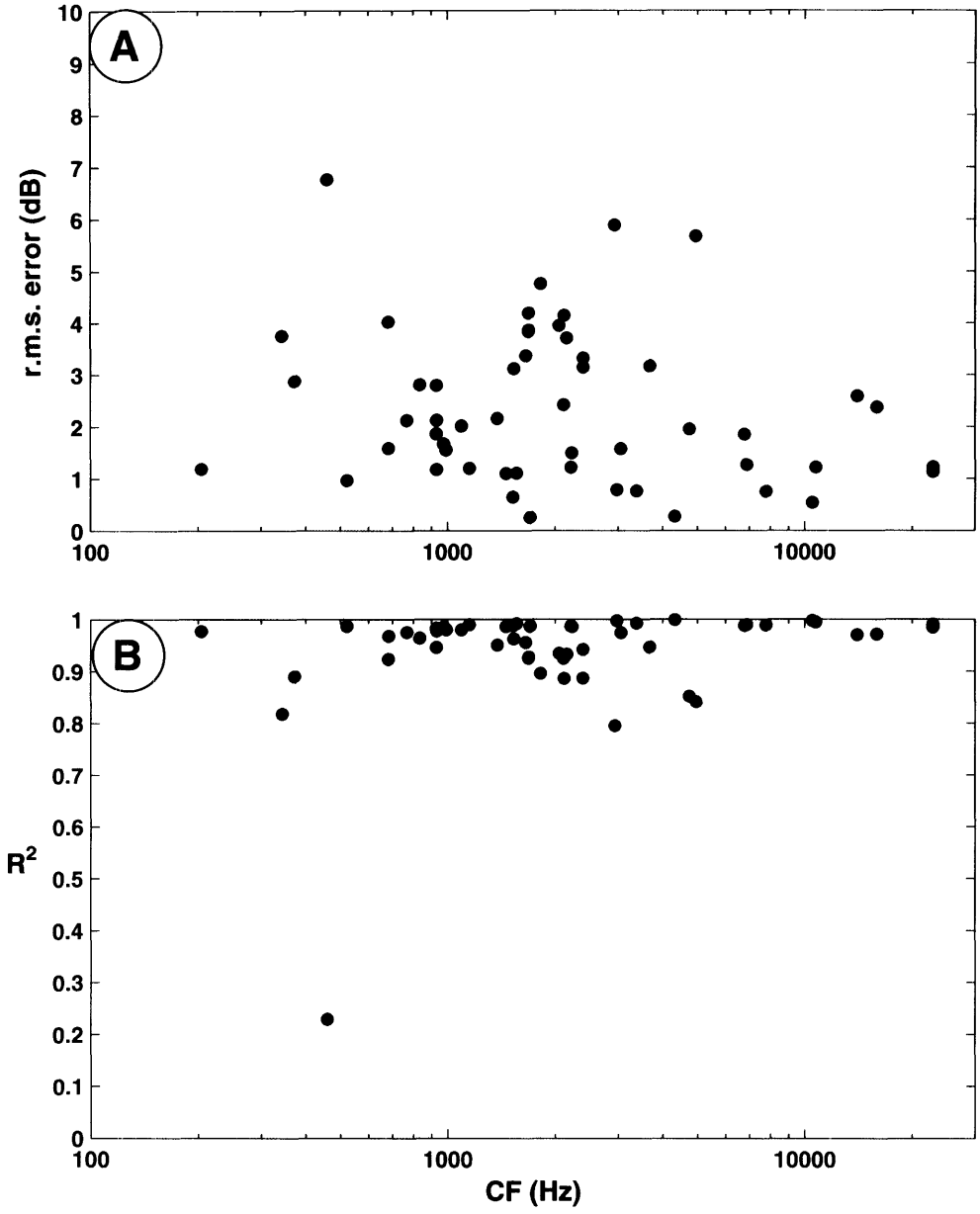


Figure 3.5. A: R.m.s. error (dB) between measured masked thresholds and Ro-Exp model predictions, plotted as a function of fiber's CF. B: R^2 of the best fitting Ro-Exp models, plotted as a function of fiber's CF.

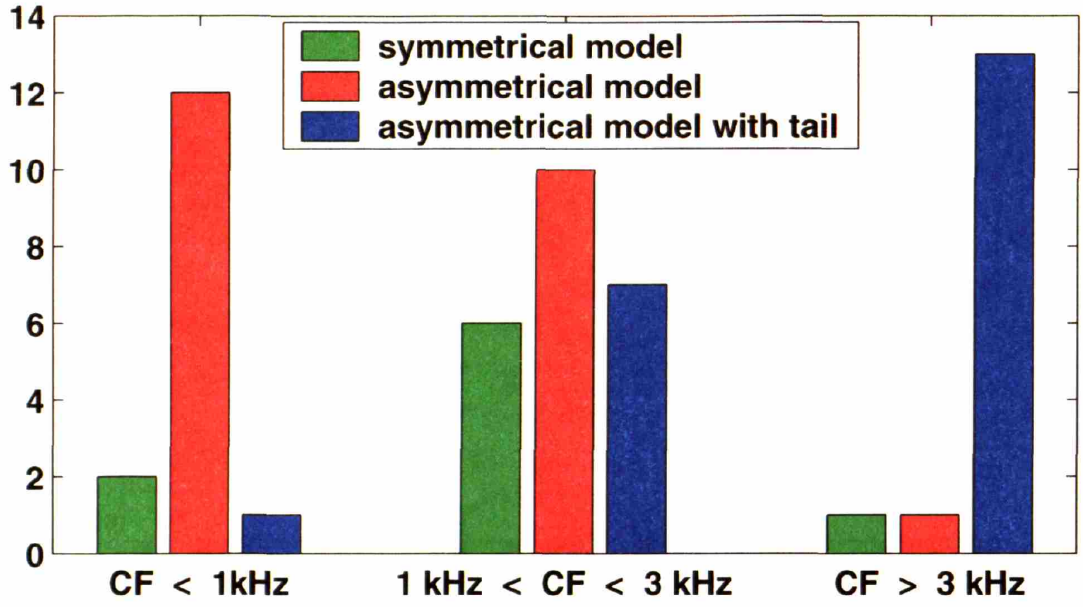


Figure 3.6. Best-fitting Ro-Exp model as a function of fiber's CF.

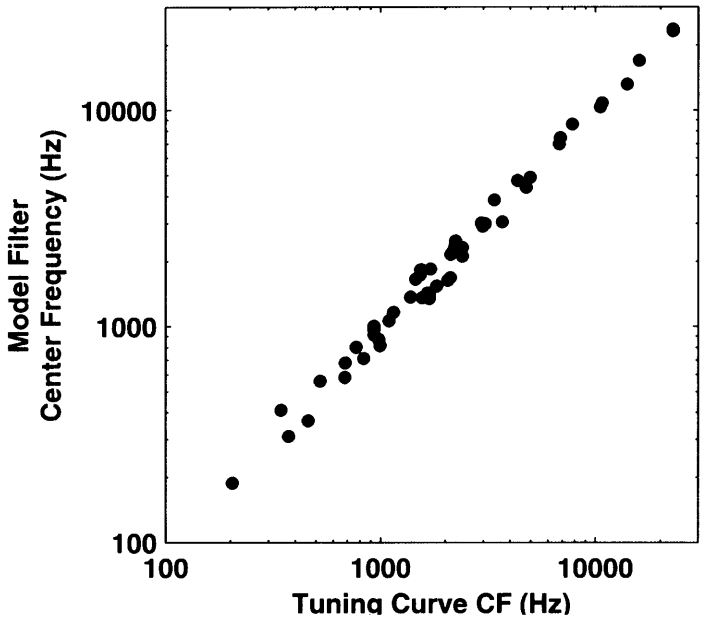


Figure 3.7. Center frequency of the best-fitting Ro-Exp model as a function of tuning curve CF.

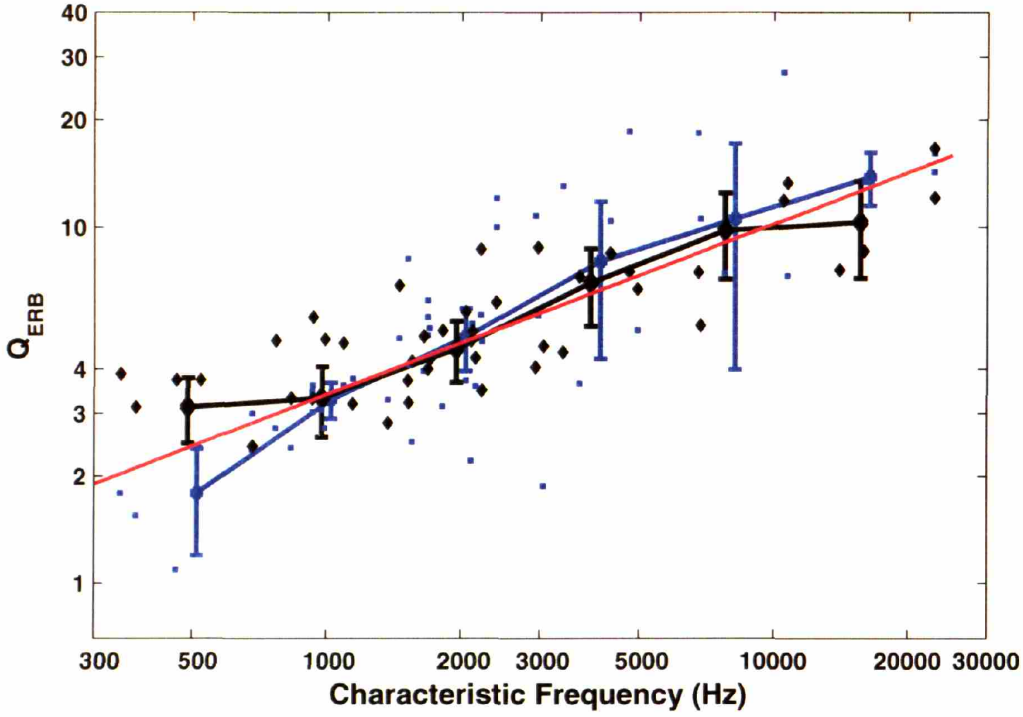


Figure 3.8. Q_{ERB}s of tuning curves (black diamonds) and model neural filters (blue squares) as a function of fibers tuning-curve CF. Medians and interquartile ranges (errorbars) of the Q_{ERB}s across 6 non-overlapping octave-wide frequency bands centered at 0.5, 1, 2, 4, 8, and 16 kHz, respectively. Red line indicates best linear fit across both groups based on analysis of covariance.

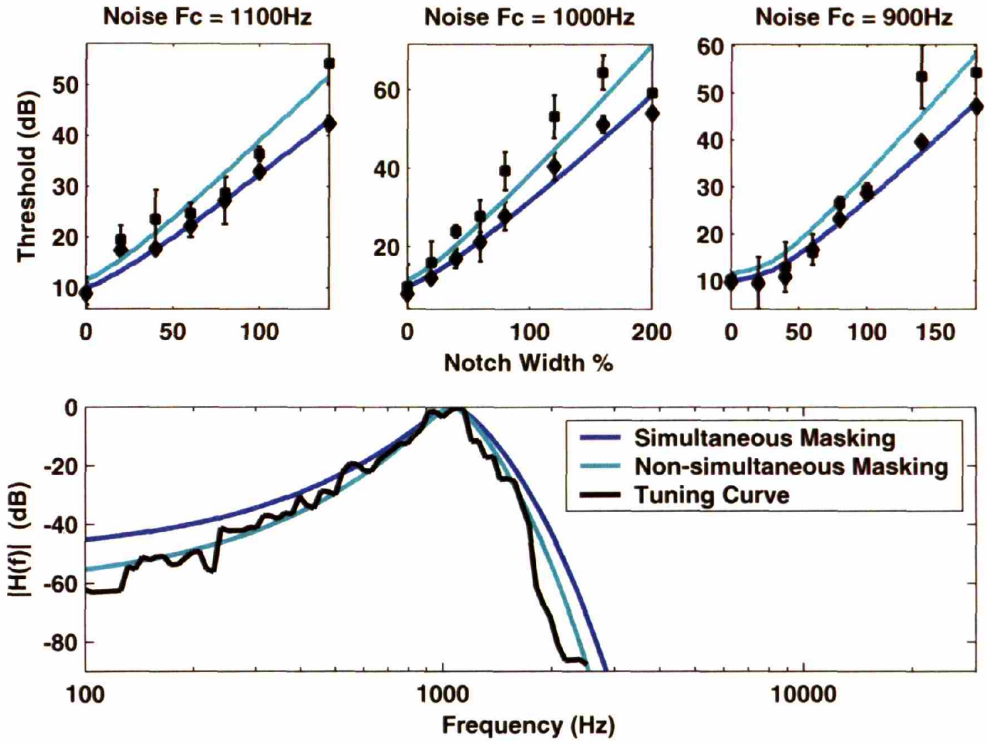


Figure 3.9. Top: comparison between simultaneous (diamonds) and non-simultaneous (squares) thresholds measured for the same fiber (CF = 1100 Hz, tone level = 20 dB above threshold at CF). Bottom: comparison between the pure-tone tuning curve and the best-fitting Ro-Exp models derived with the simultaneous (blue) and the non-simultaneous (cyan) paradigm (see Methods).

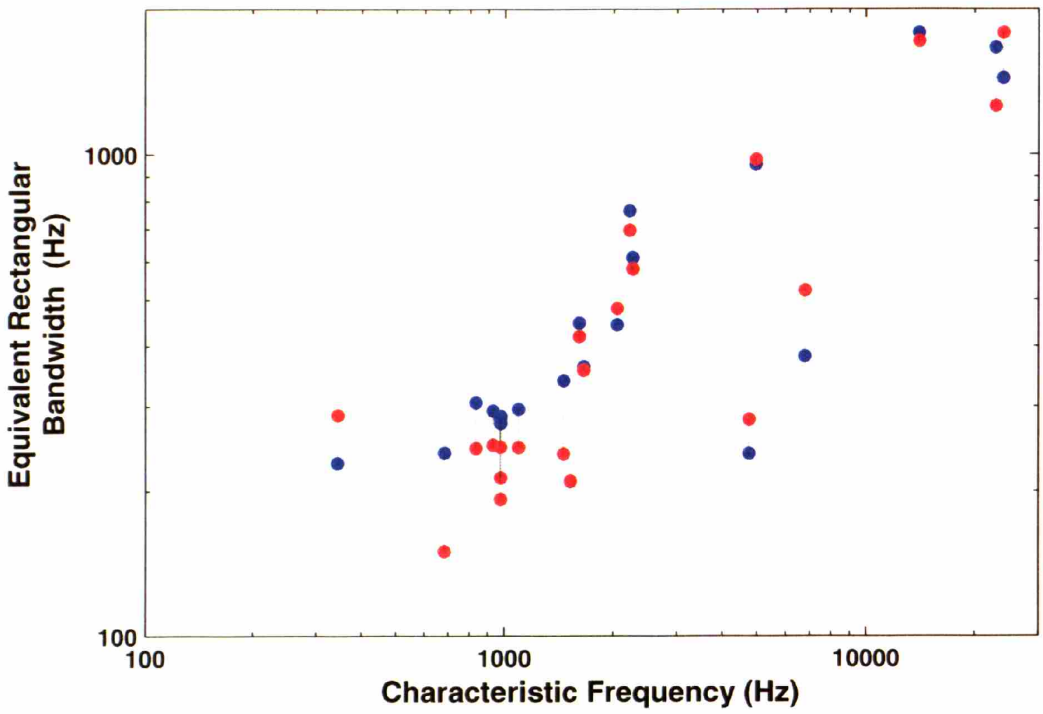


Figure 3.10. Comparison of the equivalent rectangular bandwidths of best-fitting Ro-Exp models derived with the simultaneous (blue) and non-simultaneous (red) paradigm.

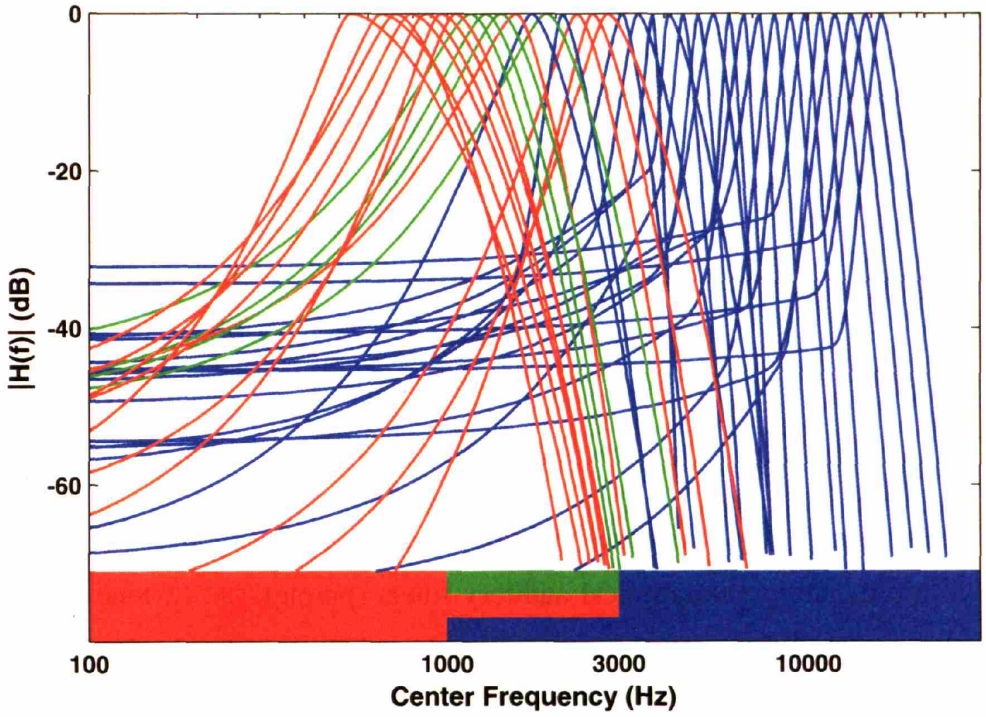
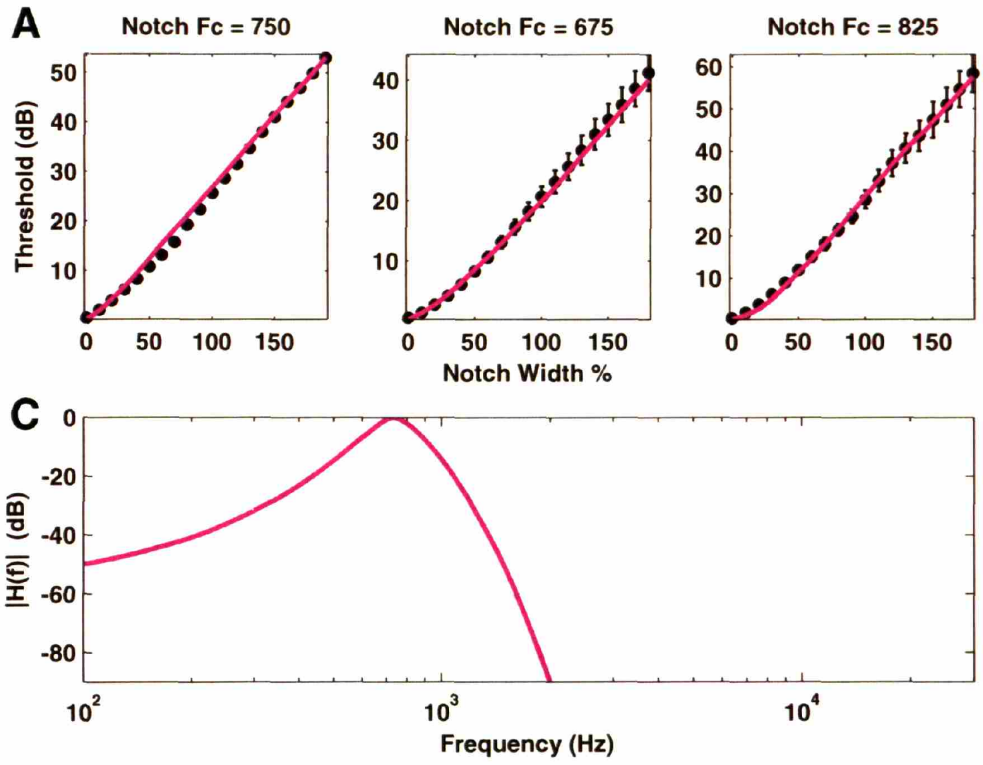
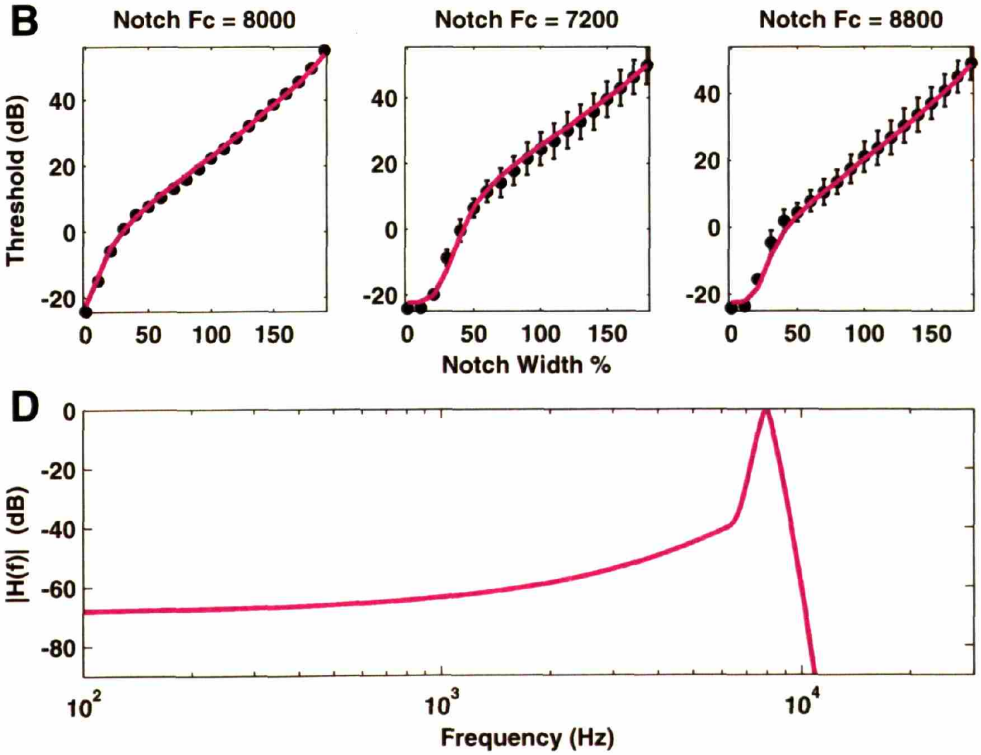


Figure 3.11. Example of a population of neural auditory filters, with center frequencies ranging from 500 Hz to 16 kHz. The shape of each of the population filters was selected to be consistent with the trend observed for individual fibers (Fig. 3.6): filters were asymmetrical without tail (orange) below 1 kHz and asymmetrical with low-frequency tail (blue) above 5 kHz, while between 1 kHz and 3 kHz filters were randomly selected among the three models.

Figure 3.12 (next page): Masked thresholds based on a neural population (black circles) and the fits of the pseudo-psychophysical auditory filters (purple). A, C: tone frequency = 750 Hz; tone level = 20 dB SPL. B, D: tone frequency = 8 kHz; tone level = 20 dB SPL. A, B: thresholds and fits are plotted as a function of notch width (expressed as a percentage of the tone frequency) in three separate panels for notches placed symmetrically and asymmetrically (10% above and 10% below) with respect to the tone frequency. C, D: best-fitting pseudo-psychophysical filters (asymmetrical without (C) and with (D) low-frequency tail).


A
Notch Fc = 750
Notch Fc = 675
Notch Fc = 825
Threshold (dB)
Notch Width %
C
|H(f)| (dB)
Frequency (Hz)


B
Notch Fc = 8000
Notch Fc = 7200
Notch Fc = 8800
Threshold (dB)
Notch Width %
D
|H(f)| (dB)
Frequency (Hz)

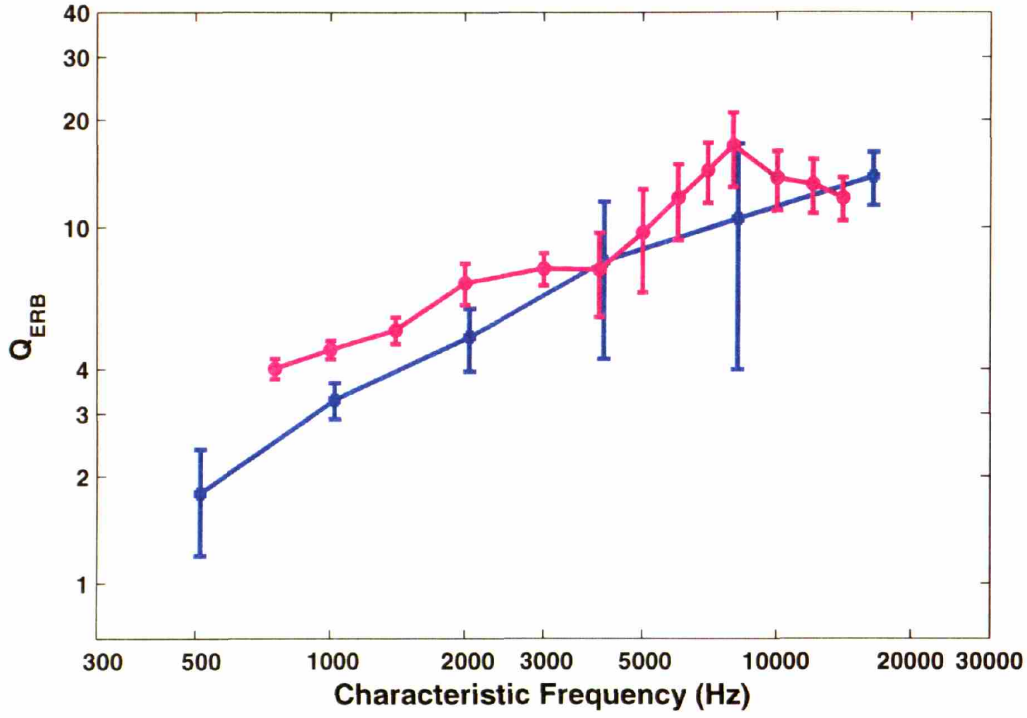


Figure 3.13. Blue: Q_{ERB}s of model neural filters as a function of CF. Medians and interquartile ranges (error bars) of the Q_{ERB}s across 6 non-overlapping octave-wide frequency bands centered at 0.5, 1, 2, 4, 8, and 16 kHz, respectively. Purple: Q_{ERB}s of pseudo-psychophysical model filters as a function of their center frequency. Error bars show interquartile ranges of the Q_{ERB}s across 100 simulated populations.

Chapter 4. Summary and Conclusions

The main focus of this thesis was to investigate possible neural representations of the pitch of harmonic complex tones, with particular attention devoted to discussing how well pitch cues provided by each representation can account for major trends in human psychophysical data on pitch perception. In psychophysics, these trends are often interpreted in terms of the extent to which harmonics are "resolved" (identified and/or processed individually) thanks to the decomposition of incoming sounds into their frequency components performed by the auditory periphery. Given the possible differences between the peripheral tuning sharpness in humans and in experimental animals, we focused on estimating peripheral frequency selectivity and defining harmonic resolvability in the species studied (in our case, the cat), thereby providing quantitative grounds for discussing the results in terms of their implications for human psychophysics. Here, we provide a summary of our main findings.

In Chapter 1, we studied harmonic resolvability based on average-rate responses of AN fibers and investigated two traditional representations of pitch: (1) a rate-place representation of pitch, based on matching "harmonic templates" to the pattern of excitation of an array of tonotopically-organized neurons, and (2) a temporal representation of pitch, based on matching "periodic templates" to the most common interspike intervals of a population of AN fibers.

We found evidence for an increase with CF of harmonic resolving "power" of single AN fibers (expressed as N_{max}, the maximum resolved harmonic number CF/F0), which closely mirrored the increase with CF of tuning sharpness (expressed as a Q-factor, the ratio CF/Bandwidth) for pure-tone tuning curves (Kiang et al. 1965; Liberman 1978; Shera et al. 2002). An increase of N_{max} with CF is also consistent with previous results obtained in the AN with rippled noise (Wilson and Evans 1971) and in the AVCN with complex tones (Smoorenburg and Linschoten 1977). Also consistent with existing AVCN data (Smoorenburg and Linschoten 1977), we found that N_{max} decreases rapidly with small (10-20 dB) increases in stimulus level, a phenomenon that we attributed mainly to rate saturation rather than to broadening of cochlear tuning with level.

We were consistently able to estimate pitch by matching harmonic templates to profiles of average discharge rate against CF for F0s above 400-500 Hz. Below this lower limit, larger errors became increasingly common, although some reliable estimates were obtained for F0s as low as 250 Hz. Both the precision and the predicted strength of rate-based pitch estimates increased with F0, consistent with an increase in the degree of harmonic resolvability due to the progressive sharpening of cochlear tuning with CF. However, since the predicted strength of rate-based pitch estimates continued increasing beyond 1400 Hz, we concluded that a purely rate-place representation cannot account for the existence of an upper limit to the perception of the pitch of missing-F0 harmonic complex tones in humans. Rate-based pitch estimates were largely independent of the relative phases among the harmonics, in agreement with results of human psychophysics experiment for complex tones containing resolved harmonics (Houtsma and Smurzynski 1990; Shackleton and Carlyon 1994). Given the rapid saturation of average-rate responses of the majority of the AN fibers with stimulus level, we deemed it unlikely that a strictly rate-based representation can account for the (relatively) stable human performance in detecting small F0-changes at high levels in the presence of resolved harmonics (Bernstein and Oxenham, 2006)

Pitch estimates derived by matching periodic templates to pooled all-order interspike-interval distributions were highly accurate and precise for F0s up to 1300 Hz. Above this upper limit, the purely temporal representation broke down, arguably due to the degradation of phase-locking to the frequencies of individual harmonics. Thus, a pitch representation based on interspike-interval distributions roughly predicts the upper limit for the pitch of the missing fundamental in humans. However, the predicted strength of interval-based pitch estimates was highest for F0s below 400-500 Hz (where harmonics are poorly resolved, based on rate-results), and decreased monotonically with F0s (as the degree of harmonic resolvability increases). This observation led us to conclude that a temporal representation cannot account for the greater salience of the pitch based upon resolved harmonics compared to that based on entirely unresolved harmonics (Bernstein and Oxenham 2003b; Houtsma and Smurzynski 1990). The interspike-interval representation was consistent with phase invariance for the pitch of stimuli containing resolved harmonics, but could not completely_account for the perceptual dependence of

pitch on the relative phase among the harmonics when these are all unresolved (Lundeen and Small 1984; Shackleton and Carlyon 1994).

Given the limitations of both a strictly place and a strictly temporal representations, in Chapter 2 we tested an alternative spatio-temporal representation of the pitch of harmonic complex tones, where cues to the frequencies of resolved harmonics arise from the spatial pattern in the phase of the phase-locked response of AN fibers (Shamma 1985). This spatio-temporal representation was investigated in single AN fibers and the results were first analyzed as a function of CF, and then re-interpreted as a function of F0 by applying the scaling invariance principle of cochlear mechanics (Zweig 1976). We found that the spatio-temporal representation is viable in the cat for F0s between 300 Hz and 1 kHz.

We attributed the failure of the spatio-temporal representation below 300 Hz to poor harmonic resolvability, and this lower limit was in broad agreement with that found for the rate-place representation in Chapter 1. Although harmonic resolvability, as measured using a strictly "spectral" definition like the one we adopted in Chapter 1, increases with F0, the effectiveness of the spatio-temporal representation decreased very rapidly with F0 above 1 kHz, a finding which we attributed to the rapid degradation of phase-locking to the harmonic frequencies. The existence of an upper limit for the spatio-temporal representation of pitch is thus in broad agreement with the upper limit for the pitch of missing-F0 complex tones in humans. For F0s between 300 Hz and 1 kHz, a range in which the spatio-temporal and the rate-place representation are both effective, we observed a degradation of both representations with stimulus level. However, spatio-temporal pitch cues remained up to 2-3 times stronger than rate-place cues at very high levels, suggesting that the spatio-temporal representation may be more consistent with psychophysical data than the rate-place representation. Finally, the spatio-temporal representation was also consistent with human psychophysical data on phase-independence for pitch based on resolved harmonics.

In Table 2, we present a summary of the different ranges of viability for the three pitch representations examined in this thesis, and of their predicted adequacy in accounting for psychophysical data. Also presented in Table 2 is a speculation on where the F0-limits observed for each representation in the cat would translate to in humans, based on the assumptions that 1) cochlear filters are three times sharper in humans than in cats (Shera et

al. 2002) and that 2) the degree of neural phase-locking and its upper limit are not significantly different in the two species.

In Chapter 3, we studied peripheral frequency selectivity by applying the notched-noise method, widely used in psychophysics, to single AN fibers. We found that, at low stimulus levels, this method yields accurate predictions of single AN fibers' "masked" thresholds for pure tones in band-reject ("notched") noises, across the entire range of CFs. We observed a similar increase in tuning sharpness (expressed as a Q-factor, the ratio CF/Bandwidth) with CF for both pure-tone tuning curves and neural auditory filters estimated with the notched-noise method. This increase was well-fit by a power function of CF with exponent equal to 0.49 ± 0.07, a value slightly higher, but not significantly different from those derived in other studies using different experiment paradigms [0.37 ± 0.10, observed by Shera et al. (2002) for the CF dependence of Q_{10} in pure-tone tuning curves from AN fibers in cats; 0.37 ± 0.09, found in Chapter 1 of the present study for the increase in N_{max} with CF]. To quantify the effect of suppression on estimated tuning sharpness, we compared the equivalent rectangular bandwidths (ERBs) of neural auditory filters derived using a simultaneous masking and a non-simultaneous masking paradigm. Although in the majority of the cases suppression contributes to an overestimation of the ERBs, in a few cases this trend is reversed, a finding for which we have no plausible explanation. We also fit pseudo-psychophysical auditory filters to masked thresholds predicted from a model population of neural auditory filters. The results showed a close agreement between the predicted sharpness of psychophysical tuning and the frequency selectivity measured physiologically, thus supporting the use of the notched-noise method in human psychophysics.

	F0 range (Hz) for cat	Predicted F0 range (Hz) for humans	Depends on resolved harmonics	Predicts the upper limit for pitch of the missing-F0	Robust at high levels	Phase-invariant for resolved harmonics
Rate-Place	≥ 450	≥ 150	yes	no	no	yes
Interspike-Interval	≤ 1300	≤ 1300	no	yes	yes	yes
Spatio-Temporal	300 – 1k	100 – 1k	yes	yes	partly	yes

Table 2: F0-ranges of viability for the rate-place, the interspike-interval and the spatio-temporal representations, and their predicted adequacy in accounting for psychophysical data.

Chapter 5. Implications and future directions

The results of our experiments indicate that three types of cues to the pitch of harmonic complex tones are available in the auditory nerve: "rate-place" cues, reflecting harmonicity of the frequency spectrum, "temporal" cues, reflecting periodicities in the time waveform and "spatio-temporal" cues, reflecting modulations in temporal information with cochlear place. A fundamental question is whether the formation of a pitch percept is based upon the exclusive use of any one of these three possible cues, or whether different mechanisms are employed to different degrees.

Resolved harmonics are almost always available to normal hearing listeners in everyday acoustic environments, and pitch based on resolved harmonics is far more salient than the pitch based on exclusively unresolved harmonics (Houtsma and Smurzynski 1990; Bernstein and Oxenham 2003b). Recent psychophysical data (Bernstein and Oxenham 2006) shows a correlation between the degraded cochlear frequency selectivity of hearing impaired listeners with sensorineural hearing loss and their deficit in F0-discrimination performance. Abnormally high F0 difference limens are also exhibited by cochlear implant users, and their performance becomes very poor above 300 Hz (Zeng 2002; Shannon 1983), presumably due to the insufficient "place" information available from electrical stimulation. These observations strongly suggest that a pitch code which privileges resolved harmonics over unresolved harmonics (i.e. rate-place or spatio-temporal) is most likely to be used in normal listening conditions. While a strictly rate-place representation degrades rapidly with level and cannot account for the existence of an upper limit to the human perception of the pitch of the missing-F0 (Chapter 1), a spatio-temporal representation may overcome this limitation (Chapter 2), thereby becoming the strongest candidate to explain several major human psychophysical phenomena.

On the other hand, a weaker pitch can be heard by hearing-impaired and normal-hearing listeners also in the presence of only unresolved harmonics, a finding that cannot be accounted for by any pitch representation dependent upon harmonic resolvability and that therefore is in favor of the hypothesis that a purely temporal pitch representation must indeed be used when place or (spatio-temporal) cues are unavailable. A purely temporal representation is also thought to be the most likely explanation for the weak pitch percepts

of cochlear implant users, because of the poor spatial resolution of the electrical stimulation. In contrast, neural responses to electric stimulation with sinusoids and pulse trains show a high degree of phase-locking, compared to that observed in response to acoustic stimuli (Dynes and Delgutte 1992; Javel and Shepherd 2000). However, despite the large amount of available temporal information, pitch is highly ambiguous and is heard only at very low F0s, consistent with only a small contribution of a temporal mechanism to pitch perception in normal conditions.

Another important issue is the fate undergone at more central stages of the auditory pathway by the cues provided at the level of the AN by the different pitch representations examined. Spatio-temporal pitch cues may be converted into place cues at some more central stage along the auditory pathway. The conversion of spatio-temporal cues into place cues requires a neural mechanism 1) receiving inputs originating from neighboring cochlear locations and 2) sensitive to differences in the relative timings of its inputs. Although there is some evidence in the AVCN for neurons that may satisfy these requirements (Carney 1990), more data are needed to shed light on the exact neural mechanisms (coincidence detection, lateral inhibition or a combination of the two are proposed explanations) performing this operation. Given the improvement in neural synchronization exhibited by some AVCN cells (Joris et al. 1994), which might actually enhance spatio-temporal cues to pitch, another possibility is that the extraction of these cues takes place at a more central auditory center.

The generation of a pitch percept from a spatial pattern of activation, whether originally generated by a purely rate-place code or produced by a transformation of spatio-temporal cues into place cues, is generally assumed to be based on a harmonic template matching mechanism. While there is strong evidence for a robust tonotopic mapping of the frequency spectrum of sounds at virtually all levels of the auditory pathway (including the auditory cortex) (Popper and Fay 1992), the process underlying the formation of preferential links among harmonically-related channels is a matter of ongoing debate. A classical hypothesis is that harmonic templates arise as a consequence of the frequent exposure to speech and natural harmonic sounds in early development (Terhardt 1974). However, young infants (7-8 months old) not only can perceive the pitch of the missing-F0, but also perform better in the presence of low-order, presumably resolved harmonics (Clarkson and Rogers 1995),

suggesting that the formation of the harmonic templates may occur before a sufficient exposure to sounds characterized by a rich harmonic structure. An appealing alternative theory has been proposed (Shamma and Klein 2000) according to which harmonic templates arise from the exposure to any kind of broadband stimulation and not necessarily to harmonic sounds. According to this model, the half-wave rectification performed by the hair cells results in a high likelihood for highly-correlated spike trains at harmonically-related CFs, which are in turn reflected in the accumulation of coincidences over time in these pairs of channels. The accumulation of coincidences between the firings in cochlear channels that are harmonically-related can conceivably explain the emergence, over time, of harmonic templates for all F0s whose resolved harmonics are in the frequency range of phase-locking.

The nature of this hypothetical mechanism underlying the formation of harmonic templates is related, but not identical to the cues to resolved harmonics generated by the variations in phase of the basilar membrane motion with cochlear place discussed in Chapter 2. Both mechanisms are "spatio-temporal" in that they depend upon a combination of place information and neural phase-locking. Both are also based on across-CF interactions, but while the harmonic template generation process relies on correlations between channels with wide separations, extraction of spatio-temporal cues to resolved harmonics only requires comparisons between spike trains in neighboring channels. Cross-frequency coincidence detection is the most likely operation at the origin of the emergence of the harmonic templates, while on the other hand both coincidence detection and lateral inhibition (the latter effectively implemented in Chapter 2) could conceivably underlie the extraction of spatio-temporal cues to resolved harmonics.

Although evidence for precise timing information (of the order of microseconds) like the one needed for the preservation of an interspike-interval or a spatio-temporal representation is weaker and weaker as one ascends the auditory pathway, phase-locking up to 1 kHz has occasionally been observed in the medial geniculate body of the thalamus (Rouiller et al. 1979) and in the cortex (de Ribaupierre et al. 1972). The existence of a central pitch code entirely or partially based on temporal information cannot therefore be completely ruled out. It has also been hypothesized that an interval code in the AN and CN could be converted into some other types of temporal code, possibly on different time

scales or even asynchronous (Cariani and Delgutte 1996b) or distributing the large amount of temporal information progressively more sparsely over a greater number of neurons (Cariani 1999). A more realistic hypothesis is that even a purely temporal peripheral code based on interspike intervals may somewhere be converted into a place code (Licklider 1951). A physiologically-realistic model has recently been proposed (Wiegrebe and Meddis 2004), according to which a temporal representation of pitch in the AN may be converted into a place representation by a population of units with sustained-chopper responses in the VCN, albeit only for low F0s (below 500 Hz).

Given the important role played by pitch in source segregation in speech and auditory scene analysis, the extent to which the proposed pitch mechanisms can represent more than one complex tone is of great interest. If based on a place code, the identification of two simultaneous pitches would require an additional amount of frequency resolution, while if based on a temporal code, segregation of two pitches would rely on simultaneous cues to different periodicities. Recent data from cat AN (Larsen et al. 2005) indicate that the rate-place and the interval representation may be complementary in their ability to represent pairs of complex tones with a difference in F0 as small as 7%: the interval representation is more effective at lower F0s and the rate-place representation at higher F0s. This performance is more than sufficient to account for the lower psychophysical limit for simultaneous vowels of a 15-25% F0-separation (Assmann and Paschall 1998), below which human listeners cannot correctly identify both pitches simultaneously. Unfortunately, data are not yet available for the spatio-temporal representation, where this task could be problematic given the additional constraints imposed by the presence of two complex tones: frequency resolution of two sets of harmonics and competition between phase-locking to different, non-harmonically related frequencies. A systematic study of the human perception of simultaneous complex tones would also be of high significance, since sounds present in natural environments are very rarely encountered in total isolation.

Finally, a spatio-temporal mechanism, which we propose to be the basis for the perception of pitch, could also in principle underlie the detection of pure tones (in the frequency range of phase-locking) in the presence of masking sounds, and has also been proposed as the basis for level discrimination (Heinz et al. 2001). This idea might have important implications for current methods used in psychophysics to estimate peripheral

frequency selectivity, which are all based on the phenomenon of masking, commonly thought of as a mechanism by which the channels tuned to the signal are either "swamped" (excitatory masking) or "suppressed" (suppressive masking) by the masker. Alternative methods, quantifying masking as the degree to which a masker can disrupt the spatio-temporal response pattern in response to a signal, might yield different results and/or improve the accuracy in estimating frequency selectivity of methods currently used.

In summary, we investigated possible neural representation of the pitch of complex tones and we proposed a spatio-temporal representation, which combines the advantages and overcomes some of the limitations of strictly place and temporal codes, as the most likely to be used in normal listening conditions. To our knowledge, this was the first study in which a possible implementation of a spatio-temporal representation has been directly tested physiologically. Much work remains to be done to better understand the central mechanisms of extraction of peripheral cues to pitch and the degree to which these mechanisms might be similar across different species.

REFERENCES

Anderson DJ, Rose JE, Hind JE, and Brugge JF. Temporal position of discharges in single auditory nerve fibers within the cycle of a sine-wave stimulus: frequency and intensity effects. *J Acoust Soc Am* 49: Suppl 2:1131+, 1971.

ANSI. American national psychoacoustical terminology. S3.20. New York. 1973.

Assmann PF and Paschall DD. Pitches of concurrent vowels. *J Acoust Soc Am* 103: 1150-1160, 1998.

Bernstein JG and Oxenham AJ. Effects of relative frequency, absolute frequency, and phase on fundamental frequency discrimination: Data and an autocorrelation model. *J Acoust Soc Am* 113: 2290, 2003a.

Bernstein JG and Oxenham AJ. Pitch discrimination of diotic and dichotic tone complexes: harmonic resolvability or harmonic number? *J Acoust Soc Am* 113: 3323-3334, 2003b.

Bernstein JGW. Pitch perception and harmonic resolvability in normal-hearing and hearing-impaired listeners. Ph.D. Thesis, MIT, 2006.

Bregman AS. *Auditory Scene Analysis: The Perceptual Organization of Sound.* Cambridge, MA: MIT Press, 1990.

Brown CH, Beecher MD, Moody DB, and Stebbins WC. Localization of primate calls by old world monkeys. *Science* 201: 753-754, 1978.

Buus S, Schorer E, Florentine M, and Zwicker E. Decision rules in detection of simple and complex tones. *J Acoust Soc Am* 80: 1646-1657, 1986.

Capranica RR and Moffat AJM. Selectivity of the peripheral auditory system of spadefoot toads (Scaphiopus couchi) for sounds of biological significance. *J Comp Physiol* 100: 231-249, 1975.

Cariani P. Temporal coding of periodicity pitch in the auditory system: an overview. *Neural Plast* 6: 147-172, 1999.

Cariani PA and Delgutte B. Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. *J Neurophysiol* 76: 1698-1716, 1996a.

Cariani PA and Delgutte B. Neural correlates of the pitch of complex tones. II. Pitch shift, pitch ambiguity, phase invariance, pitch circularity, rate pitch, and the dominance region for pitch. *J Neurophysiol* 76: 1717-1734, 1996b.

Carlyon RP. Comments on "A unitary model of pitch perception" [J. Acoust. Soc. Am. 102, 1811-1820 (1997)]. *J Acoust Soc Am* 104: 1118-1121, 1998.

Carlyon RP and Shackleton TM. Comparing the fundamental frequencies of resolved and unresolved harmonics: Evidence for two pitch mechanisms? *J Acoust Soc Am* 95: 3541-3554, 1994.

Carney LH and Yin TCT. Temporal coding of resonances by low-frequency auditory nerve fibers: single-fiber responses and a population model. *J Neurophysiol* 60: 1653-1677, 1988.

Carney LH. Sensitivities of cells in the anteroventral cochlear nucleus of cat to spatiotemporal discharge patterns across primary afferents. J Neurophysiol 64:437-456, 1990)

Carney L. Spatiotemporal encoding of sound level: Models for normal encoding and recruitment of loudness. *Hear Res* 76: 31-44, 1994.

Cedolin, L and Delgutte B. Dual representation of the pitch of complex tones in the auditory nerve. Abstr. Assoc. Res. Otolaryngol. 26, 2003.

Cedolin L and Delgutte B. Neural representations of pitch. From Sound to Sense: 50+ Years of Discoveries in Speech Communication, MIT, 2004.

Cedolin L and Delgutte B. Representations of the pitch of complex tones in the auditory nerve. In: *Auditory signal processing: Physiology, psychoacoustics, and models*, edited by Pressnitzer D, deCheveigne A, McAdams S and Collet L. New York: Springer, 2005a.

Cedolin L and Delgutte B. Spatio-temporal representation of the pitch of complex tones in the auditory nerve. Abstr. Assoc. Res. Otolaryngol. 28, 2005b.

Cedolin L and Delgutte B. Pitch of Complex Tones: Rate-Place and Interspike Interval Representations in the Auditory Nerve. *J Neurophysiol* 94: 347–362, 2005c.

Clarkson MG and Rogers EC. Infants require low-frequency energy to hear the pitch of the missing fundamental. *J Acoust Soc Am* 98: 148-154, 1995.

Cohen MA, Grossberg S, and Wyse LL. A spectral network model of pitch perception. *J Acoust Soc Am* 98: 862-879, 1994.

Colburn HS, Carney LH, and Heinz MG. Quantifying the information in auditory-nerve responses for level discrimination. *J Assoc Res Otolaryngol* 4: 294-311, 2003.

Conley RA and Keilson SE. Rate representation and discriminability of second formant frequencies for /ɛ/-like steady-state vowels in cat auditory nerve. *J Acoust Soc Am* 98: 3223-3234, 1995.

Cooper NP and Rhode WS. Mechanical responses to two-tone distortion products in the apical and basal turns of the mammalian cochlea. *J Neurophysiol* 78: 261-270, 1997.

Costalupes JA, Young ED, and Gibson DJ. Effects of continuous noise backgrounds on rate response of auditory nerve fibers in cat. *J Neurophysiol* 51: 1326-1344, 1984.

Cynx J and Shapiro M. Perception of missing fundamental by a species of songbird (Sturnus vulgaris). *J Comp Psychol* 100: 356-360, 1986.

Darwin CJ and Carlyon RP. Auditory grouping. In: *The handbook of perception and cognition, Volume 6, Hearing*, edited by Moore BCJ. London: Academic, 1995.

Darwin CJ and Hukin RW. "Effectiveness of spatial cues, prosody, and talker characteristics in selective attention," *J. Acoust. Soc. Am.* 107, 970-977, 2000.

de Boer E and Nuttal AL. The mechanical waveform of the basilar membrane. III. Intensity effects. *J Acoust Soc Am* 107: 1497-1507, 2000.

de Cheveigné A. Cancellation model of pitch perception. *J Acoust Soc Am* 103: 1261-1271, 1998.

Dear SP, Fritz J, Haresign T, Ferragamo M, and Simmons JA. Tonotopic and functional organization in the auditory cortex of the big brown bat. *J Neurophysiol* 70: 1988-2009, 1993.

Delgutte B. Some correlates of phonetic distinctions at the level of the auditory nerve. In: *The Representation of Speech in the Peripheral Auditory System*, edited by Granström RCaB. Amsterdam: Elsevier, p. 131-150, 1982.

Delgutte B. Peripheral auditory processing of speech information: Implications from a physiological study of intensity discrimination. In: *The Psychophysics of Speech Perception*, edited by Schouten M. Nijhof: Dordrecht, 1987, p. 333-353.

Delgutte B. Physiological mechanisms of masking. In: *Basic Issues in Hearing*, edited by Duifhuis H, Horst JW and Wit HP. London: Academic Press, p. 204-214, 1988.

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

Delgutte B. Two-tone rate suppression in auditory-nerve fibers: Dependence on suppressor frequency and level. *Hearing Research* 49: 225-246, 1990b.

Duifhuis H, Willems LF, and Sluyter RJ. Measurement of pitch in speech: An implementation of Goldstein's theory of pitch perception. *J Acoust Soc Am* 71: 1568-1580, 1982.

de Ribaupierre F, Goldstein Jr. MH, and Yeni-Komshian G. Cortical coding of repetitive acoustic pulses. *Brain Research* 48: 205-225, 1972.

Dynes SB and Delgutte B. Phase locking of auditory-nerve discharges to sinusoidal electric stimulation of the cochlea. *Hear Res* 58: 79-90, 1992.

Efron B and Tibshirani RJ. *An introduction to the Bootstrap*. New York: Chapman & Hall, 1993.

Evans EF, Pratt SR, Spenner H, and Cooper NP. Comparisons of Physiological and Behavioural Properties: Auditory Frequency Selectivity. In: *Auditory Physiology and Perception: Proceedings of the 9th International Symposium on Hearing*, edited by Cazals Y, Horner K and Demany L, Carcens, France. Pergamon Press, p. 159-169, 1991.

Evans E, Wilson JP. The frequency selectivity of the cochlea. In: *Basic Mechanisms in Hearing*, edited by Møller A. London: Academic, 519-554, 1973.

Fletcher H. Auditory patterns. *Reviews of Modern Physics* 12: 47-65, 1940.

Glasberg BR and Moore BCJ. Derivation of auditory filter shapes from notched-noise data. *Hear Res* 47: 103-138, 1990.

Glattke TJ and Small AM, Jr. Frequency selectivity of the ear in forward masking. *J Acoust Soc Am* 42: 154-157, 1967.

Goldstein JL. An optimum processor theory for the central formation of the pitch of complex tones. *J Acoust Soc Am* 54: 1496-1516, 1973.

Green D. Detection of multiple component signals in noise. *J Acoust Soc Am* 30: 904-911, 1958.

Greenberg S, Geisler CD, and Deng L. Frequency selectivity of single cochlear-nerve fibers based on the temporal response pattern to two-tone signals. *J Acoust Soc Am* 79(4): 1010-1019, 1986.

Hall JL. Maximum-likelihood sequential procedure for estimation of psychometric functions. J. Acoust. Soc. Am. 44, 370, 1968.

Harris, DM. Action potential suppression, tuning curves and thresholds: Comparison with single fiber data. Hear. Res. 1, 133-154, 1979.

Heffner H and Whitfield IC. Perception of the missing fundamental by cats. *J Acoust Soc Am* 59: 915-919, 1976.

Heinz MG, Colburn HS, and Carney LH. Rate and timing cues associated with the cochlear amplifier: level discrimination based on monaural cross-frequency coincidence detection. *J Acoust Soc Am* 110: 2065-2084, 2001.

Hirahara T, Cariani PA, and Delgutte B. Representation of low-frequency vowel formants in the auditory nerve. In: *Proc. Workshop on Auditory Basis of Speech Perception,*, edited by Ainsworth W. Keele, UK: ESCA, 1996, p. 83-86.

Horst JW, Javel E, and Farley GR. Coding of spectral fine structure in the auditory nerve. II. Level-dependent nonlinear responses. *J Acoust Soc Am* 88: 2656-2681, 1990.

Houtgast T. Psychophysical experiments on grating acuity. *Symposium on Hearing Theory*, IPO, Eindhoven, 1972.

Houtgast T. Psychophysical evidence for lateral inhibition in hearing. *J Acoust Soc Am* 51: 1885-1894, 1972b.

Houtsma AJM and Smurzynski J. Pitch identification and discrimination for complex tones with many harmonics. *J Acoust SocAm* 87: 304-310, 1990.

Kiang NYS, Watanabe T, Thomas EC, and Clark LF. *Discharge Patterns of Single FIbers in the Cat's Auditory Nerve*. Cambridge, MA: The MIT Press, 1965.

Javel E and Shepherd RK. Electrical stimulation of the auditory nerve. III. Response initiation sites and temporal fine structure. *Hear Res* 140: 45-76, 2000.

Johnson DH. The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones. *J Acoust Soc Am* 68: 1115-1122, 1980.

Joris PX, Carney LH, Smith PH, and Yin TCT. Enhancement of neural synchronization in the anteroventral cochlear nucleus. I. Responses to tones at the characteristic frequency. *J Neurophysiol* 71: 1022-1036, 1994.

Kanwal JS, Matsumura S, Ohlemiller K, and Suga N. Analysis of acoustic elements and syntax in communication sounds emitted by mustached bats. *J Acoust Soc Am* 96: 1229-1254, 1994.

Kiang NYS, Watanabe T, Thomas EC, and Clark LF. *Discharge Patterns of Single FIbers in the Cat's Auditory Nerve*. Cambridge, MA: The MIT Press, 1965.

Kiang NYS, Moxon EC, and Levine RA. Auditory-nerve activity in cats with normal and abnormal cochleas. In: *Sensorineural hearing loss. Ciba Found Symp*: 241-273, 1970.

Lai YC, Winslow RL, and Sachs MB. A model of selective processing of auditory-nerve inputs by stellate cells of the antero-ventral cochlear nucleus. *J Comput Neurosci* 1: 167-194, 1994.

Larsen E, Cedolin L, and Delgutte B. Coding of pitch in the auditory nerve: Two simultaneous complex tones. *Abstr. Assoc. Res. Otolaryngol.* 28, 2005.

Liberman MC. Auditory-nerve responses from cats raised in a low-noise chamber. *J Acoust Soc Am* 63: 442-455, 1978.

Licklider JCR. A duplex theory of pitch perception. *Experientia* 7: 128-134, 1951.

Licklider JCR. 'Periodicity' pitch and 'place' pitch. *J. Acoust. Soc. Am.* 26, 945, 1954.

Loeb GE, White MW, and Merzenich MM. Spatial cross-correlation. A proposed mechanism for acoustic pitch perception. *Biol Cybern* 47: 149-163, 1983.

Louage DH, van der Heijden M, Joris PX. Temporal properties of responses to broadband noise in the auditory nerve. *J Neurophysiol* 91: 2051-2065, 2004.

Lundeen C and Small AMJ. The influence of temporal cues on the strength of periodicity pitches. *J Acoust Soc Am* 75: 1578-1587, 1984.

May BJ, Huang AY, Le Prell GS, and Hienz RD. Vowel formant frequency discrimination in cats: Comparison of auditory nerve representations and psychophysical thresholds. *Aud Neurosci* 3: 135–162, 1996.

McKinney MF and Delgutte B. A possible neurophysiological basis of the octave enlargement effect. *J Acoust Soc Am* 73: 1694-1700, 1999.

Meddis R and Hewitt MJ. Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I. Pitch identification. *J Acoust Soc Am* 89: 2866-2882, 1991.

Meddis R and O'Mard L. A unitary model of pitch perception. *J Acoust Soc Am* 102: 1811-1820, 1997.

Montgomery C and Clarkson M. Infants' pitch perception: masking by low- and high-frequency noises. *J Acoust Soc Am* 102: 3665-3672, 1997.

Moore BCJ. Frequency difference limens for short-duration tones. *J Acoust Soc Am* 54: 610-619, 1973a.

Moore BCJ. Some experiments relating to the perception of complex tones. *Q J Exp Psychol* 25: 451-475, 1973b.

Moore BCJ. Psychophysical tuning curves measured in simultaneous and forward masking. *J Acoust Soc Am* 63: 524-532, 1978.

Moore BCJ and Glasberg BR. Formulae describing frequency selectivity as a function of frequency and level and their use in calculating excitation patterns. *Hear Res* 28: 209-225, 1987.

Moore BCJ. *Introduction to the Psychology of Hearing.* London: Academic, 1990.

Moore BCJ and Peters RW. "Pitch discrimination and phase sensitivity in young and elderly subjects and its relationship to frequency selectivity," *J. Acoust. Soc. Am.* 91, 2881-2893, 1992.

Moore BCJ, Glasberg BR and Baer T. A model for the prediction of thresholds, loudness and partial loudness. J. Audio Eng. Soc. 45, 224–240, 1997.

Nicastro N and Owren MJ. Classification of domestic cat (Felis catus) vocalizations by naive and experienced human listeners. *J Comp Psychol* 117: 44-52, 2003.

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

Ohm G. Über die Definition des Tones nebst daran geknüpfte Theorie der Sirene und ähnlicher tonbildender Vorrichtungen. *Ann Phys Chem* 59: 513-565, 1843.

Oxenham AJ and Shera CA. Estimates of human cochlear tuning at low levels using forward and simultaneous masking. J Assoc Res Otolaryngol. 4:541-54, 2003.

Palmer AR. The representation of the spectra and fundamental frequencies of steady-state single- and double-vowel sounds in the temporal discharge patterns of guinea pig cochlear-nerve fibers. *J Acoust Soc Am* 88: 1412-1426, 1990.

Palmer AR and Russell IJ. Phase-locking in the cochlear nerve of the guinea pig and its relation to the receptor potential of inner hair cells. *Hearing Res* 24: 1-15, 1986.

Palmer AR and Winter IM. Cochlear nerve and cochlear nucleus responses to the fundamental frequency of voiced speech sounds and harmonic complex tones. In: *Auditory Physiology and Perception*, edited by Horner K. Oxford: Pergamon, 1992, p. 231-240.

Palmer AR and Winter IM. Coding of the fundamental frequency of voiced speech sounds and harmonic complex tones in the ventral cochlear nucleus. In: *Mammalian Cochlear Nuclei: Organization and Function*, edited by Mugnaini E. New York: Plenum, 1993, p. 373-384.

Patterson R. Auditory filter shapes derived with noise stimuli. *J Acoust Soc Am* 59: 640-654, 1976.

Patterson RD, Nimmo-Smith I, Weber DL and Milroy R. The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold. J. Acoust. Soc. Am. 72, 1788–1803, 1982.

Patterson R and Holdsworth J. A functional model of neural activity patterns and auditory images. In: *Advances in Speech, Hearing and Language Processing*, edited by Ainsworth W. London: JAI, 1994.

Pfeiffer RR and Kim DO. Cochlear nerve fiber responses: Distribution along the cochlear partition. *J Acoust Soc Am* 58: 867-965, 1975.

Plomp R. The ear as a frequency analyzer. *J Acoust Soc Am* 36: 1628-1636, 1964.

Plomp R. Pitch of complex tones. *J Acoust Soc Am* 41: 1526-1533, 1967.

Preisler A and Schmidt S. Spontaneous classification of complex tones at high and ultrasonic frequencies in the bat, Megaderma lyra. *J Acoust Soc Am* 103: 2595-2607, 1998.

Pressnitzer D, Patterson RD, and Krumbholz K. The lower limit of melodic pitch. *J Acoust Soc Am* 109: 2074-2084, 2001.

Rhode WS. Interspike intervals as a correlate of periodicity pitch in cat cochlear nucleus. *J Acoust Soc Am* 97: 2413-2429, 1995.

Ritsma RJ. Frequencies dominant in the perception of the pitch of complex sounds. *J Acoust Soc Am* 42: 191-198, 1967.

Ritsma RJ and Engel FL. Pitch of frequency-modulated signals. *J Acoust Soc Am* 36: 1637-1644, 1964.

Rose JE, Brugge JR, Anderson DJ, and Hind JE. Phase-locked response to low-frequency tones in single auditory nerve fibers of the squirrel monkey. *J Neurophysiol* 30: 769-793, 1967.

Rose JE, Hind JE, Anderson DJ, and Brugge JF. Some effects of stimulus intensity on response of auditory nerve fibers in the squirrel monkey. *J Neurophysiol* 34: 685-699, 1971.

Rosen S and Baker RJ. Characterising auditory filter nonlinearity Hearing Res. 73, 231–243, 1994.

Rosen S, Baker RJ, and Darling A. Auditory filter nonlinearity at 2 kHz in normal hearing listeners. *J Acoust Soc Am* 103: 2359-2550, 1998.

Rouiller E, de Ribaupierre Y and de Ribaupierre F. Phase-locked responses to low frequency tones in the medial geniculate body. *Hear Res* 1: 213-226, 1979.

Ruggero M. Physiology and coding of sound in the auditory nerve. In: *The Mammalian Auditory Pathway: Neurophysiology.*, edited by AN Popper RF. New-York: Springer-Verlag, 1992, p. 34-93.

Ruggero MA, Rich NC, Recio A, Narayan SS, and Robles L. Basilar-membrane responses to tones at the base of the chinchilla cochlea. *J Acoust Soc Am* 101: 2151-2163, 1997.

Ruggero M and Temchin AN. Unexceptional sharpness of frequency tuning in the human cochlea. *Proc Natl Acad Sci USA* 102: 18614-18619, 2005.

Sachs MB and Kiang NYS. Two-tone inhibition in auditory-nerve fibers. *J Acoust Soc Am* 43: 1120-1128, 1968.

Sachs MB and Abbas PJ. Rate versus level functions for auditory-nerve fibers in cats: tone-burst stimuli. *J Acoust Soc Am* 56 No.6: 1835-1847, 1974.

Sachs MB and Young ED. Encoding of steady-state vowels in the auditory nerve: Representation in terms of discharge rate. *J Acoust Soc Am* 66: 470-479, 1979.

Schmiedt RA. Boundaries of two-tone rate suppression of cochlear-nerve activity. *Hearing Res* 7: 335-351, 1982.

Schouten, JF. The residue and the mechanism of hearing. *Proc. Kon. Akad. Wetenschap.* 43, 991-999, 1940.

Schroeder MR. Synthesis of low peak-factor signals and binary sequences with low autocorrelation. *IEEE Trans Inf Theory* 16: 85-89, 1970.

Schwartz DA and Purves D. Pitch is determined by naturally occurring periodic sounds. *Hear Res* 194: 31-46, 2004.

Seebeck A. Beobachtungen über einige Bedingungen der Entstehung von Tönen. *Ann Phys Chem* 53: 417-436, 1841.

Semal C and Demany L. The upper limit of "musical" pitch. *Music Perception* 8: 165-175, 1990.

Shackleton TM and Carlyon RP. The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination. *J Acoust Soc Am* 95: 3529-3540, 1994.

Shamma SA. Speech processing in the auditory system. I: The representation of speech sounds in the responses of the auditory nerve. *J Acoust Soc Am* 78: 1612-1621, 1985a.

Shamma S. Speech processing in the auditory system. II: Lateral inhibition and the central processing of speech evoked activity in the auditory nerve. *J Acoust Soc Am* 78: 1622-1632, 1985b.

Shamma S and Klein D. The case of the missing pitch templates: how harmonic templates emerge in the early auditory system. *J Acoust Soc Am* 107: 2631-2644, 2000.

Shannon RV. Multichannel electrical stimulation of the auditory nerve in man. I. Basic psychophysics. *Hearing Res* 11: 157-189, 1983.

Shera CA. Intensity-invariance of fine time structure in basilar-membrane click responses: Implications for cochlear mechanics. *J Acoust Soc Am* 110: 332-348, 2001.

Shera CA, Guinan JJ, Jr., and Oxenham AJ. Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements. *Proc Natl Acad Sci USA* 99: 3318-3323, 2002.

Shipley C, Carterette EC, and Buchwald JS. The effect of articulation on the acoustical structure of feline vocalizations. *J Acoust Soc Am* 89: 902-909, 1991.

Shofner WP. Temporal representation of rippled noise in the anteroventral cochlear nucleus of the chinchilla. *J Acoust Soc Am* 90: 2450-2466, 1991.

Popper AN, Fay RR. The mammalian auditory pathway: Neurophysiology. In: *The Springer Handbook of Auditory Research*. Vol. 2. New York: Springer-Verlag, 1992.

Siebert WM. Stimulus transformations in the peripheral auditory system. In: *Recognizing Patterns*, edited by Kollers PA and Eden M. Cambridge: MIT Press, 1968, p. 104-133.

Simmons JA, Moss CF and Ferragamo M. Convergence of temporal and spectral information into acoustic images of complex sonar targets perceived by the echolocating bat, Eptesicus fuscus. *J Comp Physiol [A]* 166: 449-470, 1990.

Smoorenburg GF and Linschoten DH. A neurophysiological study on auditory frequency analysis of complex tones. In: *Psychophysics and Physiology of Hearing*, edited by Wilson JP. London: Academic, 1977, p. 175-184.

Spiegel MF. Thresholds for tones in maskers of various bandwidths and for signals of various bandwidths as a function of signal frequency. *J Acoust Soc Am* 69: 791-795, 1981.

Srulovicz, P and Goldstein, JL. A central spectrum model: a synthesis of auditorynerve timing and place cues in monaural communication of frequency spectrum. *J Acoust Soc Am* 73, 1266-1276, 1983.

Suta D, Kvasnak E, Popelar J, and Syka J. Representation of species-specific vocalizations in the inferior colliculus of the guinea pig. *J Neurophysiol* 90: 3794-3808, 2003.

Taylor MM and Creelman CD. PEST: Efficient estimates on probability functions. *J Acoust Soc Am* 41: 782-787, 1967.

Terhardt E. Pitch, consonance, and harmony. *J Acoust Soc Am* 55: 1061-1069, 1974.

Tomlinson RWW and Schwarz DWF. Perception of the missing fundamental in nonhuman primates. *J Acoust Soc Am* 84: 560-565, 1988.

van der Heijden M and Joris PX. Cochlear phase and amplitude retrieved from the auditory nerve at arbitrary frequencies. *J Neurosci* 23: 9194-9198, 2003.

Viemeister NF. Psychophysical aspects of auditory coding. In: *Auditory Function. Neurobiological Bases of Hearing*, edited by Edelman GM, Gall WE and Cowan WM. New York: John Wiley & Sons, 1988, p. 213-241.

Ward WD. Subjective Musical Pitch. *J Acoust Soc Am* 26: 369-380, 1954.

Wightman FL. The pattern-transformation model of pitch. *J Acoust Soc Am* 54: 407-416, 1973.

Wilson J and Evans E. Grating acuity of the ear: psychophysical and neurophysiological measures of frequency resolving power. *7th Int. Congr. on Acoustics*, Budapest, 1971, p. 397-400.

Winslow RL and Sachs MB. Single-tone intensity discrimination based on auditory-nerve rate responses in backgrounds of quiet, noise, and with stimulation of the crossed olivocochlear bundle. *Hear Res* 35: 165-190, 1988.

Winter IM and Palmer AR. Intensity coding in low-frequency auditory-nerve fibers of the guinea pig. *J Acoust Soc Am* 90: 1958-1967, 1991.

Yost WA. The dominance region and ripple noise pitch: a test of the peripheral weighting model. *J Acoust Soc Am* 72: 416-425, 1982.

Yost WA. Pitch of iterated rippled noise. *J Acoust Soc Am* 100: 511-518, 1996.

Zeng FG. Temporal pitch in electric hearing. *Hear Res* 174: 101-106, 2002.

Zhang X, Heinz MG, Bruce IC, and Carney LH. A phenomenological model for the responses of auditory-nerve fibers: I. Nonlinear tuning with compression and suppression. *J Acoust Soc Am* 109: 648-670, 2001.

Zweig G. Basilar membrane motion. *Cold Spring Harbor Symp Quant Biol* 40: 619-633, 1976.