Pitch perception and harmonic resolvability in normal-hearing and hearing-impaired listeners

by

Joshua G.W. Bernstein

B.S. Electrical Engineering
Cornell University, 1999

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
MAR 3 0 2006
LIBRARIES

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

DOCTOR OF PHILOSOPHY
IN
SPEECH AND HEARING BIOSCIENCE AND TECHNOLOGY

ARCHIVES

AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
FEBRUARY 2006

©2006 Massachusetts Institute of Technology
All rights reserved.

Signature of Author: _____
Division of Health Sciences and Technology
November 22, 2005

Certified by: _____
Andrew J. Oxenham
Principal Research Scientist, Research Laboratory of Electronics
Lecturer, Division of Health Sciences and Technology
Thesis Supervisor

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

Pitch perception and harmonic resolvability in normal-hearing and hearing-impaired listeners

by

Joshua G.W. Bernstein
B.S. Electrical Engineering, Cornell University, 1999

Submitted to the Division of Health Sciences & Technology on November 22, 2005 in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Speech and Hearing Bioscience and Technology

ABSTRACT

Listeners with sensorineural hearing loss are often impaired in their ability to perceive the pitch associated with the fundamental frequency (F0) of complex harmonic sounds. Four studies investigated the relationship between F0 discrimination performance and the ability to resolve individual harmonic frequencies of a complex, testing the hypotheses (1) that the accurate F0 discrimination performance associated with low-order harmonics is due to their being resolved, and (2) that listeners with sensorineural hearing loss experience a pitch discrimination deficit due to a reduction in frequency selectivity.

The first study revealed that resolved harmonics were *not sufficient* for accurate F0 discrimination. Increasing harmonic resolvability by presenting even and odd harmonics to opposite ears did not improve pitch discrimination, raising the possibility that complex-tone pitch discrimination is not governed by harmonic resolvability *per se*, but is related to harmonic number. Based on this idea, the second study found that an autocorrelation model of pitch perception, modified to include place dependence by limiting the range of periodicities accurately processed by a given frequency channel, could account for the more accurate F0 discrimination associated with low-order harmonics without relying on harmonic resolvability.

However, further results in the third and fourth studies suggested a role for harmonic resolvability in pitch discrimination, inconsistent with the lack of dependence on resolvability of the modified autocorrelation model. In normal-hearing subjects at high stimulus levels and in hearing-impaired subjects, a wider spacing between adjacent frequency components, related to a reduction in frequency selectivity, was required to yield accurate F0 discrimination performance. Thus, resolved harmonics may be *necessary* for accurate F0 encoding, and the pitch discrimination deficit associated with sensorineural hearing loss may be related to a reduction in frequency selectivity.

These results support spectral or spectrotemporal pitch models that derive F0 from resolved harmonics, or a place-dependent temporal model whereby peripheral filter bandwidths limit the range of detectable periodicities. Because spectral processing plays an important role in pitch discrimination, hearing-impaired and cochlear-implant listeners may benefit from hearing-aid fitting procedures and cochlear-implant processing algorithms that emphasize or enhance spectral place cues.

Thesis Supervisor: Andrew J. Oxenham
Title: Principal Research Scientist, Research Laboratory of Electronics;
Lecturer, Division of Health Sciences and Technology

Biographical Note

Joshua Bernstein did his undergraduate work Cornell University's College of Engineering, Ithaca, NY (1995-1999), earning a B.S. in Electrical Engineering, Summa Cum Laude, in May, 1999. While a student at Cornell, he participated in the Engineering Cooperative Program, interning at Hughes Space & Communications, El Segundo, CA, as a satellite payload test engineer and design engineer.

He then enrolled in the Speech and Hearing Bioscience and Technology (SHBT) Doctoral Program in the Division of Health Science and Technology at the Massachusetts Institute of Technology (MIT) and Harvard University (2000-2005), with a fellowship from the National Institute of Health (NIH) through the National Institute for Deafness and Other Communication Disorders. He completed the work presented in this thesis under the supervision of Dr. Andrew Oxenham, as a member of the Auditory Perception and Cognition Group at MIT's Research Laboratory of Electronics. Portions of this work earned him Honorable Mention for the Helen Carr Peake Prize for Bioengineering Research (2003) and a Mentored Doctoral Student Research Poster Session Award from the Amercian Auditory Society and the National Institute of Health (2005). During his time as a doctoral student, he also served as a teaching assistant for the graduate level course in "Acoustics of Speech and Hearing".

He is the first author of two publications on pitch perception (Chapters 2 and 3 of this thesis):

1. Bernstein, J.G. and Oxenham, A.J. (2003). Pitch discrimination of diotic and dichotic tone complexes: Harmonic resolvability or harmonic number? *J. Acoust. Soc. Am*, **113**, 3323-3334.
2. Bernstein, J.G.W. and Oxenham, A.J. (2005). An autocorrelation model with place dependence to account for the effect of harmonic number on fundamental frequency discrimination. *J. Acoust. Soc. Am.,* **117**, 3816-3831.

and co-author of two additional publications:

3. Oxenham, A.J., Bernstein, J.G.W. and Penagos, H. (2004). Correct tonotopic representation is necessary for complex pitch perception. *Proc. Nat. Acad. Sci*, **101**, 1421-1425.
4. Oxenham, A.J., Bernstein, J.G. and Micheyl, C. (2004). "Pitch perception of complex tones within and across ears and frequency regions," in Auditory Signal Processing: Physiology, Psychoacoustics, and Models, Eds. D. Pressnitzer, A. de Cheveigne, S. McAdams, and L. Collet (Springer, New York).

Table of Contents

List of figures.....5

List of tables.....6

1. Introduction.....7
2. Sufficiency of resolved harmonics..... 14
3. An autocorrelation model with place dependence.....42
4. Necessity of resolved harmonics I: Stimulus level.....78
5. Necessity of resolved harmonics II: Sensorineural hearing loss..... 107
6. Summary and conclusions..... 147
7. Implications and future directions.....151

References..... 155

Acknowledgements..... 165

List of figures

Figure	Description	Page
2.1	Schematic of stimuli used in Experiment 2A (Hearing out harmonics)	20
2.2	Mean results, experiment 2A	24
2.3	Individual subject results, experiment 2A	24
2.4	Mean results, experiment 2B (f_0 DLs for diotic and dichotic complexes)	28
2.5	Individual results, experiment 2B	30
2.6	Mean results, experiment 2C (Perceived pitch of dichotic complexes)	34
3.1	Results of experiment 3 (f_0 DLs for fixed spectral envelope) and autocorrelation model predictions	51
3.2	Sample autocorrelation functions	62
3.3	Estimates of model-predicted d' as a function of f_0 difference	63
3.4	Effect of background noise on model predictions	64
3.5	Sample autocorrelation CF-dependent weighting functions	69
3.6	Estimates of predicted d' for the CF-dependent modified autocorrelation model	70
3.7	Modified autocorrelation model f_0 DL predictions	71
3.8	Mean summary autocorrelation functions for modified model	72
4.1	Results, experiment 4A (f_0 DL versus level for fixed spectral envelope)	84
4.2	Summary measures of f_0 discrimination and frequency selectivity derived from the results of experiments 4A, B and C	86
4.3	Auditory filter shapes across level derived from notched noise masking data of Experiment 4B	91
4.4	Sample excitation patterns	94
4.5	Excitation-pattern peak-to-valley ratios	94
4.6	Schematic of stimuli used in experiment 4C (Hearing out harmonics)	95
4.7	Results, experiment 4C	98
5.1	Audiograms	113
5.2	Results of experiment 5A: f_0 DLs	118
5.3	Results of experiment 5B: Modulation discrimination	122
5.4	Results of experiment 5C: Notched-noise masking	126
5.5	The f_0 DL phase effect (experiment 5A)	133
5.6	Correlations involving the f_0 DL transition point ($f_{0,tr}$)	136
5.7	Correlations between frequency selectivity estimates and hearing loss	136
5.8	Correlations involving f_0 DL_{min}	139

List of tables

Table	Description	Page
3.1	RMANOVA results for experiment 3	52
4.1	Auditory filter shape parameters	91
5.1	Audiological, biographical and stimulus-level information	113
5.2	Accuracy of four filter-shape models in fitting the notched-noise data	127
5.3	Best-fitting f_0 DL and frequency selectivity summary measures	129
5.4	Student t-tests comparing NH and HI listeners in experiment 5A, B and C	130
5.5	Summary of correlations between f_0 DL and frequency selectivity measures	134
5.6	Correlations between f_0 DL and frequency selectivity measures after partialling out audiometric threshold	137
6.1	Summary of pitch models	150

Chapter 1. Introduction

1.1 The missing fundamental frequency

Harmonic sounds, consisting of a sum of sinusoids, each with a frequency at a multiple of the fundamental frequency (f_0), are ubiquitous in our natural environment. Voiced human speech, sounds evoked by many musical instruments, animal vocalizations, and mechanical vibrations are all quasi-periodic signals whose frequency spectra can be approximated as a series of sinusoids at discrete harmonically related frequencies. The auditory system tends to group the individual harmonic components together into a single percept with a pitch that usually corresponds to the f_0 of the complex, even if the component at the f_0 is absent from the stimulus or is masked (Schouten, 1940; Licklider, 1954). The pitch of a complex sound is a useful attribute in an everyday listening environment. Pitch can convey, for example, musical melody, prosody in running speech, and lexical information in tonal languages. Pitch information can also provide a cue for the segregation of simultaneous talkers (e.g. Darwin and Hukin, 2000), thus aiding speech intelligibility in a noisy environment.

1.2 The dominance of low-frequency components

Although the mechanisms underlying pitch perception have been investigated for over a century (Seebeck, 1841; Ohm, 1843), they are still debated. Nevertheless, the stimulus parameters that influence complex pitch processing are well documented, and provide some clues as to the underlying mechanisms. One important aspect of complex pitch discrimination is the dominance of low-frequency components in determining the pitch of a harmonic complex (Plomp, 1967; Ritsma, 1967). Low-frequency components also tend to yield a stronger pitch sensation than high-frequency components, both in terms of the salience of musical pitch (Houtsma and Goldstein, 1972; Houtsma and Smurzynski, 1990) and in terms of the size of the just-noticeable difference in f_0 (Houtsma and Smurzynski, 1990).

Two main classes of pitch model are able to account for the dominance of low-order components, although the mechanisms by which they do so differ considerably. "Place" or "spectral" models (e.g. Goldstein, 1973; Wightman, 1973; Terhardt, 1974,1979) hypothesize that the missing f_0 is extracted by comparing the individual frequency components of a harmonic

complex to an internally stored template. These models rely on the spatial separation of frequency components along the cochlear partition, and therefore predict that pitch strength will deteriorate as the spacing between the individual components within a complex becomes so small that the individual peaks in the cochlear representation are no longer resolved. Because the components of a harmonic complex are equally spaced on a linear frequency scale, but the absolute bandwidths of auditory filters increase with increasing center frequency (CF), the density of harmonics per auditory filter increases with increasing harmonic number. As a result, low-order harmonics are resolved from one another, but higher-order harmonics begin to interact within single auditory filters and eventually become unresolved. In contrast, "temporal" models based on the autocorrelation of auditory-nerve fiber activity, pooled across the total population of fibers (e.g. Schouten, 1940; Licklider, 1951,1959; Meddis and Hewitt, 1991a,b; Cariani and Delgutte, 1996a; Meddis and O'Mard, 1997) predict poorer resolution within the model (and hence reduced performance in f_0 discrimination) as the *absolute* frequency of components increases (Cariani and Delgutte, 1996a; Carlyon, 1998), due primarily to the roll-off in the phase-locking properties of auditory-nerve fibers above about 1.5 kHz (Weiss and Rose, 1988a).

"Spectral" models account for the dominance of low-harmonics in terms of *relative* frequency (the ratio between the absolute frequency and the f_0 and the spacing between components), while "temporal" models account for the dominance in terms of *absolute* frequency. Nevertheless, it is likely that both relative and absolute frequency effects play a role in complex pitch perception. Ritsma (1967) demonstrated that the third through fifth harmonics dominated the perceived pitch for various f_0's, such that the dominant frequency region for pitch was relative to the complex's f_0. Investigating a wider range of f_0, Plomp (1967) found that the harmonics that dominated the perceived pitch also depended on the f_0. According to data of Pressnitzer *et al.* (2001) and a series of f_0 discrimination studies (Ritsma, 1962; Ritsma and Hoekstra, 1974; Cullen and Long, 1986) summarized by Krumbholz *et al.* (2000), the transition between good and poor f_0 discrimination performance occurs at a constant harmonic number in moderate absolute frequency regions, but not in low or high regions, suggesting that the pitch percept is influenced by both absolute and relative frequency effects.

These results suggest that both temporal and spectral information to extract f_0 information. "Spectrotemporal" models (Srulovicz and Goldstein, 1983; Shamma and Klein, 2000; Cedolin and Delgutte, 2005a) are similar to the purely "place" models described above, in that individual resolved harmonic frequencies are first identified, then are matched to internally stored harmonic templates to derive the pitch. However, "spectrotemporal" models propose that temporal information is used to identify the individual resolved component frequencies. Srulovicz and Goldstein (1983) suggest that the individual harmonic frequencies can be extracted by tallying first-order interspike interval histograms in each channel, an operation analogous to a within-channel autocorrelation operation. Alternatively, identification of individual resolved frequencies could be accomplished by an across-channel cross-correlation mechanism (Loeb *et al.*, 1983; Shamma, 1985; Shamma and Klein, 2000). These "spectrotemporal" models would predict the dominance of low-order harmonics based on a combination of spectral (harmonic resolvability) and temporal (phase locking) effects.

A final category of pitch model also combines temporal and spectral information by modifying a purely temporal model to include a cochlear place limitation (Chapter 3; de Cheveigné and Pressnitzer, 2005). For these models, the missing f_0 is derived from the dominant interspike interval across frequency channels, but the range of periodicities accurately processed by a given channel are limited relative to the channel's CF (Chapter 3) or impulse response duration (de Cheveigné and Pressnitzer, 2005). These pitch models account for the dominance of low-order harmonics as a result of these CF-dependent limitations (see Chapter 3). In this thesis, such models are termed "CF-dependent temporal," differentiating them from the "spectrotemporal" models described above that use both temporal and spectral information but in a "harmonic template" sense more akin to traditional "spectral" models.

1.3 Sensorineural hearing loss and pitch perception

Listeners with sensorineural hearing loss (SNHL) generally experience a pitch-processing impairment, as evidenced by a deficit in their ability to discriminate small changes in f_0 (Moore and Peters, 1992; Arehart, 1994; Moore, 1995; Arehart and Burns, 1999). This deficit may contribute to the poor speech intelligibility in noise suffered by these listeners. For example, listeners with SNHL show a reduction in the ability to segregate two simultaneous vowels with

different f_0's (Stubbs and Summerfield, 1988), a task which requires processing of the f_0 and may be important for speech intelligibility in multi-talker environments. The mechanisms underlying the pitch-processing deficit experienced by listeners with SNHL could be spectral and/or temporal in nature. This thesis focused on the hypothesis implied by "spectral" and "spectrotemporal" models of pitch, that the dominance of low-order harmonics reflects the dependence of f_0 extraction of the presence of resolved harmonics. According to this hypothesis, listeners with SNHL experience a pitch-processing deficit because of a reduction in harmonic resolvability resulting from wider peripheral filters.

1.4 Goals

The main goals of this thesis were twofold. The first goal was to explore and clarify the role of frequency selectivity and harmonic resolvability in f_0 discrimination. The second goal was to determine to what extent deficits in frequency selectivity are responsible for the deterioration in f_0 discrimination performance that accompanies sensorineural hearing loss (SNHL). The relationship between peripheral frequency selectivity, harmonic resolvability, harmonic number and f_0 processing was investigated via psychophysical studies of the relationship between f_0 discrimination and frequency selectivity in normal-hearing (NH) and hearing-impaired (HI) subjects, and physiologically-based computational modeling.

The research undertaken in this thesis addressed three main questions:

(1) To what extent does the ability to discriminate the pitch of complex sounds depend on the three related attributes of harmonic complex stimuli and the peripheral auditory system: harmonic number, harmonic resolvability and frequency selectivity?

(2) Can a temporal pitch mechanism that does not rely on harmonic resolvability *per se* account for the apparent relationship between pitch discrimination performance and the three quantities addressed in (1)?

(3) To what extent do deficits in frequency selectivity contribute to deterioration in pitch salience in listeners with sensorineural hearing loss (SNHL)?

1.5 Methodology

1.5.1. Overview

To test the hypothesis that the dependence of f_0 discrimination performance on harmonic number reflects peripheral harmonic resolvability, it was necessary to determine whether the presence of resolved harmonics is both necessary and sufficient for producing good f_0 discrimination performance. Chapter 2 addresses the question of *sufficiency*. Even and odd components of a harmonic stimulus were presented to opposite ears, thereby approximately doubling the number of available resolved harmonics. The argument that resolved harmonics are *sufficient* for good f_0 discrimination would be supported by results indicating that the transition from poor to good f_0 discrimination occurs at twice the harmonic number in the dichotic case, where twice as many resolved harmonics are available.

The results of chapter 2 indicated that the additional resolved harmonics provided by dichotic presentation did not improve f_0 discrimination performance. This raised the possibility that the f_0 discrimination does not depend on harmonic resolvability *per se*, but some other limitation related to the ratio between f_0 and absolute frequency. Chapter 3 tested one such possibility: a temporal autocorrelation-model modification proposed by Moore (1982), whereby for each auditory nerve fiber (ANF), the range of lags for which the autocorrelation is calculated is limited relative to ANF characteristic frequency (CF). Chapter 3 investigated whether such a model could explain the harmonic number dependence of f_0 DLs.

Chapters 4 and 5 address the question of whether resolved harmonics are *necessary* for accurate f_0 discrimination. The modified autocorrelation model of Chapter 3, where the harmonic number dependence of f_0 DLs did not derive from harmonic resolvability, would predict that resolved harmonics are not only insufficient for good f_0 DLs (Chapter 2), but they are also unnecessary. In contrast, spectral and spectrotemporal models of pitch depend critically on resolved harmonics for good f_0 discrimination performance. In Chapter 4, harmonic stimuli were presented at high levels, where auditory filters are known to increase in bandwidth (e.g. Rhode, 1971; Weber, 1977; Robles *et al.*, 1986; Glasberg and Moore, 1990; Rosen and Stock, 1992), yielding fewer resolved harmonics. In chapter 5, harmonic stimuli were presented to hearing-impaired listeners, who generally have widened auditory filters and therefore fewer available resolved harmonics, to

test the hypothesis that the f_0 discrimination deficit experienced by listeners with SNHL is related to a loss of peripheral frequency selectivity. The results from both chapters support the hypothesis that resolved harmonics are necessary for good f_0 discrimination. Reduced frequency selectivity, caused by both high presentation levels in normal hearing and cochlear dysfunction in impaired hearing, resulted in a decrease in the harmonic number at which the transition between good and poor f_0 discrimination occurs.

1.5.2 "Peripheral" frequency selectivity and harmonic resolvability

Chapters 2, 4 and 5 use psychophysical methods to estimate the frequency selectivity of the auditory system and the extent to which individual harmonics are resolved under the conditions described in the previous section. Harmonic resolvability, defined as the ability of the auditory system to identify the frequency of an individual component of a harmonic complex, is estimated in a "hearing out harmonics" paradigm, where the listener's task is to discriminate the frequency of a harmonic component from that of a pure-tone presented in isolation (Chapters 2 and 4). Estimates of the frequency selectivity of the auditory system, defined in terms of auditory filter bandwidths, are derived form tone-in-noise spectral masking patterns (Chapters 4 and 5) and from the effects of component phase interactions on the perception of complex tones (Chapter 5).

The inherent difficulty associated with ascribing psychoacoustical phenomena to a physiological mechanism based in a particular physical location in the anatomy of the auditory system obscures the question of whether accurate f_0 discrimination performance is dependent on the *peripheral* resolvability of harmonics. This is especially the case for the "hearing out harmonics" paradigm (Chapters 2 and 4), where the ability to identify the frequency of a particular harmonic only means that it is resolved at some level of the auditory system. Nevertheless, several results from the literature support the idea that the other frequency selectivity estimation methods in this thesis address mechanisms located in the auditory periphery – that is, at the level of the basilar membrane and auditory-nerve responses. Evans (2001) demonstrated in guinea pigs a close correspondence between the behavioral equivalent rectangular bandwidth (ERB) obtained using a notch-noise paradigm (i.e. Chapters 4 and 5) and the physiological ERB obtained from auditory nerve frequency threshold curves. Similarly, Shera *et al.* (2002) showed that physiological peripheral filter bandwidths estimated using

otoacoustic emission data in humans were similar to those obtained in a psychoacoustic forward masking notched-noise paradigm. The estimates based on phase interactions of individual frequency components (Chapter 5) are also likely to represent peripheral mechanisms, since the sharp phase transitions as a function of stimulus frequency observed in auditory nerve responses (Ruggero *et al.*, 1997) would disrupt the fine-timing comparisons that would needed to detect phase interactions across peripheral channels.

1.5.3. Summary

Overall the results from this thesis suggest that resolved harmonics are necessary but not sufficient for accurate coding of f_0 information. The final chapter of the thesis discusses the implications of the results and suggests new lines of research that could be pursued in the ongoing effort to understand how pitch is processed within the auditory system.

Chapter 2. Are resolved harmonics sufficient for accurate f_0 discrimination?

This work described in this chapter is published in the *Journal of the Acoustical Society of America.*

Bernstein, J.G. and Oxenham, A.J. (2003). Pitch discrimination of diotic and dichotic tone complexes: Harmonic resolvability or harmonic number? *J. Acoust. Soc. Am.* **113**, 3323-3334

2.1 Abstract

Three experiments investigated the relationship between harmonic number, harmonic resolvability, and the perception of harmonic complexes. Complexes with successive equal-amplitude sine- or random-phase harmonic components of a 100- or 200-Hz fundamental frequency (f_0) were presented dichotically, with even and odd components to opposite ears, or diotically, with all harmonics presented to both ears. Experiment 2A measured performance in discriminating a 3.5-5% frequency difference between a component of a harmonic complex and a pure tone in isolation. Listeners achieved at least 75% correct for approximately the first 10 and 20 individual harmonics in the diotic and dichotic conditions, respectively, verifying that only processes before the binaural combination of information limit frequency selectivity. Experiment 2B measured fundamental frequency difference limens (f_0 DLs) as a function of the average lowest harmonic number. Similar results at both f_0s provide further evidence that harmonic number, not absolute frequency, underlies the order-of-magnitude increase observed in f_0 DLs when only harmonics above about the 10th are presented. Similar results under diotic and dichotic conditions indicate that the auditory system, in performing f_0 discrimination, is unable to utilize the additional peripherally resolved harmonics in the dichotic case. In experiment 2C, dichotic complexes containing harmonics below the 12th, or only above the 15th, elicited pitches of the f_0 and twice the f_0, respectively. Together, experiments 2B and 2C suggest that harmonic number, regardless of peripheral resolvability, governs the transition between two fundamentally different pitch percepts, one based on the frequencies of individual resolved harmonics and the other based on the periodicity of the temporal envelope.

2.2 Introduction

The mechanisms underlying pitch perception have been a matter of intense debate ever since Ohm (1843) disputed Seebeck's (1841) description of the phenomenon of the missing fundamental frequency (f_0). More recently, one aspect of this debate has been concerned with the mechanisms underlying the different contributions that low- and high-frequency harmonics make to the overall perceived pitch of a harmonic complex. Early work showed a dominant frequency region for pitch that was determined by both relative and absolute frequency relations. Ritsma (1967) demonstrated that the third through fifth harmonics dominated the perceived pitch for various f_0s, such that the dominant frequency region for pitch was relative to the complex's f_0. Investigating a wider range of f_0s, Plomp (1967) found that the harmonics that dominated the perceived pitch also depended on the f_0 of the complex, suggesting that absolute frequency also influenced the dominance region.

Most models of pitch perception can account qualitatively for the dominance of low harmonics in determining the overall pitch and for the greatly reduced pitch salience observed when only high harmonics are presented. However, the mechanisms by which they do so differ considerably. For instance, models that rely on the spatial separation of frequency components along the cochlear partition (e.g. Goldstein, 1973; Wightman, 1973; Terhardt, 1974,1979) predict that pitch salience will deteriorate as the spacing between the individual components within a complex becomes so small that the individual peaks in the cochlear representation are no longer resolved. Because the components of a harmonic complex are equally spaced on a linear frequency scale, but the absolute bandwidths of auditory filters increase with increasing center frequency (CF), the density of harmonics per auditory filter increases with increasing harmonic number. As a result, low-order harmonics are resolved from one another, but higher-order harmonics begin to interact within single auditory filters and eventually become unresolved. In contrast, models based on the autocorrelation of auditory-nerve fiber activity, pooled across the total population of fibers (e.g. Meddis and Hewitt, 1991a,b; Cariani and Delgutte, 1996a,b; Meddis and O'Mard, 1997) predict poorer resolution within the model (and hence reduced performance in f_0 discrimination) as the *absolute* frequency of components increases (Cariani and Delgutte, 1996a; Carlyon, 1998), due primarily to the roll-off in the phase-locking properties of auditory-nerve fibers above about 1.5 kHz (Weiss and Rose, 1988a). These two categories of

models are often referred to as "place" and "temporal" models, respectively. However, it should be noted that the term "place model" does not necessarily imply that the frequencies of individual harmonics are encoded via a place mechanism. Instead it is possible that the frequency information at each place is encoded via a temporal mechanism (Srulovicz and Goldstein, 1983; Shamma and Klein, 2000). Nevertheless, it is important for these place models that the components are sufficiently well resolved for the frequency of each to be estimated individually.

The defining role of absolute frequency and phase locking, implied by temporal models based on the pooled autocorrelation function, has been put into question by various psychophysical experiments indicating that relative frequency relationships play an important role in the deterioration of pitch salience for high-order harmonics. Houtsma and Smurzynski (1990) estimated pitch salience, in terms of melodic interval recognition and fundamental frequency difference limens (f_0 DLs), for complex tones comprising 11 successive harmonics as a function of the lowest harmonic present. They found that for both measures, performance was much poorer when only harmonics above the 10th were presented than when at least some harmonics below the 10th were present. Although they carried out their experiment at only one f_0 (200 Hz), meaning that the respective influences of absolute and relative frequencies could not be distinguished, earlier research with two harmonics (Houtsma and Goldstein, 1972), and later research with many harmonics (Carlyon and Shackleton, 1994; Shackleton and Carlyon, 1994; Kaernbach and Bering, 2001) strongly support the idea that performance in such tasks is limited primarily by the lowest harmonic number present, and not by the lowest absolute frequency present.

While it has been generally assumed that pitch discrimination deteriorates when only high harmonics are present because the harmonics are peripherally unresolved (Houtsma and Smurzynski, 1990; Carlyon and Shackleton, 1994; Shackleton and Carlyon, 1994), certain results in the literature cast some doubt on this interpretation. Houtsma and Goldstein (1972) estimated the pitch strength of harmonic complexes consisting of two successive components by measuring performance in musical interval identification. Harmonics that are unresolved when both are presented to the same ear (monotic), become resolved when presented to opposite ears (dichotic).

If strong pitch salience required the presence of resolved harmonics, we might expect stronger pitch salience when two normally unresolved harmonics (i.e., unresolved under monotic presentation) are presented dichotically. However, the decrease in performance with increasing harmonic number was the same under monotic and dichotic presentations, suggesting that the decrease in pitch salience with increasing harmonic number may not be due to the harmonics becoming unresolved *per se*. Arehart and Burns (1999) reported similar results using three musically trained hearing-impaired listeners.

This paper further investigates the transition in f_0 DLs found in the data of Houtsma and Smurzynski (1990), to determine whether the frequency at which it occurs is defined by harmonic resolvability, harmonic number regardless of resolvability, or absolute frequency. An f_0 DL paradigm (Houtsma and Smurzynski, 1990) was used to test whether presenting normally unresolved components to opposite ears improves performance. Under diotic presentation, all components were presented to both ears, such that the peripheral spacing between components was the f_0. Under dichotic presentation, even and odd components were presented to opposite ears, such that peripheral spacing between components was $2f_0$. The approach differs from those of two earlier studies addressing this issue (Houtsma and Goldstein, 1972; Arehart and Burns, 1999) in two principal ways. First, the f_0 discrimination task does not require the musical training that is necessary for a musical interval identification task. Second, 12-component complexes yield a much stronger pitch salience than the relatively weak pitch elicited by two-tone complexes, even with low-order harmonics.

Underlying this study was the important assumption that approximately twice as many harmonics should be resolved in the dichotic conditions, where the peripheral frequency spacing between components is twice that of the diotic conditions. The first experiment was designed to test the validity of this assumption. In addition, experiment 2A addressed the discrepancy in the literature between direct and indirect estimates of harmonic resolvability, as described below.

2.3 Experiment 2A: Resolvability of individual harmonics

2.3.1 Rationale

The existing studies on pitch perception show very good consistency in terms of the locus of the transition region between good and poor f_0 discrimination (Cullen and Long, 1986; Houtsma and Smurzynski, 1990). However, as pointed out by Shackleton and Carlyon (1994), while these data sets show a transition that occurs between harmonic numbers 10 and 13, direct measures of individual component resolvability have shown that listeners are generally only able to hear out the first 5 to 8 harmonics of a harmonic complex (Plomp, 1964; Plomp and Mimpen, 1968). Similarly, Shackleton and Carlyon (1994) concluded that the limits of the resolvability of individual components within an inharmonic tone complex, as measured by Moore and Ohgushi (1993), were also lower than those estimated indirectly using f_0 difference limens for harmonic tone complexes.

One reason for this discrepancy might be the nature of the respective tasks. Musicians have been shown to have better performance than non-musicians in "hearing out" harmonics (Soderquist, 1970; Fine and Moore, 1993), while their auditory filter bandwidths are not significantly different (Fine and Moore, 1993). The difference between direct and indirect estimates of peripheral resolvability may be attributable to attentional limitations, whereby, in hearing out individual partials, subjects may have difficulty overcoming their perceptual fusion of the complex into a single auditory object. The difference could also be due to other non-peripheral limitations. In contrast to the Plomp (1964) and Moore and Ohgushi (1993) studies, which required subjects to hear out an individual partial presented simultaneously with a complex, this study gated the target harmonic on and off repeatedly within the presentation interval. This strategy was designed help overcome any non-peripheral limitations and to encourage perceptual segregation, while not affecting peripheral resolvability[1]. If good f_0 discrimination depends on

[1] Viemeister and Bacon (1982) found that a component whose onset is delayed relative to the remaining component produced more forward masking than when the entire complex is gated synchronously. If this "enhancement" effect can be thought of as "amplifying" the representation in a subset of auditory nerve fibers, this should not have any effect on peripheral resolvability, as the signal-to-noise ratio within that population would be unaffected. In fact, physiological enhancement of the response to a component of a harmonic complex with delayed onset time has been found in the cochlear nucleus (Scutt *et al.*, 1997) but not in the auditory nerve (Palmer *et al.*, 1995) of the guinea pig.

the presence of peripherally resolved harmonics, we expect that listeners should be able to hear out approximately 10 harmonics, more than the 5 to 8 measured by Plomp (1964).

2.3.2. Procedure

In this and subsequent experiments, all subjects had some degree of musical training. The least musically trained subject had 4 years of instruction in middle school, while the most musically trained were two professional musicians with more than 18 years formal training. All subjects had normal hearing (15 dB HL or less re ANSI-1969 at octave frequencies between 250 Hz and 8 kHz). Four subjects (ages 18-26, two female) participated in this experiment.

All stimuli were presented in a background noise, which we will call modified uniform masking noise (UMN_m). This noise is similar to uniform masking noise (UMN, Schmidt and Zwicker, 1991), in that it is intended to yield pure-tone masked thresholds at a constant sound pressure level (SPL) across frequency, but the spectrum is somewhat different; UMN_m has a long-term spectrum level that is flat (15 dB/Hz in our study) for frequencies below 600 Hz, and rolls off at 2 dB/octave above 600 Hz. The noise was low-pass filtered with a cutoff at 10 kHz. Thresholds for pure tones at 200, 500, 1500, and 4000 Hz in UMN_m in the left ear were estimated via a three-alternative forced-choice, 2-down, 1-up adaptive algorithm (Levitt, 1971). For each subject, pure tone thresholds in UMN_m fell within a 5 dB range at all four frequencies tested, such that harmonic components presented at equal SPL had nearly equal sensation level (SL). As an approximation, we defined 0 dB SL for each subject as the highest of the thresholds across the four frequencies tested, which ranged from 29.7 to 33 dB SPL across all subjects in this and subsequent experiments.

The stimuli were generated digitally and played out via a soundcard (LynxStudio LynxOne) with 24-bit resolution and a sampling frequency of 32 kHz. The stimuli were then passed through a programmable attenuator (TDT PA4) and headphone buffer (TDT HB6) before being presented to the subject via Sennheiser HD 580 headphones. Subjects were seated in a double-walled sound-attenuating chamber.

Each trial in the experiment consisted of two intervals, each with a 1-s duration, separated by 375 ms of silence. The first interval contained three bursts of a 300-ms sinusoid (referred to as the comparison tone), including 20-ms Hanning window onset and offset ramps, separated by 50-ms silent gaps. The second interval consisted of a harmonic complex with the first 40 successive harmonics of the f_0 with duration 1000 ms, including 20-ms Hanning window onset and offset ramps. Components were presented in random phase to order to ensure that the frequency of the target component was detectable only if the component was spectrally resolved[2]. The target component was gated on and off in the same manner as in the first interval, while all the other components were on continuously throughout the interval. Each component was presented at a nominal 15 dB SL [adjusted for each subject], such that the stimuli in this experiment were similar in level to those used in experiment 2B. The task was a two-alternative forced-choice task, where the listener was required to discriminate which of the comparison tone (interval 1) or target tone (interval 2) was higher in frequency. A schematic of the stimuli is shown in Fig. 2.1.

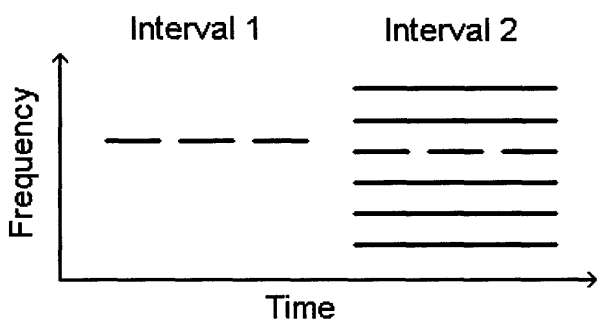

Figure 2.1. Schematic of the stimuli used in experiment 2A. Interval 2 contains a 40-component harmonic complex, with the target harmonic gated on and off to perceptually remove it from the complex. Interval 1 contains a pure-tone probe, higher or lower in frequency than the target harmonic in interval 2, gated in the same way as the target harmonic.

Four conditions were presented, for all combinations of the harmonic complex in interval 2 presented diotically or dichotically, with a 100- or 200-Hz average f_0 ($\bar{f}_0$). Fifty trials for each of ten target harmonic numbers in each condition were presented (diotic: 5 through 14, inclusive; dichotic: 11, 12, 13, 14, 16, 18, 20, 22, 25, and 28), for a total of 500 trials per condition. The

[2] The temporal waveform for several harmonics of a sine-phase complex that fall in one auditory filter is click-like, with brief peaks occurring at intervals of the f_0, separated by low-level epochs. Eliminating a spectrally unresolved harmonic component (i.e. adding it out of phase) will result in that component appearing during the low-level epochs, thereby allowing the detection of the subtracted component's frequency by "listening in the valleys," or "dip listening" (Duifhuis, 1970). Since random-phase complexes generally have much flatter temporal envelopes and are not conducive to listening in the valleys (Alcántara and Moore, 1995), this greatly reduced the possibility of dip listening.

trials were presented in runs, each consisting of 5 trials for each of the ten harmonics for one condition, presented in random order. In the dichotic conditions, the comparison and target harmonics were always presented to the same ear throughout a run, and the distribution of the even and odd harmonics of the complex in interval 2 to the left and right ears was varied accordingly. For example, for a trial where the target 14th harmonic and comparison tone were presented to the right ear, the even harmonics in interval 2 were also presented to the right ear. In the dichotic conditions, five runs were presented with the target in the left ear, and five runs were presented with the target in the right ear.

The difference (Δf) between the frequency of the comparison tone (f_{comp}) and that of the target tone (f_{targ}) was set as a proportion of f_{targ}. This is different from Plomp's (1964) experiment, where he required listeners to identify which of two pure tones was in fact a component of the complex. One comparison tone was at the frequency of one of the components, and the other was halfway between the frequencies two successive components, such that it always fell at the same place relative to the target tone on a *linear* scale. In our experiment, the comparison tone was adjusted relative to the target tone on a *logarithmic* scale, ensuring that any decrease in performance with increasing harmonic number reflects a reduction in resolvability, and not the increase in linear pure tone DLs with increasing frequency (Moore, 1973).

In each trial, f_{comp} was either higher or lower than f_{targ}, each with probability 0.5, with $\Delta f = |f_{targ} - f_{comp}|$ chosen from a uniform distribution of 3.5 to 5.0% of the f_{targ}. The value of Δf was always at least 3.5% of the f_{targ}, which is well above the frequency discrimination threshold for tones in quiet (Moore, 1973). The f_0 of the complex was randomly chosen from a uniform distribution between $0.935\,\bar{f}_0$ and $1.065\,\bar{f}_0$. Randomizing Δf was intended to prevent the listener from correctly identifying the frequency relationship without actually hearing out the target tone, by memorizing the frequency relationship between the comparison tone and the complex's f_0. Testing a large number of target harmonics (10 per condition) and randomizing f_0 further prevented this type of alternative cue[3].

[3] In fact, the combined randomizations ensured that, for a given f_{comp}, the probability that the frequency of the target harmonic was higher than f_{comp} was approximately equal to the probability of it being lower (except when $f_{comp} < f_{targ}$

Each subject began with a training phase, where runs rotated through the four conditions, during which feedback was provided. Training continued until a subject was reliably obtaining nearly 100% correct for the lowest harmonic tested in each condition. The training period varied across subjects from 15 minutes to 2 hours. During the data collection phase, feedback was not provided.

2.3.3. Results

Figure 2.2 shows the mean data. The error bars denote ± 1 standard error of the mean performance across all listeners. Although there was significant variability in performance across subjects, a systematic trend is clear in the data. Percent correct generally decreases with increasing harmonic number, with the 75% correct point corresponding roughly to the 10th harmonic in the diotic conditions, and to the 20th harmonic in the dichotic conditions. For each condition, the pooled data from all subjects were fit (solid lines in Fig. 2.2) to a complementary error function (erfc) bound to 50% and 100% correct at the extremes[4]. The non-linear least squares Gauss-Newton method was used to fit the data to Eq. 2.1 with two free parameters (n_0 and w). The estimated n_0 was taken to be the estimated limit of harmonic resolvability, in accordance with the methods of Plomp (1964). Judgments of the goodness of fit were based on a 95% confidence interval ($\pm 2\sigma$) measure of uncertainty in the n_0 estimate. The values obtained for the estimated limits of resolvability and 95% confidence interval, $n_0 \pm 2\sigma$, for the pooled data were: 9.34 ± 1.03 (diotic 100 Hz), 21.18 ± 1.65 (dichotic 100 Hz), 11.20 ± 0.74 (diotic 200 Hz), and 17.73 ± 1.91 (dichotic 200 Hz).

---

when the lowest target component was tested or $f_{comp} > f_{targ}$ when the highest target component was tested), so that subjects were prevented from answering correctly based only on the frequency of the comparison tone.

[4] Percent correct $(n) = 100\left[\frac{1}{2} + \frac{1}{2\sqrt{\pi}}\int_{n}^{\infty} e^{-[w(n'-n_0)]^2}\,dn'\right]$ (2.1)

where n is harmonic number, n' is the harmonic number integration variable, w is a factor describing the slope of the psychometric function and n_0 is the harmonic number that yields 75% correct.

Figure 2.3 shows the individual data. The left column shows data from the 100-Hz $\bar{f}_0$ and the right column shows data from the 200-Hz $\bar{f}_0$. There was considerable intersubject variability in performance, as well as certain non-monotonic trends within individual subjects. One subject (S2) had difficulty hearing out even the lowest harmonics in the 100-Hz diotic condition. Two subjects (S1 and S3) showed non-monotonicities in the diotic conditions near the 12th harmonic. In the dichotic conditions, large non-monotonicities were exhibited by one subject (S3) at the 100- and 200-Hz $\bar{f}_0$s, and by two others (S1 and S2) at the 200-Hz $\bar{f}_0$. For these subjects, performance decreased below 75% in the vicinity of the 12th to 16th harmonics, and then increased before once again dropping below 75% for higher harmonics. The non-monotonicities in the diotic and dichotic conditions in the vicinity of the 12th and 14th harmonics are also present in the mean data (Fig. 2.2).

Individual subject data in each condition were fit to the erfc function (Eq. 2.1). Fits ranged from good for subjects and conditions where the psychometric function exhibited few non-monotonicities (e.g. subject S4, diotic 200 Hz, 2σ= 0.71 harmonics), to extremely poor for subjects and conditions where the psychometric function exhibited many non-monotonicities (e.g. subject S3, dichotic 100 Hz, 2σ= 6.67 harmonics).

2.3.4. Discussion

Five aspects of the results merit attention. First, roughly twice as many harmonics can be heard out in the dichotic conditions as in the diotic conditions. This is the most important result of the experiment, as it verifies the central assumption for experiment 2B, that only processes before the combination of binaural information limit harmonic resolvability.

Second, our estimates of the limits of harmonic resolvability in the diotic conditions are greater than those reported by Plomp (1964). Our results indicate that the first 9 to 11 harmonics of a complex for $\bar{f}_0$s of 100 and 200 Hz are peripherally resolved. This estimate closely matches the indirect estimate of the limits of harmonic resolvability (Houtsma and Smurzynski, 1990; Shackleton and Carlyon, 1994), where the lowest harmonic present must be the 10th or below in order to yield small f_0 DLs. This indicates that enough harmonics are peripherally resolved to

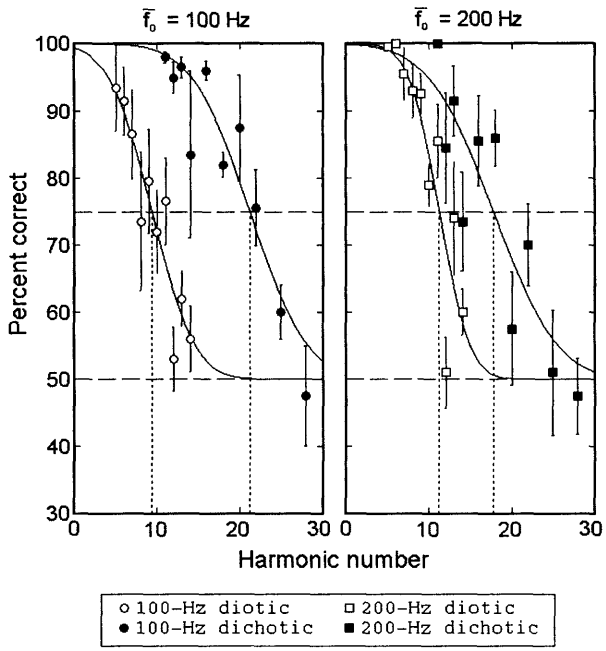

Figure 2.2. Mean results of experiment 2A, showing percent correct in identifying the probe tone as higher or lower than the target tone as a function of harmonic number. Error bars represent plus and minus one standard error across the individual scores for the four subjects. Open symbols indicate diotic conditions, with all harmonics presented to both ears; filled symbols indicate dichotic conditions, with odd and even harmonics presented to opposite ears. The left and right panels show results with f_0s of 100 Hz and 200 Hz, respectively. Solid lines represent the best fits of the erfc function (Eq. 2.1, footnote 4) to the pooled data. The limit of harmonic resolvability, defined as the harmonic that yields 75% correct performance, is depicted by a vertical dotted line. The upper and lower horizontal dashed lines indicate 75% correct (limit of harmonic resolvability) and 50% correct (chance), respectively.

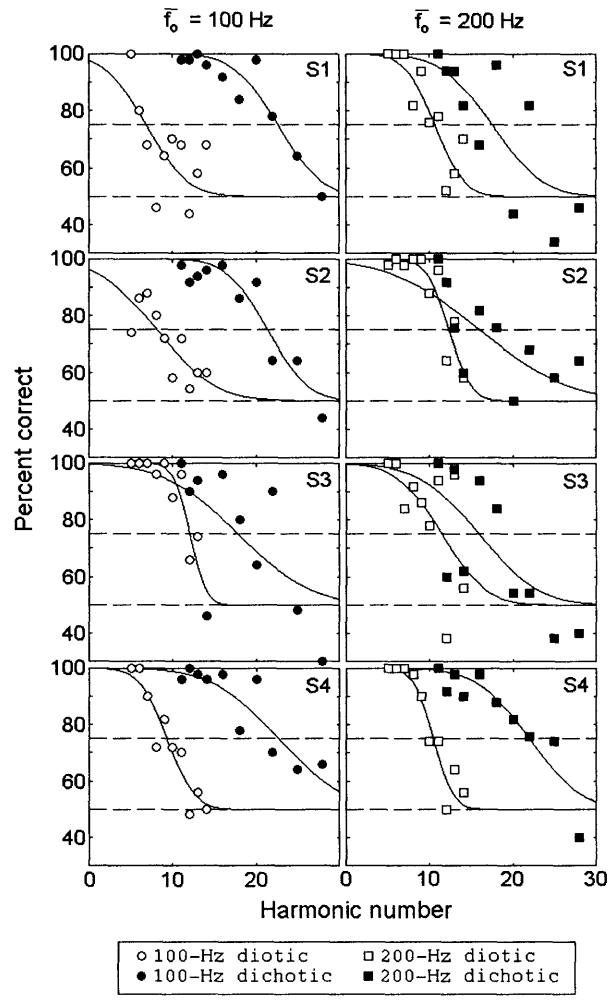

Figure 2.3. Results from the individual subjects in experiment 2A, showing percent correct in identifying the probe tone as higher or lower than the target tone as a function of harmonic number. Each data point represents performance over 50 stimulus trials. Each row represents results from one subject. The left column (circles) and right column (squares) show results with f_0s of 100 Hz and 200 Hz, respectively. The solid curves respresent best fits of the erfc function (Eq. 2.1) to the individual data. The upper and lower dashed lines in each plot represent 75 and 50% correct, respectively.

account for the limits of good f_0 discrimination, thereby resolving the apparent discrepancy between direct and indirect measures of resolvability (Shackleton and Carlyon, 1994). A caveat to this conclusion is that the "enhancement" effect (see footnote 1) may have helped to overcome some non-peripheral limitation to harmonic resolvability that occurs before the detection of pitch. Therefore, in the absence of "enhancement," all of these peripherally resolved harmonics might not be available to the pitch detector. Also, this is an operational definition of resolvability, which depends on the 3.5-5.0% Δf used in this experiment. A smaller Δf may have yielded a lower estimate of the number of resolved harmonics.

Third, there was some indication of more resolved harmonics for the 200-Hz than the 100-Hz f_0, consistent with results of Shera *et. al.* (2002) indicating that the cochlear filter bandwidths relative to CF decrease with increasing absolute frequency at low signal levels. Nevertheless, this difference was small, indicating that harmonic number largely determines resolvability. The limited range of f_0s used in this study prevents a comparison with the effects of f_0 reported by Plomp (1964), where for f_0s greater than 200 Hz, the number of resolved harmonics decreased with increasing f_0.

Fourth, some subjects experienced difficulties with even low-frequency harmonics, or displayed non-monotonic psychometric functions. For example, for subject S2 at the 200-Hz f_0 and subject S3 at both f_0s, the initial drop below 75% correct performance in the dichotic conditions occurred at a similar harmonic number as in the diotic conditions. This suggests that there may be some central limitation on resolution for these subjects and conditions that operates on both diotic and dichotic complexes. However, for all subjects, harmonics above the 14th are well resolved under dichotic presentation, and any central limitation of harmonic resolvability seems to appear only near the 14th harmonic.

Fifth, the estimate of n_0 in the dichotic 200-Hz condition had a large 95% confidence interval (±10.8%), consistent with the poor fit apparent in a visual inspection of the data. Given the high range of pure tone frequencies presented in this condition, this large uncertainty may reflect absolute frequency effects. However, even at the highest frequencies tested (5.6 kHz), the minimum Δf we used (3.5%) is still greater than the 0.5% obtained for similar frequency long-

duration tones in quiet (Moore, 1973). Although the 60 dB SPL tones used in the Moore (1973) study are not comparable to the 15 dB SL tones used in this study, Hoekstra (1979) showed that a reduction from moderate (40 dB) to low (15 dB) SLs increased DLs for a 2-kHz pure tone by less than a factor of two. This suggests that the variable results found at these very high frequencies cannot be ascribed solely to the coding limitations of individual components.

2.4 Experiment 2B: Fundamental frequency difference limens

2.4.1. Rationale

In experiment 2B we measured f_0 DLs as a function of the lowest harmonic number present for diotic and dichotic harmonic complexes. If good discrimination ability were dependent on the presence of resolved harmonics *per se*, the auditory system should be able to utilize the information provided by the additional resolved harmonics available under dichotic presentation, such that the order of magnitude increase in f_0 DLs (Houtsma and Smurzynski, 1990) would occur at twice the harmonic number as compared to diotic presentation. Alternatively, if good discrimination ability were dependent only on the presence of low-numbered harmonics, regardless of resolvability, the additional resolved harmonics should provide no benefit, such that the increase in f_0 DLs would occur at the same lowest harmonic number in both dichotic and diotic conditions.

In order to determine if the increase in f_0 DLs is due to absolute or relative frequency effects, we performed the measurements at two different f_0s (100 and 200 Hz). Based on the results of Shackleton and Carlyon (1994), suggesting that the DL shift is due to relative frequency effects (i.e. the presence or absence of resolved harmonics), we expect that the DL shift should occur at approximately the same harmonic number for both f_0s. Alternatively, if the DL shift were mainly due to absolute frequency effects as implied by many temporal pitch models, then the DL shift should occur at about the same absolute frequency, or twice the harmonic number for the 100-Hz f_0 as compared to the 200-Hz f_0 conditions. While we measured f_0 DLs with harmonics in random phase in order to allow a direct comparison with the harmonic resolvability data of experiment 2A, we also performed the measurements with harmonics in sine phase to allow a more direct comparison with earlier data.

2.4.2. Methods

Stimuli were 500-ms (including 30-ms Hanning window rise and fall) harmonic complexes with 12 successive components. Each component was presented at 10 dB SL in UMN_m background noise (see experiment 2A). This low level was used to prevent the detection of combination tones. Stimuli were presented diotically and dichotically with f_0s of 100 and 200 Hz, in sine phase and random phase, for a total of eight conditions. Discrimination thresholds were estimated for eight normal-hearing subjects. Four subjects (ages 18-24, two female), including the first author, participated in the sine-phase conditions. Two had also participated in experiment 2A. Four new subjects (ages 18-24, one female) participated in the random-phase conditions. The setup for stimulus delivery was the same as in experiment 2A.

Fundamental frequency DLs as a function of the complex's average lowest harmonic number ($\overline{N}$) were estimated via a three-alternative forced-choice, 2-down, 1-up adaptive algorithm tracking the 70.7% correct point (Levitt, 1971). The f_0 difference (Δf_0) was initially set to 10% of the f_0. The starting step size was 2% of the f_0, decreasing to 0.5% after the first two reversals, and then to 0.2% after the next two reversals. The f_0 DL was estimated as the average of the Δf_0s at the remaining six reversal points.

Two of the intervals contained harmonic complexes with a base f_0 ($f_{0,base}$), while one interval contained a complex with a higher f_0 ($f_{0,base} + \Delta f_0$). The task was to identify the interval with the higher f_0. Subjects were informed that two of the intervals had the same pitch, and one had a higher pitch, and were asked to identify the interval with the higher pitch. In order to prevent subjects from basing their judgments on the frequency of the lowest harmonic, the lowest harmonic number (N) was roved from interval to interval, such that in the three intervals it was $\overline{N}$-1, $\overline{N}$, and $\overline{N}$+1, in random order. The highest harmonic number was also roved, such that twelve components were presented in each stimulus interval. For the dichotic conditions, odd and even components were presented randomly to the left or right ear on a trial-by-trial basis. Feedback was provided after each trial. Subjects were informed that there were different sound qualities that varied from interval to interval. They were told to ignore the timbre ("treble/bass quality") of the sounds, as responses based on timbre would result in incorrect answers, and to

respond based solely on the pitch. Fundamental frequency discrimination was tested for $\overline{N}$=3, 6, 9, 12, 15, 18 and 24 in all eight conditions.

Each subject went through a training phase of at least one hour, which continued until performance was no longer showing consistent improvement. During the measurement phase, four adaptive runs were made per subject, for each value of $\overline{N}$ in each condition, and the estimated f_0 DL for a subject was taken as the mean of these four estimates. If the standard deviation across the last six reversals points in any one run was greater than 0.8%, the data for that run was excluded and the run was repeated at the end of the experiment.

2.4.3. Results

Figure 2.4 shows the estimated f_0 DLs (expressed as a percentage of the f_0) as a function of $\overline{N}$. Each data point represents the arithmetic mean and the error bars represent ± the standard error across the mean f_0 DLs measured for four subjects. The central finding of this study is that the dramatic increase in f_0 DLs occurs at the same $\overline{N}$ under diotic and dichotic presentation. Furthermore, this increase occurs at the same $\overline{N}$ at both f_0s.

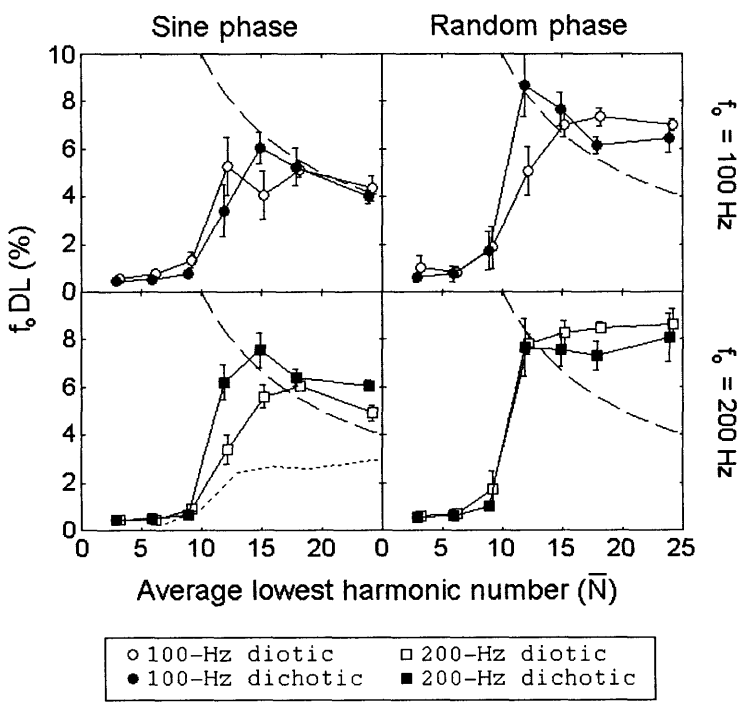

Figure 2.4. Mean results from experiment 2B. Each data point represents the mean f_0 DL (%) across four subjects; error bars denote plus and minus one standard error of the mean. The long-dashed curves show the limit of performance based only on the lower spectral edge of the complexes (see text for details). The short-dashed curve in the lower left panel shows data from Houtmsa and Smurzynski (1990) for a monotic complex with a 200-Hz f_0.

To investigate other trends in the data, an analysis of variance (ANOVA) with three within-subject factors [f_0, $\overline{N}$ and mode of presentation (diotic or dichotic)] and one between-subject factor (phase) was conducted. While the f_0 DL measurement used f_0 steps on a linear frequency scale in accordance with the methods of Houtsma and Smurzynski (1990), the statistical analysis was performed with logarithmically (log) transformed data, in an attempt to satisfy the uniform variance assumption. Only the following main effects and interactions were found to be significant ($p<0.05$). There was a main effect of $\overline{N}$ [$F(1,6) = 179.5$, $p<0.0001$], two-way interactions between f_0 and $\overline{N}$ [$F(1,6) = 5.60$, $p<0.0005$] and between f_0 and mode of presentation [$F(1,6) = 6.8$, $p<0.05$], a three-way interaction between mode of presentation, phase and f_0 [$F(1,6) = 8.32$, $p<0.05$] and a four-way interaction between all factors [$F(6,36) = 5.84$, $p<0.0005$].

The significant four-way interaction suggests caution in interpreting main effects and low-order interactions. Nevertheless, the ANOVA supports two trends in the data concerning $\overline{N}$ and f_0. First, the main effect of $\overline{N}$ clearly reflects the result that good performance in f_0 discrimination requires $\overline{N} \leq 9$. Second, the two-way interaction between f_0 and $\overline{N}$ does not reflect an absolute frequency effect on the frequency of the increase in f_0 DLs, since this transition occurs at the same $\overline{N}$ for both f_0s. Rather, this interaction probably reflects an absolute frequency effect for complexes with $\overline{N} > 9$, where larger f_0 DLs are seen for the 200-Hz f_0 as compared to the 100-Hz f_0. Interpreting the effects of mode of presentation and phase requires a closer examination of the data.

The significant higher-order interactions probably reflect the result that dichotic f_0 DLs were somewhat higher or lower than diotic f_0 DLs depending on f_0, phase and $\overline{N}$. Two trends in the difference between f_0 DLs measured under dichotic versus diotic presentation were apparent in the data. The first trend was that dichotic f_0 DLs were larger than diotic f_0 DLs presentation at $\overline{N} = 12$ or $\overline{N} = 15$ for all combinations of f_0 and phase except for the 200-Hz random phase case. This trend will be addressed further in conjunction with results of experiment 2C. The second, less apparent, trend was that dichotic f_0 DLs were slightly smaller for $\overline{N} = 18$ and $\overline{N} = 24$ at both f_0s. Differences between diotic and dichotic f_0 DLs were seen in some subjects, but not in others.

Figure 2.5 shows mean DLs for two sample subjects who participated in the random-phase conditions. At one extreme, subject S9 (right column) shows larger DLs under dichotic presentation near $\overline{N}=12$ for both f_0s. Of the 16 combinations of subject and f_0, seven showed larger dichotic DLs near $\overline{N}=12$ (four of eight in sine-phase, three of eight in random-phase). At the other extreme, subject S8 (left column) shows larger DLs under diotic presentation at $\overline{N}=18$ and 24. While no subjects showed larger diotic DLs near $\overline{N}=18$ for sine-phase stimuli, two did for random-phase stimuli.

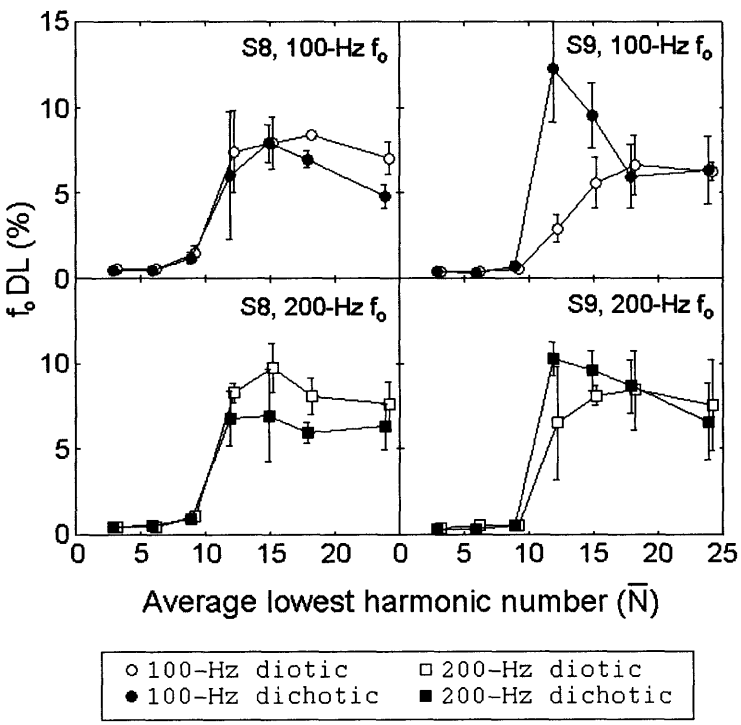

Figure 2.5. Individual results from experiment 2B for two sample subjects. Error bars show plus and minus one standard deviation across four stimulus trials. Subject S8 (left column) shows larger f_0 DLs under dichotic presentation for $\overline{N}$ = 12 and 15. Subject S9 (right column) shows smaller f_0 DLs under dichotic presentation for $\overline{N}$ = 18 and 24.

The results of Houtsma and Smurzynski (1990) suggested that the phase relationship between harmonics affected the f_0 DLs for high-order, but not low-order harmonics. While this trend also appears in our data, the ANOVA indicated no significant main effect of phase or two-way interactions between phase and any other factor ($p>0.05$). Although the f_0 DLs appear larger in the random phase conditions for $\overline{N}>9$, this difference is not statistically significant for the logarithmically transformed data. The lack of a significant phase effect in our data may be due to the fact that phase was a between-subjects factor, giving the test less statistical power than if random and sine phase complexes had been tested in the same subjects.

Another possibility is that even though the lowest harmonic number was roved from interval to interval, for large Δf_0s it is possible for listeners to achieve above chance performance without extracting f_0 information. Phase effects may not be present in any condition where f_0 information was not used to perform the task. Both in our 3IFC study and in the Houtsma and Smurzynski (1990) study, if the listener were to base their answer on the frequency of the lowest harmonic present in each interval (or the low-frequency edge of the excitation pattern for unresolved complexes), they would achieve 66.7% correct (near the 70.7% correct point approximated by the 2-up, 1-down adaptive procedure) if $\Delta f_0/f_0 > 1/\overline{N}$. Any data point falling above the DL = $(100/\overline{N})$% dashed line in Fig. 2.4 could reflect responses based on the "lowest harmonic" cue, rather than f_0 extraction. Performance is slightly worse than the "lowest harmonic" prediction for $\overline{N}=18$ and 24 in the sine-phase conditions, and much worse for $\overline{N}>12$ in all random-phase conditions. Thus in this study, listeners may be using lowest harmonic cues, rather than f_0 pitch cues, to perform f_0 discrimination for complexes with high $\overline{N}$, especially when the components are in random phase. In the Houtsma and Smurzynski (1990) study, f_0 DLs are much smaller than the lowest harmonic cue prediction, and therefore most likely reflect actual f_0 discrimination performance.

To look for possible phase effects, Scheffe post-hoc tests compared sine- and random-phase data for the four combinations of f_0 and mode of presentation for $\overline{N}=12$, which is above the resolved harmonic region, but below the region where the "lowest harmonic" cue may have influenced the results. Results indicate that f_0 DLs were significantly different ($p<0.05$) in the 100-Hz dichotic and 200-Hz diotic conditions, providing some weak evidence for the presence of phase effects in these conditions.

2.4.4. Discussion

The fact that the transition from small to large f_0 DLs occurs at the same $\overline{N}$ under diotic and dichotic presentation indicates that the auditory system is unable to utilize the information provided by the additional resolved harmonics in the dichotic case for f_0 discrimination. While two subjects did show slightly smaller f_0 DLs for dichotic complexes than for diotic complexes when $\overline{N}\geq 18$, the f_0 DLs (around 6%) are still much larger than those found for lower numbered

harmonics. This supports the hypothesis that good f_0 discrimination is not limited by harmonic resolvability, but by harmonic number, regardless of resolvability. This result also indicates that subjects cannot ignore the input from one ear in performing the f_0 discrimination task. Remember that the ear with the even harmonics contains consecutive harmonics of a fundamental $2f_0$, with a lowest harmonic around $\overline{N}/2$. If subjects were able to ignore the ear with odd harmonics, we would expect the transition between good and poor f_0 DL discrimination to occur at twice the average lowest harmonic number, i.e., around $\overline{N} = 20$. Thus, this result is consistent with the idea that pitch is derived from a combined "central spectrum" (Zurek, 1979) that prevents an independent pitch percept derived from the input to one ear. Note, however, that the odd and even harmonics were presented to left and right ears at random in each trial, making it impossible for the listener to know which ear to ignore. It is possible that if odd and even harmonics were presented consistently to the same ears, subjects may have been able to learn to ignore the input from the ear with odd harmonics.

The transition from small to large f_0 DLs occurs at the same $\overline{N}$ at both f_0s, consistent with the results of Kaernbach and Bering (2001). This confirms our expectation (Shackleton and Carlyon, 1994) that the dramatic increase in f_0 DLs is due to a relative frequency effect that depends more on harmonic number than on an absolute frequency effect, such as the roll-off of phase-locking with increasing absolute frequency. Nevertheless, effects of absolute frequency are also present, in that the f_0 DLs for $\overline{N} > 9$ are greater for the 200-Hz f_0 than for the 100-Hz f_0. These absolute frequency effects may be related to phase locking, where the additional information available from phase locking to the fine structure at a lower absolute frequency region in the 100-Hz condition aided f_0 discrimination. Also, because we tested only f_0s of 100 and 200 Hz, we did not observe the absolute frequency effects reported in other studies where the f_0 DL transition occurs at a lower $\overline{N}$ for f_0s below 100 Hz and above 200 Hz (Ritsma, 1962; Krumbholz *et al.*, 2000; Pressnitzer *et al.*, 2001).

For the diotic 200-Hz sine phase condition, f_0 DLs for complexes with $\overline{N} > 10$ are approximately twice as large as those of the monotic 200-Hz sine-phase results of Houtsma and Smurzynski (1990), depicted as a dashed line in the lower left panel of Fig. 2.4, although the transition from

small to large f_0 DLs occurs at the same $\overline{N}$ in both studies. The difference in DL between this and the earlier study can be probably explained in terms of differences in sensation level. Hoekstra (1979) showed that an increase in sensation level from the 10 dB used in our study to the 20 dB used in the Houtsma and Smurzynski (1990) study decreased f_0 DLs for harmonic complexes by a factor of two to four, depending on f_0 and subject.

2.5. Experiment 2C: Perceived pitch of dichotic stimuli

2.5.1. Rationale

Flanagan and Guttman (1960) investigated the pitch of same- and alternating-polarity click trains. A same-polarity click train has a click rate equal to the f_0, and a spectrum consisting of all the harmonics of f_0, whereas an alternating-polarity click train has a click rate that is twice the f_0 ($2f_0$), and a spectrum consisting of only the odd harmonics of the f_0. According to Flanagan and Guttman (1960), stimuli with $f_0 < 150$ Hz elicit a pitch corresponding to the click rate, regardless of polarity, while stimuli with $f_0 > 150$ Hz elicit a pitch corresponding to the f_0. This result is consistent with a two-mechanism model of pitch perception. Click trains with a high f_0 that contain resolved components in the absolute frequency dominance region for pitch (Plomp, 1967) yield a pitch at the f_0, consistent with a mechanism that extracts pitch from spectral cues. Click trains with a low f_0 that contain only unresolved components in the dominance region yield a pitch consistent with a mechanism that extracts pitch from peaks in the temporal envelope of the waveform. The temporal envelope of the alternating polarity click train repeats at the difference frequency between components of $2f_0$, whereas the waveform of the same polarity click train repeats at the f_0.

Experiment 2C estimated the perceived pitch of the dichotic stimuli used in experiment 2B. If, as suggested by the results of experiment 2B, the individual resolved components above the 10th harmonic are not used in f_0 discrimination, then the pitch of dichotic complexes with $\overline{N} > 10$ may be derived from the repetition rate of the temporal envelope. If so, these complexes should yield a perceived pitch at $2f_0$, consistent with the peripheral difference frequency between adjacent components. Alternatively, the central pitch mechanism may be able to make some, but poor, use of the higher-order resolved harmonics. If so, these dichotic stimuli should yield a pitch at

the f_0 derived from the combined central spectrum, but with the poor f_0 discrimination performance seen in experiment 2B.

2.5.2. Methods

Assuming that listeners would only perceive a pitch at the f_0 or at $2f_0$ for their alternating-phase stimuli, Shackleton and Carlyon (1994) asked listeners to identify which of two sine-phase stimuli, with fundamental frequencies equal to the f_0 or to $2f_0$ of the alternating-phase stimulus, most closely matched each alternating-phase stimulus. Similarly, we assumed that our dichotic stimuli would yield perceived pitches corresponding to either the f_0, consistent with spectral cues, or to $2f_0$, consistent with monaural temporal envelope cues. However, we used a different experimental paradigm. Subjects compared the pitch of a dichotic stimulus with that of a diotic stimulus, where the f_0 of the diotic stimulus was a half-octave (a factor of $\sqrt{2}$) higher than that of the dichotic stimulus. We assumed that the diotic sine-phase stimulus yielded a perceived pitch near its f_0. Thus, if the dichotic stimulus was judged as "higher" we assumed that the subject perceived its pitch to be $2f_0$. Similarly, if the dichotic stimulus was judged "lower", we assumed the subject perceived its pitch to be the f_0.

The dichotic stimuli were sine-phase complexes identical to those described in experiment 2B. The diotic stimuli were sine-phase harmonic complexes with f_0 half an octave above the f_0 of the dichotic stimulus in the same trial, with harmonics chosen such that the bandwidth was limited to that of the dichotic stimulus. The diotic and dichotic stimuli were presented randomly in the first

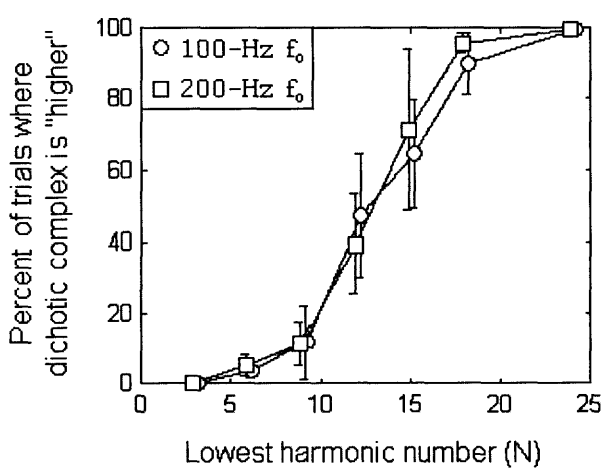

Figure 2.6. Mean results from experiment 2C, showing the percentage of trials where subjects reported a dichotic complex to have a higher pitch than a diotic complex with f_0 a factor of $\sqrt{2}$ higher. Error bars indicate plus or minus one standard error of the mean across the individual subjects. For lowest harmonic number N < 12, subjects nearly always identified the diotic complex as "higher"; for N > 15 subjects nearly always identified the dichotic complex as "higher". The transition from 0% to 100% occurs at approximately the same harmonic number for both f_0s tested.

and second intervals and the subject was asked to identify the "higher" interval. Lowest harmonic number was not roved from interval to interval. Each run consisted of seven trials for each of the seven average lowest harmonic numbers tested in experiment 2B, for a total of 49 trials. Twelve runs were performed at both the 100- and 200-Hz f_0s, such that each dichotic complex was presented a total 84 times per subject. To acquaint subjects with the task, they underwent a short (15 min) training session during which they were required to identify the higher of two pure tones separated by ½ octave. Four subjects (one female) took part in this experiment. Three had already participated in experiments 2A or 2B. The setup for stimulus delivery was identical to that described in experiment 2A.

2.5.3. Results

The results shown in Fig. 2.6 clearly indicate that subjects perceived a pitch lower than $\sqrt{2}$ times the f_0 for dichotic stimuli with a low lowest harmonic number and a pitch higher than $\sqrt{2}$ times the f_0 for dichotic stimuli with a high lowest harmonic number. The transition between the two pitch percepts occurred between lowest harmonic numbers 9 and 18, roughly the same region as was seen for the f_0 DL shift in experiment 2B. If our assumption that listeners always perceive a pitch corresponding to either the f_0 or $2f_0$ holds, then listeners are perceiving a pitch corresponding to the f_0 for complexes containing harmonics lower than the 10th, consistent with spectral cues, and a pitch corresponding to $2f_0$ for complexes containing only harmonics above the 15th, consistent with temporal envelope cues. In between, the pitch appears to be ambiguous. Further testing would be necessary to determine if this ambiguity reflects two simultaneous pitches (at the f_0 and at $2f_0$) deriving from different mechanisms.

2.5.4. Discussion

The values of N that yielded an ambiguous pitch in this experiment correspond well to the values of $\overline{N}$ that yielded elevated f_0 DLs under dichotic presentation in experiment 2B. This suggests that neither a mechanism that operates on resolved harmonics nor a mechanism that extracts the pitch from the temporal envelope responds well to dichotic stimuli in this region. Since approximately 20 harmonics are resolved under dichotic presentation (experiment 2A), listeners had difficulty extracting the f_0 from these high-order, but resolved harmonics. Since dichotic complexes have fewer components falling within an auditory filter, the resulting temporal

envelopes will be less modulated than the envelopes associated with the diotic stimuli, reducing the effectiveness of the envelope as a pitch cue.

The data show that listeners always perceived a pitch near the f_0 for $N < 12$. This result is in conflict with the results of Hall and Soderquist (1975) , where subjects reported two pitches, one at each f_0, when three successive components each of a 200-Hz and a 400-Hz f_0 were presented to opposite ears. The larger number of harmonics presented in the current study (six to each ear) may have encouraged the fusion of binaural information in processing pitch.

2.6. General discussion

2.6.1. Absolute or relative frequency?

The transitions from strong to weak pitch salience in experiment 2B, and from a perceived pitch of the f_0 to $2f_0$ in experiment 2C occur at approximately the same lowest harmonic number for both the 100-Hz and 200-Hz f_0s. These results are consistent with the idea that relative frequency relationships, such as those that govern harmonic resolvability, underlie the different pitch percepts associated with complexes containing low- and high-order harmonics (Houtsma and Smurzynski, 1990; Carlyon and Shackleton, 1994; Shackleton and Carlyon, 1994; Kaernbach and Bering, 2001). If the change in pitch salience were due to absolute frequency effects, as suggested by autocorrelation models (Cariani and Delgutte, 1996a; Carlyon, 1998), the transition should have occurred at the same absolute frequency, and not the same harmonic number, for the two f_0s.

2.6.2. Resolvability or harmonic number?

Taken together, the results from the experiments demonstrate an interrelationship between harmonic number, resolvability, and pitch. Specifically, the region around the 10th harmonic appears to be important in defining transitions in harmonic resolvability, f_0 *discrimination*, and pitch height, at least for the f_0 range between 100 and 200 Hz. First, experiment 2A showed that for diotic stimuli approximately the first 10 harmonics are resolved, while higher harmonics are unresolved. Second, consistent with Houtsma and Smurzynski (1990), experiment 2B showed that small f_0 DLs require the presence of harmonics below the 10th. Third, experiment 2C showed that a perceived pitch associated with the f_0 of the combined binaural spectrum requires

the presence of harmonics below the 10th. Taken together, these three observations are consistent with the idea that complexes containing resolved harmonics below the 10th yield fundamentally different pitch percepts from those containing only harmonics above the 10th.

Consistent with earlier data from two-tone complexes (Houtsma and Goldstein, 1972; Arehart and Burns, 1999), the interpretation that harmonic resolution *per se* is responsible for the changes in pitch perception is not supported by the comparison of the diotic and dichotic results in experiments 2B and 2C. The additional resolved harmonics in the dichotic case yield neither small f_0 DLs in experiment 2B, nor a pitch match consistent with extraction of cues from a centrally combined spectrum in experiment 2C, both of which would be expected if the shift from a salient spectral pitch to a weak temporal pitch were due to harmonics becoming unresolved. For example, although a dichotic stimulus with $N = 15$ contains many resolved components, it yields poor f_0 discrimination performance and an ambiguous pitch percept. Thus, harmonic number, regardless of resolvability, seems to underlie the changes in pitch perception.

2.6.3. Implications for pitch theories

"Harmonic template" pitch theories propose that a pitch detection mechanism codes the individual frequencies of the peripherally resolved partials and compares them to an internally stored template to derive a pitch at the f_0 (e.g. Goldstein, 1973; Terhardt, 1974,1979). The failure of these models to explain how periodicity information is extracted from complexes containing only high-order harmonics has driven an opposing view that f_0 extraction is performed by a single autocorrelation or similar mechanism that operates on all harmonics, regardless of resolvability (Licklider, 1951,1959; Meddis and Hewitt, 1991a,b; Meddis and O'Mard, 1997; de Cheveigné, 1998). Meddis and O'Mard (1997) have claimed that their model accounts for the different pitch percepts associated with resolved and unresolved harmonic complexes, due to the inherent differences in the result of the autocorrelation calculation for resolved versus unresolved harmonics, although the validity of this claim has been put into doubt by further analysis of their model (Carlyon, 1998). Alternatively, several studies have suggested that pitch may be processed via two different mechanisms, a harmonic template mechanism operating on resolved harmonics, and a separate mechanism operating on the temporal envelope

resulting from unresolved harmonics (Houtsma and Smurzynski, 1990; Carlyon and Shackleton, 1994; Shackleton and Carlyon, 1994; Steinschneider *et al.*, 1998; Grimault *et al.*, 2002).

The results of experiment 2B argue against a pitch processing mechanism that responds inherently differently to unresolved versus resolved harmonics. With such a mechanism, we would expect f_0 discrimination performance to improve when normally unresolved harmonics are artificially resolved under dichotic presentation, whereas experiment 2B showed that f_0 discrimination performance was the same or worse in the dichotic conditions. Therefore, any theory of pitch perception must account for relative frequency effects without relying on harmonic resolvability.

"Temporal" theories could account for this relative frequency effect if the autocorrelation in each channel were constrained to be sensitive to a limited range of periodicities relative to the inverse of the channel's CF, thereby limiting the range of harmonic numbers contributing to the pitch percept (Moore, 1982). This modification would also need to somehow account for a pitch derived from the temporal envelope for complexes containing only high-order components. If this requirement could be met, the modified theory would be consistent with the ambiguous pitch and elevated f_0 DLs seen for dichotic complexes with N=12 and N=15, which have relatively ineffective envelope cues (see Sec. 2.5.4.).

"Place" theories could account for this relative frequency effect if the templates that derive the pitch from low-order harmonics were constrained to consist of only those harmonics that are *normally* resolved. This is consistent both with the idea of harmonic templates learned from exposure to harmonic sounds (Terhardt, 1974) and the more recent proposal that templates for low-order harmonics may emerge from any form of wideband stimulation (Shamma and Klein, 2000). With this constraint, even though artificially resolved harmonics (above the 10th and up to the 20th partial) are available under dichotic presentation, the pitch processing mechanism will be unable to utilize these additional resolved harmonics since no template will have developed to match them.

Even with this constraint, "harmonic template" theories do not fully explain the results for dichotic complexes containing artificially resolved harmonics. For these stimuli, we would expect that the even harmonics in one ear would match a template corresponding to $2f_0$, yielding f_0 DLs on the order of those measured for complexes containing low-order harmonics. While ambiguous pitch matches suggest that listeners may sometimes perceive a pitch corresponding to $2f_0$ for these dichotic complexes, f_0 DLs are *larger* than those for diotic complexes with the same N. Apparently, the presence of the odd harmonics in the opposite ear has a substantial detrimental effect on f_0 discrimination.

One possible explanation for these results postulates the existence of inhibitory inputs to harmonic templates, tuned to partials of subharmonics of the f_0. Under normal circumstances, where all harmonics of a complex are present, such inhibition might be useful in preventing erroneous pitch percepts at multiples of the f_0. According to this scheme, while the resolved $(mn)^{th}$ partials of a complex (where m and n are integers) would facilitate a template for a pitch corresponding to n times the f_0 (nf_0) of the complex, the remaining resolved partials of the complex would inhibit this template. Thus, only the template for a pitch at the f_0 would respond to the stimulus, yielding a pitch percept corresponding to the f_0 and good f_0 discrimination. For dichotic complexes with $N > 10$, templates for pitches corresponding to nf_0 would still be inhibited, but in this case the template for a pitch corresponding to the f_0, with a limited number of harmonics represented, would not respond to the high-order harmonics. With no template available, the pitch could only be derived from temporal cues.

Another interpretation of the results is that the pitch is extracted from a combined "central spectrum" representation (Zurek, 1979) that prevents an independent pitch percept derived from the input to one ear. The additional *peripherally* resolved components might not available in the central spectrum representation used to derive pitch. Listeners may have been able to overcome this central fusion in hearing out individual harmonics in experiment 2A, but not when deriving a pitch from the sum of components in experiments 2B and 2C. The non-monotonic psychometric functions seen in some subjects in experiment 2A may reflect an inability to overcome the binaural fusion even in the "hearing out" task.

2.7. Summary and conclusions

In experiment 2A approximately twice as many harmonics are resolved under dichotic as compared to diotic presentation, verifying that harmonic resolvability is not limited by binaural interactions. A direct estimate of the limits of harmonic resolvability indicated that approximately 9 and 11 harmonics are resolved for 100- and 200-Hz f_0s, respectively. The results from our direct measure, which minimizes non-peripheral limitations by gating the target component on and off, resolve the discrepancy between previous direct estimates that only 5 to 8 harmonics are resolved (Plomp, 1964), and indirect estimates suggesting that approximately 10 harmonics are resolved (Houtsma and Smurzynski, 1990; Shackleton and Carlyon, 1994).

In experiment 2B, listeners were unable to utilize the additional resolved harmonics available under dichotic presentation for f_0 discrimination. This implies that the deterioration in f_0 DLs with increasing lower cutoff frequency is due not to harmonics becoming unresolved *per se*, but instead to the increasing lowest harmonic number, regardless of resolvability. This result suggests constraints to both "place and "temporal" models of pitch perception. For a "harmonic template" theory to account for the data, only those harmonics that are *normally* resolved should be represented in the templates. For an "autocorrelation" theory to do so, the range of periodicities to which the autocorrelation in each channel is sensitive should be CF-dependent (Moore, 1982). Moore's (1982) suggestion will be explore in an analysis of the ability of the Meddis and O'Mard (1997) autocorrelation model to account for the dependence of f_0 discrimination on harmonic number in Chapter 3.

The results of experiment 2B and experiment 2C are consistent with a two-mechanism model of pitch perception (e.g. Carlyon and Shackleton, 1994). When harmonics below the 10th are present, a harmonic template mechanism is able to extract pitch from the resolved components, yielding small f_0 DLs and a pitch consistent with spectral cues. When only harmonics above the 10th are present, the auditory system relies on temporal envelope cues for pitch extraction, regardless of resolvability, yielding some ambiguous pitch percepts for dichotic complexes, and poor f_0 discrimination performance in all cases. A temporal model, constrained as described above, may nevertheless be able to account for these results within the framework of a single autocorrelation mechanism.

2.8 Segue

The results of the three experiments presented in this chapter raise the possibility that harmonic resolvability *per se* is not responsible for the observed transition from small to large f_0 DLs with increasing harmonic number. Chapter 3 investigated whether a temporally-based autocorrelation model of pitch perception can account for the dependence of f_0 DLs on harmonic number in such a way that does not rely on peripheral resolvability. The Meddis and O'Mard (1997) autocorrelation model was modified to include such a harmonic dependence by limiting the range of periodicities to which a given channel can respond relative to that channel's characteristic frequency (CF).

Chapter 3. An autocorrelation model with place dependence

This work described in this chapter is published in the *Journal of the Acoustical Society of America*.

Bernstein, J.G.W. and Oxenham, A.J. (2005a). An autocorrelation model with place dependence to account for the effect of harmonic number on fundamental frequency discimnination. *J. Acoust. Soc. Am.* **117**, 3816-3831.

3.1 Abstract

Fundamental frequency (f_0) difference limens (DLs) were measured as a function of f_0 for sine- and random-phase harmonic complexes, bandpass filtered with 3-dB cutoff frequencies of 2.5 and 3.5 kHz (low region) or 5 and 7 kHz (high region), and presented at an average 15 dB sensation level (approximately 48 dB SPL) per component in a wideband background noise. Fundamental frequencies ranged from 50 to 300 Hz and 100 to 600 Hz in the low- and high-spectral regions, respectively. In each spectral region, f_0 DLs improved dramatically with increasing f_0 as approximately the tenth harmonic appeared in the passband. Generally, f_0 DLs for complexes with similar harmonic numbers were similar in the two spectral regions. The dependence of f_0 discrimination on harmonic number presents a significant challenge to autocorrelation (AC) models of pitch, in which predictions generally depend more on spectral region than harmonic number. A modification involving a "lag window" is proposed and tested, restricting the AC representation to a limited range of lags relative to each channel's characteristic frequency (CF). This modified unitary pitch model was able to account for the dependence of f_0 DLs on harmonic number, although this correct behavior was not based on peripheral harmonic resolvability.

3.2 Introduction

Psychophysical studies of the pitch of harmonic tone complexes have demonstrated a relationship between the ability to discriminate small differences in fundamental frequency (f_0), and the harmonic numbers presented, i.e. the ratios between the frequencies of the individual harmonic components and f_0 of the stimulus (Chapter 2; Houtsma and Goldstein, 1972; Houtsma and Smurzynski, 1990; Shackleton and Carlyon, 1994). Harmonic complexes containing components with frequencies less than 10 times the f_0, i.e. harmonic numbers below the 10th,

generally yield good f_0 discrimination performance, while those containing only harmonics above the 10th yield poorer f_0 discrimination performance, at least for f_0's in the 100- to 200-Hz range. The different f_0 discrimination results yielded by low- and high-order harmonics has traditionally been explained in terms of harmonic resolvability (Carlyon and Shackleton, 1994; Shackleton and Carlyon, 1994). The individual frequency components of a harmonic complex are spaced linearly in frequency, while auditory filter bandwidths are approximately proportional to the filters' characteristic frequencies (CFs). The frequency spacing between low-order harmonics will be wider than the bandwidths of the auditory filters they excite. As a result, one low-order harmonic will dominate the output of a single auditory filter, and will therefore be resolved by the auditory system. Conversely, multiple high-order harmonics fall within a single auditory filter and will therefore be unresolved by the auditory system. To estimate the f_0, the individual frequencies of resolved low-order resolved components, derived from either rate-place or temporal phase-locking cues, could be compared to an internally stored harmonic template (Goldstein, 1973; Wightman, 1973; Terhardt, 1974,1979; Srulovicz and Goldstein, 1983). A separate temporal mechanism could estimate the f_0 for unresolved harmonics, by acting on the temporal envelope, resulting from the interaction of several components within a single auditory filter, which has a periodicity corresponding to the f_0 (Moore, 1977; Shackleton and Carlyon, 1994; Cariani and Delgutte, 1996a).

Certain results in the literature have provided evidence that f_0 discrimination performance is related to harmonic resolvability. One such result concerns the effect of phase on f_0 discrimination. Houtsma and Smurzynski (1990) showed that both the magnitude and the phase-dependency of f_0 DLs varied with harmonic number in the same way. Complexes containing low-order harmonics yielded small f_0 DLs that were not affected by the relative phase of the individual harmonics, whereas complexes containing only high-order components yielded large f_0 DLs that were phase-dependent. The phase relationship between harmonics should only affect f_0 discrimination if the harmonics are unresolved and interact within a single auditory filter (Moore, 1977; Shackleton and Carlyon, 1994). Therefore, the co-occurrence of large and phase-dependent f_0 DLs suggests that f_0 discrimination performance also depends on harmonic resolvability. Another important result concerns the ability to hear out the frequency of an individual harmonic of a complex, which is a more direct estimate of harmonic resolvability.

Chapter 2 found that f_0 DLs showed the same dependence on harmonic number as listeners' abilities to hear out harmonic frequencies. Below about the 10th harmonic, f_0 DLs were small and the frequency of an individual could be heard out. Above the 10th harmonic, f_0 DLs became large, and individual component frequencies were no longer discriminable from nearby pure-tone frequencies

These studies have shown that f_0 discrimination performance has the same dependence on harmonic number as two different measures that clearly depend on harmonic resolvability. Nevertheless, this is not conclusive evidence that f_0 discrimination is directly dependent on harmonic resolvability. In fact, several results in the literature suggest that f_0 discrimination performance does not depend on harmonic resolvability *per se*. Chapter 2 showed that the dichotic presentation of harmonic complexes, where even and odd components were presented to opposite ears, did not increase the harmonic number of the transition between good and poor f_0 discrimination, even though twice as many peripherally resolved components were available. Similar results were shown with the dichotic presentation of two-tone complexes in normal-hearing (Houtsma and Goldstein, 1972) and hearing-impaired listeners (Arehart and Burns, 1999). These results raise the possibility that the correlation between the dependencies of f_0 DLs and harmonic resolvability on harmonic number is epiphenomenal and not causal.

As an alternative to the harmonic template theories described above, pitch could be derived from a single temporal mechanism that acts on timing information from all frequency channels, regardless of resolvability (Licklider, 1951; Meddis and Hewitt, 1991a,b; Cariani and Delgutte, 1996a; Meddis and O'Mard, 1997; de Cheveigné, 1998). A recent implementation of these timing-based models is the Meddis and O'Mard (1997) unitary autocorrelation (AC) model of pitch perception. The Meddis and O'Mard model performs an AC of the probability of firing as a function of time in each simulated auditory nerve fiber (ANF). These individual autocorrelation functions are then summed across all frequency channels to produce a summary autocorrelation function (SACF). The AC in each channel contains peaks at a period equal to the inverse of the f_0 whether it responds to the envelope of the waveform of several interacting components or to an individual resolved frequency component at a multiple of the f_0. Therefore, the SACF will contain a large peak at the inverse of the f_0, allowing the extraction of the f_0. This

mathematical formulation is analogous to calculating the all-order interval histogram based on spike times in the auditory nerve (Cariani and Delgutte, 1996a). Meddis and O'Mard (1997) have argued that this autocorrelation model can account for the effect of harmonic number on f_0 discrimination indicated by the psychophysical results of Shackleton and Carlyon (1994). They claimed that the AC responds inherently differently to resolved and unresolved harmonics, yielding the requisite f_0 discrimination behavior.

However, Carlyon (1998) disputed this claim, suggesting that any deterioration in f_0 discrimination seen in the AC model is a function of the roll-off of phase locking with absolute frequency, as was seen in the physiological recordings of Cariani and Delgutte (1996a), and not a function of harmonic number as seen in psychophysical studies (Shackleton and Carlyon). According to Carlyon (1998), the most important shortcoming of the Meddis and O'Mard (1997) AC model is that it fails to predict the effect of harmonic number on f_0 discrimination seen in the psychophysics: two harmonic complexes with different f_0's, band-pass filtered in the same spectral region, yielded very different f_0 discrimination performance when one complex contains low-order harmonics and the other does not (Shackleton and Carlyon, 1994).

The present study addressed this controversy. The Meddis and O'Mard (1997) unitary AC model of pitch perception was tested for its ability to account for the effects of harmonic number on f_0 discrimination. A psychophysical experiment measuring f_0 DLs as a function of f_0 for fixed spectral regions was performed in order to provide more data points than the six (two f_0's times three spectral regions) tested by Shackleton and Carlyon (1994). The same stimuli were then passed through the Meddis and O'Mard (1997) AC model to determine its ability to predict the experimental results. Overall, the AC model failed to predict the experimental results. Whereas the experimental results (described in Section 3.3), showed decreasing f_0 DLs with increasing f_0, the model predictions (described in Section 3.4) showed the opposite trend. A number of possible modifications to the model were then tested. Of these, the most successful was one similar to that suggested by Moore (Moore, 1982), in which place dependency is introduced into the model, such that each frequency channel responded only to a limited range of periodicities related to the channel's CF.

3.3. Experiment 3: f_0 DLs with a fixed spectral envelope

3.3.1. Rationale

This experiment was intended to provide a larger data set than that provided by Shackleton and Carlyon (1994) with which to test the ability of the AC model to account for the effects of harmonic number on f_0 discrimination. This experiment also addressed two issues surrounding the mechanisms underlying pitch processing: the roles of phase and temporal fine-structure in f_0 discrimination.

(3.3.1.1) Phase

Previous results have shown that the phase relationship between harmonics affects f_0 DLs (Houtsma and Smurzynski, 1990): harmonic stimuli with "peakier" waveforms, such as sine- or cosine-phase complexes, yield smaller f_0 DLs than those with "flatter" waveforms, such as random- or negative Schroeder-phase (Schroeder, 1970) complexes. However, this phase effect was not apparent in the results of Chapter 2. There are two possible reasons for this discrepancy. First, different groups of listeners were tested for the two phase relationships (random- and sine-phase) in Chapter 2, yielding an analysis of variance (ANOVA) with less statistical power than would be expected if the same subjects had been tested for both phase relationships. Second, as discussed in Chapter 2, listeners may have performed the f_0 discrimination task without extracting the f_0, by listening for a change in the frequency of the lowest harmonic present. Although the lowest harmonic number presented was randomized from interval to interval, a large enough change in f_0 would overcome this small amount of randomization. Data analysis showed that for complexes containing only high-order harmonics, f_0 DLs were large enough that subjects may have been using the lowest harmonic cue, rather than f_0 cues, to perform the task, especially for the random-phase stimuli. If subjects were not using f_0 to perform the task, then the effects of phase on f_0 extraction would not be apparent in the results.

To address the possibility that the lack of a significant phase effect resulted from different groups of listeners participating in two phase conditions, all subjects in the present study participated in both the sine-phase and random-phase conditions. To address the possibility that listeners had used the frequency of the lowest harmonic rather than f_0 cues to perform the f_0 discrimination task, the experiment described below attempted to eliminate lowest harmonic cues by using

harmonic stimuli with a fixed spectral envelope, and measuring f_0 DLs as a function of f_0. Although the frequency of the lowest harmonic increases with increasing f_0, a lower-numbered harmonic will also begin to appear at the low end of the spectrum. Thus, the cochlear excitation pattern will remain roughly constant, at least for those complexes containing only unresolved harmonics where the lowest harmonic cue may have played a role. As f_0 increases, the lowest harmonic number present in the pass-band decreases, allowing a direct comparison with the f_0 DL measurements of Chapter 2. Results indicating larger f_0 DLs in this experiment would indicate that subjects may have been using the lowest harmonic cue in the previous study.

(3.3.1.2) Temporal fine structure

The effects of phase on f_0 discrimination have provided evidence that the pitch of complexes containing unresolved harmonics is derived from the repetition rate of peaks in the temporal envelope. Negative Schroeder-phase complexes, which have flatter envelopes than sine-phase complexes, yield larger f_0 DLs (Houtsma and Smurzynski, 1990). When unresolved harmonics are presented in alternating sine- and cosine- phase, yielding temporal envelopes with two peaks per period, the resulting pitch percept is judged to be twice the f_0 (Shackleton and Carlyon, 1994). Still, this does not rule out that periodicity information could be extracted from the fine structure of unresolved harmonic complexes in some conditions. Hall *et al.* (2003) demonstrated that phase manipulations affected amplitude-modulation (AM) rate discrimination performance for unresolved components in a high spectral region, but had little effect in a relatively low spectral region. Their interpretation was that fine-structure cues, which are unaffected by phase manipulations, are used in the low-frequency region, while envelope cues, which are affected by phase manipulations, are used in the high-frequency region where there is little phase-locking to the fine-structure. Similarly, Chapter 2 found that for unresolved complexes containing the same harmonic numbers, f_0 DLs were larger for a 200- than a 100-Hz f_0, which may reflect reduced fine-structure information in the higher spectral region occupied by the 200-Hz complexes. Furthermore, deterioration in phase-locking to the frequencies of individual partials could affect f_0 DLs for complexes containing resolved harmonics.

This experiment tested whether the presence of phase-locking to the fine-structure in the low region aided performance, in a task more closely related to pitch processing than the AM rate

discrimination task of Hall *et al.* (2003). Fundamental frequency DLs were measured in two conditions: a "low spectrum" condition (2.5-3.5 kHz), in which phase-locking to fine structure is thought to be more available, and a "high spectrum" condition (5-7 kHz), in which phase-locking to the fine-structure information is greatly reduced, at least in mammalian species that have been tested so far (Rose *et al.*, 1968; Johnson, 1980; Palmer and Russell, 1986; Weiss and Rose, 1988b). Testing f_0 DLs in two different frequency regions also provided a control to verify that f_0 discrimination performance depends primarily on harmonic number, and not f_0 *per se*.

3.3.2 Methods

Five subjects participated in the experiment (ages 18-21, 3 female). All subjects had normal hearing (15 dB HL or less re ANSI-1969 at octave frequencies between 250 Hz and 8 kHz) and were self-described amateur musicians with at least 5 years of experience singing or playing a musical instrument.

All stimuli were presented in modified uniform masking noise (UMN_m; see Chapter 2). This noise is similar to uniform masking noise (UMN, Schmidt and Zwicker, 1991), in that it is intended to yield pure-tone masked thresholds at a constant sound pressure level (SPL) across frequency, but the spectrum is somewhat different; UMN_m has a long-term spectrum level that is flat (15 dB/Hz SPL in our study) for frequencies below 600 Hz, and rolls off at 2 dB/octave above 600 Hz. The noise was low-pass filtered with a cutoff at 16 kHz. Thresholds for pure tones at 200, 500, 1500, and 4000 Hz in UMN_m in the left ear were estimated via a three-alternative forced-choice, 2-down, 1-up adaptive algorithm (Levitt, 1971). For each subject, pure tone thresholds in UMN_m fell within a 5 dB range at all four frequencies tested, such that harmonic components presented at equal SPL had nearly equal sensation level (SL). As an approximation, we defined 0 dB SL for each subject as the highest of the thresholds across the four frequencies tested, which ranged from 31 to 33.3 dB SPL across all subjects.

The stimuli were generated digitally and played out via a soundcard (LynxStudio LynxOne) with 24-bit resolution and a sampling frequency of 32 kHz. The stimuli were then passed through a programmable attenuator (TDT PA4) and headphone buffer (TDT HB6) before being presented

to the subject via Sennheiser HD 580 headphones. Subjects were seated in a double-walled sound-attenuating chamber.

Fundamental frequency DLs as a function of a complex's f_0 were estimated via a three-alternative forced-choice, 2-down, 1-up adaptive algorithm tracking the 70.7% correct point (Levitt, 1971). The f_0 difference (Δf_0) was initially set to 10% of the f_0. The starting step size was 2% of the f_0, decreasing to 0.5% after the first two reversals, and then to 0.2% after the next two reversals. The f_0 DL was estimated as the average of the Δf_0s at the remaining six reversal points. If the standard deviation of the last six reversal points was greater than 0.8%, the data were rejected and the run repeated. In each trial, two of the three intervals contained harmonic complexes with a base f_0 ($f_{0,base}$), while the other interval contained a complex with a higher f_0 ($f_{0,base} + \Delta f_0$). Subjects were informed that two of the intervals had the same pitch, while the third interval had a higher pitch, and were asked to identify the interval with the higher pitch. DLs were estimated for six different f_0's in each spectral condition (low: 50, 75, 100, 150, 200 and 300 Hz; high: 100, 150, 200, 300, 400 and 600 Hz). The f_0's tested in the high condition were double those tested in the low condition such that the harmonic numbers presented were the same in each spectral region. Measurements were repeated four times per subject for each combination of frequency region, phase and f_0, except for one subject who completed only two runs for the random-phase conditions.

Stimuli were re-synthesized for each trial of the experiment. First, diotic harmonic complexes containing equal-amplitude harmonics of the f_0 up to 10 kHz were synthesized. These harmonic complexes were then filtered with both 4th-order low-pass and 4th-order high-pass digital Butterworth filters. The 3-dB filter cutoff frequencies for the high- and low-pass filters, respectively, were 2.5 and 3.5 kHz in the low condition, and 5 and 7 kHz in the high condition. The filter weights for the high-pass filters were scaled such that the double filtering operation gave a 0 dB maximum amplitude response. The duration of the stimulus in each trial of the experiment was 500 ms, including 30-ms Hanning window onset and offset ramps.

Following the filtering operations, the stimulus in the interval with the higher f_0 was scaled in amplitude to have equal root-mean-square (rms) power to that of the two other intervals with the

base f_0. The complexes were presented at an average level per component (before filtering) of 15 dB SL per component (adjusted individually based on tone-in-noise detection thresholds). In order to prevent the use of loudness cues, amplitude randomization was applied by roving the amplitude of the complex in each interval by ±5 dB, uniformly distributed. On the average, the following -15 dB (re max) frequency bands contained harmonics above threshold: 1.56-5.35 kHz and 3.28-9.37 kHz in the low and high spectral conditions, respectively.

The resulting signals were then added to the UMN_m noise. Because of the rms normalization step, the average presentation level *per harmonic* was somewhat higher for the interval with ($f_{0,\text{base}} + \Delta f_0$) then for the intervals with $f_{0,\text{base}}$. However, this difference was quite small relative to the 10-dB random amplitude variation, reaching only about 0.6 dB for the largest measured f_0 DL of 15%. Complexes were presented in either sine or random phase. For the random-phase stimuli, the phase of each harmonic was newly chosen from a uniform random distribution ranging from $-\pi$ to $+\pi$ in each interval of the experiment.

3.3.3. Results and discussion

For each frequency region condition and f_0, the lowest detectable harmonic number (N) was estimated by dividing the average lowest detectable frequency in the pass-band (1.56 and 3.28 kHz in the low and high conditions, respectively) by the f_0. Figure 3.1a shows the estimated f_0 DLs as a function of N. The corresponding f_0's in the low- and high-spectrum conditions are shown along the top axis. Figure 3.1b shows the f_0 DLs predicted by the autocorrelation model, which will be discussed in Section 3.4 below. The main findings of this experiment are (i) f_0 DLs increase with increasing N (decreasing f_0), independent of spectral region, (ii) the relative phase relationship between partials affected f_0 DLs for high, but not low N, and (iii) there was a small but significant effect of spectral region on f_0 DLs. Each of these effects will be discussed in turn.

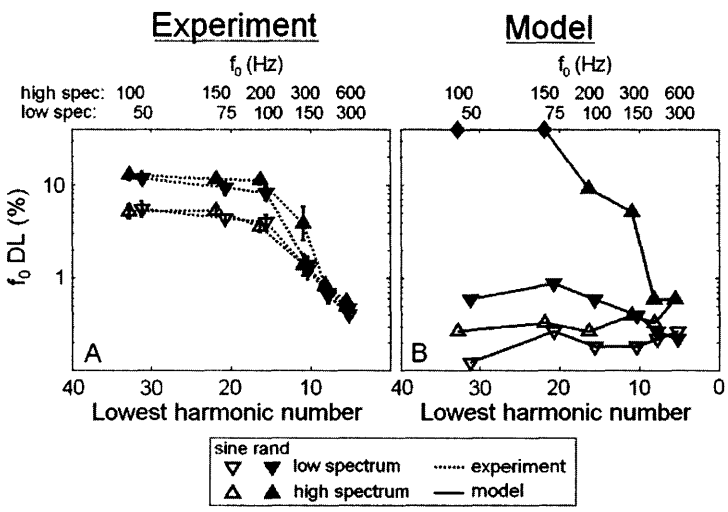

Figure 3.1. Fundamental frequency DLs (a) measured psychophysically and (b) predicted by the optimal detector autocorrelation model, as a function of the lowest harmonic number present within the passband. Stimulus f_0's corresponding to the lowest harmonic numbers are listed at the top of the figure. Optimal model predictions in (b) are calculated as the minimum value of δ such that d' exceeds the value of d_0'=190 depicted in Fig. 3.3. Diamonds (♦) plotted along the top horizontal axis indicate that d_0' was not reached at the maximum tested value of δ=0.3.

A repeated-measures analysis of variance (RMANOVA) with three within-subject factors (spectral region, phase and N) was conducted in order to determine the influence of each factor on f_0 discrimination. Values of $p<0.05$ were taken to be statistically significant. The RMANOVA was performed with logarithmically transformed data in an attempt to satisfy the equal-variance assumption, and the Greenhouse-Geisser (Geisser and Greeenhouse, 1958) correction for sphericity was included wherever necessary, with corrected values for degrees of freedom reported. However, neither manipulation affected the statistical significance of any main effect or interaction. Data from the subject who completed only two runs in the random-phase conditions were excluded from the RMANOVA. While the six f_0's tested in the high-spectrum conditions were exactly double those tested in the low-spectrum condition, the low edge frequency in the high-spectrum conditions was not exactly double that of the low-spectrum conditions. As a result, N's differed by approximately 5% in the two spectral conditions. Nevertheless, for the purpose of performing the RMANOVA, we assumed that the N's were identical. For example a 100-Hz low-spectrum stimulus was assumed to have the same N as a 200-Hz high-spectrum stimulus. This small 5% shift in the value of N was unlikely to affect the RMANOVA results. The results of this analysis are shown in Table 3.1.

Table 3.1. Results of the RMANOVA for the f_0 DL experiment. Asterisks indicate statistical significance (p<0.05). Degrees of freedom are adjusted based on the Geisser-Greenhouse correction.

	Effect	F	df	p
Main Effects	*N*	161	(2.12, 6.37)	<0.0005*
	phase	4180	(1, 3)	<0.0005*
	spectral region	12.9	(1, 3)	0.037*
2-way Interactions	*N* * spectral region	0.827	(1.45, 4.35)	0.827
	N * phase	25.1	(2.37, 7.11)	<0.0005*
	phase * spectral region	0.144	(1, 3)	0.318
3-way Interaction	*N* * phase * spectral region	0.226	(1.87, 5.59)	0.15

There is clear transition from large to small f_0 DLs as f_0 increases (*N* decreases) in both the low- and high-spectrum conditions. The dependence of f_0 DLs on *N* is supported by a significant main effect of *N*. The transition to small f_0 DLs occurs as the approximately 10th harmonic (the highest resolved harmonic as estimated in Chapter 2) begins to appear at the low end of the pass-band, consistent with previous results (Chapter 2; Houtsma and Smurzynski, 1990). When plotted as a function of *N*, the low- and high-spectrum data overlap, indicating that f_0 DLs in these conditions depend mainly on harmonic number and not on f_0 or spectral region. This conclusion is supported by the fact that there was no significant interaction between spectral region and *N*.

Phase effects are apparent in these results, but only for those complexes with *N*>10, where random-phase f_0 DLs are larger than sine-phase f_0 DLs, consistent with previous findings (Chapter 2; Houtsma and Smurzynski, 1990). The significant main effect of phase and a significant interaction between phase and *N* is consistent with the observation that phase effects were only observed for stimuli with high *N*. For low f_0's (*N*>10), the random-phase relationship of the harmonics gives f_0 DLs of 11-13%, which are much poorer than had been previously measured for random-phase complexes containing only high-order harmonics (Chapter 2). This result indicates that the previous estimates of f_0 DLs in the 6-8% range for high-order, random-phase complexes likely reflected the influence of the "lowest harmonic present" cue (see Chapter 2). The relatively small f_0 DLs (~4-6%) measured for the sine-phase, high-order complexes were approximately the same as those measured in Chapter 2, suggesting that the lowest harmonic cue

did not play a role in the sine-phase conditions. With the elimination of the confounding "lowest harmonic" cue that affected random-phase but not sine-phase f_0 DLs, the effects of phase on f_0 discrimination are found to be significant, in line with Houtsma and Smurzynski (1990). The large f_0 DLs would make music perception based on unresolved complexes difficult, since musical semitones are only 6% apart in frequency.

While there was a significant main effect of spectral region, f_0 DLs for the same N did not generally appear to be different between the low- and high-spectrum conditions, with one exception: performance was notably worse for the high-spectrum stimulus in the random-phase, $N \approx 10$ case. This difference was only observed for two of the five subjects, one of whom showed very large variability across runs, and does not constitute a general trend in the data. Although there was neither a significant two-way interaction between spectral region and either N or phase, nor a significant three-way interaction, the main effect of spectral region disappeared when the $N \approx 10$ data were excluded from the RMANOVA analysis [$F(1,3) = 4.8$, $p=0.12$]. This implies that phase locking to the stimulus fine structure did not play a significant role overall in f_0 discrimination for the stimuli used in this experiment.

The lack of a main effect of spectral region or a significant interaction between N and spectral region conflicts with the results of Hoekstra (1979), who also measured f_0 DLs as a function of f_0 for bandpass-filtered harmonic complexes in various spectral regions. Comparing similar spectral regions to those used in the current experiment, Hoekstra found that f_0 DLs were larger at higher spectral regions for complexes with small N, but not large N, suggesting that phase-locking to the stimulus fine-structure is more important for low-order, resolved harmonics. The discrepancy between the results of Hoekstra (1979) and the current study may be related to the bandwidths of the spectral regions used in the two studies. Hoekstra's 1/3-octave filters yielded only one audible partial for those stimuli with a low enough N to yield small f_0 DLs, while the approximately one-octave filters used in the current study produced multiple audible partials for all stimuli. The different results obtained in the two studies suggest that phase-locking to the stimulus fine structure may be more important for pure-tone frequency discrimination than for complex-tone f_0 discrimination. Alternatively, it may be that temporal fine-structure information is important for complex-tone f_0 discrimination, but that a large effect of spectral region was not

observed in the present study because of the frequency ranges chosen for the two spectral conditions. The 3-dB bandpass-filter cutoff frequencies were chosen such that phase-locking should have been greatly reduced in the high-spectrum condition relative to the low-spectrum condition. However, the filter slopes yielded an audible frequency range in the high-spectrum condition that extended down to 3.28 kHz, where phase-locking to the stimulus fine structure might still have been available.

3.4 Simulations with the autocorrelation model

3.4.1 Introduction

Meddis and O'Mard (1997) showed that the autocorrelation model successfully accounted for the results of Houtsma and Smurzynski (1990): for stimuli with a fixed f_0, f_0 DLs increased as the order of the harmonics increased. Carlyon (1998) suggested that the model's successful prediction was due not to its dependence on harmonic number and harmonic resolvability, but to the reduction of phase-locking with increasing absolute frequency. Because Houtsma and Smurzynski (1990) tested only one stimulus f_0 of 200 Hz, it was not clear from their results whether the increase in f_0 DLs was due to effects of harmonic number and resolvability, or to effects of spectral region. Consistent with earlier studies (Chapter 2; Shackleton and Carlyon, 1994; Kaernbach and Bering, 2001), the experiment described above, which measured f_0 DLs in two different spectral regions, demonstrated that f_0 discrimination performance depended mainly on harmonic number, and not spectral region or f_0. These data provide a basis for testing the Meddis and O'Mard (1997) autocorrelation model to determine its ability to predict the dependence of f_0 discrimination on harmonic number.

3.4.2 Model description

The stimuli from the above psychophysical experiment were passed through the Meddis and O'Mard (1997) autocorrelation model to determine its ability to account for the psychophysical f_0 discrimination results. This model consists of an outer/middle ear bandpass filter, a basilar membrane gammatone filterbank (Patterson *et al.*, 1992), inner hair cell half-wave rectification and low-pass filtering, and the translation of the inner hair cell membrane potential into a probability of firing versus time in the auditory nerve fiber. The model used to generate ANF firing information in these simulations was identical to that used by Meddis and O'Mard (1997),

except for the following two changes. First, forty channels, consisting of only those CFs falling within the stimulus passband (1.5-5 kHz and 3-10 kHz for the low- and high-spectrum conditions, respectively) were used, with CFs spaced according to the Greenwood's (1961) human scale. CFs falling outside these ranges, where the harmonic complex stimuli would not be detectable in the psychophysical experiment, were not included. Second, the inner hair-cell and auditory nerve models were replaced by a newer model (Sumner *et al.*, 2002) that allowed for stochastic spike generation. All ANFs were modeled as high-spontaneous rate fibers. The bandwidths of the model's gammatone filters were derived from the equivalent rectangular bandwidth (ERB_N) formula described by Glasberg and Moore (1990), just as in the Meddis and O'Mard study. Because the only physiologically-derived cochlear mechanical filtering data available for humans (Shera *et al.*, 2002) are only appropriate for very low-level stimuli, the psychophysically bandwidths derived by Glasberg and Moore (1990) form a reasonable substitute.

Two different methods for converting from ANF firing to a psychophysical f_0 DL estimate were tested. The first method was that used by Meddis and O'Mard (1997), whereby discriminability was estimated by the Euclidean distance (D) between autocorrelation functions (ACFs) calculated from the ANF probabilities of firing as a function of time, $p(t)$. The second method was an optimal detector model based on stochastic firing of the ANFs. Both methods are described below.

(3.4.2.1) Euclidean distance measure

Meddis and O'Mard's (1997) procedure for estimating discrimination thresholds was also used here. The main difference was that whereas Meddis and O'Mard based all of their computations on $p(t,k)$, the probability of firing (p) as a function of time (t) for each ANF channel index (k), the current simulations were based on stochastic ANF responses. This allowed for the possible influence of ANF refractoriness on the results. The inner hair cell / auditory nerve complex was set to "spike" mode (Sumner *et al.*, 2002), yielding stochastic boolean responses $s(t,k)$, whereby a one or a zero represented the presence or absence of a spike at each point in time. Each stimulus was re-synthesized and presented to the model n=15 times (although n was increased to

60 and 100 for the simulations described below in Section 3.5.2 and 3.5.5, respectively) and $p(t,k)$ were estimated by averaging across the n outputs $s(t,k)$ obtained for each k.

The autocorrelation function (ACF) of $p(t,k)$ was then calculated in each fiber according to the formulation of Meddis and O'Mard:

$$h(t_0,l,k) = \frac{1}{\tau}\sum_{i=1}^{\infty} p(t_0 - T,k)\,p(t_0 - T - l,k)e^{-T/\tau}dt \qquad (3.1)$$

where $h(t,l,k)$ is the channel's ACF, t_0 is the point in time at which the autocorrelation was measured, l is the autocorrelation lag, τ is the autocorrelation time constant, dt is the sampling interval, 25 μs, and $T = idt$. Because of the exponential window used in the ACF formulation, the autocorrelation will tend to fluctuate with time. In these simulations, t_0 was chosen to be an integer number of periods of each stimulus, just before the beginning of the offset ramp. This is in contrast to the Meddis and O'Mard study, where a "snapshot" of the SACF was taken at the end of the stimulus. The only other difference in the autocorrelation calculation in this study as compared to Meddis and O'Mard (1997) was that here τ was selected to be 25 ms, whereas Meddis and O'Mard used a shorter τ of 10 ms. The τ used in the current study, being longer than the period corresponding to the minimum f_0 tested, 50 Hz, tended to smooth out the SACF variation across time. A summary autocorrelation function, $SACF(f_0,l)$, was computed by summing the individual channel ACFs. The range of lags was fixed throughout the modeling from zero to a maximum lag (l_{max}) of 25 ms. This value of l_{max} corresponds to a minimum frequency of 40 Hz, which is below the minimum f_0 of 50 Hz used in the psychophysical experiment described above.

For each combination of f_0, spectral region, and phase, ACFs and SACFs were calculated for stimuli with f_0 increased by small perturbations, Δf_0, with 30 values of $\delta = \Delta f_0 / f_0$ logarithmically spaced across the range $0.001 \leq \delta \leq 0.3$. Following Meddis and O'Mard (1997), the squared Euclidean distance between the SACFs of the unperturbed stimulus (δ=0) and each of the perturbed stimuli was then calculated:

$$D^2(f_0,\delta) = \sum_{i=0}^{l_{max}/dt_s}[SACF((1+\delta)f_0, idt) - SACF(f_0, idt)]^2 \qquad (3.2)$$

The procedure to convert from the D^2 statistic to an estimate of the f_0 DL was to choose a criterion based on a threshold D^2 (D_0^2), which served as a free parameter in fitting the model predictions to the psychophysical data. The lowest value of δ producing a D^2 that exceeded D_0^2 was taken to be the estimated f_0 DL. (In practice, to reduce erroneous results due to noise in the data, D^2 was judged to exceed D_0^2 only if it did so for two consecutive values of δ.) Because D_0^2 was allowed to vary as a free parameter, the D^2 measure was unable to predict an absolute f_0 DL that could be directly compared with experimental data. Rather, this statistic yielded a measure of the relative discriminability between stimulus pairs, providing a way to compare trends in the SACF and trends in measured f_0 DLs across different conditions.

(3.4.2.2) Optimal detector model

The D^2 measure is a potentially flawed decision variable. Because D^2 is simply the distance between two SACF functions, it is likely to be sensitive to changes in stimulus dimensions that are unrelated to the stimulus pitch. For example, whereas psychophysical f_0 discrimination performance is fairly robust to changes in stimulus bandwidth, Pressnitzer *et al.* showed that such changes affect the SACF amplitude, and therefore model predictions based on the D^2 statistic. Similarly, Carlyon (1998) demonstrated that the D^2 statistic is susceptible to changes in stimulus amplitude, such as those introduced by level roving in the current study. Although calculating the D^2 between SACF functions averaged across many stimulus trials would reduce the influence of level roving on the model predictions, such a strategy would be likely to fail on a trial-by-trial basis due to its sensitivity to SACF amplitude fluctuations. An optimal detector model, with the ability to incorporate the variance associated with level roving into the decision statistic, was tested as a possible alternative.

The operation of the optimal detector was based on signal-detection theory of Green and Swets (1966). Up to four different sources of noise were present in the model: 1) the stochastic firing of the ANF; 2) stimulus level roving; 3) the background noise; and 4) phase randomization. Only the first two noise sources were always present. For the initial simulations, background noise was not used, while phase randomization was only present in the random-phase conditions. These noise sources produced SACF variation at each lag, allowing the performance of an optimal detector to be computed based on the statistical properties of the SACF variation.

The decision variable was assumed to be a vector $\Delta\overline{SACF}(f_{0A}, f_{0B})$ containing the SACF differences ($\Delta SACF$) yielded at each lag by two stimuli with different f_0's (f_{0A} and f_{0B}):

$$\Delta SACF(f_{0A}, f_{0B}, l) = SACF(f_{0A}, l) - SACF(f_{0B}, l) \quad (3.3)$$

In this model, optimal detection strategy – the weighting of the information obtained at different lags – will vary depending on the f_0 and Δf_0. As in the D^2 model, each stimulus was presented n=15 times for the each combination of f_0, frequency region, phase and δ. Each $s(t,k)$ was substituted for $p(t,k)$ in equation 3.1 to yield stochastic individual channel ACFs, which were then summed across channels to yield n stochastic SACFs.

The performance (d') achieved by an optimal detector for discriminating stimuli on the basis of f_0 was estimated to be:

$$(d')^2 = \Delta\overline{m}^T G^{-1} \Delta\overline{m} \quad (3.4)$$

where $\Delta\overline{m}$ was the mean of the $\Delta\overline{SACF}$ s across the n stimulus trials, and G is the covariance matrix, calculated from the n $\Delta\overline{SACF}$ s (Van Tress, 2001). In practice, both the mean and variance of $\Delta SACF$ were nearly zero for a subset of lags, such that G was often nearly singular and not easily invertible. To resolve this problem, a very small amount of independent noise (variance = 10^{-8}) was added to each lag by augmenting the variances along the diagonal of G.

Because the d' estimates obtained from Eq. 3.4 will vary depending on the number of nerve fibers and the number of lag points used in the simulation, no attempt was made to predict the experimental d' value of 1.26 (2-up, 1-down, 3AFC, Hacker and Ratcliff, 1979) using the model simulations. The extremely large d' estimates reported below are a result of the large number of individual observations of f_0-related activity available across the lag range, and are not reliable estimates of absolute performance. Instead, a similar procedure to the D^2 method was used, whereby a d' criterion (d_0') was chosen in order to predict an f_0 discrimination threshold, allowing relative performance comparisons across conditions.

3.4.3 Stimuli

The stimuli were produced in the same manner as those in the experiment above, including random level roving and phase randomization applied independently to each of the n stimulus presentations. There were three main differences between the stimuli used in the experiment and those used in the modeling simulations. First, the stimuli used in the modeling component were reduced in duration to 200 ms in order to reduce computational load. The shorter duration should have no effect on the model predictions, since the autocorrelations were calculated only near the end of each stimulus, with a relatively short $\tau = 25$ ms and an l_{max} of 25 ms. Furthermore, decreasing the stimulus duration has little effect on f_0 DLs until durations fall below about 100 ms (Plack and Carlyon, 1995). Thus, it can be assumed that these 200-ms stimuli would yield similar results to the 500-ms stimuli used in the psychophysical experiment described above.

Second, no background noise was used in the initial model simulations. The main reasons for using a background noise in the psychophysical experiments (to mask distortion products and to promote the fusion of individual components into a single object) are not issues for the autocorrelation model with linear gammatone filters. However, because the presence of a background noise may still affect the ANF response to the complex tone stimuli, the possible influence of a background noise on the simulation results is examined in section 3.4.5.

Third, the method of setting the signal levels differed from the psychophysical experiment. Because the model contained only high-spontaneous rate ANFs, the dynamic range available to human listeners was not available to the model. Stimulus levels similar to those actually used in the experiment tended to saturate the ANF outputs. To determine a reasonable operating level for the modeling simulations, it was assumed that for a given stimulus level, an optimal detector would choose to use those ANFs that yield the best possible performance, and discard those ANFs that yield little information, as in the "selective listening hypothesis" (Delgutte, 1982,1987; Lai *et al.*, 1994). In these simulations, rather than adjusting the model ANF spontaneous rates and thresholds to find those that yielded the optimal performance for a given stimulus level, the ANF parameters were kept fixed and the stimulus level was adjusted. Pilot tests indicated that the best overall performance (in terms of both D^2 and d') occurred when the

firing rate (r) of an ANF with CF at the center of the stimulus passband was at approximately the 90% point of the operating range, that is, when $r = r_{sp} + 0.90(r_{\max} - r_{sp})$, where r_{sp} and r_{max} are the spontaneous and maximum ANF firing rates, respectively. Therefore, in the simulation results shown below, all stimulus levels were set such that a pure tone at the level and frequency of a harmonic component at the center of the stimulus passband yielded an r at 90% of the operating range of an ANF with CF at the tone frequency. Athough the absolute model performance was best at this stimulus level, the relative performance of the model across the various conditions was generally unaffected by the stimulus level, provided the stimuli were above rate threshold.

3.4.4. Model Results

The two main findings of the simulations are that 1) the D^2 and optimal-detector formulations of the model yield virtually identical predictions, and 2) neither formulation was successful in accounting for the psychophysical results, especially for the sine-phase conditions.

(3.4.4.1) Comparison of the D^2 and d' measures

The Euclidean distance and optimal detector procedures produced virtually identical results. Because both procedures yield the same results, only the optimal detector model will be shown and discussed for the remainder of the paper. That these two procedures yielded similar results is perhaps not surprising, since both measures involve taking the sum of the squares of the differences between SACF functions. The main difference between the two methods is that the d' method weights these differences based on the variances at different lags across stimulus trials, whereas the D^2 statistic weights each lag equally. The similar results seen for the two methods suggests that the weighting was of little consequence – lags falling between SACF peaks added little to the sum of squared differences between SACFs, regardless of the weighting strategy. The finding implies that the D^2 measure was in fact sensitive to f_0-related activity in the SACFs, and that weighting the lags equally yields results similar to those yielded by an optimal strategy.

It is important to note that in these simulations, the Euclidean distance procedure was not challenged with level roving, which was essentially eliminated by averaging SACFs across

stimulus trials. On a trial-by-trial basis, the simple Euclidean distance measure might be more sensitive to the level roving than to the changes in f_0, prohibiting it from detecting changes in f_0. In contrast, the optimal detector formulation took into account the variance due to level roving. The similarity of the two sets of results suggests that the optimal detector model was able to ignore level roving effects in discriminating f_0.

(3.4.4.2) Optimal detector predictions

Figure 3.2 shows SACFs and individual channel ACFs for low-spectrum complexes with three different f_0's. Sine-phase stimulus responses are shown in the top row. For the lowest f_0 of 50 Hz, harmonics are all unresolved and interact within each model filter, such that the ACFs in each channel are phase-locked to the stimulus envelope. For the middle f_0, 150 Hz, harmonics begin to be resolved for the lowest CFs, and ACFs in these channels become phased-locked to individual sinusoids rather than stimulus envelopes. At 300 Hz, harmonic resolvability extends further, up to about 2.4 kHz. Amplitudes of SACF peaks are largest for the 50-Hz case where the f_0 appears to be coded mainly by the envelope, and diminish with increasing f_0, as resolved harmonics appear. A similar effect was observed in the high-spectrum conditions, where the SACF peaks were even smaller (not shown).

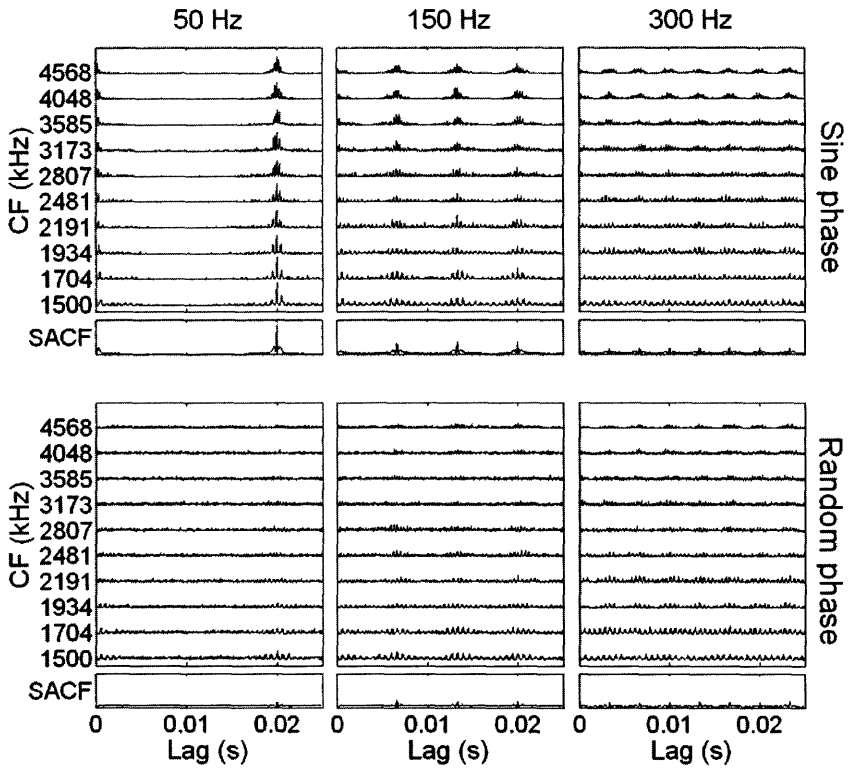

Figure 3.2. Sample ACFs (top 10 plots in each panel) for a subset of model ANFs with CFs as indicated along the vertical axis, for low-spectrum stimuli with three selected f_0's under both phase conditions. The corresponding SACFs are shown in the bottom plot of each panel.

The observed decrease in SACF-peak amplitude with increasing f_0 for sine-phase stimuli is reflected in the model's f_0 DL predictions. Figure 3.3 shows the model's predicted d' as a function of δ, the fractional change in f_0. Figure 3.1b shows the minimum values of δ such that $d'>d_0'$, where $d_0'=190$ was arbitrarily selected (horizontal dashed line in Fig. 3.3) to yield predicted f_0 DLs in the general range of the psychophysical results. For the sine-phase stimuli (open symbols), predicted f_0 DLs generally *increase* with increasing f_0, opposite to the trend seen in the psychophysical data. This is the case in both spectral regions. Note that this trend would occur independently of the chosen d_0', since the $d'(\delta)$ functions (Fig. 3.3) rarely cross. These results indicate that phase-locking to the envelope of unresolved harmonics was stronger than phase-locking to individual resolved harmonics, yielding smaller predicted f_0 DLs for lower stimulus f_0's. This result may depend on the relatively high stimulus spectral regions tested. Phase-locking to resolved components would most likely be stronger for stimuli with energy below 1.5 kHz, the frequency at which phase locking begins to roll off in the guinea pig-based model used here.

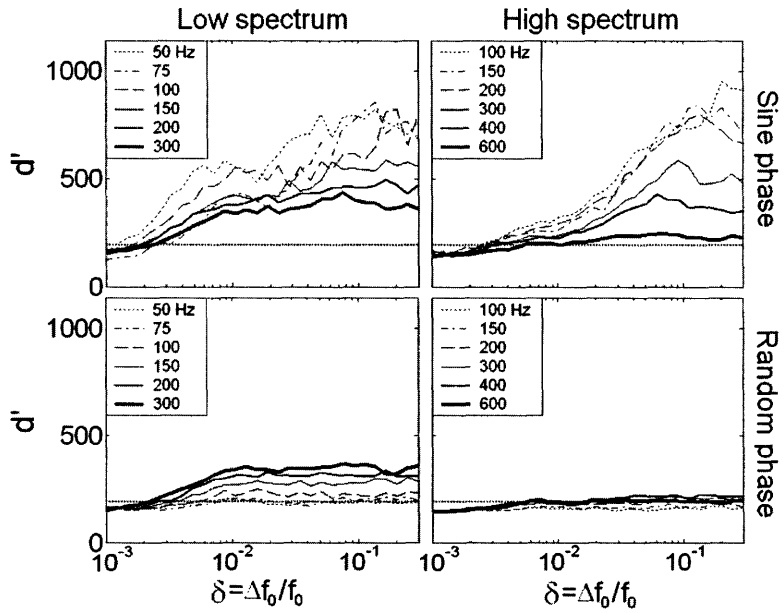

Figure 3.3. Plots of the estimated d' as a function of δ, the fractional change in f_0, as predicted by an optimal detector model. For sine-phase stimuli, slopes decrease with increasing f_0, while for random-phase stimuli, slopes increase with increasing f_0. Horizontal dotted lines indicate the arbitrary d_0'=190 used to predict f_0 discrimination thresholds plotted in Fig. 3.1b. The plots rarely cross, indicating that the predicted f_0 DL vs. f_0 trend is independent of the chosen value of d_0'.

For random-phase stimuli (filled symbols), f_0 DLs predicted by the model tended to decrease with increasing f_0, consistent with the general trend seen in the psychophysical results. Diamonds indicate that d' failed to exceed d_0' for the largest tested value of δ=0.3. The heights of the SACF peaks did not appear to change substantially with f_0 (Fig. 3.2, bottom row), suggesting that the decrease in f_0 DLs is most likely a result of the additional SACF peaks present for stimuli with larger f_0's. This correct behavior for the random-phase conditions is a result of a very large phase effect that is present mainly for low f_0's, where the predicted f_0 DLs for the same f_0 are drastically different between the two phase conditions. The presence of such a phase effect in the model (albeit much larger than that seen in the data) is consistent with previous studies that have found phase effects in the AC for harmonic complexes containing high-order harmonics, but not for those containing low-order harmonics (Patterson *et al.*, 1995; Meddis and O'Mard, 1997; Carlyon and Shamma, 2003). Since the autocorrelation operation discards relative timing information across channels, but remains sensitive to timing information within each channel, we expect the relative phase of harmonics to affect the resulting SACFs only in cases where the harmonics are unresolved by the cochlear filters, i.e. for the lowest f_0's presented.

For similar harmonic numbers present in the passband, the AC model predicts larger f_0 DLs in the high-spectrum conditions (upward-pointing triangles) than in the low-spectrum conditions

(downward-pointing triangles), suggesting an effect of spectral region in the model that was not seen in the psychophysical data. This is consistent with Carlyon's (1998) conclusion that, in contrast to the psychophysical results, the AC model is sensitive to spectral region effects, as a result of the decline in phase-locking with increasing absolute frequency.

3.4.5 Effects of added noise

The above simulations were performed without the presence of background noise. To test the possibility that background noise could affect the model simulation results, a subset of the above simulations were repeated with background noise present. In the psychophysical experiment described above, the background noise level was held fixed and the stimulus level set relative to the detection threshold for a pure tone in the noise. Repeating a similar strategy to determine an appropriate noise level for the modeling simulations would require a model for signal-in-noise detection based on ANF responses, which is outside of the scope of this paper. Instead, we chose to examine the influence of background noise over a range of levels. The nominal signal level was the same as that used in the above simulations. The background noise levels were chosen such that the signal-to-noise ratio (SNR) ranged from -10 dB to +∞ (no noise) relative to the average SNR used in the experiment (SNR_{exp}). The background noise was turned on 100 ms before, and off 100 ms after, the harmonic stimulus.

Figure 3.4 shows the predicted f_0 DLs at various SNRs (*re* SNR_{exp}) for the sine-phase conditions. Low-spectrum and high-spectrum results are plotted in the left and right panels, respectively. The predictions are largely unaffected by the background noise until the SNR reaches the SNR_{exp}. Interestingly, for a narrow window of SNRs

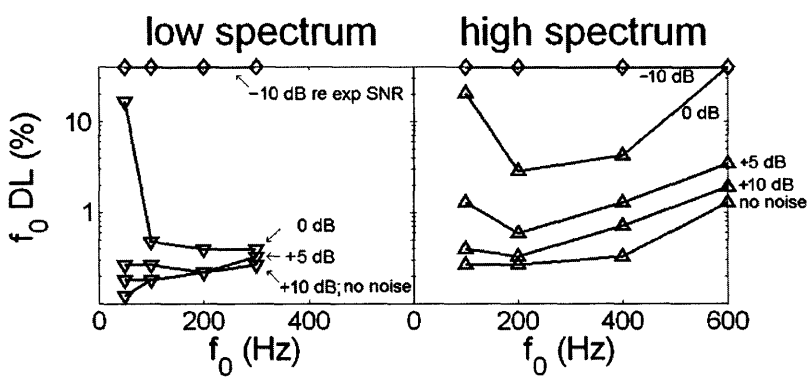

Figure 3.4. Effects of the introduction of background noise on model predictions. Signal level was held constant while the noise level was adjusted; SNR (dB) are described relative to the SNR used in the psychophysical experiment. For SNRs 5 dB or greater than that used in the experiment, the background noise has little effect on model predictions. As in Fig. 3.1, diamonds plotted along the top axis indicate that d_0' was not reached for the highest δ tested of 0.3.

near SNR_{exp}, the trend in f_0 DLs as a function of f_0 actually switches, and f_0 DLs decrease with increasing f_0 as in the experimental data. One aspect of this behavior with respect to SNR is consistent with previous psychophysical data. Hoekstra (1979) showed that f_0 DLs generally increased with decreasing SNR, and this effect was most pronounced in a given fixed frequency region for low f_0's at low SNRs. In the model simulations, the predicted f_0 DLs increase more rapidly with decreasing SNR for low f_0's than for high f_0's. However, Hoekstra (1979) also showed that the general trend for f_0 DLs to improve with increasing f_0 for a fixed spectral region was unaffected by SNR. In contrast, the model only shows a trend for f_0 DLs to increase with f_0 for a narrow range of SNRs, and is therefore unsatisfactory as a predictor of f_0 DL data.

Overall, this analysis shows that the model predictions are relatively unaffected by the presence of background noise, provided the SNR is above a certain threshold. For the remainder of the simulations described below, no noise background was used.

3.5 Model modifications

To account for a variety of psychophysical effects, various modifications to autocorrelation models of pitch have been suggested. These include SACF normalization (Patterson *et al.*, 1996; Yost *et al.*, 1996; Patterson *et al.*, 2000), SACF weighting functions (Pressnitzer *et al.*, 2001; Krumbholz *et al.*, 2003; Cedolin and Delgutte, 2005b), a lag-dependent AC time constant (Wiegrebe, 2001) a nonlinear filterbank (Lopez-Poveda and Meddis, 2001), and a CF-dependent ACF weighting function (Moore, 1982). In the model simulations described below, the CF-dependent weighting function was the most successful in accounting for the effect of harmonic number observed in these psychophysical results of Section 3.3. Each of these possible modifications will be discussed in turn.

3.5.1. SACF Normalization

The height of the SACF peak normalized to the value at zero lag has been successful in predicting the pitch strength of iterated ripple noises (Patterson *et al.*, 1996; Yost *et al.*, 1996; Patterson *et al.*, 2000). Cariani and Delgutte (1996a; 1996b) performed an analysis similar to SACF normalization by using the peak-to-background ratio in the all-order interval histogram as a neural estimate of the pitch salience. They were able to successfully account for a wide range

of psychophysical pitch phenomena using this type of analysis. However, when the optimal detector model was adjusted to include SACF normalization (results not shown), there was virtually no change from the results seen in Fig. 3.3. The reason for this is that the optimal detector inherently normalizes the SACF function to the standard deviation at each lag. In essence, the extra normalization step scales the mean and standard deviation of the SACF equally, leaving d' unaffected. SACF normalization did serve to reduce the noise associated with level roving, increasing the overall d'. However, this effect was similar across all conditions, such that when d_0' was adjusted accordingly, normalized and unnormalized SACFs yielded virtually identical f_0 DL predictions.

3.5.2. SACF weighting function

An SACF weighting function that generally gives more weight to short lags should yield a larger estimated d' for high-f_0 stimuli that contain SACF peaks at short lags. Thus, such a weighting function may have the potential to account for the better discrimination performance observed for high f_0's. For example, Pressnitzer *et. al.* (2001) found that the Meddis and O'Mard (1997) model, modified to include a linear SACF weighting function, successfully predicted an increase in the lowest f_0 that could convey melody for higher spectral regions. In the optimal detector formulation, weighting the SACF would have no effect, since the weights would alter both the mean and standard deviations by the same factor, thus not affecting d'. Instead, independent noise with variance $\sigma_w^2(l)$ was added along the diagonal of the covariance matrix G in equation 3.4, according to

$$\sigma_w^{\ 2}(l) = w(l)^{-2} \tag{3.5}$$

where $w(l)$ is the analogous SACF weighting function. Three different versions of $w(l)$ were tested: a linear function, $w = 1 - l / l_{\max}$ (Pressnitzer *et al.*, 2001), a power function, $w = 1 - (l / l_{\max})^{\alpha}$ with α ranging from 1/64 to 1 (Krumbholz *et al.*, 2003), and an exponential function, $w = \exp(-l / \lambda)$ with λ ranging from 0.3 to 30 (Cedolin and Delgutte, 2005b). For each w, the model was tested both with and without SACF normalization. The most promising results were produced by the combination of an exponential $w(l)$ with $3<\lambda<4$ms, and SACF normalization. For low-spectrum stimuli, this modified model yielded f_0 DLs that decreased with increasing f_0 for low-spectrum stimuli, consistent with the experimental data (results not

shown). However, this combination of modifications was unable to account for the high-spectrum data, and was therefore unsatisfactory. None of the other functions produced desirable results.

3.5.3 A lag-dependent time constant

Another lag-dependent AC modification was suggested by Wiegrebe (2001), whereby the AC time constant (τ in Eq. 3.1) increases with increasing lag. Like the SACF weighting function of Pressnitzer *et al.* (2001), a lag-dependent τ would affect the SACF differently for different stimulus f_0's, and could therefore influence the model's f_0 DL predictions. However, this modification would most likely not account for the results of the experiment described in section 3.3, because the longer time constant associated with low f_0's would tend to increase the amplitudes of peaks in the SACF, yielding smaller f_0 DLs than for high f_0's. Thus, Wiegrebe's (2001) modification would be likely to skew the model predictions even more heavily in favor of low f_0's.

3.5.4 A nonlinear filterbank

The model simulations described above used a bank of linear gammatone filters (Patterson *et al.*, 1992) to represent the basilar membrane. A more accurate nonlinear filter model that includes the compressive input-output function observed at the level of the basilar membrane (Rhode, 1971; Ruggero *et al.*, 1997) has been shown to be important for a number of psychophysical phenomena (Oxenham and Bacon, 2003), and might better account for the f_0 DL data. The inclusion of a basilar membrane nonlinearity (e.g. Lopez-Poveda and Meddis, 2001) might compress the "peaky" sine-phase waveform more than the "flat" random-phase waveform yielded by interacting unresolved harmonics (Carlyon and Datta, 1997), possibly reducing the size of the phase effect predicted by the AC model. However, simulations using the dual-resonance nonlinear (DRNL) filterbank (Lopez-Poveda and Meddis, 2001) yielded unsatisfactory results (not shown), similar to those seen with the gammatone model. Thus, although the compression offered by this model is similar to that observed physiologically, it was not substantial enough to account for these data.

3.5.5 A CF-dependent "lag window"

Section 3.4 showed that for sine-phase stimuli, the Meddis and O'Mard (1997) AC model responded preferentially to low f_0's for stimuli bandpass filtered in fixed spectral regions. Therefore, to successfully predict the improved f_0 discrimination for higher f_0's seen in the human performance, the AC model must be modified in such a way as to impair performance for low f_0's within a given spectral region. One way to accomplish this is to limit the range of lags for which the autocorrelation is calculated in each frequency channel in a CF-dependent manner (Moore, 1982). With this lag-window limitation, the AC will respond best to f_0's that have certain harmonic numbers falling within each channel's bandwidth. Schouten (1970) first proposed the idea that "each pitch extractor has a limited range of measurable time intervals" in order to account for Ritsma's (1967) demonstration of the dominance of low-order harmonics in complex pitch perception. Moore (1982) further quantified the lag window, suggesting that a mechanism based on first-order interspike intervals operates over a range of lags between about 0.5/CF and 15/CF. Thus the AC in a particular channel will respond to f_0's that are 1/15 to 2 times the channel's CF. In other words, the AC will respond to a given f_0 only if at least one of the f_0's 1st-15th harmonics fall near the CF. Ghitza (1986) implemented a similar idea, whereby the interspike interval analysis window length was roughly inversely proportional to each channel's CF.

After experimenting with various possibilities, we found that a piecewise-linear weighting function was able to account for the psychophysical data with some success. The CF-dependent weighting function consisted of four segments:

$$w_{acf}(l,CF) = \begin{cases} 0, & l < 0.5/CF \\ CF^2/CF_0, & 0.5/CF \le l < N_C/CF \\ CF^2/CF_0 - m\left(l - \dfrac{N_C}{CF}\right), & N_C/CF \le l < (N_C + N_\Delta)/CF \\ A - \dfrac{A}{l_0}l, & l \ge (N_C + N_\Delta)/CF \end{cases} \qquad (3.6)$$

where l is the lag, CF_0 = 1500 Hz, the lowest CF used in the simulations, N_C is the cutoff between the second and third segments relative to CF, N_Δ is the width of the third segment

relative to CF, A is the amplitude of the fourth segment at l=0, l_0 is the lag for which the fourth segment reaches zero, and m, the slope of the third segment, is defined as

$$m = \frac{CF^2 - A + \frac{A}{l_0}\left(\frac{N_C + N_\Delta}{CF}\right)}{N_\Delta / CF} \tag{3.7}$$

The fourth segment, independent of CF, is identical to the linear SACF weighting function of Pressnitzer *et al.* (2001). The zero-crossing of this segment (l_0) was set to 33 ms as suggested by Pressnitzer *et al.*, consistent with a 30-Hz lower limit of melodic pitch. Finally, in some conditions, the estimated d' reflected activity at low lags completely unrelated to the stimulus f_0. To prevent this phenomenon, w_{acf} for each CF was set to zero for all values of l < 0.875 ms. Sample w_{acf} functions for various CF are shown in Fig. 3.5. (The linear segments of the functions appear curved because they are plotted on a logarithmic scale.) The lag window was applied to the ACF for each simulated ANF, and these windowed ACFs were summed to create the SACF just as before.

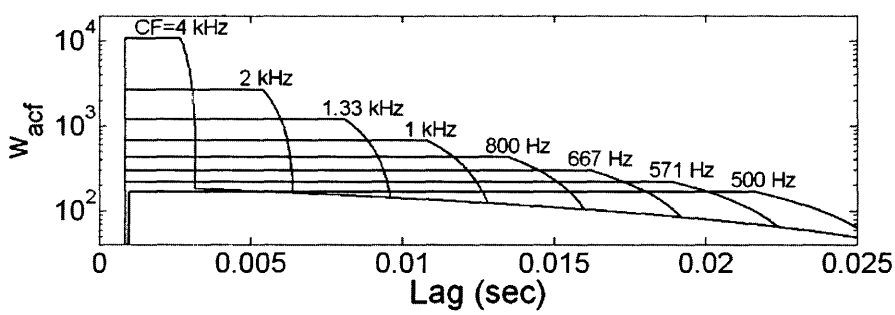

Figure 3.5. Sample w_{acf}'s (Eq. 3.6) for a range of CFs, with parameters that yielded the best fit to the experimental data as shown in Fig. 3.7.

The CF-dependent windowing procedure described here was notably different from the SACF weighting described in section 3.5.2. There, the addition of independent noise was used as a substitute for an SACF-weighting function, which would have scaled the mean and standard deviation equally, yielding no net effect on d'. Here, the weighting functions (w_{acf}) were applied to the individual ACFs before summing them to produce the SACF. Thus, the statistical properties of the SACF at each lag tended to reflect the statistical properties of the ACFs for channels that were most heavily weighted at that lag.

Estimates of d' were generally noisier than in the unmodified model because the lag window tended to reduce the total number of ANF spikes that were used in the calculation. Therefore,

two minor modifications were made. First, the number of stimulus repetitions n was increased to 100. Second, d' was determined to exceed threshold only if it did not fall below d_0' again for a higher value of δ. This ensured that the threshold was not exceeded due to random fluctuations in the d' estimates.

The modified AC model was fit to the sine-phase experimental data of section 3.3 with four free parameters (N_C, N_Δ, A and d_0'). The two most important aspects of the experimental data were the dependence of f_0 DLs on N, and the lack of an effect of spectral region on f_0 DLs. Therefore, the fitting procedure minimized the sum of two error measures: the root-mean-squared difference between the logarithms of predicted and actual f_0 DLs, and the root-mean-squared difference between the logarithms of the predicted f_0 DLs for stimuli with equivalent N's in the low- and high-spectrum conditions. The strong model nonlinearities and limited range of δ values tested prohibited the successful use of an automated fitting procedure, such as the Nelder-Mead simplex method used by MATLAB's fminsearch function. Instead, a parameter-space search method was used, where coarse step-sizes allowed for a reduction in computation time. Thus, we caution that a somewhat different set of parameters may yield a better fit than those reported here.

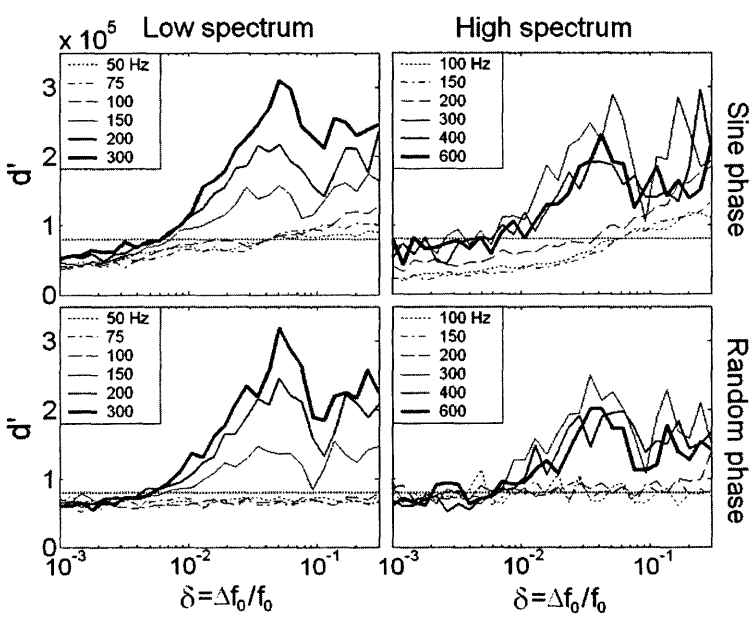

Figure 3.6. Model estimates of d' versus δ using the lag windows described in Eq. 3.6 and pictured in Fig. 3.5, with parameters N_C =10.8, N_Δ=2 and A=200 that best fit the sine-phase data. Stimulus f_0's are clearly divided into two groups, with lower f_0's yielding gradual d' slopes, and higher f_0's yielding steeper d' slopes.

Figure 3.6 shows d' as a function of δ for the modified model with parameters that yielded the best fit to the sine-phase experimental data: N_C =10.8, N_Δ=2 and A=200. The sample w_{acf}

functions depicted in Fig. 3.5 reflect these parameter values. The best-fitting d_0' of 7.91×10^4 is depicted as a horizontal dashed line in each panel of Fig. 3.6. Figure 3.7b shows the modified model's f_0 DL predictions as a function of N, based on these best-fitting parameters. The f_0's corresponding to the N's are shown along the top axis. The psychophysical results from Fig. 3.1a are replotted in Fig. 3.7a for direct comparison with the model predictions. The modified model yielded a reasonable fit to both sets of data, and captured three main features of the data. First, f_0 DLs generally decrease with increasing f_0. Second, the model predictions for the two spectral regions overlap when plotted as a function of N, such that f_0 DL are mainly dependent on harmonic number. The separation of stimuli into two groups based on N is clearly seen in Fig. 3.5: those stimuli with low f_0's, such that $N>12$, have shallow d' versus δ slopes, yielding large f_0 DLs, while those with high f_0's, such that $N<12$, have steeper slopes, yielding small f_0 DLs. Third, phase effects are only present for complexes with large N. For small N, sine- and random-phase stimuli yield similar f_0 DL predictions.

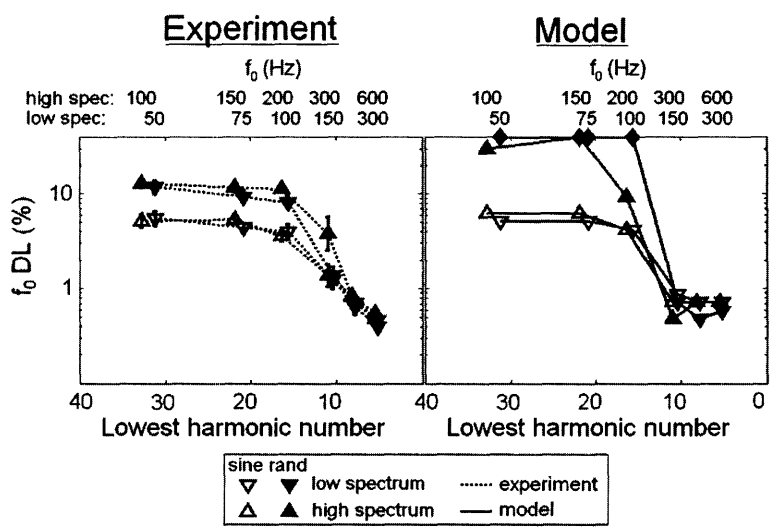

Figure 3.7. (a) Psychophysical f_0 DLs from Fig. 3.1a are replotted for direct comparison with the model predictions. (b) Model f_0 DL predictions based on d' estimates shown in Fig. 4.6 using the lag window (Eq. 4.6), plotted as a function of N. As in Fig. 3.1, f_0's corresponding to values of N for the low- and high-spectrum conditions are plotted above each panel, and diamonds (♦) plotted along the top horizontal axis indicate that d_0' was not reached at the maximum tested value of δ=0.3. Both experimental and model f_0 DLs generally overlap for stimuli with the same N, indicating the modified model successfully accounts for effects of N on f_0 discrimination performance.

The one major failure of the modified model is that it overpredicted the phase effect for low f_0's. The variability in the envelopes associated with low-f_0, unresolved, random-phase complexes was so large relative to the mean envelope that d' was not affected by increasing δ. Thus, the model failed to reach threshold at the highest tested value of δ=0.3, and was unable to predict discrimination thresholds for these complexes. This problem was also observed for the original, unmodified model.

The inclusion of a compressive nonlinearity in the model might help to reduce the magnitude of this phase effect by compressing "peaky" sine-phase envelope more than "flat" random-phase envelopes. However, because substituting DRNL filters (Lopez-Poveda and Meddis, 2001) for gammatone filters did not greatly affect the predictions of the unmodified model (Section 3.5.4), it is also unlikely to greatly influence the predictions of the modified model.

Non-monotonicities were observed in d' estimates at the three highest f_0's tested in each condition. For values of δ near 0.1, d' estimates suddenly decreased then increased. This non-monotonic behavior can be understood by examining the sample SACF functions in Fig. 3.8. For the relatively high f_0 of 200 Hz, the SACF contains multiple sharp peaks at lags near $1/f_0$, reflecting the stimulus fine structure. As δ increases, these closely spaced peaks move in and out of alignment with one another, yielding the observed non-monotonic behavior. In contrast, the SACF representations for low f_0's (e.g. 50 and 100 Hz) are dominated by a single large peak at each multiple of $1/f_0$, with relatively small side bands. As a result, non-monotonic behavior is not observed for these stimuli. This analysis suggests that the model uses fine-structure information to discriminate f_0 for low-order, but not for high-order harmonics. Regardless, these non-monotonicities occur for f_0 separations well above the discrimination threshold, and therefore do not impact the model's f_0 DL predictions.

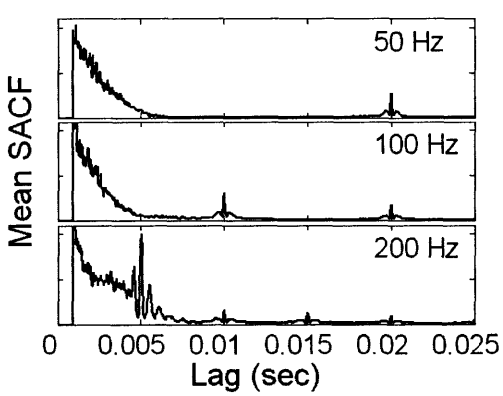

Figure 3.8. Mean SACFs produced with the lag-window modification (Eq. 3.6; Fig. 3.5) for sine-phase stimuli with various f_0's. For higher f_0's (e.g. 200 Hz), the large SACF peak at l=5 ms contains large fine-structure side peaks, causing the non-monotonic behavior of d' observed in Fig. 3.6. For lower f_0's (50 and 100 Hz), the SACF side peaks are small relative to the central SACF peak; non-monotonic d' behavior is not observed for these stimuli.

3.6 Discussion

The analysis of section 3.4 showed that the Meddis and O'Mard unitary AC model of pitch perception is unable to account for the dependence of f_0 DLs on harmonic number. Whereas experimental data presented both here and elsewhere (Hoekstra, 1979; Houtsma and Smurzynski, 1990; Carlyon and Shackleton, 1994; Shackleton and Carlyon, 1994; Kaernbach and Bering, 2001) show that discrimination performance deteriorates with increasing lowest harmonic number within a given passband, the Meddis and O'Mard model predicts just the opposite for the stimuli used here. This result is consistent with the results of Cedolin and Delgutte (2005b) who estimated pitch salience based on all-order interval analysis of cat ANF spikes. They found that pitch salience estimated in this way was maximal for the lowest f_0's tested, where individual harmonics are not well resolved by the cat's auditory periphery.

We have shown (section 3.5) that this failure of the AC model is not fatal to the idea that a single mechanism based on temporal information can account for the perceived pitch based on both resolved and unresolved harmonics. With the introduction of a CF-dependent lag window similar to that described by Moore (1982), the model was able to predict the dependence of f_0 discrimination on harmonic number. This was achieved because the modification reverses the original model's "preference" for high-order harmonics by applying a weighting function that amplifies the AC response to low-order harmonics, and attenuates the response to high-order harmonics.

The success of the modified AC model where the original model has failed supports the idea that temporal information alone is not enough to yield a salient pitch percept, and that the temporal information must be presented at the correct place on the cochlear partition in order to yield good f_0 discrimination performance (Oxenham *et al.*, 2004). The lag window modification effectively codes "place" information into the AC model by weighting each channel's contribution based on its relationship to the stimulus f_0. For a given CF, a range of lags between 0.5/CF and N_C/CF are weighted most heavily. The ACF will respond most readily to a certain range of stimulus f_0's that contain peaks falling within this lag range.

It is important to note that the correct behavior of the modified model with regard to the effects of harmonic number is *not* based on harmonic resolvability. The modified model responds preferentially to complexes containing low harmonics because of the introduction of the CF-dependent w_{acf}. This could be considered a major failing of the model, if good f_0 discrimination performance were directly dependent on the presence of resolved harmonics. On the other hand, the direct dependence of the modified model's f_0 DL predictions on harmonic number is consistent the results of several studies (described in the Introduction) suggesting that f_0 discrimination performance may depend only on harmonic number, and not on harmonic resolvability *per se* (Chapter 2; Houtsma and Goldstein, 1972; Arehart and Burns, 1999).

The AC model was modified to fit the f_0 discrimination data described in section 3.3, and has not yet been tested on other data sets. Nevertheless, the dependence of predicted f_0 performance on harmonic number is a direct result of the w_{acf} modification, suggesting that the modified model should be able to account at least qualitatively for the results of other studies that have shown an increase in f_0 DLs with increasing N. These include f_0 discrimination studies with bandpass-filtered harmonic complexes (Houtsma and Goldstein, 1972; Hoekstra, 1979; Houtsma and Smurzynski, 1990; Shackleton and Carlyon, 1994), as well as those that manipulate N for complexes with a fixed f_0 (Chapter 2; Houtsma and Goldstein, 1972; Houtsma and Smurzynski, 1990; Shackleton and Carlyon, 1994). Furthermore, because this modification relies on harmonic number rather than peripheral resolvability, it is likely to account for results indicating that the diotic presentation of alternating harmonics does not improve f_0 discrimination performance (Chapter 2; Houtsma and Goldstein, 1972; Arehart and Burns, 1999) despite the improvement in peripheral resolvability (Chapter 2).

In contrast to the behavior of the modified model with respect to N, its correct behavior with respect to phase effects is most likely based on harmonic resolvability. The original model predicted a large effect of phase on f_0 DLs for low f_0's containing unresolved harmonics, where the SACF mainly reflects phase-locking to the envelope (Fig. 3.2). In these conditions, the envelope resulting from the interaction of multiple harmonics within one filter was much "peakier" with sine-phase complexes than with random-phase complexes, yielding smaller predicted f_0 DLs. While the modified model predicts that stimuli yielding large f_0 DLs should

also yield phase-dependent f_0 DLs, the two effects rely on different processes. The dependency of f_0 DLs on harmonic number derives from the w_{acf} modification. The dependency on phase derived from inherent differences in the way the model processes resolved and unresolved harmonics, and was correctly predicted by both the original and modified AC models.

How is the mathematical formulation of a lag window to be interpreted in terms of physiological mechanisms? Licklider (1951) formulated an AC model of pitch perception in terms of a system of neurons, where every cochlear frequency channel is associated with its own bank of AC neurons, and each neuron in the bank is tuned to one of a wide range of periodicities. The ACF (Eq. 3.1) represents the responses of each of the neurons in the bank, and the lag window is a weighting function applied to these responses. In the physiological interpretation, a larger number of neurons associated with a given lag will reduce the noise in the periodicity representation, yielding smaller predicted f_0 DLs for the f_0 associated with that lag.

In a manner similar to that described in the harmonic template model of Shamma and Klein (2000), the autocorrelation mechanism might develop over time to detect only those temporal correlations that tend to occur in the outputs of individual ANFs in response to generic wideband stimuli. The CF-dependent lag windows described here (Eq. 3.6 and Fig. 3.5) could emerge naturally based on the statistical properties of ANF outputs in response to such stimuli. Since the temporal extent of the impulse response of a bandpass filter is inversely proportional to the filter's bandwidth, the narrower filters associated with lower CFs will yield a wider range of lags over which a filtered wideband input stimulus will correlate with itself. Mirroring the properties of these naturally-occurring autocorrelations, the system would be tuned to detect ANF response correlations at longer lags for low CFs than for high CFs. de Chevigné and Pressnitzer (2005) have proposed a similar idea that relates filter impulse response times to pitch processing.

With the addition of a CF-dependent lag window, a single pitch mechanism based on temporal information can account for the poorer f_0 discrimination performance associated with high N. However, it does not address other evidence relating to frequency modulation (FM) detection and temporal integration that points to the possible existence of two separate pitch mechanisms. Plack and Carlyon (1995) showed that f_0 discrimination was affected by decreasing stimulus

durations below 100 ms more for unresolved than for resolved complexes. They suggested that the exceptionally poor FM detection performance (relative to the f_0 DL) measured for unresolved complexes resulted from an absence of the longer integration time needed to extract the f_0. Because the modified autocorrelation model needs the same integration time for a given f_0 (i.e. somewhat longer than a single pitch period, in order to yield an SACF peak at $l=1/f_0$) regardless of resolvability, it is not likely to account for this result. It may be possible to interpret the CF-dependent weighting function as a manifestation of two pitch mechanisms. In this interpretation, the second segment of the lag-window (Eq. 3.6) corresponds to the mechanism for low-order resolved harmonics, the CF-independent fourth segment represents the more poorly performing mechanism for high-order, unresolved harmonics and the third- segment represents the transition between the two.

The autocorrelation model outlined here and elsewhere (e.g.Meddis and Hewitt, 1991a,b; Cariani and Delgutte, 1996a,b; Meddis and O'Mard, 1997) takes into account all-order intervals between ANF spikes. Kaernbach and Demany (1998) challenged the view that the f_0 detection mechanism takes into account anything but first-order interspike intervals. They showed that a click-train with f_0 information in its first-order interspike interval statistics was easier to discriminate from a random click train than a click train containing f_0 information in its second- and higher-order interval statistics, even though the waveform autocorrelation showed a similar peak at a lag corresponding to the f_0 in both cases. However, Pressnitzer *et al.* (2002) showed that an all-order autocorrelation based on simulated ANF responses, rather than the raw waveform, may be able to account for this phenomenon, as a result of the auditory filtering and neural transduction present in the model.

3.7 Summary and conclusions

Measurements of f_0 DLs for bandpass-filtered harmonic stimuli demonstrated that f_0 discrimination performance depends largely on harmonic number: as the ratio of a complex's f_0 to the frequency of its lowest component increases, f_0 discrimination improves. The Meddis and O'Mard (1997) unitary AC model of pitch perception fails to predict this effect of harmonic number on f_0 discrimination. While psychophysical measurements show an improvement in f_0 discrimination with increasing f_0 for bandpass filtered harmonic stimuli, the AC model predicts

the opposite behavior, at least for sine-phase complexes. In order for the model to correctly predict the psychophysical results, an *ad hoc* modification was made, whereby the lags for which the AC was measured in each frequency channel were weighted in a CF-dependent manner. This yielded f_0 DL predictions that decreased with increasing f_0, and depended mainly on harmonic number effects, consist with the data. This modification works by forcing the model to respond preferentially to low numbered harmonics. The correct behavior of the model in no way reflects a preference for resolved harmonics, *per se*. Instead, the model introduces a dependence on harmonic number, without regard to harmonic resolvability.

In conclusion, this study has shown that a single autocorrelation mechanism, modified to include CF dependency, is sufficient to account for the dependence of f_0 DLs on harmonic number. Consequently, two pitch mechanisms may not be needed to explain this effect. Nevertheless, the modified autocorrelation model may not account for other evidence for two pitch mechanisms, such as the differences observed between resolved and unresolved harmonics in the temporal integration of f_0 information (Plack and Carlyon, 1995).

3.8 Segue

Because the "lag window" modified autocorrelation model does not depend on peripheral harmonic resolvability to account for the dependence of f_0 DLs on harmonic number, it is generally consistent with the results of Chapter 2, where an increase in the number of peripheral resolved harmonics did not shift the f_0 DL transition to a higher harmonic number. Similarly, the modified model should not predict a shift in the f_0 DL transition with a *reduction* in the number of resolved components. In the following two chapters, f_0 DLs and frequency selectivity were measured in two conditions where fewer resolved harmonics should be available due to a reduction in peripheral frequency selectivity: high stimulus levels (Chapter 4) and sensorineural hearing loss (Chapter 5).

Chapter 4: Necessity of resolved harmonics for accurate f_0 discrimination: Stimulus level

This work described in this chapter has been submitted to the *Journal of the Acoustical Society of America.*

Bernstein, J.G. and Oxenham, A.J. (2005b). The relationship between harmonic resolvability and pitch discrimination: Effects of stimulus level. *J. Acoust. Soc. Am.* (submitted).

4.1 Abstract

Three experiments tested the hypothesis that fundamental frequency (f_0) discrimination depends on the resolvability of harmonics within a tone complex. Auditory filter bandwidths increase with stimulus level, providing a tool to investigate effects of reduced frequency selectivity. Fundamental frequency difference limens (f_0 DLs) were measured for random-phase harmonic complexes with eight f_0's between 75 and 400 Hz, bandpass filtered between 1.5 and 3.5 kHz, and presented at 12.5 dB/component average sensation level in threshold equalizing noise (TEN) with levels of 10, 40, and 65 dB SPL per equivalent rectangular auditory filter bandwidth (ERB_N). With increased level, the transition from large (poor) to small (good) f_0 DLs shifted to a higher f_0. This shift corresponded to a decrease in harmonic resolvability, as estimated in the same listeners with excitation patterns derived from measures of auditory filter shape and with a more direct measure that involved hearing out individual harmonics. The results are consistent with the idea that resolved harmonics are necessary for good f_0 discrimination. Additionally, f_0 DLs for high f_0's increased with stimulus level in the same way as pure-tone frequency DLs, suggesting that for this frequency range, f_0 is more poorly encoded at higher levels, even when harmonics are well resolved.

4.2. Introduction

Harmonic sounds, consisting of a sum of sinusoids, each with a frequency at a multiple of the fundamental frequency (f_0), are ubiquitous in our natural environment. Voiced human speech, sounds of many musical instruments, animal vocalizations, and mechanical vibrations are all periodic signals whose frequency spectra are made up of sinusoids at discrete harmonically related frequencies. The auditory system tends to group the individual harmonic components together into a single percept with a pitch that usually corresponds to the f_0 of the complex, even

if the component at the f_0 is absent from the stimulus or is masked (Schouten, 1940; Licklider, 1954).

Recent debates surrounding pitch perception have focused on the dependence of f_0 discrimination on harmonic number. The just-noticeable difference in the f_0 of a harmonic complex (the f_0 difference limen, or f_0 DL) has been shown to be smallest for complexes containing low-order harmonics, below about the 10th (Chapters 2 and 3; Houtsma and Smurzynski, 1990; Shackleton and Carlyon, 1994; Krumbholz *et al.*, 2000; Kaernbach and Bering, 2001). This dependence of f_0 DL on harmonic number has generally been ascribed to resolvability. On a linear frequency scale, individual components of a harmonic complex are equally spaced, whereas auditory filter bandwidths increase with increasing center frequency. As a result, low-order harmonics, spaced wider than filter bandwidths along the basilar membrane, are resolved by the auditory periphery, whereas multiple high-order harmonics fall within the bandwidth of a single auditory filter and are therefore unresolved.

Although many models of pitch perception are able to account for the dependence of f_0 discrimination on harmonic number, they do so in different ways. "Spectral" (Goldstein, 1973; Terhardt, 1974; 1979) and some "spectro-temporal" (e.g. Shamma and Klein, 2000; Cedolin and Delgutte, 2005a) models of pitch propose that the f_0 of a harmonic complex is identified by comparing the frequencies of individual harmonics to an internally stored "harmonic template". Because spectral models require spectrally resolved components to extract the f_0, they predict performance to worsen with increasing harmonic number due to a reduction in harmonic resolvability. "Temporal" models of pitch typically discard place information and extract f_0 information based on an autocorrelation or all-order interval analysis of auditory-nerve firing patterns, pooled across the population of fibers (Meddis and Hewitt, 1991b,1992; Cariani and Delgutte, 1996a,b; Meddis and O'Mard, 1997). These models predict a deterioration in f_0 discrimination with increasing absolute frequency (Cariani and Delgutte, 1996a; Carlyon, 1998), as a result of the roll-off of phase-locking in the auditory nerve (Weiss and Rose, 1988a,b), but are unable to account for the dependence of f_0 discrimination on harmonic number *per se* (Carlyon, 1998). A recent modification of the autocorrelation model (Chapter 3) was designed to account for the effects of harmonic number by limiting the range of periodicities over which the

autocorrelation function is calculated relative to each filter's CF (Schouten, 1970; Moore, 1982; Ghitza, 1986). However, the resulting dependence on harmonic number is not determined at the periphery and so is not based on harmonic resolvability. Although the inclusion of place information in this modified temporal model renders it "spectrotemporal" in nature, this type of model is referred to as "CF-dependent temporal" to differentiate it from the harmonic-template spectrotemporal models discussed above (e.g. Srulovicz and Goldstein, 1983; Shamma and Klein, 2000; Cedolin and Delgutte, 2005a).

The question addressed in this study is whether the increase in f_0 DL with increasing lowest harmonic number is directly related to a decrease in the resolvability of the harmonics (as predicted by spectral models), or whether the increase is related only to harmonic number, independent of resolvability (as predicted by a CF-dependent temporal model). The study exploited the fact that frequency selectivity, measured both physiologically (e.g., Rhode, 1971; Robles *et al.*, 1986) and psychophysically (e.g., Weber, 1977; Pick, 1980; Moore and Glasberg, 1987; Glasberg and Moore, 1990; Rosen and Stock, 1992; Hicks and Bacon, 1999), broadens at high levels, at least for frequencies at or above 1 kHz. The link between harmonic resolvability and pitch perception was examined by measuring the effects of stimulus level on complex-tone f_0 discrimination, pure-tone frequency discrimination, auditory filter bandwidths, and the ability to hear out the frequencies of individual harmonics in the same normal-hearing listeners.

4.3. Experiment 4A: Fundamental frequency discrimination

4.3.1. Rationale

Fundamental frequency DLs are known to increase with increasing lowest harmonic number present. Low-order harmonics yield small f_0 DLs (< 1% of the f_0) and high-order harmonics yield large f_0 DLs (2 to 12% of the f_0 depending on the sensation level and phase relationships of the harmonics), with a steep transition between the two regions occurring in the vicinity of the 10th harmonic. This transition is seen whether harmonic complexes are bandpass filtered into a fixed spectral region and the f_0 adjusted (Chapter 3; Hoekstra, 1979; Shackleton and Carlyon, 1994) or the f_0 is held constant and the harmonic number adjusted (Chapter 2; Houtsma and Smurzynski, 1990). Experiment 4A aimed to determine whether the point at which this f_0 DL transition takes places varies with stimulus level. Fundamental frequency DLs were measured as

a function of f_0 for harmonic complexes. By using this paradigm instead of holding f_0 constant and adjusting harmonic number (Chapter 2; Houtsma and Smurzynski, 1990), the spectral region remained constant for all stimuli, eliminating the potentially confounding effects of absolute frequency on the level-dependence of frequency selectivity (Baker *et al.*, 1998). If small f_0 DLs are associated with resolved harmonics, then the transition between small and large f_0 DLs should occur at a higher f_0 (lower harmonic number) at high stimulus levels where frequency selectivity is poorer, because a wider frequency spacing would be required to resolve individual harmonics.

4.3.2. Methods

All stimuli were presented in threshold equalizing noise (TEN, Moore *et al.*, 2000), for four reasons. First, and most importantly, the use of a background noise enables the presentation of stimuli at a constant SL while varying the absolute SPL over a wide range. This reduced the possibility that stimulus SL could have a confounding influence on f_0 DLs (Hoekstra, 1979). Second, TEN is intended to yield detection thresholds for pure tones in noise that are constant across frequency such that harmonics with equal sound pressure level (SPL) will also have equal sensation level (SL). Third, the noise serves to mask any possible combination tones in the frequency region below the stimulus frequency range. Fourth, the presentation of harmonic complexes in a background noise is thought to aid the fusion of individual partials into a single perceptual object (Grose *et al.*, 2002), enabling the listeners to focus their attention more easily on the stimulus pitch.

The stimulus levels presented in this experiment were referenced to the pure-tone detection thresholds for three TEN levels: 10, 40, and 65 dB SPL per equivalent rectangular auditory filter bandwidth (ERB_N, Glasberg and Moore, 1990). Because some of the stimuli presented in the high-level noise were uncomfortably loud for one of the subjects (GW), the highest-level noise was reduced to 62 dB SPL/ERB_N for this subject. To determine the sensation level (SL) reference, pure-tone detection thresholds were measured for each subject at each noise level for 1.5-, 2.5- and 3.5-kHz tones, frequencies that correspond to the lower-frequency cutoff, center frequency, and upper-frequency cutoff of the bandpass filter used in the f_0 discrimination experiment. Although TEN was intended to yield constant pure-tone detection thresholds, there

was some small variation in the threshold SPL at the three frequencies tested. Therefore, we defined 0 dB SL for each noise level as the maximum of the thresholds measured across the three tested frequencies. Across subjects, the 0-dB SL reference ranged from 5.5 to 10.7, 36 to 38.5, and 60.5 to 64.5 dB SPL for the 10, 40 and 65 dB SPL/ERB_N TEN levels, respectively. Harmonic complex stimuli were presented at an average 12.5 dB SL per component. The absolute stimulus SPLs corresponding to this SL for the three levels of background TEN are referred to as the low, mid, and high levels, respectively. Although the across-frequency variation was greater at the low level (across-subject mean of the across-frequency standard deviation of 2.2 dB) than the mid and high levels (standard deviation 1.0 dB in each case), this resulted in an average SL only 0.5 dB higher at the low level than at the mid and high levels.

The stimuli for this experiment consisted of 500-ms (including 30-ms raised-cosine rise and fall ramps) bandpass-filtered random-phase harmonic complexes. A new set of phases were selected independently from a uniform distribution for each stimulus. The large f_0 DLs produced by random-phase complexes for unresolved harmonics (Chapter 3; Micheyl *et al.*, 2005) should maximize the difference between f_0 DLs associated with low and high f_0's, thus providing the best opportunity to observe the transition between these two regions. The bandpass filter was held constant throughout the experiment, with 1.5- and 3.5-kHz corner frequencies and 50 dB/octave low- and high-frequency slopes. The filtering operation was implemented in the spectral domain by first adjusting the amplitude of each sinusoidal component, then summing all of the components together. Stimuli were presented in the TEN background at each the three levels described above. Fundamental frequency discrimination was tested for eight different average f_0's (75, 125, 150, 175, 200, 250, 325 and 400 Hz), at each of the three level conditions, with four repetitions per data point, for a total of 96 runs per subject.

The experimental method was similar to that described in Chapter 3. Fundamental frequency DLs were estimated in a three-interval three-alternative forced-choice (3I-3AFC) adaptive procedure, using a two-down, one-up algorithm to track the 70.7% correct point on the psychometric function (Levitt, 1971). Two intervals contained a stimulus with a base f_0 ($f_{0,base}$) and the other interval contained a complex with a higher f_0. The listener's task was to identify the interval containing the complex with the higher pitch. The f_0 difference (Δf_0) was initially set

to 20% of the f_0, changed by a factor of 1.59 until the second reversal, and then changed by a factor of 1.26 for six more reversals. The f_0 DL was estimated as the geometric mean of the Δf_0's at the last six reversal points.

To reduce the effectiveness of loudness as an alternative discrimination cue, the root-mean-squared (RMS) power was first equalized across the three intervals by increasing the stimulus level for the interval containing the higher f_0, and then a random level perturbation was added to each interval, chosen from a uniform distribution of ±2.5 dB. In addition, $f_{0,\mathrm{base}}$ was roved from trial to trial within a run, chosen from a uniform distribution between ±2.5% of the average f_0. This was intended to encourage subjects to compare the pitches of the stimuli in each of the intervals of one trial, rather than comparing the pitch of each interval with some internally stored representation of the $f_{0,\mathrm{base}}$, although the f_0 roving may not have been effective for low f_0's where measured f_0 DLs were relatively large (8% or more).

After the f_0 DL measurements were completed, frequency DLs (FDLs) were measured for a pure tone with average frequency 1500 Hz, presented at 12.5 dB average SL in the three levels of background TEN. The 1500-Hz frequency was chosen for the FDL measurement because it represented the lower corner frequency of the bandpass-filtered harmonic tones, and was thus the most likely frequency region to yield resolved harmonics for the stimuli used in the f_0 DL measurements. FDL measurements were repeated four times at each level for each subject, using the same procedure as the f_0 DL measure, including level roving and frequency randomization.

Four normal-hearing subjects (one female) participated. Ages ranged from 22 to 30 years. All had audiometric thresholds of 15 dB HL or less *re* ANSI-1996 at octave frequencies between 250 and 8000 Hz. Two subjects (GW and MC) were professional musicians with more than ten years of formal musical training, and two (AW and AR) were amateur musicians with at least three years of musical training. Each subject completed a training period of at least four hours, which continued until FDLs and f_0 DLs no longer showed steady improvement.

The stimuli were generated digitally and played out via a soundcard (LynxStudio LynxOne) with 24-bit resolution and a sampling frequency of 32 kHz. The stimuli were then passed through a

programmable attenuator (TDT PA4) and headphone buffer (TDT HB6) before being presented to the subject via Sennheiser HD 580 headphones. Subjects were seated in a double-walled sound-attenuating chamber. Intervals were marked by colored boxes on a computer screen, and feedback (correct/incorrect) was provided following each response.

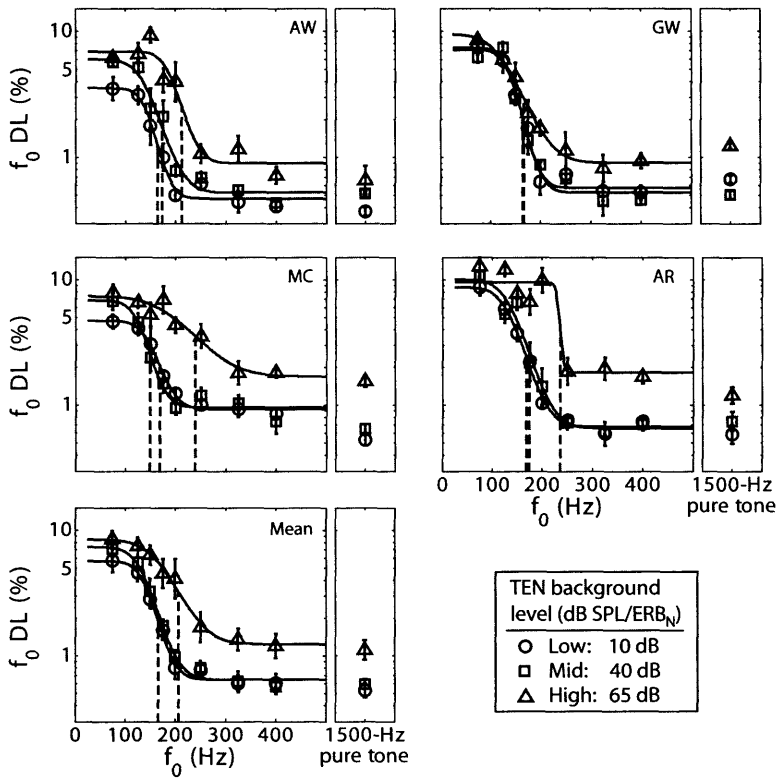

Figure 4.1. Fundamental frequency DLs as a function of f_0 and FDLs for a 1500-Hz pure tone for the four individual subjects (upper four panels) and the mean data across the four subjects (lower panel). Stimuli were presented at an average 12.5 dB SL per component in the three TEN background levels specified in the legend. Error bars indicate ±1 standard error of the mean f_0 DL across the five runs for each condition, or across the four subjects in the case of the mean data. Solid lines represent the sigmoid functions (Equation 4.1) that best fit the data at each stimulus level. Vertical dashed lines indicate the f_0 DL transition derived from the sigmoid fit, where f_0 DLs are halfway (on a log scale) between minimum and maximum.

4.3.3. Results

Figure 4.1 plots f_0 DLs and pure-tone FDLs for each of the individual subjects in the experiment (upper four panels) and the mean across the four subjects (lower panel). For all stimulus levels, f_0 DLs generally decreased with increasing f_0 (decreasing harmonic number), with a steep transition between large f_0 DLs for low f_0's and small f_0 DLs for high f_0's, consistent with previous findings discussed in the introduction (Chapters 2 and 3; Hoekstra, 1979; Houtsma and Smurzynski, 1990; Shackleton and Carlyon, 1994). There were two effects of level on f_0 DLs, both of which occurred only as the level increased from mid to high. First, there was an increase in the f_0 at which the f_0 DL transition occurred. This effect was observed in the mean data as well as for three of out the four individual subjects. In the mean data, f_0 DLs decreased to a low plateau level for a 200-Hz f_0 at the low and mid stimulus levels, but not until the f_0 reached

approximately 250 Hz at the high level. Second, both the 1500-Hz pure-tone FDL and the minimum f_0 DL (f_0 DL_{min}) achieved at the highest f_0's were elevated at the high level, an effect that was apparent in all four subjects and the mean data.

The f_0 DL data were parametrized to quantify and statistically test the two observed effects of level on f_0 DLs. A sigmoid function[1] with four free parameters was fitted to the log-transformed f_0 DLs. The sigmoid functions that best fit the f_0 DL data are depicted as solid curves in each panel of Fig. 4.1. The f_0 transition point ($f_{0,tr}$, vertical dashed lines in Fig. 4.1) was defined as the f_0 for which f_0 DLs were halfway (on a log scale) between the f_0 DL_{min} (achieved at high f_0's) and the maximum f_0 DL (f_0 DL_{max}, achieved at low f_0's) for a given stimulus level.

The effect of level on $f_{0,tr}$ and f_0 DL_{min} were analyzed statistically using bootstrap resampling to derive 95% confidence intervals (CIs) surrounding the parameter estimates (Efron and Tibishirani, 1993). For each individual subject and level, 1000 estimates of each parameter were generated, with each estimate obtained by fitting the sigmoid function to a random resampling (with replacement) of five f_0 DL estimates at each f_0. For the group data, the bootstrap estimates were generated by randomly resampling (with replacement) four individual subject mean f_0 DLs at each f_0. Two estimates were determined to be statistically significant if their 95% CIs, derived empirically from the bootstrap estimates, did not overlap. The $f_{0,tr}$ estimates are replotted as round symbols in each panel Fig. 4.2, with error bars representing the 95% CIs (the other symbols represent frequency selectivity estimates from experiments 4B and 4C, which will be described in Sections 4.4 and 4.5, respectively). Significantly different parameter estimates are identified by common small symbols along the bottom of each panel of Fig. 4.2. For example, for subject AW, the $f_{0,tr}$ was significantly different between the low and high conditions (small

---

[1] The sigmoid function was defined as

$$10 \log_{10}\left[f_0 DL\ (\%)\right] = DL_{min} + \frac{(DL_{max} - DL_{min})}{\sqrt{\pi}} \int_{m(f_0 - f_{0,tr})}^{\infty} \exp\left[-(f_0')^2\right] df_0' \qquad (4.1)$$

where DL_{max} and DL_{min} are the maximum and minimum values of $10\log_{10}(f_0$ DL (%)) achieved at very low and very high f_0's, respectively, m is the slope of the function, and $f_{0,tr}$ is the f_0 that yields an f_0 DL halfway (on a log scale) between DL_{min} and DL_{max}.

filled round symbols) and between the mid and high conditions (small open round symbols), but not between the low and mid conditions.

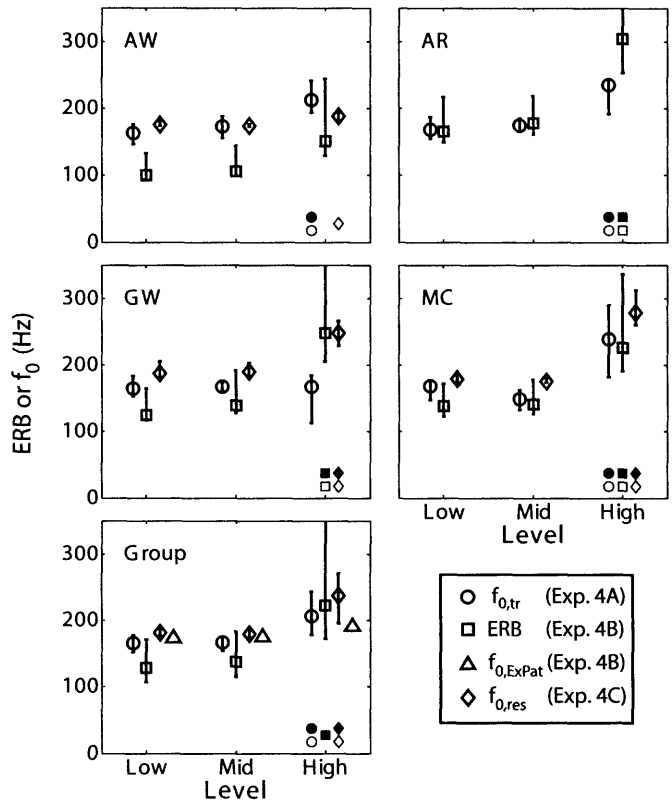

Figure 4.2. Summary f_0 DL and frequency selectivity measures for each individual subject (upper four panels) and for the data pooled across subjects (lower panel). Small symbols at the bottom of the plot underneath the high-level data indicate that a given high-level parameter estimate was significantly different from the corresponding low- (closed symbols) or mid-level estimate (open symbols). Error bars represent the 95% confidence intervals obtained from bootstrap resampling of each measure (not available for the $f_{0,ExPat}$ estimates based on the group data).

The high-level $f_{0,tr}$ was significantly larger ($p < 0.05$) than both the low- and mid-level $f_{0,tr}$'s for three out of four subjects (AW, AR and MC). Only subject GW showed no significant differences in $f_{0,tr}$ across stimulus level. In the group analysis, the high-level $f_{0,tr}$ was significantly greater than both the low- and mid-level $f_{0,tr}$, and the low/mid comparison did not show a significant difference. The log-transformed f_0 DL_{min} (Fig. 4.1) was significantly larger at the high than at the low and mid levels, both in the group data and in three out of four individual subjects. Only one subject (AR) showed no significant differences in f_0 DL_{min} across level. A similar effect was seen in the 1500-Hz pure-tone FDLs (Fig. 4.1). Fischer LSD t-tests on the mean data showed FDLs to be larger at the high level than the low level, although significant differences were not seen in the low/mid and mid/high comparisons. Fischer LSD t-tests also

showed FDLs to be significantly greater at the high level than the low level for three individual subjects (subject AW did not show this effect).

4.3.4. Discussion

The shift toward higher f_0's of the transition from large to small f_0 DLs at the high level is consistent with the hypothesis that good f_0 discrimination performance is associated with resolved harmonics. With reduced frequency selectivity at higher stimulus levels, higher f_0's would be needed to yield resolved harmonics in the considered spectral region. This hypothesis is tested further in experiments 4B and 4C by comparing the results with estimates of frequency selectivity. GW, the one subject who did not show this effect, was the subject tested at high stimulus level that was slightly lower than for the other subjects (TEN level 62 dB instead of 65 dB SPL/ERB_N), and may have shown results more similar to the other subjects if tested at the higher level.

The increase from the mid to the high level led not only to an increase in $f_{0,tr}$, but also to an increase in the f_0 DL_{min}. One possible interpretation for the increased f_0 DL_{min} is that frequency selectivity was reduced to such a degree that individual harmonics were not well resolved, even at the 400-Hz f_0, yielding poor f_0 DLs. However, two effects observed in the data argue against this conclusion. First, it appears that a plateau was in fact reached, whereby f_0 DLs no longer decreased for f_0's above 250 Hz. Second, a similar effect was observed for the FDL of the 1500-Hz pure tone, which, being the only tone present, is resolved by definition. These two observations suggest another explanation, namely that the increase in f_0 DL_{min} and FDL at the high level reflects a deterioration in the frequency coding for the individual resolved components. A similar effect of level on FDLs for pure tones presented in background noise was observed by Dye and Hafter (Dye and Hafter, 1980) at higher frequencies. Possible implications of this finding are addressed in the General Discussion.

4.4. Experiment 4B: Auditory filter shapes

4.4.1 Rationale

The central hypothesis of this study was that the increase in auditory filter bandwidths at higher signal levels would decrease the number of available resolved harmonics, thereby shifting the

transition from large to small f_0 DLs to higher f_0's. This experiment was designed to verify the first part of this hypothesis, that auditory filter bandwidths increase with increasing level, which we expect given previous results (Weber, 1977; Pick, 1980; Moore and Glasberg, 1987; Rosen and Stock, 1992; Hicks and Bacon, 1999). Measurements of auditory filter bandwidths in this experiment allowed us to compare the predicted variation in frequency selectivity with stimulus level to compare with the f_0 DLs measured in the same subjects in experiment 4A.

A version of the notched-noise method (Patterson, 1976) described by Rosen and Baker (1994) measured the level of a notch noise that just masked a pure tone presented at a constant level, as a function of the width of the noise's spectral notch. A model auditory-filter shape was then fitted to the data. Stimulus levels and durations were similar to those of experiment 4A to ensure that the auditory filter shapes estimated in this experiment were as similar as possible to those presumably used by subjects for f_0 discrimination in the previous experiment. Auditory filter shapes were estimated using simultaneous rather than forward masking to mimic the simultaneous masking between components that occurs with the simultaneous presentation of the harmonics of a complex.

4.4.2. Methods

The notched-noise level that just masks a pure tone was measured as a function of the masker notch width. Throughout the experiment, the pure-tone signal had a constant frequency (f_{sig}) of 1500 Hz, corresponding to the low-frequency edge of the passband in experiment 4A, where harmonics were most likely to be resolved. Three level conditions were tested (low, mid, and high), whereby the signal was fixed at the SPL level corresponding to 10 dB SL (adjusted for each subject) re. one of the TEN levels that was used in experiment 4A. This signal level was at the lower end of the 10-15 dB SL per component level range that was used in experiment 4A. Although the signal SPL was adjusted relative to the detection threshold in TEN, the TEN background was not used in this experiment.

Each trial in the experiment consisted of three intervals, each with a 700-ms duration, separated by 500-ms silent gaps. Two of the intervals contained only a 700-ms noise burst (including 10-ms raised-cosine onset and offset ramps). The other interval also contained a 500-ms pure-tone

signal (including 30-ms raised-cosine onset and offset ramps), temporally centered within the noise burst. The listeners' task was to identify which of the three intervals contained the pure-tone signal. A 3I-3AFC procedure with a two-up, one-down adaptive algorithm tracked the 70.7% correct point (Levitt, 1971). The spectrum level of the noise was initially set to -25, 5, and 30 dB SPL/Hz, in the low, mid, and high conditions, respectively, and changed by 8 dB for the first two reversals, 4 dB for the next two reversals, and 2 dB for the last eight reversals. Threshold was estimated as the mean of the noise levels at the last eight reversal points.

To reduce the overall level of the masking noise at a given masker spectrum level, the two bandpass noises making up the noise masker had narrower bandwidths than in the Rosen and Baker (1994) study (200 Hz, or $0.13f_{sig}$, compared to $0.8f_{sig}$ used by Rosen and Baker). The notch width was defined in terms of the deviations from the signal frequency, expressed as a proportion of f_{sig}, of the high-frequency edge of the lower-frequency noise band (Δf_l) and the low-frequency edge of the upper-frequency noise band (Δf_u). The maximum notch deviation relative to the signal frequency was also limited relative to the Rosen and Baker study ($\pm 0.2f_{sig}$ as compared to $\pm 0.4f_{sig}$). The limited range of notch widths reduced our ability to estimate the filter tail shapes, but was necessary to avoid the uncomfortably loud masker levels that would have been necessary to mask the signal at wider notch widths. Three symmetrical notch conditions were tested, with equal Δf_l and Δf_u values of 0, 0.1 and $0.2f_{sig}$. To allow for the possibility of asymmetrical filters, there were also two asymmetric conditions, one with $\Delta f_l = 0.1f_{sig}$ and $\Delta f_u = 0.2\,f_{sig}$, and the other with $\Delta f_l = 0.2\,f_{sig}$ and $\Delta f_u = 0.1\,f_{sig}$. An additional modification to the Rosen and Baker (1994) paradigm was the addition of a low-pass noise to mask any possible low-frequency combination bands (Greenwood, 1972) that could facilitate the detection of the signal. The low-pass noise had a cutoff frequency equal to the low-frequency edge of the lower-frequency noise band and a spectrum level 20 dB below that of the notched noise.

The same subjects took part in this experiment as had taken part in experiment 4A. All subjects underwent a short training period where one masked threshold for each of the 15 (five notch conditions at three levels) conditions was estimated. Subjects then completed four measurements for each data point, for a total of 60 runs.

4.4.3. Results and discussion

A standard fitting procedure was used to derive auditory filter shapes (Glasberg and Moore, 1990). Because of the small number of conditions tested (5 notch widths x 3 stimulus levels = 15 conditions), it was desirable to limit the number of free parameters in the filter shape model that were used to fit the data. This was accomplished by assuming the filter-tip shape (defined by the slope p) to be symmetrical and invariant across stimulus level, and assuming that both p and k (the signal-to-noise ratio in the filter output required for signal detection) were level invariant. Thus, the only parameter that was allowed to vary across level was the dynamic range limitation (r). The dynamic range limitation was applied only to the low-frequency side of the filter, representing the wide low-frequency tails observed in auditory-nerve fiber tuning curves (Kiang *et al.*, 1965). With these constraints, the 15 data points per subject were fit with five free parameters: p, k and three values of r. Limiting the number of free parameters to five only marginally sacrificed the overall goodness of fit. The root-mean-squared (rms) fitting error (resulting from fitting each the individual subject's data with a separate set of filters) was 1.95 dB in the five free parameter case and 1.45 dB in the case where r, p and k were all allowed to vary with level (twelve free parameters).

The assumed level invariance of the filter tip is similar to the approach taken by Glasberg *et al.* (1999) and Glasberg and Moore (2000), whereby the filter tail and high-frequency slope of the filter tip were held constant across level, and only the gain of the filter tip and its low-frequency slope varied with stimulus level. In the present study, the dynamic range limitation, r, models the relative gain of the filter-tip relative to that of the flat filter tail. The use of a level invariant filter-tip shape models the physiologically-observed two-part response of the basilar membrane (Ruggero, 1992; Ruggero *et al.*, 1997), with one broadly tuned linear filter corresponding to the passive mechanical properties of the basilar membrane, and one narrowly tuned variable-gain filter representing the active mechanism thought to be governed by the outer hair cells (Ruggero and Rich, 1991).

The fitting procedure took into account the Sennheiser H580 transfer function, the middle-ear transfer function, and the possibility of off-frequency listening and variations in filter bandwidth with CF, as described by Glasberg and Moore (1990). Although the filter tip was assumed to be

symmetrical, the asymmetrical application of the dynamic-range limitation and the combination of off-frequency listening and proportional variation in filter bandwidth with CF accounted for unequal threshold measurements in the two asymmetric notch conditions

The means and standard deviations (across the four subjects) of each of the best fitting model parameters are listed in Table 4.1[1]. Figure 4.3 shows the fitted filter shapes for each stimulus level based on the mean filter parameters shown in Table 4.1. The main finding of this analysis is that the filter shape changed continuously with increasing level: the filter's dynamic range decreased as the stimulus level increased from the low to the mid to the high level. This trend is notably different from that observed in experiment 4A, where f_0 DLs remained unchanged as the stimulus level increased from low to mid and then increased as the stimulus level increased from the mid to the high level.

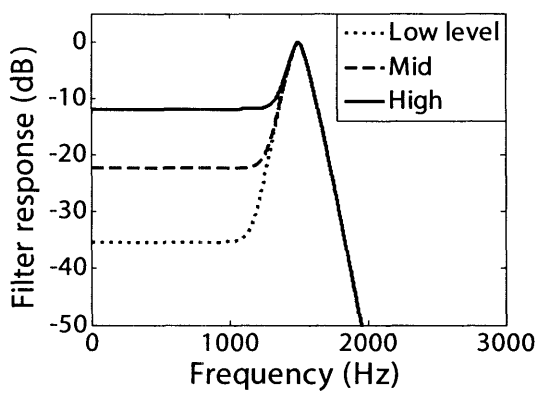

Figure 4.3. Auditory filters at the three stimulus levels, based on the mean of each of four best-fitting filter parameters (p and three values of r) across the four subjects.

Parameter	Level	Mean Value	Standard Deviation
p	All	45.61	11.26
r	Low	-48.05	22.83
	Mid	-22.90	2.53
	High	-12.25	2.07
k	All	-1.68	1.01

Table 4.1. Means and standard deviations across subjects of the filter-model parameters that best fit the notched-noise data of experiment 4B.

4.4.4. Comparing f_0 DLs to estimates of frequency selectivity

To directly test the hypothesis that small f_0 DLs are associated with resolved harmonics, each estimate of the $f_{0,tr}$ (see section 4.3.3) was compared to two estimates of frequency selectivity based on the fitted filter shapes of experiment 4B. First, the equivalent rectangular bandwidth

[1] The set of filter parameters that best fit the masking data pooled across subjects (not shown) were similar to the mean of each filter parameters fit to the individual subjects shown in Table 4.1.

(ERB) of the fitted filter shapes provided a rough estimate of frequency selectivity (Section 4.4.4.1). Second, harmonic resolvability was estimated based on peak-to-valley ratios (PVRs) in excitation patterns derived from the fitted filter shape functions and the spectra of the harmonic stimuli (Section 4.4.4.2).

(4.4.4.1) Equivalent rectangular bandwidth (ERB)

Filter ERBs, calculated by integrating the fitted filter shape across frequency (Hartmann, 1998), are plotted as square symbols in Fig. 4.2, with error bars indicating the 95% CIs derived from 1000 bootstrap estimates of each ERB. As for the $f_{0,tr}$ and f_0 DL_{min}, there was little difference between the low- and mid-level ERBs, but an increase in the ERB from the mid to the high level. This was confirmed statistically, where ERBs were significantly greater ($p < 0.05$) at the high than the mid level for three out of four subjects, with only subject AW showing no significant across-level ERB differences. (The intersubject variability in ERBs did not match that observed in the $f_{0,tr}$, where only subject GW did not show significant effects of level). When the data were pooled across subjects, ERBs were significantly larger at the high than the low level. The difference in ERB between the low and mid conditions was not significant, and the difference between mid and high, although generally apparent in the individual and mean data, failed to reach significance.. Overall the similar pattern of results for the effect of level on f_0 DLs and ERBs in the mean data supports the idea that f_0 discrimination performance is related to frequency selectivity and that the worse f_0 discrimination performance observed at the high level is related to the reduction in frequency selectivity associated with these stimuli.

ERBs (Fig. 4.2, square symbols) varied with level in a different manner than the auditory filter shapes (Fig. 4.3). Whereas auditory filter shapes change consistently with level, while the resulting ERBs remain unchanged until the stimulus level increased above the mid level. This result is related to the fact that only the dynamic range limitation (r) was allowed to vary with level. When r has a large negative value (in dB), the tail of the filter has little effect on the overall ERB. It is not until the dynamic range of the filter decreases substantially that the energy in the tail of the filter begins to affect the filter's ERB.

(4.4.4.2) Excitation Pattern Model

The ERB measure provides a general statistic describing the frequency selectivity of the auditory system for comparison with the f_0 discrimination data. To test the hypothesis more directly that good f_0 discrimination is associated with resolved harmonics, harmonic resolvability was estimated based on the PVRs in peripheral excitation patterns calculated using the fitted filters described above. The filterbank was produced using the mean filter parameters that best fit the notched-noise masking data for individual subjects (Table 4.1). The filter parameter estimates for the 1.5-kHz CF were assumed to be scalable to other CFs along the cochlear partition, such that filters were identical on a logarithmic frequency scale across CF. The filterbank consisted of 501 model filters with CFs logarithmically spaced between 100 Hz and 10 kHz. Excitation patterns were calculated in the spectral domain, such that the excitation for each CF in the filterbank was the output power of the filter in response to the power spectrum of the harmonic stimulus plus background noise. The background noise was set at 10, 40 or 65 dB SPL/ERB_N and the signal level was set at 12.5 dB SL, where the 0 dB SL reference was averaged across the four subjects. A different excitation pattern was produced for each f_0 and stimulus level that was tested in experiment 4A, with the appropriate filterbank used at each stimulus level.

Sample excitation patterns for the mid-level conditions are shown in Fig. 4.4 for 75-, 200- and 400-Hz stimulus f_0's. For the lowest f_0 of 75 Hz, there are no discernable peaks present in the excitation pattern, because the frequency spacing between adjacent harmonic components is too narrow for the harmonics to be spectrally resolved by the filter bank. As the f_0 increases, peaks in the excitation pattern appear and become more prominent as individual components become increasingly spectrally resolved. The PVR quantified the degree to which harmonics were resolved by the filterbank. The PVR was measured between the first peak in the excitation pattern occurring at a CF $\geq$1.5 kHz (vertical dashed lines in Fig. 4.4), and the valley at a higher CF immediately adjacent to the peak.

Figure 4.5 shows PVRs as a function of f_0 for each stimulus level. PVRs are roughly equal for the low and mid stimulus levels, and are smaller at the high stimulus level, a trend similar to that observed for the ERB and $f_{0,tr}$ estimates. To directly compare the PVRs to the f_0 DL data, a threshold PVR (PVR_{th}), defined as the minimum PVR that yielded resolved harmonics, was

varied as a free parameter to fit the PVR estimates to the $f_{0,tr}$ estimated derived from the pooled f_0 DL data of experiment 4A (round symbols in Fig. 4.2, lower panel)[2]. The threshold PVR (PVR_{thres}, horizontal dashed line in Fig. 4.5) was adjusted to minimize the least squares difference between the the f_0 needed to achieve the PVR_{th} (termed $f_{0,ExPat}$, vertical dashed lines in Fig. 4.5) and the $f_{0,tr}$ estimates derived from the f_0 DL data. PVRs for f_0's between those tested in experiment 4A were linearly interpolated as shown in Fig. 4.5. The fitted PVR_{th} represents the estimate of the PVR needed in the excitation pattern to yield f_0 DLs halfway (on a log scale) between f_0 DL_{max} and f_0 DL_{min}.

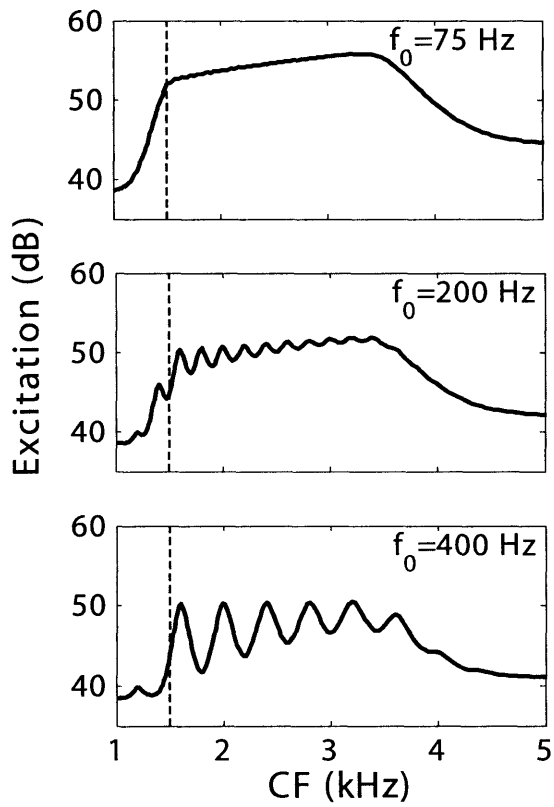

Figure 4.4. Sample excitation patterns for three f_0's presented at the mid level. PVRs (Fig 4.5) were calculated based on the first peak occurring at a CF greater than 1500 Hz (dashed lines).

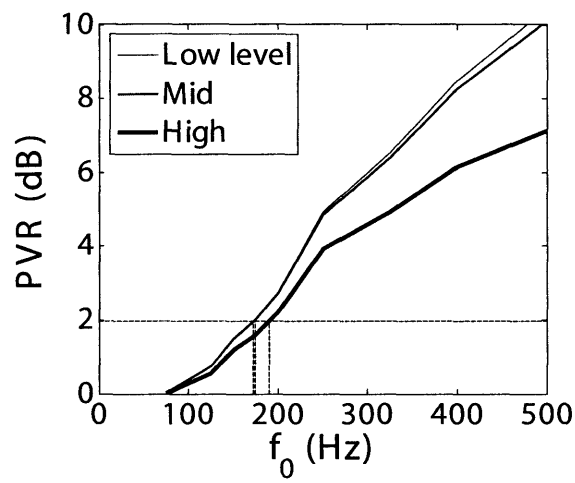

Figure 4.5. PVRs as a function of stimulus f_0 for the first excitation pattern peak occurring at a CF of 1500 Hz or higher. The horizontal dashed line indicates the PVR_{th} = 1.98 dB that minimized the mean-squared difference between the corresponding $f_{0,res}$ (vertical dashed lines and Fig 4.2, lower panel, triangles) and the $f_{0,tr}$ estimates (Fig. 4.2, lower panel, round symbols) derived from the f_0 DL data.

[2] Excitation pattern model fits were not performed for each individual subject. Since the intersubject variability in the $f_{0,tr}$ did not match the intersubject variability in the ERB, we would not expect the excitation pattern model based on the same data to closely fit the f_0 DL data on an individual subject basis.

The estimated $f_{0,\mathrm{ExPat}}$ at which the PVR was equal to the best-fiting $\mathrm{PVR_{th}}$ of 1.98 dB are plotted as triangles in the lower panel of Fig. 4.2. The pattern of results was qualitatively similar to the effect of stimulus level on the $f_{0,\mathrm{tr}}$, with $f_{0,\mathrm{ExPat}}$ remaining roughly constant between the low and mid levels (172.8 and 174.2 Hz, respectively), but increasing at the high level (190.8 Hz). The same trend was seen in the PVR versus f_0 plots of Fig. 4.5. As with the ERB, the changes in the filter tail did not affect the PVR until the high level. Quantitatively, the excitation pattern model did not show as large of an effect of level on the $f_{0,\mathrm{ExPat}}$ as was observed in the $f_{0,\mathrm{tr}}$ (approximately a 10% and 25% change, respectively, from the mid to the high level).

.

(4.4.4.3) Allowing p to vary with level

Several filter-shape models that allowed the filter tip to vary with level were also investigated (results not shown). Three such models variations were tested: (1) symmetrical tip and dynamic range limitation, (2) asymmetrical tip and dynamic range limitation, and (3) asymmetrical tip and no dymaic range limitation. For all three variations, the filter ERB increased regularly with increasing level, inconsistent with the trend observed in the $f_{0,\mathrm{tr}}$. The trend in frequency selectivity was consistent with the trend in $f_{0,\mathrm{tr}}$ as a function of level only when the filter tip was held constant across level as described in Section 4.4.3.

4.5 Experiment 4C: Hearing out harmonics

4.5.1 Rationale

This experiment measured the ability of subjects to hear out the frequency of individual harmonics. A method similar to that of Chapter 2 measured performance in discriminating the frequency of a target harmonic embedded in a complex from that of a pure tone presented in isolation. The target harmonic was gated on and off repeatedly in order to draw listeners' attention to it without affecting peripheral resolvability. This

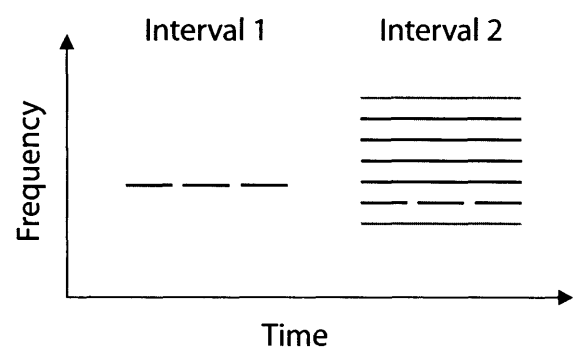

Figure 4.6. Schematic of the stimulus paradigm used in experiment 4C. Subjects compared the frequencies of the gated comparison tone presented in isolation (Interval 1) with the frequency of a gated harmonic component of a bandpass filtered complex (Interval 2). Shading represents the amplitude of each frequency component; components falling within the filter skirt are shown in lighter greyscale.

approach was successful in our earlier study, estimating that approximately 9 to 11 harmonics are resolved for 100- and 200-Hz tone complexes, a number that closely corresponded to the transition from large to small f_0 DLs.

4.5.2. Methods

A schematic of the stimulus paradigm is shown in Fig. 4.6. Each trial consisted of two intervals, each with duration 500 ms, separated by 375 ms. The second interval contained a bandpass-filtered harmonic complex, identical to that of experiment 4A, except that one harmonic (the target tone) was gated on and off in time, with three bursts of a 150-ms sinusoid (comparison tone), including 30-ms raised-cosine onset and offset ramps, separated by 25-ms silent gaps. The first interval contained a single stimulus frequency (the comparison tone) gated on and off in the same manner as the target tone. Harmonic complexes were presented in random phase. This reduced the possibility that the frequency of an unresolved harmonic could be detected based on the Duifhuis (1970) effect, whereby a sinusoid at the frequency of the missing harmonic may appear in the waveform during the temporal dips associated with sine-phase complexes (see Chapter 2, footnote 2, for a discussion). Both intervals were presented in the same wideband TEN background as experiment 4A, which was turned on 250 ms before the start of the first interval, and turned off 250 ms following the end of the second interval. Each component (before filtering, where applicable) was presented at 12.5 dB SL (adjusted for each subject). Level randomization was not used in this experiment, because loudness variations would not have provided a useable cue. The frequency of the comparison tone (f_{comp}) was 3.5% either higher or lower (each with probability 0.5) than the frequency of the target tone (f_{targ}) . A two-alternative forced-choice task was used, where the listener was required to discriminate whether the target or comparison tone was higher in frequency. Feedback was provided following each response.

The target harmonic was chosen such that its nominal frequency fell between 1600 and 1750 Hz. For six of the nominal f_0's, only one harmonic fell in this range. For the 75- and 125-Hz f_0's, two harmonics fell within this range, and the total number of trials were evenly divided between the two possibilities. The limited range of f_{targ}'s created the possibility that listeners could obtain correct responses based on the absolute frequency of the comparison tone alone, without

comparing it to the frequency of the target tone. In fact, an f_{comp} below 1600 Hz or above 1750 Hz could only be lower or higher, respectively, than an f_{targ}. Three steps were taken to reduce the likelihood that subjects would use such a strategy. First, the presentation order of the eight nominal f_0 values of the complex was randomized. Second, the actual value of the f_0 was roved, with the frequency of the target component chosen from a uniform distribution ranging from 50 Hz below to 50 Hz above the nominal target frequency, and the f_0 set accordingly. Third, dummy trials were added in such a way that the probability across dummy and non-dummy trials that a given comparison tone frequency was higher than the target tone was roughly 50% for all possible comparison-tone frequencies. For example, in the f_{comp} > 1750 Hz range where in the non-dummy trials f_{comp} would always be larger than f_{targ}, dummy trials were added where a similar f_{comp} was always lower than f_{targ}. The dummy trials were selected to have a mix of f_0's, comprising unresolved (f_0 <100 Hz), partly unresolved (f_0 = 150 to 200 Hz), and mostly resolved (f_0 >300 Hz) harmonics. The dummy target tones, which sometimes had frequencies below 1500 Hz, were not subjected to the slope of the bandpass filter. If listeners were responding based on the f_{comp} alone, then responses would have been biased toward "lower" and "higher" for low and high f_{comp}, respectively. However, an analysis of the data (not shown) indicated that this was not the case. Across all dummy and non-dummy trials, listeners responded "lower" or "higher" with a probability of roughly 0.5 across the entire range of f_{comp}'s presented.

Each run consisted of 72 trials for one stimulus level condition. There were 48 non-dummy trials (six trials for each of the eight f_0's tested in experiment 4A), plus 24 dummy trials (two trials each for twelve combinations of f_0 and target frequency). There were seventeen runs for each of the three stimulus levels, for a total of 102 non-dummy trials for each f_0 and stimulus level. The stimulus level for each run was randomly selected, without replacement, until three runs were completed, then the process repeated.

Three of the four subjects from experiments 4A and 4B participated in this experiment (subject AR did not participate). Each was given at least one hour of additional training.

4.5.3. Results and Discussion

The percentage correct as function of f_0 and level are plotted for each of the three individual subjects and for the mean data in Fig. 4.7. For each subject and stimulus level, there was a transition from chance performance (or below) at the lowest f_0's to near-perfect performance for the highest f_0's, consistent with the interpretation that harmonics are unresolved for low f_0's, and resolved for high f_0's. To compare with the f_0 DL data of experiment 4A, the data were fit to a psychometric function fixed to 50% and 100% correct at the extremes (solid curves in Fig. 4.7), and the f_0 required to reach 75% correct was defined as the limit of harmonic resolvability, termed the $f_{0,res}$ (vertical dashed lines in Fig. 4.7).

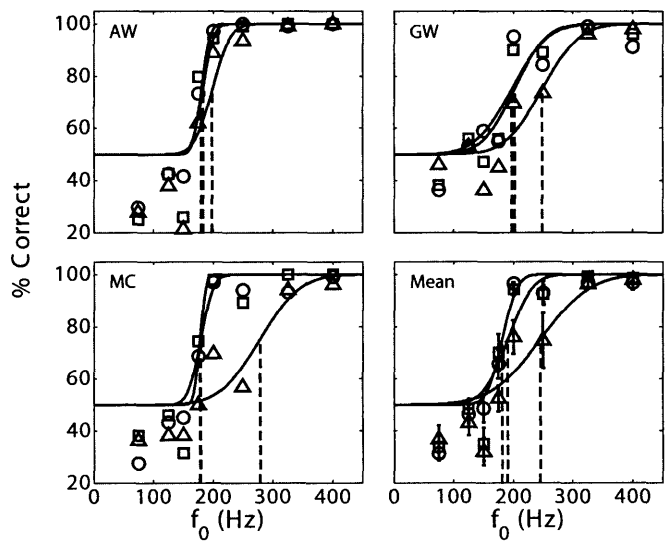

Figure 4.7. Results of experiment 4C showing the percent correct in hearing out the frequency of an individual harmonic in the 1600-1750 Hz range as a function of stimulus f_0 for three stimulus levels for each individual subject (two upper panels and lower left panel) and for the mean data across the three subjects who participated in this experiment (lower right panel). Solid lines indicate the sigmoid function that best fit the data at each stimulus level, while vertical dashed lines represent the estimate of the limit of harmonic resolvability ($f_{0,res}$) based on the 75% correct point. Error bars for the mean data represent ± one standard error across the three subjects.

The $f_{0,res}$ estimates are plotted as diamonds in each panel of Fig. 4.2 for direct comparison with the previous experiments. Bootstrap estimates of the 95% confidence intervals (error bars in Fig. 4.2) were performed for each subject by random resampling (with replacement) from the 102 responses (correct/incorrect) at each f_0. The estimated $f_{0,res}$ were significantly larger at the high than the mid level for all three subjects that participated in the experiment (and significantly larger at the high level than the low level for two subjects, GW and MC).

In the group data (lower panel of Fig. 4.2), the effect of level on the $f_{0,res}$ was similar to the level effects observed in the previous experiments. As for the $f_{0,tr}$ (circles) and ERB (squares), the $f_{0,res}$ (diamonds) remained constant at the low and mid levels, and then seemed to increase for the high level. The observed behavior of $f_{0,res}$ with respect to level was supported by a bootstrap

statistical analysis of the data pooled across subjects, where 1000 estimates of the $f_{0,res}$ at each level were produced by resampling the group data across f_0's (95% CIs denoted by error bars in the lower panel of Fig. 4.2). The $f_{0,res}$ was significantly greater at the high level than at either the low or mid levels, and there was no significant difference in $f_{0,res}$ between the low and mid conditions. The similarity in the overall pattern of results across the three experiments provides further evidence that the effects of stimulus level on f_0 DLs are related to frequency selectivity and the resolvability of individual harmonics.

Finally, below-chance performance was observed for the lowest f_0's suggesting that listeners may have had access to information regarding the frequency of the gated target harmonic, but that they used this information incorrectly. Similar below-chance performance was sometimes observed for similar conditions in Chapter 2. One possibility is that subjects were comparing the pure-tone frequency to that of a harmonic adjacent to the gated harmonic that became unmasked during the "off" intervals. However, the fact that subjects were not able to use this cue to produce better performance, despite feedback, suggests that the cue was unreliable and was fundamentally different from that used at the high f_0's, presumably reflecting a difference in harmonic resolvability.

4.6. General discussion

4.6.1. The relationship between f_0 discrimination and frequency selectivity

The results of experiment 4A show that increasing the stimulus level has two detrimental effects on f_0 discrimination for harmonic complexes in the 1.5 to 3.5 kHz range tested in this experiment. First, increasing the stimulus level can lead to an increase in the harmonic spacing needed to reach f_0 DL_{min}; second, even if harmonic spacing is wide (or infinite, in the case of the pure tone), frequency discrimination performance becomes worse with increasing stimulus level when the tones are presented in a background noise to keep their SL constant. That both of these aspects of the f_0 DL data were affected by level in the same way as the ERB and hearing out harmonics data suggests that both f_0 DL level effects may be related to a loss of frequency selectivity at the high stimulus level.

Some intersubject variability was observed in the experimental results, in that not all subjects showed an effect of level for each f_0 discrimination and frequency selectivity measure. Moreover, on an individual subject basis the lack of an effect of level in one measure did not correspond to the lack of an effect of level in another. There was no significant effect of level on the $f_{0,tr}$ and f_0 DL_{min} for subjects GW and MC, respectively, whereas subject AW showed no significant effect of level on the ERB. Nevertheless, each of these measures behaved similarly with respect to level when the data was averaged across subjects.

As pointed out by Shackleton and Carlyon (1994), a quantitative measure of the limit of harmonic resolvability may depend on the task, and the detection criteria used in any given experiment. The current study to some extent avoids the problem of comparing results from different paradigms by comparing patterns of results, as a function of an independent variable (in this case level), rather than single values. Both the excitation pattern model (experiment 4B) and the hearing out harmonics paradigm (experiment 4C) found that harmonic resolvability depended on stimulus level in a similar way to the f_0 DL transition point in experiment 4A. There was little or no change from the low to the mid stimulus level, but an approximately 15% increase (from around 175 Hz to 200 Hz) in both the minimum f_0 for which harmonics could be heard out and the f_0 DL transition point from the mid to the high stimulus level. That the two estimates of harmonic resolvability and the f_0 DL transition point depended on stimulus level in the same way provides support for the hypothesis that resolved harmonics are associated with good f_0 performance.

The results shown here are in apparent conflict with the results of Krumbholz *et al.* (2000) and Pressnitzer *et al.* (2001), who found that the harmonic number associated with the $f_{0,tr}$ increased sharply with increasing lower stimulus cutoff frequencies up to 1 kHz. This effect would not be predicted by spectral and spectrotemporal pitch models where the $f_{0,tr}$ is determined by the auditory filter bandwidth and harmonic resolvability. This apparent discrepancy may reconciled if for these low f_c's, the $f_{0,tr}$ is limited by an absolute minimum repetition rate, on the order of 30-50 Hz, depending on the f_0 DL criterion that defines the $f_{0,tr}$. For f_c's greater than 1 kHz, the $f_{0,tr}$ falls well above this minimum repetition rate and would instead be governed by harmonic

resolvability limitations. This implies that both spectral (current study) and temporal factors (Krumholz *et al.*, 2000; Pressnitzer *et al.*, 2001) may limit f_0 discrimination performance.

Finally, the results from the present study can be compared to those of Chapter 2, where an increase in harmonic resolvability by presenting even and odd harmonics to opposite ears (dichotic presentation) failed to improve f_0 discrimination performance. Taken together, the results can be interpreted as implying that resolved harmonics may be *necessary* for good f_0 discrimination performance (current experiments), but not *sufficient* (Chapter 2). These apparently contradictory results could be reconciled in one of several ways. Two possibilities (discussed in Chapter 2) are that f_0 is derived by comparing the stimulus spectrum to internally stored harmonic templates that only contain the first approximately 10 "normally" resolved harmonics, or that f_0 is derived from the "central spectrum" (Zurek, 1979) where additional resolved harmonics may not be available under dichotic presentation. Another possibility is that the spectral f_0 discrimination limitation is related to the frequency selectivity of the auditory system but not to harmonic resolvability *per se*. For example, f_0 discrimination may be limited by the cochlear filter impulse response (de Cheveigné and Pressnitzer, 2005, see Section 4.6.2.1), a quantity that is known to covary with filter bandwidth .

4.6.2. Implications for theories of pitch perception

The results presented here, identifying a correspondence between f_0 discrimination performance and frequency selectivity, have implications for models of pitch perception, as discussed below.

(4.6.2.1) Temporal Models

"Temporal" models of pitch perception estimate f_0 using ANF temporal firing patterns. Chapter 3 showed that a temporal autocorrelation model of pitch (Meddis and O'Mard, 1997) can account for the effect of harmonic number on f_0 discrimination performance if it is modified to include CF dependence, thus rendering it no longer a *purely* temporal model. In the modified model, "lag windows" limited the range of f_0's to which the model would respond relative to a given channel's CF, thereby forcing a dependence on harmonic number. However, this successful behavior was not a consequence of harmonic resolvability or frequency selectivity. Instead, the f_0 discrimination performance predicted by the model depended mainly on the ratio between CF

and f_0. Therefore, this model may not predict a detrimental effect of increased filter bandwidths on f_0 discrimination performance. The results of the current study suggest that a model for which f_0 discrimination depends on harmonic number and not on frequency selectivity *per se* is not sufficient.

A CF-dependent autocorrelation model similar to the one described in Chapter 3 would be more likely to predict the effects of stimulus level on the f_0 DL transition point if the model's lag windows were a direct consequence of peripheral frequency selectivity. This has been proposed in a novel implementation of the autocorrelation model by de Cheveigné and Pressnitzer (2005), who demonstrate how the long delays that are needed to compute an autocorrelation function (e.g. <10 ms for f_0's lower than 100 Hz) could be effectively achieved by a combination of short phase shifts across multiple channels with a common CF. A consequence of their formulation is that the range of lags that can be achieved for a given CF is limited by the duration of the impulse response, and therefore the bandwidth of the filter. Thus, in their implementation it is likely that the ability to efficiently code f_0 will depend on filter bandwidth, and hence on level, in a way similar to that found in our data. Furthermore, as pointed out by de Cheveigné and Pressnitzer (2005), their model is also likely to account for the results of Chapter 2 indicating that resolved harmonics are not *sufficient* for good f_0 discrimination performance. This is because the model's dependence on harmonic number derives from the frequency selectivity of the auditory periphery, which would not be affected by the presentation of even and odd harmonics to opposite ears, but not from the auditory system's ability to resolve individual harmonics.

Because temporal models, including the CF-dependent temporal model of de Cheveigné and Pressnitzer (2005), do not use the spectral resolution of the system to directly extract frequency information, it is not clear whether they could to account for the increase in f_0 DL_{min} and pure-tone FDLs observed at the high stimulus level based on a loss of frequency selectivity. One possibility is that the reduced frequency selectivity at high stimulus levels could have disrupted the temporal coding of individual frequencies by allowing additional background noise energy to pass through the filter. However, the stimulus level was set relative to the detection threshold for a pure tone at each background noise level, such that any additional noise energy passing through a filter was accompanied by a corresponding increase in the level of each frequency component

to maintain a constant detection threshold. Physiological evidence shows that when tones are presented in a background noise, the degree of phase locking is mainly dependent on the SNR, remaining constant when the signal and noise levels are manipulated simultaneously to maintain constant SNR (Rhode *et al.*, 1978; Abbas, 1981).

Temporal or CF-dependent temporal) models may be able to account for the increase in the f_0 DL_{min} and pure-tone FDL in terms of a deterioration of temporal coding at high stimulus levels. Unlike spectral information based on ANF firing rates (see Section 4.6.2.2), ANF phase locking does not generally deteriorate at high levels, (Johnson, 1980), even in the presence of a background noise (Rhode *et al.*, 1978; Abbas, 1981), implying that a temporal code is unlikely to explain the high-level increase in pure-tone FDLs and f_0 DL_{min} in terms of coding by different spontaneous-rate ANF populations (Liberman, 1978). However, a temporal code could deteriorate at high levels via a phenomenon known as peak splitting. At high stimulus levels, ANFs can sometimes fire during the condensation or rarefaction portion of the sinusoidal stimulus cycle, thus reducing the cycle-by-cycle precision in temporal fidelity (Johnson, 1980). In the absence of a background noise, where pure-tone FDLs generally improve with level (e.g. Wier *et al.*, 1977), peak splitting might not affect psychophysical FDLs at high levels due to the availability of off-frequency channels not affected by this phenomenon. However, this effect has been shown to occur mainly at low frequencies (below 1 kHz), where pure-tone FDLs do not increase with stimulus level (Dye and Hafter, 1980).

(4.6.2.2) Spectral models

Most spectral models of pitch are based on the concept that individual resolved frequencies are first identified and then compared to an internally stored template to derive the f_0. Models in this category are generally consistent with the observed shift in the f_0 DL transition point toward higher f_0's at a higher stimulus level. With increased filter bandwidths, higher f_0's will be needed to yield the increased separation between adjacent partials needed for resolved harmonics. Of course, all models that require resolved harmonics fail to predict the (albeit poor) pitch perception elicited by unresolved harmonics, and would therefore require a separate temporal envelope pitch extraction mechanism to account for these percepts.

The increase in f_0 DL_{min} with stimulus level is similar to the increase with level in FDLs for a 4-kHz pure tone presented near the 85% correct detection threshold in a background noise (Dye and Hafter, 1980). In that study, FDLs increased with level at 4 kHz, but decreased at 500 Hz or 1 kHz. This was interpreted in terms of reduced phase locking at 4 kHz, leading to frequency being encoded via spectral cues that are affected by the reduction in frequency selectivity with level. Although 1.5 kHz is lower than the frequency limits normally associated with a roll-off in phase locking in frequency discrimination (e.g., Moore and Sek, 1995), the similarity in the effects of level for the 4-kHz tone in the Dye and Hafter (1980) study and both the high-f_0 f_0 DLs and the 1.5-kHz pure-tone FDLs in the current study suggests a role for spectral or spectro-temporal (as opposed to purely temporal) encoding in the 1.5 to 3.5-kHz frequency region tested in the current experiment. Whether a purely spectral model (e.g. Goldstein, 1973; Wightman, 1973; Terhardt, 1974,1979) could successfully account for the effect of level on f_0 DL_{min} and the pure-tone FDL would depend in part on the exact implementation. For instance, if a change in frequency is detected by a change in excitation on the steep low-frequency slope of the excitation pattern, as in Zwicker's (1970) model for frequency modulation detection, then the predicted effects of level may remain small, because the slope of the upper-frequency skirt of the auditory filter (which determines the low-frequency slope of the excitation pattern) is largely level invariant (Ruggero *et al.*, 1997; Glasberg *et al.*, 1999). A spectral model might also explain the increase in terms of coding by different spontaneous-rate ANF populations (Liberman, 1978). At high stimulus levels, the relatively large population of high spontaneous-rate fibers would become saturated, leaving only a relatively small population of medium or low spontaneous-rate fibers for rate-based frequency encoding.

(4.6.2.3) Harmonic-template spectrotemporal models

This class of pitch model uses a combination of both place and temporal information to extract the individual frequencies of the harmonic components and calculate the f_0. Because of this, they would be likely to correctly predict an increase in $f_{0,tr}$ with increasing level. In one type of model in this class, frequency information is extracted from the rapid transitions in the filter phase response near CF (Shamma, 1985; Cedolin and Delgutte, 2005a). These models might also explain the observed effect of level on f_0 DL_{min}, in that phase transitions, as measured on the basilar membrane tend to become more gradual with the increased filter bandwidths at high

stimulus levels (Rhode and Cooper, 1996; Ruggero *et al.*, 1997) thereby reducing pitch discrimination performance for resolved harmonics and pure tones.

Carney (1994) in examining the possibility that the filter phase transitions encode stimulus level, posited that the slope of the phase response changes continuously with stimulus level over the wide dynamic range (approx. 30 to 100 dB SPL) of the basilar membrane compressive nonlinearity (Yates *et al.*, 1990; Ruggero, 1992; Ruggero *et al.*, 1992). Such a continuous change in slope would probably not account for the non-continuous change in f_0 DL_{min} with level observed here. However, physiological data from the chinchilla (Rhode and Cooper, 1996; Ruggero *et al.*, 1997) suggest that the phase response is fairly constant with level until a threshold of approximately 60-70 dB SPL is reached, at which point the phase transition become shallower around CF. This is roughly line with the observed change in f_0 DL_{min} from the mid to the high level in the current study.

A second class of spectrotemporal model involves the extraction of individual resolved frequencies from phase-locking information in the auditory nerve (Goldstein and Srulovicz, 1977; Srulovicz and Goldstein, 1983). Because this type of spectro-temporal model, like the temporal models described above, does not depend on cochlear frequency selectivity to identify the frequencies of individual components, it is not clear whether they could account for the observed increased in f_0 DL_{min}. These models would probably account for the increased $f_{0,tr}$ in terms of a disruption of the temporal coding of individual frequency components due to the interaction of unresolved harmonics, although Delgutte (1984) argued the identification of individual frequency components based on a Fourier analysis of ANF responses would be robust to peripheral filtering.

4.7. Summary and conclusions

With increased stimulus level, f_0 DL performance for bandpass-filtered harmonic complexes deteriorates in two ways. First, the transition from high (poor) to low (good) f_0 DLs shifts to a higher f_0, implying that a larger spacing between adjacent harmonics is needed for good f_0 discrimination performance. The pattern in the shift in the f_0 DL transition point as function of level matched the pattern observed in estimates of harmonic resolvability based on measures of

auditory filter shapes and hearing out harmonics, supporting the hypothesis that f_0 discrimination performance is related to frequency selectivity. Second, the minimum f_0 DL (at the highest f_0 tested) increased with increasing stimulus level, as did pure-tone DLs at a comparable frequency. Overall the results provide evidence in favor of models of pitch perception that depend on frequency selectivity to encode f_0 information.

4.8 Segue

The f_0 DL transition point and three estimates of frequency selectivity each behaved in a similar manner with respect to stimulus level, remaining constant from the low to mid levels, and increasing at the high level. This result is consistent with the idea that resolved harmonics are necessary for good f_0 discrimination performance. Nevertheless, the similar trends observed across level have not established a significant correlation between the two types of measure. The following study addresses the same hypothesis by seeking a correlation between the f_0 DL transition point and estimates of frequency selectivity in listeners with sensorineural hearing loss (SNHL) across a range of severity. This study tested the hypothesis that the f_0 discrimination deficit experienced by listeners with SNHL is related to a reduction in frequency selectivity accompanying the hearing loss.

Chapter 5: Necessity of resolved harmonics for accurate f_0 discrimination: Sensorineural hearing loss

5.1 Abstract

Sensorineural hearing loss (SNHL) often results in impaired fundamental frequency (f_0) processing, but it remains unclear what aspects of hearing loss are responsible for this impairment. This study used ten listeners with moderate SNHL and three normal-hearing (NH) listeners to test whether the minimum spacing between harmonics necessary for good f_0 discrimination is related to frequency selectivity. Fundamental frequency difference limens (f_0 DLs) were measured for sine- and random-phase harmonic complexes, bandpass filtered between 1.5 and 3.5 kHz, with f_0's ranging from 75 to 500 Hz (or higher). All listeners showed a transition between small (good) f_0 DLs at high f_0's and large (poor) f_0 DLs at low f_0's, although the f_0 at which this transition occurred ($f_{0,tr}$) varied across subjects. Three measures thought to reflect frequency selectivity were significantly correlated to both the $f_{0,tr}$ and the minimum f_0 DL achieved at high f_0's: (1) the maximum f_0 for which f_0 DLs were phase-dependent, (2) the maximum modulation frequency for which amplitude modulation and quasi-frequency modulation were discriminable, and (3) equivalent rectangular bandwidths (ERBs) of auditory filters, estimated using the notched-noise method. These results provide evidence of a relationship between f_0 discrimination performance and frequency selectivity in listeners with SNHL, supporting "spectral" and "spectro-temporal" theories of pitch perception that rely on sharp tuning in the auditory periphery to extract f_0 information.

5.2. Introduction

Harmonic sounds are ubiquitous in the natural environment. The perceived pitch of such sounds, usually corresponding to their fundamental frequency (f_0), is a useful attribute in an everyday listening environment. Pitch can convey, for example, musical melody, prosody in running speech, and linguistic information in Asiatic tonal languages. Pitch information can also provide a cue for the segregation of simultaneous talkers (e.g. Darwin and Hukin, 2000), thus aiding speech intelligibility in a noisy environment.

Listeners with SNHL are faced with an impaired ability to discriminate the f_0 of complex sounds (Hoekstra and Ritsma, 1977; Hoekstra, 1979; Moore and Glasberg, 1988,1990; Moore and Peters, 1992; Arehart, 1994; Moore, 1995; Arehart and Burns, 1999; Moore and Moore, 2003). The mechanisms underlying the pitch processing deficit that accompanies SNHL remain unknown. One possible cause is the reduction in peripheral frequency selectivity that often accompanies SNHL (Glasberg and Moore, 1986). "Spectral" (e.g. Goldstein, 1973; Wightman, 1973; Terhardt, 1974,1979) and some "spectro-temporal" (e.g. Shamma and Klein, 2000; Cedolin and Delgutte, 2005a) models of pitch propose that individual harmonics of a complex tone must be resolved within the peripheral auditory system for the f_0 to be successfully extracted (for a recent review, see de Cheveigné, 2005). These models would predict that reduced harmonic resolvability in listeners with SNHL due to the broadening of peripheral filters (e.g., Tyler *et al.*, 1983; Glasberg and Moore, 1986; Moore *et al.*, 1999) should impair pitch processing.

Certain results in NH listeners provide evidence for a role of frequency selectivity and harmonic resolvability in f_0 discrimination. First, "good" f_0 discrimination performance (f_0 DLs of around 1% or less) is only found in the presence of low-order harmonics, which are more likely to be peripherally resolved. As f_0 is held fixed while harmonic number is increased (Chapter 2; Houtsma and Smurzynski, 1990; Moore *et al.*, 2005) or the absolute frequency region is held fixed while f_0 is decreased (Chapters 3 and 4; Hoekstra, 1979; Shackleton and Carlyon, 1994), f_0 DLs transition from small (good) to large (poor). In both paradigms, the f_0 DL transition occurs when only harmonics above about the 10th harmonic are present, which corresponds roughly to the limits of harmonic resolvability (Chapter 2), although the estimated limit of resolvability

varies somewhat, depending on the method used (Plomp, 1964; Moore and Ohgushi, 1993; Shackleton and Carlyon, 1994; Moore *et al.*, 2005). Second, the phase relationships between harmonic components only affects f_0 discrimination for those stimuli that yield large (poor) f_0 DLs (Chapter 3; Houtsma and Smurzynski, 1990). Because f_0 DLs should only depend on phase when harmonics are unresolved and interact within individual peripheral filters, the co-occurrence of poor and phase-dependent f_0 DLs is consistent with the idea that good f_0 DLs require resolved harmonics. Third, the transition from large to small f_0 DLs with increasing f_0 occurs at a higher f_0 at high stimulus levels than at lower levels, corresponding well with increased auditory-filter bandwidths at high levels (Chapter 4), which again supports the idea that f_0 discrimination performance is related to peripheral frequency selectivity.

The results from NH listeners allow us to propose a specific effect that poorer frequency selectivity associated with SNHL should have on pitch discrimination: for stimuli within a given spectral region, the *transition* from large to small f_0 DLs should shift to a higher f_0, because a wider harmonic spacing would be needed for resolved harmonics.

Several studies have specifically investigated the relationship between frequency selectivity and f_0 discrimination performance in listeners with SNHL, but none has identified an across-subject correlation between these two types of perceptual measure (Hoekstra, 1979; Moore and Glasberg, 1990; Moore and Peters, 1992). Moore and Glasberg (1990) measured f_0 discrimination in listeners with unilateral SNHL for harmonic complexes containing both low- and high-order harmonics (1-12) and for complexes containing only high-order, less well resolved harmonics (6-12). They also estimated frequency selectivity by measuring auditory filter shapes (Glasberg and Moore, 1986). A significant correlation between frequency selectivity and f_0 discrimination was not observed, although none of the listeners with poor frequency selectivity showed normal f_0 discrimination. In a related study, Moore and Peters (1992) investigated frequency selectivity and f_0 discrimination in both young and elderly NH and hearing-impaired (HI) listeners. While they found both reduced frequency selectivity and reduced f_0 discrimination performance for many of the impaired subjects, there was only a weak correlation between these two deficits.

The weak evidence for a relationship between reduced frequency selectivity and reduced f_0 discrimination in the studies of Moore and Glasberg (1990) and Moore and Peters (1992) may be due in part to the fact that they chose harmonic number 6 as the cutoff between their low- and high-order conditions, based on the results of Plomp (1964) and Moore *et. al.* (1984) suggesting that NH listeners can hear out only the first approximately five or six individual components of a harmonic complex. More recent studies of the relationship between f_0 discrimination and harmonic number have suggested that the cutoff between large and small f_0 DLs occurs near harmonic number 10 (Chapters 2 and 3; Houtsma and Smurzynski, 1990; Kaernbach and Bering, 2001). Thus, all conditions in the Moore and Peters (1992) and Moore and Glasberg (1990) studies contained harmonics that are known to yield good f_0 discrimination in NH listeners, preventing a comparison of the effects of hearing impairment on pitch discrimination for complexes containing low-order (resolved) versus high-order (unresolved) components.

Hoekstra (1979) and Hoekstra and Ritsma (1977) also measured f_0 discrimination in listeners with hearing impairment believed to be of cochlear origin. They used harmonic complexes with a range of f_0's, bandpass filtered into a fixed spectral region. They found that f_0 DLs transitioned from large to small with increasing f_0, but that this transition occurred at higher f_0's for HI listeners. This result is consistent with the idea that listeners with SNHL require a larger spacing between components to yield resolved harmonics and therefore good f_0 discrimination. While Hoekstra (1979) and Hoekstra and Ritsma (1977) estimated frequency selectivity psychophysical tuning curves (PTCs) in a subset of the SNHL listeners, the number of subjects in this subset was too small to permit a correlational analysis of the relationship between f_0 discrimination and frequency selectivity. Nevertheless, Hoekstra and Ritsma (1977) noted that those listeners with abnormally high f_0 DL transition f_0's also demonstrated abnormal PTCs, suggesting a relationship between the two measures. Arehart (1994) also found that the f_0 DL transition occurred at a lower harmonic number for listeners with SNHL, but did not relate these measures to estimates of peripheral frequency selectivity.

The goal of current study was to test the hypothesis that the f_0 (and hence the harmonic spacing) where f_0 DLs transition from large to small is dependent on peripheral frequency selectivity. What distinguishes this study from previous investigations of the relationship between f_0

discrimination and frequency selectivity is that it (1) focuses on the point of *transition* between large and small f_0 DLs (Experiment 5A), and (2) relates this transition to measures of frequency selectivity (Experiments 5B and 5C) in a sufficiently large and diverse population of SNHL subjects to enable a correlational analysis. The harmonic complexes were filtered into a passband extending from 1500 to 3500 Hz. Single-frequency measures of frequency selectivity and frequency discrimination used a signal frequency of 1500 Hz. This value, representing the lower cutoff of the passband for the complexes, was selected because it was the point at which the harmonics within the complex should have been best resolved, and is likely to represent the upper limit of performance (Houtsma and Smurzynski, 1990).

5.3. Experiment 5A: f_0 DLs

5.3.1. Rationale

Experiment 5A measured f_0 DLs as a function of f_0 for harmonic complexes bandpass filtered into a fixed spectral region in listeners with SNHL. By investigating the dependence of f_0 DLs on harmonic resolvability while keeping the frequency region constant, this paradigm avoided a possible confounding factor of SNHL that varies across frequency regions, which might arise in a paradigm that kept f_0 constant while varying harmonic number (e.g. Chapter 2; Houtsma and Smurzynski, 1990).

Harmonic stimuli were presented in both sine and random phase to give an estimate of harmonic resolvability based on the phase dependence of f_0 DLs. In NH listeners, the phase relationships between harmonic components have been shown to affect f_0 DLs for complexes containing only high-order harmonics but not for those containing low-order harmonics (Chapter 3; Moore, 1977; Houtsma and Smurzynski, 1990). This result is generally interpreted in terms of harmonic resolvability. For high-order, unresolved harmonics that interact within individual peripheral filters, the phase relationship between components affects the temporal envelope and therefore the pitch percept associated with these complexes. In contrast, low-order resolved harmonics do not interact appreciably within individual peripheral filters, such that the pitch percept associated with these harmonics is not affected by phase manipulations.

5.3.2. Subjects

Ten subjects (four female) with SNHL participated in the study. Pure-tone audiograms were measured using an AD229e diagnostic audiometer (Interacoustics) and TDH39 headphones. Bone-conduction threshold measurements (Radioear B-71) verified that the hearing loss for each subject was sensorineural in nature, based on the absence of any air-bone gaps larger than 10 dB. All subjects had moderate (30-65 dB HL re ANSI-1996) losses at audiometric frequencies between 1.5 and 4 kHz, the relevant frequencies for the stimulus frequency range (1.5 to 3.5 kHz) used in the study[1], with two exceptions. Subject I1 had a low-frequency loss with near-normal thresholds at 3 and 4 kHz (10 and 20 dB HL, respectively), but impaired thresholds at lower frequencies (50 dB HL at 1 and 2 kHz). Subject I2 had a notched loss, with impaired thresholds at 1.5 and 2 kHz (45 and 50 dB HL, respectively), but a mild loss of 25 dB HL at 4 kHz. A threshold equalizing noise (TEN) test (Moore *et al.*, 2000) for octave frequencies between 250 and 8 kHz verified the absence of "dead regions" in each subject. Ages ranged from 27 to 78 years, with a mean of 49.7 and a median of 50.5 years. In each subject, measurements were made for the ear that showed the most hearing loss that also fell in the 30- to 65-dB HL range (i.e., if one ear showed a maximum loss of 70 dB HL and the other a maximum of 50 dB HL in the specified range, the ear with the 50 dB loss was tested). Listeners with symmetrical losses were given their choice of test ear. One subject with an asymmetrical loss (I9) was tested in both ears. For simplicity, the two ears of this subject are treated as if they were from separate subjects [I9(l) and I9(r)] for the remainder of this paper, such that the total number of SNHL subjects was considered to be eleven. Audiograms for the 11 ears with SNHL are shown in Fig. 5.1, with vertical lines indicating the 1.5-3.5 kHz stimulus region tested in this study. Three NH subjects (one female) also participated in the study. Normal-hearing subjects had audiometric thresholds of 15 dB HL or less *re* ANSI-1996 at octave frequencies between 250 Hz and 8 kHz. The ages of the NH subjects were 19, 20, and 52 years. Each NH subject was tested with stimuli presented to the left ear. Table 5.1 lists audiological information for each NH and HI subject who participated in the study, as well as information regarding the stimulus level(s) tested in each of the three experiments (see Section 5.3.3.2). All of the subjects were paid for their time.

[1] Some subjects with audiometric thresholds at 1 and 2 kHz within 5 dB HL of each other were not tested at 1.5 kHz. For these subjects, the 1.5-kHz threshold is taken as the mean of the 1- and 2-kHz thresholds.

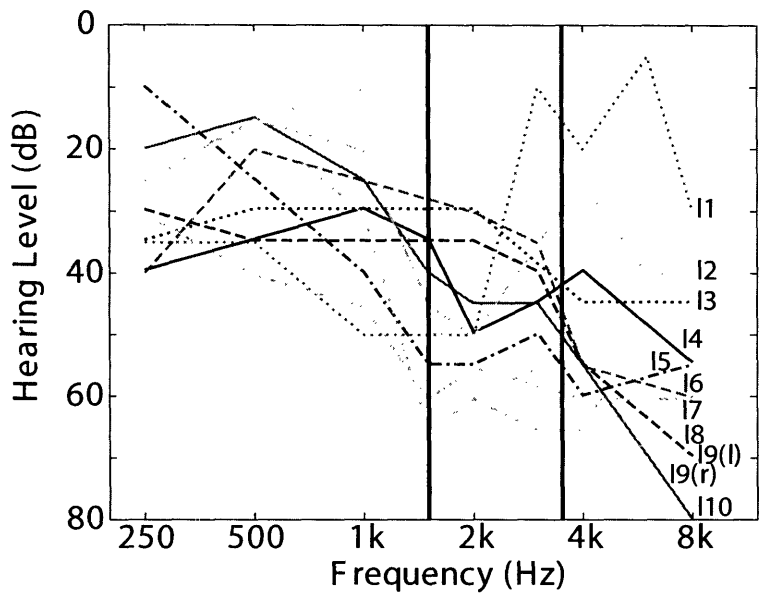

Figure 5.1. Audiograms for the ten HI subjects (11 ears) who participated in the study. Vertical lines represent the 1.5- to 3.5-kHz stimulus region used in experiments 5A, B and C.

Table 5.1. Audiological, biographical and stimulus level information for the 10 HI subjects (11 ears) and 3 NH subjects who participated in the study. (*) Measurements were made in both the left and right ears of subject I9. (**) Absolute thresholds were not measured in NH subjects using the adaptive technique and HD 580 headphones that were used in experiments 1-3.

	Subject	Age	Sex	Test Ear	Audiometric threshold @1.5 kHz (dB HL) TDH 39	Max. adaptive threshold, 1.5-3.5 kHz (dB SPL) HD 580	TEN level (dB SPL / ERB_N)
Hearing-Impaired	I1	29	F	R	50	49.3	50
	I2	58	M	L	45	46.3	50
	I3	67	M	R	30	36.5	50
	I4	58	F	L	35	44.8	50
	I5	52	M	R	55	48.2	50
	I6	78	F	R	27.5	47	50
	I7	27	M	L	60	57	60
	I8	33	M	L	35	50.2	50
	I9(l)	46	F	L	45	58.7	60
	I9(r)*	46	F	R	65	61	62
	I10	49	M	L	40	46	50
Normal-Hearing	N1	19	M	L	-5	**	50, 65
	N2	20	M	L	-2.5	**	50, 65
	N3	52	F	L	5	**	50, 65

5.3.3. Methods

The stimuli and methods for estimating f_0 DLs were similar to those used in Chapter 4.

(5.3.3.1) Stimuli

Fundamental frequency DLs were measured as a function of f_0 for 500-ms (including 30-ms raised-cosine rise and fall ramps) random- and sine-phase harmonic complexes, bandpass filtered between 1.5 and 3.5 kHz, with 50 dB/oct. slopes. The filtering operation was implemented in the spectral domain by first adjusting the amplitude of each sinusoidal component, and then summing all of the components together.

Random-phase harmonic complexes were chosen because they are known to yield very poor f_0 DLs (on the order of 5-10% of the f_0) (Chapters 2, 3 and 4; Micheyl *et al.*, 2005) when the harmonics are unresolved, thus producing a large f_0 DL difference between low and high f_0's and providing the best opportunity to observe the transition from large to small f_0 DLs. Sine-phase conditions were included to give an estimate of harmonic resolvability based on the phase dependence of f_0 DLs. At least nine f_0's were tested for each HI subject (50, 75, 125, 150, 175 200, 250, 325, 400 and 500 Hz). Higher f_0's (750 Hz and, in one case, 1500 Hz) were tested in conditions where the 500-Hz f_0 DL appeared to be larger than the 1500-Hz pure-tone frequency difference limen (FDL) in a pilot run, suggesting that the f_0 DL had not reached its asymptotic value. For NH subjects, nine f_0's (75-500 Hz) were tested in the random-phase conditions. The 500-Hz f_0 was not tested in the sine-phase conditions, resulting in a total of eight f_0's tested in these subjects.

(5.3.3.2) Stimulus level

To keep both the sensation level (SL) and the overall sound pressure level (SPL) similar for both NH and HI listeners, all stimuli were presented in a background of threshold equalizing noise (TEN; Moore *et al.*, 2000). TEN is intended to yield detection thresholds in noise that are approximately constant across frequency, such that pure tones presented at equal SPL in TEN will have roughly equal SL also. For each HI listener, the TEN was set to a level that, in a NH listener, would yield tone-in-noise detection thresholds at least as high as the HI subject's detection thresholds in quiet. Initially, the TEN level was intended to be the same for all

subjects. Tone-in-quiet thresholds, estimated using a three-interval three-alternative forced-choice, two-down, one-up adaptive procedure, were no higher than 50 dB SPL in the 1.5-3.5 kHz range for each of the first five HI subjects recruited for the study. For these subjects, the background noise was set to 50 dB SPL per ERB_N, where ERB_N is the equivalent rectangular bandwidth of peripheral filters in NH listeners as described by Glasberg and Moore (1990). After measurements had been made for the initial five subjects, five additional subjects (six ears) were recruited for the study. Three of these six additional ears had at least one tone-in-quiet threshold in the 1.5-3.5 kHz range above 50 dB SPL. For these ears, the TEN level (Table 5.1, column 7) was set to 60 dB SPL (two ears) or 62 dB SPL (one ear) depending on the maximum absolute threshold measured in the 1.5 to 3.5 kHz range (Table 5.1, column 6)

To set stimulus levels, detection thresholds were measured for pure tones in the TEN background. The 0 dB SL reference was determined for each subject individually, and was taken as the mean of six threshold measurements, two estimates each for pure tone-frequencies of 1.5, 2 and 3 kHz. Across the HI subjects, the 0-dB SL reference ranged (in dB SPL) from -2.9 to +3.5 dB re the TEN level in dB SPL/ERB_N.

Because of the different levels tested in the HI subjects, NH subjects were tested with both 50 dB SPL/ERB_N and 65 dB SPL/ERB_N TEN, to ensure that stimuli were presented at a level at least as high as the highest level presented to HI subjects. Due to a lack of testing time, sine-phase conditions were not tested at the higher level in NH listeners. Across the NH listeners, the 0-dB SL reference ranged (in dB SPL) from -3.8 to -1.8 dB and from -3.1 to -1.0 dB re the 50 and 65 dB SPL/ERB_N TEN levels, respectively.

Stimuli were presented at a nominal 12.5 dB SL per component in a TEN background at the level specified in Table 5.1, or for NH listeners, at each of two levels specified in Table 5.1. Exceptions to the procedure for setting the stimulus level were made for three subjects (two NH, one HI) where the stimuli presented in 62-65 dB SPL/ERB_N noise were uncomfortably loud at the 75- and/or 125-Hz f_0. For the subjects and conditions where this occurred, stimulus levels

were reduced somewhat, or the conditions were eliminated[2]. To prevent contralateral detection of the stimuli, TEN was presented to the non-test ear at 20 dB below the level of that presented to the test ear.

(5.3.3.3) Procedure

Five f_0 DL measurements were made for each combination of f_0 and phase and, for NH listeners, stimulus level. For each subject, all of the random-phase conditions were tested before the sine-phase conditions, because the decision to include sine-phase measurements was not made until the random-phase data had already been collected for the first five HI subjects. FDLs for a 1500-Hz pure tone were measured as an additional condition interspersed with the f_0 DL estimates for harmonic complexes. Five FDL measurements each were interspersed with the measurements involving random- and sine-phase stimuli.

Fundamental frequency DLs and FDLs were estimated in a three-interval three-alternative forced-choice (3I-3AFC) adaptive procedure, using at two-down, one-up algorithm to track the 70.7% correct point on the psychometric function (Levitt, 1971). Two intervals contained a stimulus with a base f_0 ($f_{0,base}$) and the other interval contained a complex with a higher f_0. The listener's task was to identify the interval containing the complex with the higher pitch. The f_0 difference (Δf_0) was initially set to 20% of the f_0, changed by a factor of 1.59 until the second reversal, and then changed by a factor of 1.26 for six more reversals. The f_0 DL was estimated as the geometric mean of the Δf_0's at the last six reversal points.

To reduce the effectiveness of loudness as an alternative cue to choose the correct response without extracting pitch information, the root-mean-squared (RMS) power was first equalized across the three intervals by increasing the stimulus level for the interval containing the higher f_0, and then a random level perturbation was added to each interval, chosen from a uniform distribution of ±2.5 dB. In addition, $f_{0,\mathrm{base}}$ was roved from trial to trial within a run, chosen from a uniform distribution between ±2.5% of the average f_0. This was intended to encourage subjects

[2] For NH subject N1, the TEN and stimulus levels were each reduced by 5 dB for the 75- and 125-Hz f_0's. For NH subject N3, the stimulus level was reduced by 3 dB for the 75-Hz f_0, but the TEN was kept at 65 dB SPL/$\mathrm{ERB_N}$. For I9(r), who completed more f_0 conditions than the other subjects, the 75-Hz f_0 conditions were not tested.

to compare the pitches of the stimuli across each of the intervals of one trial, rather than comparing the pitch in each interval with some internally stored representation of the $f_{0,\mathrm{base}}$, although the f_0 roving may not have been effective, especially for low f_0's where measured f_0 DLs were relatively large (8% or more).

(5.3.3.4) Apparatus

The stimuli were generated digitally and played out via a soundcard (LynxStudio LynxOne) with 24-bit resolution and a sampling frequency of 32 kHz. The stimuli were then passed through a programmable attenuator (TDT PA4) and headphone buffer (TDT HB6) before being presented to the subject via one earpiece of a Sennheiser HD 580 headset. Subjects were seated in a double-walled sound-attenuating chamber. Intervals were marked by colored boxes on a computer screen, and visual feedback (correct/incorrect) was provided following each response.

5.3.4. Results

Figure 5.2 plots f_0 DLs as a function of f_0 for six sample HI subjects [I5, I6, I7, I9(l), I9(r) and I10], representing the range of results observed across the eleven HI ears, and the mean f_0 DLs across the three NH subjects at each of the two stimulus levels. "Low" and "high" levels for NH listeners refer to stimuli presented in 50 and 65 dB SPL/$\mathrm{ERB_N}$ background TEN, respectively. Random-phase conditions are denoted by round symbols, and sine-phase conditions by square symbols. The solid lines in Fig. 5.2 represent fitted functions to the random-phase data, and the dashed vertical lines represent mid-points of the transitions in the functions, as described in Sec. 5.7.1.

Four main findings are apparent in the results. First, for most subjects and phase conditions, f_0 DLs generally transitioned from large to small f_0 DLs with increasing f_0. This is consistent with previous results in NH listeners (Chapters 2 and 3; Hoekstra, 1979; Shackleton and Carlyon, 1994) and is thought to reflect the transition from all unresolved to some resolved harmonics. Second, the f_0 where the f_0 DL transition occurred (the f_0 transition point, $f_{0,tr}$), varied across subjects. This may reflect differences in harmonic resolvability. Consistent with this hypothesis is the fact that NH listeners and listeners with relatively mild hearing loss (e.g., I6 and I10), show the transition at a relatively low f_0 of around 200 Hz, whereas subjects with more

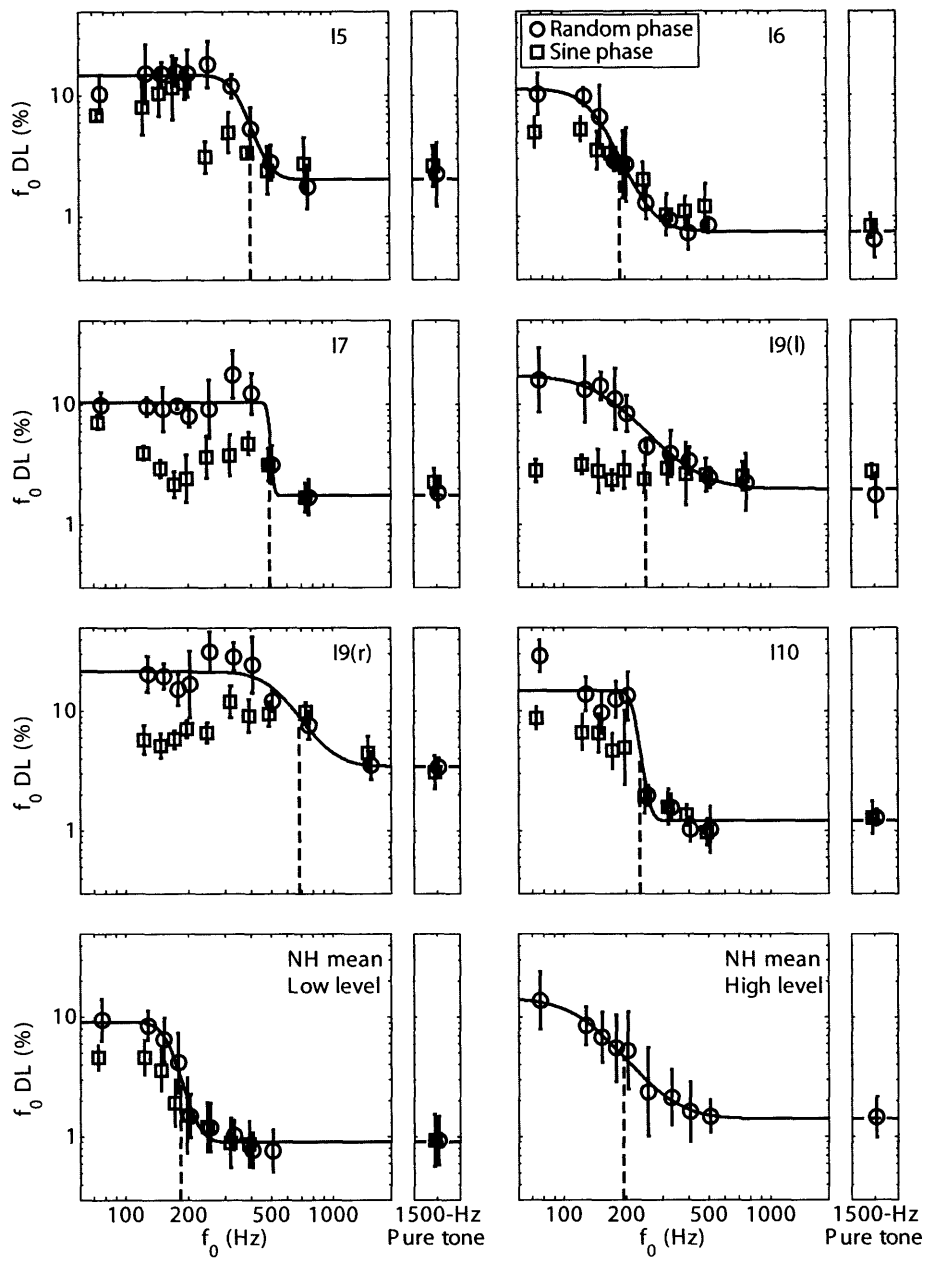

Figure 5.2. (Top six panels) f_0 DLs plotted as a function of f_0 and FDLs for a 1500-Hz pure-tone FDL for six example HI subjects. (Bottom two panels) mean f_0 DLs and FDLs across the three NH subjects at two stimulus levels. Error bars indicate the standard deviation of the five f_0 DL or FDL measurements for each HI listener, and the standard deviation across the three mean f_0 DLs or FDLs for NH listeners. Dashed curves indicate the sigmoid function[3] that best fit the random-phase data, for each individual HI subject or the mean NH data (see Sec. 5.7.1). Vertical dashed lines indicate the $f_{0,tr}$, defined as the f_0 that yielded random-phase f_0 DLs halfway (on a log scale) between maximum and minimum.

severe hearing loss [e.g., I7 and I9(r)] tend to have transitions at higher f_0's of 500 Hz or more. The hypothesis that the across-subject variability in the f_0 DL transition point is related to frequency selectivity will be examined quantitatively in Section 5.7 by comparing these results to estimates of frequency selectivity. Third, the results from some HI subjects show substantial non-monotonicities in the pattern of results [e.g. I5, I7 and I9(r)]. For these subjects. f_0 DLs are elevated at moderate f_0's, which may result from unresolved harmonics that yield an envelope repetition rate too high to be processed efficiently (e.g., Kohlrausch *et al.*, 2000). Overall, flat or non-monotonic f_0 DL functions were observed in three and five HI subjects (out of 11) in the random- and sine-phase conditions, respectively. Fourth, for most subjects, the phase relationships between harmonics affected f_0 DLs for low f_0's, but not high f_0's, consistent with previous results in NH hearing listeners (Chapter 2; Houtsma and Smurzynski, 1990) and with the idea that complexes with high f_0's contain unresolved harmonics.

5.4. Experiment 5B: Modulation discrimination

5.4.1. Rationale

The effect of phase on f_0 DLs in experiment 5A provides an estimate of peripheral frequency selectivity that can be compared to the $f_{0,\mathrm{tr}}$. However, the phase effect is not completely independent of the $f_{0,\mathrm{tr}}$ because both relate to the same random-phase f_0 DL data. A stronger test of this relationship would rely on an estimate of frequency selectivity that is independent of f_0 discrimination.

The present experiment estimates frequency selectivity by measuring listeners' ability to discriminate between sinusoidal amplitude modulation (SAM) and quasi-frequency modulation (QFM, Zwicker, 1952). Both types of waveform were generated by three-tone complexes with identical amplitude spectra but different relative phases between components. Previous results have shown that HI listeners can perform this task out to higher modulation frequencies than can NH listeners (Nelson and Schroder, 1995), and that performance improves with increased stimulus level for NH listeners (Nelson, 1994). This may be because the wider peripheral filters associated with SNHL and high stimulus levels in NH listeners increases the likelihood of peripheral interactions between components. The hypothesis that small f_0 DLs require resolved harmonics suggests that *better* SAM/QFM discrimination (i.e., a higher maximum modulation

frequency) should correlate with *poorer* f_0 discrimination performance (i.e. a higher $f_{0,tr}$). One advantage of testing this prediction is that it postulates higher levels of performance in HI listeners than in NH listeners and so should not be confounded by any more general higher-level perceptual or processing deficits by the HI listeners (who were, on the average, older than the NH listeners). Also, this measurement estimates frequency selectivity by varying the spacing between adjacent harmonics, a situation analogous to that of the f_0 DL measures in Experiment 5A.

5.4.2. Methods

As in the studies of Nelson (1994) and Nelson and Schroder (1995), the current experiment measured the percentage correct in discriminating between SAM and QFM complexes as a function of the modulation frequency (f_m). The carrier frequency (f_c) was fixed at 1500 Hz (the lower cutoff frequency of the bandpass filter used in experiment 5A). In this three-interval three-alternative forced-choice task, two intervals contained SAM complexes and the third contained a QFM complex. Listeners were asked to identify which interval containing the stimulus that was different from the other two. Visual feedback (correct/incorrect) was provided. Each interval consisted of a three-component tone complex, with frequencies $f_c - f_m$, f_c, and $f_c + f_m$, and duration 500 ms (including 30-ms raised cosine onset and offset ramps). The intervals were separated by silent gaps of 375 ms.

The wideband background TEN was not used in this experiment because it was found to be too detrimental to performance in the modulation discrimination task. Instead, a low-pass TEN with a cutoff frequency of ($f_c - 1.95f_m$) was used to mask any distortion products occurring at frequencies of $f_c - 2f_m$ or below. The lowpass TEN had the same spectral characteristics as the wideband TEN of experiment 5A for frequencies below its cutoff. The lowpass TEN was turned on 250 ms before the first interval and turned off 250 ms following the offset of the third interval, giving it a total duration of 2750 ms.

Although the wideband TEN was not used, the amplitude of the center component was set at a SPL equal to the 12.5 dB SL level that was used in experiment 5A, adjusted for each subject. The amplitude of each side band was 6 dB below that of the center component, producing 100%

modulation in the SAM case. SAM complexes were generated by setting the three components to be in sine starting phase. QFM complexes were identical to the SAM complexes except that the center component was phase-shifted by 90°.

The f_m's were set to the f_0 values tested in experiment 5A (eight f_m's for NH subjects, nine or ten f_m's for HI subjects). Each run included four trials for each f_m, presented in random order. Each HI subject completed 13 runs for a total of 52 stimulus presentations for each f_m, with two exceptions, detailed below. Each NH subject was tested with the stimulus level set relative to the detection threshold in both the 50 and 65 dB SPL/ERB_N TEN, with 13 runs presented at each level. The same 11 HI and 3 NH subjects from experiment 5A participated in this experiment. Each subject completed at least one hour of practice before the measurement period began.

Two HI subjects (I6 and I8) were unable to achieve much above chance performance even for the lowest f_m tested of 75 Hz, unless the randomization of f_m within a run was greatly diminished. For these two subjects, eight trials, each of two f_m's, were presented within a run, and seven runs were completed for each pair of f_m's for a total of 56 stimulus trials per f_m. With this modification, one of the subjects (I6) still failed to achieve 100% correct for the 75-Hz f_m. Two additional f_m's (25 and 50 Hz) were added for this subject, who achieved near-perfect performance at 25 Hz.

5.4.3. Results

The upper six panels of Fig. 5.3 show the percent correct as a function of f_m for the same six sample HI subjects who were shown in Fig. 5.2. The lower two panels of Fig. 5.3 show the percent correct for the three NH listeners for each f_m at the low and high stimulus levels. Each subject showed qualitatively similar results, with performance decreasing from near 100% correct for the lowest f_m tested to near chance (33%) for the highest f_m tested. The solid lines represent sigmoidal fits, and the vertical dashed lines represent estimates of the 66.7% correct point based on the fitted functions, as described in Sec. 5.7.1.4. No consistent non-monotonicities were observed in the results, suggesting that the non-monotonicities observed by Nelson and Schroder (1995) may have derived from combination tones that were masked by the lowpass noise in the current experiment. The data shown in Fig. 5.3 generally support the

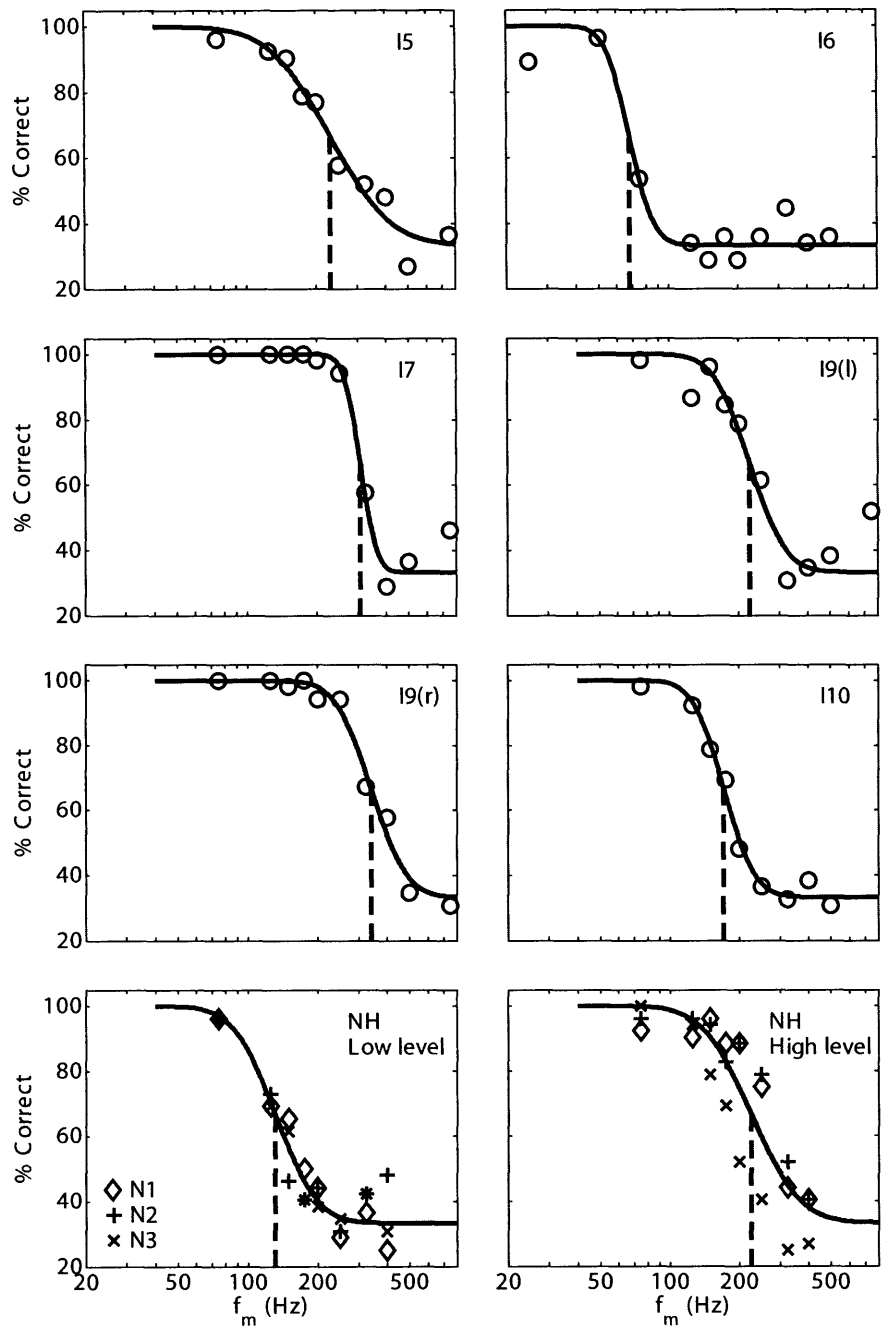

Figure 5.3. Results of experiment 3B, showing the percentage correct in discriminating between SAM and QFM as a function of f_m, the frequency spacing between components, for the six sample HI subjects of Fig. 5.2 (top six panels), and for the three NH listeners at the two stimulus level (two lower panels). Dashed curves indicate the sigmoid function that best fit the data for each HI subject or for the pooled NH data at each level. Vertical dashed lines indicate the estimate of the threshold f_m ($f_{m,tr}$) yielding 67% correct performance, halfway between perfect performance

hypothesis that listeners with poorer frequency selectivity can perform better than normal (i.e., out to a higher modulation frequency) in the modulation discrimination task. NH listeners and HI listeners with mild hearing loss (e.g. I6 and I10) who performed best at f_0 discrimination (Fig. 5.2) performed worst at discriminating f_m (Fig. 5.3). Conversely, subjects with more moderate-to-severe hearing loss who were poor discriminators of f_0 performed best at discriminating f_m at high modulation frequencies [e.g. I5, I7, I9(l) and I9(r)]. However, this was not always the case. For example, I8 had a relatively high $f_{0,tr}$ but still showed difficulty in performing this task. The relationship between performance in this task and f_0 DLs is evaluated in more detail for all subjects in Sec. 5.7.

5.5. Experiment 5C: Auditory filter shapes

5.5.1. Rationale

The f_0 DL phase effect (Experiment 5A) and modulation discrimination data (Experiment 5B) provide estimates of frequency selectivity. Nevertheless, the current standard for evaluating peripheral frequency selectivity in the spectral domain is the notched-noise method of auditory filter-shape estimation (Patterson, 1976). This experiment used a "fixed signal level" version of the notched-noise method described by Rosen and Baker (1994) to estimate auditory filter bandwidths in the NH and HI listeners who participated in experiments 5A and 5B. The level of the notched-noise masker that just masked a pure tone was measured as a function of the masker's spectral notch width. At threshold, this paradigm is thought to deliver roughly constant overall (signal plus noise) power across notch widths to the auditory filter in question, thus reducing the possible confounding influence of variations in filter shape with input level.

5.5.2. Methods

Throughout the experiment, the pure-tone signal had a constant frequency (f_{sig}) of 1500 Hz, corresponding to the low-frequency edge of the passband in experiment 5A. The signal was fixed at the SPL level corresponding to 10 dB SL (adjusted for each subject individually) re. the TEN level that was used in experiment 5A. Although the signal SPL was adjusted relative to the detection threshold in TEN, the TEN background was not used in this experiment. The NH listeners were only tested with the signal at one level (10 dB SL re. the 50 dB SPL/ERB_N TEN

level) because the higher signal level could not be comfortably masked for wide notch widths in these listeners.

Each trial in the experiment consisted of three intervals, each with a 700-ms duration, separated by 500-ms silent gaps. Two of the intervals contained only a 700-ms noise burst (including 10-ms raised-cosine onset and offset ramps). The other interval also contained a 500-ms pure-tone signal (including 30-ms raised-cosine onset and offset ramps), temporally centered within the noise burst. The listeners' task was to identify which of the three intervals contained the pure-tone signal. A 3I-3AFC procedure with a two-up, one-down adaptive algorithm tracked the 70.7% correct point (Levitt, 1971). The spectrum level of the noise (dB SPL/Hz) was initially set to -35 dB re. the TEN noise level (dB SPL/ERB_N) for each subject, and changed by 8 dB for the first two reversals, 4 dB for the next two reversals, and 2 dB for the last eight reversals. Threshold was estimated as the mean of the noise levels at the last eight reversal points. Reported thresholds are the mean of three such threshold estimates.

The noise masker consisted of two bandpass noises, each with a bandwidth of 200 Hz. The notch width was defined in terms the deviations from the signal frequency, expressed as a proportion of f_{sig}, of both the high-frequency edge of the lower-frequency noise band (Δf_l) and the low-frequency edge of the upper-frequency noise band (Δf_u). Five symmetrical notch conditions were presented, with equal Δf_l and Δf_u values of 0, 0.1, 0.2, 0.3, and $0.4f_{sig}$. To allow for the possibility of asymmetrical filters, there were also four asymmetric conditions [(Δf_l, Δf_u) = ($0.1f_{sig}$, $0.3f_{sig}$), ($0.2f_{sig}$, $0.4f_{sig}$), ($0.3f_{sig}$, $0.1f_{sig}$) and ($0.4f_{sig}$, $0.2f_{sig}$)]. More notch-conditions were presented here than in Chapter 4 because the lower stimulus levels (and the generally wider auditory filters for HI subject tested at a higher stimulus level) limited to a safe range the noise level needed to mask the signal tone for the wider notch deviations (i.e. $0.3f_{sig}$ and $0.4f_{sig}$). A low-pass noise was also included to mask any possible low-frequency combination bands (Greenwood, 1972) that could facilitate the detection of the signal, with cutoff frequency equal to the low-frequency edge of the lower-frequency noise band and a spectrum level 20 dB below that of the notched noise.

5.5.3. Analysis

A standard fitting procedure was used to derive auditory filter shapes from the data (Glasberg and Moore, 1990). The lower side of the rounded-exponential (roex) filter was defined by the filter slope (p) and the dynamic range limit (r). The upper slope had no dynamic range limit, simulating the uniformly steep upper slope often found in auditory-nerve and basilar-membrane tuning curves (e.g. Kiang *et al.*, 1965; Sellick *et al.*, 1982). The fitting procedure took into account the Sennheiser H580 transfer function, the middle-ear transfer function, and the possibility of off-frequency listening and variations in filter bandwidth with CF, as described by Glasberg and Moore (1990). The filter tip was assumed to be symmetrical, but the asymmetrical application of the dynamic-range limitation and the combination of off-frequency listening and proportional variation in filter bandwidth with CF accounted well for unequal threshold measurements in the four asymmetric notch conditions (rms fitting error 1.16 dB versus 1.31 dB in the symmetric conditions). The equivalent rectangular bandwidths (ERB) of the filters were derived from the fitted parameters.

5.5.4. Results

Figure 5.4 shows the notched-noise masking data along with the masking predictions based on the best-fitting filter functions (solid lines) for each of the six example HI subjects (upper six panels) from Figs. 5.2 and 5.3, and for the mean of the three NH subjects (lower panel). Round symbols represent conditions with symmetrical noise notches, while left- and right-pointing triangles represent asymmetrical conditions where Δf_l was less than and greater than Δf_u, respectively. The simple filter-shape model yielded a reasonable fit to the data for each subject. As expected, HI subjects generally showed broader frequency selectivity than did NH subjects. This can be seen in Fig. 5.4 by the generally shallower increase in masker level as a function of notch width in the HI than in the NH listeners. Although for illustrative purposes fits are shown for the mean NH data in the two lower panels of Fig. 5.4, fits were performed for each individual NH subject for the across-subject analyses described in Sec. 5.7.

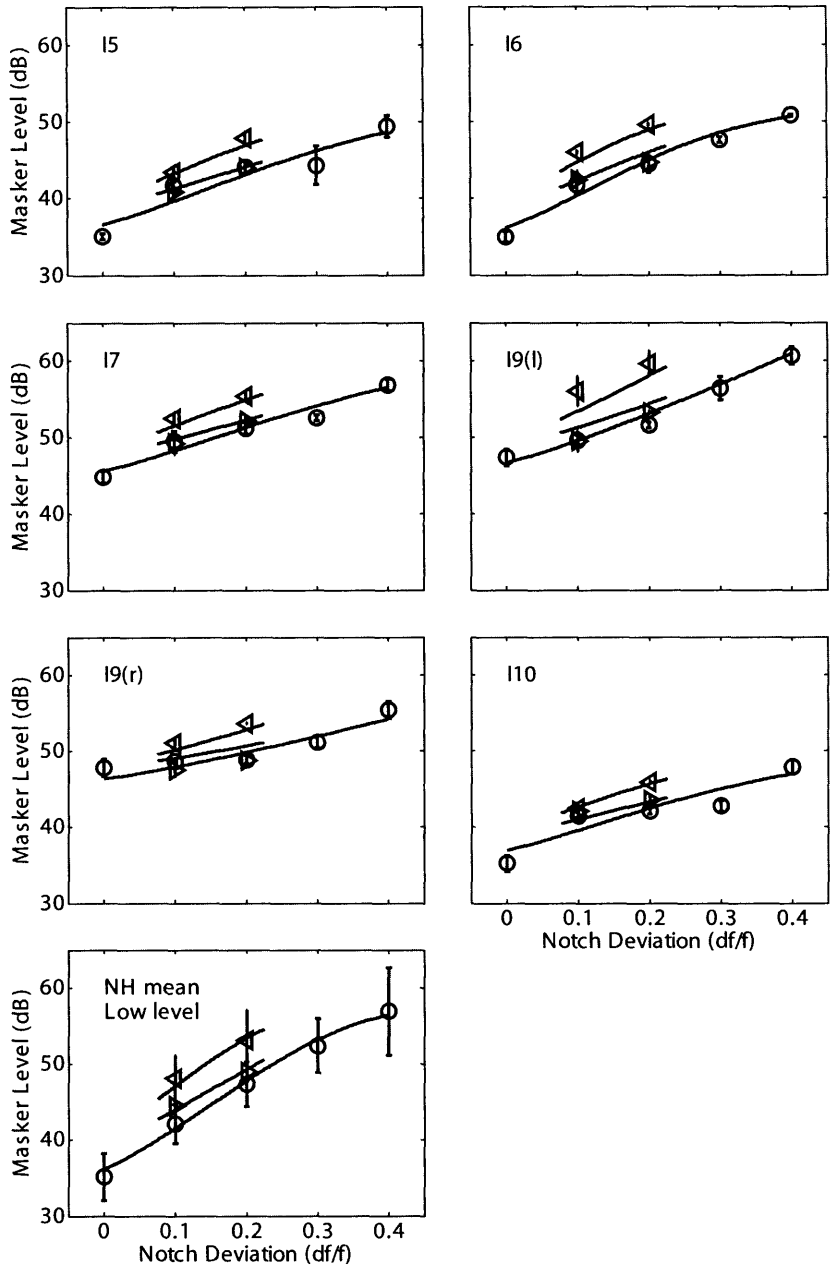

Figure 5.4. The notched-noise masking level needed to just mask a 1500-Hz probe tone presented at 10 dB re the threshold in TEN at the level indicated in Table 5.1. Round symbols indicate notches that are symmetrical around the probe-tone frequency. Left- and right-pointing triangles indicate asymetrical notches, shifted toward lower and higher frequencies, respectively. For the asymmetrical conditions data are plotted according to the notch edge closest to the probe frequency, and the second notch edge was 0.2/f farther away from the probe frequency. Solid lines indicate the predicted masker levels based on the best fitting auditory filter shape. Error bars indicate the standard deviation across the three masking measurements for the HI listeners, or across the three NH listeners.

5.5.5. Alternative filter models

The notched-noise data were also fit with three alternative filter-shape models. Thus, all four permutations of symmetrical or asymmetrical filter tips (same or different upper- and lower-frequency slopes) and the presence or absence of a low-frequency dynamic range-limiting filter tail were tested. All four filter models yielded similar rms fitting errors (Table 5.2), although there was a small advantage for filter models incorporating a larger number of free parameters. Because the dynamic range limitation was the crucial variable in the general correspondence between f_0 DLs and frequency selectivity in the NH data of Chapter 4, but the rms error was only marginally improved by the addition of a fourth free parameter in the asymmetric case, the filter model with a symmetrical tip and dynamic range limitation was chosen to characterize frequency selectivity for comparisons with the f_0 DL data detailed below in Section 5.7.

Filter tip	Dynamic range limitation?	Free parameters	RMS fitting error (dB)
Symmetric	Yes	3	1.24
Asymmetric	Yes	4	1.14
Symmetric	No	2	1.42
Asymmetric	No	3	1.31

Table 5.2. The accuracy of four auditory filter models in fitting the notched-noise masking data across 14 NH and HI subjects.

5.6. Hearing out harmonics

We attempted to use a hearing out harmonics paradigm (Chapters 2 and 4) to investigate the degree to which individual harmonics were resolved as a function of the stimulus f_0. The stimuli and experimental method were identical to those described in Chapter 4. The task was to compare the frequency of a comparison pure-tone presented in isolation to that of a reference harmonic embedded within the complex, with both the comparison and reference tone gated on and off to encourage perceptual segregation. Four HI listeners (I2, I3, I8 and I10) attempted this task and none were able to achieve above-chance performance for any f_0 condition, even after two to four hours of training. As a result, this paradigm was abandoned as a method for estimating harmonic resolvablity for the remaining listeners. The reason for the difficulty experienced by these listeners in this task are unknown, and could be related to their hearing loss, or to their lack of musical training.

5.7. Analysis

5.7.1. Summary measures

Each of the experiments described above yielded summary values that were then used to derive correlations between the measures of pitch discrimination (experiment 5A) and the measures of frequency selectivity (experiments 5A, B and C). The different summary measures, and the way they are derived, are described below. The values for each of these measures in each subject are shown in Table 5.3, with boldface entries indicating HI values that fell more than two standard deviations above the mean NH values estimates at a comparable stimulus level[3]..

Three measures of pitch discrimination performance were derived from the random-phase f_0 DL data. A sigmoid function[4] was fit to the log-transformed f_0 DLs versus log-transformed f_0 data (Fig. 5.2). A fit was made to the data for each HI subject and separately to the data at each of the two stimulus levels for each NH subject. The fitting procedure adjusted four free parameters, representing estimates of (1) the maximum (f_0 DL_{max}) and (2) the minimum f_0 DL (f_0 DL_{min}) attained at very low and very high f_0's, respectively, (3) the f_0 at which f_0 DLs transitioned from large to small ($f_{0,tr}$) and (4) the slope (m) of the transition. The FDL data measured for the 1.5-kHz pure tone was included in the fitting procedure, with f_0 set to infinity, because a pure tone is "infinitely" resolved. The assumption that the pure-tone case will yield the smallest possible DL may be questionable because the presence of additional resolved harmonics would yield more f_0 information. However, two-tailed t-tests indicated that for each subject and level, with one exception[5], the f_0 DL at the largest f_0 tested (500, 750 or 1500 Hz) was not significantly different ($p>0.1$) from the FDL for the 1500-Hz pure tone. With the f_0 set to infinity, the FDL data should

---

[3] HI $f_{0,PE}$ and ERB estimates measured at the high stimulus level were compared to the NH low-level estimates because no high-level NH data was available (asterisks in Table 5.3).

[4] The sigmoid function was defined as:

$$\log[f_0 DL\,(\%)] = \log(f_0 DL_{max}) + \frac{1}{\sqrt{\pi}}\log\left(\frac{f_0 DL_{min}}{f_0 DL_{max}}\right)\int_{m\log(f_0/f_{0,tr})}^{\infty} e^{-[\log(f_0')]^2}\, d[\log(f_0')] \qquad (5.1)$$

[5] For subject I1, the f_0 DL at 750 Hz, the largest f_0 tested in this subject, was significantly *smaller* than the 1500-Hz pure-tone FDL ($p<0.005$).

only directly influence the estimate of the parameter f_0 DL_{min}, since changes in $f_{0,tr}$, f_0 DL_{max} and m do not affect the value of the sigmoid function at an f_0 of infinity. Because the non-monotonicities observed in the sine-phase f_0 DL data prohibited a satisfactory fit for some subjects, only the random-phase data were analyzed in this way. The sigmoid functions that best fit the random-phase data are shown as solid curves in Fig. 5.2. While fitted curves are shown for the mean NH data in the lower two panels in Fig. 5.2, fits were made for each individual NH subject for the regression analyses described in Sec. 5.7.2, below.

Table 5.3. Best-fit estimates for individual subjects of three aspects of the f_0 DL data and three frequency selectivity estimates. Logarithmic transformations of the data shown here were used in the correlation analyses of Figs. 5.6-5.8 and Tables 5.4 and 5.5. Boldface entries indicate values for HI subjects that fell more than two deviations above the mean NH values at a comparable level. (*) In cases where no NH data was available at the high level, the high-level HI values were compared to the low-level NH mean.

		f_0 discrimination			Frequency selectivity		
	Subject	$f_{0,tr}$ (Hz)	f_0 DL_{min} (%)	f_0 DL_{max} (%)	$f_{0,PE}$ (Hz)	$f_{m,tr}$ (Hz)	ERB (Hz)
Hearing-impaired	I1	182.7	**1.84**	**32.86**	**275.7**	**287.7**	395.6
	I2	**214.9**	**1.44**	13.93	**246.2**	**143.3**	**406.2**
	I3	187.3	1.29	**21.61**	154.4	**166.0**	**482.0**
	I4	**203.5**	**1.38**	**25.31**	167.2	**153.2**	**517.2**
	I5	**403.2**	**2.02**	14.84	**489.0**	**230.7**	**543.1**
	I6	189.6	0.75	11.16	165.3	68.0	**409.4**
	I7	**493.6**	1.76	10.42	**446.9***	308.4	**618.7***
	I8	**345.1**	**3.60**	**23.86**	**330.9**	**152.0**	**454.0**
	I9(l)	250.5	**1.99**	17.17	**259.8***	223.9	**564.7***
	I9(r)	**687.4**	**3.45**	21.42	**462.7***	**341.0**	**941.5***
	I10	**231.9**	1.21	14.91	**241.1**	**171.6**	**623.0**
Normal-hearing low level	N1	192.2	0.66	5.87	186.2	128.5	343.8
	N2	173.8	0.96	9.82	175.4	136.8	347.7
	N3	184.1	1.10	11.68	204.7	131.4	282.5
	Mean	183.4	0.91	9.12	188.8	132.2	324.7
	St. Dev.	9.2	0.22	2.97	14.8	4.2	36.6
Normal-hearing high level	N1	196.4	1.06	5.79		269.4	
	N2	246.5	1.59	13.40		260.0	
	N3	160.8	1.48	31.40		174.5	
	Mean	201.2	1.38	16.86		234.6	
	St. Dev.	43.1	0.28	13.15		52.3	

(5.7.1.1) The f_0 DL transition point ($f_{0,tr}$)

The $f_{0,tr}$, one of the parameters in the sigmoid fitting procedure, provides an estimate of the f_0 for which f_0 DLs were halfway (on a log scale) between maximum and minimum. One-tailed student t-tests compared $f_{0,tr}$ estimates (and subsequent summary measures) between NH and HI groups, with the degrees of freedom (*df*) adjusted for unequal sample variances wherever necessary (Table 5.4).[6] With the data pooled across levels, $f_{0,tr}$ were significantly greater in the HI than the in NH group. The difference in $f_{0,tr}$ remained significant even when the two level ranges were analyzed separately, with stimuli presented in TEN at 50 dB and in the 60-65 dB SPL/ERB_N range grouped into the "low-level" and "high-level" conditions, respectively. Seven out of eleven HI ears had an $f_{0,tr}$ than was more than two standard deviations above the NH mean at a comparable level.

Table 5.4. One-tailed t-test analyses comparing three aspects of the f_0 DL data between groups of NH and HI listeners. Analyses were performed for NH and HI data as a whole and for data grouped by stimulus level. Boldface entries indicate a significant difference in a given variable between NH and HI listeners. Degrees of freedom (*df*) are adjusted for unequal sample variances.

Data subset	Number of data points NH	HI	Variable	t	df	p
All data	6	11	$f_{0,tr}$	2.56	13.2	**0.012**
	6		f_0 DL_{min}	2.36	13.8	**0.017**
	6		f_0 DL_{max}	1.78	6.84	0.060
	3		$f_{0,PE}$	2.75	11.8	**0.009**
	6		$f_{m,tr}$	0.37	13.2	0.359
	3		ERB	4.67	7.84	**<0.001**
Low level	3	8	$f_{0,tr}$	2.25	7.94	**0.027**
			f_0 DL_{min}	2.47	6.65	**0.022**
			f_0 DL_{max}	3.10	3.73	**0.020**
			$f_{0,PE}$	1.67	8.21	0.066
			$f_{m,tr}$	1.28	7.21	0.121
			ERB	4.28	4.81	**0.004**
High level	3	3	$f_{0,tr}$	2.48	2.67	**0.050**
			f_0 DL_{min}	2.18	3.32	0.055
			f_0 DL_{max}	0.28	2.74	0.399
			$f_{m,tr}$	1.16	3.97	0.156

[6] The application of a Bonferoni correction would mean that only one of the t-tests shown in Table 5.4 would yield a significant difference (ERB, all data). However, such a correction factor is excessively conservative for this analysis, because (1) it was hypothesized that HI listeners would show a deficit for most of the measures, and (2) without the Bonferoni correction, 9 out of 16 t-tests showed significant differences, an outcome with a very small probability of random occurrence with the $p<0.05$ criterion.

(5.7.1.2) Maximum and minimum f_0 DL values (f_0 DL_{max} and f_0 DL_{min})

These values, also derived from the sigmoidal fit to data in experiment 5A, provide estimates of the f_0 discrimination performance associated with unresolved and resolved harmonics, respectively. They did not form part of the original hypothesis regarding pitch discrimination and frequency selectivity, but certain relationships were found between f_0 DL_{min} and the other measures, which are described in the correlational analyses below. One-tailed student t-tests (Table 5.4) showed that the f_0 DL_{min} was significantly greater in HI than NH listeners when pooled across level and for the low-level data alone, but just failed to reach significance at the high level, perhaps because of the small number of subjects. The elevated f_0 DL_{min} in HI listeners suggest that the frequencies of individual resolved harmonics are more poorly encoded, consistent with previous studies of pure-tone frequency discrimination in listeners with SNHL (e.g. Tyler *et al.*, 1983; Moore and Glasberg, 1986; Moore and Peters, 1992). The effect of hearing loss on f_0 DL_{max} was less clear. There was no significant difference between NH and HI groups for the data pooled across level, although a significant difference emerged at the lower level when it was analyzed separately, and f_0 DL_{max} estimates were more than two standard deviations above the NH mean for four HI subjects.

(5.7.1.3) The phase-effect transition point ($f_{0,PE}$)

This measure is also derived from experiment 5A (Figure 5.2), but relates to the effect of phase on f_0 DLs, providing an estimate of harmonic resolvability based on the idea that the relative phase between successive components should only affect f_0 DLs if the components are unresolved and interact within individual auditory filters. In all subjects, f_0 DLs were larger in the random- than the sine-phase conditions for low but not for high f_0's, consistent with the idea that complexes with high f_0's contain resolved harmonics. This observation was confirmed by two-factor (f_0 and phase) ANOVAs performed on the f_0 DL data for each individual subject, with all 17 subjects showed a significant ($p<0.05$) interaction between f_0 and phase.

The f_0 DLs phase effect (PE) was defined as the ratio between the f_0 DLs measured in random- and sine-phase conditions. Resampling was performed to obtain all possible estimates of the PE by recalculating the PE 25 times for each f_0, once for each combination of the five repeated f_0 DL

measurements made for each phase relationship. The mean and standard deviations of the 25 PE estimates for each f_0 are plotted in Fig. 5.5 for two of the sample HI listeners for whom the f_0 DL data was plotted in Fig. 5.2. Because the random-phase conditions were always tested before the sine-phase conditions, general differences between the random- and sine-phase conditions could be attributed to learning effects. To control for this possibility, "phase-effect" ratios were calculated between the 1500-Hz pure-tone FDL measurements that were interspersed in each f_0 DL phase condition.

For all subjects, the PE was generally greater than one for low f_0's (f_0 DLs affected by component phase), and approximately equal to one for high f_0's (no phase effect). To estimate the transition f_0 at which phase no longer affected f_0 DLs for each subject, a sigmoid function with four free parameters was fit to the log-transforms of the 25 PE estimates at each f_0 (solid curves in Fig. 5.5)[7]. As with the fits to the random-phase f_0 DL data, the PE estimates for the 1500 Hz pure tone were included in the fitting procedure, with f_0 set to infinity. This was done instead of setting the value of the sigmoid function to zero for infinite f_0's to allow for some flexibility in the value of the PE function at high f_0's depending on the variance in the pure-tone FDLs and possible learning effects. The PE transition f_0 ($f_{0,\mathrm{PE}}$) was defined as the f_0 for which the PE was halfway between its maximum and minimum values (vertical dashed lines in Fig. 5.5). One-tailed t-tests (Table 5.4) showed the $f_{0,\mathrm{PE}}$ to be significantly greater ($p<0.05$) in HI than NH subjects for the data taken as a whole, but not for the low-level data alone (sine-phase f_0 DLs were not measured for NH subjects at the high level). Eight out of eleven HI ears had $f_{0,\mathrm{PE}}$'s more than two standard deviations above the NH low-level mean (Table 5.3).

[7] For the NH listeners, the data at 500 Hz were not included in the PE analysis because sine-phase measurements were not performed at that f_0.

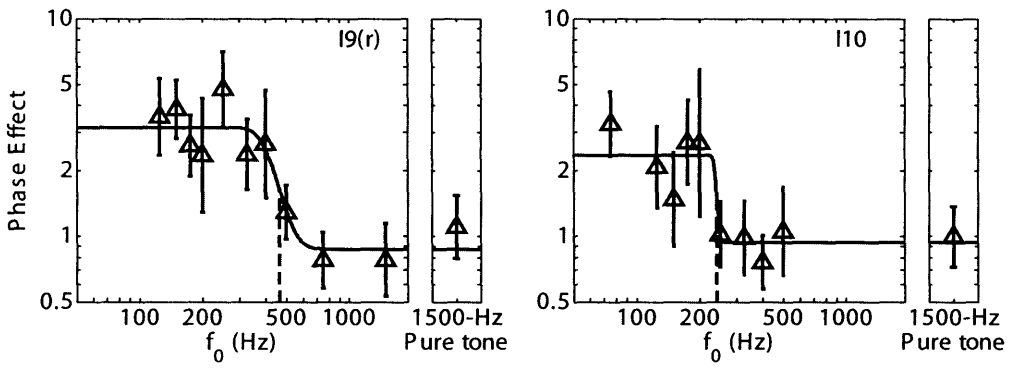

Figure 5.5. The f_0 DL phase effect (PE), defined as the ratio between random- and sine-phase f_0 DLs, for two sample HI subjects. Error bars indicate the standard deviation across the 25 PE estimates for each HI listener. Dashed curves indicate the sigmoid functions that best fit the PE data. Vertical dashed lines indicate the phase-effect transition f_0 ($f_{0,PE}$) defined as the f_0 for which the phase effect was halfway (on a log scale) between maximal and minimal.

(5.7.1.4) Maximum modulation frequency

This measure, derived from the modulation discrimination data of Experiment 5B, provided another estimate of component resolvability, based on the idea that the phase difference between the QFM and SAM should only be detectable based on the peripheral interaction of unresolved components. A sigmoid function fixed at 100% and 33% correct at the extremes was fit (minimum least squares) to the percentage correct data as a function of the log-transformed f_m's (dashed curves in Fig. 5.3). The 67% correct point of this function was taken as the estimate of the transition f_m ($f_{m,tr}$) between resolved and unresolved harmonics (vertical dashed lines in Fig. 5.3). Contrary to the hypothesis that HI listeners with wider peripheral filters should perform better than NH listeners, one-tailed t-tests (Table 5.4) did not find the $f_{m,tr}$ to be significantly greater in HI listeners for the data as a whole or for each level anaylzed separately. This could be due to a deficit in modulation processing by some HI listeners that offset any performance benefit deriving from wider auditory filters. Nevertheless, the $f_{m,tr}$ was more than two standard deviations above the NH mean at a comparable level for eight out of the eleven HI ears (Table 5.3).

(5.7.1.5) Equivalent rectangular bandwidth (ERB)

The ERB of the filter shape that best fit the notched-noise masking data (Experiment 5C) provided a third estimate of peripheral frequency selectivity. ERBs were significantly greater in

HI than NH listeners in one-tailed t-tests (Table 5.4) based on all of the data and for the low-level data alone (ERBs were not measured for NH subjects at the high level).

5.7.2. Regression analyses

Table 5.5 lists the results of single regression analyses performed with each of the three measures of f_0 discrimination ($f_{0,tr}$, f_0 DL_{min} and f_0 DL_{max}) treated as the dependent variable and each of the three estimates of frequency selectivity ($f_{0,PE}$, $f_{m,tr}$ and ERB) or the degree of hearing loss at 1.5 kHz ($HL_{1.5k}$) treated as the independent variable. The results of correlational analyses between each of the three measures of frequency selectivity are also shown. R^2 values are listed, along with an indication of the significance of each correlation (boldface indicates $p<0.05$), for analyses conducted with one data point per subject (N=14), only the HI subjects (N=11) and, where applicable, all data including two stimulus levels for each NH subject (N=17). N/A indicates that NH listeners were only tested at the low level for one of the measures in a given correlation, such that 17 data points were not available. The R^2 and p shown in each correlation plot (Figs. 5.6, 5.7 and 5.8) are based on fourteen data points, one for each NH subject tested at the low level and one for each HI subject.

Table 5.5. Summary of correlations between f_0 DL and frequency selectivity measures. Boldface entries indicate significant ($p<0.05$) correlations. 'N/A' indicates that NH listeners were tested at only one level for at least one of the measures associated with a given cell, such that 17 data points were not available.

		f_0 DL measures			Frequency selectivity measures		
Data included in analysis	Independent Variable	$f_{0,tr}$	f_0 DL_{min}	f_0 DL_{max}	$f_{0,PE}$	$f_{m,tr}$	ERB
All subjects	$HL_{1.5k}$	**0.50**	**0.52**	**0.38**	**0.51**	**0.52**	**0.68**
N=14	$f_{0,PE}$	**0.80**	**0.57**	0.04		**0.56**	**0.37**
	$f_{m,tr}$	**0.44**	**0.48**	0.17			**0.41**
	ERB	**0.62**	**0.38**	0.14			
HI subjects	$HL_{1.5k}$	**0.58**	0.34	0.01	**0.64**	**0.79**	**0.50**
only	$f_{0,PE}$	**0.76**	**0.52**	0.02		**0.51**	0.28
N=11	$f_{m,tr}$	**0.37**	**0.44**	0.08			0.35
	ERB	**0.60**	0.21	0.01			
Two levels	$HL_{1.5k}$	**0.56**	**0.39**	**0.25**	N/A	0.15	N/A
for each NH	$f_{0,PE}$	N/A	N/A	N/A		N/A	N/A
subject	$f_{m,tr}$	**0.31**	**0.37**	0.03			N/A
N=17	ERB	N/A	N/A	N/A			

(5.7.2.1) The relationship between the f_0 DL transition point and frequency selectivity

Figure 5.6 shows the data and regression line for the log-transformed $f_{0,tr}$ plotted as a function of $HL_{1.5k}$ and each of the three log-transformed frequency selectivity estimates. The $f_{0,tr}$ was significantly correlated to $HL_{1.5k}$ [Fig. 5.6(a)], further supporting the conclusion that the deficit in f_0 discrimination performance is related to hearing impairment. With TEN level partialled out of the analysis, the log-transformed $f_{0,tr}$ was still significantly correlated with $HL_{1.5k}$ ($p<0.005$), suggesting that the dependence of $f_{0,tr}$ on hearing loss was not an epiphenomenon of a dependence on stimulus level. The $f_{0,tr}$ was significantly correlated to each of the three estimates of peripheral frequency selectivity, even when only the HI data were included in the analysis (N=11). However, the significance of the correlation between $f_{0,tr}$ and ERB was dependent on a single subject, I9(r), in the HI-only analysis. With I9(r) was excluded from the analysis (total N=10), the correlation became non-significant ($p=0.095$).

These significant correlations do not prove that the relationship between frequency selectivity and $f_{0,tr}$ is a direct one. Because $f_{0,tr}$ [Fig. 5.6(a)] and the measures of frequency selectivity (Fig. 5.7) were each significantly correlated with $HL_{1.5k}$[8], the correlations between $f_{0,tr}$ and each estimate of frequency selectivity could be epiphenomena of their common dependencies on $HL_{1.5k}$. Partial regression analyses were performed to investigate this possibility, whereby the variance in the $f_{0,tr}$ or f_0 DL_{min} accounted for by variance in $HL_{1.5k}$ was removed from the analysis (Table 5.6). None of the partial correlations involving $f_{m,tr}$ or ERB as the frequency selectivity variable were significant. However, with $f_{0,PE}$ as the frequency variable, the partial correlation with $f_{0,tr}$ was statistically significant, suggesting that the correlation between $f_{0,tr}$ and $f_{0,PE}$ was not an epiphenomenon of their common dependence on $HL_{1.5k}$.

[8] Although the correlation between $f_{m,tr}$ and $HL_{1.5k}$ became non-significant ($p=0.15$) when the high-level NH data was included in the analysis (total N=17), this is not a fair analysis because the $HL_{1.5k}$ data is not affected by level.

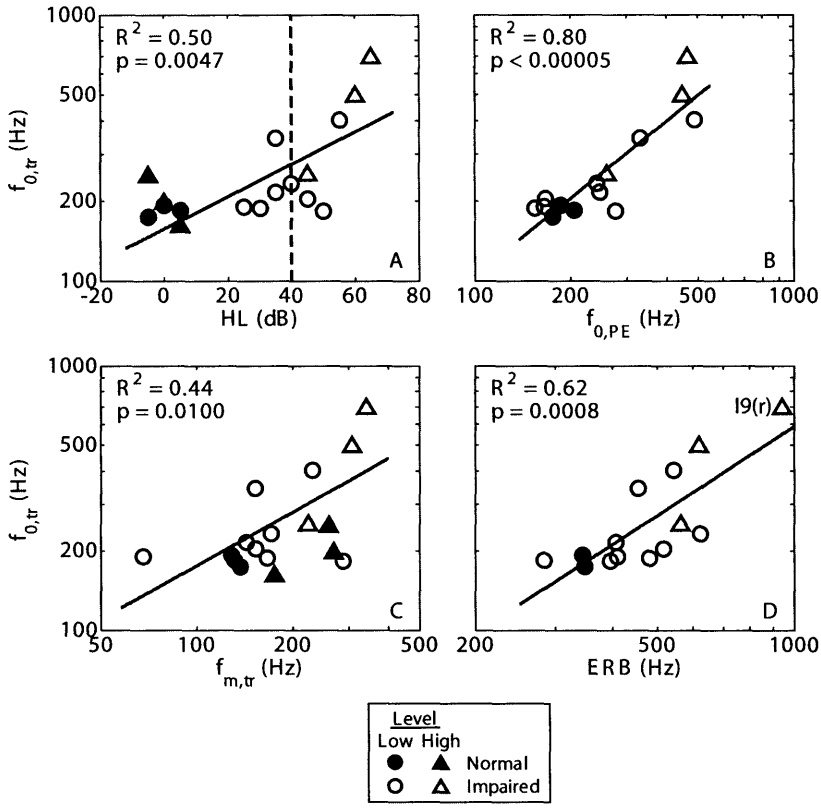

Figure 5.6. The $f_{0,tr}$ was significantly correlated with (a) the audiometric threshold at 1.5 kHz ($HL_{1.5k}$), and each of the three estimates of frequency selectivity: (b) $f_{0,PE}$, (c) $f_{m,tr}$ and (d) ERB. The NH data for stimuli presented at the high level (filled triangles) were not included in the regression analyses. Vertical dashed line in (a) represents the cutoff between "normal-to-mild" and "moderate" hearing loss groups that yielded significantly different regression coefficients (see Sec. 5.7.2.1).

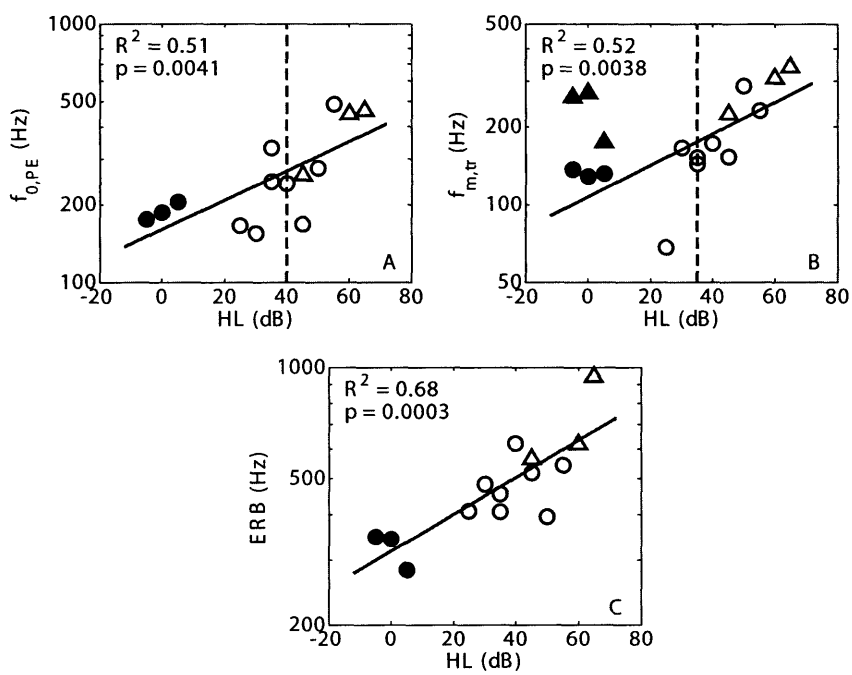

Figure 5.7. The three estimates of frequency selectivity, (a) $f_{0,PE}$, (b) $f_{m,tr}$ and (c) ERB, were each significantly correlated to $HL_{1.5k}$. Vertical dashed lines in (a) and (b) represent the cutoffs between "normal-to-mild" and "moderate" hearing loss groups that yielded significantly different regression coefficients (see Section 5.7.2.1). See the legend of Fig. 5.6 for symbol definitions.

Table 5.6. Results of partial-correlation analyses obtained by removing the variance accounted for by variance in $HL_{1.5k}$ from correlations between f_0 DL and frequency selectivity summary measures. Boldface entries indicate statistical significance (p<0.05).

f_0 DL Variable	Frequency selectivity variable	R^2	p
$f_{0,tr}$	$f_{0,PE}$	0.78	**0.002**
	$f_{m,tr}$	0.31	0.299
	ERB	0.48	0.100
f_0 DL_{min}	$f_{0,PE}$	0.49	0.088
	$f_{m,tr}$	0.37	0.219
	ERB	0.07	0.834

One additional aspect of the data supports the conclusion that the $f_{0,tr}$ depends on frequency selectivity *per se*. Figure 5.6(a) shows that although the $f_{0,tr}$ depends on HL, the relationship between the two measures is not linear. The $f_{0,tr}$ estimates remain roughly constant with hearing loss until HL increases above approximately 30-40 dB HL. To quantify this observation, the data were divided into two categories based on the degree of hearing loss at 1.5 kHz. Subjects with audiometric thresholds < 40 dB HL [vertical dashed line in Fig. 5.6(a)] were assigned to the "normal-to-mild" group (N=7), while those with thresholds ≥ 40 dB HL were assigned to the "moderate" group (N=7). A Porthoff (1966) analysis found the regression coefficients to be statistically different between the two groups (p<0.01). Furthermore, the correlation was significant for the "moderate" group (R^2=0.78, p<0.01), but not the "normal-to-mild" group (p=0.16), suggesting that hearing loss (dB HL) is a good predictor of the $f_{0,tr}$ only for hearing losses ≥40 dB HL.

In line with earlier studies of frequency selectivity in HI listeners (Tyler *et al.*, 1983; Nelson, 1991; Moore, 1998; Moore *et al.*, 1999), two of the three measures of frequency selectivity showed a similar behavior in that hearing loss had little effect on $f_{0,PE}$ or $f_{m,tr}$ until the loss exceeded approximately 40 dB. For the $f_{0,PE}$ vs. HL correlation, regression coefficients were significantly different (p<0.05) between the "normal-to-mild" and "moderate" groups, with the cutoff between the two groups defined at 40 dB HL [vertical dashed line in Fig. 5.7(a)]. The difference between regression coefficients just failed to reach significance for the $f_{m,tr}$ vs. HL

correlation (p=0.066) with a 40 dB HL cutoff between the two groups, although the difference became significant (p<0.05) when the cutoff was defined at 35 dB HL [vertical dashed line in Fig. 5.7(b)]. For both $f_{0,\mathrm{PE}}$ and $f_{m,tr}$, the correlations between frequency selectivity and dB HL were significant for the "moderate" group ($f_{0,\mathrm{PE}}$: R^2=0.70, p<0.05; $f_{m,\mathrm{tr}}$: R^2=0.72, p<0.05), but not for the "normal-to-mild" group ($f_{0,\mathrm{PE}}$: p=0.34; $f_{m,\mathrm{tr}}$: p=0.91). Similar results were not observed statistically for the ERB measure, were the regression coefficients were not significantly different between the two groups (p=0.55). The similar behavior as a function of dB HL for $f_{0,tr}$ and for two of the three frequency selectivity measures suggests that the relationship between $f_{0,tr}$ and frequency selectivity is not explained by audiometric thresholds alone. Further supporting this conclusion is the fact that the relationships between $f_{0,\mathrm{tr}}$ and the three measures of frequency selectivity were generally more linear than those between $f_{0,\mathrm{tr}}$ and $HL_{1.5k}$. There were no significant differences between regression coefficients between the two groups whether the cutoff was defined at 35 or 40 dB HLK. .

(5.7.2.2) f_0 DL_{max} and f_0 DL_{min}

The f_0 DL_{min} was significantly correlated with $HL_{1.5k}$ and each of the three measures of frequency selectivity (Fig. 5.8), although the correlations with $HL_{1.5k}$ and with ERB were somewhat weak, becoming non-significant when the NH data was removed from the analysis. In a multiple regression analysis with f_0 DL_{min} as the dependent variable and $HL_{1.5k}$ and TEN level as independent variables, f_0 DL_{min} was found to be correlated with $HL_{1.5k}$ (p<0.005) but not TEN level (p=0.087), suggesting that f_0 DL_{min} was mainly dependent on hearing loss.

As with $f_{0,\mathrm{tr}}$, the significant correlations between f_0 DL_{min} and frequency selectivity do not prove the existence of a direct relationship between peripheral frequency selectivity and f_0 discrimination performance for resolved complexes. However, when the contribution of $HL_{1.5k}$ was partialled out of the analyses, correlations between f_0 DL_{min} each of the frequency selectivity estimates were not significant (Table 5.6), leaving open the possibility that f_0 DL_{min} is dependent on audiometric thresholds and not frequency selecitivty *per se.*

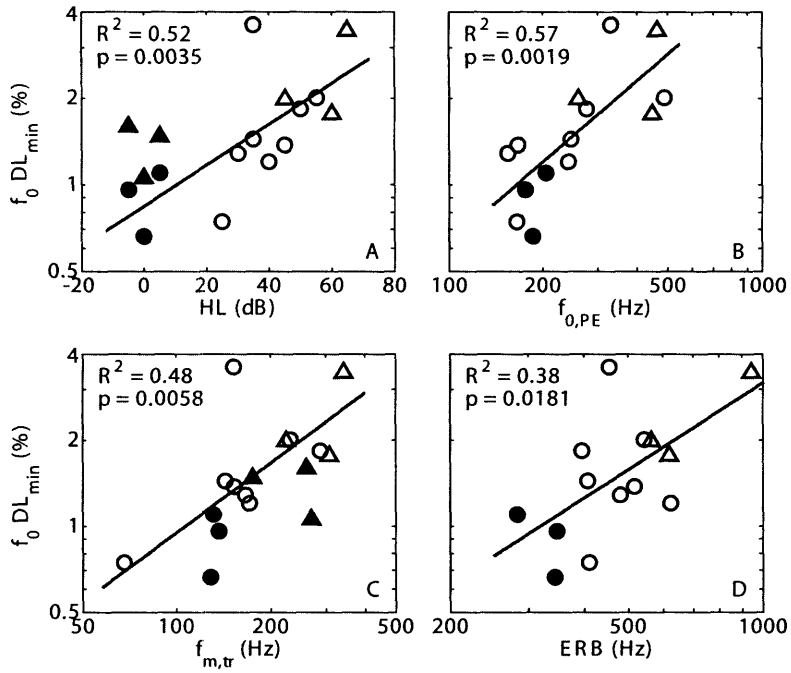

Figure 5.8. The f_0 DL_{min} was significantly correlated with (a) the audiometric threshold at 1.5 kHz ($HL_{1.5k}$), and each of the three estimates of frequency selectivity: (b) $f_{0,PE}$, (c) $f_{m,tr}$ and (d) ERB. See the legend of Fig. 5.6 for symbol definitions.

Estimates of f_0 DL_{max} were significantly correlated with $HL_{1.5k}$ (not shown), suggesting that SNHL is associated with an impairment in f_0 discrimination with unresolved harmonics. However, a multiple regression analysis with f_0 DL_{max} as the dependent variable and both $HL_{1.5k}$ and TEN level as independent variable, showed that neither regression coefficient was significant (p=0.065 and p=0.93, respectively), an ambiguous result leaving open the possibility that f_0 DL_{max} was dependent on stimulus level and not hearing loss *per se*. Furthermore, the f_0 DL_{max} was not significantly correlated with any of the three measures of peripheral frequency selectivity (not shown). This suggests that if there is a deficit in f_0 DL_{max} related to HI, that some mechanism not directly related peripheral frequency selectivity, such as the ability to process envelope modulations, may be responsible.

5.8 Discussion

5.8.1. The relationship between f_0 discrimination and frequency selectivity

The results show a strong correlation between frequency selectivity and f_0 discrimination performance in listeners with SNHL. The $f_{0,tr}$ was correlated across the 11 HI and three NH subjects to each of the three measures of frequency selectivity, supporting the hypothesis that the spacing between harmonics required for good f_0 discrimination performance is related to peripheral frequency selectivity, The significant correlations between f_0 discrimination and

frequency selectivity were not a result of generally poor performance in psychoacoustic tasks by HI listeners, because *good* performance in modulation discrimination (Experiment 5B, large $f_{m,tr}$) was correlated with *poor* performance in f_0 discrimination (Experiment 5A, large $f_{0,\mathrm{tr}}$ and f_0 $\mathrm{DL}_{\mathrm{min}}$).

Several pieces of evidence support the idea of a direct relationship whereby f_0 discrimination performance is dependent on frequency selectivity. First, in multiple regression analyses with both HL and $f_{0,\mathrm{PE}}$ as dependant variables, the regression coefficient was significantly different from zero for $f_{0,\mathrm{PE}}$ but not for dB HL. Second, two measures of frequency selectivity behaved in a similar nonlinear fashion as function of dB HL as the $f_{0,tr}$, where there was little change in any of these variables until hearing loss exceeded approximately 35 dB. Frequency selectivity and the $f_{0,tr}$ behaved similarly across the range of audiometric thresholds, with little effect of HL on either measure until below 40 dB HL, suggesting that dB HL is not a good predictor of performance in the "normal-to-mild" range. Finally, the data of Chapter 4 in NH subjects show that stimulus level affected both frequency selectivity and the $f_{0,\mathrm{tr}}$ in the same way as hearing loss in the current study. Thus, peripheral frequency selectivity, the common denominator between these two studies, the increases in $f_{0,\mathrm{tr}}$ and f_0 $\mathrm{DL}_{\mathrm{min}}$.

This study finds a significant correlation between frequency selectivity and f_0 discrimination performance, while others than have looked for such a relationship have not (e.g. Moore and Glasberg, 1990; Moore and Peters, 1992). An important difference is that the present study sought a relationship between the transition between large and small f_0 DLs and frequency selectivity. If the main role of peripheral spectral tuning is to resolve individual harmonic frequencies, as proposed by spectral and some spectrotemporal models of pitch perception, then the transition measure may be more directly related to frequency selectivity than absolute f_0 DL. A measure of the transition point also takes into account the possibility that other factors, such as the ability to process temporal phase-locking information, could influence the absolute f_0 DL The only previous study to compare this aspect of f_0 discrimination performance with estimates of peripheral frequency selectivity in HI listeners did not estimate the latter in a sufficient number of listeners to garner a correlation (Hoekstra, 1979).

The current study also found significant correlations between each of the three estimates of frequency selectivity and the f_0 DL_{min}. As the f_0 DL_{min} estimate was tightly coupled to the 1500-Hz pure-tone FDLs, the log-transformed FDLs were also significantly correlated to each of the three log-transformed estimates of frequency selectivity ($f_{0,PE}$: R^2=0.55, p<0.005; $f_{m,tr}$: R^2=0.51, p<0.005; ERB: R^2=0.32, p<0.05). This result conflicts with several previous studies of the relationship between pure-tone frequency discrimination and peripheral frequency selectivity that have found only weak or non-significant correlations between the two types of measure (e.g. Tyler *et al.*, 1983; Moore and Glasberg, 1986; Moore and Peters, 1992). There are several possible reasons for this discrepancy.

One possible basis for this discrepancy concerns the method of frequency selectivity estimation. Tyler *et al.* (1983) used PTCs to estimate frequency selectivity, but tested only three data points on the PTC to limit the measurement time, which may have limited the accuracy of their measure. Moore and Peters (1992) and Moore and Glasberg (1986) characterized frequency selectivity using notched-noise masking data, as in experiment 5C of the current study. Of the three methods of frequency selectivity in the current study, the ERB based on notched-noise masking data yielded the weakest correlation with the FDL data, becoming non-significant when only the HI subjects were included in the analysis. A weak ERB-FDL correlation could be due to differences in stimulus type, or perhaps the additional step of fitting the masking data to a model auditory filter increased the variability in the frequency selectivity estimates.

The use of background noise in the current study may also underlie the departure from previous investigations that found weak or absent correlations between FDLs and frequency selectivity when stimuli were presented in isolation. Investiagtions of FDLs as a function of stimulus level in NH listeners suggest that frequency selectivity may have more influence on frequency discrimination when pure tones are presented in a background noise. For pure-tones presented in isolation, FDLs generally decrease (improve) with increasing stimulus level (Wier *et al.*, 1977). In contrast, when pure tones are presented to NH subjects at constant SL in background noise, FDLs worsen with increasing stimulus level at frequencies of 1.5 kHz (Chapter 4) or above (Dye and Hafter, 1980), consistent with the observed reduction in peripheral frequency selectivity (Chapter 4). It may be that in the absence of a background noise, a higher-level stimulus excites

a larger number of auditory nerve fibers, thereby distributing information for frequency discrimination over a broader tonotopic regions and improving frequency discrimination (Green and Luce, 1974) in such a way that offsets the effects of a reduction in frequency selectivity. Florentine and Buus (1981) invoked a similar idea involving the spread of excitation to explain the deviation from Weber's law in pure-tone intensity discrimination.

The use of TEN in the current study to reduce differences in SL and SPL across subjects may have also reduced the contribution of absolute- (e.g. Chapter 4) and sensation-level (e.g. Hoekstra, 1979) influences on FDLs. Moore and Glasberg (1986) and Tyler *et al.* (1983) presented pure tones to HI listeners at a constant SPL (80 and 94 dB, respectively), with the equivalent SL ranging approximately 50 dB and 80 dB across subjects, respectively. Moore and Peters (1992) presented tones at a constant 25 dB SL, yielding an approximately 50-dB SPL range across the HI subjects.

Finally, it is worth pointing out that the pure-tone frequency discrimination measurements reported here were only performed at a single frequency, 1.5 kHz. It is not known whether similar effects would be obtained with hearing loss at lower frequencies. Dye and Hafter (1980) showed that for lower-frequency tones (500 Hz and 1 kHz), increasing the level of both the tone and the background noise tended to improve rather than impair frequency discrimination performance, suggesting that frequency selectivity may have less effect on lower-frequency tones. On the other hand, level is known to have less effect on frequency selectivity at low frequencies (1 kHz and below) than at high frequencies (Baker *et al.*, 1998). Thus, the results of Dye and Hafter (1980) at low frequencies may reflect the absence of an effect of level on frequency selectivity rather than the absence of an effect of frequency selectivity on FDLs.

5.8.2. Modulation discrimination

Although $f_{m,tr}$ and $f_{0,\mathrm{tr}}$ were significantly correlated, HI subjects performed worse at modulation discrimination relative to their f_0 discrimination performance than the NH listeners. In Fig. 5.6(c), the $f_{m,\mathrm{tr}}$ for NH listeners tested at the higher level (filled triangles) generally fell to the right of the regression line, indicating that NH subjects performed better at modulation discrimination at this level than would be predicted from their f_0 DL data based on the

relationship between $f_{0,tr}$ and $f_{m,tr}$ for the other 14 data points. This observation was supported by a significant one-tailed independent-sample t-test, adjusted for unequal variances, comparing NH and HI listeners based on the log-transformed ratio $f_{0,tr}$ / $f_{m,tr}$ [t(13.7) = 1.84, p<0.05]. One interpretation of this result is that HI listeners may have some deficit in modulation processing that reduces their discrimination performance below what they might achieve based on peripheral filter bandwidths alone. Although HI subjects do not generally show deficits in modulation processing when signals are presented to NH and HI listeners at an equal SL (Bacon and Gleitman, 1992), the wideband background noise was not used in experiment 5B, such that signals were presented at a higher SL for NH listeners. Alternatively, the relatively small $f_{m,tr}$ (relative to $f_{0,tr}$) in some HI listeners may reflect an absolute upper f_m limitation whereby modulation processing performance begins to deteriorate even in NH listeners for f_m's greater than about 150 Hz (Kohlrausch *et al.*, 2000). Either way, a modulation processing limitation in HI listeners may partially to the lack of a significant difference in $f_{m,\mathrm{tr}}$ that was observed between the NH and HI groups (Table 5.4) despite the significant positive correlation between $f_{m,\mathrm{tr}}$ and $HL_{1.5k}$ [Fig. 5.7(b)].

Another aspect of the modulation discrimination data that may be related to a limitation in modulation processing, due either to SNHL or an absolute f_m limit, is that $f_{m,\mathrm{tr}}$ estimates were generally smaller than $f_{0,tr}$ estimates. This result is reflected in the regression analysis, where the estimate of the linear regression coefficient (B_1) was significantly less than one (0.56 with 95% confidence interval ±0.22). An analysis of this result interpreted in terms of harmonic resolvability alone would suggest that the limit of harmonic resolvability, as estimated by the modulation discrimination task, occurs at a lower f_0 (*higher* harmonic number) than the $f_{0,tr}$. One possible interpretation of this discrepancy is that the $f_{m,tr}$, which relies on wide peripheral filters for good performance, provides an *upper* limit on the extent of harmonic resolvability, whereas estimates based on listeners' ability to hear out harmonics (Chapter 2; Plomp, 1964; Moore and Ohgushi, 1993) or phase effects on f_0 DLs (Section 5.7.1.3 and Moore *et al.*, 2005) which rely on narrow filters for better performance, provide a *lower* limit.

5.8.3. Comparing estimates of frequency selectivity

All three estimates of frequency selectivity were correlated to one another (Table 5.5), suggesting that each measures a related aspect of the functioning of the auditory periphery. Nevertheless, there were important differences between each of the measures with regard to the measurement time, the proportion of variance accounted for in the f_0 DL data, and the robustness of the correlations with the f_0 DL data. Of the three measures, the f_0 DL phase effect was the most time consuming test of peripheral frequency-selectivity, necessitating a total training and testing time on the order of 10 hours per subject, but it also accounted for the largest proportion of the variance in the $f_{0,tr}$, although the difference between R^2 values was only significant ($p<0.05$) between $f_{0,\mathrm{PE}}$ and $f_{m,\mathrm{tr}}$ and not between $f_{0,\mathrm{PE}}$ and ERB. Furthermore, $f_{0,\mathrm{PE}}$ was the only estimate of frequency selectivity that was significantly correlated with $f_{0,\mathrm{tr}}$ when HL was partialled out in the multiple regression analysis described in the previous section. The modulation-discrimination estimate took the least amount of test time (approximately 2-3 hours per subject) but also accounted for the smallest proportion of the variance in $f_{0,tr}$ (although the difference between R^2 estimates was not significant between the correlations with $f_{m,\mathrm{tr}}$ and ERB). The ERB estimates fell in between, necessitating approximately 4 hours training and test time per subject, and accounting for a moderate proportion of the variance in $f_{0,tr}$ that was not significantly different from either of the other two measures. However, the ERB was an especially weak predictor of the f_0 discrimination data when only the HI data was included in the analysis, relying on a single HI subject for the significant correlation with $f_{0,tr}$, as discussed in Section 5.7.2.1. The proportions of the variance in f_0 $DL_{\min}$ accounted for were not statistically different between the three frequency selectivity estimates when all 14 NH and HI subjects were included in the analysis, although ERB was not significantly correlated with the f_0 $DL_{\min}$ in the 11 HI subjects.

In summary, the f_0 DL phase effect was the strongest predictor of the $f_{0,\mathrm{tr}}$. This may be because $f_{0,\mathrm{tr}}$ and $f_{0,\mathrm{PE}}$ are both derived from f_0 DL data, and obtained using identical stimuli and measurement procedures. The more independent measures of frequency selectivity were less robust, which may be partially explained by the fact that the stimuli and tasks in these tests were not identical to those in the f_0 DL experiment. In the case of the $f_{m,\mathrm{tr}}$, the results may have been influenced by limitations in modulation processing, as discussed in the previous section. In the

case of the ERB, inaccuracies in the intervening step of model fitting may have introduced additional variability. Despite the nuances of each measure of frequency selectivity, each was correlated to both $f_{0,tr}$ and f_0 DL_{min}, suggesting that each of these aspects of the f_0 DL data is related to peripheral frequency selectivity.

5.8.4. Perceptual implications for HI listeners

The results shown here indicate that listeners with SNHL experience a deficit in f_0 processing, directly related to a loss of peripheral frequency selectivity, that manifests itself in at least three ways. First, a larger spacing between adjacent harmonics is needed to yield the smallest possible f_0 DLs for a given subject and spectral region. This means that in everyday listening conditions, a larger proportion of possible stimulus f_0's will yield a weak pitch percept in these listeners. Second, even when harmonics are widely separated, the f_0 DLs are larger (poorer) than in NH listeners. Finally, the results of experiment 5A also show that listeners with SNHL had a higher $f_{0,PE}$ than normal, meaning that these are subjects will experience a detrimental effect of component phase on f_0 discrimination for a larger range of f_0's. As discussed in Section 5.5 this is because individual harmonics are more likely to interact within the wider auditory filters associated with SNHL. This effect is of particular importance in a reverberant environment, where a heterogeneous mixture of reflection delays (i.e. random phase) tends to "smear" the temporal envelopes (Houtgast *et al.*, 1980; Steeneken and Houtgast, 1980) at the output of auditory filters excited by unresolved harmonics (Qin and Oxenham, 2005). With wider filters, listeners with SNHL will be more susceptible to a negative impact of reverberation phase randomization on f_0 discrimination.

5.8.5. Implications for pitch models

The current findings corroborate the previous findings of Chapter 4 showing that in NH listeners, f_0 DL_{min} and $f_{0,tr}$ increased as a function of stimulus level in the same way as peripheral frequency selectivity. The current study extends this finding by establishing a relationship between f_0 discrimination and frequency selectivity in a large enough population of NH and HI subjects to yield significant correlations between the two measures. Because the findings of the two studies are similar with respect to the relationship between f_0 discrimination and frequency selectivity, the implications for models of pitch perception of the current HI results are the same

as those discussed in the previous manuscript (for full discussion, see Chapter 4). To summarize, these results are consistent with any pitch model that relies on peripheral frequency selectivity to explain why low-order harmonics yield better f_0 DLs than high-order harmonics. This includes "spectral" and "spectrotemporal" models that use place or timing information to extract the frequencies of individual resolved harmonics, as well as a recent version of the autocorrelation model (de Cheveigné and Pressnitzer, 2005) that depends on temporal response characteristics of auditory filters that are related to the filter bandwidths.

5.9. Summary and conclusions

Listeners with SNHL experience a deficit in f_0 discrimination that manifests itself in terms of an increase in the minimum spacing between harmonics required for f_0 DLs to transition from large (poor) to small (good). The f_0 DL transition point was significantly correlated to three different estimates of peripheral frequency selectivity, supporting the hypothesis that good f_0 discrimination performance depends on sharp peripheral frequency selectivity, and that listeners with SNHL experience a deficit in f_0 processing due to a reduction in frequency selectivity. Additionally, the best f_0 discrimination performance achieved by HI listeners was worse than that attained by NH listeners even when harmonics were spaced widely enough in frequency to yield relatively good f_0 discrimination performance associated with resolved harmonics. This effect, also observed for pure tones, was also correlated with two estimates of peripheral frequency selectivity in HI listeners, suggesting a role for place information in the frequency encoding of individual resolved harmonics. These results support "spectral" and "spectrotemporal" theories of pitch perception that rely on peripheral frequency selectivity to extract the frequencies of individual resolved harmonics.

Chapter 6. Summary and conclusions

1) In Chapter 2, the presentation of alternating harmonics to opposite ears did not shift the f_0 DL transition to a higher harmonic number, even though this presentation mode doubled the number of peripherally resolved harmonics. This result suggested that the f_0 DL transition may not reflect the peripheral resolvability of harmonics *per se*, but instead relies on harmonic number.

2) Based on this result, Chapter 3 tested an ad-hoc modification applied to a temporal autocorrelation model of pitch discrimination (Meddis and O'Mard, 1997), designed to account for the harmonic number dependence of f_0 DLs at a stage following peripheral processing, and therefore in a manner not based on peripheral resolvability. In this modified model, the range of autocorrelation lags, and therefore the range of stimulus f_0's for which a given channel could relay periodicity information, was limited relative to that channel's characteristic frequency (CF). This "CF-dependent temporal" model was able to account for the dependence of f_0 DLs on harmonic number whereas the standard autocorrelation model could not. Because the success of the modified model did not rely on peripheral harmonic resolvability, this model was consistent with an interpretation of the results of Chapter 2 that f_0 DLs are dependent on harmonic number but not on harmonic resolvability *per se*.

3) Because the modified model of Chapter 3 did not directly depend on the frequency selectivity of the auditory periphery to account for the harmonic number dependence, it would therefore be unlikely to predict an effect of the broadening of peripheral filters on the harmonic number dependence of f_0 DLs. Chapters 4 and 5 tested this prediction by measuring the influence of broadened filters on the harmonic number corresponding to the transition between good and poor f_0 discrimination performance. For harmonic stimuli bandpass filtered into a fixed spectral region, the f_0 DL transition shifted toward a higher f_0 (lower harmonic number) in two situations where frequency selectivity is known to be reduced: at high stimulus levels (Chapter 4) and in hearing-impaired listeners (Chapter 5). Furthermore, the shift f_0 DL transition point corresponded to (and in the case of the impaired listeners, were significantly correlated with) estimates of frequency selectivity under similar

conditions. These results are consistent with the idea that resolved harmonics are necessary for good f_0 discrimination performance.

4) The results of this thesis demonstrate an important role of frequency selectivity in the discrimination of the f_0 of complex sounds, suggesting that complex pitch is not derived from temporal information alone, but must also include spectral place information. A shift in the f_0 DL transition point did not occur with an artificial increase in peripheral resolvability resulting from the dichotic presentation of harmonic stimuli (Chapter 3), but did occur with a reduction in frequency selectivity (Chapters 4 and 5). The latter result argues against the CF-dependent autocorrelation model of Chapter 3, which did not depend on peripheral frequency selectivity to explain the dependence of f_0 DLs on harmonic number. Instead, this combination of results would be consistent with the following pitch models, as summarized in Table 6.1:

   a. A spectral or spectrotemporal harmonic template model that requires resolved harmonics for good f_0 discrimination performance, with the added constraint that harmonics must not only be resolved, but low-ordered such they would be resolved under normal circumstances. Some examples, discussed in Chapter 2, include harmonic templates that develop in response to naturally-occurring stimuli (Terhardt, 1974) or in response to across-channel coincidences between any form of wideband stimulation (Shamma and Klein, 2000), or templates that respond to a "central spectrum" representation (Zurek, 1979), whereby additional *peripherally* resolved harmonics provided under dichotic presentation in Chapter 2 would not be available to the central pitch processor. With these constraints, these spectral template models would account for the lack of a shift in the f_0 DL transition point under dichotic presentation (Chapter 2), but would still account for the shift in the f_0 DL transition point that accompanied reduced frequency selectivity (Chapters 4 and 5) in terms of a reduction in the number of resolved harmonic components.

   b. A CF-dependent temporal model whereby the range of periodicities that would yield good f_0 discrimination performance are limited relative to each channel's CF, like that

of Chapter 3, with the additional requirement that this periodicity limitation is related to the peripheral channel's bandwidth. In a recent model that satisfies these requirements (de Cheveigné and Pressnitzer, 2005), the lower limit on a range of detectable periodicities is limited by the duration of the impulse response of a peripheral filter. This model would account for the shift in the f_0 DL transition point that accompanied a reduction in frequency selectivity in Chapters 4 and 5 based on the shorter impulse response durations that are associated with wider filter bandwidths. This model would also account for the lack of a shift in the f_0 DL transition point with dichotic presentation because its ability to account for the harmonic number dependence of f_0 DLs stems not from the resolvability of harmonics, but from the characteristics of the auditory periphery that remains unchanged under dichotic stimulus presentation.

5) The important role for spectral information in f_0 discrimination is consistent with the idea that hearing-impaired listeners experience difficulty in discriminating pitch due to a loss of frequency selectivity. This suggests that efforts should be made to provide hearing-impaired listeners with spectral information and resolved harmonics in order to maximize the pitch discrimination abilities. Several possible ways of achieving this goal are discussed in the following chapter.

Table 6.1. Summary of implications of experimental results for models of pitch perception.

Model type	Examples	Would account for effect of		Additional constraint needed to account for Chapter 2 results
		Harmonic Number	Reduced Frequency Selectivity	
Spectral	Goldstein (1973) Wightman (1973) Terhardt (1974; 1979)	Yes	Yes	"Central spectrum" (Zurek, 1979) *or* Templates contain only first ten "normally" resolved harmonics
Spectro-temporal (Harmonic template)	Shamma & Klein (2000) Cedolin & Delgutte (2005)	Yes	Yes	
Temporal	Shouten (1940) Licklider (1951; 1959) Cariani & Delgutte (1996) Meddis & O'Mard (1997)	No	No	N/A – cannot account for harmonic dependence
CF-dependent temporal	Chapter 3	Yes	No	None needed
	de Cheveigné & Pressnitzer (2005)	Yes	Yes	None needed

Chapter 7. Implications and future directions

7.1 Improving the spectral pitch information provided to impaired listeners

7.1.1. Hearing aids

The present results suggest that resolved harmonics are important in providing f_0 information. Since harmonics are most likely to be resolved at low frequencies, where auditory filter bandwidths are the narrowest, information at these frequencies may be vital to accurate pitch discrimination. While many hearing-impaired listeners have relatively normal audiometric thresholds at low frequencies, closed hearing aid inserts that physically block the ear canal tend to prevent the relaying of low-frequency information. This occurs because the occlusion of the ear canal attenuates direct sound across all frequencies, while hearing aids tend to not amplify and deliver low-frequency (long wavelength) sound because of their small size and because hearing-aid fitting procedures will minimize amplification due to the absence of a hearing loss at low frequencies. Open air hearing-aid inserts are more likely to relay spectral pitch cues at low-frequencies to these listeners via natural air conduction. Such devices are already used to preventing the "booming voice" phenomenon (Moore, 1997), whereby low-frequency components in the patient's own voice is passed to the ear canal via bone conduction and amplified due to the closed cavity created by the occlusion .

An additional step that can be taken to optimize pitch information is to avoid overamplification. The results of Chapter 4 suggest that high stimulus levels adversely affect f_0 discrimination in NH listeners due to a decrease in peripheral frequency selectivity. If a patient with SNHL has some residual filter nonlinearity, then stimuli amplified to high levels could further reduce frequency selectivity, negatively impacting f_0 discrimination performance. Therefore the benefits generally associated with higher stimulus levels (e.g. audibility speech intelligibility) should to be balanced with the possible negative impacts of increased stimulus level on frequency selectivity.

7.1.2. Cochlear implants

The current results demonstrate that place information is important for the processing of f_0 information. While f_0 information can be relayed by temporal modulations in CI users where place-pitch information is unavailable, the f_0 discrimination performance associated with such

stimuli is much poorer than that associated with well-resolved harmonics (Carlyon *et al.*, 2002). The results shown here imply that to give CI users the best chance at receiving strong, useful f_0 information, place cues are of high importance. Recently, a novel approach was taken by Geurts and Wouters (2004) to give rate-place pitch information to CI users using triangular analysis filters. CI users were successful in discriminating f_0 based on place cues alone, although it is not clear whether this successful behavior was based on the extraction of the missing f_0 *per se*. "Current steering", whereby the place of cochlear stimulation is finely tuned by adjusting the relative amount of current presented simultaneously to adjacent electrodes, also has the potential to increase spectral resolution (Frohne-Büchner *et al.*, 2005; Koch *et al.*, 2005). Such a stimulation strategy could conceivably relay enhanced spectral complex pitch information to cochlear implantees, although the number of resolved harmonics that could be delivered would still be limited by the electrode frequency spacing, because this strategy would not allow the creation of multiple excitation pattern peaks between two electrodes. Finally, in cases where some low-frequency acoustic hearing is available, the preservation of this acoustic frequency range, where harmonics are most likely to be resolved even in an impaired ear, could provide useful f_0 information to CI users (Turner *et al.*, 2004; Kong *et al.*, 2005; Qin, 2005).

7.2 Future directions

The research described in this thesis has identified a link between peripheral frequency selectivity and complex pitch discrimination. Some fundamental and important questions regarding the coding of complex pitch remain. Two possible directions for future research are described below.

7.2.1. The role of temporal information

The "spectral" versus "temporal" pitch processing debate is sometimes portrayed in terms of black and white, either one or the other. However, the fact that peripheral frequency selectivity is linked to complex pitch perception does not rule out temporal processing. As discussed in the Introduction (Chapter 1), a third category of pitch model ("spectrotemporal") posits a role for both spectral and temporal processing in pitch discrimination. Several results from the psychophysical literature suggest a role for temporal processing pitch perception. Temporal processes must be at work for unresolved complexes, where spectral information is absent, as

evidenced by the influence of relative phase on f_0 DLs (e.g. Chapters 3 and 5; Houtsma and Smurzynski, 1990; Moore *et al.*, 2005). Temporal information may also play a role for pure-tone and resolved-harmonic pitch. One piece of evidence in support of this idea is the general worsening in performance at high absolute frequencies (e.g. Moore, 1973; Krumbholz *et al.*, 2000; Moore *et al.*, 2005) where neural phase locking to temporal fine structure is known to deteriorate. Some evidence exists in support of the idea that hearing-impaired listeners show a reduction in phase-locking to temporal fine structure. For example, Lacher-Fougère and Demany (2005) recently showed that cochlear hearing loss resulted in impaired detection of interaural phase-differences (IPD) in the carrier fine structure. The question remains whether a deficit in temporal processing contributes to the deterioration in pitch processing that accompanies SNHL. To address this issue a measure of temporal processing could be compared to the f_0 DL_{max} (associated with unresolved harmonics) and the f_0 DL_{min} (associated with resolved harmonics) across listeners with SNHL to determine whether a correlation exists.

IPD detection (Lacher-Fougère and Demany, 2005), one possible temporal processing measure, has the benefit of differentiating between fine-structure and envelope processing depending on the aspect of the signal to which the IPD is applied. Another possible measure of temporal processing ability is the resilience of frequency modulation (FM) detection performance at low modulation rates (f_m's) to the influence of added amplitude modulation (AM) (Moore and Sek, 1996; Moore and Skrodzka, 2002). Moore and Sek (1996) showed that at low f_m's (e.g. 2 Hz) and for low carrier frequencies (f_c's) below about 4 kHz were phase locking information is available, FM detection is relatively immune to the addition of AM to the signal, and is therefore thought to be regulated by temporal mechanisms. At higher f_m' (e.g. 20 Hz), FM detection is impaired by the added AM, suggesting that FM detection is spectrally mediated based on amplitude changes in the auditory filter outputs in response to FM (termed FM-induced AM). This may be because the temporal FM detection mechanism is "sluggish" and cannot follow the rapid changes in carrier frequency (Carlyon *et al.*, 2000). Thus, comparing the influence of added AM on FM detection at low versus high f_m rates can provide a measure of the extent to which HI subjects are able to use temporal cues for FM detection at the low f_m (Moore and Skrodzka, 2002)

7.2.2 The role of spectral pitch cues in complex environments

Pitch is thought to be an important attribute in hearing out simultaneous streams of information, such as those experienced in a multi-talker environment (e.g. Darwin and Hukin, 2000). HI listeners are known to experience difficulty in such complex listening situations. This thesis has demonstrated that HI listeners experience a difficulty in processing complex pitch associated with a loss of frequency selectivity. These effects of SNHL may be related. Recent work by Micheyl *et al.* (2005) has shown that the ability of NH listeners to make use of f_0 differences and onset/offset asynchronies to hear out the pitch of one harmonic complex in the presence of another depends to some extent on the spectral resolvability of the individual components in each complex. Using cochlear-implant simulations in NH listeners, Qin and Oxenham (2005) showed that in the absence of spectral cues, f_0 differences did not aid concurrent vowel segregation, despite the presence of temporal envelope pitch cues. Finally, Gilbert & Micheyl (2005) showed that for speech intelligibility in the presence of interfering speech babble, listeners depended on low frequencies (< 750 Hz), where harmonics are most likely to be resolved.

These results suggest that *spectral* pitch cues may be needed to hearing out complex pitch in a multiple-source environment. Thus, listeners with SNHL may experience difficulty in multi-talker situations due to the deterioration, demonstrated in this thesis, of spectral pitch cues, related to a loss of peripheral frequency selectivity. A tactic similar to the one taken in Chapters 4 and 5 could test this hypothesis by making use of the variation in peripheral frequency selectivity with stimulus level and SNHL. The idea would be to investigate whether there is a significant correlation between peripheral frequency selectivity and the ability to utilize f_0 differences for pitch discrimination and speech segregation in an environment consisting of multiple harmonic sources.

References

Abbas, P.J. (1981). "Auditory-nerve fiber responses to tones in a noise masker," *Hear. Res.* **5**, 69-80.

Alcántara, J.I. and Moore, B.C.J. (1995). "The identification of vowel-like harmonic complexes: Effects of component phase, level, and fundamental frequency," *J. Acoust. Soc. Am.* **97**, 3813-3824.

Arehart, K. H. and Burns, E. M. (1999). "A comparison of monotic and dichotic complex-tone pitch perception in listeners with hearing loss," *J. Acoust. Soc. Am.* **106**, 993-997.

Arehart, K.H. (1994). "Effects of harmonic content on complex-tone fundamental-frequency discrimination in hearing-impaired listeners," *J. Acoust. Soc. Am.* **95**, 3574-3585.

Bacon, S.P. and Gleitman, R.M. (1992). "Modulation detection in subjects with relatively flat hearing losses," *J. Speech Hear. Res.* **35**, 642-653.

Baker, R.J., Rosen, S. and Darling, A.M. (1998). "An efficient characterisation of human auditory filtering across level and frequency that is also physiologically reasonable," in *Psychophysical and Physiological Advances in Hearing*, edited by A. Rees A. R. Palmer, A. Q. Summerfield and R. Meddis (Whurr, London).

Bernstein, J.G. and Oxenham, A.J. (2003). "Pitch discrimination of diotic and dichotic tone complexes: Harmonic resolvability or harmonic number?," *J. Acoust. Soc. Am.* **113**, 3323-3334.

Bernstein, J.G.W. and Oxenham, A.J. (2005a). "An autocorrelation model with place dependence to account for the effect of harmonic number on fundamental frequency discimnination," *J. Acoust. Soc. Am.* **117**, 3816-3831.

Bernstein, J.G.W. and Oxenham, A.J. (2005b). "The relationship between harmonic resolvability and pitch discrimination: Effects of stimulus level," *J. Acoust. Soc. Am.*, (submitted).

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

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

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

Carlyon, R. P., Moore, B. C. and Micheyl, C. (2000). "The effect of modulation rate on the detection of frequency modulation and mistuning of complex tones," *J Acoust Soc Am* **108** (1), 304-15.

Carlyon, R.P. and Shackleton, T.M. (1994). "Comparing the fundamental frequencies of resolved and unresolved harmonics: Evidence for two pitch mechanisms?," *J. Acoust. Soc. Am.* **95**, 3541-3554.

Carlyon, R.P. and Datta, A.J. (1997). "Excitation produced by Schroeder-phase complexes: Evidence for fast-acting compression in the auditory system," *J. Acoust. Soc. Am.* **101**, 3636-3647.

Carlyon, R.P. and Shamma, S. (2003). "An account of monaural phase sensitivity," *J. Acoust. Soc. Am.* **114**, 333-348.

Carlyon, R.P., van Wieringen, A., Long, C.J., Deeks, J.M. and Wouters, J. (2002). "Temporal pitch mechanisms in acoustic and electirc hearing," *J. Acoust. Soc. Am.* **112**, 621-633.

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

Cedolin, L. and Delgutte, B. (2005a). "Spatio-temporal representation of the pitch of complex tones in the auditory nerve," *Assoc. Res. Otolaryngol. Abs.*, 1195.

Cedolin, L. and Delgutte, B. (2005b). "Pitch of complex tones: Rate-place and interspike interval representations in the auditory nerve," *J. Neurophysiol.* **94**, 347-362.

Cullen, J.K. and Long, G.R. (1986). "Rate discrimination of high-pass-filtered pulse trains," *J. Acoust. Soc. Am.* **79**, 114-119.

Darwin, C.J. and Hukin, R.W. (2000). "Effectiveness of spatial cues, prosody, and talker characteristics in selective attention," *J. Acoust. Soc. Am.* **107**, 970-977.

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

de Cheveigné, A. (2005). "Pitch perception models," in *Springer Handbook of Auditory Research: Pitch Perception*, edited by A.J. Oxenham C.J. Plack, A.N. Popper and R. Fay (Springer, Berlin).

de Cheveigné, A. and Pressnitzer, D. (2005). "The case of the missing delay lines: cross-channel phase interaction," *J. Acoust. Soc. Am.*, (in press).

Delgutte, B. (1982). "Some correlates of phonetic distinctions at the level of the auditory nerve," in *The Representation of Speech in the PeripheralAuditory System*, edited by Granstrom RCaB (Elsevier, Amsterdam), pp. 131-150.

Delgutte, B. (1984). "Speech coding in the auditory nerve: II. Processing schemes for vowel-like sounds," *J. Acoust. Soc. Am.* **75**, 879-886.

Delgutte, B. (1987). "Peripheral auditory processing of speech information: Implications from a physiological study of intensity discrimination," in *The Psychophysics of Speech Perception*, edited by M.E.H. Schouten (Nijhoff, Dordrecht, The Netherlands), pp. 333-353.

Duifhuis, H. (1970). "Audibility of high harmonics in a periodic pulse," *J. Acoust. Soc. Am.* **48**, 888-893.

Dye, R.H. and Hafter, E.R. (1980). "Just-noticeable differences of frequency for masked tones," *J. Acoust. Soc. Am.* **67**, 1746-1753.

Efron, B. and Tibishirani, R.J.. (1993). *An Introduction to the Bootstrap*. (Chapman & Hall, New York).

Evans, E.F. (2001). "Latest comparisons between physiological and behavioural frequency selectivity," in *Physiological and Psychophysical Bases of Auditory Function*, edited by A. J. M. Houtsma J. Breebaart, A. Kohlrausch, V. F. Prijs and R. Schoonhoven (Shaker, Maastricht), pp. 382-387.

Fine, P.A. and Moore, B.C.J. (1993). "Frequency analysis and musical ability," *Music Percept.* **11**, 39-53.

Flanagan, J.L. and Guttman, N. (1960). "On the pitch of periodic pulses," *J. Acoust. Soc. Am.* **32**, 1308-1319.

Florentine, M. and Buus, S. (1981). "An excitation-pattern model for intensity discrimination," *J. Acoust. Soc. Am.* **70**, 1646-1654.

Frohne-Büchner, C., Büchner, A., Brendel, M., Stöver, T. and Lenarz, T. (2005), "Investigation on improvement of the frequency resolution via current steering," Conference on implantable auditory prostheses, Pacific Grove, California.

Geisser, S. and Greeenhouse, S.W. (1958). "An extension on Box's results on the use of the *F* distribution in mulivariate analysis," *Ann. Math. Stat.* **29**, 885-891.

Geurts, L. and Wouters, J. (2004). "Better place-coding of the fundamental frequency in cochlear implants," *J. Acoust. Soc. Am.* **115**, 844-852.

Ghitza, O. (1986). "Auditory nerve representation as a front-end for speech recognition in a noisy environment," *Comput. Speech Lang.* **1**, 109-130.

Gilbert, G. and Micheyl, C. (2005). "Influence of competing multi-talker babble on frequency-importance functions for speech measured using a correlational approach," *Acta Acustica* **91**, 145-154.

Glasberg, B. R, Moore, B.C.J. and Stone, M.A. (1999). "Modelling changes in frequency selectivity with level," in *Psychophysics, Physiology and Models of Hearing*, edited by V. Hohmann T. Dau, and B. Kollmeier (World Scientific, Singapore).

Glasberg, B.R. and Moore, B.C.J. (1986). "Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments," *J. Acoust. Soc. Am.* **79**, 1020-1033.

Glasberg, B.R. and Moore, B.C.J. (1990). "Derivation of auditory filter shapes from notched-noise data," *Hear. Res.* **47**, 103-138.

Glasberg , B.R. and Moore, B.C.J. (2000). "Frequency selectivity as a function of level and frequency measured with uniformly exciting notched noise," *J. Acoust. Soc. Am.* **108**, 2318-2328.

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

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

Green, D.M. and Swets, J.A.. (1966). *Signal Detection Theory and Psychophysics*. (Krieger, New York).

Green, D.M. and Luce, R.D. (1974). "Counting and timing mechanisms in auditory discrimination and reaction time," in *Contemporary developments in mathematical psychology*, edited by R. C. Atkinson D. H. Krantz, R. D. Luce and P. Suppres (Freeman, San Francisco).

Greenwood, D.D. (1961). "Critical bandwidth and the frequency coordinates of the basilar membrane," *J. Acoust. Soc. Am.* **33**, 1344-1356.

Greenwood, D.D. (1972). "Masking by combination bands: Estimation of the levels of the combination bands $(n+1)f_l - nf_h$," *Journal of the Acoustical Society of America* **52**, 1144-1154.

Grimault, N., Micheyl, C., Carlyon, R.P. and Collet, L. (2002). "Evidence for two pitch encoding mechanisms using a selective auditory training paradigm," *Percept. Psychophys.* **64** (2), 189-197.

Grose, J.H., Hall, J.W. and Buss, E. (2002). "Virtual pitch integration for asynchronous harmonics," *J. Acoust. Soc. Am.* **112**, 2956-2961.

Hacker, M.J. and Ratcliff, R. (1979). "A revised table of d' for M-alternative forced choice," *Percept. Psychophys.* **26**, 168-170.

Hall, J.W. and Soderquist, D.R. (1975). "Encoding and pitch strength of complex tones," *J. Acoust. Soc. Am.* **58**, 1257-1261.

Hall, J.W., Buss, E. and Grose, J.H. (2003). "Modulation rate discrimination for unresolved components: Temporal cues related to fine structure and envelope," *J. Acoust. Soc. Am.* **113**, 986-993.

Hartmann, W.M.. (1998). *Signals, Sound, and Sensation.* (Springer, New York).

Hicks, M. L. and Bacon, S. P. (1999). "Psychophysical measures of auditory nonlinearities as a function of frequency in individuals with normal hearing," *J. Acoust. Soc. Am.* **105**, 326-338.

Hoekstra, A., (1979)."Frequency discrimination and frequency analysis in hearing," Ph.D. Thesis, Institute of Audiology, University Hospital, Groningen, Netherlands.

Hoekstra, A. and Ritsma, R.J. (1977). "Perceptive hearing loss and frequency selectivity," in *Psychophysics and Physiology of Hearing*, edited by E.F. Evans and J.P. Wilson (Academic, London).

Houtgast, T., Steeneken, H.J.M. and Plomp, R. (1980). "Predicting speech intelligibility in rooms from the modulation transfer function. I. General room acoustics," *Acustica* **46**, 60-72.

Houtsma, A.J.M and Goldstein, J.L. (1972). "The central origin of the pitch of pure tones: Evidence from musical interval recognition," *J. Acoust. Soc. Am.* **51**, 520-529.

Houtsma, A.J.M. and Smurzynski, J. (1990). "Pitch identification and discrimination for complex tones with many harmonics," *J. Acoust. Soc. Am.* **87**, 304-310.

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

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

Kaernbach, C. and Bering, C. (2001). "Exploring the temporal mechanisms involved in the pitch of unresolved complexes," *J. Acoust. Soc. Am.* **110**, 1039-1048.

Kiang, N.Y.-S., Watanabe, T., Thomas, E.C. and Clark , L.F.. (1965). *Discharge Patterns of Single Fibres in the Cat's Auditory Nerve.* (MIT Press, Cambridge, Mass.).

Koch, D.B., Downing, M. and Litvak, L. (2005), "Current steering and spectral channels in HiResolution bionic ear users: Cochlear-implant place/pitch perception," Conference on implantable auditory prostheses, Pacific Grove, California.

Kohlrausch, A., Fassel, R. and Dau, T. (2000). "The influence of carrier level and frequency on modulation and beat- detection thresholds for sinusoidal carriers," *J. Acoust. Soc. Am.* **108**, 723-734.

Kong, Y.Y., Stickney, G.S. and Zeng, F.G. (2005). "Speech and melody recognition in binaurally combined acoustic and electric hearing," *J. Acoust. Soc. Am.* **117**, 1351-1361.

Krumbholz, K., Patterson, R. D. and Pressnitzer, D. (2000). "The lower limit of pitch as determined by rate discrimination," *J. Acoust. Soc. Am.* **108**, 1170-1180.

Krumbholz, K., Patterson, R. D., Nobbe, A. and Fastl, H. (2003). "Microsecond temporal resolution in monaural hearing without spectral cues?," *J. Acoust. Soc. Am.* **113**, 2790-2800.

Lacher-Fougère, S. and Demany, L. (2005). "Consequences of cochlear damage for the detection of interaural phase differences," *J. Acoust. Soc. Am.* **118**, 2519-2526.

Lai, Y.C., Winslow, R.L. and Sachs, M.B. (1994). "A model of selective processing of auditory-nerve inputs by stellate cells of the antero-verntral cochlear nucleus," *J. Comput. Neurosci.* **1**, 167-194.

Levitt, H. (1971). "Transformed up-down methods in psychoacoustics," *J. Acoust. Soc. Am.* **49**, 467-477.

Liberman, M.C. (1978). "Auditory-nerve response from cats raised in a low-noise chamber," *J. Acoust. Soc. Am.* **63**, 442-455.

Licklider, J.C.R. (1951). "A duplex theory of pitch perception," *Experientia* **7**, 128-133.

Licklider, J.C.R. (1954). "'Periodicity' pitch and 'place' pitch," *J. Acoust. Soc. Am.* **26**, 945.

Licklider, J.C.R. (1959). "Three auditory theories," in *Psychology, a Study of Science*, edited by S. Koch (McGraw-Hill, New York), pp. 41-144.

Loeb, G.E., White, M.W. and Merzenich, M.M. (1983). "Spatial cross correlation: A proposed mechanism for acoustic pitch perception," *Biological Cybernetics* **47**, 149-163.

Lopez-Poveda, E.A. and Meddis, R. (2001). "A human nonlinear cochlear filterbank," *J. Acoust. Soc. Am.* **110**, 3107-3118.

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

Meddis, R. and Hewitt, M. (1991b). "Virtual pitch and phase sensitivity studied of a computer model of the auditory periphery. II: Phase sensitivity," *J. Acoust. Soc. Am.* **89**, 2882-2894.

Meddis, R. and Hewitt, M. (1992). "Modeling the identification of concurrent vowels with different fundamental frequencies," *J. Acoust. Soc. Am.* **91**, 233-245.

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

Micheyl, C., Bernstein, J.G.W. and Oxenham, A.J. (2005). "Detection and F0 discrimination of concurrent harmonic complex tones," *J. Acoust. Soc. Am.*, (submitted).

Moore, B. C. J. and Skrodzka, E. (2002). "Detection of frequency modulation by hearing-impaired listeners: Effects of carrier frequency, modulation rate, and added amplitude modulation," *J. Acoust. Soc. Am.* **111**, 327-335.

Moore, B.C.J. (1973). "Frequency difference limens for short-duration tones," *J. Acoust. Soc. Am.* **54**, 610-619.

Moore, B.C.J. (1977). "Effects of relative phase of the components on the pitch of three-component complex tones," in *Psychophysics and Physiology of Hearing*, edited by E.F. Evans and J.P. Wilson (Academic Press, London), pp. 349-358.

Moore, B.C.J.. (1982). *An Introduction to the Psychology of Hearing, 2nd ed.* (Academic, London).

Moore, B.C.J.. (1995). *Perceptual Consequences of Cochlear Damage.* (OUP, Oxford).

Moore, B.C.J.. (1997). *An Introduction to the Psychology of Hearing, 4th ed.* (Academic Press, London).

Moore, B.C.J.. (1998). *Cochlear Hearing Loss.* (Whurr, London).

Moore, B.C.J. and Glasberg, B.R. (1986). "The relationship between frequency selectivity and frequency discrimination for subjects with unilateral and bilateral cochlear impairments," in *Auditory frequency selectivity*, edited by B.C.J. Moore and R.D. Patterson (Plenum, New York).

Moore, B.C.J. and Glasberg, B.R. (1987). "Formulae describing frequency selectivity as a function of frequency and level and their use in calculating excitation patterns," *Hear. Res.* **28**, 209-225.

Moore, B.C.J. and Glasberg, B.R. (1988). "Pitch perception and phase sensitivity for subjects with unilateral and bilateral cochlear hearing impairments," in *Clinical audiology*, edited by A. Quaranta (Laterza, Bari, Italy).

Moore, B.C.J. and Glasberg, B.R. (1990). "Frequency selectivity in subjects with cochlear loss and its effects on pitch discrimination and phase sensitivity," in *Advances in audiology*, edited by F. Grandori, G. Cianfrone, and D.T. Kemp (Karger, Basel), pp. 187-200.

Moore, B.C.J. and Peters, R.W. (1992). "Pitch discrimination and phase sensitivity in young and elderly subjects and its relationship to frequency selectivity," *J. Acoust. Soc. Am.* **91**, 2881-2893.

Moore, B.C.J. and Ohgushi, K. (1993). "Audibility of partials in inharmonic complex tones," *J. Acoust. Soc. Am.* **93**, 452-461.

Moore, B.C.J. and Sek, A. (1995). "Effects of carrier frequency, modulation rate, and modulation waveform on the detection of modulation and the discrimination of modulation type (amplitude modulation versus frequency modulation)," *J. Acoust. Soc. Am.* **97**, 2468-2478.

Moore, B.C.J. and Sek, A. (1996). "Detection of frequency modulation at low modulation rates: Evidence for a mechanism based on phase locking," *J. Acoust. Soc. Am.* **100**, 2320-2331.

Moore, B.C.J. and Moore, G.A. (2003). "Discrimination of the fundamental frequency of complex tones with fixed and shifting spectral envelopes by normally hearing and hearing-impaired subjects," *Hear. Res.* **182**, 153-163.

Moore, B.C.J., Glasberg, B.R. and Shailer, M.J. (1984). "Frequency and intensity difference limens for harmonics within complex tones," *J. Acoust. Soc. Am.* **75**, 550-561.

Moore, B.C.J., Glasberg , B.R. and Flanagan, H. (2005). "Frequency discrimination of complex tones as a function of harmonic number and component phase: Teasing out the role of component resolvability and temporal fine structure," *Assoc. Res. Otolaryngol. Abs.*, 465.

Moore, B.C.J., Vickers, D.A., Plack, C.J. and Oxenham, A.J. (1999). "Inter-relationship between different psychoacoustic measures assumed to be related to the cochlear active mechanism," *J. Acoust. Soc. Am.* **106**, 2761-2778.

Moore, B.C.J., Huss, M., Vickers, D.A., Glasberg, B.R. and Alcantara, J.I. (2000). "A test for the diagnosis of dead regions in the cochlea," *Br. J. Audiol.* **34**, 205-224.

Nelson, D.A. (1991). "High-level psychophysical tuning curves: Forward masking in normal-hearing and hearing-impaired listeners," *J. Speech Hear. Res.* **34**, 1233-1249.

Nelson, D.A. (1994). "Level-dependent critical bandwidth for phase discrimination," *Journal of the Acoustical Society of America* **95**, 1514-1524.

Nelson, D.A. and Schroder, A.C. (1995). "Critical bandwidth for phase discrimination in hearing-impaired listeners," *J. Acoust. Soc. Am.* **98**, 1969-1976.

Ohm, G.S. (1843). "Über die Definition des Tones, nebst daran geknüpfter Theorie der Sirene und ähnlicher tonbildender Vorrichtungen [On the definition of the tone and the related theory of the siren and similar tone-producing devices]," *Ann. Phys. Chem.* **59**, 513-565.

Oxenham, A.J. and Bacon, S.P. (2003). "Cochlear compression: Perceptual measures and implications for normal and impaired hearing," *Ear Hearing* **24**, 352-366.

Oxenham, A.J., Bernstein, J.G.W. and Penagos, H. (2004). "Correct tonotopic representation is necessary for complex pitch perception," *Proc. Nat. Acad. Sci.* **101**, 1421-1425.

Palmer, A.R. and Russell, I.J. (1986). "Phase-locking in the cochlear nerve of the guinea-pig and its relation to the receptor potential of inner hair cells," *Hear. Res.* **24**, 1-15.

Palmer, A.R., Summerfield, Q. and Fantini, D.A. (1995). "Responses of auditory-nerve fibers to stimuli producing psychophysical enhancement," *J. Acoust. Soc. Am.* **97**, 1786-1799.

Patterson, R. D., Yost, W. A., Handel, S. and Datta, A. J. (2000). "The perceptual tone/noise ratio of merged iterated rippled noises," *J. Acoust. Soc. Am.* **107**, 1578-1588.

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

Patterson, R.D., Allerhand, M.H. and Giguère, C. (1995). "Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform," *J. Acoust. Soc. Am.* **98**, 1890-1894.

Patterson, R.D., Handel, S., Yost, W.A. and Datta, A.J. (1996). "The relative strength of the tone and noise components in iterated rippled noise," *J. Acoust. Soc. Am.* **107**, 1578-1588.

Patterson, R.D., Robinson, K., Holdsworth, J., McKeown, D., Zhang, C. and Allerhand, M. (1992). "Complex sounds and auditory images," in *Auditory Physiology and Perception*, edited by Y. Cazals, L. Demany, and K. Horner (Pergamon, Oxford).

Pick, G. (1980). "Level dependence of psychophysical frequency resolution and auditory filter shape," *J. Acoust. Soc. Am.* **68**, 1085-1095.

Plack, C.J. and Carlyon, R.P. (1995). "Differences in frequency modulation detection and fundamental frequency discrimination between complex tones consisting of resolved and unresolved harmonics," *J. Acoust. Soc. Am.* **98**, 1355-1364.

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

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

Plomp, R. and Mimpen, A.M. (1968). "The ear as a frequency analyzer II," *J. Acoust. Soc. Am.* **43**, 764-767.

Potthoff, R.F. (1966). "Statistical aspects of the problem of biases in psychological tests. (Institute of Statistics Mimeo Series No 479.)," Chapel HIll: University of North Carolina, Department of Statistics, Chapel Hill.

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

Pressnitzer, D., de Cheveigné, A. and Winter, I.M. (2002). "Perceptual pitch shift for sounds with similar waveform autocorrelation," *Acoust. Res. Lett. Online* **3** (1), 1-6.

Qin, M.K., (2005)."Role of F0 in speech reception in the presence of interference: simulating aspects of cochlear-implant processing," Ph.D. Thesis, MIT, Cambridge, MA.

Qin, M.K. and Oxenham, A.J. (2005). "Effects of envelope-vocoder processing on F0 discrimination and concurrent-vowel identification," *Ear Hear.*, (in press).

Rhode, W.S. (1971). "Observations of the vibration of the basilar membrane in squirrel monkeys using the Mössbauer technique," *J. Acoust. Soc. Am.* **49**, 1218-1231.

Rhode, W.S. and Cooper, N.P. (1996). "Nonlinear mechanics in the apical turn of the chinchilla cochlea in vivo," *Auditory Neurosci.* **3**, 101-121.

Rhode, W.S., Geisler, C.D. and Kennedy, D.T. (1978). "Auditory nerve fiber responses to wide-band noise and tone combinations," *J. Neurophysiol.* **41**, 692-704.

Ritsma, R.J. (1962). "Existence region of the tonal residue. I.," *J. Acoust. Soc. Am.* **34**, 1224-1229.

Ritsma, R.J. (1967). "Frequencies dominant in the perception of the pitch of complex sounds," *J. Acoust. Soc. Am.* **42**, 191-198.

Ritsma, R.J. and Hoekstra, A. (1974). "Frequency selectivity and the tonal residue," in *Facts and Models in Hearing*, edited by E. Zwicker and E. Terhardt (Springer, Berlin), pp. 156-163.

Robles, L., Ruggero, M.A. and Rich, N.C. (1986). "Basilar membrane mechanics at the base of the chinchilla cochlea I. Input-output functions, tuning curves, and phase responses," *J. Acoust. Soc. Am.* **80**, 1364-1374.

Rose, J.E., Brugge, J.F., Anderson, D.J. and Hind, J.E. (1968). "Patterns of activity in single auditory nerve fibres of the squirrel monkey," in *Hearing Mechanisms in Vertebrates*, edited by A.V.S. de Reuck and J. Knight (Churchill, London).

Rosen, S. and Stock, D. (1992). "Auditory filter bandwidths as a function of level at low frequencies (125 Hz-1 kHz)," *Journal of the Acoustical Society of America* **92**, 773-781.

Rosen, S. and Baker, R.J. (1994). "Characterising auditory filter nonlinearity," *Hear. Res.* **73**, 231-243.

Ruggero, M.A. (1992). "Responses to sound of the basilar membrane of the mammalian cochlea," *Curr. Opin. Neurobiol.* **2**, 449-456.

Ruggero, M.A. and Rich, N.C. (1991). "Furosemide alters organ of Corti mechanics: Evidence for feedback of outer hair cells upon the basilar membrane," *J. Neurosci.* **11**, 1057-1067.

Ruggero, M.A., Robles, L. and Rich, N.C. (1992). "Two-tone suppression in the basilar membrane of the cochlea: Mechanical basis of auditory-nerve rate suppression," *J. Neurophysiol.* **68**, 1087-1099.

Ruggero, M.A., Rich, N.C., Recio, A., Narayan, S.S. and Robles, L. (1997). "Basilar-membrane responses to tones at the base of the chinchilla cochlea," *J. Acoust. Soc. Am.* **101**, 2151-2163.

Schmidt, S. and Zwicker, E. (1991). "The effect of masker spectral asymmetry on overshoot in simultaneous masking," *J. Acoust. Soc. Am.* **89**, 1991.

Schouten, J.F. (1940). "The residue and the mechanism of hearing," *Proc. Kon. Akad. Wetenschap.* **43**, 991-999.

Schouten, J.F. (1970). "The residue revisited," in *Frequency Analysis and Periodicity Detection in Hearing*, edited by R. Plomp and G.F. Smoorenburg (Sijthoff, Lieden, The Netherlands), pp. 41-54.

Schroeder, M.R. (1970). "Synthesis of low peak-factor signals and binary sequences with low autocorrelation," *IEEE Trans. Inf. Theory* **16**, 85-89.

Scutt, M.J., Palmer, A.R. and Summerfield, A.Q. (1997), "Psychophysical and physiological responses to signals which are enhanced by temporal context," Assoc. Res. Otolaryngol. MidWinter Meeting.

Seebeck, A. (1841). "Beoubachtungen über einige Bedingungen der Entstehun von Tönen [Observations on some conditions for the creation of tones]," *Ann. Phys. Chem.* **53**, 417-436.

Sellick, P.M., Patuzzi, R. and Johnstone, B.M. (1982). "Measurement of basilar membrane motion in the guinea pig using the Mössbauer technique," *J. Acoust. Soc. Am.* **72**, 131-141.

Shackleton, T.M. and Carlyon, R.P. (1994). "The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination," *J. Acoust. Soc. Am.* **95**, 3529-3540.

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

Shamma, S.A. (1985). "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.

Shera, C.A., Guinan, J.J. and Oxenham, A.J. (2002). "Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements," *Proc. Nat. Acad. Sci.* **99**, 3318-3323.

Soderquist, D.R. (1970). "Frequency analysis and the critical band," *Psychonomic Science* **21**, 117-119.

Srulovicz, P. and Goldstein, J.L. (1983). "A central spectrum model: a synthesis of auditory-nerve timing and place cues in monaural communication of frequency spectrum," *J. Acoust. Soc. Am.* **73**, 1266-1276.

Steeneken, H.J.M. and Houtgast, T. (1980). "A physical method for measuring speech-transmission quality," *J. Acoust. Soc. Am.* **69**, 318-326.

Steinschneider, M., Reser, D.H., Fishman, Y.I., Schroeder, C.E. and Arezo, J.C. (1998). "Click train encoding in primary auditory cortex of the awake monkey: Evidence for two mechanisms subserving pitch perception," *J. Acoust. Soc. Am.* **104**, 2935-2955.

Stubbs, R.J. and Summerfield, A.Q. (1988). "Evaluation of two voice-separation algorithms using normally-hearing and hearing impaired listeners," *J. Acoust. Soc. Am.* **84**, 1236-1249.

Sumner, C.J., Lopez-Poveda, E.A., O'Mard, L.P. and Meddis, R. (2002). "A revised model of the inner-hair cell and auditory-nerve complex," *J. Acoust. Soc. Am.* **111**, 2178-2188.

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

Terhardt, E. (1979). "Calculating virtual pitch," *Hear. Res.* **1**, 155-182.

Turner, C.W., Gantz, B.J., Vidal, C., Behrens, A. and Henry, B.A. (2004). "Speech recognition in noise for cochlear implant listeners: Benefits of residual acoustic hearing," *J. Acoust. Soc. Am.* **115**, 1729-1735.

Tyler, R.S., Wood, E.J. and Fernandes, M.A. (1983). "Frequency resolution and discrmination of constant and dynamic tones in normal and hearing-impaired listeners," *J. Acoust. Soc. Am.* **74**, 1190-1199.

Van Tress, H.L.. (2001). *Detection, Estimation, and Modulation Theory, Part I.* (Wiley, New York).

Viemeister, N.F. and Bacon, S.P. (1982). "Forward masking by enhanced components in harmonic complexes," *J. Acoust. Soc. Am.* **71**, 1502-1507.

Weber, D.L. (1977). "Growth of masking and the auditory filter," *J. Acoust. Soc. Am.* **62**, 424-429.

Weiss, T.F. and Rose, C. (1988a). "Stages of degradation of timing information in the cochlea - a comparison of hair-cell and nerve fiber responses in the alligator lizard," *Hear. Res.* **33**, 167-174.

Weiss, T.F. and Rose, C. (1988b). "A comparison of synchronization filters in different auditory receptor organs," *Hear. Res.* **33**, 175-180.

Wiegrebe, L. (2001). "Searching for the time constant of neural pitch extraction," *J. Acoust. Soc. Am.* **109**, 1082-1091.

Wier, C.C., Jesteadt, W. and Green, D.M. (1977). "Frequency discrimination as a function of frequency and sensation level," *J. Acoust. Soc. Am.* **61**, 178-184.

Wightman, F.L. (1973). "The pattern-transformation model of pitch," *J. Acoust. Soc. Am.* **54**, 407-416.

Yates, G.K., Winter, I.M. and Robertson, D. (1990). "Basilar membrane nonlinearity determines auditory nerve rate-intensity functions and cochlear dynamic range," *Hear. Res.* **45**, 203-220.

Yost, W.A., Patterson, R.D. and Sheft, S. (1996). "A time domain description for the pitch strength of iterated rippled noise," *J. Acoust. Soc. Am.* **99**, 1066-1078.

Zurek, P.M. (1979). "Measurements of binaural echo suppression," *J. Acoust. Soc. Am.* **66**, 1750-1757.

Zwicker, E. (1952). "Die Grenzen der Hörbarkeit der Amplitudenmodulation und der Frequenzmodulation eines Tones [The limits of audibility of amplitude and frequency modulation of tones]," *Acustica* **2**, 125-133.

Zwicker , E. (1970). "Masking and psychological excitation as consequences of the ear's frequency analysis," in *Frequency Analysis and Periodicity Detection in Hearing*, edited by R. Plomp and G.F. Smoorenburg (Sijthoff, Leiden), pp. 376-394.

Acknowledgements

Many people were instrumental is helping me to successfully complete my Doctoral Dissertation. I gratefully thank and acknowledge:

Jill Bernstein, my wife and best friend, for being by my side though it all.

My parents, Mike and Marilyn Bernstein, for their loving support.

Andrew Oxenham, my mentor, who, with just the right balance of engaged and hands-off guidance, taught me to be an independent research scientist.

My labmates, Christophe Micheyl, Craig Lewiston, Xuedong Zhang, Andrea Simonson, Courtenay Wilson, Alexander Gutshalk, Michael 'Q' Qin and Peninah Rosengard, for interesting and lively discussions, moral support and for making the lab an enjoyable place to work. I especially thank Andrea Simonson for her help in the recruiting and screening of hearing-impaired research subjects.

My thesis committee, Louis Braida, Bertrand Delgutte, Gerald Kidd and Christophe Micheyl, for the time and energy they spent in helping me to steer my dissertation path, and for the helpful criticisms that they provided along the way.

The editors and reviewers of the Journal of the Acoustical Society of America, for their comments and criticisms that helped to improve the quality of the manuscripts that compose this dissertation: Neal Viemeister, John Grose, Bob Carlyon, Peter Cariani, Lutz Wiegrebe and four anonymous reviewers. Also thanks to Chris Plack, Ray Meddis, Peninah Rosengard and Ray Goldsworthy for informally reviewing portions of my thesis.

The 26 anonymous research subjects who provided the data in this thesis, for their time and effort.

Ray Meddis and Lowel O'Mard, for inviting and hosting me for a week at the University of Essex, UK, and for teaching me the ins and outs of their computational Auditory Modeling System, the focus of Chapter 4.

Marco Pelizzone and everyone at the Centre Romand d'Implants Cochleaires, for welcoming me and giving me a place to work during my time in Geneva.

The rest of the students and faculty of the Harvard-MIT Speech and Hearing Bioscience and Technology program, for creating an academic and research environment that is not only stimulating and rigorous, but also warm and supportive.

The National Institutes of Deafness and Other Communication Disorders, for their financial support (NIH Grants R01-DC-05216 and 5T32-DC-00038).