The relationship between frequency selectivity and pitch discrimination: Effects of stimulus level

Joshua G. W. Bernstein[a] and Andrew J. Oxenham[b]
Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139 and Harvard-MIT Division of Health Sciences and Technology, Speech and Hearing Bioscience and Technology Program, Cambridge, Massachusetts 02139

(Received 7 June 2005; revised 12 September 2006; accepted 25 September 2006)

Three experiments tested the hypothesis that fundamental frequency (f_0) discrimination depends on the resolvability of harmonics within a tone complex. 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 with levels of 10, 40, and 65 dB SPL per equivalent rectangular auditory filter bandwidth. With increasing 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, the frequencies of harmonics are more poorly encoded at higher levels, even when harmonics are well resolved. [DOI: 10.1121/1.2372451]

PACS number(s): 43.66.Hg, 43.66.Fe, 43.66.Ba [JHG] Pages: 3916–3928

I. 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, the 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 tenth (Houtsma and Smurzynski, 1990; Shackleton and Carlyon, 1994; Krumbholz *et al.*, 2000; Kaernbach and Bering, 2001; Bernstein and Oxenham, 2003; 2005). This dependence of f_0 DLs on harmonic number has generally been ascribed to harmonic 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" (Srulovicz and Goldstein, 1983; Shamma and Klein, 2000; Cedolin and Delgutte, 2005) models of pitch propose that the f_0 of a harmonic complex is identified by comparing the frequencies of individual harmonics to internally stored "harmonic templates." 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, 1991; 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, 1988), but are unable to account for the dependence of f_0 discrimination on harmonic number *per se* (Carlyon, 1998). A recent modification of the

[a]Author to whom correspondence should be addressed. Present address: Army Audiology and Speech Center, Walter Reed Army Medical Center, 6900 Georgia Avenue NW, Washington, DC 20307. Electronic mail: joshua.bernstein@amedd.army.mil
[b]Present address: Department of Psychology, University of Minnesota, 75 East River Road, Minneapolis, MN 55455. Electronic mail: oxenham@umn.edu

autocorrelation model (Bernstein and Oxenham, 2005) 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 characteristic frequency (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 "spectro-temporal" in nature, this type of model is referred to as "CF-dependent temporal" to differentiate it from the harmonic-template spectro-temporal models (e.g., Srulovicz and Goldstein, 1983; Shamma and Klein, 2000; Cedolin and Delgutte, 2005) mentioned above.

The question addressed in this study is whether the increase in f_0 DLs 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 to harmonic number, independent of resolvability (as predicted by a CF-dependent temporal model). Our study exploits 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 of 1 kHz and above. 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. A companion paper (Bernstein and Oxenham, 2006) investigates the relationship between frequency selectivity and pitch discrimination in listeners with sensorineural hearing loss.

II. EXPERIMENT 1: FUNDAMENTAL FREQUENCY DISCRIMINATION

A. 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 around the tenth harmonic. This transition is seen whether harmonic complexes are bandpass filtered into a fixed spectral region and the f_0 adjusted (Hoekstra, 1979; Shackleton and Carlyon, 1994; Bernstein and Oxenham, 2005) or the f_0 is held constant and the harmonic number adjusted (Houtsma and Smurzynski, 1990; Bernstein and Oxenham, 2003). The aim of this experiment was to determine whether the f_0 DL transition point varies with stimulus level. Fundamental frequency DLs were measured as a function of f_0 for bandpass-filtered harmonic complexes. By using this paradigm instead of holding f_0 constant and adjusting harmonic number (Houtsma and Smurzynski, 1990; Bernstein and Oxenham, 2003), 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.

B. Methods

All stimuli were presented in threshold equalizing noise (TEN; Moore *et al.*, 2000), for four reasons. First, the use of a background noise enables the presentation of stimuli at a constant sensation level (SL) while varying the absolute sound-pressure level (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 SPL will also be equal in terms of 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 f_0 of the stimulus.

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 listeners (S2), the highest-level noise was reduced by 3 dB to 62 dB SPL/ERB_N for this listener. To determine the SL reference, pure-tone detection thresholds were measured for each listener at each noise level for 1.5-, 2.5-, and 3.5-kHz tones, frequencies that, respectively, 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 listeners, 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=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 was selected independently from a uniform distribution for each stimulus. The large f_0 DLs produced by random-phase complexes for unresolved harmonics (Bernstein and Oxenham, 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 the 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 the components together. Stimuli were presented to the left ear in the TEN background at each of the three levels described above. Uncorrelated TEN with the same frequency content and level was also presented to the contralateral (right) ear to prevent detection of the stimulus in the contralateral ear via acoustic or electric crosstalk. Fundamental frequency discrimination was tested for eight average f_0's (75, 125, 150, 175, 200, 250, 325, and 400 Hz), at each of the three levels, with five repetitions per data point, for a total of 120 runs per listener. It is possible that with this bandpass-filter f_0 DL paradigm, listeners could have tracked individual frequencies rather than the f_0 in the case of resolved harmonics (Houtsma and Goldstein, 1972). However, Moore and Glasberg (1990) demonstrated that for harmonic complexes, listeners tend to base their pitch comparisons on the missing f_0 rather than on the frequencies of individual resolved components, even in cases where it would be advantageous to ignore the missing f_0. This suggests that listeners in the current study were also likely to base their responses on the f_0 and not on individual resolved frequencies.

The experimental method was similar to that described by Bernstein and Oxenham (2005). 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 ($f_{0,\text{base}}$) 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), which 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 and then a random level perturbation was added to each interval, chosen from a uniform distribution of ±2.5 dB. In addition, $f_{0,\text{base}}$ was roved from trial to trial within a run, chosen from a uniform distribution between ±5% of the average f_0. This was intended to encourage listeners 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 $f_{0,\text{base}}$, 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 an average frequency of 1500 Hz, presented at an average SL of 12.5 dB in each of 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 listener, using the same procedure as the f_0 DL measure, including level roving and frequency randomization.

Four normal-hearing listeners (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 listeners (S2 and S3) were professional musicians with more than 10 years of formal musical training, and two (S1 and S4) were amateur musicians with at least 3 years of musical training. Each listener completed a training period of at least 4 h, 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 listener via Sennheiser HD 580 headphones. Listeners 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.

C. Results

Figure 1 plots f_0 DLs and pure-tone FDLs for each of the individual listeners in the experiment (upper four panels) and the mean across the four listeners (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 (Hoekstra, 1979; Houtsma and Smurzynski, 1990; Shackleton and Carlyon, 1994; Bernstein and Oxenham, 2003; 2005). 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 the four individual listeners. 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$_{\text{min}}$) achieved at the highest f_0's were elevated at the high level, an effect that was apparent in all four listeners and the mean data.

The f_0 DL data were parametrized to quantify and sta-

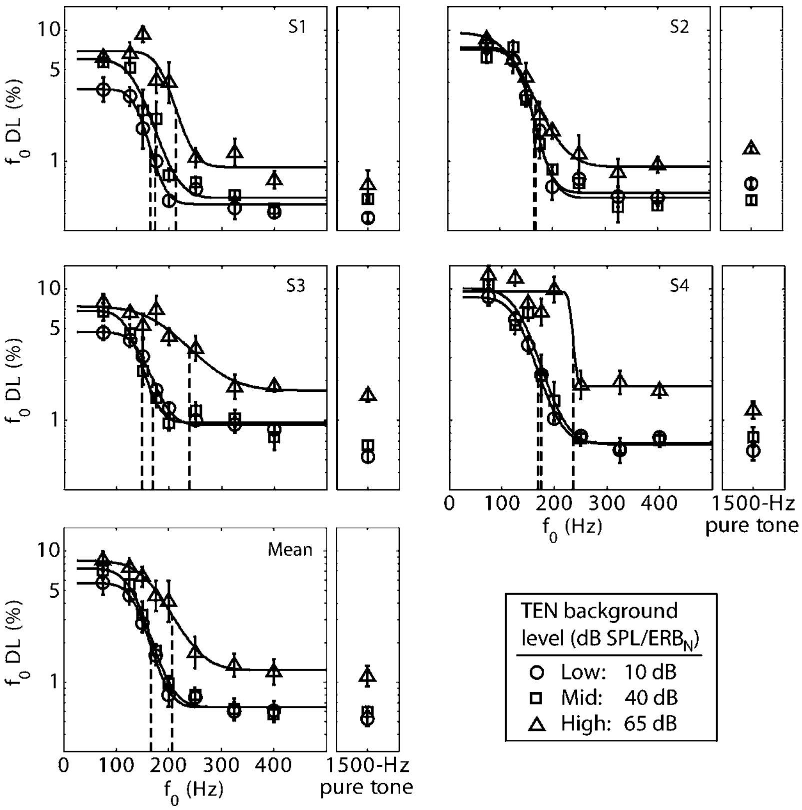

FIG. 1. Fundamental frequency DLs as a function of f_0 and FDLs for a 1500-Hz pure tone for the four individual listeners (upper four panels) and the mean data across the four listeners (lower panel). Stimuli were presented at an average 12.5 dB SL per component in the three TEN background levels specified 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 listeners in the case of the mean data. Solid lines represent the sigmoid functions [Eq. (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.

tistically 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. 1. The f_0 transition point ($f_{0,\mathrm{tr}}$, vertical dashed lines in Fig. 1) was defined as the estimate of the f_0 for which f_0 DLs were halfway (on a log scale) between the $f_0\,\mathrm{DL}_{\mathrm{min}}$ (achieved at high f_0's) and the maximum f_0 DL ($f_0\,\mathrm{DL}_{\mathrm{max}}$, achieved at low f_0's) for a given stimulus level.

The effects of level on $f_{0,\mathrm{tr}}$ and $f_0\mathrm{DL}_{\mathrm{min}}$ were analyzed statistically using bootstrap resampling methods (Efron and Tibishirani, 1993). For each individual listener 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 listener mean f_0 DLs at each f_0. Bootstrapped fits yielding $f_{0,\mathrm{tr}}$ estimates that fell outside of the 75 to 400-Hz range of f_0's tested were discarded. Two estimates were deemed to be significantly different if the 95% confidence interval (CI) of the difference between them, derived empirically by pairwise subtraction of the 1000 bootstrap estimates for each, did not overlap zero. The $f_{0,\mathrm{tr}}$ estimates are shown as circles in each panel of Fig. 2, with error bars representing the 95% CIs of each estimate (the other symbols represent frequency selectivity estimates from experiments 2 and 3, which will be described in Secs. III and IV, respectively). Significant differences between the high- and either the mid- (open symbols) or low-level (closed symbols) parameter estimates are identified by small symbols beneath the high-level estimates along the bottom of each panel of Fig. 2

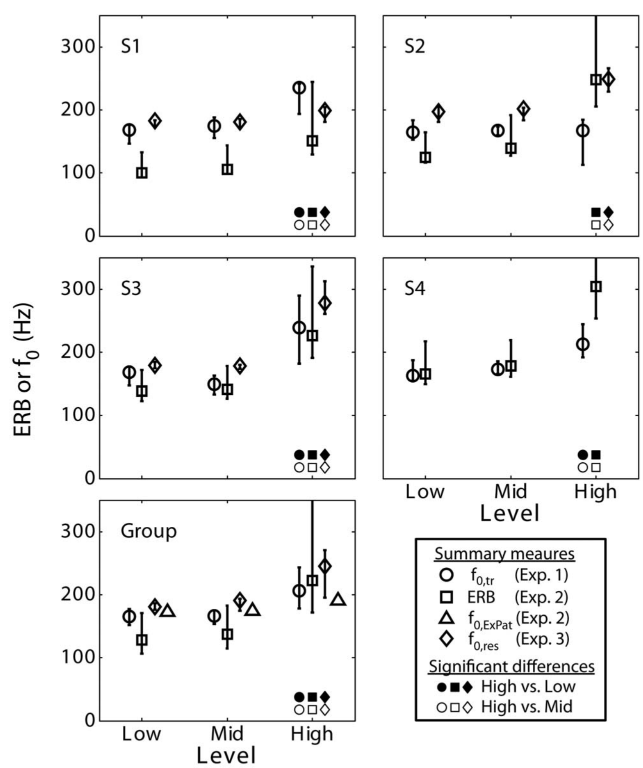

FIG. 2. Summary f_0 DL (experiment 1) and frequency selectivity measures (experiments 2 and 3) for each individual listener (upper four panels) and for the data pooled across listeners (lower panel). Small symbols at the bottom of the plot underneath the high-level data indicate that a given high-level 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 $f_{0,\mathrm{ExPat}}$ estimates based on the group data). Listener S4 did not participate in experiment 3.

(parameter estimates were never found to be significantly different between the mid and low levels). For example, for listener S1, the $f_{0,\mathrm{tr}}$ was significantly different between the low and high conditions (small filled circle) and between the mid and high conditions (small open circle).

The high-level $f_{0,\mathrm{tr}}$ was significantly larger ($p < 0.05$) than both the low- and mid-level $f_{0,\mathrm{tr}}$'s both for the group analysis and for three out of four listeners (S1, S3, and S4). Only listener S2 showed no significant differences in $f_{0,\mathrm{tr}}$ across stimulus level. The log-transformed f_0 $\mathrm{DL}_{\mathrm{min}}$ (Fig. 1) was also significantly larger at the high than at the low and mid levels in the group data and in three (S1, S2, and S4) out of four individual listeners (bootstrap CI estimates not shown). Although listener S3 showed no significant differences in f_0 $\mathrm{DL}_{\mathrm{min}}$ across level, this may simply reflect the poor sigmoid fit for a large number of bootstrap resamplings of the high-level data for this subject, as an effect of level is visually apparent in the f_0 DL data for this subject. The observation that f_0 DL level effects only occurred between the mid and high conditions (where significant differences were observed) and not between the low and mid conditions (where no significant differences were observed) was confirmed by an additional bootstrap analysis comparing the high-mid versus mid-low differences in $f_{0,\mathrm{tr}}$ and f_0 $\mathrm{DL}_{\mathrm{min}}$. The high-mid difference was significantly greater ($p < 0.05$) than the mid-low difference for both group analyses ($f_{0,\mathrm{tr}}$ and f_0 $\mathrm{DL}_{\mathrm{min}}$) and for two (S3 and S4) and three (S1, S2, and S4) out of four individual subjects for the $f_{0,\mathrm{tr}}$ and f_0 $\mathrm{DL}_{\mathrm{min}}$ estimates, respectively.

An effect of level similar to that observed in the f_0 $\mathrm{DL}_{\mathrm{min}}$ was also observed in the 1500-Hz pure-tone FDLs (Fig. 1). Fisher LSD *t*-tests on the mean data showed that pure-tone FDLs were larger at the high level than at both the mid and low levels.

D. 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 2 and 3 by comparing the results with estimates of frequency selectivity. S2, the one listener who did not show this effect, was also the listener tested at a slightly lower level in the high-level condition than the other listeners (TEN level 62 dB instead of 65-dB SPL/ERB_N), and may have shown results more similar to the other listeners if tested at the slightly higher level.

The increase from the mid to the high level led not only to an increase in $f_{0,\mathrm{tr}}$, but also to an increase in the f_0 $\mathrm{DL}_{\mathrm{min}}$. One possible interpretation for the increased f_0 $\mathrm{DL}_{\mathrm{min}}$ 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 aspects of 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 and presented at an average level 12.5 dB above threshold, is resolved by definition. These two observations suggest another explanation, namely that the increase in f_0 $\mathrm{DL}_{\mathrm{min}}$ and FDL at the high level reflects a deterioration in the frequency coding of the individual resolved components. A similar effect of level on FDLs for pure tones presented in background noise was observed by Dye and Hafter (1980), but at higher frequencies. Possible implications of this finding are addressed in Sec. V. Regardless of the mechanism by which individual pure-tone frequencies are encoded, it may be that a similar increase in pure-tone FDLs with level is not observed in the absence of background noise (e.g., Wier *et al.*, 1977) because information from off-frequency channels is available, which would not be affected by spectral or temporal distortions that may occur at high levels.

III. EXPERIMENT 2: AUDITORY FILTER SHAPES

A. Rationale

The central hypothesis of this study was that a reduction in frequency selectivity 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 frequency selectivity becomes worse with increasing level, which we expect given previous results (e.g., Weber, 1977; Pick, 1980; Moore and Glasberg, 1987; Rosen and Stock, 1992; Hicks and Bacon, 1999). Measurements of auditory filters in this experiment allowed a comparison between the variation in frequency selectivity with stimulus level and the f_0 DLs measured in the same listeners in experiment 1.

A version of the notched-noise method (Patterson, 1976), described by Rosen and Baker (1994), was used to measure the level of a 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 1 to ensure that the auditory filter shapes estimated in this experiment were as similar as possible to those presumably used by listeners 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.

B. Methods

The notched-noise level that just masked a pure-tone signal 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 1, 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 listener) *re* one of the TEN levels that was used in experiment 1. This signal level was at the lower end of the 10–15-dB SL per component level range that was used in experiment 1. 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 $0.13f_{\mathrm{sig}}$, compared to a range of 0.4 to $0.8f_{\mathrm{sig}}$ 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 ($\pm 0.2f_{\mathrm{sig}}$ as compared to $\pm 0.4f_{\mathrm{sig}}$). 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 needed to mask the signal at wider notch widths. Three symmetrical notch conditions were tested, with equal Δf_l and Δf_u values of 0 (no notch), 0.1, and $0.2f_{\mathrm{sig}}$. To allow for the possibility of asymmetrical filters, there were also two asymmetric conditions, one with $\Delta f_l = 0.1f_{\mathrm{sig}}$ and $\Delta f_u = 0.2f_{\mathrm{sig}}$, and the other with $\Delta f_l = 0.2f_{\mathrm{sig}}$ and $\Delta f_u = 0.1f_{\mathrm{sig}}$. A further 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. To prevent the use of the contralateral ear for signal detection based on acoustic or electric crosstalk, an uncorrelated contralateral noise was presented, centered on f_{sig} with a bandwidth three times the ERB_N (Glasberg and Moore, 1990) and a spectrum level 40 dB below the signal level.

The same listeners took part in this experiment as had taken part in experiment 1. All listeners underwent a short training period where one masked threshold for each of the 15 (five notch widths at three levels) conditions was estimated. Listeners then completed four measurements for each data point, for a total of 60 runs.

C. 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 × 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 listener were fitted 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 listener'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 (12 free parameters).

The assumed level invariance of the filter-tip shape 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, only the gain of the filter tip relative to the flat filter tail is allowed to vary with level, as modeled by the dynamic range limitation, r. The use of a level-invariant filter-tip shape models the physiologically observed two-component 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 level-dependent active mechanism thought to be governed by the outer hair cells (Ruggero and Rich, 1991).

The fitting procedure took into account the Sennheiser HD580 transfer function, the middle-ear transfer function, the possibility of off-frequency listening and variations in filter bandwidth in proportion with CF, as described by Glasberg and Moore (1990). Although the filter tip was assumed to be symmetric, 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 listeners) of each of the best-fitting model parameters are listed in Table I.[2] Figure 3 shows the fitted filter shapes for each stimulus level based on the mean filter parameters from

TABLE I. Means and standard deviations across listeners of the filter-model parameters that best fit the notched-noise data of experiment 2.

Parameter	Level	Mean value	Standard deviation
p	All	47.18	9.99
r	Low	−35.30	2.85
	Mid	−22.23	2.63
	High	−11.82	1.97
k	All	−1.25	0.59

Table I. 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 1, where f_0 DLs remained unchanged as the stimulus level increased from the low to the mid level, and then increased as the stimulus level increased from the mid to the high level.

D. Quantifying frequency selectivity

To directly test the hypothesis that small f_0 DLs are associated with resolved harmonics, two summary measures of frequency selectivity were derived and compared to the estimates of the $f_{0,\mathrm{tr}}$ (see Sec. II C). The first measure was the equivalent rectangular bandwidth (ERB) of the fitted filter shapes (Sec. III D 1). The second measure was an estimate of harmonic resolvability, based on peak-to-valley ratios (PVRs) in excitation patterns derived from the fitted filter shape functions and the spectra of the harmonic stimuli (Sec. III D 2).

1. Equivalent rectangular bandwidth (ERB)

Filter ERBs, calculated by integrating the fitted filter shape across frequency (Hartmann, 1998), are plotted as squares in Fig. 2 for individual listeners (upper four panels) and for filters derived from the data pooled across listeners (lower panel). Error bars indicate the 95% CIs derived from 1000 bootstrap estimates of each ERB. As for the $f_{0,\mathrm{tr}}$ and f_0 DL$_{\mathrm{min}}$, 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 at the high than both the mid

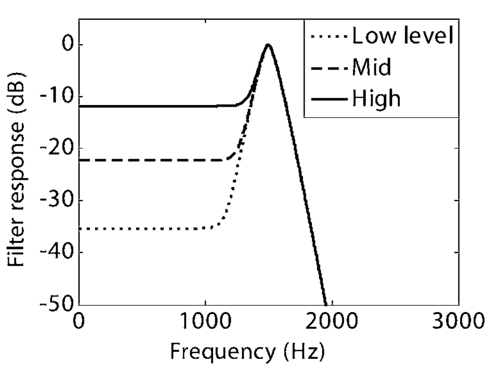

FIG. 3. Auditory filters at the three stimulus levels, based on the mean of each of four best-fitting filter parameters (p and three values of r) across the four listeners.

(small filled squares) and the low levels (small open squares) for the group analysis and for all four listeners. Providing further support for this observation, the high-mid ERB difference was significantly larger than the mid-low difference in three out of four subjects (S2, S3, and S4), although this comparison failed to reach significance ($p=0.07$) in the group analysis. (The intersubject variability in ERBs did not match that observed in the $f_{0,\mathrm{tr}}$, where only listener S2 did not show a high-mid difference that was significantly greater than the mid-low difference.) Overall, the similar pattern of results for the effect of level on f_0 DLs and ERBs supports the idea that f_0 discrimination performance is related to frequency selectivity and that the poorer f_0 discrimination performance observed at the high level is related to the reduction in frequency selectivity associated with these stimuli.

The ERBs (Fig. 2, squares) varied with level in a different manner than the auditory filter shapes (Fig. 3). Whereas auditory filter shapes changed continuously with level, 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.

2. Excitation pattern model

The ERB measure provided a general statistic describing the frequency selectivity of the auditory system for comparison with the f_0 discrimination data. To more directly test the hypothesis 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 listeners (Table I). 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 listeners. A different excitation pattern was produced for each f_0 and stimulus level tested in experiment 1, with the appropriate filterbank used at each stimulus level, based on the filters derived in experiment 2.

Sample excitation patterns for the mid-level conditions are shown in Fig. 4 for 75-, 200-, and 400-Hz stimulus f_0's. For the lowest f_0 of 75 Hz, there are no discernible 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

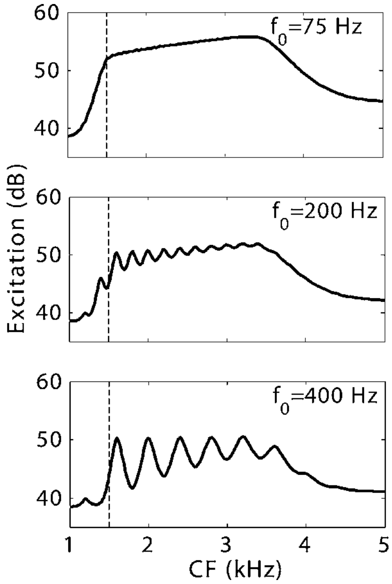

FIG. 4. Sample excitation patterns for three f_0's presented at the mid level.

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$\geqslant$1.5 kHz (vertical dashed lines in Fig. 4), and the valley at a higher CF immediately adjacent to the peak.

Figure 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 $f_{0,\mathrm{tr}}$ estimates. To directly compare the PVRs to the f_0 DL data, a threshold PVR ($\mathrm{PVR}_{\mathrm{th}}$), defined as the minimum PVR that yielded resolved harmonics, was varied as a single free parameter to fit the PVR estimates to the $f_{0,\mathrm{tr}}$ estimates derived from the pooled f_0 DL data of experiment 1 (circles in Fig. 2, lower panel).[3] The $\mathrm{PVR}_{\mathrm{th}}$ (horizontal dashed line in Fig. 5) was adjusted to minimize the least-squares difference between the f_0 needed to achieve the $\mathrm{PVR}_{\mathrm{th}}$ (termed $f_{0,\mathrm{ExPat}}$, vertical dashed lines in Fig. 5) and the $f_{0,\mathrm{tr}}$ estimates derived from the f_0 DL data. The PVRs for f_0's between those tested in experiment 1 were linearly interpolated as shown in Fig. 5. The fitted $\mathrm{PVR}_{\mathrm{th}}$ represents the estimate of the PVR needed in the excitation pattern to yield f_0 DLs halfway (on a log scale) between $f_0\mathrm{DL}_{\mathrm{max}}$ and $f_0\mathrm{DL}_{\mathrm{min}}$.

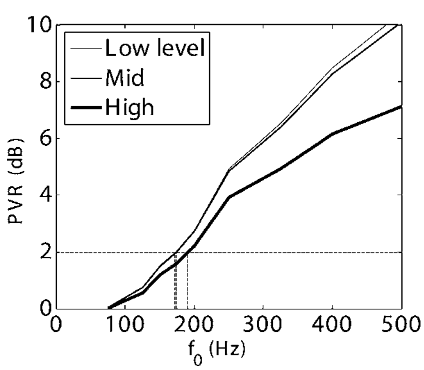

FIG. 5. Peak-to-valley ratios (PVRs) as a function of stimulus f_0 for the first excitation pattern peak occurring at a CF of 1500 Hz or higher. The horizontal dashed line indicates the $\mathrm{PVR}_{\mathrm{th}}$=1.98 dB that minimized the mean-squared difference between the corresponding $f_{0,\mathrm{res}}$ (vertical dashed lines) and the $f_{0,\mathrm{tr}}$ estimates derived from the f_0 DL data (Fig. 1).

The estimated $f_{0,\mathrm{ExPat}}$ at which the PVR was equal to the best-fitting $\mathrm{PVR}_{\mathrm{th}}$ of 1.98 dB are plotted as triangles in the lower panel of Fig. 2. The pattern of results is 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. 5. As with the ERB, 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 $f_{0,\mathrm{ExPat}}$ (approximately a 10% change) relative to that observed in the $f_{0,\mathrm{tr}}$ (approximately a 25% change from the mid to the high level). It may be that f_0 discrimination is based on a complex analysis of several resolved harmonics, and not just the first resolved harmonic as modeled here. Supporting this idea, there was a smaller difference between the level effects on the $f_{0,\mathrm{ExPat}}$ and $f_{0,\mathrm{tr}}$ when the PVR for the $f_{0,\mathrm{ExPat}}$ estimate was calculated for the second or third rather than the first peak in the excitation pattern (simulation results not shown).

3. Allowing the auditory filter tip shape 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 with low-frequency dynamic range limitation; (2) asymmetrical tip and low-frequency dynamic range limitation; and (3) asymmetrical tip without dynamic 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 Sec. III C.

IV. EXPERIMENT 3: HEARING OUT HARMONICS

A. Rationale

This experiment measured the ability of listeners to hear out the frequencies of individual harmonics. A method similar to that of Bernstein and Oxenham (2003) 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 approach was successful in the earlier study (Bernstein and Oxenham, 2003), 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.

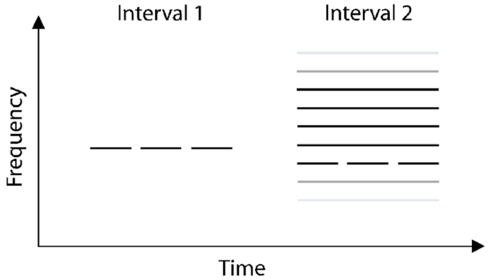

FIG. 6. Schematic of the stimulus paradigm used in experiment 3. Listeners compared the frequencies of the gated comparison tone presented in isolation (interval 1) with the frequency of a gated harmonic component of a bandpass filtered complex (interval 2). Shading represents the amplitude of each frequency component; components falling within the filter skirt are shown in lighter gray.

B. Methods

A schematic of the stimulus paradigm is shown in Fig. 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 1, 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 Bernstein and Oxenham, 2003, footnote 2, for a discussion). Both intervals were presented to the left ear in the same wideband TEN background as experiment 1, 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. Again, uncorrelated TEN was presented to the contralateral ear to prevent the detection of the signal via acoustic or electric crosstalk. Each component (before filtering, where applicable) was presented at 12.5 dB SL (adjusted for each listener). Level randomization was not used in this experiment, because overall loudness variations would not have provided a usable cue. The frequency of the comparison tone (f_{comp}) was 3.5% higher or lower (each with probability 0.5) than the frequency of the target tone (f_{targ}). The listener was required to discriminate whether the target was higher or lower in frequency than the comparison tone. 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 was 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 listeners 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 nondummy 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 $f_{\mathrm{comp}} > 1750$ Hz range where in the nondummy 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 ($f_0 = 150-200$ Hz), and mostly resolved ($f_0 > 300$ Hz) harmonics. The dummy comparison 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 nondummy trials, listeners responded "lower" or "higher" with a probability of roughly 0.5 for the entire range of f_{comp}'s presented.

Each run consisted of 72 trials for one stimulus level condition. There were 48 nondummy trials (six trials for each of the eight f_0's tested in experiment 1), plus 24 dummy trials (two trials each for 12 combinations of f_0 and target frequency). There were 17 runs for each of the three stimulus levels, for a total of 102 nondummy trials for each f_0 and stimulus level. The stimulus level for each run was randomly selected, without replacement, until three runs were completed, and then the process was repeated.

Three of the four listeners from experiments 1 and 2 participated in this experiment (listener S4 did not participate). Each was given at least 1 h of additional training.

C. Results and discussion

The percentage correct as function of f_0 and level are plotted for each of the three individual listeners and for the mean data in Fig. 7. For each listener 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 1, the data were fitted to a psychometric function fixed to 50% and 100% correct at the extremes (solid curves in Fig. 7), and the f_0 required to reach 75% correct was defined as the limit of harmonic resolvability, termed the $f_{0,\mathrm{res}}$ (vertical dashed lines in Fig. 7).

The effect of level on the $f_{0,\mathrm{res}}$ (diamonds in each panel of Fig. 2) was similar to the level effects observed in the previous experiments. As for the $f_{0,\mathrm{tr}}$ (circles) and ERB

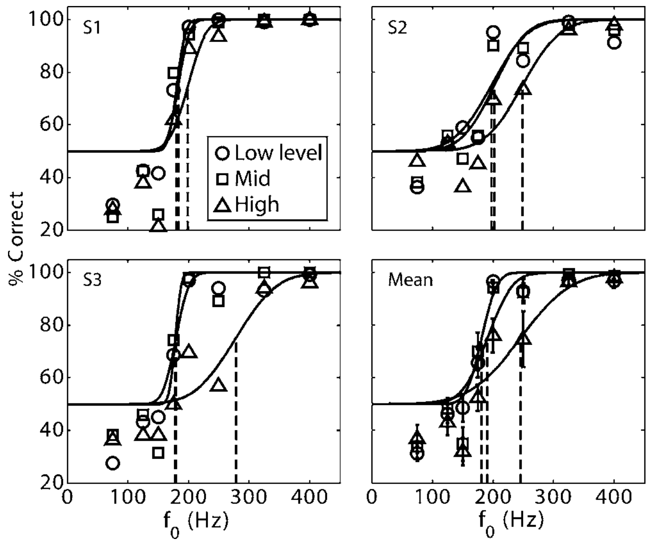

FIG. 7. Results of experiment 3 showing the percent correct in hearing out the frequency of an individual harmonic in the 1600–1750-Hz range as a function of stimulus f_0 for three stimulus levels for each individual listener (two upper panels and lower left panel) and for the mean data across the three listeners 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 ±1 standard error across the three listeners.

(squares), the $f_{0,res}$ was similar at the low and mid levels, and then increased at the high level. This observation was supported by bootstrap resampling performed for each individual listener by randomly resampling (with replacement) 102 responses (correct/incorrect), and for the group data by resampling (with replacement) four individual-subject percent correct scores at each f_0 for a given level (95% CIs are shown as error bars in Fig. 2). The estimated $f_{0,res}$ values were significantly larger at the high than both the mid and low levels for the group data and all three individual listeners (open diamonds at the bottom of each panel in Fig. 2). Furthermore, the high-mid $f_{0,res}$ difference was significantly greater than the mid-low difference for all three subjects and the group analysis. 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 the Bernstein and Oxenham (2003) study. One possibility is that listeners were comparing the pure-tone frequency to that of a harmonic adjacent to the gated harmonic that sometimes became unmasked during the "off" intervals. However, the fact that listeners 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.

V. GENERAL DISCUSSION

A. The relationship between f_0 discrimination and frequency selectivity

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. Predictions based on the excitation pattern model (experiment 2) and the more direct hearing-out-harmonics paradigm (experiment 3) both led to the conclusion that harmonic resolvability depends on stimulus level in a similar way to the f_0 DL transition point in experiment 1. There was little or no change from the low to the mid stimulus level, but a similar increase 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. The fact 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.

Some intersubject variability was observed in the experimental results, in that not all listeners showed an effect of level or a significantly greater high-mid versus mid-low difference for each f_0 discrimination and frequency selectivity measure. Moreover, on an individual listener basis the lack of a significant difference in one measure did not always correspond to the lack of significant difference in another. Nevertheless, each of these measures behaved similarly with respect to level when the data were averaged across listeners.

The results shown here may appear to be in conflict with the results of Krumbholz *et al.* (2000; see also Pressnitzer *et al.*, 2001). They concentrated on the lower limit of f_0 discrimination, and found that it corresponded more to a constant f_0 between 32 and 64 Hz, than to a constant harmonic number. Our conclusions can be reconciled with theirs if one assumes that pitch perception is constrained by two separate limits, one determined by the lower absolute limit of periodicity pitch (which appears to be between 30 and 60 Hz), and one determined by the spectral content, which is dependent on harmonic number, at least for f_0's between about 100 and 400 Hz.

B. Frequency selectivity and the coding of pure tones

In addition to the $f_{0,tr}$ effect, this study also found that both the pure-tone FDL and the f_0 DL_{min} (i.e. the f_0 DL associated with resolved harmonics) behaved in a similar manner to the various estimates of frequency selectivity, remaining constant from the low to the mid levels then increasing at the high stimulus level. It could be that level affected the two aspects of f_0 discrimination (i.e., increased f_0 DL_{min} and increased $f_{0,tr}$) in a similar manner simply because both rely on the frequency selectivity of the auditory system. This would be predicted by place-based models of pitch and frequency discrimination. On the other hand, the frequency coding of pure tones may rely on different mechanisms from those that underlie harmonic resolvability, but which also deteriorate at high levels. In this case, the question arises as to whether the increase in the transition point, $f_{0,\text{tr}}$, really reflects harmonic resolvability, or whether it simply reflects poorer coding of individual harmonics (by a mechanism other than frequency selectivity, such as poorer temporal coding), making them more susceptible to interference by neighboring harmonics. We tested this possibility using a simple signal-detection model, with two independent additive noises representing the encoding accuracy of individual partials and the influence of neighboring partials, respectively. Using this simple model, we found that it was not possible to predict the observed shift in $f_{0,\text{tr}}$ by simply increasing the noise associated with the coding of individual harmonics. Instead, an increase in both types of noise was required to predict the changes in both $f_{0,\text{tr}}$ and f_0 DL_{min}. Thus, while we cannot completely rule out influences other than frequency selectivity and harmonic resolvability on $f_{0,\text{tr}}$, it appears reasonable to assume that a change in frequency selectivity is responsible for the increase in $f_{0,\text{tr}}$ at the high level.

C. 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.

1. Temporal models

"Temporal" models of pitch perception estimate f_0 using temporal information, often derived from auditory nerve firing patterns. Bernstein and Oxenham (2005) 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" limit the range of f_0's to which the model responds, relative to a given channel's CF, thereby forcing a dependence on harmonic number. However, because the dependence on harmonic number is not a consequence of harmonic resolvability or frequency selectivity, this model would be unlikely to predict a detrimental effect of reduced frequency selectivity on the f_0 DL transition point.

In a novel implementation of the autocorrelation model by de Cheveigné and Pressnitzer (2006), the range of lags achieved for a given CF is limited by the duration of the impulse response, and therefore the bandwidth, of the filter. For this model, the ability to efficiently code f_0 would be likely to depend on filter bandwidth, and hence on level, in a way similar to that found in our data. Furthermore, this model is also likely to resolve the apparent discrepancy between the current results that show a relationship between accurate f_0 and harmonic resolvability and those of Bernstein and Oxenham (2003) where increasing peripheral resolvability by presenting even and odd harmonics to opposite ears did not improve f_0 discrimination performance. The de Cheveigné and Pressnitzer (2006) model is in principle consistent with both sets of results because the model's dependence on harmonic number derives from the frequency selectivity of the auditory periphery, which would not be affected by dichotic presentation, but would be affected by changes in stimulus level.

It is not clear how temporal models, including that of de Cheveigné and Pressnitzer (2006), could account for the increase in f_0 DL_{min} and pure-tone FDLs observed at the high stimulus level, as phase locking in the auditory nerve does not generally deteriorate at high levels (Johnson, 1980). Furthermore, physiological evidence shows that when tones are presented in a background noise, the degree of phase locking is mainly dependent on the SNR (Rhode *et al.*, 1978; Abbas, 1981), rather than absolute level. In our experiment, stimuli were presented at equal SL across level, which is likely to have yielded a roughly constant SNR at the filter output. One possibility, suggested by a reviewer, is that the accuracy of phase locking to a pure-tone frequency could be disrupted at high stimulus levels by a relative increase in the ANF response to low-frequency background-noise energy resulting from the reduced filter tip-to-tail ratio.

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 (e.g., Goldstein, 1973; Wightman, 1973; Terhardt, 1974; 1979). 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 the f_0 DL_{min} and the pure-tone FDL with level may also be consistent with spectral pure-tone frequency encoding. Dye and Hafter (1980) observed a similar effect of level on FDLs, but only for a relatively high-frequency (4 kHz) tone. For a 500-Hz or 1-kHz tone, FDLs decreased. 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 similar effects of level observed in the Dye and Hafter and the current studies, together with the parallel dependence of frequency selectivity on level (experiment 2), suggest a possible 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. A spectral model might account for the effect of level on the f_0 DL_{min} and the pure-tone FDL in terms of a broadening of the excitation pattern (e.g., Zwicker, 1970) or in terms of rate saturation of the relatively large population of high spontaneous-rate auditory-nerve fibers (Liberman, 1978) at high levels, leaving only a relatively small population of medium or low spontaneous-rate fibers for rate-based frequency encoding. However, a purely spectral model is unlikely to account for the observed deterioration in pure-tone frequency encoding (e.g., Moore and Sek, 1995) and resolved complex-tone f_0 discrimination performance (Moore *et al.*, 2006) at high absolute frequencies, thought to reflect the roll-off of phase locking in the auditory nerve (Weiss and Rose, 1988).

3. Harmonic-template spectro-temporal models

This class of pitch model uses a combination of place and temporal information to extract the individual frequencies of the harmonic components and calculate the f_0. Like purely spectral harmonic template models, these models also require the resolution of individual harmonic frequencies, and would therefore be likely to correctly predict an increase in $f_{0,\mathrm{tr}}$ 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, 2005). These models might also explain the observed effect of level on f_0 $\mathrm{DL}_{\mathrm{min}}$, 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).

Like the temporal models described above, spectro-temporal models that involve the extraction of individual resolved frequencies from phase-locking information in the auditory nerve (e.g., Srulovicz and Goldstein, 1983) do not depend on cochlear frequency selectivity to identify the frequencies of individual components. It is therefore not clear whether they could account for the observed increased in f_0 $\mathrm{DL}_{\mathrm{min}}$ with level. While these models might account for the increased $f_{0,\mathrm{tr}}$ in terms of a disruption of the temporal coding of individual frequency components due to the interaction of unresolved harmonics, Delgutte (1984) argued that the identification of individual frequency components based on a Fourier analysis of auditory-nerve responses would be robust to peripheral filtering.

VI. SUMMARY AND CONCLUSIONS

At high stimulus levels, 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 shift in the f_0 DL transition point as a function of level closely matches estimates of harmonic resolvability based on measures of auditory filter shapes and hearing out harmonics, consistent with the idea that resolved harmonics are necessary for accurate f_0 discrimination performance. Second, the minimum f_0 DL (at the highest f_0 tested) increases with increasing stimulus level, as does pure-tone FDLs at a comparable frequency, indicating that even resolved harmonic frequencies may be more poorly encoded at high stimulus levels. Overall, the results provide evidence in favor of spectral, spectro-temporal, or CF-dependent temporal models of pitch perception that depend on peripheral frequency selectivity to encode f_0 information, but not necessarily to the exclusion of temporal information.

ACKNOWLEDGMENTS

This work was supported by NIH Grants R01 DC 05216 and 5T32 DC 00038. An earlier version of this manuscript formed part of a Ph.D. thesis submitted by the first author to the Massachusetts Institute of Technology. We thank Christophe Micheyl, Louis Braida, Bertrand Delgutte, Gerald Kidd, John Grose, and two anonymous reviewers for their helpful comments on earlier versions of this manuscript.

[1]The sigmoid function was defined as

$$10 \log_{10}[f_0 \mathrm{DL}(\%)] = \mathrm{DL}_{\mathrm{min}} + [(\mathrm{DL}_{\mathrm{max}} - \mathrm{DL}_{\mathrm{min}})]/\sqrt{\pi} \times \int_{m(f_0 - f_{0,tr})}^{\infty} \exp[-(f_0')^2] df_0' \tag{1}$$

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

[2]The set of filter parameters that best fit the masking data pooled across listeners (not shown) was similar to the mean of each filter parameters fit to the individual listeners shown in Table I.

[3]Excitation pattern model fits were not performed for each individual listener. Since the intersubject variability in the $f_{0,\mathrm{tr}}$ level effects did not match the intersubject variability in the ERB level effects, 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.

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

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. (**2005**). "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. (**2006**). "The relationship between frequency selectivity and pitch discrimination: Sensorineural hearing loss," J. Acoust. Soc. Am. **120**, 3929–3945.

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.

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

de Cheveigné, A., and Pressnitzer, D. (**2006**). "The case of the missing delay lines: Synthetic delays obtained by cross-channel phase interaction," J. Acoust. Soc. Am. **119**, 3908–3918.

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

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 fre-

quency 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).
Ghitza, O. (**1986**). "Auditory nerve representation as a front-end for speech recognition in a noisy environment," Comput. Speech Lang. **1**, 109–130.
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.
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).
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.
Greenwood, D. D. (**1972**). "Masking by combination bands: Estimation of the levels of the combination bands $(n+1)f_1-nf_h$," J. Acoust. Soc. Am. **52**, 1144–1154.
Grose, J. H., Hall, J. W., and Buss, E. (**2002**). "Virtual pitch integration for asynchronous harmonics," J. Acoust. Soc. Am. **112**, 2956–2961.
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.
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 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 Fibers in the Cat's Auditory Nerve* (MIT Press, Cambridge, MA).
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.
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. (**1954**). "'Periodicity' pitch and 'place' pitch," J. Acoust. Soc. Am. **26**, 945.
Meddis, R., and Hewitt, M. (**1991**). "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.
Moore, B. C. J. (**1982**). *An Introduction to the Psychology of Hearing*, 2nd ed. (Academic, London).
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. (**1990**). "Frequency discrimination of complex tones with overlapping and non-overlapping harmonics," J. Acoust. Soc. Am. **87**, 2163–2177.
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., Glasberg, B. R., Flanagan, H., and Adams, J. (**2006**). "Frequency discrimination of complex tones; assessing the role of component resolvability and temporal fine structure," J. Acoust. Soc. Am. **119**, 480–490.
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.
Patterson, R. D. (**1976**). "Auditory filter shapes derived with noise stimuli," J. Acoust. Soc. Am. **59**, 640–654.
Pick, G. (**1980**). "Level dependence of psychophysical frequency resolution and auditory filter shape," J. Acoust. Soc. Am. **68**, 1085–1095.
Pressnitzer, D., Patterson, R. D., and Krumbholz, K. (**2001**). "The lower limit of melodic pitch," J. Acoust. Soc. Am. **109**, 2074–2084.
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*," Aud. 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.
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.
Rosen, S., and Baker, R. J. (**1994**). "Characterising auditory filter nonlinearity," Hear. Res. **73**, 231–243.
Rosen, S., and Stock, D. (**1992**). "Auditory filter bandwidths as a function of level at low frequencies (125 Hz–1 kHz)," J. Acoust. Soc. Am. **92**, 773–781.
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., 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.
Schouten, J. F. (**1940**). "The residue and the mechanism of hearing," Proc. K. Ned. Akad. Wet. **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.
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. 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.
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.
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.
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.
Weber, D. L. (**1977**). "Growth of masking and the auditory filter," J. Acoust. Soc. Am. **62**, 424–429.
Weiss, T. F., and Rose, C. (**1988**). "A comparison of synchronization filters in different auditory receptor organs," Hear. Res. **33**, 175–180.
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.
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, The Netherlands), pp. 376–394.