Pitch perception for mixtures of spectrally overlapping harmonic complex tones

Christophe Micheyl,[a] Michael V. Keebler, and Andrew J. Oxenham
Department of Psychology, University of Minnesota, Minneapolis, Minnesota 55455-0344

This study measured difference limens for fundamental frequency (DLF0s) for a target harmonic complex in the presence of a simultaneous spectrally overlapping harmonic masker. The resolvability of the target harmonics was manipulated by bandpass filtering the stimuli into a low (800–2400 Hz) or high (1600–3200 Hz) spectral region, using different nominal F0s for the targets (100, 200, and 400 Hz), and different masker F0s (0, +9, or −9 semitones) relative to the target. Three different modes of masker presentation, relative to the target, were tested: ipsilateral, contralateral, and dichotic, with a higher masker level in the contralateral ear. Ipsilateral and dichotic maskers generally caused marked elevations in DLF0s compared to both the unmasked and contralateral masker conditions. Analyses based on excitation patterns revealed that ipsilaterally masked F0 difference limens were small ($<2\%$) only when the excitation patterns evoked by the target-plus-masker mixture contained several salient (>1 dB) peaks at or close to target harmonic frequencies, even though these peaks were rarely produced by the target alone. The findings are discussed in terms of place- or place-time mechanisms of pitch perception.
© 2010 Acoustical Society of America. [DOI: 10.1121/1.3372751]

I. INTRODUCTION

Many sounds, including voiced speech, some animal vocalizations, and the sounds produced by most musical instruments, are spectrally complex and temporally periodic, or quasi-periodic. The "prototype" of such sounds is the harmonic complex tone (HCT), which consists of several sinusoidal components or harmonics with frequencies at integer multiples of the fundamental frequency (F0). The percept of an HCT is not usually that of a collection of individual tones, but rather a coherent sound with a unitary pitch, corresponding to the F0. Pitch plays a crucial role in music: sequences of pitches over time form melodies, and simultaneous combinations of pitches form the basis of harmony. Pitch also plays a role in the perception of speech, conveying cues regarding speaker identity, as well as prosodic and (in tone languages) lexical information. Finally, pitch provides a perceptual dimension along which different sources may be distinguished and followed or "tracked" over time. For instance, pitch may facilitate listening selectively to the speech of one talker in the presence of one or several competing talkers (Brokx and Nooteboom, 1982; Bird and Darwin, 1998; Darwin *et al.*, 2003), or following one melody in the presence of other melodies (Butler, 1979; Deutsch, 1979; Oxenham and Simonson, 2009).

This study addresses the question of how well changes in the pitch of one HCT can be discriminated in the presence of another HCT that is presented simultaneously in the same spectral region. The results are then related to the degree to which frequency components of the target and masker can be considered separated, or "resolved," in the auditory periph-ery. The question is not merely of theoretical interest. Reduced harmonic resolvability resulting from reduced frequency resolution in individuals with hearing loss of cochlear origin (Glasberg and Moore, 1986) could explain some of the listening difficulties experienced by these individuals in situations that involve concurrent harmonic sounds, such as voices and music (Moore and Carlyon, 2005; Oxenham, 2008).

Relatively few studies have examined the relationship between harmonic resolvability and pitch perception with concurrent harmonic sounds (Beerends and Houtsma, 1986; Beerends, 1989; Beerends and Houtsma, 1989; Carlyon, 1996a, 1996b; Micheyl *et al.*, 2006; Bernstein and Oxenham, 2008). Findings from these and other studies have been reviewed recently by Oxenham (2008) and Micheyl and Oxenham (2010), and are discussed briefly below.

Beerends and Houtsma (1989) measured listeners' ability to recognize the pitches of two simultaneously presented pairs of contiguous harmonics of different F0s, drawn randomly from a relatively small closed set. They found that if none of the components were "aurally resolved," performance (measured as the percentage of correct identifications of either one or both notes) was close to chance. Beerends and Houtsma (1989) did not provide a precise definition of aurally resolved, but referred to studies suggesting that the accurate perception of F0 is only possible when harmonics below about the tenth are present (Terhardt, 1970; Houtsma and Goldstein, 1972; Plomp, 1976).

Carlyon (1996a) measured difference limens for F0 (DLF0s) for bandpass-filtered harmonic complexes in the presence and absence of a simultaneous, spectrally overlapping masker. The masker had a fixed F0, intermediate between the F0s of the two targets presented on each trial. The target and masker either both contained resolved, or both

---

[a] Author to whom correspondence should be addressed. Electronic mail: cmicheyl@umn.edu

contained only "unresolved" harmonics according to the criteria defined by Carlyon and Shackleton (1994), whereby a HCT was considered as resolved if the average number of harmonics in the 10-dB bandwidth of auditory filters with center frequencies within the stimulus pass-band was lower than 2, and unresolved if that number was higher than 3.25. Carlyon (1996a) found that, when the target and masker complexes were both resolved prior to mixing, listeners could reliably discriminate relatively small changes in the target F0; performance was only moderately poorer in the presence of the masker than in the unmasked condition. In contrast, when the target and masker complexes were both unresolved according to the above definition, listeners heard the resulting mixture as a noise-like "crackle," and they were unable to distinguish two pitches (see also Carlyon, 1996b).

Rather than using equal-level targets and maskers, as was done in the earlier studies, Micheyl et al. (2006) measured the target-to-masker ratio (TMR) required for listeners to discriminate fixed differences in the target F0 at predefined levels of performance (70.7% or 79.4% correct). Stimuli were bandpass-filtered between 1200 and 3600 Hz, and the three nominal target F0s (100, 200, and 400 Hz), in conjunction with three average separations between the target and masker F0s (0, −7, and +7 semitones), yielded conditions with varying degrees of harmonic resolvability. In that study (as in Shackleton and Carlyon, 1994), a harmonic was considered resolved if no other component fell within the 10-dB bandwidth of the auditory-filter centered on that harmonic frequency. The results revealed that, when resolved target harmonics were present in the mixture, the threshold TMR (defined as the TMR corresponding to 70.7% or 79.4% correct) was usually negative, indicating that listeners could successfully segregate the target from the masker, and they could then listen selectively for changes in the target F0. In contrast, when all target and masker harmonics were unresolved prior to mixing, listeners required a positive TMR in order to reliably discriminate changes in the F0 of the target, suggesting that the target pitch could only be reliably tracked when the target dominated the overall sensation evoked by the mixture. Interestingly, in conditions where the target contained resolved harmonics before but not after mixing with the masker, negative threshold TMRs were occasionally observed. This might suggest that accurate F0 discrimination is sometimes possible even when no resolved harmonics are present. A similar conclusion was reached by Bernstein and Oxenham (2008), who showed that introducing a 3% difference in F0 between the odd and even harmonics of an HCT containing only unresolved harmonics (i.e., harmonics above the tenth) improved F0 discrimination performance to the point where it nearly equaled that achieved with only the even (resolved) harmonics present.

The present study sought to explore further the relationship between harmonic resolvability and listeners' ability to accurately perceive changes in the pitch of a target HCT in the presence and absence of a spectrally overlapping simultaneous masker, the F0 of which was fixed across observation intervals. A range of resolvability conditions was produced by filtering the stimuli into two different spectral regions, and by using three nominal (or average) F0s for the

targets (ranging from 100 to 400 Hz) and three relative masker F0s (equal to, 9 semitones above, or 9 semitones below the nominal target F0). The presence of resolved harmonics was determined based on excitation patterns (EPs) (Glasberg and Moore, 1990). This EP-based approach provides a more direct measure of harmonic resolvability than estimates based on component-spacing and auditory-filter-bandwidth considerations (Shackleton and Carlyon, 1994; Micheyl et al., 2006), and also takes into account the relative level of target and masker components at the output of auditory filters, which is the primary determinant of energetic masking.

To help distinguish between peripheral and more central effects, the binaural properties of the masker and target were varied. If listeners' ability to discriminate the F0 of the target complex depends on the spacing and level relationships of harmonics *within the same ear*, and listeners can selectively attend to the target ear, a contralateral harmonic masker should have little or no influence on performance. However, if listeners cannot make use of ear separation in pitch perception tasks, as suggested by some earlier studies (Houtsma and Goldstein, 1972; Gerson and Goldstein, 1978; Zurek, 1979; Beerends and Houtsma, 1989; Bernstein and Oxenham, 2003), then the impairment in pitch discrimination performance may be similar, regardless of whether the target and masker are presented to the same or different ears.

II. METHODS

A. Listeners

Five listeners (aged 20–26 years) took part in this experiment, all of whom had audiometric thresholds of 20 dB HL or better at octave frequencies between 250 and 8000 Hz. All listeners had received some musical education, and played a musical instrument at some point in their life, and one was a professional piano teacher and a practicing musician.

Before formal testing, the listeners were given the opportunity to familiarize themselves with the pitch discrimination task. The listeners had no difficulty understanding the instructions, and most of them needed very little practice before their DLF0s fell in the same range as those of two of the authors (both of whom had extensive experience with pitch discrimination tasks), as measured during pilot tests. For one of the listeners, the measured DLF0s on the first two runs were higher than expected based on data in the literature. That listener performed two additional practice runs before actual data collection began; this was sufficient to bring her DLF0s in line with those of the other listeners, and with data in the literature.

B. Procedure

DLF0s were measured using a two-interval two-alternative forced-choice (2I2AFC) procedure. On each trial, two 400-ms "target" harmonic complex tones differing in F0 were presented, separated by an interval of 500 ms. The higher-F0 complex was presented either first or second, with equal probability. The listener's task was to indicate whether the higher-F0 target occurred first or second. Responses were

Micheyl *et al.*: Pitch perception of concurrent harmonic complexes

given by pressing the "1" or "2" key on a computer numeric keypad. Visual feedback ("correct" or "false") was provided on the computer screen following each trial.

The F0s of the two target tones were geometrically centered on a nominal F0 (100, 200, or 400 Hz), and the amount by which they differed, $\Delta F0$ (expressed as a percentage of the lower F0) was varied adaptively using a two-down one-up rule, which tracked the 70.7%-correct point on the psychometric function (Levitt, 1971). The value of $\Delta F0$ was set to 90% (i.e., slightly less than an octave) at the beginning of each run. It was divided by a factor of 4 after two consecutive correct responses, and multiplied by that same factor after each incorrect response, until the first reversal from increasing to decreasing. A factor of 2 was used for the following two reversals, after which the step-size was fixed at a factor of $\sqrt{2}$. The value of $\Delta F0$ was not permitted to exceed 90%. If the tracking procedure called for a higher value than this, the value was set to the maximum, and the tracking procedure continued. If the maximum level was reached on eight (not necessarily consecutive) occasions during a run, the run was terminated, and no threshold estimate was returned. Each adaptive run terminated after six reversals were obtained using the final step-size. The geometric mean of the $\Delta F0$ values (in percent) at the last six reversals was taken as the threshold estimate for the run. Except for one listener, the mean DLF0s used in the plots and statistical analyses below are based on a minimum of (and usually more than) four threshold estimates per condition per listener. For one listener who dropped out of the study before completion, only two threshold measurements were obtained in some of the conditions. In runs that were terminated early due to the largest $\Delta F0$ value allowed in the tracking procedure (90%) being reached, the run was not discarded, which would have increased any under-estimation bias. Instead, each "unmeasured" threshold was replaced by the maximum allowed $\Delta F0$ value (90%) before averaging across runs. Any mean DLF0s that include such "replaced" estimates from any subject are identified in the results as not being reliably below 90%. All reported means and standard errors across runs or listeners are geometric.

Depending on the condition being tested, the target complex was either presented in isolation (condition "None") or accompanied by another complex, the "masker," which had an F0 equal to, 9 semitones below, or 9 semitones above, the "nominal" F0 of the target, defined as the geometric mean of the F0s of the two targets presented on a trial (100, 200, or 400 Hz); for brevity, the latter two conditions are referred to as the −9- and +9-semitone masker conditions. The target was always presented monaurally to the left ear. The masker was presented to the same ear as the target ("Ipsi" condition), to the opposite ear ("Contra" condition), or to both ears but with the level in the contralateral ear raised by 20 dB relative to that in the target ear, so that the masker was clearly lateralized to the opposite side from the target ("Dichotic" condition). The four masker conditions (None, Ipsi, Contra, and Dichotic) were tested in a partly randomized blocked fashion, so that one threshold measurement was obtained in each masker condition at a given nominal F0 and spectral region, before another F0-region combination was tested. Within

each block, the four masker conditions were presented in randomized order, with the exception that condition None was always tested first, i.e., the no-masker condition was presented first, followed by the Ipsi, Contra, and Dichotic masker conditions in random order. This was done to provide listeners with the opportunity to hear the target complex in isolation before the masker was introduced. The 0-semitone, −9-semitone, and +9-semitone masker-F0 conditions were tested in separate blocks, randomly intermingled within each test session.

C. Stimuli

The target HCTs had a total duration of 400 ms, including 20-ms raised-cosine ramps. The maskers, when present, were gated synchronously with the targets. The F0s of the two targets presented in each trial were smaller and larger than the nominal F0 by a factor of $\sqrt{1+\Delta F0/100}$. In this way the geometric mean of the two target F0s presented on each trial was equal to the nominal F0, while the difference between them was equal to $\Delta F0$ in percent, relative to the lower-F0 target. The starting phases of the harmonics were drawn randomly and independently from a uniform distribution spanning $0° - 360°$ on each presentation. The complexes were presented at a level of 50 dB SPL per component prior to filtering. Pink noise with a spectrum level of 20 dB (re 20 μPa) at 1 kHz was also presented. It was digitally lowpass-filtered in the spectral domain, using a rectangular filter with a corner frequency adjusted to coincide with the lower cutoff frequency of the complex tone filter (800 or 1600 Hz, depending on the spectral region being tested). The purpose of this background noise was to prevent listeners from detecting distortion products, which could have confounded the interpretation of the results by introducing resolved components in otherwise unresolved conditions. A fresh noise sample was generated on each trial. The noise was presented binaurally[1] during the presentation of the complex tones and was gated on and off with 20-ms raised-cosine ramps. In each trial the noise was turned on 400 ms before the onset of the first target complex in a trial and was turned off 400 ms after the offset of the second target complex.

The complexes were digitally bandpass-filtered using an eighth-order Butterworth filter with 6-dB cutoff frequencies of either 800 and 2400 Hz (LOW spectral region), or 1600 and 3200 Hz (HIGH spectral region), yielding a constant half-amplitude bandwidth of 1600 Hz. These two spectral regions (LOW and HIGH) were combined with the three nominal F0s (100, 200, and 400 Hz) to yield six conditions, which are referred to as, e.g., "100-LOW" for "100-Hz F0 in the LOW spectral region." The use of multiple spectral regions and F0 conditions was motivated by the consideration that the resolvability of frequency components in a HCT depends not only on the frequency spacing between the components, which is determined by F0, but also on the bandwidth of the peripheral auditory filters, which depends on spectral region. As pointed out by Carlyon and Shackleton (1994), by varying spectral region and F0 independently, one can separate the effects of harmonic resolvability from those of F0 or spectral region alone.
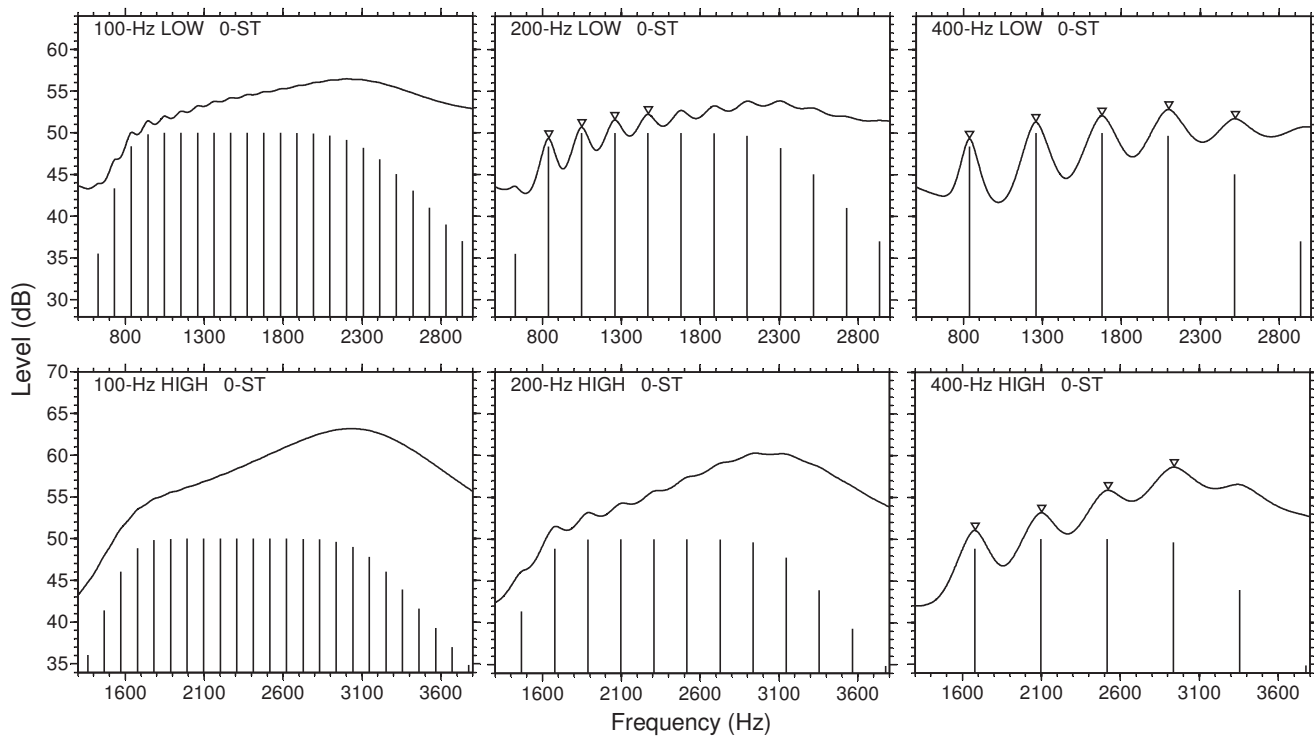
FIG. 1. Excitation patterns evoked by isolated target HCTs for the different stimulus conditions. Each panel corresponds to a different combination of spectral region and nominal F0, as indicated by the key. The downward-pointing triangles indicate EP peaks larger than 1 dB, when more than one such peak was detected. The magnitude spectra of the target complex before application of the middle-ear and headphone corrections is also shown in each panel (solid lines). For these simulations, the F0 of the target was set to $F0_{nom}\sqrt{1+\Delta F0/100}$, with $F0_{nom}$ equal to the nominal F0, and $\Delta F0=10\%$.

D. Apparatus

A Madsen Conera™ Diagnostic Audiometer (GN Otometrics, A/S) was used for pure-tone audiometry. During the experiments proper, stimulus presentation and response collection were controlled using the AFC software package (Stefan Ewert, Universität Oldenburg) under MATLAB (The MathWorks, Inc.). The stimuli were generated digitally and played out via a soundcard (LynxStudio L22) with 24-bit resolution and a sampling frequency of 32 kHz. They were presented to the listener via Sennheiser HD 580 headphones while seated in a double-walled sound-attenuating chamber (IAC).

E. Excitation pattern simulations

As indicated in the Introduction, there are different approaches to quantifying harmonic resolvability. Here we used EP simulations. The EPs were computed using the formulas given in Glasberg and Moore (1990). The characteristic frequencies of the simulated (roex) auditory filters were spaced 0.1 ERB_N apart. To improve peak-estimation accuracy, EPs were interpolated with a resolution of 0.001-ERB_N using cubic splines. Prior to the computation of EPs, the levels of the components were corrected to reflect the transfer functions of the middle-ear and of the HD580 headphones. The simulations also included pink noise with the same level as in the experiments.

A harmonic was considered resolved if it produced a separate EP peak with a level more than 1 dB above the levels of the two adjacent valleys on its upper and lower sides. According to this 1-dB criterion, for the stimuli used here (including the pink noise background), harmonics of the 200-Hz nominal-F0 complex were resolved up the seventh; the eighth and higher harmonics were unresolved. This is broadly consistent with the conclusions of several psychoacoustic studies in which direct measures of the ability to hear out harmonics were obtained (Plomp, 1964; Moore and Ohgushi, 1993; Moore et al., 2006), and one harmonic below that at which Bernstein and Oxenham (2006) estimated that the transition region between good and poor DLF0s occurred for F0s of around 175 Hz at moderate levels.[2] We also tested other values for the criteria. We found that using a criterion of 2 dB led to declaring harmonics higher than the fifth unresolved, while using a criterion of 0.5 dB led to declaring harmonics up to the 11th resolved, neither of which is in accord with our current understanding of resolvability. Consequently, the 1-dB criterion was used in all subsequent analyses.

Figure 1 shows EPs evoked by a target HCT for each of the different spectral region and nominal-F0 combinations, as indicated within each panel. For these simulations, the F0 of the target was set to $F0_{nom}\sqrt{1+\Delta F0/100}$ with $F0_{nom}$ equal to the nominal F0, and $\Delta F0=10\%$. Peaks in the EP larger than 1 dB are indicated by downward-pointing triangles. A 10% $\Delta F0$ is larger than the largest mean unmasked DLF0 measured in the experiment. This shows that, in the 100-LOW, 100-HIGH, and 200-HIGH conditions, the two target HCTs presented on a trial never contained resolved harmonics. In contrast, in the 200-LOW, 400-LOW, and 400-HIGH conditions, the target HCTs always contained at least three (and up to four) resolved harmonics, prior to mixing with the masker.
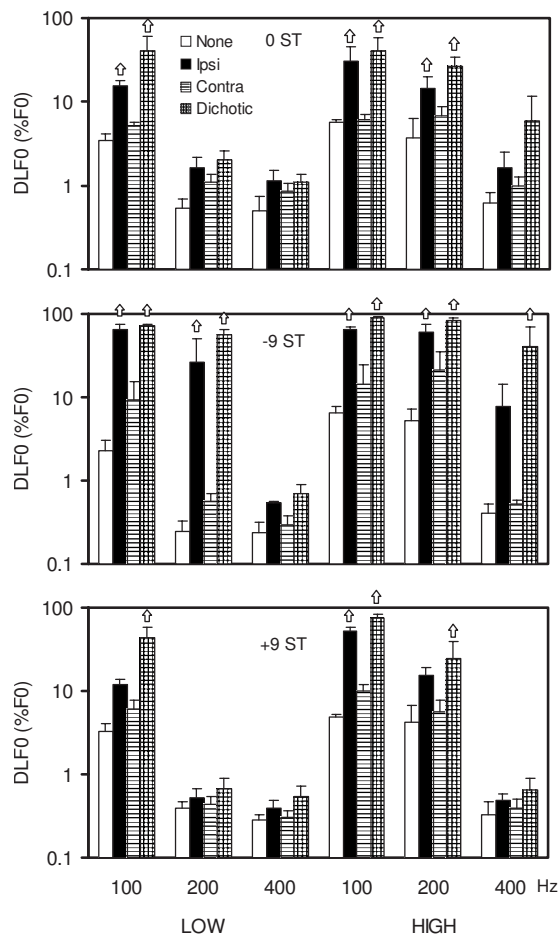
FIG. 2. Mean DLF0s expressed as a percentage of the lower F0. The different conditions are presented along the *x*-axis. The three panels correspond to the three masker-F0 conditions: masker F0 equal to the nominal F0 of the targets (0 ST, top panel); masker F0 9 semitones below the nominal target F0 −9 ST, middle panel); masker F0 9 semitones above the nominal target F0 (+9 ST, bottom panel). The different masker type conditions are indicated by different histogram-bar fillings: open for None, solid for Ipsi, striped for Contra, and tiled for Dichotic. Upward arrows represent DLF0s that were not reliably below the maximum value of 90%.

In addition to EPs evoked by isolated complexes, we computed EPs for target-plus-masker mixtures. To facilitate comparisons with the experimental results, the ΔF0s between the two target HCTs in these simulations were set based on the DLF0s measured in the experiment. Therefore, the resulting EPs are presented after the description of the experimental results.

III. RESULTS

The mean DLF0s of the five listeners in the different stimulus conditions are shown in Fig. 2. The upper panel shows DLF0s obtained when the F0 of the masker (when present) was equal to the nominal F0 of the target. The middle and lower panels show DLF0s when the masker F0 was 9 semitones below (middle panel) or above (lower panel) the nominal F0 of the target. The filled and textured bars show DLF0s measured with the masker present. Each panel also shows unmasked DLF0s (open bars). Although these unmasked DLF0s were measured under identical stimulus conditions in all three panels, they are shown sepa-

rately to indicate that they were obtained in different blocks of trials. These unmasked DLF0s displayed a consistent pattern across the three panels. Consistent with previous studies (Houtsma and Smurzynski, 1990; Shackleton and Carlyon, 1994), the DLF0s were below 1% (mean=0.37%) for the three conditions in which the targets contained resolved harmonics (i.e., 200-LOW, 400-LOW, and 400-HIGH), and between 2% and 7% (mean=4.2%) for the three conditions in which the targets contained only unresolved harmonics (i.e., 100-LOW, 100-HIGH, and 200-HIGH conditions). The following two sections consider the influence of the masker.

A. Masker F0 equal to nominal target F0

1. Ipsilateral masker

Comparing the open and solid bars in the upper panel of Fig. 2, it can be seen that the ipsilateral masker with an F0 equal to the nominal F0 of the target generally produced elevated DLF0s relative to the unmasked condition. On average across all combinations of spectral region and F0, masked DLF0s were more than three times larger than the corresponding unmasked DLF0s. This effect was confirmed statistically by the results of a three-way (spectral region $\times$ F0 $\times$ masker presence) repeated-measures analysis of variance (RMANOVA) on the log-transformed[3] DLF0s, which showed a significant main effect of masker presence [$F(1,4)=74.60$, $p=0.001$]. The upward-pointing arrows indicate conditions in which DLF0s sometimes reached the maximum allowed ΔF0 value of 90%, and may therefore be an underestimate of the "true" DLF0. For the ipsilateral masker, this occurred in the three conditions in which the targets contained no resolved harmonics before mixing with the masker, i.e., the 100-LOW, 100-HIGH, and 200-HIGH conditions. Thus, in these conditions, we can only place a lower bound on thresholds. Based on the data shown in Fig. 2, this lower bound seems to be about 15%. Therefore, we can conclude that ΔF0s of 15% or more could not be reliably discriminated with 70.7% accuracy. This value of 15% is larger than two musical semitones, and about four times greater than DLF0s in quiet.

In contrast, in the three conditions in which the targets contained resolved harmonics prior to mixing (i.e., the 200-LOW, 400-LOW, and 400-HIGH conditions), DLF0s in the presence of the masker were less than 2% on average.

2. Contralateral masker

DLF0s measured in the presence of the contralateral masker (horizontal-striped bars) were also significantly higher than DLF0s measured in the absence of a masker (open bars) [main effect of contralateral masker presence in a three-way (spectral region $\times$ F0 $\times$ masker presence) RMANOVA: $F(1,4)=28.39$, $p=0.006$]. However, this effect, which corresponded to a factor of 1.56 on average, was significantly smaller than that produced by the ipsilateral masker [as indicated by a significant main effect of masker type in a three-way (F0 $\times$ spectral region $\times$ masker type: ipsilateral vs. contralateral) RMANOVA on the difference in DLF0s between masked and unmasked conditions: $F(1,4)=75.41$, $p=0.001$]. The contralateral masker only had a sig-

nificant effect in the 100-, 200-, and 400-LOW conditions [$3.10 < t(4) < 5.22$; $0.006 < p < 0.036$]. In the HIGH region, the effect of the contralateral masker was either non-significant [100-HIGH: $t(4)=0.48$, $p=0.656$; 200-HIGH: $t(4)=1.94$, $p=0.125$], or borderline [400-HIGH: $t(4)=2.76$, $p=0.051$, for the 400-HIGH condition].

3. Dichotic masker

The DLF0s measured in the presence of the dichotic masker (tiled bars) were much higher than the corresponding unmasked DLF0s [main effect of dichotic masker presence in a three-way (spectral region $\times$ F0 $\times$ dichotic-masker presence) RMANOVA: $F(1,4)=37.99$, $p=0.004$]. On average, these DLF0s were larger than those measured in the presence of the ipsilateral masker [main effect of masker type in a three-way (masker type $\times$ spectral region $\times$ F0) RMANOVA on the log-transformed masked DLF0s: $F(1,4)=75.41$, $p=0.001$]. These results indicate that perceiving the target and masker at opposite sides of the head did not reduce interference. Taken together with the results for the contralateral masker condition, the results suggest a peripheral locus for the interference effects observed with the ipsilateral masker.

B. Masker F0 9 semitones below or above the nominal target F0

1. Ipsilateral masker

The ipsilateral masker with an F0 9 semitones below the nominal F0 of the two targets produced significant increases in DLF0s relative to the unmasked condition [main effect of masker presence in a three-way (masker presence $\times$ spectral region $\times$ F0) RMANOVA on the DLF0s: $F(1,4)=26.61$, $p=0.006$]; the difference in DLF0s was significant for all combinations of spectral region and target F0 [Fisher's LSD tests, $3.87 < t(4) < 5.46$, $p < 0.05$] except 400-LOW [$t(4)=1.75$, $p=0.108$]. The ipsilateral masker with an F0 9 semitones above the nominal target F0 also caused a significant elevation in DLF0s [$F(1,4)=15.92$, $p=0.016$]. However, when tested for individual combinations of spectral region and F0, the effect of this masker was statistically significant only for the 100-LOW [$t(4)=4.39$, $p=0.012$] and 100-HIGH conditions [$t(4)=5.66$, $p=0.005$].

Overall, DLF0s were larger in the presence of the lower-F0 than higher-F0 ipslateral masker [main effect of relative masker F0 in a three-way (relative masker F0 $\times$ spectral region $\times$ F0) RMANOVA on the ipsilaterally masked DLF0s: $F(1,4)=29.22$, $p=0.006$]. DLF0s measured in the presence of the lower- and higher-F0 ipsilateral masker were compared for each condition of spectral region and nominal target F0 separately. The results revealed significant differences in all conditions [$2.94 < t(4) < 5.22$, $0.006 < p < 0.043$], except for the 400-LOW [$t(4)=1.31$, $p=0.262$] and 100-HIGH [$t(4)=1.35$, $p=0.249$] conditions.

2. Contralateral masker

Although the contralateral masker with an F0 9 semitones below the nominal target F0 caused a statistically significant increase in DLF0s relative to those for the unmasked

condition [main effect of masker presence in a three-way (masker presence $\times$ spectral region $\times$ F0) RMANOVA: $F(1,4)=9.10$, $p=0.039$], comparisons performed on each spectral region and F0 combination separately showed a significant effect only for the 200-HIGH condition [$t(4)=3.10$, $p=0.036$]; in all other conditions, the effect was not significant [$0.61 < t(4) < 1.89$, $0.132 < p < 0.576$]. The contralateral masker with an F0 9 semitones above the nominal target F0 did not cause a statistically significant increase in DLF0s overall.

3. Dichotic masker

The dichotic masker with an F0 9 semitones below the nominal target F0 caused a significant elevation in DLF0s compared to the baseline [main effect of masker presence in a three-way in a three-way (masker presence $\times$ spectral region $\times$ F0) RMANOVA: $F(1,4)=29.51$, $p=0.006$]. This effect was significant for every combination of spectral region and F0 [$4.59 < t(4) < 5.91$, $p < 0.05$] except 400 LOW [$t(4)=2.17$, $p=0.096$]. The higher-F0 dichotic masker also caused a significant increase in DLF0s [$F(1,4)=18.37$, $p=0.013$], but the effect was significant only for some of the spectral region and F0 conditions, namely, the 100-LOW, 100-HIGH, and 200-HIGH conditions [$4.38 < t(4) < 6.73$, $p < 0.05$].

IV. DISCUSSION

A. Excitation pattern simulations

To aid the interpretation of the results in terms of resolvability, EPs were computed for the target-plus-masker mixtures of HCTs that were used in the experiment. The EPs were computed for both intervals of a 2IAFC trial, with the ΔF0 adjusted to equal the mean threshold measured in the corresponding condition (as shown in Fig. 2). However, to avoid clutter in the figures, only the EPs evoked by mixtures containing the higher-F0 target (with an F0 equal to $F0_{nom}\sqrt{1+\Delta F0/100}$) are shown.

The resulting EPs are shown in Fig. 3 (LOW spectral region) and Fig. 4 (HIGH spectral region). Each panel corresponds to a given nominal F0 and relative masker-F0 condition, as indicated by the key in each panel. The magnitude spectra of the target and masker are superimposed and are represented by solid and dashed vertical lines, respectively. The solid curves show the EPs evoked by the mixture. The downward-pointing triangles mark EP peaks that have a level more than 1 dB higher than that the adjacent troughs on both sides of the peak.

For the three conditions in which the ipsilateral masker was found to increase DLF0s by a large amount, i.e., 100-Hz LOW, 100-Hz HIGH, and 200-Hz HIGH, the EPs evoked by target-plus-masker mixtures never contained more than one peak greater than 1 dB. In contrast, in the three conditions for which the ipsilateral masker had a relatively small effect, and masked DLF0s remained relatively small ($<2\%$), i.e., 200-LOW, 400-LOW, and 400-HIGH, the EPs displayed at least three peaks of more than 1 dB. These observations suggest that the ability of listeners to discriminate F0 accurately
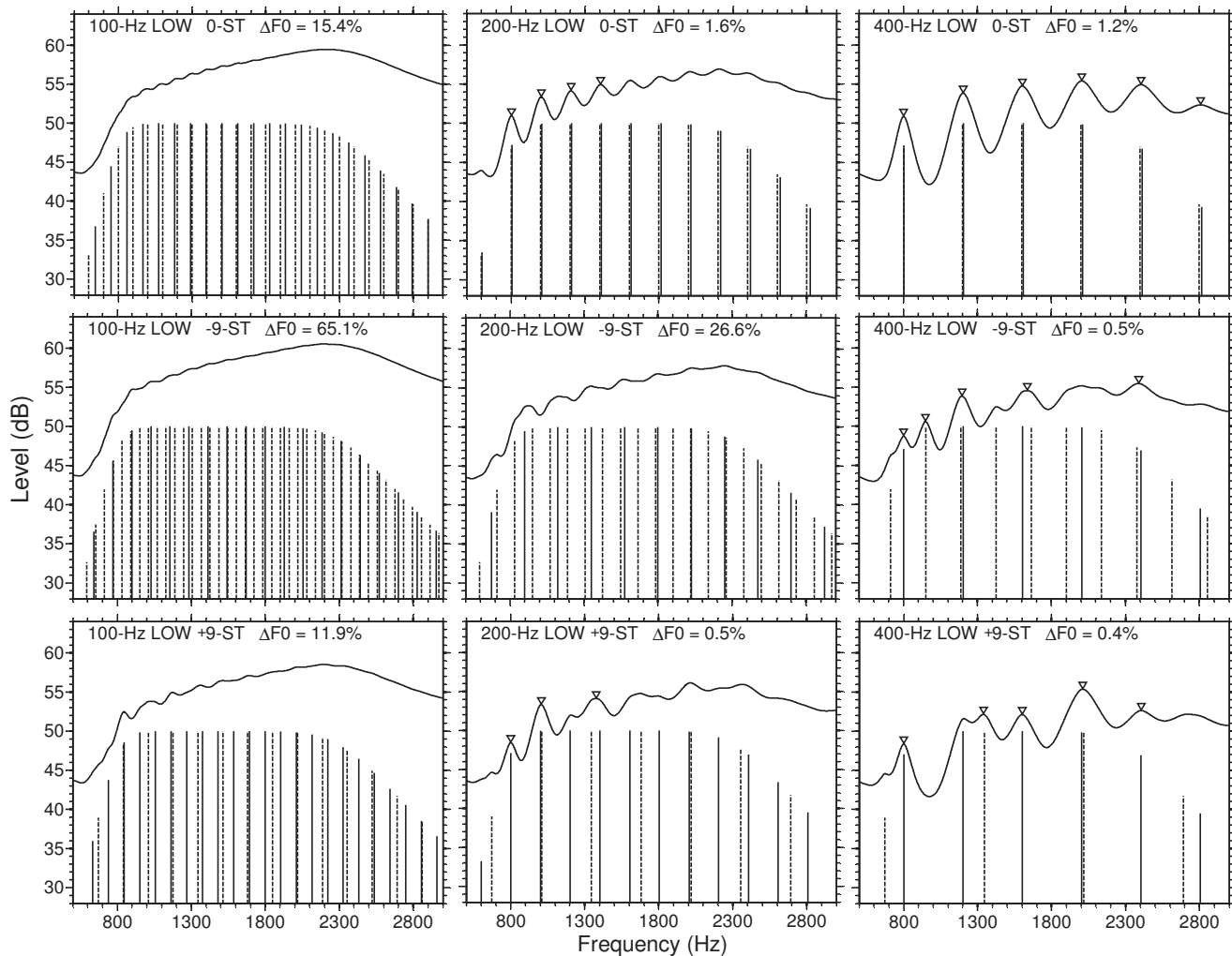
FIG. 3. EPs evoked by target-plus-masker mixtures filtered into the LOW spectral region. Each panel corresponds to a different combination of spectral region, nominal target F0, and relative masker F0, as indicated by the key. The downward-pointing triangles indicate EP peaks larger than 1 dB. The magnitude spectra of the target and masker complexes (before application of the middle-ear and headphone corrections) are shown as solid and dashed lines, respectively. For these simulations, the F0 of the target was set to $F0_{nom}\sqrt{1+\Delta F0/100}$, with $F0_{nom}$ equal to the nominal F0, and $\Delta F0$ equal to the mean. DLF0 measured in the corresponding experimental condition. The F0 of the masker was equal (top row) to, 9 semitones below (middle row), or 9 semitones above (lower row), the nominal target F0. The nominal target F0, $\Delta F0$, and masker-F0 position relative to the nominal target F0 (0, −9, or +9 semitones) are indicated in each panel.

in the presence of the ipsilateral masker is related to whether the EP evoked by the target-plus-masker mixture contains several salient (>1 dB) peaks.

Interestingly, EP peaks larger than 1 dB were rarely evoked by individual target or masker harmonics. More often, they reflected a mixture of two very closely spaced harmonics, one from the target and one from the masker. Yet listeners were able to achieve low DLF0s, as indicated by the results for the 200-LOW, 400-LOW, and 400-HIGH conditions. This suggests that DLF0s in the masked F0-discrimination task did not depend critically on whether or not harmonics of the target and masker fell into different auditory filters, and evoked separate EP peaks—as implied by some definitions of "resolvability." Instead, it seems that masker harmonics could in some cases combine with target harmonics to create a single peak that was used by the auditory system to extract the target pitch. In the following two sections, we consider whether F0-estimation schemes based

solely on place representations, or a combination of place and time information, can account for these results.

B. Place-based F0-estimation schemes for single and concurrent complexes

Place-based F0-estimation schemes [Wightman, 1973; Terhardt, 1974; Duifhuis *et al.*, 1982; for a review, see de Cheveigné (2005)] typically involve two stages. In the first stage, the frequencies of individual harmonics are estimated. In the second stage, these frequencies are used to estimate F0. A commonly used method for estimating F0 based on a set of observed frequencies involves dividing each of the frequencies by successive integers, and computing a histogram of the resulting values; the highest frequency corresponding to a mode of the histogram is the F0 estimate (Schroeder, 1968).

To determine whether this simple place-based F0-estimation scheme could explain the experimental results, we
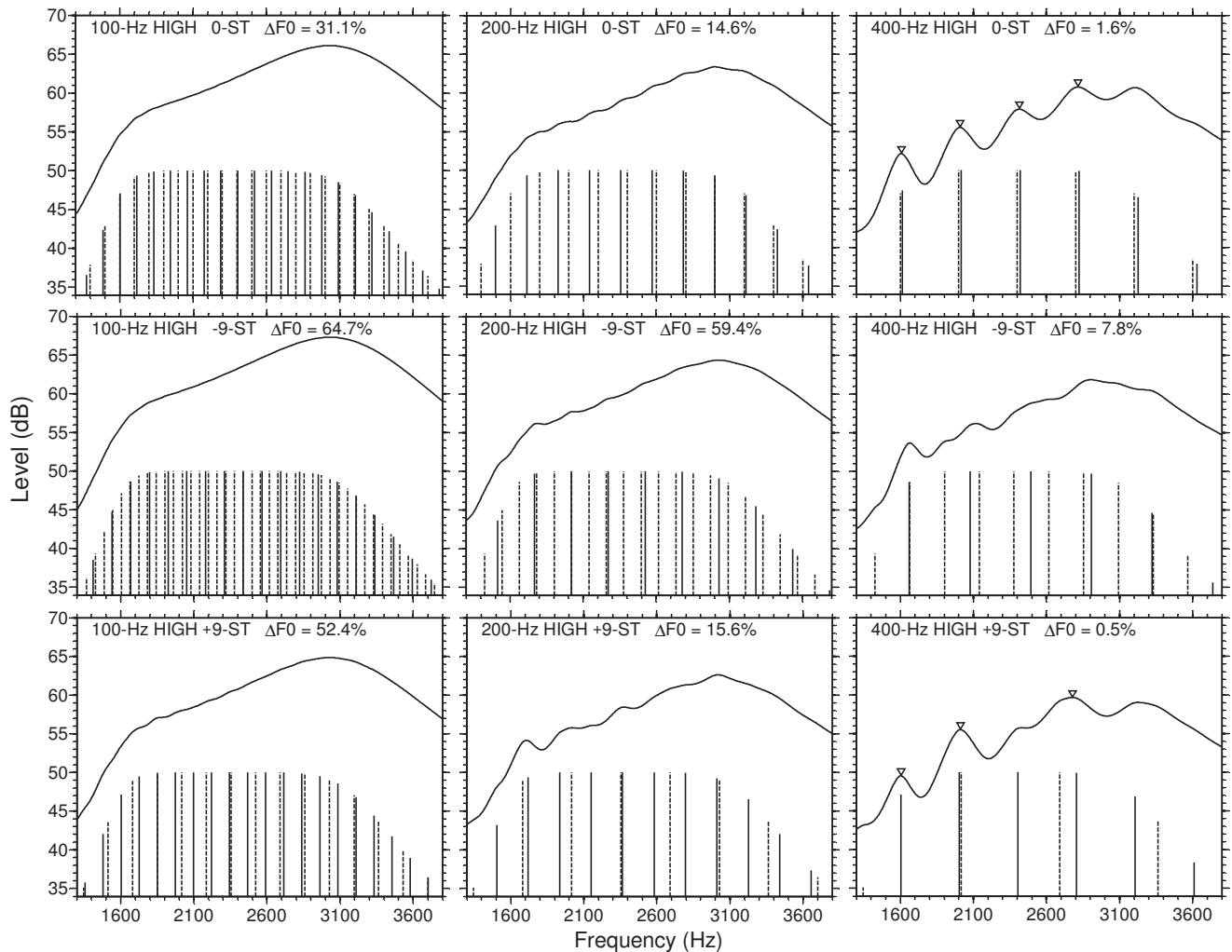
FIG. 4. EPs evoked by target-plus-masker mixtures filtered into the HIGH spectral region. For further details, see Fig. 3.

computed Schroeder histograms based on the frequencies of peaks larger than 1 dB in the EPs shown in Figs. 3 and 4. To estimate F0, the frequencies of the peaks were divided by successive integers between 1 and 100, and the resulting list of frequencies was used to build a histogram. The centers of the bins in the histogram were spaced regularly on a log scale going from 50 to 700 Hz, encompassing the range of target and masker F0s that could possibly occur in the experiment. The spacing between consecutive bin centers on the log scale was chosen to correspond to a step of 0.1% of the F0. The highest bin center corresponding to a mode of the histogram was selected as the estimated F0. These "raw" F0 estimates are reported in Table I. Even for isolated HCTs, F0 estimates derived using this technique are sometimes equal to an integer multiple or sub-multiple of the true F0, other than 1 (Stubbs and Summerfield, 1988). To remedy this problem, we computed integer multiples and sub-multiples of the estimated F0, and picked the value closest to the actual F0 of the target or masker in the corresponding stimulus condition. The resulting "corrected" F0 estimates are reported in Tables II and III.

1. Masker F0 equal to the nominal target F0

First, consider the conditions in which the F0 of the masker was equal to the nominal target F0. The F0s that were estimated in these conditions are shown in the first column of Tables I–III. While the raw estimates (Table I) were often in error, reflecting the susceptibility of the Schroeder-histogram method to octave confusions mentioned above, the corrected estimates were less than 1% away from the true target F0 (Table II), and masker F0 (Table III). This can be understood based on the observation in Figs. 3 and 4 that, even though corresponding harmonics of the target and masker were too close to each other to evoke separate EP peaks, pairs of harmonics from the two HCTs were distant enough from neighboring pairs to produce a salient peak. The frequencies of these peaks were intermediate between the harmonic frequencies of the two HCTs. Therefore, while these frequencies did not equal precisely those of the target harmonics, they were slightly but consistently shifted toward them. Specifically, the corrected F0 estimates were 0.6%–0.8% higher for mixtures containing the higher-F0 target (shown in Figs. 3 and 4) than for mixtures containing the lower-F0 target (not shown in Figs. 3 and 4). Although such changes are small, they are comparable with DLF0s for single complexes containing resolved harmonics in their passband, which according to the present study, and earlier ones (Shackleton and Carlyon, 1994; Micheyl and Oxenham, 2004), are around 0.5%.

TABLE I. F0s estimated from the frequencies of salient peaks in the EPs shown in Figs. 3 and 4. These F0 estimates were obtained from the frequencies of salient (>1 dB) peaks in the EP evoked by each target-plus-masker mixture, using the Schroeder-histogram method, as described in the text. The spectral region (LOW, HIGH) is indicated in the first column. The nominal F0 is indicated in the second column. The second column indicates whether the estimates reported on the corresponding line were obtained from target-plus-masker mixtures containing the lower-F0 target or the higher-F0 target. The last three columns show the estimated target F0s in the corresponding stimulus condition, for the three relative masker-F0 conditions (0 ST, −9 ST, and +9 ST). Empty cells correspond to conditions in which one or both mixtures contained no EP peak larger than 1 dB, preventing estimation of the F0. Rows corresponding to combinations of spectral region and nominal F0 for which no F0 estimate could be obtained are not shown.

Region	F0$_{nom}$ (Hz)	Tgt F0	0 semitones (Hz)	−9 semitones (Hz)	+9 semitones (Hz)
LOW	200	Lower	67		687
		Higher	201		690
	400	Lower	80	63	399
		Higher	201	599	401
HIGH	400	Lower	400		107
		Higher	403		96

If the frequencies of EP peaks evoked by pairs of neighboring target and masker harmonics were approximately equal to the average frequency of the two harmonics, masked DLF0s in these conditions should be roughly double those measured in the corresponding unmasked conditions. This prediction is not very far off: on average across the 200-LOW, 400-LOW, and 400-HIGH conditions, masked DLF0s were 2.6 times larger than unmasked DLF0s. The slightly larger-than-predicted effect of the masker could be due to the fact that EP peaks evoked conjointly by two harmonics separated by a few Hz were somewhat wider than EP peaks evoked by a single harmonic, so that their frequency could not be estimated quite as accurately.

These observations are consistent with the hypothesis that, in conditions in which the masker F0 equaled the nominal target F0, and target and masker harmonics were very close in frequency, performance was based on the discrimination of changes in the F0 estimated from the frequencies of salient peaks in place representations of the target-plus-masker mixture, or on shifts in the EP slopes surrounding each peak (Zwicker, 1952).

2. Masker F0 9 semitones away from the nominal target F0

Next, consider the conditions in which the masker F0 was 9 semitones below or above the nominal target F0. The F0s that were estimated from the frequencies of EP peaks in these conditions are indicated in the middle and last (right-hand) columns of Tables II and III. Except for the 400-LOW condition with the masker F0 9 semitones above the nominal target F0, these estimates were at least 4% (and up to 34%) away from the true (lower and higher) target F0s (Table II). Such large estimation errors are due to the fact that in these conditions, the EPs contained peaks, the frequencies of which were intermediate between those of target and masker harmonics separated by several percent. This is especially apparent in the panels corresponding to the 200-LOW and 400-HIGH conditions with the masker F0 9 semitones above the nominal F0 of the targets, and to the 400-HIGH condition with the masker 9 semitones below the target, in Figs. 3 and 4. These EP peaks, which did not correspond precisely to a target harmonic, introduced spurious entries into the Schroeder-histogram, resulting in F0 estimates that corresponded neither to the target F0, nor to the masker F0.

Deviations between the estimated and true target F0s might not necessarily prevent accurate performance in the F0-discrimination task, as long as the difference between the estimated F0s is large enough to be detected, and is of the same sign as the difference between the true target F0s—so that the direction of the F0 change between the first and second intervals can be identified correctly. However, this was not always the case. For instance, in the 400-HIGH condition with the masker F0 9 semitones above the nominal target F0, the estimated F0 of the lower-F0 target was *higher* than the estimated F0 of the higher-F0 target. Yet in this condition, the listeners achieved very small DLF0s (0.4% on average). This indicates that the human auditory system is more effective at estimating the pitches of concurrent harmonic complexes than predicted by the EP model and Schroeder-histogram. The failure of the simple F0 estimation

TABLE II. Corrected F0 estimates, and corresponding deviations from the true target F0s. These corrected estimates are integer multiples or sub-multiples of the raw F0 estimates shown in Table I. The integer multiple that fell closest to the actual target F0 in the corresponding condition was selected. These estimates represent the best (i.e., closest) estimate of the target F0 that could be obtained from the measured frequencies of salient EP peaks after eliminating octave confusions in the Schroeder-histogram method. The columns are as in Table I.

Region	F0$_{nom}$ (Hz)	Tgt F0	0 semitones [Hz (%)]	−9 semitones [Hz (%)]	+9 semitones [Hz (%)]
LOW	200	Lower	200 (0.6)		229 (14.8)
		Higher	201 (0.2)		230 (14.7)
	400	Lower	399 (0.4)	377 (5.7)	399 (0.0)
		Higher	402 (0.2)	300 (33.9.3)	401 (0.1)
HIGH	400	Lower	400 (0.7)		426 (6.8)
		Higher	403 (0.1)		384 (4.5)

TABLE III. Corrected F0s estimates and corresponding deviations from the true masker F0s. These corrected estimates are integer multiples or sub-multiples of the raw F0 estimates shown in Table I. The integer multiple that fell closest to the actual masker F0 in the corresponding condition was selected. They represent the best (i.e., closest) estimate of the masker F0 that could be obtained from the measured frequencies of salient EP peaks after eliminating octave confusions in the Schroeder-histogram method. The columns are as in Table I.

Region	F0$_{nom}$ (Hz)	Tgt F0	0 semitones [Hz (%)]	−9 semitones [Hz (%)]	+9 semitones [Hz (%)]
LOW	200	Lower	200 (0.2)		344 (2.1)
		Higher	201 (0.6)		345 (2.6)
	400	Lower	399 (0.2)	252 (5.8)	798 (18.7)
		Higher	402 (0.4)	200 (19.1)	802 (19.3)
HIGH	400	Lower	400 (0.1)		639 (5.2)
		Higher	403 (0.7)		671 (0.2)

scheme described above does not necessarily imply that place-based models are inconsistent with the experimental data. However, it indicates that in order to account for these data, a more sophisticated F0-estimation scheme is required. One approach that has been proposed for estimating the F0s of two concurrent sounds involves computing two F0 estimates successively: first, based on the frequencies of all peaks present in the place representation; then, using only frequencies that are not candidate harmonics of the F0 estimated at the first stage (Parsons, 1976). One limitation of this approach is that, when harmonics from the two sounds are relatively close in frequency, candidate harmonics of both F0s are eliminated. Another potential problem with this method is that, if the majority of peaks in an EP were produced by pairs of nearby harmonics from the target and masker, the first estimated F0 (based on all peaks present) may fit neither the true masker F0, nor the true target F0; if this is the case, using integer multiples of that first estimated F0 to reject peaks may not help much in estimating either of the two F0s present.

Another strategy that has been devised for estimating the F0s of two simultaneous tones involves searching simultaneously for *two* harmonic sieves, which *conjointly* best describe the EP, or other place representation, evoked by two concurrent harmonic sounds. This approach was used by Scheffers (1983) to simulate the identification of concurrent vowels by human listeners. More recently, Larsen *et al.* (2008) applied a joint F0-estimation algorithm to recover the F0s of two concurrent HCTs based on rate-place profiles at the level of the auditory nerve. These authors used a form of analysis-by-synthesis, in which rate-profiles evoked by a mixture of two sounds were matched with broad templates generated by a simple model of auditory nerve responses. This scheme could estimate accurately the F0s of *both* HCTs even when their harmonics were so close in frequency that each pair of harmonics evoked a single peak in the rate-place profiles—similar to the EPs for the 200 and 400 Hz F0s in the top row of Fig. 3. Therefore, an F0-estimation scheme of the type proposed by Larsen *et al.* predicts relatively accurate F0 discrimination of the target even in conditions in which all harmonics of the target are close in frequency to a harmonic from the masker, as found in the present results. In the relevant conditions (200-LOW, 400-LOW, and 400-HIGH, with the masker F0 at 0 ST), relatively small DLF0s (between 1% and 2%) were observed in the presence of the ipsilateral masker.

According to Larsen *et al.* (2008), the only situations in which their scheme fails are when the spectral components of the two sounds are too unresolved, leading to difficulties in fitting even broad templates. Thus, the model is expected to fail in conditions for which the harmonics of the target and masker are already unresolved prior to mixing, as was the case in the 100-LOW and 100-HIGH conditions of the present study. It is likely that the algorithm would also fail in other conditions in which the EPs contained no salient peaks, such as the 200-HIGH condition, or the 200-LOW condition with the masker F0 9 semitones below the nominal target F0. This prediction would be consistent with our finding that, in

these conditions, listeners were not consistently able to discriminate the target F0, or had very high DLF0s.

To summarize, a simple place-based scheme that uses salient (>1 dB) peaks in EPs evoked by mixtures of HCTs to estimate an overall F0 can potentially explain our finding of relatively small ($<2\%$) DLF0s in conditions that involve target and maskers with similar F0s, even though none of these harmonics was individually resolved. However, such a simple scheme cannot explain the thresholds obtained in conditions in which the masker F0 was 9 semitones below or above the nominal target F0. In these conditions, a more elaborate template-matching scheme, such as that proposed by Larsen *et al.* (2008), may be needed to account for human listeners' ability to accurately discriminate pitch in mixtures of concurrent harmonic complexes based on EPs. Alternatively, this ability may rely on more accurate place representations than predicted by the EP model, or on a combination of place and time information, as discussed in the following section.

C. Place-time models of concurrent sound perception

While the above analysis was cast in terms of place models, it should not be taken to imply that the results are in any way inconsistent with temporal models of pitch perception that estimate periodicities in the input signal based on waveforms at the output of peripheral auditory filters [Meddis and Hewitt, 1992; de Cheveigné, 1993; Cariani, 2001; for a review, see de Cheveigné (2006)]. For instance, Meddis and Hewitt's (1992) computational model of concurrent-vowel perception involves an initial stage that simulates peripheral filtering, followed by the computation of autocorrelation functions (ACFs) at the output of each filter. Although the ACFs are summed across all channels to estimate a first F0, this estimate is subsequently used to sort the channels into two groups depending on whether the periodicity that dominates their output matches the first estimated F0 or not. While this scheme was used to model the identification of concurrent vowels, it could be modified to model F0 discrimination of a target harmonic complex in the presence of a harmonic masker. de Cheveigné's (1993) "cancellation" model uses the estimate of the F0 of the masker to create a temporal "sieve" at the corresponding periodicity, which is then used to "cancel out" the masker F0, and facilitate the estimation of the target F0. Cariani's (2001) "timing nets" can also be described as "temporal sieves," which extract common or recurrent spike patterns in the input, and use these patterns to automatically extract concurrent F0s.

While implementing these models and testing their predictions on the stimuli used in the current study is beyond the scope of this article, it is relatively clear *a priori* that place-time models are in no way inconsistent with the present finding of a generally good correspondence between stimulus conditions in which discrimination of the target F0 remained relatively accurate after the masker was introduced, and conditions in which salient EP peaks were present. The presence of salient EP peaks corresponding to individual target harmonics is an indication that there exist peripheral channels in which the target-to-masker ratio is relatively high. A higher target-to-masker ratio should facilitate the estimation of the

frequencies of individual target harmonics and, consequently, of the target F0, by a central processor, regardless of whether this processor operates on the basis of place information or temporal information (Goldstein, 1973; Srulovicz, and Goldstein, 1983).

Therefore, the results of this study should not be interpreted as providing evidence against temporal models of F0 perception and concurrent F0 extraction. However, the findings do provide further evidence that listeners' ability to discriminate relatively accurately the F0 of a harmonic target in the presence of a concurrent harmonic masker is not independent of peripheral resolution. Whether this influence is mediated by a place-based mechanism, or by a mechanism that operates on temporal information at the output of peripheral auditory filters remains an open question.

D. Relationship with earlier studies on the perception of concurrent HCTs

1. Influence of frequency resolution and harmonic resolvability on concurrent F0 perception

In line with earlier studies (Carlyon, 1996a, 1996b, 1997; Micheyl et al., 2006), when both the target and masker contained only unresolved harmonics before mixing, performance in discriminating F0 differences was very poor, with DLF0s well above 10% and in most cases not reliably below our limit of 90%. In fact, listeners in Carlyon's (1996a) experiment were still able to discriminate the target F0 at levels above chance, probably because the average periodicity of the combined complex co-varied with the F0 of the target complex. This "mean rate" cue, discussed by Carlyon (1996a, 1997), was available in the earlier experiments using harmonics with constant (sine) starting phases, but appears to have been less available in our experiment, presumably due to our use of random-phase complexes [see Micheyl et al. (2006) for a more detailed discussion].

An interesting and less well-researched area involves target complexes that included resolved harmonics before mixing, but not after mixing. Micheyl et al. (2006) noted some conditions in which F0 discrimination was relatively good, but the target harmonics were unlikely to have been resolved after mixing with the masker. Here we explored this question in more detail by using EP simulations. We found that DLF0s were only low (<2%) when several salient (>1 dB) spectral peaks were present close to frequencies corresponding to target harmonics. Thus, even if the target harmonics are not resolved in the sense of not being close to other components, it seems that salient spectral peaks in the EP representation corresponding to target harmonic frequencies may be a necessary prerequisite for low DLF0s. A good example of a target that produces clear EP peaks before, but not after, mixing is the 200-Hz LOW condition in the presence of the masker with an F0 that is 9 semitones below the nominal target F0. In this case, the DLF0 went from between 0.2% and 0.3% in the absence of the masker to an average of 26.6%, and not always measurable, with the −9-semitone masker. The poor DLF0s coincide with the elimination of salient EP peaks by the addition of the masker. The results are therefore consistent with the idea that salient spectral peaks are necessary for good pitch perception in the presence of a masking harmonic complex.

This conclusion is at odds with that of a recent paper by Bernstein and Oxenham (2008). In their study, mistuning the odd harmonics from the even harmonics improved DLF0s, even though all the harmonics apparently remained unresolved. Bernstein and Oxenham concluded that resolved harmonics were not necessary for good pitch discrimination, and they were able to simulate their data using a variant of the autocorrelation model (Bernstein and Oxenham, 2005). More work is needed to provide a fully satisfactory explanation for this apparent discrepancy.

2. Influence of relative ear of presentation of the target and masker

Although the elevations in DLF0s produced by ipsilateral maskers likely originate in peripheral interactions between harmonics from the target and masker, it was unclear *a priori* how a masker presented to the opposite ear would affect DLF0s. Although some previous studies have measured thresholds or performance in F0-discrimination or pitch-identification tasks with concurrently presented harmonics, the sets of harmonics that were presented to the left and right ears in these studies had the same F0 (Houtsma and Goldstein, 1972; Bernstein and Oxenham, 2003), or they contained only two components (Beerends and Houtsma, 1989), or they were filtered into different spectral regions (Gockel et al., 2009). The complexes in the present study occupied the same spectral region but had different F0s in the two ears. Our results show that, under such conditions, DLF0s are much less affected by contralateral maskers than by ipsilateral maskers, and that this difference is not due to perceived lateralization differences, but rather to interactions between components in the same ear.

Although this pattern of results is consistent with the hypothesis that the effect of the ipsilateral masker was due to a large extent to peripheral interactions between target and masker components, an influence of more central factors cannot be ruled out, because DLF0s were still elevated compared to the baseline when the masker was presented in the contralateral ear only. This central interference is unlikely to reflect confusion between the pitches of the target and masker, because it was observed even in conditions in which the F0s of the target and masker were 9 semitones apart on average. It could reflect an unavoidable, albeit partial, aggregation of information across the two ears prior to the computation of pitch, or a mechanism comparable to that responsible for pitch discrimination interference (Gockel et al., 2009).

Using different stimuli and a different task than those used here, Beerends and Houtsma (1989) found little influence of the ear of presentation on listeners' ability to recognize the pitches of two simultaneously presented pairs of harmonics. Here, we found that thresholds for the F0 discrimination of a target complex in the presence of a spectrally overlapping complex were significantly smaller when the target and masker were presented to opposite ears. It would be interesting to determine in future studies whether a similar beneficial effect of ear separation between the target

and masker complexes can also be observed in a pitch-recognition task similar to that used by Beerends and Houtsma, but with complex tones such as those used in the current study, which contained more than two harmonics each.

V. CONCLUSIONS

The ability of normal-hearing listeners to discriminate small changes in the F0 of a bandpass-filtered target HCT was measured with and without a simultaneous spectrally overlapping masker HCT. A range of nominal target F0s (100, 200, and 400 Hz) was tested with masker F0s that were either similar to, or 9 semitones below or above, the target F0. The degree to which harmonic frequencies were spectrally resolved was assessed using an EP model (Glasberg and Moore, 1990). For the range of conditions tested here, good F0 discrimination, with DLF0s of less than 2%, was achieved only in conditions that produced severaly salient (>1 dB) peaks in the EP at or near target harmonic frequencies. In many cases the EP peaks reflected a summation of both a masker and target harmonic, so the target harmonics were not resolved in the mixture. Nevertheless, the combined peaks seemed sufficient to produce good F0 discrimination abilities. In cases where no salient peaks remained in the EP representation after the target and masker were mixed, DLF0s were mostly poor and were always greater than a semitone (6%). Thus, based on the present results, it seems that salient spectral peaks may be necessary for pitch perception of one harmonic sound in the presence of another. Further study will be necessary to determine the generality of this conclusion.

ACKNOWLEDGMENTS

[1]The five listeners who took part in the study were initially tested with the same noise routed to each ear. However, the combination of diotic noise and a monaural target allowed the possibility of some binaural masking release, so that the target was presented at a level that was further above its threshold in noise that would have been the case in monaural or uncorrelated noise. To check whether the use of diotic noise influenced the results, three of the five listeners were re-tested using uncorrelated noise at the two ears. No differences in the pattern of results between the two modes of noise presentation were observed, as confirmed by analyses of variance. Accordingly, the data from those listeners who were tested with both diotic and dichotic noise were pooled together.

[2]Bernstein and Oxenham (2006) estimated that the threshold target-to-valley ratio for the lowest-frequency EP peak corresponding to a resolved harmonic in their stimuli was about 2 dB. However, these authors used a different auditory-filter model for their EP simulations than the *roex* function used in Glasberg and Moore (1990) and in the present study.

[3]Consistent with the use of a logarithmic scale in the adaptive procedure as well as in the plots, all statistical tests and error bars were calculated using log-transformed DLF0s.

Beerends, J. G. (**1989**). "The influence of duration on the perception of pitch in single and simultaneous complex tones," J. Acoust. Soc. Am. **86**, 1835–1844.

Beerends, J. G., and Houtsma, A. J. M. (**1986**). "Pitch identification of simultaneous dichotic two-tone complexes," J. Acoust. Soc. Am. **80**, 1048–1055.

Beerends, J. G., and Houtsma, A. J. M. (**1989**). "Pitch identification of simultaneous diotic and dichotic two-tone complexes," J. Acoust. Soc. Am. **85**, 813–819.

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., 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.

Bernstein, J. G., and Oxenham, A. J. (**2006**). "The relationship between frequency selectivity and pitch discrimination: Effects of stimulus level," J. Acoust. Soc. Am. **120**, 3916–3928.

Bernstein, J. G., and Oxenham, A. J. (**2008**). "Harmonic segregation through mistuning can improve fundamental frequency discrimination," J. Acoust. Soc. Am. **124**, 1653–1667.

Bird, J., and Darwin, C. J. (**1998**). "Effects of a difference in fundamental frequency in separating two sentences," in *Psychophysical and Physiological Advances in Hearing*, edited by A. R. Palmer, A. Rees, A. Q. Summerfield, and R. Meddis (Whurr, London), pp. 263–269.

Brokx, J. P. L., and Nooteboom, S. G. (**1982**). "Intonation and the perceptual separation of simultaneous voices," J. Phonetics **10**, 23–36.

Butler, D. (**1979**). "A further study of melodic channeling," Percept. Psychophys. **25**, 264–268.

Cariani, P. A. (**2001**). "Neural timing nets," IEEE Trans. Neural Netw. **14**, 737–753.

Carlyon, R. P. (**1996a**). "Encoding the fundamental frequency of a complex tone in the presence of a spectrally overlapping masker," J. Acoust. Soc. Am. **99**, 517–524.

Carlyon, R. P. (**1996b**). "Masker asynchrony impairs the fundamental-frequency discrimination of unresolved harmonics," J. Acoust. Soc. Am. **99**, 525–533.

Carlyon, R. P. (**1997**). "The effect of two temporal cues on pitch judgments," J. Acoust. Soc. Am. **102**, 1097–1105.

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.

Darwin, C. J., Brungart, D. S., and Simpson, B. D. (**2003**). "Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers," J. Acoust. Soc. Am. **114**, 2913–2922.

de Cheveigné, A. (**1993**). "Separation of concurrent harmonic sounds: Fundamental frequency estimation and a time-domain cancellation model of auditory processing," J. Acoust. Soc. Am. **93**, 3271–3290.

de Cheveigné, A. (**2005**). "Pitch perception models," in *Pitch. Neural Coding and Perception*, edited by C. J. Plack, A. J. Oxenham, R. Fay, and A. N. Popper (Springer, New York), pp. 169–233.

de Cheveigné, A. (**2006**). "Multiple F0 estimation," in *Computational Auditory Scene Analysis. Principles, Algorithms, and Applications*, edited by D. Wang and G. J. Brown (Wiley, Hoboken, New Jersey), pp. 45–80.

Deutsch, D. (**1979**). "Binaural integration of melodic patterns," Percept. Psychophys. **25**, 399–405.

Duifhuis, H., Willems, L. F., and Sluyter, R. J. (**1982**). "Measurement of pitch in speech: An implementation of Goldstein's theory of pitch perception," J. Acoust. Soc. Am. **71**, 1568–1580.

Gerson, A., and Goldstein, J. L. (**1978**). "Evidence for a general template in central optimal processing for pitch of complex tones," J. Acoust. Soc. Am. **63**, 498–510.

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.

Gockel, H. E., Hafter, E. R., and Moore, B. C. J. (**2009**). "Pitch discrimination interference: The role of ear of entry and of octave similarity," J. Acoust. Soc. Am. **125**, 324–327.

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.

Houtsma, A. J. M., and Goldstein, J. L. (**1972**). "The central origin of the pitch of complex 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.

Larsen, E., Cedolin, L., and Delgutte, B. (**2008**). "Pitch representations in the auditory nerve: Two concurrent complex tones," J. Neurophysiol. **100**, 1301–1319.

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

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

Micheyl, C., Bernstein, J. G., and Oxenham, A. J. (**2006**). "Detection and F0 discrimination of harmonic complex tones in the presence of competing tones or noise," J. Acoust. Soc. Am. **120**, 1493–1505.

Micheyl, C., and Oxenham, A. J. (**2004**). "Sequential F0 comparisons between resolved and unresolved harmonics: No evidence for across-pitch-mechanisms translation noise," J. Acoust. Soc. Am. **116**, 3038–3050.

Micheyl, C. and Oxenham, A. J. (**2009**). "Pitch, harmonicity, and concurrent sound segregation: Psychoacoustical and neurophysiological findings," Hear. Res. (in press).

Moore, B. C. J., and Carlyon, R. P. (**2005**). "Perception of pitch by people with cochlear hearing loss and by cochlear implant users," in *Pitch: Neural Coding and Perception*, edited by C. J. Plack, A. J. Oxenham, R. Fay, and A. N. Popper (Springer, New York).

Moore, B. C. J., Glasberg, B. R., Low, K.-E., Cope, T., and Cope, W. (**2006**). "Effects of level and frequency on the audibility of partials in inharmonic complex tones," J. Acoust. Soc. Am. **120**, 934–944.

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

Oxenham, A. J. (**2008**). "Pitch perception and auditory stream segregation: Implications for hearing loss and cochlear implants," Trends Amplif. **12**, 316–331.

Oxenham, A. J., and Simonson, A. M. (**2009**). "Masking release for low- and high-pass filtered speech in the presence of noise and single-talker interference," J. Acoust. Soc. Am. **125**, 457–468.

Parsons, T. (**1976**). "Separation of speech from interfering speech by means of harmonic selection," J. Acoust. Soc. Am. **60**, 911–918.

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

Plomp, R. (**1976**). *Aspects of Tone Sensation* (Academic, London).

Scheffers, M. T. M. (**1983**). "Sifting vowels: Auditory pitch analysis and sound segregation," Ph.D. thesis, Groningen University, The Netherlands.

Schroeder, M. R. (**1968**). "Period histogram and product spectrum: New methods for fundamental-frequency measurement," J. Acoust. Soc. Am. **43**, 829–834.

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.

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.

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

Terhardt, E. (**1970**). "Frequency analysis and periodicity detection in the sensations of roughness and periodicity pitch," in *Frequency Analysis and Periodicity Detection in Hearing*, edited by R. Plomp and G. F. Smoorenbug (Sijthoff, Leiden, The Netherlands).

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

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

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," Acustica **2**, 125–133.