Pitch strength decreases as F0 and harmonic resolution increase in complex tones composed exclusively of high harmonics[a)]

D. Timothy Ives[b)] and Roy D. Patterson
Centre for the Neural Basis of Hearing, Department of Physiology, Development and Neuroscience, University of Cambridge, Downing Street, Cambridge, CB2 3EG, United Kingdom

(Received 15 January 2007; revised 4 February 2008; accepted 7 February 2008)

A melodic pitch experiment was performed to demonstrate the importance of time-interval resolution for pitch strength. The experiments show that notes with a low fundamental (75 Hz) and relatively few resolved harmonics support better performance than comparable notes with a higher fundamental (300 Hz) and more resolved harmonics. Two four note melodies were presented to listeners and one note in the second melody was changed by one or two semitones. Listeners were required to identify the note that changed. There were three orthogonal stimulus dimensions: F0 (75 and 300 Hz); lowest frequency component (3, 7, 11, or 15); and number of harmonics (4 or 8). Performance decreased as the frequency of the lowest component increased for both F0's, but performance was better for the *lower* F0. The spectral and temporal information in the stimuli were compared using a time-domain model of auditory perception. It is argued that the distribution of time intervals in the auditory nerve can explain the *decrease* in performance as F0, and spectral resolution *increase*. Excitation patterns based on the same time-interval information do not contain sufficient resolution to explain listener's performance on the melody task.
 [DOI: 10.1121/1.2890737]

PACS number(s): 43.66.Ba, 43.66.Hg, 43.66.Lj [RAL] Pages: 1–XXXX

I. INTRODUCTION

A series of experiments with filtered click trains and harmonic complexes has shown that pitch strength decreases as the lowest harmonic of a complex increases. The phenomenon has been demonstrated for the lowest harmonics using magnitude estimation (Fastl and Stoll, 1979; Fruhmann and Kluiber, 2005), and for higher harmonics using a variety of pitch discrimination tasks [e.g., Ritsma and Hoekstra, 1974; Cullen and Long, 1986; Houtsma and Smurzynski, 1990; see Krumbholz *et al.* (2000) for a review]. This paper reports an experiment that makes use of this phenomenon to demonstrate the importance of time-interval resolution for pitch strength. A harmonic complex with eight, adjacent components was used to measure performance on a melodic pitch task (Patterson *et al.*, 1983; Pressnitzer and Patterson, 2001), as a function of the frequency of the lowest harmonic in the complex. The important variable was the fundamental (F0) of the complex which was either low (75 Hz) or high (300 Hz), and the main empirical question was which fundamental supports better performance on the melodic pitch task?

In time-domain models of peripheral processing, the reduction in pitch strength with increasing harmonic number is associated with the loss of phase locking at high frequencies (e.g., Patterson *et al.*, 2000; Krumbholz *et al.*, 2000; Pressnitzer *et al.*, 2001). As a result, time-domain models predict that performance based on complexes limited to high harmonics will be worse for the *higher* fundamental (300 Hz); the higher harmonics occur above 3000 Hz for the 300 Hz fundamental, where the internal representation of the time interval information is smeared by the loss of phase locking.

In spectral models of pitch perception, the reduction in pitch strength with increasing harmonic number is associated with the loss of harmonic resolution at high harmonic numbers. This occurs because the frequency spacing between components of a harmonic complex is fixed, whereas the bandwidth of the auditory filter increases with filter center frequency. Thus, for all fundamentals, harmonic resolution (harmonic-spacing/center-frequency) decreases as harmonic number increases. It is also the case that the frequency resolution of the auditory filter improves somewhat with filter center frequency, where filter resolution is defined as the ratio of the center frequency (f_c) to the bandwidth (bw); it is referred to as the quality (Q) of the filter ($Q=f_c/\mathrm{bw}$). As a result, spectral models, which ignore the effects of phase locking, predict that performance will be worse for the *lower* fundamental with the lower value of Q.

The results of the experiment show that performance on the melodic pitch task is worse for the higher fundamental in support of the view that it is time-interval resolution rather than harmonic resolution that imposes the limit on pitch strength for these harmonic complexes.

Spectral and temporal summaries of the pitch information in complex sounds. The logic of the experiment will be illustrated using a time-domain model of auditory processing, since such models make it possible to compare the spec-

[a)]Portions of this work were presented in "Why pitch strength decreases with increasing harmonic number in complex tones" at the 153rd Meeting of the Acoustical Society of America, Salt Lake City, 2007.
[b)]Electronic mail: dti20@cam.ac.uk

tral and temporal information that is assumed to exist in the auditory system at the level of the auditory nerve. There are a number of different time-domain models which are typically referred to by the representation of sound that they produce, for example, the "correlogram" (Slaney and Lyon, 1990), the "autocorrelogram" (Meddis and Hewitt, 1991), and the "auditory image" (Patterson *et al.*, 1992, 1995). The example is based on the auditory image model (AIM) and the specific implementation is that described in Bleeck *et al.* (2004). The first three stages of AIM are typical of most time-domain models of auditory processing. A bandpass filter simulates the operation of the outer and middle ears, and then an auditory filterbank simulates the spectral analysis performed in the cochlea by the basilar partition. The shape of the auditory filter is typically derived from simultaneous noise-masking experiments, rather than pitch experiments. In this case, it is the gammatone auditory filterbank of Patterson *et al.* (1995). The simulated membrane motion is converted into a simulation of the phase-locked, neural activity pattern (NAP) that flows from the cochlea in response to the sound; the simulated NAP represents the probability of neural firing, it is produced by compressing, half-wave rectifying and low-pass filtering the membrane motion, separately in each filter channel. The NAPs produced by AIM are very similar to those produced by correlogram and autocorrelogram models of pitch perception (e.g., Slaney and Lyon, 1990; Meddis and Hewitt, 1991; Yost *et al.*, 1996).

The NAPs produced in response to two complex sounds composed of harmonics 3–10 of a 300 Hz fundamental and a 75 Hz fundamental are shown in Figs. 1(a) and 1(b), respectively. The dimensions of the NAP are time (the abscissa) and auditory-filter center frequency on a quasilogarithmic axis (the ordinate). Figure 1 covers the frequency range from 50 to 12 000 Hz. The time range encompasses three periods of the corresponding fundamental; so for the 300 Hz F0 the range is 10 ms, and for the 75 Hz F0 the range is 40 ms. The vertical and horizontal side panels to the right and below each figure show the average of the activity in the NAP across one of the dimensions. The average over time is shown in the vertical (or spectral) profile; the average over frequency is shown in the horizontal (or temporal) profile. The spectral profiles are often referred to as excitation patterns (e.g., Glasberg and Moore, 1990), and they show that there are more resolved harmonics in the NAP of the sound with the higher F0 (300 Hz) [Fig. 1(a)] than for the lower F0 (75 Hz) [Fig. 1(b)]. This suggests that using the spectral profiles to predict pitch strength would lead to a higher value of pitch strength for the higher F0. The spectral summaries derived from other time-domain models and the spectral summaries used in spectral models of auditory processing would all lead to the same, qualitative, prediction. With regard to temporal information, the NAPs reveal faint ridges in the activity, which occur every 3.3 ms for the 300 Hz NAP and every 13.3 ms for the 75 Hz NAP. However, it is difficult to see the strength of the temporal regularity in the NAP because the propagation delay in the cochlea means that the temporal pattern in the lower channels is progressively shifted in time. Similarly, the temporal profiles provide only a poor representation of the temporal regularity in these sounds. This is a general limitation of time-frequency representations of the information in the auditory nerve.

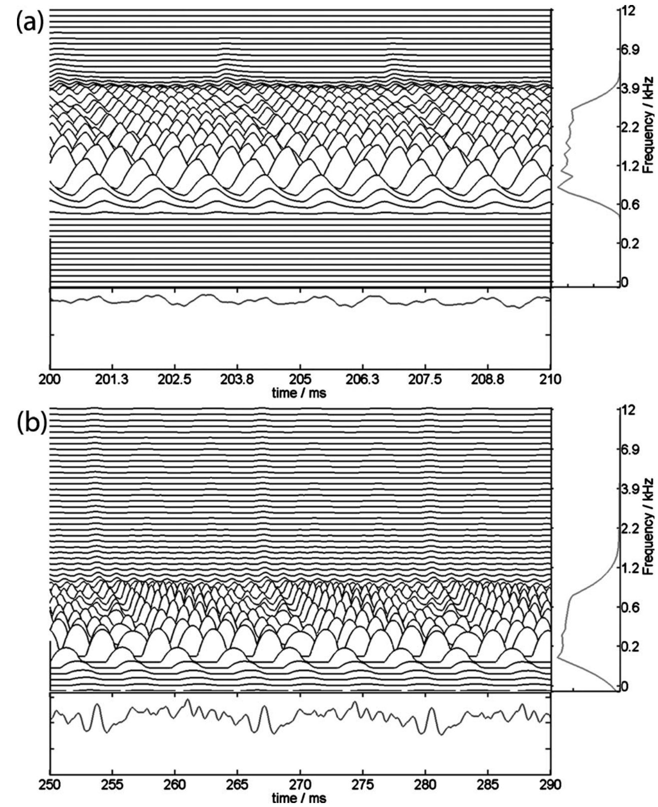

FIG. 1. Neural activity patterns (NAPs) for harmonic complex sounds composed of the third to tenth harmonics of (a) an F0 of 300 Hz, and (b) an F0 of 75 Hz. Side panels show the spectral profiles (vertical) and temporal profiles (horizontal) of the NAP.

The temporal information in the NAP concerning how the sound will be perceived is not coded by time, *per se*, but rather by the time intervals between the peaks of the membrane motion. For this reason, time-domain models include an extra stage, in which autocorrelation (e.g., Slaney and Lyon, 1990) or strobed temporal integration (Patterson *et al.*, 1992) is applied to the NAP to extract and stabilize the phase-locked, repeating neural patterns produced by periodic sounds. Broadly speaking, the time intervals between peaks within a channel are calculated and used to construct a form of time-interval histogram for that channel of the filterbank, and the complete array of time-interval histograms is the correlogram (Slaney and Lyon, 1990), or auditory image (Patterson *et al.*, 1992), of the sound. The histogram is dynamic and events emerge in, and decay from, the histogram with a half life on the order of 30 ms. It is argued that these representations provide a better description of what will be heard than the NAP. They have the stability of auditory perception (Patterson *et al.*, 1992) and they do not contain the between-channel phase information associated with the propagation delay which we do not hear (Patterson, 1987). However, all that matters in the current study is that they reveal the precision of the time-interval information in the auditory nerve and make it possible to produce a simple summary of the temporal information in the form of a temporal profile.

PROOF COPY 002805JAS

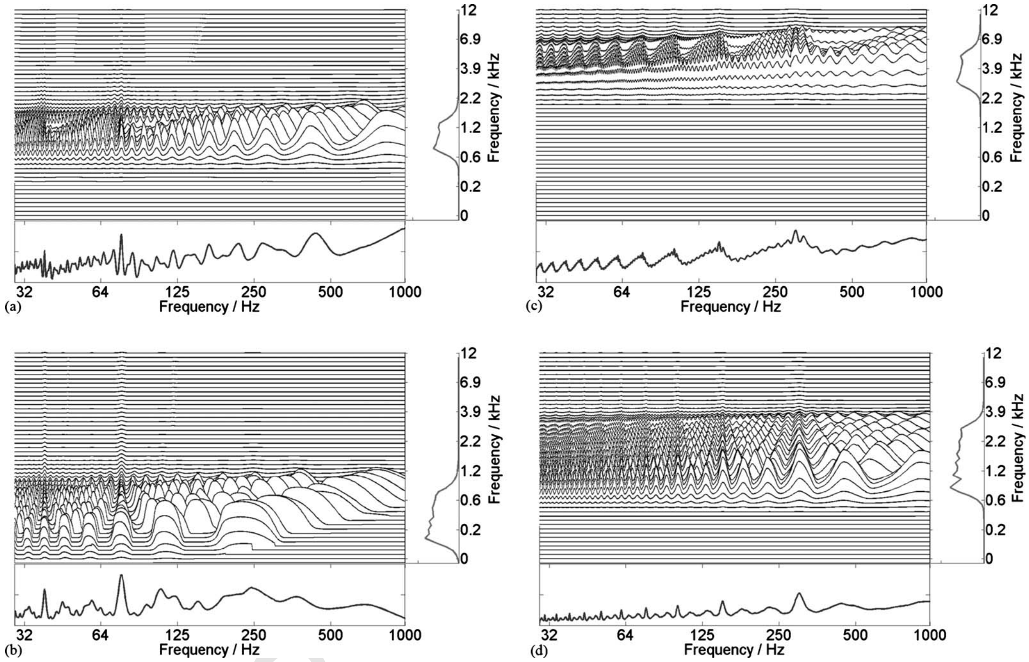

FIG. 2. Stabilized auditory images (SAI) of four harmonic complexes. All stimuli have eight consecutive harmonics; they differ in their fundamentals and lowest components. (a) F0=75 Hz, harmonics 11–18; (b) F0=75 Hz, harmonics 3–10; (c) F0=300 Hz, harmonics 11–18; (d) F0=300 Hz, harmonics 3–10. Side panels show the spectral profiles (vertical) and temporal profiles (horizontal) of the auditory images.

The auditory images of four harmonic complexes simulated by AIM are shown in Fig. 2. The stimuli all have eight consecutive harmonics but they differ in fundamental (F0) and/or lowest component (LC) as follows: (a) F0=75 Hz, LC=11; (b) F0=75 Hz, LC=3; (c) F0=300 Hz, LC=11; (b) F0=300 Hz, LC=3. The auditory images corresponding to the NAPs in Figs. 1(a) and 1(b) are shown in Figs. 2(d) and 2(b), respectively. In each channel of all four panels, there is a local maximum at the F0 of the stimulus, and together these peaks produce a vertical ridge in each panel that corresponds to the pitch that the listener hears. In the upper panels [Figs. 2(a) and 2(b)], where the lowest component is the 11th, and the auditory filters are wide relative to component density, the interaction of the components within a filter is clearly manifested by the asymmetric modulation of the pattern at the F0 rate. The corresponding correlograms of Slaney and Lyon (1990) and the autocorrelograms of Meddis and Hewitt (1991) would have a similar form in as much as there would be local peaks at F0 and prominent modulation for the stimuli where the lowest component is the 11th, but the pattern of activity within the period of the sound would be blurred and the envelope of the modulation would be more symmetric.

The vertical and horizontal side panels to the right and below each sub-figure show the average of the activity in the auditory image across one of the dimensions. The average over time interval is shown in the vertical, or spectral, profile; the average over frequency (or channels) is shown in the horizontal, or temporal, profile. The unit on the time-interval axis is the frequency equivalent of time interval, that is, time interval^{-1}. It is used to make the spectral and temporal profiles directly comparable. The spectral profile of the auditory image is very similar to that of the corresponding NAP. The temporal profile of the auditory image shows that the timing information in the neural pattern of these stimuli is very regular, and if the auditory system has access to this information it could be used to explain pitch perception. The advantage of time-domain models of auditory processing is that the spectral and temporal profiles are derived from a common simulation of the information in the auditory nerve, which facilitates comparison of the spectral and temporal pitch models based on such profiles. Moreover, the parameters of the filterbank are derived from separate, masking experiments, so the resulting models have the potential to explain pitch and masking within a unified framework.

In the spectral profile, when the lowest component is increased from three to eleven, the profile ceases to resolve individual components. This is shown by comparing the peaky spectral profile for the stimulus with a LC of 3 in Fig. 2(d), with the smoother profile for the stimulus with a LC of 11 in Fig. 2(c). The effect of increasing LC is similar for the lower F0 in the left column, but the harmonic resolution is reduced in both cases. In the temporal profile, when the lowest component is increased from three to eleven, the pronounced peak at 75 Hz in the left-hand column remains; compare Figs. 2(b) and 2(a). The 300 Hz peak in the temporal profile in the right-hand column becomes much less pronounced relative to the surrounding activity, [compare Figs. 2(d) and 2(c)] but there is still a small peak in Fig. 2(c).

As F0 is increased from 75 to 300 Hz, activity in the spectral profile shifts up along the frequency axis. For the stimuli with higher order components [Figs. 2(a) and 2(c)], there is little change in the resolution of the spectral profile when F0 is changed; the harmonic resolution remains poor. But for the stimuli with lower order components [Figs. 2(b) and 2(d)], the increase from a fundamental of 75 Hz to one of 300 Hz is accompanied by an increase in harmonic resolution, which is due to the increase in the Q of the filter with center frequency. As a result, a model based on spectral profiles would predict (a) that performance for stimuli with higher order components will be poor independent of F0, and (b) that performance for stimuli with lower order components will be better for the higher F0 (300 Hz).

As F0 is increased from 75 to 300 Hz, the peak in the temporal profile shifts to the right along the time-interval axis. For stimuli with lower order components [Figs. 2(b) and 2(d)], the ratio of the magnitude of the F0 peak to the magnitude of the neighboring trough is large, and a model based on temporal profiles would predict good performance in both conditions. For stimuli with higher order components [Figs. 2(a) and 2(c)], the peak to trough ratio is still reasonably large for the lower F0, but it is much reduced for the higher F0. So, a model based on temporal profiles would predict reasonable performance for the low F0 and poorer performance for the higher F0. Thus, there is a clear difference between the predictions of the two classes of model.

II. MAIN EXPERIMENT

The melody task is based on the procedure described previously by Patterson *et al.* (1983) and revived by Pressnitzer *et al.* (2001). Listeners were presented with two successive melodies. The second melody was a repetition of the first but had one of the notes changed by one diatonic interval up or down. The task for the listener was to identify which note had changed in the second melody. Melodies consisted of four notes from the diatonic major scale. The structure of the notes was such that only the residue pitch was consistent with the musical scale, and that sinusoidal pitch could not be used to make judgments. A melody task was used rather than a pitch discrimination task as it is a better measure of pitch strength.

A. Stimuli

The notes in the melodies were synthesized from a harmonic series whose lowest components were missing. The pitch of the note corresponded to the F0 of the harmonic series. The harmonics were attenuated by a low-pass filter with a slope of 6 dB/octave relative to the lowest component present in the complex. Performance on a melody task was measured as a function of three parameters: fundamental frequency (F0); average, lowest component number (ALC); and number of components (NC). There were two nominal F0's 75 and 300 Hz; the F0 was subject to a rove of half an octave. The ALC was 3, 7, 11, or 15. The NC was either 4 or 8. Stimuli were generated using MATLAB; they had a sampling rate of 48 kHz and 16 bit amplitude resolution. They were played using an Audigy-2 soundcard. The duration of each note was 500 ms, which included a 100 ms raised cosine onset and a 333 ms raised cosine offset. Stimuli were presented diotically using AKG K240DF Studio-Monitor headphones at a level of approximately 60 dB SPL. Difference tones in the region of F0 and its immediate harmonics (Pressnitzer and Patterson, 2001) were masked by bandpass filtered white noise; the frequency range was 20–160 Hz for

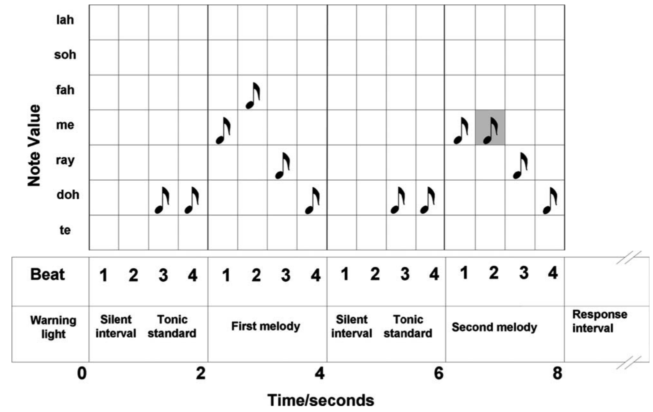

FIG. 3. Schematic of the procedure of the melody task, adapted from Patterson *et al.* (1983). One note changes by a single diatonic interval between the first and second presentations of the melody, and the listener has to identify the changed note, marked here by a grey square.

the lower F0 and 50–400 Hz for the higher F0. The level of the noise was 50 dB SPL. Cubic difference tones just below the lowest harmonic were not masked as this would involve inserting a loud noise that would overlap in the spectrum with the stimulus. Cubic difference tones might increase pitch strength slightly in all conditions, but they would not be expected to contribute a distinctive cue to the melody that would affect performance differentially for a particular F0 or lowest harmonic number. The experiment was run in an IAC double-walled, sound-isolated booth.

B. Subjects

Three listeners participated in the first experiment; their ages ranged from 20 to 26 years. All listeners had normal hearing thresholds at 500 Hz, 1, 2, and 4 kHz. Listeners were not chosen on the basis of musical ability, but two of the listeners were trained musicians. All listeners were paid at an hourly rate. Listeners were trained on the melody task over a 2 h period, although they would be allowed to take frequent breaks so the actual training time was somewhat less than 2 h. The training program varied between listeners. Typically it involved starting with an easy condition having eight components, an ALC of three, and no roving of the lowest component. The difficulty of the task was then increased by including stimuli with fewer components (i.e., four), adding the rove, and finally presented stimuli with higher values of ALC. Three potential listeners were rejected after the training period because they were unable to learn the task sufficiently well within the allotted time.

C. Procedure

Listeners were presented with two consecutive four note melodies. The second melody had one of the notes changed, and listeners had to identify the interval with the changed note. The procedure is illustrated schematically in Fig. 3 as four bars of music: The two melodies are presented in the second and fourth bars; the tonic, which defines the scale for the trial, is presented twice before each of the melodies, as a pick up in the third and fourth beats of the first and third bars. After the presentation of the second melody, there was an indefinite response interval, which was terminated by the listener's response. Feedback was then given as to which note actually changed, then another trial begun. In the example shown in Fig. 3, it is the second note that has changed in the second melody as shown by the gray square.

The notes of the melodies were harmonic complexes without their lowest components. The melody was defined as the sequence of fundamentals (that is, the residue pitch) rather than the sequence of intervals associated with any of the component sinusoids. On each trial, the F0 of the tonic was randomly selected from a half-octave range, centered logarithmically on F0. The actual ranges were 63–89 and 252–357 Hz. The F0's of the other notes in the scale were calculated relative to the F0 of the tonic using the following frequency ratios: $2^{-1/12}$ (te); 1 (doh); $2^{2/12}$(ray); $2^{4/12}$(me); $2^{5/12}$(fah); $2^{7/12}$ (soh); and $2^{9/12}$(lah). Note that a ratio of $2^{1/12}$ produces an increase in frequency of one semitone on the equal temperament scale. The intervals are musical but, due to the randomizing of the F0, the notes of the melodies are only rarely the notes found on the A440 keyboard. The purpose of randomizing the F0 of the tonic was to force the listeners to using musical intervals rather than absolute frequencies to perform the task. The notes of the first melody of a trial were drawn randomly, with replacement, from the first five notes of the diatonic scale based on the randomly chosen tonic for that trial. The melody was repeated in the same key, and one of the notes was shifted up or down by a single diatonic interval. This shift can result in either a tone or a semitone change, since the size of a diatonic interval depends on its position in the scale.

The LC of each note in each melody was subjected to a restricted rove, the purpose of which was to preclude the use of the sinusoidal pitch of one of the components to perform the task. The degree of rove was one component, and so, the LC in each tone was either LC or LC+1. There were two further restrictions on the value of the LC: First, adjacent notes in a melody were precluded from having the same LC; second, each note had a different LC in the second melody from that which it had in the first melody. With these restrictions, it sufficed to alternate between the LC and the one above it using one of the patterns 1–0–1–0 or 0–1–0–1 for the first melody and the other pattern for the second melody.

The note-synthesis parameters were combined to produce 16 conditions ($2\times F0, 2\times NC, 4\times ALC$). The order of these 16 conditions was randomized, and together they constituted one replication of the experiment. The listeners performed three or four replications in a 20 min block, with four or five blocks in a 2 h session. All listeners completed 45 or 46 replications.

D. Results of main experiment

AQ: #2

The average results for the three listeners are shown in Fig. 4; the pattern of results was the same for all three listeners as shown by the analysis of variance (ANOVA) in Table I. The abscissa shows the ALC of the harmonic series; the ordinate shows the probability of the listener correctly identifying which of the notes changed in the second melody. Performance is plotted separately for the two NCs and the two F0's. The black and grey lines show the results for the 75 and 300 Hz F0's, respectively. The solid and dashed lines show the results for the four- and eight-harmonic stimuli, respectively. Figure 4 shows that, as ALC is increased, performance decreases, i.e., the probability of identifying which note changed in the second melody decreases in all conditions. However, the effect is much more marked for the 300 Hz F0, where performance decreases abruptly as ALC increases beyond 7. This is the most important result, as it differentiates the spectral and temporal models: Strictly spectral models would predict that there should be no reduction in listener performance when F0 is increased; indeed, performance should improve slightly with increasing F0 because the auditory filter becomes relatively narrower at higher center frequencies. Temporal models predict that there will be a decrease in performance with increasing F0 because of the progressive reduction in the phase locking of nerve fibers. The effect of increasing NC from four to eight had no consistent effect on listener performance.

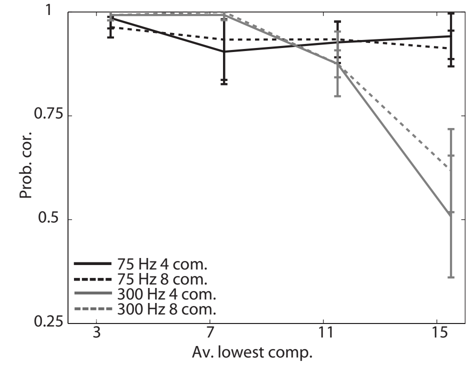

FIG. 4. Performance on the melody task with the 75 and 300 Hz fundamentals. The abscissa shows the average lowest component and the ordinate shows the probability of the listener correctly identifying the note which changed. Performance is plotted for each NC condition as a function of average lowest component. The black and grey lines show the results for the 75 and 300 Hz F0's, respectively. The solid and dashed lines show the results for the four- and eight-harmonic stimuli, respectively.

An ANOVA was performed on the data; the results are presented in Table I, which confirms that the above-described effects are statistically significant at the $P<0.01$ level (bold type in Table I). There is a main effect of ALC, and one interaction, $FO\times ALC$. The interaction of F0 with ALC shows that ALC has a greater effect on performance for the higher F0.

III. ANCILLARY EXPERIMENTS

Prior to running the main experiment, two similar ancillary experiments were performed. They are presented briefly here inasmuch as they provide additional data concerning the effects observed in the main experiment, and they provide data on the effects of a larger component rove.

A. Method

The experimental task and the procedure were the same as those described for the main experiment in Sec. II. The design was slightly different. The F0 was 300 Hz in the first ancillary experiment and 75 Hz in the second. The ALC val-

TABLE I. Results of an ANOVA of performance data (Dependent variable: SCORE). There is one significant ($P<0.01$) main effect and one significant interaction, both of which are shown in bold type; they are ALC, and F0 $\times$ ALC.

Source	Type III sum of squares	df	Mean square	F	Sig.	Partial eta squared
F0	0.079	1	0.079	4.011	0.183	0.667
ALC	**0.412**	**3**	**0.137**	**18.477**	**0.002**	**0.902**
NC	0.002	1	0.002	3.728	0.193	0.651
SUB	0.012	2	0.006	0.273	0.782	0.186
F0*ALC	**0.347**	**3**	**0.116**	**27.632**	**0.001**	**0.933**
F0*NC	0.003	1	0.003	2.222	0.275	0.526
F0*SUB	0.039	2	0.020	3.834	0.080	0.545
ALC*NC	0.004	3	0.001	5.355	0.039	0.728
ALC*SUB	0.045	6	0.007	1.910	0.245	0.693
NC*SUB	0.001	2	0.001	0.443	0.717	0.421
F0*ALC*NC	0.012	3	0.004	7.181	0.021	0.782
F0*ALC*SUB	0.025	6	0.004	7.440	0.014	0.882
F0*NC*SUB	0.003	2	0.001	2.648	0.150	0.469
ALC*NC*SUB	0.002	6	0.000	0.480	0.803	0.324

ues were the same in the two ancillary experiments and the values were the same as in the main experiment, namely, 3, 7, 11, or 15. The number of components was 4 or 8, as in the main experiment; however, the ancillary experiments also included a condition with just two components. In the ancillary experiments, the lowest-component rove (LCR) was either one component (as in the main experiment) or three components. The LCR was subject to the same restrictions as to those in the main experiment. Specifically, for a rove of three, a random permutation of the four rove values was calculated for the first melody (e.g., 0–2–3–1) and recalculated for the second melody such that none of the notes in the second melody had the same lowest component as the corresponding note in the first melody (e.g., 1–0–2–3). Four listeners participated in each of the ancillary experiments, and three of the listeners were the same in the two experiments.

In the conditions where there were only two components in the sound, the pitch is ambiguous and the form of the ambiguity differs between musical and nonmusical listeners (Seither-Preisler *et al.*, 2007). The problem is that the sinusoidal pitches of the individual components are strong relative to the residue pitch produced by two components; this, in turn, makes it difficult for nonmusical listeners to focus on the residue pitch and not be distracted by the sinusoidal pitches. These problems reduced performance in the two-component conditions; the reduction was larger for the lower F0, and larger for the less musical listeners, but there was not enough data to quantify the interaction of F0 and listener. While it might be interesting to study how the pitch of the residue builds up with number of components, while the sinusoidal pitches of the individual components become less salient, that was not the purpose of these experiments. Consequently, the two-component condition was dropped from the design of the main experiment, and the two-component results from the ancillary experiments are omitted from further discussion.

B. Results

The remaining results of the two ancillary experiments are plotted together in Fig. 5; the pattern of results was the same for the four listeners in each of the experiments, so the figure shows performance averaged across listeners. The abscissa shows the ALC of the harmonic series; the ordinate shows the probability of the listener correctly identifying which of the notes changed in the second melody, as before. Performance is plotted separately for the two LCRs and the two F0's. Performance was averaged over number of components (four and eight) because the variable did not affect performance; the same noneffect was later observed in the

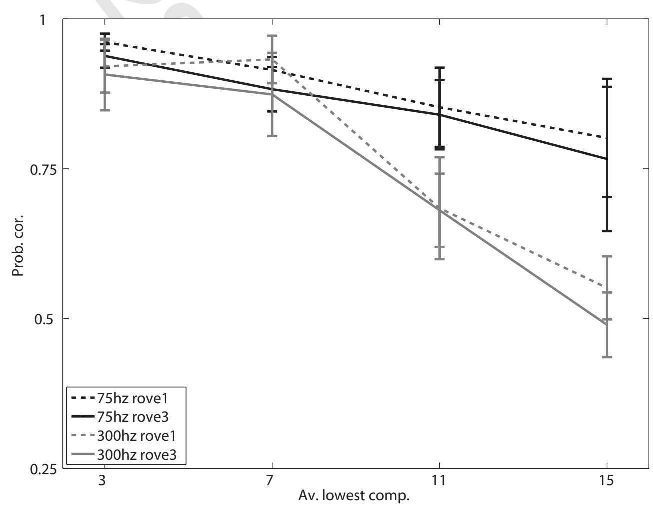

FIG. 5. Performance on the melody task in the ancillary experiments with the 75 Hz fundamental (black lines) and the 300 Hz fundamental (grey lines). The abscissa shows the average lowest component and the ordinate shows the probability of the listener correctly identifying the note which changed. Performance is plotted separately for the two rove conditions. The dashed and solid lines show the results for LCR values of one and three, respectively.

main experiment. The black and grey lines show the results for the 75 and 300 Hz F0's, respectively. The dashed and solid lines show the results for LCR values of one and three, respectively.

Consider first, the effect of roving the lowest component; compare the solid lines for a rove of one component with the dashed lines for the rove of three components. Although performance is slightly better for the rove of one, the pattern of results is the same, and the effect of rove magnitude is not significant. With this observation in mind, the results in Fig. 5 are seen to support the conclusions of the main experiment. The overall performance in the ancillary experiments is slightly lower overall, perhaps because two of the three listeners in the main experiment were trained musicians. However, the pattern of results is the same; whereas, performance decreases only slowly with increasing ALC when the F0 is 75 Hz, it decreases rapidly with ALC in the region above seven for an F0 of 300 Hz.

The comparison of performance for the two F0's must be made with some caution in this case, since three of the four listeners were common to the two ancillary experiments, and these three listeners performed the 300 Hz experiment before the 75 Hz experiment. However, there were more than 40 replications of all conditions for each listener in each ancillary experiment, *after* the initial training in the melody task, and an analysis showed that there was essentially no learning over the 40 replications in either experiment. It is also the case that the one listener who only participated in the 75 Hz experiment showed no learning over the course of the experiment, and had the same average level of performance as the other listeners in that experiment, indicating that training on the higher F0 was not required to produce good performance with the lower F0. Thus, it seems likely that the elevation of performance in the 75 Hz experiment for the higher ALC values is not simply due to learning, and probably represents the same effect as observed in the main experiment. Accordingly, in Sec. IV the data from the main and ancillary experiments are combined, so that the performance of the trained musicians is moderated by that of the rest of the listeners to provide the best estimate of what performance would be in the population.

IV. MODELING PITCH STRENGTH WITH DUAL PROFILES

The spectral and temporal profiles of the auditory image both describe aspects of the frequency information in a sound. They can be combined into a dual profile that facilitates comparison of the two kinds of frequency information by inverting the time-interval dimension of the temporal profile (Bleeck and Patterson, 2002). The dual profile for a typical stimulus in the current experiment is shown in Fig. 6. It had the following parameters: NC=4; ALC=3; and F0 =300 Hz. The temporal profile is the blue (darker gray) line with its maximum at 300 Hz; and the spectral profile is the red (lighter gray) line with its maximum at 900 Hz. The peak in the temporal profile at 300 Hz is the F0 of the harmonic series; the position of the peak is independent of the experimental parameters NC and ALC. Should the auditory system have a representation like the temporal profile, it would provide a consistent cue to the temporal pitch of these sounds. The spectral profile has four peaks at 900, 1200, 1500, and 1800 Hz. These peaks are at the four components of the signal, i.e., the third, fourth, fifth, and sixth harmonics of 300 Hz. The spectral profile shows that these four components are resolved, which means that a spectral model would be able to extract the F0 from the component spacing of this stimulus using a more central mechanism that computes subharmonics from a set of spectral peaks. As ALC increases, component resolution decreases and pitch strength decreases. In the following, we use the dual profile to assess the relative value of these spectral and temporal summaries of the sound as predictors of the data from the current experiment.

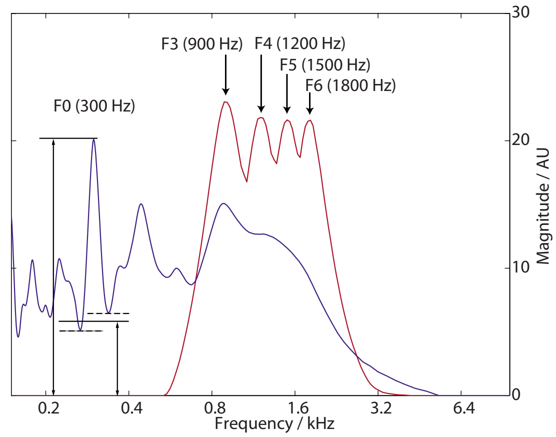

FIG. 6. The dual profile for a stimulus with four resolved harmonics: NC =4, ALC=3, and F0=300 Hz. The temporal profile is the blue (dark) line and the spectral profile is the red (light gray) line. The F0 is represented in the temporal profile by the locating of the largest peak. In the spectral profile the F0 is represented by the spacing of the peaks.

A. The gammatone auditory filterbank

The dual profile shown in Fig. 6 was produced using a gammatone auditory filterbank (GT-AFB) (Patterson *et al.*, 1995) and the version of AIM described in Bleeck *et al.* (2004). The GT-AFB provides a linear simulation of the spectral analysis performed in the cochlea by the basilar partition. The dual profiles for all of the stimuli with F0's of 75 and 300 Hz are shown in Fig. 7. Figures 7(a)–7(h) show the profiles for an F0 of 75 Hz and Figs. 7(i)–7(p) show the profiles for an F0 of 300 Hz. Each row in Fig. 7 shows dual profiles with a constant NC; the value is eight for the top row, four for the middle row, eight for the second from bottom row, and four for the bottom row. Each column shows dual profiles for stimuli with a constant ALC, with values of three for the leftmost column, seven and eleven for the middle columns, and fifteen for the rightmost column. Thus, Fig. 7(a) is the dual profile for the stimulus with an F0 of 75 Hz, consisting of eight harmonics beginning from the third, and Fig. 7(p) is for an F0 of 300 Hz, consisting of four harmonics beginning from the fifteenth. Figure 7 shows that, generally, the spectral profiles do not contain many resolved harmonics for stimuli with lowest components above seven; this is shown in the three rightmost columns of Fig. 7.

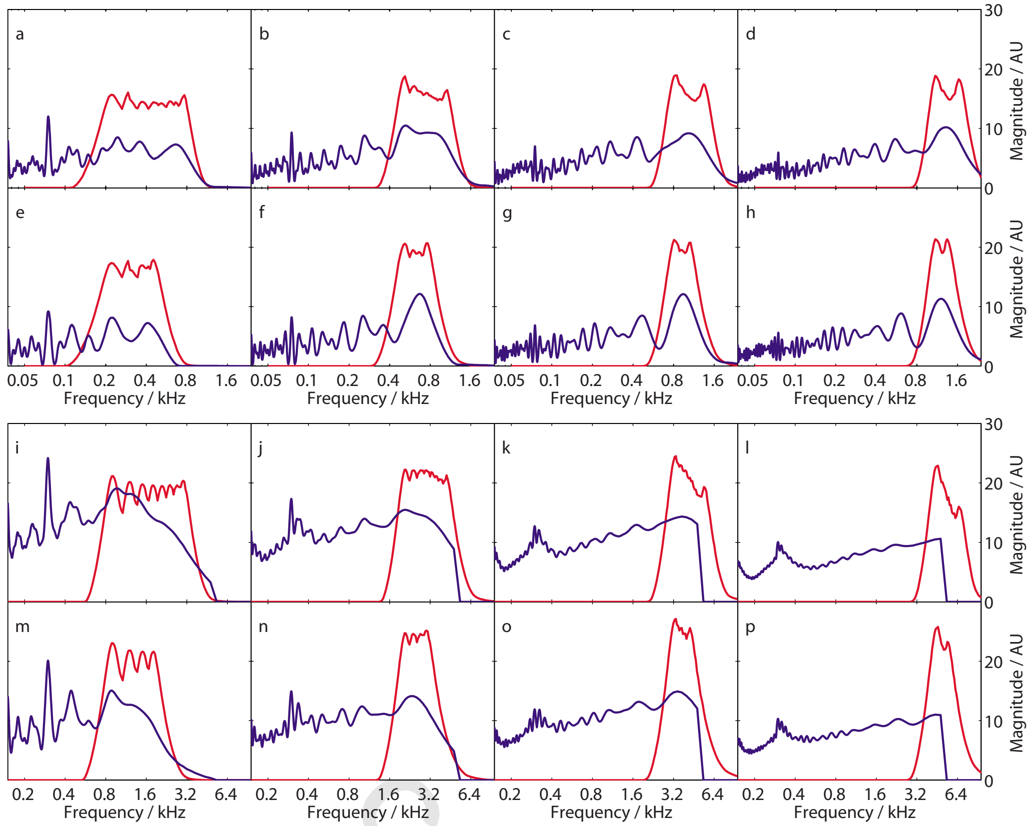

FIG. 7. Dual profiles produced with the GT-AFB for stimuli with F0's of 75 and 300 Hz. Panels (a)–(h) Profiles for an F0 of 75 Hz. (i)–(p) Profiles for an F0 of 300 Hz. Each row shows dual profiles with a constant NC; the value is eight for the top row, four for the middle row, eight for the second from bottom row, and four for the bottom row. Each column shows dual profiles for stimuli with a constant ALC, with values of three for the leftmost column, seven and eleven for the middle columns, and fifteen for the rightmost column.

The temporal profile always has a peak at the F0 of the harmonic series, 75 or 300 Hz, depending on the stimulus. The peak at F0 is not always the largest peak in the temporal profile; however, any other large peaks are spaced well away from the F0 in frequency. Sometimes the peaks are up to four octaves away, and as such, are far enough away not to interfere with the F0 peak. The peak used in the modeling was the largest peak within a two octave range centered on the fundamental. Thus, the temporal profile marks the F0 value by the location of a single peak and there is no need for a more central subharmonic generator. The height of the peak relative to the adjacent troughs can be used to estimate the strength of the pitch (Patterson *et al.*, 1996; Patterson *et al.*, 2000) and to explain the lower limit of pitch for complex harmonic sounds (Pressnitzer *et al.*, 2001). The pitch strength metric is illustrated in Fig. 6 by the faint lines; it is the height of the peak at F0, measured from the abscissa, minus the average of the trough values on either side of the peak, again measured from the abscissa. The effect of the loss of phase locking on this metric can be readily observed in the lower two rows of Fig. 7, where the F0 is 300 Hz and NC is either 8 or 4. As ALC increases from panel to panel across each row, the peak to trough ratio decreases progressively. The effect is much smaller in the upper rows where the energy of the stimulus is concentrated in the region below 2000 Hz, where phase locking is more precise.

There is a ceiling effect in the perceptual data at the lowest ALC values (3 and 7). Accordingly, the maximum value of the peak-to-trough ratio was limited to 7 in the modeling of pitch strength. This had the effect of limiting the model's estimate of pitch strength so that it did not rise further as ALC decreased from 7 to 3. The solid black and grey lines in Fig. 8 show the pitch strength estimates as a function of ALC for the 75 and 300-Hz F0's, respectively. The pitch

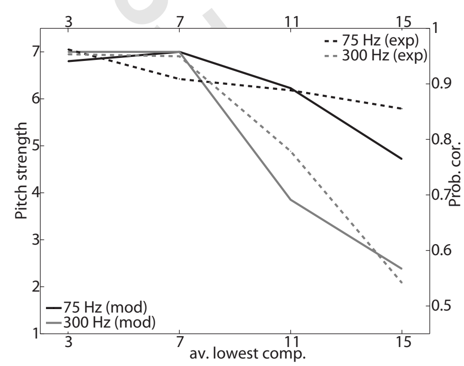

FIG. 8. Comparison of the experimental results with pitch-strength estimates from the dual profile model, based on a gammatone auditory filterbank, for an F0 of 75 Hz (black lines) and 300 Hz (grey lines). Dashed lines are the average experimental data plotted using the right ordinate (probability of correct identification as a function of average lowest component). Solid lines are the model values plotted using the left ordinate (pitch strength as a function of the average lowest component).

strength values were averaged over the two NC conditions for each value of ALC. Figure 8 also shows the perceptual data for the listeners from both the main and the ancillary experiments averaged over NC for each F0. The perceptual data are presented separately for the two F0's with dashed black and grey lines for 75 and 300 Hz, respectively. Figure 8 shows that the model can explain the more rapid fall off in pitch strength with increasing ALC at the higher F0.

B. The dynamic, compressive gammachirp auditory filterbank

Unoki *et al.* (2006) have argued that the compressive GammaChirp auditory filter (cGC) of Irino and Patterson (2001) provides a better representation of cochlear filtering than the linear GT auditory filter in models of simultaneous masking. The magnitude characteristic of the cGC filter is asymmetric and level dependent with resolution similar to that described by Ruggero and Temchin (2005). At normal listening levels for speech and music, the bandwidth of the auditory filter is greater than the traditional ERB values reported by Glasberg and Moore (1990) as noted in Unoki *et al.* (2006). Moreover, Irino and Patterson (2006) have recently described a ***dynamic*** version of the cGC filter with fast-acting compression which suggests that AIM can be extended to explain two-tone suppression and forward masking, as well as simultaneous noise masking. In an effort to increase the generality of the modeling, a version of AIM with the nonlinear dcGC filterbank was used to produce dual profiles for the stimuli in the experiment, to determine the pitch-strength values that would be derived from the temporal profiles of this more realistic time-domain model of auditory processing. The dual profiles produced with the dcGC filterbank were quite similar to those produced with the gammatone filterbank, primarily because the nonlinearities do not distort the time-interval patterns produced in the cochlea simulation as noted in Irino *et al.* (2007). The temporal profiles exhibited somewhat more pronounced peaks for the higher values of ALC, and the spectral profiles contained even less information, as would be expected with a broader auditory. But the differences were not large, and so the pattern of performance predicted for the melodic pitch task is quite similar for AIM with the dcGC filterbank. The results indicate that AIM with the dcGC filterbank would have the distinct advantage of being able to explain temporal pitch, masking, and suppression within one time-domain framework.

V. SUMMARY AND CONCLUSIONS

The decrease in pitch strength that occurs as the components of a harmonic complex are increased in frequency was used to demonstrate the importance of temporal fine structure in pitch perception. Performance on a melodic pitch task was shown to be better when the fundamental was lower (75 Hz) rather than higher (300 Hz), despite the fact that the internal representation of the harmonic complex has more resolved components when the fundamental is higher. A time-domain model of auditory processing (AIM) (Patterson *et al.*, 1995; Bleeck *et al.*, 2004) was used to simulate the neural activity produced by the stimuli in the auditory nerve and to compare the spectral and temporal information in the simulated neural activity in the form of the spectral and temporal profiles of the auditory image. Peaks in the time-interval profile can explain the *decrease* in performance as F0 *increases*. The corresponding spectral profiles show that spectral resolution increases when F0 increases, which suggests that spectral models based on excitation patterns would predict that performance on the melody task would improve as F0 increases, which is not the case.

The temporal profiles produced by the traditional version of AIM with the gammatone filterbank are similar to those produced by the most recent version of AIM, with a dynamic, compressive gammachirp filterbank. The latter model offers the prospect of being able to explain pitch, masking, and two-tone suppression within one time-domain framework.

ACKNOWLEDGMENTS

Research supported by the U.K. Medical Research Council (G0500221, G9900369). We would like to thank Steven Bailey, a project student, for his assistance in running the experiment, and his participation as a listener. The authors would also like to thank Alexis Hervais-Adelman for assistance with the ANOVA calculations.

Bleeck, S., Ives, T., and Patterson, R. D. (**2004**). "Aim-mat: The auditory image model in MATLAB," Acta Acust. **90**, 781–788.

Bleeck, S., and Patterson, R. D. (**2002**). "A comprehensive model of sinusoidal and residue pitch," poster presentation at Pitch: Neural Coding and Perception, Delmenhorst, Germany, 14–18 August.

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

Fastl, H., and Stoll, G. (**1979**). "Scaling of pitch strength," Hear. Res. **1**, 293–301.

Fruhmann, M., and Kluiber, F. (**2005**). "On the pitch strength of harmonic complex tones," DAGA 05, Munchen, edited by H. Fastl and M. Fruhmann, DEGA, Berlin, Vol **II**, pp. 467–468.

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

Houtsma, A. J. M., and Smurzynski, J. (**1990**). "The central origin of the pitch of complex tones: Evidence from musical interval recognition," J. Acoust. Soc. Am. **87**, 304–310.

Irino, T., and Patterson, R. D. (**2001**). "A compressive gammachirp auditory filter for both physiological and psychophysical data," J. Acoust. Soc. Am. **109**, 2008–2022.

Irino, T., and Patterson, R. D. (**2006**). "A dynamic, compressive gammachirp auditory filterbank," IEEE Audio, Speech Lang. Proc. **14**, 2222–2232. AQ: #3

Irino, T., Walters, T. C., and Patterson, R. D. (**2007**). "A computational auditory model with a nonlinear cochlea and acoustic scale normalization," Proceedings of the 19th International Congress on Acoustics, Madrid.

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.

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

Patterson, R. D. (**1987**). "A pulse ribbon model of monaural phase perception," J. Acoust. Soc. Am. **82**, 1560–1586.

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

Patterson, R. D., Handel, S., Yost, W. A., and Datta, J. A. (**1996**). "The relative strength of the tone and noise components in iterated rippled noise," J. Acoust. Soc. Am. **100**, 3286–3294.

Patterson, R. D., Peters, R. W., and Milroy, R. (**1983**). "Threshold duration for melodic pitch," in *Hearing-Physiological Bases and Psychophysics*, edited by R. Klinke and R. Hartmann, Proceedings of the Sixth International Symposium on Hearing (Springer, Berlin), pp. 321–326.

Patterson, R. D., Robinson, K., Holdsworth, J., McKeown, D., Zhang, C., and Allerhand, M. (**1992**). "Complex sounds and auditory images," in *Auditory Physiology and Perception*, Proceedings of the Ninth International Symposium on Hearing, edited by Y. Cazals, L. Demany, and K. Horner (Pergamon, Oxford), pp. 429–446.

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

Pressnitzer, D., and Patterson, R. D. (**2001**). "Distortion products and the pitch of harmonic complex tones," in *Proceedings of the 12th International Symposium on Hearing, Physiological and Psychophysical Bases of Auditory Function*, edited by D. Breebaart, A. Houtsma, A. Kohlrausch, V. Prijs, and R. Schoonhoven (Shaker BV, Maastrict), pp. 97–104.

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

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

Ruggero, M. A., and Temchin, A. N. (**2005**). "Unexceptional sharpness of frequency tuning in the human cochlea," Proc. Natl. Acad. Sci. U.S.A. **102**, 18614–18619.

Seither-Preisler, A., Johnson, L., Krumbholz, K., Nobbe, A., Patterson, R. D., Seither, S., and Lütkenhöner, B. (**2007**). "Observation: Tone sequences with conflicting fundamental pitch and timbre changes are heard differently by musicians and non-musicians," J. Exp. Psychol. Hum. Percept. Perform. **33**, 743–751.

Slaney, M., and Lyon, R. F. (**1990**). "Visual representations of speech—A computer model based on correlation," J. Acoust. Soc. Am. **88**, S23–■■■. AQ: #4

Unoki, M., Irino, T., Glasberg, B. R., Moore, B. C. J., and Patterson, R. D. (**2006**). "Comparison of the roex and gammachirp filters as representations of the auditory filter," J. Acoust. Soc. Am. **120**, 1474–1492.

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

PROOF COPY 002805JAS

PROOF COPY 002805JAS

AUTHOR QUERIES — 002805JAS

#1 AQ: Check word
#2 AQ: Check edit
#3 AQ: Provide full journal title, CODEN and/or ISSN number for Ref. c9
#4 AQ: Supply page range