Pitch discrimination with mixtures of three concurrent harmonic complexes

Jackson E. Graves^{a)} and Andrew J. Oxenham

Department of Psychology, University of Minnesota, 75 East River Parkway, Minneapolis, Minnesota 55455, USA

(Received 13 July 2018; revised 19 February 2019; accepted 13 March 2019; published online 17 April 2019)

In natural listening contexts, especially in music, it is common to hear three or more simultaneous pitches, but few empirical or theoretical studies have addressed how this is achieved. Place and pattern-recognition theories of pitch require at least some harmonics to be spectrally resolved for pitch to be extracted, but it is unclear how often such conditions exist when multiple complex tones are presented together. In three behavioral experiments, mixtures of three concurrent complexes were filtered into a single bandpass spectral region, and the relationship between the fundamental frequencies and spectral region was varied in order to manipulate the extent to which harmonics were resolved either before or after mixing. In experiment 1, listeners discriminated major from minor triads (a difference of 1 semitone in one note of the triad). In experiments 2 and 3, listeners compared the pitch of a probe tone with that of a subsequent target, embedded within two other tones. All three experiments demonstrated above-chance performance, even in conditions where the combinations of harmonic components were unlikely to be resolved after mixing, suggesting that fully resolved harmonics may not be necessary to extract the pitch from multiple simultaneous complexes. © 2019 Acoustical Society of America. <https://doi.org/10.1121/1.5096639>

[VRM]

Pages: 2072–2083

I. INTRODUCTION

Human listeners are generally able to perceive multiple pitches at the same time without great difficulty. In Western music in particular, the presence of three or more concurrent pitches is the rule, not the exception (e.g., [Parncutt et al., 2019](#)). Nevertheless, despite an extensive body of psychoacoustic research on pitch perception for single harmonic complexes ([Plack and Oxenham, 2005](#)), and some on the perception of two-complex mixtures (e.g., [Beerends and Houtsma, 1989](#); [Carlyon, 1996](#); [Micheyl et al., 2006, 2010](#); [Wang et al., 2012](#)), the perception of three or more concurrent complexes has received relatively little attention. Besides being ecologically valid, mixtures of three or more concurrent complexes can provide a strong test for models of pitch perception. More specifically, testing the perception of such mixtures may help to distinguish between the two most prevalent classes of pitch models, based either on rate-place or temporal coding.

Rate-place models (also known as place, template, or pattern-recognition models) are based on the tonotopic representation of harmonics along the basilar membrane and the pattern of average firing rate (excitation pattern) produced by them in the auditory nerve ([Goldstein, 1973](#); [Wightman, 1973](#); [Terhardt, 1974](#); [Cohen et al., 1995](#); [Shamma and Klein, 2000](#); [Cedolin and Delgutte, 2005](#)). These models require the presence of some spectrally resolved harmonics, or peaks in the excitation pattern, in order to extract pitch;

therefore, when no resolved harmonics are contained within a single complex or in a mixture, they will fail ([Shamma and Klein, 2000](#); [Cedolin and Delgutte, 2005](#)). In contrast, temporal models are generally based on the time intervals between neural spikes within and across auditory nerve fibers that share the same (or similar) characteristic frequency, and are typically instantiated via an autocorrelation function ([Licklider, 1951](#); [Meddis and Hewitt, 1991](#); [Cariani and Delgutte, 1996](#); [de Cheveigné, 1998](#)). These models do not rely on spectrally resolved components, and so in principle can account for pitch perception even in the absence of any spectrally resolved harmonics ([Carlyon, 1998](#); [Bernstein and Oxenham, 2005](#); for a recent review, see [Oxenham, 2018](#)).

Pitch discrimination abilities and pitch salience decrease dramatically when harmonics of a complex below about the tenth are removed ([Houtsma and Smurzynski, 1990](#); [Shackleton and Carlyon, 1994](#)). This decrease in performance coincides with the point at which none of the remaining harmonics are spectrally resolved ([Bernstein and Oxenham, 2003](#)). However, it is not clear whether it is spectral resolvability or harmonic number *per se* that underlies this phenomenon. In other words, is it the lack of spectrally resolved peaks in response to the stimulus that leads to poorer performance (as predicted by place models of pitch) or is it something else that covaries with the lowest harmonic number in each complex? [Bernstein and Oxenham \(2003\)](#) showed that spectrally resolved harmonics are not in themselves sufficient to produce a salient pitch: when peripheral resolvability was artificially increased by presenting alternating harmonics to opposite ears, fundamental-frequency (F_0) discrimination did not improve. Both [Bernstein and](#)

^{a)}Current affiliation: Laboratoire des systèmes perceptifs, Département d'études cognitives, École normale supérieure, PSL University, CNRS, 29 rue d'Ulm, 75005 Paris, France. Electronic mail: grave276@umn.edu

Oxenham (2005) and de Cheveigné and Pressnitzer (2006) have presented temporal pitch models that do not rely on harmonic resolvability but can still account for poorer pitch perception when only high-numbered (>10) harmonics are present.

Experiments that use single harmonic complex tones cannot determine whether harmonic resolvability underlies the transition from good to poor pitch perception with increasing lowest harmonic number, because changes in resolvability co-vary with changes in harmonic number. Presenting more than one harmonic tone complex at once should help to dissociate peripheral resolvability from other properties associated with harmonic number, by decreasing the resolvability of the individual harmonics while maintaining the same harmonic numbers in each complex tone. In other words, the resolvability of the components within each complex tone after mixing with other complex tones is likely to be different from the resolvability of the components before mixing.

To date, studies using multiple complexes have reached differing conclusions concerning the role of resolvability. Carlyon (1996) measured listeners' ability to discriminate the pitch of a bandpass-filtered tone complex in the presence of a masker tone complex, filtered into the same spectral region, with an F_0 at the geometric center of the target F_0 s in the two intervals of the two-interval forced-choice task. Performance was good only in conditions when the target contained resolved harmonics. However, the presence of the masker did not necessarily affect the resolvability of the target harmonics, as the masker and target harmonics did not produce individually resolvable spectral peaks at the small F_0 differences tested by Carlyon (1996). Thus, these results do not help in determining whether resolvability before or after mixing is critical. Beerends and Houtsma (1989) presented two harmonic tones with different F_0 s and used only two consecutive harmonics from each complex, presented in various combinations within and between the two ears. They found that performance only dropped when all four components were presented to the same ear, suggesting a role for harmonic resolvability after mixing. Consistent with this conclusion, Micheyl *et al.* (2010) found that accurate F_0 discrimination of a complex embedded within a masker complex in the same spectral region was only possible when at least some spectral peaks corresponding to the target harmonic frequencies were resolved. On the other hand, Bernstein and Oxenham (2008) found that taking a single harmonic complex and mistuning the odd harmonics relative to the even harmonics by 3%, thereby essentially creating two complexes with differing F_0 s, improved F_0 difference limens (F_0 DLs) without improving resolvability. It may be that the combination of just two complex tones does not sufficiently dissociate harmonic number from spectral resolvability to permit a definitive conclusion.

The aim of the present study was to better dissociate harmonic number from spectral resolvability by using combinations of three concurrent harmonic complexes, filtered into the same spectral region. The spectral region into which the complexes were filtered was manipulated in order to control the degree of resolvability of components within these

mixtures, before and after mixing. In all three experiments, listeners were capable of discriminating pitch changes of one semitone (ST) or less, even in conditions where resolved components were not likely to be present in the mixture, suggesting that spectrally resolved components are not necessary for pitch perception at this level of accuracy.

II. EXPERIMENT 1: MAJOR AND MINOR TRIAD DISCRIMINATION

A. Rationale

The goal of this experiment was to test whether listeners can perceive three simultaneous pitches with a precision of at least one ST in a three-complex mixture, in conditions where the individual complexes contain resolved harmonics before mixing (and thus include low-numbered harmonics), but are less likely to contain resolved harmonics after mixing. One ST was selected as the F_0 difference to be discriminated, as this is a functionally relevant threshold for many forms of music. If resolvability (after mixing) is the determining factor in pitch strength and accuracy, then the task should become impossible in conditions where the mixture contains no resolved harmonics. If, on the other hand, harmonic number is the determining factor, then performance should remain high as long as low-numbered harmonics are included, even when no resolved harmonics remain after mixing.

B. Methods

1. Listeners

Thirty normal-hearing listeners (21 female and 9 male) were initially recruited for the experiment, ranging in age from 19 to 44 years (mean = 24.9), and in years of musical experience from 0 to 35 (mean = 9.03). All listeners had normal hearing, defined here (and in all subsequent experiments) as audiometric thresholds of 20 dB hearing level or better at octave frequencies between 250 and 8000 Hz. Out of the recruited 30 listeners, only nine passed the screening test described in Sec. II B 3 and continued on to complete the complex-tone task and F_0 DL measurements. These 9 participants, 6 female and 3 male, ranged in age from 22 to 30 years (mean = 26.1), and ranged in years of musical experience from 2 to 19 (mean = 10.4).

2. Stimuli

In the main experiment, listeners were presented with major and minor triads in root position, first inversion, and second inversion. In Western music theory, these are combinations of three pitches that are separated from each other by specific numbers of STs, defined in the equal-temperament tuning system as a ratio of $2^{1/12}$ between F_0 s, such that an octave (doubling of F_0) is divided equally into 12 STs. In any of its three inversions, a major triad differs from a minor triad by only one note, and that note differs by only 1 ST. In root position, the F_0 s of a major triad are 0, 4, and 7 STs above the root F_0 , while the corresponding pattern for a minor triad is 0, 3, and 7 STs. In the first inversion, the major

triad has notes that are 4, 7, and 12 STs above the root, and minor is 3, 7, and 12 STs (corresponding to 0, 3, and 8 STs and 0, 4, and 9 STs, respectively, with reference to the lowest note of the triad). In the second inversion, major is 7, 12, and 16 STs, and minor is 7, 12, and 15 STs above the root (corresponding to 0, 5, and 9 STs and 0, 5, and 8 STs, respectively, with reference to the lowest note). On each trial, listeners were presented with one triad and were asked to indicate whether it was major or minor. The inversion (root, first, or second) was selected at random on each trial with equal probability. On each trial, the F_0 of the lowest note present was roved between 200 and 230 Hz using a uniform probability distribution on a logarithmic scale. This lowest note was either the root, third, or fifth, depending on which inversion was presented. The highest F_0 in the chord was between 7 and 9 STs above the lowest F_0 , corresponding to F_0 s between 300 Hz (7 STs above 200 Hz) and 387 Hz (9 ST above 230 Hz).

Because of the randomization of the inversion (root, first, or second inversion) and the roving of the lowest F_0 present, identifying any two of the three pitches is not sufficient to perform this task. For example, even after determining that the upper two pitches are 3 STs apart, a listener must also hear the third (lowest) pitch to know whether the chord is major, root position (0, 4, 7) or minor, second inversion (0, 5, 8). This is true for any combination of two out of the three pitches, so long as all three inversions are possible and the absolute pitch range is roved.

For the initial pure-tone screening task, each trial consisted of a single interval containing three concurrent pure tones whose frequencies formed either a major or minor triad, with all frequencies limited to between 200 and 387 Hz. Each pure tone had a level of 50 dB sound pressure level (SPL), and was 500 ms in duration, including 10-ms raised-cosine onset and offset ramps.

For the main experiment, each trial consisted of a single interval containing three concurrent harmonic complex tones whose F_0 s formed a major or minor triad, with all F_0 s between 200 and 387 Hz. The tones were filtered into one of four different spectral regions, with lower corner frequencies of 0.5, 2, 3, and 4 kHz. The upper corner frequency was 2 kHz for the condition with the 0.5 kHz lower corner frequency and was 8 kHz for the remaining conditions. Filter slopes were 24 dB per octave on either side of the passband. Tone durations were 500 ms, including 10 ms raised-cosine onset and offset ramps.

In addition to the main experiment, F_0 DLs were measured for single tones using an adaptive tracking procedure (see Sec. II B 3). In the F_0 DL task, the components were added in either sine (zero) starting phase or random phase, with the starting phase of each component selected at random from between 0 and 2π independently and with uniform distribution on each presentation. In the main experiment, the components were always added in random phase.

For both the F_0 DL task and the main experiment, all complexes were presented at 40 dB SPL per component within the passband, which is below the level at which combination tones can normally be heard (Plomp, 1965). Furthermore, in order to mask any potential distortion

products, and to render tones outside the passband mostly inaudible, all complexes were embedded in threshold-equalizing noise (TEN) (Moore *et al.*, 2000), presented at 30 dB SPL within the equivalent rectangular bandwidth (ERB) (Glasberg and Moore, 1990) centered on 1 kHz, and gated on and off with 10 ms raised-cosine ramps. In the main experiment, the noise started 300 ms before tone onset, and ended 200 ms after tone offset. In the F_0 DL task, the noise began 300 ms before the first interval and ended 200 ms after the end of the second interval.

The filter slopes of 24 dB/octave were chosen in order to avoid the possibility of edge pitches associated with sharp spectral edges (Kohlrausch and Houtsma, 1992) and to avoid cues based on the lowest harmonic present in the stimulus. Because our filter slopes were not infinite, the low corner frequency of the filter in each spectral region is not equivalent to the lowest audible component of a complex filtered into that region. Because stimuli were embedded in TEN at a level of 10 dB per ERB below the level per component in the passband, the threshold of audibility is approximately 10 dB below the level of within-passband components. With filter slopes of 24 dB/octave, this corresponds to 75% of the lower cutoff frequency. Thus in the 2–8 kHz spectral region, for example, the lowest audible component was somewhere between 1.5 and 2 kHz.

The combinations of passbands and F_0 ranges in this experiment were selected to attempt to generate conditions that contained either (1) harmonics that were resolved both before and after the mixing of the three complexes; (2) harmonics that were resolved before mixing but became unresolved after the three complexes were combined; or (3) harmonics that were unresolved even before the three complexes were mixed together. The estimates for whether harmonics were resolved or unresolved after mixing were obtained using the following reasoning.

It is generally accepted that harmonics higher than the tenth are spectrally unresolved for single complex tones (Bernstein and Oxenham, 2003). The upper limit for resolved harmonics appears to be somewhere between the sixth and the tenth, depending on the definition and the method of measurement (Plomp, 1964; Houtsma and Smurzynski, 1990; Shackleton and Carlyon, 1994; Bernstein and Oxenham, 2003). Adding three harmonic complexes within the same spectral bandpass spectral region, all with F_0 s within the same range, results in roughly 3 times as many components within the same bandpass region, thereby reducing the average spacing between adjacent components by about a factor of 3. Simulated excitation patterns are plotted in Fig. 1, computed according to the filter shapes and bandwidths proposed by Glasberg and Moore (1990). A previous approach to estimating spectral resolvability of individual components used the level of a peak in the excitation pattern, relative to its two adjacent valleys (Micheyl *et al.*, 2010). Taking a conservative approach and assuming that the tenth harmonic of a single complex is the highest to be resolved, it can be calculated from the excitation patterns that the peak-to-valley ratio of the excitation pattern surrounding the tenth harmonic for F_0 s between 200 and 400 Hz (the range used in this experiment) is about 0.6 dB.

FIG. 1. Excitation patterns evoked by stimuli used in experiments 1–3. Stimuli were embedded in TEN presented to mask distortion products. The top row shows single complex tones with F_0 s of 275 Hz. The bottom row shows major root-position triads with F_0 s of 218, 275, and 327 Hz. Columns show the different spectral regions used for bandpass filtering. A stimulus was considered to have resolved components if it produced more than one peak of excitation greater than 0.6 dB above the neighboring valleys. In this way, the first ten harmonics of single complexes are considered to be spectrally resolved. For stimuli with resolved components, these peaks are indicated with arrows. In the 2–8 kHz spectral region, a single complex tone may contain resolved components, but a complex triad mixture in the same region likely does not.

Thus, we can assume that only spectral peaks within a mixture of three complexes that exceed this peak-to-valley ratio will be resolved. The arrows in Fig. 1 point to such spectrally resolved peaks. The advantage of this approach is that it is not completely dependent on the assumptions made regarding the bandwidths of the auditory filters, which appear to be sharper in humans than in other commonly studied species, and may be sharper than assumed in the Glasberg and Moore (1990) model (Shera *et al.*, 2002; Sumner *et al.*, 2018; Verschooten *et al.*, 2018). Our approach is not dependent on the assumed filters because it is calibrated relative to the resolvability of a single harmonic complex. In other words, if the underlying filters were sharper, then that would simply mean that a larger peak-to-valley ratio surrounding the highest resolved harmonic would be needed as the decision criterion, but the predicted pattern of resolvability would remain similar. Note also that by assuming that even the tenth harmonic is spectrally resolved, we are adopting a very conservative approach to declaring harmonics unresolved, ensuring that any peaks that could potentially be resolved are counted.

See Table I for a summary of F_0 s and lowest components in experiments 1–3. The passbands of the lowest two spectral regions (0.5–2 and 2–8 kHz) always included components lower than or equal to the tenth for single complexes with F_0 s between 200 and 400 Hz. Accordingly, single complexes in these two regions produced multiple peaks of excitation above the peak-to-valley ratio criterion of 0.6 dB (chosen to match the tenth harmonic for stimuli in our F_0 range), and so were considered to contain resolved harmonics. In the 4–8 kHz region, no components lower than the tenth were ever present, and these stimuli were found to have no resolved harmonics using this criterion. In the 3–8 kHz region, resolvability varied with F_0 (the example complex shown in Fig. 1, with an F_0 of 275 Hz, contains no

resolved harmonics using this criterion). Subjecting mixtures of three simultaneous complexes to the 0.6 dB rule calibrated for single complex tones, resolved components were present in the 0.5–2 kHz region even after mixing, but the other three spectral regions contained no resolved components after mixing. In mixtures of three complex tones, the average spacing between components is decreased by a factor of 3 relative to a single complex tone (see Table I). This explains why no resolved components were present for mixtures in the 2–8 kHz region, even though this region did contain resolved components for single complexes before mixing.

Additional F_0 DL measurements were carried out with single complex tones in isolation. In the F_0 DL task, each trial consisted of two intervals, each containing 200 ms harmonic complex tones, separated by a 100 ms gap. The F_0 s of the two tones were geometrically centered on 260 Hz. Tones were filtered into the same four spectral regions as in the main experiment.

All stimuli were generated within MATLAB (The Mathworks, Natick, MA), using a 24-bit L22 soundcard (LynxStudio, Costa Mesa, CA), and were presented diotically through HD650 headphones (Sennheiser USA, Old Lyme, CT) at a sampling rate of 44.1 kHz.

3. Procedure

Listeners first performed a pure-tone screening task in which they were instructed to distinguish major from minor triads, produced with pure tones. The criterion to pass this screening was a performance level of at least 80% correct for each inversion (root, first, second) after a maximum of 5 training blocks, with each block containing 60 trials (20 trials per inversion). The purpose of this training/screening task was to ensure that listeners could reliably distinguish a major chord from a minor chord in simple conditions where

TABLE I. Fundamental frequencies (F_0), tenth component frequencies, and lowest components in the passband for stimuli used in experiments 1–3. “Mixture avg.” refers to the average spacing between components in the mixture, obtained by averaging the three F_0 s in the mixture, then dividing by three. Lowest component ranges are in bold if the passband always contains the tenth component or lower, and italicized if the passband never contains the tenth component or lower.

	Experiment 1					
	Components		Lowest component in passband by low cutoff			
	F_0 (Hz)	10th (kHz)	0.5 kHz	2 kHz	3 kHz	4 kHz
Highest tone	300–387	3.0–3.9	2nd	6th–7th	8th–11th	<i>11th–14th</i>
Middle tone	238–307	2.4–3.1	2nd–3rd	7th–9th	10th–13th	<i>14th–17th</i>
Lowest tone	200–230	2.0–2.3	3rd	9th–11th	<i>14th–16th</i>	<i>18th–21st</i>
Mixture avg.	82–103	0.8–1.0	5th–7th	<i>20th–25th</i>	<i>30th–37th</i>	<i>39th–49th</i>
	Experiment 2					
	Components		Lowest component in passband by low cutoff			
	F_0 (Hz)	10th (kHz)	0.5 kHz	2 kHz	3 kHz	4 kHz
High masker	283–487	2.8–4.9	2nd	5th–8th	7th–11th	9th–15th
Target	231–335	2.3–3.4	2nd–3rd	6th–9th	9th–13th	<i>12th–18th</i>
Low masker	200–224	2.0–2.2	3rd	9th–11th	<i>14th–16th</i>	<i>18th–21st</i>
Mixture avg.	79–116	0.8–1.2	5th–7th	<i>18th–26th</i>	<i>26th–38th</i>	<i>35th–51st</i>
	Experiment 3					
	Components		Lowest component in passband by low cutoff			
	F_0 (Hz)	10th (kHz)	0.5 kHz	2 kHz		
Highest masker	371–588	3.7–5.9	<i>F0–2nd</i>	4th–6th		
High masker	312–416	3.1–4.2	2nd	5th–7th		
Target	258–298	2.6–3.0	2nd	7th–8th		
Low masker	185–247	1.9–2.5	3rd	9th–11th		
Lowest masker	131–208	1.3–2.1	3rd–4th	10th–16th		
<i>Mixture avg. by condition</i>						
Target lowest	105–144	1.1–1.4	4th–5th	<i>14th–20th</i>		
Target middle	84–106	0.8–1.1	5th–6th	<i>19th–24th</i>		
Target highest	64–83	0.6–0.8	7th–8th	<i>25th–32nd</i>		

harmonic resolvability did not play a role. Listeners who passed the pure-tone screening then completed the same task using complex tones that were all filtered into one of four spectral regions (see Sec. II B 2 above). Each listener completed a total of 100 trials per inversion in each spectral region. The inversion varied pseudorandomly from trial to trial, and spectral region changed every block of 60 trials, with block order randomized between subjects and repetitions. Finally, listeners’ F_0 DLs for individual complexes were measured in the same four spectral regions, using a 2-down 1-up adaptive-tracking procedure (Levitt, 1971). For each run, the starting value for F_0 difference was 20%, and the initial step size by which the F_0 difference was changed was a factor of 1.58. After each downward reversal, the step size decreased to a factor of 1.26, 1.19, and finally 1.1. The run then continued for another six reversals at the final step size, and the F_0 DL for that run was defined as the geometric mean F_0 difference at the last six reversal points. The final F_0 DL for each subject was then defined as the geometric mean F_0 DL from three runs. Throughout all stages of the experiment, listeners received visual feedback after every

trial indicating whether their response was correct or incorrect.

C. Results

The results of the pure-tone screening task are shown in the left panel of Fig. 2. Beginning in the first block of training, listener performance was worse for first-inversion triads than for the other two inversions. Out of the 30 recruited listeners, 9 sufficiently improved their performance in all 3 inversions to be above the 80% performance criterion by the final block, while 21 did not. All other results refer only to the data from the nine listeners who met the screening criteria.

As shown in the middle panel of Fig. 2, average F_0 DLs were below 1% in the lower two spectral regions, but increased in the higher two spectral regions, with DLs approaching 1 ST ($\sim 6\%$) in the highest region, where an effect of component phase (sine vs random) was also clearest. A repeated-measures analysis of variance (ANOVA) on the log-transformed F_0 DLs found a main effect of phase [$F(1,8) = 12.8$, $p = 0.007$], a main effect of spectral region

FIG. 2. (Color online) Results from experiment 1. Left: results from pure-tone training on the major/minor discrimination task, for the 9 listeners who met the criterion and the 21 listeners who did not meet the criterion. Only the nine listeners who passed went on to complete the rest of the experiment. Center: $F0$ DLs from nine listeners for single complex tones in four spectral regions, with components added in sine or random phase. Right: Results from 9 listeners in the major/minor discrimination task with complex tones. Error bars show ± 1 standard error of the mean (SEM).

[$F(3,8) = 85.3$, $p < 0.001$], and an interaction [$F(3,24) = 4.97$, $p = 0.008$]. Paired comparisons showed that the effect of phase was significant only for the highest spectral region ($p < 0.001$).

The results from the major–minor discrimination task are shown in the right panel of Fig. 2. A repeated-measures ANOVA on percent correct scores, transformed into rationalized arcsine units (RAU; [Studebaker, 1985](#)), found a main effect of spectral region [$F(3,24) = 33.2$, $p < 0.001$], and a main effect of inversion [$F(2,16) = 5.7$, $p = 0.01$], but no interaction [$F(6,48) = 0.3$, $p = 0.93$]. Averaging across the three inversions for each subject, performance was significantly above chance in all but the highest spectral region (single-sample one-tailed t -tests, $p < 0.0125$ with Bonferroni correction for multiple comparisons). In the highest spectral region, performance was not significantly different from chance [mean RAU = 52.63, $t(8) = 1.78$, $p = 0.057$].

D. Discussion

The pattern of results was similar for both the single- and three-complex tasks: for the single-complex task, $F0$ DLs were low and, for random-phase harmonic complexes, only approached 1 ST in the highest spectral region, where none of the harmonics would be considered resolved; for the three-complex task, performance was high for conditions in the lower spectral regions and decreased to near chance (reflecting an inability to discriminate a 1 ST difference) in the highest spectral region. Most importantly, performance in the three-complex task remained above chance in the two middle spectral regions, where the harmonics were likely unresolved in the mixture. Therefore, the results of experiment 1 suggest that resolved harmonics may not be necessary for discriminating major from minor triads.

For the three-complex task, our finding that only 9 out of the 30 initially recruited listeners met the 80% correct criterion after 5 blocks of pure-tone training corresponds well with previous findings that performance for major/minor chord quality discrimination is bimodally distributed, with about 30% of listeners performing near ceiling and the rest near chance ([Chubb et al., 2013](#); [Mednicoff et al., 2018](#)). By limiting the listeners in our study to only those who could reliably distinguish major

from minor triads with pure tones, we ensured that errors in the complex-tone version of the task were not due to a basic inability to distinguish major from minor chord quality.

While the major/minor task is not directly comparable to single-complex discrimination, performance in the former is likely limited by performance in the latter: if a listener cannot discriminate a 1 ST difference, they should be unable to tell major from minor. For this reason, because $F0$ DLs for single complexes deteriorate in the 3–8 and 4–8 kHz regions, to approximately 1 ST in the 4–8 kHz region, we would necessarily expect to see some deterioration of performance on the major/minor task in these regions as well.

For the single-complex task, the elevated DLs and the emergence of a phase effect in the highest spectral region confirm that the single complexes in the highest region may not contain resolved harmonics. This emergence of a phase effect as the lowest component within the passband changed from about the eighth (in the second spectral region) to about the 16th (in the fourth and highest spectral region), coinciding with a sharp increase in DL, agrees with previous findings on resolvability of single complexes ([Houtsma and Smurzynski, 1990](#); [Shackleton and Carlyon, 1994](#); [Bernstein and Oxenham, 2003](#)).

Triads in the first inversion were consistently more difficult to discriminate than those in the second inversion or in root position. Since this effect is apparent even in the pure-tone training version of the task, it seems unlikely that it is related to complex pitch perception, and perhaps results from higher-level factors related to experience or knowledge of music theory. For instance, the lower two notes in the first inversion have the opposite relationship compared to the lower two notes in the root position, with a 3 ST interval implying a major triad in the first inversion but implying a minor triad in the more common root position.

III. EXPERIMENT 2: HEARING OUT ONE COMPLEX IN THE PRESENCE OF TWO OTHERS

A. Rationale

The results of experiment 1 support the idea that low harmonic numbers, but not necessarily resolved harmonics,

are necessary for accurate pitch perception. However, it is possible that participants may have been listening for some emergent “holistic” cue that distinguishes major from minor triads, rather than explicitly coding all three pitches in the combination. For example, major triads are rated as slightly more consonant than minor triads, which could conceivably be due to differing overall acoustic properties between the two chord qualities (McDermott *et al.*, 2010). This seems unlikely, given that the chord inversion (root, first, or second) was randomized from trial to trial and absolute F_0 was also roved. Taken together, these two randomizations would have made any individual F_0 , or even two F_0 s, an unreliable cue. Due to the same two factors, patterns of fluctuations in the temporal envelope, caused by beating between nearby components, would also have been inconsistent. Thus, there are no obvious cues on which any putative holistic quality related to major or minor could be based. Nevertheless, given this potential concern, along with the fact that only one-third of the screened listeners were able to perform the task satisfactorily even with pure tones, a second experiment was run. In this experiment, a more explicit test of pitch discrimination was provided by adapting a paradigm that has been used previously to study pure-tone pitch perception (Demany *et al.*, 2011) and auditory enhancement (Feng and Oxenham, 2015, 2018).

B. Methods

1. Listeners

Twenty-two normal-hearing listeners participated in the experiment, none of whom had participated in experiment 1. Out of the recruited 22 listeners, 15 passed the initial pure-tone screening task, described in Sec. III B 3 below, and these 15 continued on to complete the single- and multiple-pitch complex-tone tasks. These 15 participants, 10 female and 5

male, ranged in age from 18 to 26 (mean = 20.7), and ranged in years of musical experience from 0 to 19 (mean = 5).

2. Stimuli

The paradigm for experiment 2 is illustrated on the left side of Fig. 3. On each trial, listeners heard a single 300 ms reference tone, followed after a 600 ms gap by a second 300 ms target tone (pure-tone screening and single-pitch condition) or by a combination of three simultaneous 300 ms tones with different F_0 s (multiple-pitch condition), where the middle tone was the target and the higher and lower tones were maskers. The F_0 s of the reference and target tone were either 1 or 0.5 ST apart, and the direction of the pitch change was either up or down. Listeners had to identify the direction of this pitch change. The low masker’s F_0 was roved between 200 and 224 Hz. The reference tone and target tone F_0 s were geometrically centered on a nominal F_0 that was roved from 3 to 6 STs above the low masker’s F_0 , and so always fell between 238 and 317 Hz. The high masker’s F_0 was roved from 3 to 7 STs above the nominal F_0 , and so always fell between 283 and 475 Hz. All complex tones were filtered and embedded in TEN as in experiment 1, with the TEN beginning 300 ms before the first tone and ending 200 ms after the second tone within each trial.

3. Procedure

Using the same spectral regions as in experiment 1, listeners in experiment 2 identified the direction of an F_0 change of 1 or 0.5 ST between two complex tones, when the second tone was presented concurrently with two other complex tones, one with a higher F_0 and one with a lower F_0 . Listeners completed 100 trials per F_0 difference in each spectral region. The F_0 difference varied (either 0.5 or 1 ST) from trial to trial, and spectral region changed every block of

FIG. 3. Example trials from experiments 2 and 3. Listeners identified the direction of a pitch change between reference tone(s) and target tone, with the distance between the two tones held constant at either 0.5 or 1 ST. The target tone was either presented alone, or as the lowest, middle, or highest of a three-tone mixture. Dashed vertical lines show onset and offset of TEN.

40 trials, with the block order randomized. As a control condition, listeners' performance was also measured in a single-pitch condition where the second tone was presented alone, with no concurrent masking complex tones. Listeners were screened with a task equivalent to the single-pitch condition using pure tones before advancing to other tasks, to ensure they could reliably perform the task. The criterion for the pure-tone screening task was performance of 80% correct or above in both the 1 and 0.5 ST conditions, after a maximum of 3 training blocks, each consisting of 100 trials per condition. As in experiment 1, feedback was provided after every trial in all stages.

C. Results

The results from experiment 2 are shown in Fig. 4. On their final training block, the 15 listeners who passed the pure-tone screening task responded correctly on average to 98.9% of trials in the 1 ST condition, and 94.9% in the 0.5 ST condition. For the single-pitch task, performance deteriorated in the two highest spectral regions, where no harmonics below the tenth were present in the passband for at least some (3–8 kHz condition) or all (4–8 kHz) trials. However, performance remained well above chance for all spectral regions in the single-pitch task, for both the 1 and 0.5 ST F_0 difference between the reference and target tone. For the multiple-pitch task, performance was above chance in the lowest 3 spectral regions, for both F_0 differences (single-sample t -tests, $p < 0.0063$ with Bonferroni correction for multiple comparisons). In the highest spectral region, performance was not significantly different from chance for the multiple-pitch task, in either the 1 ST [mean RAU = 53.6, $t(14) = 2.15$, $p = 0.025$] or the 0.5 ST [mean RAU = 52.4, $t(14) = 1.78$, $p = 0.05$] conditions. A repeated-measures ANOVA on the RAU-transformed scores confirmed a main effect of spectral region [$F(3,42) = 81.7$, $p < 0.001$], a main effect of task (single vs multiple tones) [$F(1,14) = 169.2$, $p < 0.001$], a main effect of F_0 difference (0.5 vs 1 ST) [$F(1,14) = 37.9$, $p < 0.001$], and an interaction of task by spectral region [$F(3,42) = 9.74$, $p < 0.001$], presumably caused by ceiling effects in the single-pitch conditions in the lowest spectral regions, but no other interactions.

FIG. 4. (Color online) Results from experiment 2. Listeners identified the direction of a pitch change in single-complex-tone and multiple-complex-tone conditions, with a constant F_0 difference of either 1 or 0.5 ST. Target F_0 was roved between 238 and 317 Hz. Error bars show ± 1 SEM.

Pairwise comparisons revealed that all spectral regions differed significantly from each other ($p < 0.001$ in all cases). *Post hoc* comparisons investigating the task by spectral region interaction revealed that all spectral regions differed significantly from each other within each task ($p < 0.001$ in all cases) except for the lowest two spectral regions in the single-pitch task, which were not significantly different from each other ($p = 0.18$), again presumably due to ceiling effects.

D. Discussion

The results suggest that listeners are able to discriminate a 0.5 ST F_0 difference between two complexes, even when they are filtered into a spectral region where the mixture is unlikely to contain any spectrally resolved harmonics. However, overall performance was poorer in the multiple-pitch task than in the single-pitch task. One interpretation of the overall degradation in performance for the combination task relative to the single task is as an attention-related difficulty occurring centrally; this would be a similar effect to that observed by Beerends and Houtsma (1986) in their tests of dichotically presented simultaneous two-tone complexes.

If the only source of listener error on these tasks were sensory noise, or a failure to accurately encode the pitch of the tones, we would expect listener sensitivity to be directly proportional to F_0 difference. This is not the case: although the effect of F_0 difference on accuracy is significant, it is generally quite small. For example, in the lowest spectral region in the combination task, doubling the F_0 difference from 0.5 to 1 ST increases performance from 75.73% correct to 79.93%. For comparison, with no bias, a doubling of sensitivity (d') from 1.35 to 2.7 would increase accuracy from 75% to 91% correct. This suggests another non-sensory source of listener error on this task, perhaps related to difficulty in attending to the target tone and ignoring the maskers. In other words, listeners may be distracted by the maskers, independent of any peripheral interference between the signal representations in the rate-place code. Under this interpretation, the effect could be compared to other findings of masking not explainable by sensory interference, usually called informational masking (Pollack, 1975; Neff *et al.*, 1993; Oxenham *et al.*, 2003).

Besides interpretations based on attention or informational masking, another interpretation of the difference between the single-pitch task and the combination task is as an effect of decreased spectral resolvability due to the addition of the masker tones. In order to evaluate these two possible interpretations, we modified the paradigm in the final experiment to further facilitate attention to the target tone. We also introduced a control condition with pure tones to evaluate the difference in performance between single- and multiple-pitch conditions when all stimuli are clearly spectrally resolved. In addition, the final experiment also investigated whether systematic differences exist between the ability to discriminate the F_0 of the target, depending on whether it was the lowest, middle, or highest tone within a chord.

IV. EXPERIMENT 3: HEARING OUT THE LOW, MIDDLE, OR HIGH VOICE

A. Rationale

The middle voices in polyphonic music are not typically as easy to follow as the outer voices, particularly the highest voice (Trainor *et al.*, 2014). Since listeners in experiment 2 were tasked with hearing out the middle voice of three, it is possible that performance in the multiple-pitch task could be improved if the task instead involved attending to the lowest or highest voice in the mixture. Previous studies have shown that listeners are behaviorally most sensitive to pitch changes in the high voice (Palmer and Holleran, 1994), and even that early brain responses to unexpected stimuli are stronger if the unexpected stimulus comes in the high voice (Fujioka *et al.*, 2008). However, some of these previously observed effects might be due to spectral differences between the voices, rather than pitch differences, as previous studies did not filter their stimuli into the same spectral region. It is also unclear whether the effect is due to the relative or absolute pitch of the high voice: is the advantage for higher F_0 s in absolute terms, or for F_0 s that are just higher than their concurrent neighbors? In experiment 3 we controlled for these factors by filtering the stimuli into the same spectral regions (as in experiments 1 and 2), and by holding the target F_0 constant while varying the F_0 s of the maskers.

Beyond the problem of listening to the middle voice, the increased difficulty of the multiple-pitch task in experiment 2, relative to the single-pitch task, may be due in part to the increased demand on attention-related processes involved selecting the target tone from within the complex mixture. In an attempt to reduce this potential effect, we used a paradigm that encourages listeners to hear out one tone in the presence of others by repeating the target tone beforehand to create a perceptual stream, as has been done in previous studies (e.g., Darwin *et al.*, 1995; Oxenham and Dau, 2001; Shinn-Cunningham *et al.*, 2007).

B. Methods

1. Listeners

Fifteen normal-hearing listeners participated in the experiment, 10 female and five male, ranging in age from 18 to 64 years (mean = 30.7), and in years of musical experience from 2 to 13 (mean = 6). None of these listeners had participated in either experiment 1 or 2. All 15 listeners passed the pure-tone screening task and completed the other tasks.

2. Stimuli

The right side of Fig. 3 illustrates the paradigm used in experiment 3, which was similar to that of experiment 2. Tones were 250 ms in duration with 50 ms gaps in between. The reference tone was presented 3 times before the target tone, which was presented either alone or was gated with maskers. In experiment 3, the reference and target F_0 s always differed by exactly 0.5 ST. The nominal F_0 of the reference and target was roved between 262 and 294 Hz, and the difference in F_0 between any adjacent pair of tones in

the three-tone combination was roved between 3 and 6 ST, meaning that masker F_0 s ranged from 131 Hz (12 ST below the minimum nominal F_0) to 586 Hz (12 ST above the maximum nominal F_0). Importantly, the rove range for reference and target F_0 s stayed constant and did not change as a function of the target position. Instead, the “low target” condition used a mixture of complexes with a higher average F_0 and the “high target” condition used a mixture of complexes with a lower average F_0 . The tone levels, gating, and spectral composition were the same as in experiments 1 and 2, as was the TEN, which was gated on 300 ms before the first tone in the sequence, and gated off 200 ms after the final tone.

3. Procedure

Instead of the four spectral regions from the other two experiments, tones in this experiment were presented either as complexes bandpass filtered between 0.5 and 2 kHz, complexes bandpass filtered between 2 and 8 kHz, or as pure tones. Listeners completed 100 trials total per tone type for each target position (and the single-pitch task). Tone type changed every block of 20 trials (with block order randomized between listeners and repetitions). Listeners completed all 300 trials for each target position (or single-pitch task) before moving onto the next target position, and were explicitly informed in advance of which target position to listen for. The single-pitch task was always completed before the multiple-pitch tasks, but the order of target positions for the multiple-pitch tasks was counterbalanced between subjects. As in experiments 1 and 2, feedback was provided after every trial in all stages.

C. Results

The results from experiment 3 are shown in Fig. 5. Accuracy was near ceiling for all tone types in the single-pitch condition, which was therefore excluded from statistical analysis. A repeated-measures ANOVA on the RAU-transformed scores in the multiple-pitch conditions, considering tone type (pure tones, complexes at 0.5–2 kHz, and complexes at

FIG. 5. (Color online) Results from experiment 3. Listeners identified the direction of a 0.5 ST pitch change between reference and target tones, for pure tones or complex tones in one of two spectral regions, with the target tone presented alone, or as the lowest, middle, or highest tone in a three-tone mixture. Error bars show ± 1 SEM.

2–8 kHz) and target location (high, middle, and low), found a main effect of tone type [$F(2,28) = 20.9$, $p < 0.001$], a main effect of target position [$F(2,28) = 4.9$, $p = 0.02$], and an interaction [$F(4,56) = 4.06$, $p = 0.006$]. *Post hoc* pairwise comparisons revealed that performance with the low- F_0 target tended to be higher than for the middle- F_0 target ($p = 0.03$), but this trend did not reach significance once a Bonferroni correction for multiple comparisons had been applied ($\alpha = 0.017$). Listeners performed significantly more poorly with complexes in the 2–8 kHz spectral region than with complexes in the 0.5–2 kHz regions or with pure tones ($p < 0.005$ in both cases). Average performance was significantly above chance for all combinations of tone type and target position ($p < 0.001$ in all cases).

D. Discussion

Complexes in the 2–8 kHz region, where the combination of tones should have resulted in unresolved components after mixing, were discriminated more poorly than complexes in the 0.5–2 kHz region, where the combination still included some resolved harmonics. This could be due to increased harmonic numbers, or to decreased resolvability. However, the fact that a 0.5 ST pitch change was discriminated at above-chance levels, even in the 2–8 kHz region, suggests that resolved harmonics are not strictly necessary for obtaining small FODLs.

Importantly, the difference in performance between single- and multiple-pitch conditions is apparent even for pure tones. Since the decrease in performance on the multiple-pitch task with pure tones cannot be attributed to peripheral resolvability, multiple-pitch performance with pure tones could be thought of as a functional ceiling for this task, taking into account higher-level limitations of attention. Performance was equivalent to this ceiling level for complex tones in the 0.5–2 kHz region, and decreased slightly in the 2–8 kHz region, but remained well above chance.

No strong effect of high-voice superiority was observed, possibly suggesting that earlier effects (Palmer and Holleran, 1994; Fujioka *et al.*, 2008; Trainor *et al.*, 2014) may have been due to spectral differences rather than relative F_0 differences between the voices. In fact, the trend we observed was in the direction of low-voice superiority, which in our study may be due to the fact that the mixtures in the low-voice condition had a higher average F_0 , and thus better resolvability and lower harmonic numbers. In addition, because our complexes were generated with equal amplitudes per component, lower- F_0 complexes had a slightly higher overall level, due to a larger number of components falling within the passband. However, these explanations seem unlikely to explain the effect, as the same trend was observed with the pure-tone stimuli.

V. GENERAL DISCUSSION

The results of our behavioral experiments generally suggest that sub-ST pitch discrimination is possible even for mixtures of three concurrent complex tones filtered into spectral regions where the mixture is unlikely to contain any spectrally resolved harmonics. The addition of simultaneous

spectrally overlapping complex tones does not eliminate the ability to discriminate small pitch changes, even though the mixture is unlikely to contain any individually resolved components. This outcome has interesting implications for rate-place and temporal models of pitch perception, which have diverging explanations for the phenomenon of sharply degraded pitch perception for harmonic complex tones when only harmonics above about the tenth are presented (Houtsma and Smurzynski, 1990). The standard explanation for this phenomenon, grounded in the rate-place code, holds that the degraded pitch perception is due to loss of resolved harmonics due to broadening auditory filters, relative to the spacing of the harmonics. The present study provides additional evidence for the argument that pitch discrimination is not strictly limited by harmonic resolvability, by more clearly dissociating resolvability from harmonic number with the use of three-complex mixtures. Listeners generally performed above chance in conditions where single complexes would be resolved, even when mixing the tones together should have led to substantially decreased resolvability.

The interpretation of these findings depends to some extent on the definition of peripheral resolvability. For example, a template-matching strategy such as that used by Larsen *et al.* (2008) may define the boundary between resolved and unresolved differently. Indeed, it would be informative to compare the human behavioral responses reported in this paper to responses to these stimuli from different models of pitch perception. Rate-place models, whether involving identification of individual peaks of excitation (e.g., Micheyl *et al.*, 2010) or matching of harmonic templates (e.g., Larsen *et al.*, 2008), may produce different responses to these stimuli than temporal models based on summary autocorrelation (Meddis and O'Mard, 1997; Bernstein and Oxenham, 2005). Differing responses from these different kinds of models may provide insight into the neural mechanisms used to extract individual pitches from a multiple-pitch mixture. Nevertheless, our definition of resolvability, illustrated in Fig. 1, was a deliberately conservative one, which considered the tenth harmonic of a single complex as still being resolved. The fact that the harmonics within the three-tone mixtures were deemed unresolved despite this conservative definition, lends support to the claim that F_0 discrimination performance remained high despite a lack of resolved harmonics in the mixture.

The present study reaches a different conclusion from that of Micheyl *et al.* (2010), who found that accurate pitch discrimination of a complex tone in the presence of one spectrally overlapping masker was only possible when the mixture contained resolved components. Because our addition of another masking complex (for a total of three) reduced resolvability even more than a single masker, we created stimuli that were less likely to contain useful rate-place information, using a similar definition as the one proposed by Micheyl *et al.*, without increasing the harmonic numbers of the components in the individual complexes. Specifically, Micheyl *et al.* (2010) found that FODLs for dyads (targets with a single masker) deteriorated from reliably below 1 ST to reliably above 1 ST as the ratio of low

cutoff to target F_0 increased from about 5 to about 10. This corresponds to the ratio of low cutoff to average component spacing in the dyad increasing from about 10 to about 20. The present study used triads to achieve a stronger dissociation between target F_0 and average component spacing. For triads, we generally observed discrimination of 1 ST changes deteriorating from ceiling to chance as the ratio of low cutoff to target F_0 increased from about 5 to about 15. This corresponds to the ratio of low cutoff to average component spacing in the triad increasing from about 15 to about 45.

We conclude that even though the presence of resolvable components predicted good performance in the study by Micheyl *et al.* (2010), the greater reduction of resolvability in our study represents a new kind of stimulus for which sub-ST pitch discrimination is possible even without resolved components. In this way, our results are more in line with the conclusions of Bernstein and Oxenham (2008), who found that mistuning the odd harmonics of a single complex tone, relative to even harmonics, led to improved pitch perception, even without improving the spectral resolvability of individual components, presumably because it led to the perception of two tones, one of which with a pitch corresponding to that of the even harmonics ($2F_0$).

The resolvability of individual components within a harmonic complex can, to some extent, be measured behaviorally (Moore *et al.*, 1984). But focusing only on resolvability misses the possibility that components may also be hidden within complexes by informational masking; they can be made more salient by mistuning (Moore *et al.*, 1986), or by pulsing of a probe tone (Bernstein and Oxenham, 2003; Moore *et al.*, 2012), in a strategy similar to the one tried in experiment 3 by pulsing a complex probe before a complex mixture. Future research could examine whether the resolvability of the complex mixtures used in the present study, when measured behaviorally in this manner, agrees with predictions of resolvability from rate-place models. If poor resolvability were observed despite accurate pitch discrimination of the complexes within the mixture, the argument would be more convincing for a spectrotemporal model, or at least a model that goes beyond a temporally averaged rate-place representation.

Our results have implications for models and computational algorithms of pitch estimation. Current computational algorithms for the estimation of a single pitch (Noll, 1967; de Cheveigné and Kawahara, 2002) are far more effective than algorithms for estimating multiple simultaneous pitches (de Cheveigné and Kawahara, 1999; Klapuri, 2008; Yeh *et al.*, 2010), with generally a higher accuracy of pitch identification for single-voice than multiple-voice algorithms. A better understanding of the mechanisms used by humans for multiple pitch estimation should lead to improved computer listening capabilities for the same tasks, if computer algorithms can be improved by implementing the same strategies used by the human auditory system.

In summary, this study expanded the set of stimuli used in psychoacoustic studies of pitch to include mixtures of three simultaneous spectrally overlapping harmonic complex tones. The results demonstrate that functional pitch perception was possible within such mixtures, even when the

stimuli were filtered to fall within the same overlapping spectral region and when it was unlikely that any spectrally resolved harmonics remained within the mixture. The results should provide a strong test of models of pitch perception and point at a way to dissociate harmonic number from spectral resolvability more clearly than has been achieved previously, while providing mixtures that are more similar to those encountered in everyday musical environments.

ACKNOWLEDGMENTS

We thank Alex Oster and John Koch for assistance in collecting the data. This work was supported by NIH Grant No. R01 DC005216.

- Beerends, J. G., and Houtsma, A. J. M. (1986). "Pitch identification of simultaneous dichotic two-tone complexes," *J. Acoust. Soc. Am.* **80**, 1048–1056.
- 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. W., and Oxenham, A. J. (2003). "Pitch discrimination of diotic and dichotic tone complexes: Harmonic resolvability or harmonic number?," *J. Acoust. Soc. Am.* **113**, 3323–3334.
- Bernstein, J. G. W., and Oxenham, A. J. (2005). "An autocorrelation model with place dependence to account for the effect of harmonic number on fundamental frequency discrimination," *J. Acoust. Soc. Am.* **117**, 3816–3831.
- Bernstein, J. G. W., and Oxenham, A. J. (2008). "Harmonic segregation through mistuning can improve fundamental frequency discrimination," *J. Acoust. Soc. Am.* **124**, 1653–1667.
- Cariani, P., and Delgutte, B. (1996). "Neural correlates of the pitch of complex tones. I. Pitch and pitch salience," *J. Neurophysiol.* **76**, 1698–1716.
- Carlyon, R. P. (1996). "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. (1998). "Comments on 'A unitary model of pitch perception' [J. Acoust. Soc. Am. **102**, 1811–1820]," *J. Acoust. Soc. Am.* **104**, 1118–1121 (1997).
- Cedolin, L., and Delgutte, B. (2005). "Pitch of complex tones: Rate-place and interspike interval representations in the auditory nerve," *J. Neurophysiol.* **94**, 347–362.
- de Cheveigné, A. (1998). "Cancellation model of pitch perception," *J. Acoust. Soc. Am.* **103**, 1261–1271.
- de Cheveigné, A., and Kawahara, H. (1999). "Multiple period estimation and pitch perception model," *Speech Commun.* **27**, 175–185.
- de Cheveigné, A., and Kawahara, H. (2002). "YIN, a fundamental frequency estimator for speech and music," *J. Acoust. Soc. Am.* **111**, 1917–1930.
- de Cheveigné, A., and Pressnitzer, D. (2006). "The case of the missing delay lines: Synthetic delays obtained by cross-channel phase interaction," *J. Acoust. Soc. Am.* **119**, 3908–3918.
- Chubb, C., Dickson, C. A., Dean, T., Fagan, C., Mann, D. S., Wright, C. E., Guan, M., Silva, A. E., Gregersen, P. K., and Kowalsky, E. (2013). "Bimodal distribution of performance in discriminating major/minor modes," *J. Acoust. Soc. Am.* **134**, 3067–3078.
- Cohen, M. A., Grossberg, S., and Wyse, L. L. (1995). "A spectral network model of pitch perception," *J. Acoust. Soc. Am.* **98**, 862–879.
- Darwin, C. J., Hukin, R. W., and Al-Khatib, B. Y. (1995). "Grouping in pitch perception: Evidence for sequential constraints," *J. Acoust. Soc. Am.* **98**, 880–885.
- Demany, L., Semal, C., and Pressnitzer, D. (2011). "Implicit versus explicit frequency comparisons: Two mechanisms of auditory change detection," *J. Exp. Psychol. Hum. Percept. Perform.* **37**, 597–605.
- Feng, L., and Oxenham, A. J. (2015). "New perspectives on the measurement and time course of auditory enhancement," *J. Exp. Psychol. Hum. Percept. Perform.* **41**, 1696–1708.
- Feng, L., and Oxenham, A. J. (2018). "Auditory enhancement and the role of spectral resolution in normal-hearing listeners and cochlear-implant users," *J. Acoust. Soc. Am.* **144**, 552–566.

- Fujioka, T., Trainor, L. J., and Ross, B. (2008). "Simultaneous pitches are encoded separately in auditory cortex: An MMNm study," *Neuroreport* **19**, 361–366.
- Glasberg, B. R., and Moore, B. C. J. (1990). "Derivation of auditory filter shapes from notched-noise data," *Hear. Res.* **47**, 103–138.
- 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., and Smurzynski, J. (1990). "Pitch identification and discrimination for complex tones with many harmonics," *J. Acoust. Soc. Am.* **87**, 304–310.
- Klapuri, A. (2008). "Multipitch analysis of polyphonic music and speech signals using an auditory model," *IEEE Trans. Audio. Speech. Lang. Process.* **16**, 255–266.
- Kohlerausch, A., and Houtsma, J. M. (1992). "Pitch related to spectral edges of broadband signals," *Philos. Trans. Biol. Sci.* **336**, 375–382.
- 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.
- Licklider, J. C. R. (1951). "A duplex theory of pitch perception," *Experientia* **7**, 128–134.
- McDermott, J. H., Lehr, A. J., and Oxenham, A. J. (2010). "Individual differences reveal the basis of consonance," *Curr. Biol.* **20**, 1035–1041.
- Meddis, R., and Hewitt, M. J. (1991). "Virtual pitch and phase sensitivity of a computer model of the auditory periphery I: Pitch identification," *J. Acoust. Soc. Am.* **89**, 2866–2882.
- Meddis, R., and O'Mard, L. (1997). "A unitary model of pitch perception," *J. Acoust. Soc. Am.* **102**, 1811–1820.
- Mednicoff, S., Mejia, S., Rashid, J. A., and Chubb, C. (2018). "Many listeners cannot discriminate major vs minor tone-scrambles regardless of presentation rate," *J. Acoust. Soc. Am.* **144**, 2242–2255.
- Micheyl, C., Bernstein, J. G. W., 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., Keebler, M. V., and Oxenham, A. J. (2010). "Pitch perception for mixtures of spectrally overlapping harmonic complex tones," *J. Acoust. Soc. Am.* **128**, 257–269.
- Moore, B. C. J., Glasberg, B. R., and Oxenham, A. J. (2012). "Effects of pulsing of a target tone on the ability to hear it out in different types of complex sounds," *J. Acoust. Soc. Am.* **131**, 2927–2937.
- Moore, B. C. J., Glasberg, B. R., and Peters, R. W. (1986). "Thresholds for hearing mistuned partials as separate tones in harmonic complexes," *J. Acoust. Soc. Am.* **80**, 479–483.
- Moore, B. C. J., Glasberg, B. R., and Shailer, M. J. (1984). "Frequency and intensity difference limens for harmonics within complex tones," *J. Acoust. Soc. Am.* **75**, 550–561.
- Moore, B. C. J., Huss, M., Vickers, D. A., Glasberg, B. R., and Alcántara, J. I. (2000). "A test for the diagnosis of dead regions in the cochlea," *Br. J. Audiol.* **34**, 205–224.
- Neff, D., Dethlefs, T., and Jesteadt, W. (1993). "Informational masking for multicomponent maskers with spectral gaps," *J. Acoust. Soc. Am.* **94**, 3112–3126.
- Noll, A. M. (1967). "Cepstrum pitch determination," *J. Acoust. Soc. Am.* **41**, 293–309.
- Oxenham, A. J. (2018). "How we hear: The perception and neural coding of sound," *Ann. Rev. Psychol.* **69**, 27–50.
- Oxenham, A. J., and Dau, T. (2001). "Modulation detection interference: Effects of concurrent and sequential streaming," *J. Acoust. Soc. Am.* **110**, 402–408.
- Oxenham, A. J., Fligor, B. J., Mason, C. R., and Kidd, G. (2003). "Informational masking and musical training," *J. Acoust. Soc. Am.* **114**, 1543–1549.
- Palmer, C., and Holleran, S. (1994). "Harmonic, melodic, and frequency height influences in the perception of multivoiced music," *Percept. Psychophys.* **56**, 301–312.
- Parncutt, R., Reisinger, D., Fuchs, A., and Kaiser, F. (2019). "Consonance and prevalence of sonorities in Western polyphony: Roughness, harmonicity, familiarity, evenness, diatonicity," *J. New Music Res.* **48**, 1–20.
- Plack, C. J., and Oxenham, A. J. (2005). "The psychophysics of pitch," in *Pitch Neural Coding Perception* (Springer, New York), pp. 7–55.
- Plomp, R. (1964). "The ear as a frequency analyzer," *J. Acoust. Soc. Am.* **36**, 1628–1636.
- Plomp, R. (1965). "Detectability threshold for combination tones," *J. Acoust. Soc. Am.* **37**, 1110–1123.
- Pollack, I. (1975). "Auditory informational masking," *J. Acoust. Soc. Am.* **57**, S5–S5.
- Shackleton, T. M., and Carlyon, R. P. (1994). "The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination," *J. Acoust. Soc. Am.* **95**, 3529–3540.
- Shamma, S. A., and Klein, D. (2000). "The case of the missing pitch templates: How harmonic templates emerge in the early auditory system," *J. Acoust. Soc. Am.* **107**, 2631–2644.
- Shera, C. A., Guinan, J. J., and Oxenham, A. J. (2002). "Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements," *Proc. Natl. Acad. Sci. U.S.A.* **99**, 3318–3323.
- Shinn-Cunningham, B. G., Lee, A. K. C., and Oxenham, A. J. (2007). "A sound element gets lost in perceptual competition," *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12223–12227.
- Studebaker, G. A. (1985). "A 'rationalized' arcsine transform," *J. Speech Hear. Res.* **28**, 455–462.
- Sumner, C. J., Wells, T. T., Bergevin, C., Sollini, J., Kreft, H. A., Palmer, A. R., Oxenham, A. J., and Shera, C. A. (2018). "Mammalian behavior and physiology converge to confirm sharper cochlear tuning in humans," *Proc. Natl. Acad. Sci. U.S.A.* **115**(44), 11322–11326.
- Terhardt, E. (1974). "Pitch, consonance, and harmony," *J. Acoust. Soc. Am.* **55**, 1061–1069.
- Trainor, L. J., Marie, C., Bruce, I. C., and Bidelman, G. M. (2014). "Explaining the high voice superiority effect in polyphonic music: Evidence from cortical evoked potentials and peripheral auditory models," *Hear. Res.* **308**, 60–70.
- Verschooten, E., Desloovere, C., and Joris, P. X. (2018). "High-resolution frequency tuning but not temporal coding in the human cochlea," *PLoS Biol.* **16**(10), e2005164.
- Wang, J., Baer, T., Glasberg, B. R., Stone, M. A., Ye, D., and Moore, B. C. J. (2012). "Pitch perception of concurrent harmonic tones with overlapping spectra," *J. Acoust. Soc. Am.* **132**, 339–356.
- Wightman, F. L. (1973). "The pattern-transformation model of pitch," *J. Acoust. Soc. Am.* **54**, 407–416.
- Yeh, C., Roebel, A., and Rodet, X. (2010). "Multiple fundamental frequency estimation and polyphony inference of polyphonic music signals," *IEEE Trans. Audio, Speech Lang. Process.* **18**, 1116–1126.