# On musical interval perception for complex tones at very high frequencies

Hedwig E. Gockel and Robert P. Carlyon
Cambridge Hearing Group
MRC Cognition and Brain Sciences Unit, University of Cambridge, 15 Chaucer Rd., Cambridge CB2 7EF, UK

Email addresses of all authors:
Hedwig.gockel@mrc-cbu.cam.ac.uk, bob.carlyon@mrc-cbu.cam.ac.uk

Running title: Musical pitch at high frequencies

Corresponding author:
Hedwig E. Gockel
MRC Cognition and Brain Sciences Unit, University of Cambridge, 15 Chaucer Road, Cambridge CB2 7EF, UK
E-mail : hedwig.gockel@mrc-cbu.cam.ac.uk
Tel : +44 1223 769488
Fax : +44 1223 300984

# ABSTRACT

Listeners appear able to extract a residue pitch from high-frequency harmonics for which phase locking to the temporal fine structure is weak or absent. The present study investigated musical interval perception for high-frequency harmonic complex tones using the same stimuli as Lau, Mehta, and Oxenham [J. Neurosci. **37**, 9013-9021 (2017)]. Nine young musically trained listeners with especially good high-frequency hearing adjusted various musical intervals using harmonic complex tones containing harmonics 6-10. The reference notes had fundamental frequencies (F0s) of 280 or 1400 Hz. Interval matches were possible, albeit markedly worse, even when all harmonic frequencies were above the presumed limit of phase locking. Matches showed significantly larger systematic errors and higher variability and subjects required more trials to finish a match for the high than for the low F0. Additional absolute pitch judgements from one subject with absolute pitch, for complex tones containing harmonics 1-5 or 6-10 with a wide range of F0s, were perfect when the lowest frequency component was below about 7 kHz, but at least 50% of responses were incorrect when it was 8 kHz or higher. The results are discussed in terms of the possible effects of phase-locking information and familiarity with high-frequency stimuli on pitch.

**Key words**: musical interval adjustment, absolute pitch, phase locking, tonotopic information.

# I. INTRODUCTION

It has been widely argued that the perception of tone chroma, and especially of musical intervals, depends at least partly on the use of information derived from the pattern of phase locking in the auditory nerve (Cariani and Delgutte, 1996; Meddis and O'Mard, 1997; de Cheveigné, 1998). If this is the case, then the ability to judge and match musical intervals should be markedly worse for complex tones whose frequency components fall at very high frequencies (≥8.4 kHz in the context of the present study), for which phase locking is weak or absent (Johnson, 1980; Palmer and Russell, 1986). The present study tested this prediction by assessing the ability of musically trained listeners to adjust the fundamental frequency (F0) of complex tones so that there was a specific musical interval between them, using complex tones with harmonics in two frequency regions; a low frequency region where phase locking is robust and a high frequency region where phase locking is usually assumed to be severely reduced or absent. Interval-adjustment tasks provide arguably the most demanding test of musical pitch perception and can provide information both on consistency and biases in pitch perception.

The exact upper limit of phase locking in the auditory nerve (AN) in humans is unknown and consensus on this is currently lacking (Verschooten *et al.*, 2019). Phase locking has generally been assumed to be weak or absent for frequencies above about 4-5 kHz (Johnson, 1980; Palmer and Russell, 1986; Weiss and Rose, 1988). However, some studies have suggested that weak phase locking to temporal fine structure (TFS) might be available for frequencies up to about 7-8 or possibly even 10 kHz, with the usable limit depending, among other things, on the task used (Heinz *et al.*, 2001; Moore and Sek, 2009; Kale and Heinz, 2012; Moore and Ernst, 2012). Others argue for a limit around 3.5-4.5 kHz in the AN, with a much lower limit of about 1.4 kHz as the highest frequency usable by the central nervous system (Joris and Verschooten, 2013; Verschooten *et al.*, 2015; Verschooten *et al.*, 2018).

While the predominant view is that perception of musical pitch relies at least partly on the presence of phase locking in the AN, there is some evidence indicating that musical pitch might be perceived in the absence of phase locking. For pure-tone stimuli, Ward (1954) found that while most subjects were unable to adjust the frequency of one tone to be one octave higher than that of a reference tone when the reference frequency was above 2.7 kHz, two of his subjects were able to do so even when the reference frequency was 5 kHz, and thus the octave match was around 10 kHz, where phase locking was assumed to be absent. However, subjects needed more time at these high frequencies than at the lower frequencies. Similarly, all three musically trained subjects of Burns and Feth (1983) were able to adjust various musical intervals for reference frequencies of 1 and 10 kHz. However, the within-subject standard deviations (SDEVs) of the adjustments were about 3.5-5.5 times larger for the 10-kHz than for the 1-kHz reference tone. Thus, experiments with pure tones have indicated that, although musical pitch perception may be possible at very high frequencies, performance in pitch-related tasks is usually much worse than at lower frequencies, where phase locking is assumed to be strong.

Reasonably good pitch perception has been observed in experiments using complex tones consisting of only high-frequency components but with a “missing fundamental” frequency that is much lower. Oxenham *et al.* (2011) showed that even when all audible harmonics were above 6 kHz, a residue pitch (a pitch corresponding to the missing fundamental) was evoked, and melody discrimination for the high-frequency complex tones was as good as that for low-frequency pure tones. Carcagno *et al.* (2019) also observed good performance in a melody discrimination task for high-frequency complex tones with all audible frequency components above 6 kHz, and reported that the pattern of consonance ratings of various musical intervals for complex-tone dyads was similar to (albeit less distinct than) that observed for the same notes with lower frequency components.

Lau *et al.* (2017) used complex tones whose lowest component had an even higher frequency (at or above 8.4 kHz). They measured difference limens for fundamental frequency (F0DLs) and difference limens for frequency (FDLs) for the individual harmonics presented in isolation. They observed surprisingly small F0DLs (around 5%) given that the FDLs were much larger (around 20-30%), and argued that this could be explained by the existence of central harmonic template neurons that receive rate-place information. Gockel and Carlyon (2018) and Gockel *et al.* (2020) reported even smaller F0DLs (around 2%) for the same complex tones as those used by Lau *et al.* (2017). However, neither study assessed whether these tones were able to convey musical pitch.

The objective of the current study was to assess musical pitch perception in a stricter sense for complex tones having all components at or above 8.4 kHz. To do this subjects were required to make musical interval adjustments, and, for one subject, absolute pitch judgements. Musical interval adjustments are generally thought of as a stricter test of pitch perception than F0 discrimination or pitch matches to unison, since accurate musical interval judgments require precise frequency-ratio information and not just the ordinal properties of pitch (see e.g. Burns and Feth, 1983). Furthermore, a musical interval adjustment task is likely to be more sensitive to changes in pitch salience than a melody discrimination task, because a change in melody might be detected even if the size of the musical intervals is not precisely perceived. The mean error and the variability of the musical interval adjustments as well as the time (the number of trials) needed to make the adjustments was analyzed. Performance for these high-frequency complex tones was compared with that for lower frequencies, measured for the same subjects. If performance for the high-frequency complex tones was found to be not markedly worse than that for the low-frequency complex tones, this would extend previous results on musical pitch for complex tones to a higher frequency region. Relative performance in the two frequency regions would indicate the relative salience of musical pitch in a low frequency region and in high frequency region where phase locking is presumed to be very weak or absent.

# II. METHODS

## A. Subjects

Nine young normal-hearing musically trained subjects (5 females and 4 males) between 19 and 28 years of age (mean age of 22.1 years) participated in the experiment proper; many more were initially screened (see below). One of them had absolute pitch, i.e. was able to name notes without a reference (Bachem, 1937). None of them was a professional musician. The average number of years of musical training/practice was 16 (ranging from 13-21 years). Subjects 1, 2, 3, 8 and 9 started playing the violin or cello from age 7 years or earlier, and had played for at least 10 years. Subject 9, who had absolute pitch, started violin and piano training at the age of 3 years and had played for about 11 years. Subjects 2, 4, 5, and 7 started playing piano from age 7, 5, 8 and 9 years, and had played for at least 12 years. All of them except subject 4 had singing lessons for at least 6 years and most of them were still singing in choirs.

To ensure audibility of the high-frequency tones and basic pitch-discrimination ability, subjects had to pass a three-stage screening, as in Lau *et al.* (2017) and Gockel *et al.* (2020), to be eligible for the main part of the study: (1) Pure-tone audiometric thresholds at 0.25, 0.5, 1, 2, 4, 6, and 8 kHz had to be < 20 dB HL. (2) Masked thresholds were measured for 210-ms pure tones at 10, 12, 14 and 16 kHz in a continuous threshold-equalizing noise (TEN; Moore *et al.*, 2000), extending from 0.02 - 22 kHz. At 1 kHz, the TEN had a level of 45 dB SPL/ERB$_N$, the same as used in the experiment (see below), where ERB$_N$ stands for the average value of the equivalent rectangular bandwidth of the auditory filter for young normal-hearing listeners tested at low sound levels (Glasberg and Moore, 1990). Masked thresholds had to be ≤ 45 dB SPL up to 14 kHz, and ≤ 50 dB SPL at 16 kHz. (3) F0DLs and FDLs for the same stimuli as in the main experiment but without the TEN and without level randomization had to be < 6% and < 20% in the low and high frequency regions, respectively (see below). The geometric mean DLs for those subjects who passed the screening were 0.29% across frequencies in the low frequency region and 2.5% across frequencies in the high spectral region. These values were smaller than the mean DLs reported for a similar initial pitch-discrimination screening in Lau *et al.* (2017) by factors of 1.9 and 1.8 for the low- and high spectral regions, respectively. Some of the subjects took part in some other experiment(s) involving high-frequency tones, not presented here, before data collection for the present study commenced, and thus had some previous experience with high-frequency tones. All subjects confirmed that they were familiar with musical intervals and that they had learned them as part of their musical training. There was no additional screening for the ability of subjects to perform musical interval adjustments, as the relevant outcome was the within-subject comparison between performance in the high and low frequency regions.

Initially 29 musically trained subjects between 19-28 years old were tested, nine of whom passed all screening stages. Three dropped out at the first stage, 13 at the second stage, and four at the last stage of the screening. Informed consent was obtained from all subjects. This study was carried out in accordance with the UK regulations governing biomedical research and was approved by the Cambridge Psychology Research Ethics Committee.

## B. Screening procedure

Pure-tone audiometric thresholds in quiet were measured at octave frequencies from 0.25 kHz to 8 kHz and at 6 kHz, using a Midimate 602 audiometer (Madsen Electronics, Minneapolis, MN). Masked thresholds for the high-frequency (> 8 kHz) 210-ms pure tones (including 10-ms onset and offset hanning-shaped ramps) were measured for each ear using a two-interval two-alternative forced-choice task (2I-2AFC) with a 3-down 1-up adaptive procedure estimating the 79.4% correct point on the psychometric function (Levitt, 1971). The step size was 5 dB until two reversals occurred and 1 dB thereafter. The adaptive track terminated after 10 reversals, and the threshold was determined as the mean of the levels at the last six reversal points. The final threshold was the mean of the thresholds from three adaptive tracks.

F0DLs were measured in quiet for diotically presented complex tones containing harmonics 6-10 with an F0 of 280 or 1400 Hz (the same tones as used in the main experiment, i.e. with edge component levels that were 6 dB below that of the other components, but without level randomization; see below), and FDLs were measured for the components of the complex tones presented in isolation. A 2I-2AFC task with a 3-down 1-up adaptive procedure was used. Subjects had to indicate the tone with the higher pitch. For both F0DLs and FDLs, the nominal F0 or frequency was fixed within a given adaptive run, but varied across adaptive runs. The F0s (or frequencies) of the two tones presented within a trial were geometrically centered on the nominal F0 (or frequency). The signal duration was 210 ms (including 10-ms onset and offset hanning-shaped ramps) and the inter-stimulus interval (ISI) was 500 ms. Initially, the difference in F0 (or frequency) was 20%. This was reduced (or increased) by a factor of two for the first two reversals, by a factor of $\sqrt{2}$ for the next two reversals and by a factor of 1.2 thereafter. The adaptive track terminated after 12 reversals, and the threshold was determined as the geometric mean of the frequency differences at the last eight reversal points. The final threshold was the geometric mean of the thresholds from three adaptive tracks.

## C. Musical interval adjustments

Subjects had to adjust the F0 of a complex tone so that its pitch was a certain musical interval (target interval) below that of a preceding reference complex tone. Target intervals were a perfect fifth (“Fifth”, 7 semitones down), a major third (“Third”, 4 semitones down) and a major second (“Second”, 2 semitones down). In addition, subjects were asked to match to Unison. The reference tones had an F0 of 1400 Hz (“High”) or 280 Hz (“Low”), and all complex tones (reference and adjustable) contained harmonics 6-10 only. The frequency of the lowest component was 8400 Hz for the 1400-Hz F0 reference, so phase locking should have been absent or very weak, while in the Low-F0 condition phase locking should have been strong. The errors and the variability of the musical interval adjustments for the 1400-Hz and the 280-Hz F0 were compared. Also, the number of trials taken to make a match, i.e. the number of times subjects listened to the stimuli, was used as an indicator of the degree of difficulty (Ward, 1954; Cardozo, 1965; Gockel and Carlyon, 2016).

The reference tone was presented either diotically (“Dio”) or dichotically (“Dic”). For the latter, odd harmonics were presented to the left and even harmonics to the right ear. At low F0s, this manipulation is not expected to affect pitch discrimination for resolved-harmonic stimuli (Bernstein and Oxenham, 2003). While the temporal envelope rate of 1400 Hz was expected to be too high to lead to a pitch percept (Burns and Viemeister, 1976; Macherey and Carlyon, 2014), dichotic presentation of components would have reduced possible envelope cues to pitch even further due to the doubling of the frequency spacing between components in each ear, which would double the envelope repetition rate. The adjustable tone complex was always presented diotically. For each presentation, the starting phases of all components were randomized and individual component levels were randomized by ±3 dB about the mean component level, which was 55 dB SPL for harmonics 7-9 and 49 dB SPL for the two edge components. This was done to further weaken envelope cues, and to minimize edge pitches (Fastl, 1971; Klein and Hartmann, 1981). The tones were presented in a background of continuous TEN, extending from 0.02 to 22 kHz and with a level of 45 dB SPL/ERB$_N$ at 1 kHz, to mask possible distortion products. When the reference tone was diotic, the TEN was presented diotically as well, and when the reference tone was dichotic, an independent TEN was presented to each ear. These stimuli were similar to the ones used by Lau *et al.* (2017), except that they used gated rather than continuous TEN, and were identical to those used by Gockel *et al.* (2020).

One match consisted of several trials, and subjects could take as many trials as they wanted to finish a match. A match was finished when the subject indicated by button press that s/he was satisfied with the adjustment. No feedback was provided as to the precision of the adjustment. In each trial, subjects first heard the reference tone, whose F0 was fixed until the match was completed, followed by the adjustable tone. Both tones had a duration of 500 ms (including 10-ms onset and offset hanning-shaped ramps). The ISI was 500 ms. After cessation of the adjustable tone, subjects could adjust its F0 to form the desired musical interval (main task), and adjust its level to produce roughly equal loudness to that of the reference tone (in case of obvious differences in loudness) by button presses, and/or initiate the next trial. In practice, the loudness of the tones was perceived as roughly equal most times, and no level adjustments were made for most matches. Only for the unison adjustments, when the reference complex was presented dichotically, did the level adjustment of the diotic complex, averaged across subjects, reach about −1 dB. In each trial, the subject was allowed an unlimited number of button presses before s/he initiated the next trial. The number of trials taken for a match (“n_listen”) was counted, and was visible to the subject. The starting F0 of the adjustable complex was randomly chosen to be between 0.5 and 1 times the F0 of the reference tone. The F0 could be adjusted upwards or downwards via virtual button presses with mean step sizes of 4, 1, 1/4, and 1/16 semitones. The actual step size associated with each button was randomly varied across matches within the range 0.75-1.25 times the mean step size. This was done to discourage subjects from calculating – after the first sound exposure or after first matching to Unison – a sequence of button presses deemed to give the desired musical interval, rather than actually listening to and comparing the sounds in each trial. Subjects were informed about the random jitter, and it was clear from observation of the matching behaviour of the subjects and from subjects’ reports that subjects did not use this strategy[1].

Before data collection proper started, subjects received at least two hours of training in which they got accustomed to the procedure and stimuli and completed on average two matches for each of the 16 conditions (4 musical intervals × 2 F0s × 2 modes of presentation). The matches from the training were discarded. In the experiment proper, each subject completed at least 20 matches for each of the 16 conditions, which took on average 7.4 sessions of two hours each (including breaks). The number of sessions needed varied across subjects, and ranged from 5 to 10. The very slight variation in number of matches was the result of completing full 2-hour sessions. The order of conditions was randomized with the restriction that within a session no condition was repeated before a match was completed for all other conditions.

## D. Unison adjustments with non-overlapping harmonics

This was a control experiment to verify that the pitch evoked by the 1400-Hz F0 complex tone containing harmonics 6-10 corresponded to its F0, rather than, for example, to the frequency of the lower edge component. Subjects had to adjust the F0 of a complex tone containing harmonics 1-5 so that its pitch was the same as that of a reference tone. The reference tone contained harmonics 6-10 only and, for each match, its F0 was drawn randomly from a set of eight F0s, equally spaced on a logarithmic scale, and ranging from 280 to 1400 Hz. For the reference tone, individual component levels were randomized by ±3 dB about the mean component level, as for the musical interval adjustments. For the adjustable tone, the levels were not randomized. Both tones were presented diotically. Otherwise, the stimuli and methods were the same as for the musical interval adjustment experiment. Subjects needed between three and four two-hour sessions to complete at least 22 matches for each F0.

## E. Equipment

All stimuli were generated digitally in MATLAB (The Mathworks, Natick, MA) with a sampling rate of 48 kHz. Four separate stimuli were generated: two continuous background noise stimuli (one for each ear) and, for each trial, two complex tone stimuli (one for each ear); in the diotic conditions the stimuli were identical across ears. They were played out through four channels of a Fireface UCX (RME, Haimhausen, Germany) soundcard using 24-bit digital-to-analog conversion, and were attenuated independently with four Tucker-Davis Technologies (Alachua, FL) PA4 attenuators. They were mixed with two Tucker-Davis Technologies SM5 signal mixers, and fed into a Tucker-Davis HB 7 headphone driver, which also applied some attenuation. Stimuli were presented via Sennheiser HD 650 headphones (Wedemark, Germany), which have an approximately diffuse-field response. The specified sound levels are approximate equivalent diffuse-field levels. Subjects were seated individually in a double-walled, sound insulated booth (IAC, Winchester, UK).

## F. Analysis

For statistical analysis, repeated-measures analyses of variance (RM-ANOVA) were calculated using SPSS (Chicago, IL). Throughout the paper, if appropriate, the Huynh-Feldt correction was applied to the degrees of freedom (Howell, 1997). In such cases, the original degrees of freedom and the corrected significance value are reported. The Unison matches were analyzed separately from the musical interval adjustments. Before statistical analysis of the musical interval adjustments, the mean error and the within-subject SDEV of the adjustments were log-transformed to make them more normally distributed. Shapiro-Wilk tests confirmed that the (transformed) data were approximately normally distributed.

# III. RESULTS

## A. Musical interval adjustments

The expected F0 for each matched interval was determined on the equal-temperament scale; for the perfect fifth, major third, and major second, the expected F0 was exactly seven semitones (factor of 1/1.498), four semitones (factor of 1/1.26), and two semitones (factor of 1/1.122), respectively, below the F0 of the reference harmonic complex. Figures 1 and 2 show, for all subjects and conditions, the mean (across 20 or more repetitions) deviation of the adjusted F0 from the expected F0 in units of cents, where one cent is equal to 1/100th of a semitone; we refer to this value as the mean error (ME). The error bars show the within-subject SDEVs of the adjustments. Note the scale difference between the two figures; Figs. 1 and 2 show adjustments for a group of five subjects with relatively poor performance and a group of four subjects with relatively good performance, respectively. The group mean (and the corresponding SDEVs across subjects) for the MEs and for the within-subject SDEVs are shown in Fig. 3(a) and Fig. 3(b), respectively. In addition, Fig. 3(c) shows the group mean (and the corresponding SDEVs across subjects) for the absolute values of the MEs (AMEs); the AME gives, for each subject and condition, the size of the mean deviation from the target value regardless of its direction.

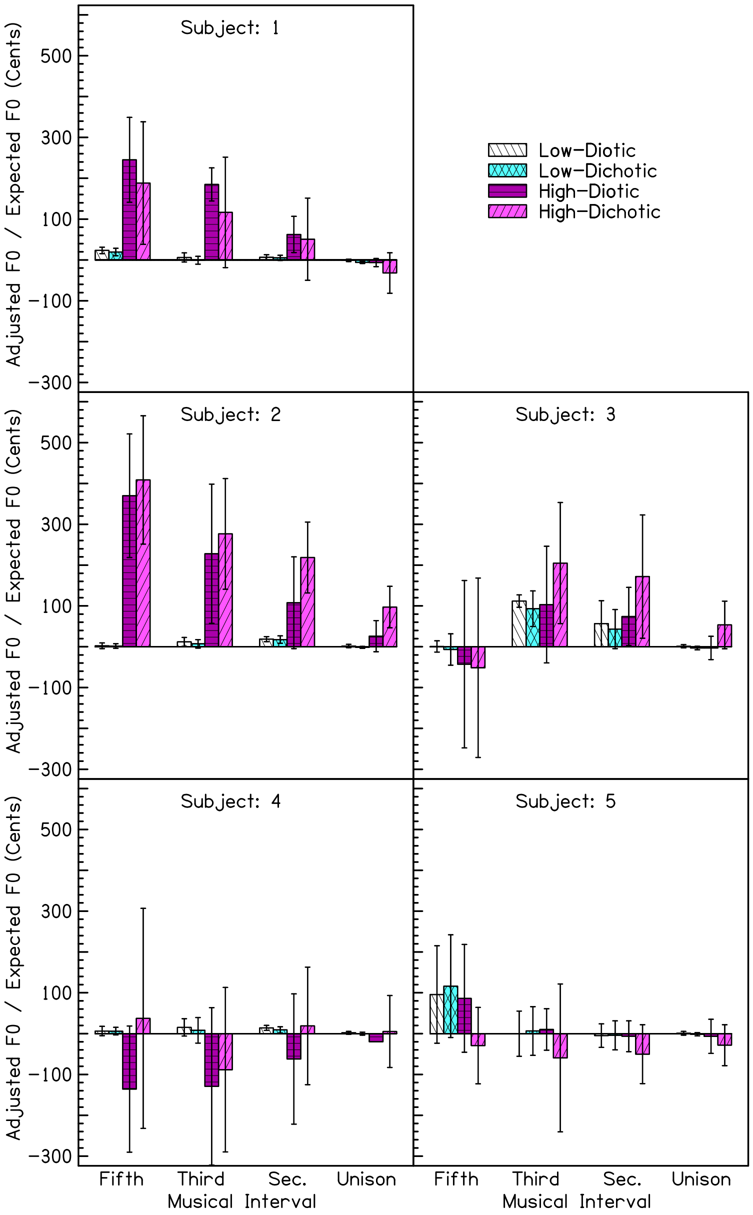

**FIG. 1.** (Color online) Mean deviation of adjusted F0 from expected F0 (in cents) for musical interval or unison adjustments for five out of nine subjects with relatively poor performance. Error bars show the within-subject SDEVs. Each group of four bars shows the results for one target musical interval. Within each group of bars, the left-hand two show results for the F0 of 280 Hz and the right-hand two show results for the F0 of 1400 Hz. All complex tones contained harmonics 6-10. In condition diotic (1st and 3rd bars in each group) all harmonics were presented diotically. In condition dichotic (2nd and 4th bars in each group)

even harmonics of the reference tone were presented to the right and odd harmonics to the left ear (see Methods).

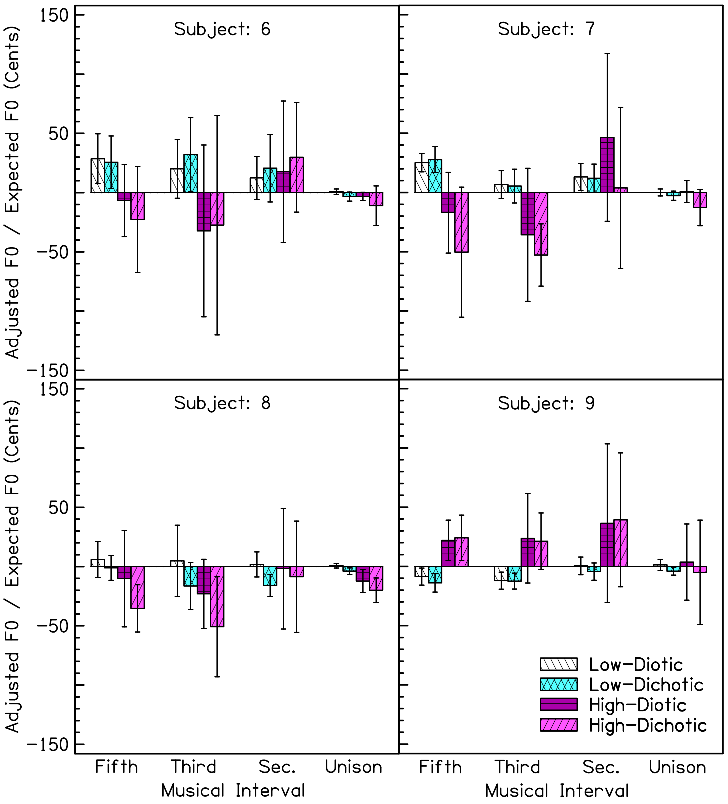

**FIG. 2.** (Color online) As Fig. 1, but for the remaining four (out of the nine) subjects, who showed better performance. Note the difference in scales between the two figures.

Musical-interval adjustments were mostly better, i.e. MEs were closer to zero and within-subject SDEVs were smaller, in the Low-F0 conditions (left two bars within each group of four bars) than in the High-F0 conditions (right two bars within each group). For the High-F0 conditions, there were large differences between subjects. For example, for subject 2 the mean adjusted F0 exceeded the expected F0 by up to 400 cents for the High-F0 perfect fifth, while in the same condition the deviation between expected and adjusted F0 was around 20 cents for subject 9, even though both subjects showed excellent performance for the Low-F0 condition. For the five subjects in Fig. 1, the mean deviation of adjusted from expected F0 often exceeded ±100 cents, mostly for the High-F0 conditions, while for subjects 6-9 in Fig. 2 they were mostly below ±100 cents. It is important to note that, for the Low-F0 conditions, all subjects were able to match all musical intervals well, with two exceptions (subject 3 for the major third and subject 5 for the fifth). Performance was often, but not always, worse for the dichotic than for the diotic reference for the High-F0 conditions.

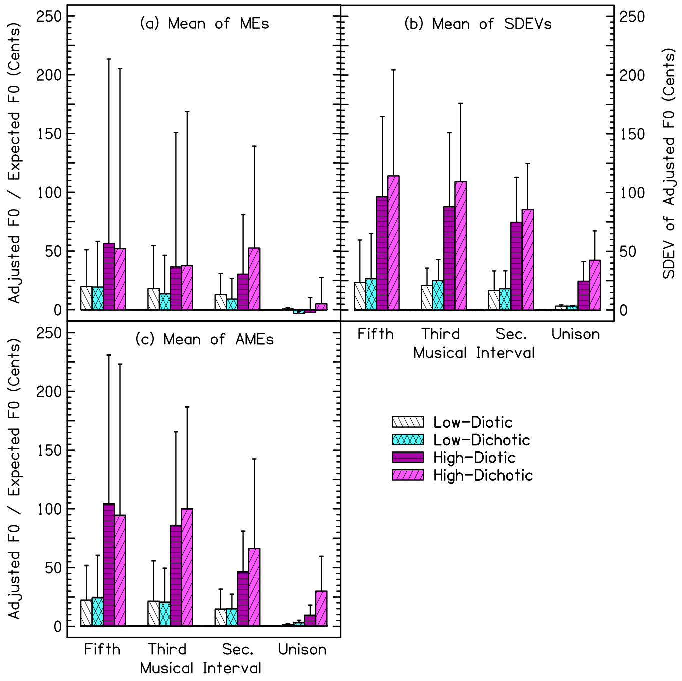

**FIG. 3.** (Color online) Group means of three measures. Error bars show SDEVs of each measure across subjects. Panel (a) shows the MEs, i.e. the systematic errors. Panel (b) shows the within-subject SDEVs. Panel (c) shows the AMEs, i.e. the absolute values of the systematic errors.

If subjects were completely unable to match musical intervals and had responded randomly, then the expected value of the adjusted F0 would be 5.3 semitones below the F0 for all conditions[2]. Thus, chance performance would lead to expected MEs of 170, –130, and –330 cents for the perfect fifth, major third, and major second, respectively. The observed MEs did not follow this pattern. In addition, the observed within-subject SDEVs were smaller than expected under the assumption of random button presses. The expected within-subject SDEV depends on the number of button presses: the more random button presses, the larger the expected SDEV. Simulations showed that for 10 and 20 random button presses the expected within-subject SDEV was about 740 and 990 cents, respectively. The observed performance was much better than this, indicating that subjects did not guess randomly in any condition.

To compare the accuracy of the musical interval adjustments across F0s, the MEs and the within-subject SDEVs of the adjustments were analyzed separately. The former is a measure of any systematic error (or bias) while the latter is a measure of the precision of the adjustments. To compare the size of the MEs across F0s, their absolute values i.e. the AMEs were used, because the interest was in the size of the mean deviation from the target value regardless of its direction. A three-way RM-ANOVA (with factors: musical interval (excluding Unison), F0 and type of presentation of the reference complex) was calculated on the log-transformed AMEs. The main effect of F0 was highly significant [$F(1,8)=18.34$, $p=0.003$]. There was no other significant main effect or interaction ($p>0.3$ in all cases). For the Unison adjustments, AMEs were also significantly larger for the High- than the Low-F0 conditions [RM-ANOVA, $F(1,8)=8.66$, $p=0.019$] and significantly larger for dichotic than diotic reference tones [$F(1,8)=6.54$, $p=0.034$]. The interaction was not significant

[F(1,8)=4.39, p=0.069]. There was no significant rank-order correlation between the (signed) MEs across F0s (Spearman's rho < 0.55 and p>0.12 for all intervals).

Consider next the variability of the matches. The within-subject SDEVs, shown by the error bars in Figs. 1 and 2, were mostly very small for the Low-F0 conditions (mean of 21.8 cents) and substantially larger for the High-F0 conditions (mean of 94.9 cents); see also Fig.3(b) for the group means of the within-subject SDEVs. Figure 4 shows, for each of the nine subjects, the ratio of the SDEV of the adjustments for the High-F0 to the SDEV for the corresponding Low-F0 condition. The geometric mean of this ratio and the standard deviation across subjects are shown in the bottom right panel. The ratios are, with few exceptions, larger than 1 and they range from about 0.75 for subject 5 for the perfect fifth to 29 for subject 4 for the perfect fifth. The few individual cases of small ratios were mostly associated with unusually large SDEVs in the corresponding Low-F0 condition as opposed to unusually small SDEVs in the High-F0 condition. For example, for subject 5 and the perfect fifth, the MEs and variability were unusually large for the low F0 (see error bars for low-F0 conditions in Figs. 1 and 2). On average (geometric mean ratio) the SDEVs were a factor of 5 larger for the High-F0 than for the Low-F0 condition. Note that subject 6, for whom the mean deviation of adjusted from expected F0 was most similar across the two F0s, produced more variable adjustments for the High-F0 than for the Low-F0 condition, like the other subjects. A three-way RM-ANOVA with factors musical interval (excluding Unison), F0 and mode of presentation, with log-transformed within-subject SDEVs as input data gave a significant main effect of F0 [F(1,8)=30.64, p=0.001]. There was no other significant main effect or interaction (p>0.12 in all cases). For the Unison adjustments, SDEVs were also significantly larger for the High-F0 than the Low-F0 [significant main effect of F0: F(1,8)=21.49, p=0.002]. In addition, there was a significant main effect of mode of presentation [F(1,8)=13.85, p=0.006], which was driven by larger SDEVs for dichotic than diotic presentation for the High-F0 but not for the Low-F0, as shown by the significant interaction between F0 and mode of presentation [ F(1,8)=13.55, p=0.006].

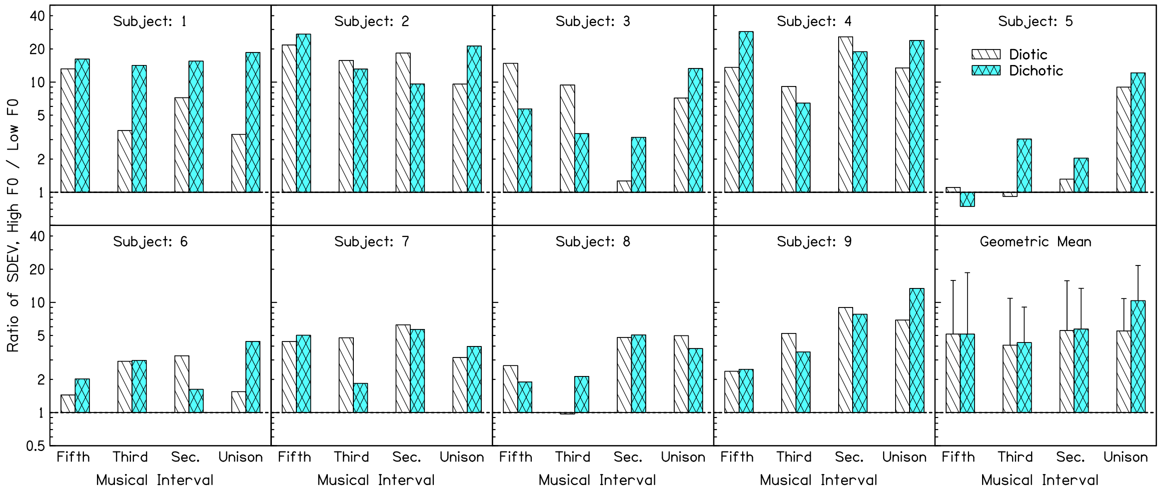

**FIG. 4.** (Color online) Ratio of the within-subject SDEVs (High F0/Low F0) of musical interval or unison adjustments (across a minimum of 20 matches for each condition). The bottom right panel shows the geometric mean (and the SDEVs) of this ratio across subjects.

Next, consider the number of trials taken to make a musical interval adjustment as an indicator of the degree of difficulty. This varied substantially across subjects, ranging from about 11 trials per adjustment (subjects 2 and 7) to about 30 trials (subject 8). Figure 5 shows the ratios of n_listen, High-F0/ Low-F0, for each condition. The ratios are mostly larger than one, indicating that subjects took longer in the High-F0 than in the corresponding Low-F0 condition to be satisfied with their musical interval adjustments. This was reflected in subjective reports; subjects described the pitch of the high-F0 (reference) tones as unclear and ambiguous. A three-way RM-ANOVA on the values of n_listen gave a significant main effect of F0 [F(1,8)=20.08, p=0.002,]. There was no other significant main effect or interaction. For the Unison adjustments, both main effects [F0: F(1,8)=17.62, p=0.003; mode of presentation: F(1,8)=32.27, p<0.001] and the interaction [F(1,8)=10.08, p=0.013] were significant; n_listen was higher for dichotic than diotic presentation, and significantly more so for the High-F0 than for the Low-F0.

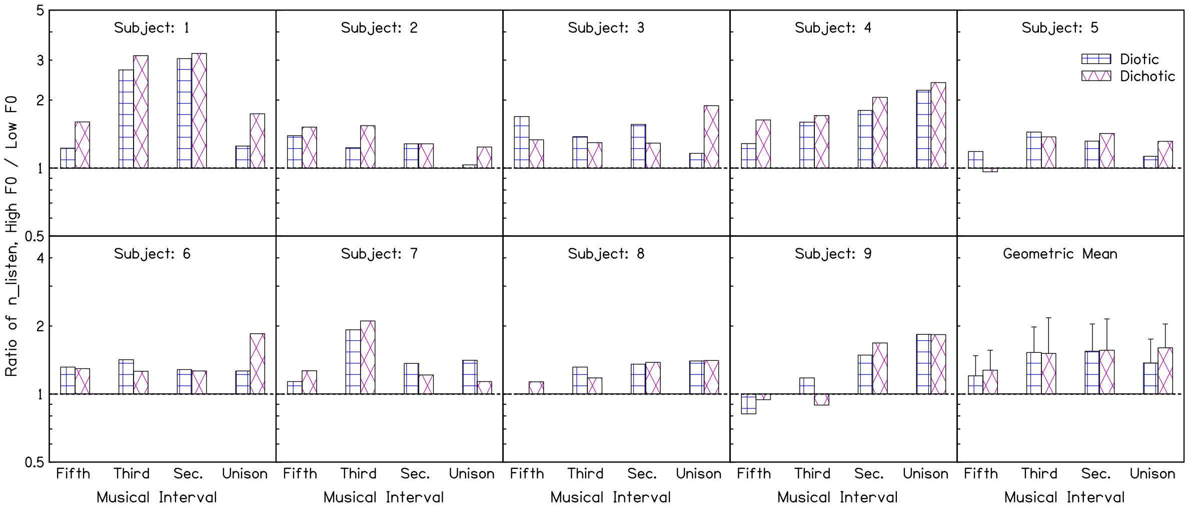

**FIG. 5.** (Color online) Ratio of the average number of trials taken to make a musical interval or unison adjustment for reference complex tones with F0s of 1400 and 280 Hz. The bottom right panel shows the geometric mean (and the SDEVs) of this ratio across subjects.

Overall the results showed that musical interval adjustments were not random. However, they were significantly more biased (had larger AMEs) and were more variable for the High-F0 than for the Low-F0, despite the fact that n_listen was usually larger for the high-F0.

## B. Unison adjustments with non-overlapping harmonics and absolute pitch judgements

It was assumed that subjects perceived a pitch corresponding to the F0 of the reference tones, even for the High-F0 conditions (see Oxenham *et al.*, 2011) and that musical interval adjustments were based on this pitch rather than the pitch of any individual harmonic. A control experiment with three subjects (subjects 5, 6, and 8), who did relatively well in the musical-interval adjustment tasks for the high F0, assessed whether the pitch of the complex tones used here did indeed correspond to its F0. Subjects adjusted the F0 of a complex tone with harmonics 1-5 to have the same pitch as a reference tone containing harmonics 6-10, with F0s ranging from 280-1400 Hz. Responses were scored as correct when they fell within ±25 cents of the reference F0 or of an F0 one or more octaves above or below the reference F0[3]. Figure 6 shows the percent correct matches as a function of the frequency of the lowest component in the reference tone. Chance performance was at 4.2% correct.

Performance ranged from good (70 to 80% correct) to very good (>95% correct) for reference complex tones whose lowest component had a frequency up to 5303 Hz. Performance worsened for all subjects when the frequency of the lowest harmonic in the complex was 6674 Hz, and became even worse for a lowest frequency of 8400 Hz, which was the same as that in the High-F0 condition of the musical interval adjustment experiment. Nevertheless, performance was above chance throughout, in agreement with the findings of Oxenham *et al.* (2011). There was no indication in the distribution of the individual matches that subjects perceived a pitch corresponding to the frequency of an individual harmonic. For the two highest F0s employed here, percent-correct values were somewhat lower than those observed by Oxenham *et al.* (2011). This is probably because in that study the individual component levels of the reference complex tone were not randomized and edge components were not reduced in level by 6 dB.

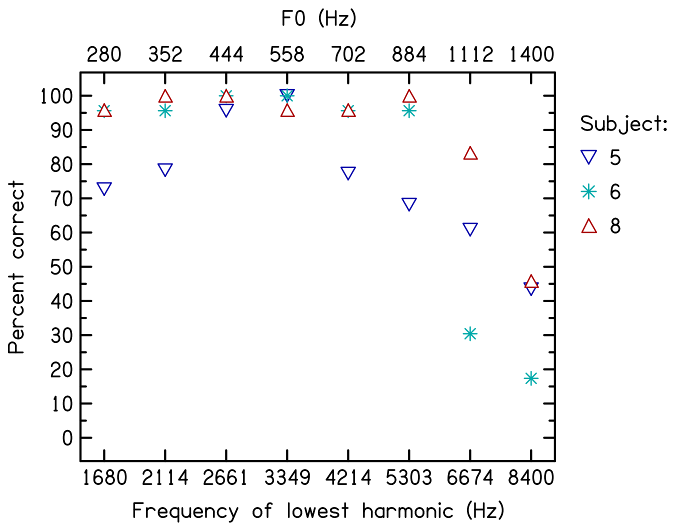

**FIG. 6.** (Color online) Average percent of pitch matches to unison, for complex tones with non-overlapping harmonics, that were within ±0.25 semitones of the F0 of the reference complex tone or one (or two) octaves below or above, as a function of the frequency of the lowest component present in the reference complex. The reference complex always contained harmonics 6-10. The variable complex contained harmonics 1-5. Chance performance corresponds to 4.2%.

Overall, these data show that the subjects perceived a pitch corresponding to the F0 rather than a pitch corresponding to an individual harmonic of the high-F0 complex. However, the pitch of the high-F0 reference note with harmonics 6-10, as employed in the musical interval adjustment experiment, was less salient than that of the low-F0 reference note.

Subject 9 possessed absolute pitch and was asked to name note chroma and the register (octave number) of the note for harmonic complex tones with a wide range of F0s and of the frequency of the lowest harmonic present in the complex (see

Appendix). Performance was perfect when the frequency of the lowest harmonic in the complex was below 7000 Hz. When the lowest frequency was at or above 7911 Hz, at least 50% of the chroma responses were incorrect. The pattern of responses indicated that the perceived pitch corresponded to the F0 of the complex. It also showed that while absolute pitch judgements were possible and perfect for medium-high component frequencies, performance markedly deteriorated when the frequency of the lowest harmonic was above about 7.5 kHz. This contrasts with the ability of the same subject to adjust musical intervals in the main experiment for a diotic reference tone whose lowest harmonic had a frequency of 8.4 kHz; the AMEs of her musical-interval adjustments were below 37 cents for all target intervals, and had a mean (excluding the unison judgements) of 27.3 cents.

# IV. GENERAL DISCUSSION

## A. Overview

In the Low-F0 conditions, most subjects were able to match musical intervals with small systematic errors and with small SDEVs for all intervals. The observed mean errors and within-subject SDEVs were similar to those reported previously for musically trained subjects (Burns and Feth, 1983; Rakowski, 1990; Burns, 1999), except for the major third for subject 3 and for the fifth for subject 5. In both cases, the adjustments were one semitone above the expected F0, leading to a smaller interval than expected, i.e. to a minor third and a diminished fifth. Subjective reports indicated that the systematic match to a minor third rather than a major third could be explained by subject 3 wrongly anchoring the reference tone as note C and, going down two notes from there on the major scale, i.e. from note C to note A. Note that the upwards major third interval corresponds to two whole-note steps from note C on the major scale. It is unclear what caused the systematic mismatch of the perfect fifth for subject 5. Musical interval adjustments were not significantly worse in the dichotic than in the diotic condition. This is in agreement with the finding that F0DLs were similar for dichotic and diotic presentation for these types of complex tones (Lau *et al.*, 2017; Gockel and Carlyon, 2018), and indicates that the (musical) pitch of these tones does not depend on the temporal envelope rate of the stimulus.

The main finding was that musical interval adjustments were possible for both F0s, even though, for the high F0, components with frequencies up to at least 9.8 kHz were required for F0 perception. For frequencies as high as this, phase locking is presumably weak or absent (Verschooten *et al.*, 2019). However, performance was clearly worse for the high than the low F0: The matches showed significantly larger systematic errors and larger within-subject SDEVs for the High-F0 than for the Low-F0 condition, despite the fact that subjects usually took more trials to make the adjustments for the former, probably because High-F0 conditions were perceived as more difficult. Thus, the poorer performance in the High-F0 condition cannot be attributed to subjects putting in less effort for this condition. On the contrary, performance likely would have been even worse in the High-F0 condition if listeners had not taken more trials in the High-F0 than the Low-F0 condition. The high-frequency complex tones clearly had a much less salient pitch than the low-frequency complex tones, and this was also obvious in the unison adjustments with non-overlapping harmonics (control experiment).

In the present study, in order to avoid distracting differences in timbre, the number of the lowest harmonic present was not roved across presentations. Conditions were designed to be as easy as possible, whilst still requiring genuine interval adjustments, as it was not *a priori* obvious how well the subjects would be able to perceive musical intervals for the High-F0 condition. Roving of the number of the lowest harmonic is sometimes employed to discourage listeners from using unwanted but useful cues based on the pitches of individual harmonics. Given that FDLs for the individual frequency components used in the High-F0 condition are substantially larger than the F0DL for the complex (Lau *et al.*, 2017; Gockel *et al.*, 2020), the pitch of an individual harmonic is unlikely to have provided a useful cue on which to base musical interval adjustments in the High-F0 condition. For the Low-F0 condition, FDLs for the individual harmonics are not smaller than the F0DL for the complex, so here too it is unlikely that musical interval adjustments would improve by using the pitch of an individual harmonic rather than that of the complex.

## B.Comparison to previous results

The present results contrast with those of Oxenham *et al.* (2011) on melody discrimination for high-frequency complex tones (their Experiment 2a). Oxenham *et al.* (2011) reported that the ability to discriminate between random melodies was equally good for high-frequency complex tones, where all audible harmonics were above 6 kHz, and for low-frequency pure tones. Several factors might contribute to the different findings. Firstly, in the present study the frequency of the lowest audible component in the complex was higher than in their study and phase locking presumably is weaker at 8.4 than at 6 kHz. Related to this, the level of the edge components was 6 dB lower than that of the inner harmonics in the present study, but not in the study of Oxenham *et al.* (2011), likely reducing the contribution of the 8.4 kHz component and shifting upwards the frequency of the most salient harmonic. Secondly, individual component levels were randomized by ±3 dB about the mean for each presentation in the present study, but not in the study of Oxenham *et al.* (2011).

Randomization of component levels might have affected the salience of the pitch of the high-frequency complex tones more than that of the low-frequency complexes, for which phase locking would be available. Thirdly, a melody discrimination task is likely to be less sensitive to changes in pitch salience than a musical interval adjustment task; a change in melody might be perceived even if the size of the musical intervals is not precisely perceived. Oxenham *et al.* (2011) also collected Unison matches between a pure tone and high-frequency complex tones (their Experiment 1) over a range of F0s and frequency regions. Performance deteriorated only when the frequency of the lowest harmonic in the complex was above 10 kHz. In the present study, Unison matches of complex tones with non-overlapping harmonics (control experiment) deteriorated for lower frequencies of the lowest harmonic present (8.4 kHz). Factors contributing to this difference might be the 6-dB decrease in the level of the edge components and the level randomization of the individual components applied in the present study, but not in the study of Oxenham *et al.* (2011).

To the best of our knowledge, there are no previous data on musical interval adjustments for high-frequency complex tones. In the following, we compare the present data with previous studies on musical interval adjustments with medium- and high-frequency pure tones. For the present high-frequency complex tones, the within-subject SDEVs of the musical interval adjustments were on average, a factor of 5 larger for the High-F0 than for the Low-F0. For the unison adjustments (main experiment), SDEVs increased on average by a factor of 5 in the diotic condition and by a factor of 10 in the dichotic condition. Presumably, unison adjustments were harder in the dichotic than the diotic condition due to the differences in timbre between the dichotic reference tone and the diotic adjusted tone in the former condition, which may have arisen from differences in suppression between components within each ear (Ruggero *et al.*, 1992) and in inhibition across ears (Boudreau and Tsuchitani, 1968).

Burns and Feth (1983) obtained musical interval adjustments for pure tones with reference frequencies of 1 and 10 kHz. Matches were less accurate for the high- than for the low-frequency tone, and the within-subject SDEVs increased on average by a factor of about 4-5, which is similar to the increase observed here. In the study of Burns and Feth (1983), musical intervals were adjusted upwards, so for the high-frequency condition both the reference tone and the adjusted tone were above 10 kHz, and thus phase locking would have been very weak or absent for both. In the present study, musical intervals were adjusted downwards to ensure audibility of the harmonics with higher ranks. Therefore, the F0 of the adjusted tone was below that of the reference tone by a factor as big as 1/1.498 for the perfect fifth, the largest musical interval used. The frequency of the lowest harmonic present in the adjusted tone complex would have been about 5.6, 6.7, and 7.5 kHz for the fifth, the major third, and the major second, respectively. The pitch of the adjustable complex probably was more salient than that of the reference complex. If we had used an upward-interval task like Burns and Feth (1983), the increase of the SDEVs might have been even larger than the observed factor of about 5. Note however that, in the present study, there was no indication that the increase in the SDEVs for the High-F0 relative to the Low-F0 condition was affected by the frequency of the lowest harmonic in the adjustable complex, as there was no significant interaction between musical interval and F0. This was presumably because performance was limited by the accuracy with which the pitch of the reference complex was encoded.

Gockel and Carlyon (2016) asked subjects to adjust pure tones downwards to form various musical intervals with a preceding Zwicker tone (ZT). A ZT is a tonal auditory afterimage that starts when a band-stop noise is turned off and can persist for 5-6 s (Zwicker, 1964). It is generally assumed to be a neural phenomenon, involving a release from neural lateral inhibition in the cochlear nucleus or higher levels in the auditory pathway, and phase locking in the AN to the frequency corresponding to the perceived pitch of the afterimage at the time of the percept is assumed to be absent (Wiegrebe *et al.*, 1995; Wiegrebe *et al.*, 1996; Gockel and Carlyon, 2016). In the study of Gockel and Carlyon (2016), the mean error of the musical interval adjustments with a ZT as reference was similar to that observed when the reference tone was a pure tone; in a first stage, the pure tones had been matched in frequency, level, and decay time so that they sounded similar to the ZTs. However, the within-subject SDEVs of the musical interval adjustments were a factor of about 1.9 larger for the ZT than for the pure tone reference, and subjects took equal time/trials to make the matches. The increase of the SDEVs relative to that in the reference condition was clearly smaller for the ZTs than for the high-frequency pure tones in the study of Burns and Feth (1983), and smaller than for the high-frequency complex tones in the present study. Note, that in the reference conditions the size of the SDEVs was very similar across the three studies (22 cents or 1.3% for the low-frequency complex tones in the present study, 20 cents or 1.2% for the pure tones ranging from 2.2 to 4.2 kHz in the ZT study, and 20 cents or 1.2% for the 1-kHz tone in the study of Burns and Feth).

While phase locking in the AN to the frequency corresponding to the perceived pitch of the ZT at the time of the percept is assumed to be absent, its relevance in the debate about the role of phase locking in pitch perception needs some qualification.

This is because for the ZT there would be phase locking to components of the band-stop noise, which might be used in creating a central rate-place representation that in turn leads to the ZT percept. This is a different situation from tones with very high frequencies, for which it is mostly assumed that phase locking is absent or very weak, and for which therefore phase locking to the stimulus at a peripheral level does not play a role either in the formation of templates or in the subsequent generation of the pitch.

Overall the present data show that while at least some of the subjects seemed to be able to adjust musical intervals for the high-frequency complex tones with “reasonable” accuracy (AMEs smaller than 53 cents and within-subject SDEVs smaller than 93 cents were observed for four of the nine subjects), performance was worse for all subjects for the High-F0 than for the Low-F0. Furthermore, the increase in SDEVs for the High-F0 relative to the Low-F0 was as large as that observed by Burns and Feth (1983) for musical interval adjustments for high frequency pure tones relative to that for low-frequency pure tones.

One of our subjects possessed absolute pitch, and additional absolute pitch judgements were collected for complex tones with a wide range of F0s and of the frequency of the lowest harmonic present. When making absolute pitch judgements, the subject listened to the stimulus only once before her response was recorded, while in the musical interval adjustment task she could listen many times before recording her response. This might have increased the difficulty of the former task, explaining why her performance for absolute pitch judgements declined more than for musical interval adjustments when the frequency of the lowest harmonic was at or above 8.4 kHz. Overall, the results of the absolute pitch judgements were very much in agreement with those of the musical interval adjustments, showing that musical pitch was much weaker for complex tones with a lowest harmonic frequency around 8.4 kHz than for complex tones with components at lower frequencies.

We are not aware of any previous data on chroma identification for high-frequency complex tones. Ohgushi and Hatoh (1992) investigated the ability of 93 music students to identify the pitch name of 1-s pure tones with frequencies corresponding to notes in the standard tempered scale ranging from C6 (1047 Hz) to C10 (16774 Hz). Up to C8 (4186 Hz), the highest note on the piano, more than 50% of all responses were correct for each tone. Above that, performance decreased markedly and so results were  broadly consistent with previous reports suggesting that musical pitch has an upper frequency limit near 5 kHz (Bachem, 1948; Ward, 1954; Attneave and Olson, 1971). However, some subjects performed above chance level beyond 5 kHz, not unlike in the study of Ward (1954), who measured octave adjustments for pure tones. Ohgushi and Hatoh (1992) showed confusion matrices for two exceptionally good subjects who could perform the task for frequencies up to about 7-8 kHz. Thus, performance for the two best subjects in Ohgushi and Hato (1992) was only slightly worse than for the present subject who named complex tones with high component frequencies, and was one of the better ones in the high-frequency musical interval task.

## C. Explanations for the deterioration in pitch perception at high frequencies

Next we consider possible explanations for our observations. The first is that the reduction (or absence) of phase locking information underlies the deterioration of performance in the high frequency region. It has been suggested that the perception of the residue pitch of complex tones containing resolved components involves some type of central harmonic template mechanism (Goldstein, 1973; Terhardt, 1974; Cohen *et al.*, 1995; Shamma and Klein, 2000). This does not mean that phase-locking information is not necessary or discarded. For example, Goldstein (1973) explicitly did not rule out the use of phase-locking information as the measure of the constituent frequencies of complex-tone stimuli in his optimum processor theory, while the model of Shamma and Klein (2000) requires exposure to sounds within the phase-locking range for the harmonic templates to initially form; frequencies for which there is no phase-locking do not contribute to the formation of a template and thus would not activate it at a later time.

The present stimuli were similar to the ones used by Lau *et al.* (2017). They observed surprisingly small F0DLs (around 5%), given that the FDLs were much larger (around 20-30%). They argued that these results could be explained by the existence of central harmonic template neurons that receive rate-place information. A single high-frequency component will not (or only weakly) activate this central template neuron, but a series of harmonics will, and so can lead to a pitch percept. There is some physiological evidence for the existence of neurons that might serve this role. Feng and Wang (2017) reported single-unit sensitivity in the auditory cortex of marmosets to harmonic structure, i.e. higher firing rates to a combination of harmonically related components than to an individual component, across the entire range of hearing, beyond the limits of peripheral phase locking. If one assumes that the pitch of complex tones is mediated by a central harmonic template mechanism, then the present results together with the findings of Lau *et al.* could be explained either by assuming that central harmonic templates get less activated by stimuli with components above the limits of phase locking because temporal fine structure information, when it is  available, provides a “better” input than purely spectral information, and/or by assuming a relative paucity of central harmonic templates receiving input from stimuli above the limits of phase locking because these high frequency input pathways have never been formed due to weak or absent phase locking in this high frequency region (Shamma and Klein, 2000).

Overall, the present results are consistent with a role of phase locking information in the production of a salient musical pitch percept that supports precise musical-interval perception. However, while phase locking information might be beneficial, it seems not to be strictly necessary to evoke a musical pitch of complex tones since all subjects performed above chance and some subjects achieved reasonable levels of performance. The latter conclusion is based on the assumption that there is no usable phase-locking information for frequencies above about 8.4 kHz (if phase locking information about all harmonics is supposed to be absent) or above about 9.8 kHz (if phase locking information for all but the lowest harmonic is supposed to be absent). As described in the introduction, whether or not this is the case is still under debate (Verschooten *et al.*, 2019). For their pure tone data, Burns and Feth (1983) concluded that their “results were not incompatible with a temporal basis” and noted that Goldstein and Srulovicz (1977) “have recently demonstrated that there is sufficient temporal information in eighth-nerve firing patterns to explain psychophysical frequency DLs at high frequencies. It is not necessary, therefore, to postulate that a separate (tonotopic) mechanism mediates discrimination above 5 kHz”.  Heinz, in Verschooten *et al.* (2019) noted “the degredation in frequency-discrimination performance as frequency increases is consistent with the ability of human listeners to use phase-locking information at high frequencies (up to ~10000 Hz)”. In contrast, Joris and Verschooten in Verschooten *et al.* (2019) argued for an upper limit of phase locking in the AN of humans of about 3.5-4.5 kHz, with a much lower limit of about 1.4 kHz as the highest frequency usable by the central nervous system. Either way, the present results contribute to the growing evidence that musical interval perception is possible with either very weak or absent phase locking, but they also show that performance is worse for these very high frequencies.

Another possible explanation for the deterioration of performance at very high frequencies is lack of familiarity with high-frequency tones. Studies of the pitch of pure tones have often used this reasoning (Ward, 1954; Attneave and Olson, 1971). Gockel and Carlyon (2016) mentioned that this might have contributed to the finding that musical interval adjustments were more precise for the ZTs, which had a lower pitch (matched frequencies between 2.2-4.2 kHz ) than for the high-frequency pure tones of Burns and Feth (1983). However, for the high-frequency complex tones used here, the F0 was relatively low at 1.4 kHz, and so the pitch itself would not be unfamiliar. Furthermore, there is at least one study that casts doubt on an explanation in terms of lack of familiarity and lack of exposure to tones with very high F0s. Jacoby *et al.* (2019) investigated musical pitch perception for members of a remote tribe, the Tsimaneʹ, who live in relative isolation from Western culture. The F0s of their musical instruments all fall below 2000 Hz, much lower than in the Western culture where F0s reach just above 4000 Hz. Moreover, Tsimaneʹ songs typically have notes at the lower end of the F0 range of their instruments. Jacoby *et al*. (2019) assessed the accuracy of the sung reproduction of musical intervals defined by two pure tones that were presented in a wide range of registers. Despite lack of experience of the Tsimaneʹ with high-frequency tones, their accuracy of interval reproduction started to deteriorate above about 4 kHz, the same frequency as for subjects from a Western culture. As argued by Jacoby *et al*. (2019), these results are consistent with biological constraints on the upper limit of musical pitch, for example the breakdown in phase locking for higher frequencies, rather than with constraints imposed by culture and exposure. However, it cannot be ruled out that a lack of exposure to (and familiarity with) resolved components in the very high frequency region, rather than a lack of exposure to high F0s, contributes to the deterioration in performance observed in the present study. In addition, there may be other (yet undiscovered) factors that co-vary with frequency region and that may underlie the observed effects.

## V. SUMMARY AND CONCLUSIONS

The ability of musically trained subjects to adjust musical intervals for reference complex tones with an F0 of 1.4 kHz and harmonic frequencies ≥ 8.4 kHz was compared to that for reference complex tones with an F0 of 280 Hz and harmonic frequencies from 1680 Hz to 2800 Hz. There were large individual differences in performance for the high-frequency complex. Musical interval adjustments were possible for both F0s, even though for the high F0 all harmonic frequencies were above the presumed limit of phase locking. However, performance was markedly worse for the high F0. The mean error and the within-subject SDEV of the adjustments were significantly larger for the high-frequency than for the low-frequency complex even though subjects took more trials for the former to make the adjustments. Absolute pitch judgements from one of the subjects were perfect for harmonic complex tones with lower component frequencies, but deteriorated once the frequency of the lowest component exceeded 7-8 kHz. The results are consistent with the idea that the salience of musical pitch is greater for tones for which phase-locking information is available, but pitch perception at high frequencies may alternatively or additionally be degraded by a lack of exposure to the upper harmonics (the sixth and above) of complex tones with high F0s.

## ACKNOWLEDGEMENTS

This research was supported by the Medical Research Council UK (SUAG/042/G101400). We thank Brian Moore for helpful discussions and comments. In compliance with our open access requirements, data from this study are available at https://www.mrc-cbu.cam.ac.uk/publications/opendata/.

# APPENDIX

### A.1. Methods for absolute pitch judgements

Subject 9, who possessed absolute pitch, was asked to name the note chroma and the register (octave number) of the note for a wide range of stimuli. This was done by choosing one of 12 virtual chroma buttons labelled C, C#, D, D#, E, F, F#, G, G#, A, A#, or B, and one of 8 virtual register buttons labelled from 1 to 8 on the computer screen. No feedback was provided.

In the first two experiments of this type, complex tones with F0s corresponding to piano keys 39-71 (33 F0s ranging from B3=246.94 Hz to G6=1567.98 Hz in one-semitone steps) were used. Piano key 69 (F6) with an F0 of 1396.91 Hz corresponds most closely to the 1400-Hz F0 used in the musical interval adjustment tasks. The complex tones contained either harmonics 1-5 or harmonics 6-10. This allowed assessment of the effect of the lowest frequency present in the complex on absolute pitch judgements. In each trial, one of the 66 stimuli was chosen at random for presentation. Tones were presented at the same level and in the same TEN as for the musical interval adjustments. In the first experiment, the stimulus duration was 1 s and there were 20 repetitions for each condition. In the second experiment, the stimulus duration was 210 ms and there were 22 repetitions per condition.

In a third experiment, the stimulus range was extended to higher F0s and various lower harmonic ranks, to assess whether, in this extended high-F0 range, the rank of the lowest harmonic in a tone complex influences performance independently from its frequency. F0s corresponding to piano keys 72-85 (14 F0s ranging from G#6=1661.22 Hz to A7=3520 Hz in one-semitone steps) were used. The complex tones always contained five consecutive harmonics. The rank of the lowest harmonic present in a complex tone with fixed F0 was varied from 1 to 6, with the restriction that the frequency of the highest harmonic was always below 18 kHz, to ensure that at least 4 components would have been audible. This resulted in 45 complex tones, for which the frequencies of the lowest-rank harmonics ranged from 1661.22 Hz (1st harmonic of G#6) to 10560 Hz (6th harmonic of A6). The stimulus duration was 210 ms and there were 22 repetitions per condition. Nine 2-hour sessions were needed to complete all three experiments.

### A.2. Results of absolute pitch judgements

Figure 7 shows the mean deviation of the responses from the true note (in semitones) across the 20 trials completed for each condition as a function of the F0 of the 1-s stimulus (x-axis, bottom) and as a function of the frequency of the lowest harmonic present in the stimulus (x-axis, top). The left and right panels show results for the complexes containing harmonics 1-5 and 6-10, respectively. The upward-pointing blue triangles (“uncorrected”) are based on the raw response values, and give an indication of overall biases; the large negative values observed for high F0s when harmonics 6-10 were present indicate a response bias towards lower registers. The circles (“corrected, absolute”) are based on responses after correcting for possible octave confusions; all responses that differed by more than six semitones from the true note were adjusted by ± *n* octaves, where *n* was the smallest integer number that would give an absolute difference between adjusted response and true note smaller than or equal to six semitones. The mean deviations were calculated from the absolute values of the deviations between true note and octave-corrected responses. For random responses, the expected mean deviation based on these octave-corrected absolute deviations is three semitones. More systematic mistakes can produce larger or smaller mean deviations. The results show that, after correcting for possible octave confusions, performance was perfect for all F0s tested when the lower harmonics were present and for F0s up to about 1100 Hz when the higher harmonics were present. For F0s above 1100 Hz, i.e. when the lowest frequency present was above 6600 Hz, the mean deviations increased first gradually and then more steeply when the lowest frequency component fell above 7900 Hz (four right-most circles in panel b).

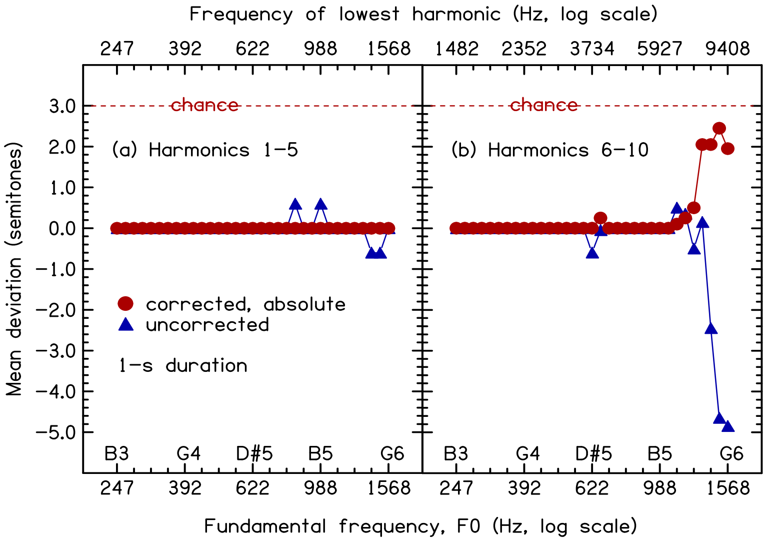

**FIG. 7.** (Color online) Results of absolute pitch judgments by subject 9 for a stimulus duration of 1 second. The mean deviation of the responses from the

"correct" note is plotted as a function of the F0 of the complex tone stimulus (the note chroma and register) on the bottom axis and as a function of the frequency of the lowest harmonic present on the top axis. The complex tone contained harmonics 1-5 (Panel a) or harmonics 6-10 (Panel b). The (red) circles are based on octave-corrected responses, while the (blue) triangles are based on uncorrected responses.

Figure 8 shows a "confusion matrix" (based on octave-corrected responses) for complex tones with harmonic ranks 6-10 for the 13 highest notes used. The color codes the number of times (out of 20) each chroma response (y-axis) occurred for a given stimulus (x-axis). Responses were 100% correct for all notes up to and including C6, for which the frequency of the lowest component fell at 6279 Hz. Once the frequency of the lowest component was at or above 7911 Hz, at least 50% of the chroma responses were incorrect. In addition, there was a bias towards responding "A".

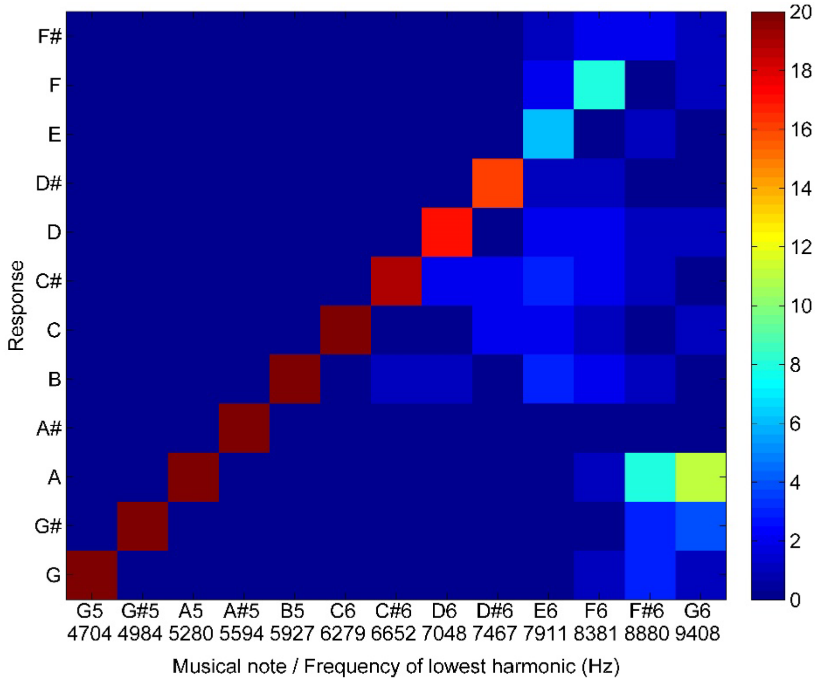

**FIG. 8.** (Color online) Confusion matrix (based on octave-corrected responses) for absolute pitch judgements of 1-s complex tones with harmonic ranks 6-10 for the 13 highest F0s shown in Fig. 7. The color codes the number of times (out of 20) each chroma response (y-axis) occurred for a given stimulus (x-axis).

The experiment was repeated with a shorter stimulus duration of 210 ms. Figures 9 and 10 show a very similar pattern of results for this duration; performance was only slightly worse. Performance deteriorated once the frequency of the lowest harmonic was above 7000 Hz and chroma identification ability appeared to have been lost for frequencies above about 8400 Hz.

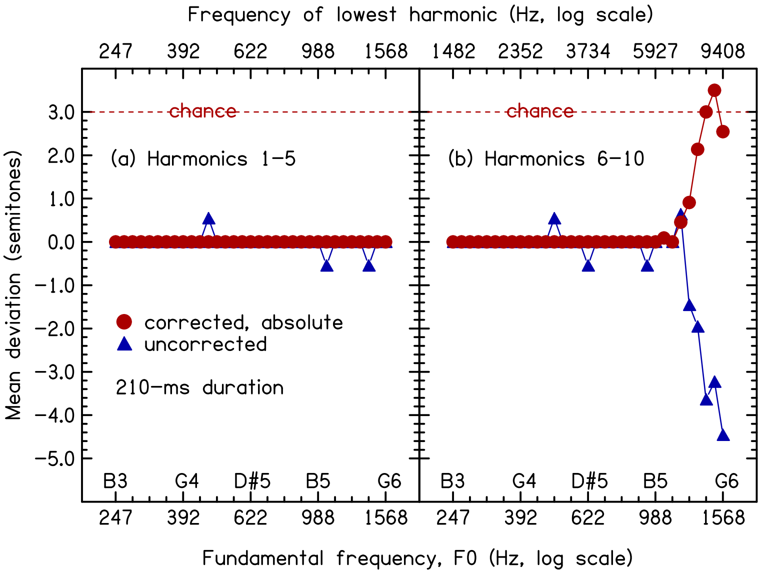

**FIG. 9.** (Color online) Results of absolute pitch judgments by subject 9 for a stimulus duration of 210 ms. Otherwise as Fig. 7.

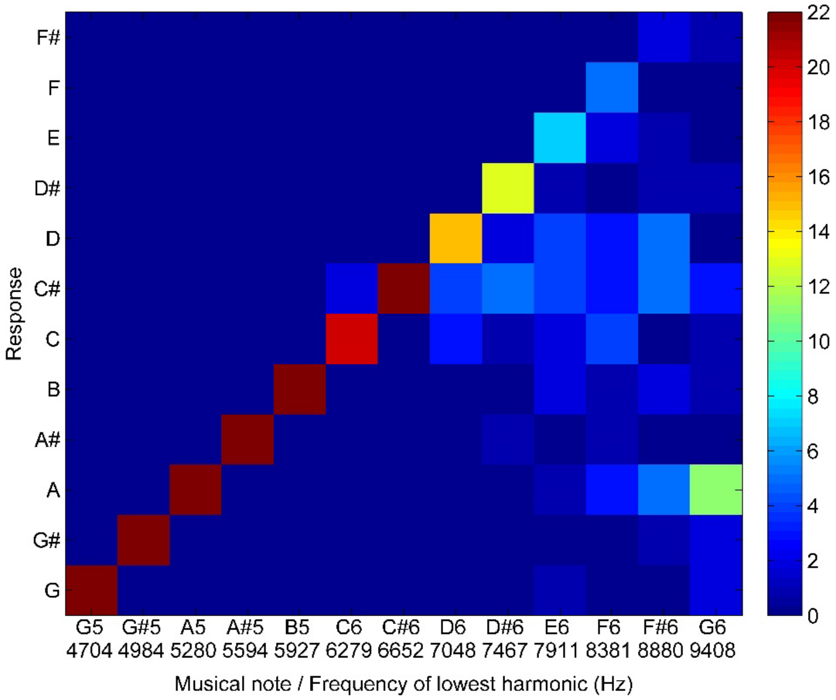

**FIG. 10.** (Color online) Confusion matrix (based on octave-corrected responses) for absolute pitch judgments of 210-ms complex tones with harmonic ranks 6-10 for the 13 highest F0s shown in Fig. 9. Otherwise as Fig. 8.

In a third experiment, a higher F0 range (14 notes from G#6=1661.22 Hz to A7=3520 Hz in one semitone steps) was used and the lowest harmonic rank was varied. Figure 11 shows the mean absolute deviation of the octave-corrected responses (across 22 trials for each condition) from the correct chroma as a function of the frequency of the lowest harmonic. Note, data points are shown only for stimuli whose lowest component had a frequency above 6 kHz; performance was perfect for complex tones with lowest-component frequencies below 6 kHz. The results of the second absolute-pitch experiment, with lowest harmonic rank equal to six, are replotted for comparison. The rank of the lowest harmonic present in the stimulus is indicated by the different symbols (see legend).

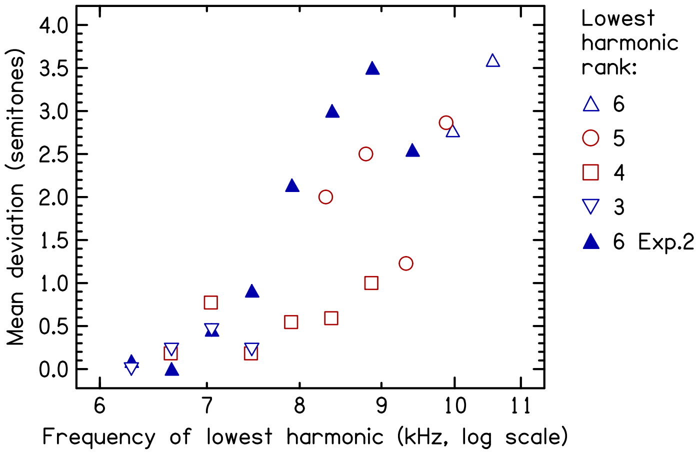

**FIG. 11.** (Color online) Results of absolute pitch judgements for the extended high-frequency range with 210-ms stimulus duration. The mean deviation of the responses from the "correct" note is plotted as a function of the frequency of the lowest harmonic present. The complex tones (the notes) always contained five consecutive harmonics, and the rank of the lowest harmonic present (see legend) and the F0 were varied.

In addition to the clear increase in deviation with increasing frequency, there was a tendency towards larger deviations with increasing harmonic rank. Unfortunately, the possible stimulus space was restricted, as frequencies above 16 kHz were unlikely to be audible, and there are not many informative comparisons between data points with different lowest harmonic rank, i.e. data points above floor and below ceiling performance levels. In addition, comparison of data points across experiments conceivably might be affected by the different context of notes tested within each experiment. Therefore, unfortunately, no clear conclusion can be drawn about the role of harmonic rank.

The main conclusion to be drawn from these absolute pitch judgements is that performance deteriorated markedly as the frequency of the lowest harmonic increased above about 7000 Hz. When that frequency was 8381 Hz (Figs. 7b and 9b, 3rd data point from the end), errors were extremely large, despite the ability of this subject to make relatively accurate musical-interval adjustments with this stimulus, with mean errors less than 30 cents, in the main part of the study (Fig. 2).

## Footnotes

1. Several additional analyses indicated that the strategy used by subjects to make musical interval adjustments was not one to first match to unison and then to adjust the F0 to a “mathematically known” ratio using a calculated sequence of button presses. This will be referred to hereafter as the “alternative strategy”. Firstly, if subjects had used the alternative strategy instead of directly matching to their “internal template” of the expected musical interval, n_listen for musical interval adjustments would be expected to be higher than n_listen for the unison matches. This was not the case. The number of trials taken for the musical interval adjustments was similar to that taken for the unison matches; the geometric mean ratio [± 1 standard deviation] across subjects (n_listen for musical interval adjustments divided by n_listen for unison matches in the corresponding condition) was 1.01 [0.79, 1.29] and 0.98 [0.85, 1.13] for the low F0 and the high F0, respectively. Secondly, if subjects had used the alternative strategy, n_listen should be higher for matches where the starting F0 was further away from unison than for matches where the starting F0 was close to unison (the starting F0 was randomly chosen between F0 and 0.5 F0): Spearman’s rank correlation, *rho*, between the starting F0 and n_listen should be negative. This also was not the case. For the four conditions that involved adjusting to a perfect fifth, *rho* was negative in 11 out of the 36 cases (9 subjects X 4 conditions), and was significant in only 1 case, i.e. in 3% of the cases. In contrast, for the four conditions where subjects had to match to unison, *rho* was negative in 29 out of the 36 cases, and was significant in 22% of the cases. Thirdly, if subjects did not use the alternative strategy, but matched directly to their “template” for the target musical interval,

n_listen should be smaller for matches where the randomly chosen starting F0 was closer to the final matched F0 than for matches where the starting F0 was further away from the matched F0. To assess this, *rho* was calculated between n_listen and the absolute difference between the random starting F0 and the final matched F0. If subjects had directly matched to the target F0, this correlation should be positive. This was the case to a similar extent for all musical intervals and for unison: For conditions that involved matching a perfect fifth, a major third, a major second and unison, *rho* was positive (significant) in 72% (22%), 69% (25%), 64% (25%) and 81% (19%) of the cases, respectively. Note that for the latter two analyses, correlations between n_listen and frequency differences were not expected to be very high as subjects probably used bigger step sizes when the perceived difference between the starting F0 and the target F0 was large than when it was small.

2. If subjects make random adjustments for each match, then the expected adjusted value corresponds to the starting F0 itself. For all conditions, the starting F0 of the adjustable complex was randomly chosen to be between 0.5 and 1 times the F0 of the reference tone (uniformly distributed on a linear frequency scale). The mean of the logarithms of all possible starting F0s is 5.3 semitones below the F0 of the reference tone.

3. Octave confusions are quite common in pitch-matching experiments (Davis *et al.*, 1951). Correcting for octave confusions by dividing or multiplying the adjusted F0 by a factor of 2, so that the adjusted F0 never differs by more than six

semitones from the true F0, allows correct chroma responses to be counted as correct while ignoring tone height (register) errors.

Attneave, F., and Olson, R. K. (**1971**). "Pitch as a medium: A new approach to psychophysical scaling," Am. J. Psychol. **84**, 147-166.

Bachem, A. (**1937**). "Various types of absolute pitch," J. Acoust. Soc. Am. **9**, 146-151.

Bachem, A. (**1948**). "Chroma fixation at the ends of the musical frequency scale," J. Acoust. Soc. Am. **20**, 704-705.

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

Boudreau, J. C., and Tsuchitani, C. (**1968**). "Binaural interaction in the cat superior olive S segment," J. Neurophysiol. **31**, 442-454.

Burns, E. M. (**1999**). "Intervals, scales, and tuning," in *The Psychology of Music*, edited by D. Deutsch (Academic Press, Amsterdam), pp. 215-264.

Burns, E. M., and Feth, L. L. (**1983**). "Pitch of sinusoids and complex tones above 10 kHz," in *Hearing - Physiological Bases and Pyschophysics*, edited by R. Klinke and R. Hartmann (Springer, Berlin), pp. 327-333.

Burns, E. M., and Viemeister, N. F. (**1976**). "Nonspectral pitch," J. Acoust. Soc. Am. **60**, 863-869.

Carcagno, S., Lakhani, S., and Plack, C. J. (**2019**). "Consonance perception beyond the traditional existence region of pitch," J. Acoust. Soc. Am. **146**, 2279-2290.

Cardozo, B. L. (**1965**). "Adjusting the method of adjustment: SD vs DL," J. Acoust. Soc. Am. **37**, 786-792.

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

Cohen, M. A., Grossberg, S., and Wyse, L. L. (**1995**). "A spectral network model of pitch perception," J. Acoust. Soc. Am. **98**, 862-879.

Davis, H., Silverman, S. R., and Mcauliffe, D. R. (**1951**). "Some observations on pitch and frequency," J. Acoust. Soc. Am. **23**, 40-42.

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

Fastl, H. (**1971**). "Über Tonhöhenempfindungen bei Rauschen ("On sensations of pitch evoked by noise")," Acustica **25**, 350-354.

Feng, L., and Wang, X. Q. (**2017**). "Harmonic template neurons in primate auditory cortex underlying complex sound processing," Proc. Natl. Acad. Sci. USA **114**, E840-E848.

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

Gockel, H. E., and Carlyon, R. P. (**2016**). "On Zwicker tones and musical pitch in the likely absence of phase locking corresponding to the pitch," J. Acoust. Soc. Am. **140**, 2257-2273.

Gockel, H. E., and Carlyon, R. P. (**2018**). "Detection of mistuning in harmonic complex tones at high frequencies," Acta Acust. united Ac. **104**, 766-769.

Gockel, H. E., Moore, B. C. J., and Carlyon, R. P. (**2020**). "Pitch perception at very high frequencies: On psychometric functions and integration of frequency information," J. Acoust. Soc. Am. **148**, 3322-3333.

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

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

Heinz, M. G., Colburn, H. S., and Carney, L. H. (**2001**). "Evaluating auditory performance limits: I. One-parameter discrimination using a computational model for the auditory nerve," Neural Computation **13**, 2273-2316.

Howell, D. C. (**1997**). *Statistical Methods for Psychology* (Duxbury, Belmont, CA), pp. 464-466.

Jacoby, N., Undurraga, E. A., McPherson, M. J., Valdes, J., Ossandon, T., and McDermott, J. H. (**2019**). "Universal and non-universal features of musical pitch perception revealed by singing," Curr. Biol. **29**, 3229-3243 e3212.

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

Joris, P. X., and Verschooten, E. (**2013**). "On the limit of neural phase locking to fine structure in humans," Adv Exp Med Biol **787**, 101-108.

Kale, S., and Heinz, M. G. (**2012**). "Temporal fine structure coding at high frequencies following noise-induced hearing loss," Assoc. Res. Otolaryngol. Abs. **35**, 364.

Klein, M. A., and Hartmann, W. M. (**1981**). "Binaural edge pitch," J. Acoust. Soc. Am. **70**, 51-61.

Lau, B. K., Mehta, A. H., and Oxenham, A. J. (**2017**). "Superoptimal perceptual integration suggests a place-based representation of pitch at high frequencies," J. Neurosci. **37**, 9013-9021.

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

Macherey, O., and Carlyon, R. P. (**2014**). "Re-examining the upper limit of temporal pitch," J. Acoust. Soc. Am. **136**, 3186-3199.

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

Moore, B. C. J., and Ernst, S. M. A. (**2012**). "Frequency difference limens at high frequencies: Evidence for a transition from a temporal to a place code," J. Acoust. Soc. Am. **132**, 1542-1547.

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," Brit. J. Audiol. **34**, 205-224.

Moore, B. C. J., and Sek, A. (**2009**). "Sensitivity of the human auditory system to temporal fine structure at high frequencies," J. Acoust. Soc. Am. **125**, 3186-3193.

Ohgushi, K., and Hatoh, T. (**1992**). "The musical pitch of high-frequency tones," in *Adv Biosci*, edited by Y. Cazals, L. Demany and K. Horner (Pergamon, Oxford), pp. 207-213.

Oxenham, A. J., Micheyl, C., Keebler, M. V., Loper, A., and Santurette, S. (**2011**). "Pitch perception beyond the traditional existence region of pitch," Proc. Natl. Acad. Sci. USA **108**, 7629-7634.

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

Rakowski, A. (**1990**). "Intonation variants of musical intervals in isolation and in musical contexts," Psychology of Music **18**, 60-72.

Ruggero, M. A., Robles, L., Rich, N. C., and Recio, A. (**1992**). "Basilar membrane responses to two-tone and broadband stimuli," Phil. Trans. R. Soc. Lond. B. **336**, 307-315.

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

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

Verschooten, E., Desloovere, C., and Joris, P. X. (**2018**). "High-resolution frequency tuning but not temporal coding in the human cochlea," Plos Biol **16**, doi: 10.1371/journal.pbio.2005164.

Verschooten, E., Robles, L., and Joris, P. X. (**2015**). "Assessment of the Limits of Neural Phase-Locking Using Mass Potentials," J. Neurosci. **35**, 2255-2268.

Verschooten, E., Shamma, S., Oxenham, A. J., Moore, B. C. J., Joris, P. X., Heinz, M. G., and Plack, C. J. (**2019**). "The upper frequency limit for the use of phase locking to code temporal fine structure in humans: A compilation of viewpoints," Hear. Res. **377**, 109-121.

Ward, W. D. (**1954**). "Subjective musical pitch," J. Acoust. Soc. Am. **26**, 369-380.

Weiss, T. F., and Rose, C. (**1988**). "A comparison of synchronization filters in different auditory receptor organs," Hear. Res. **33**, 175-179.

Wiegrebe, L., Kössl, M., and Schmidt, S. (**1995**). "Auditory sensitization during the perception of acoustical negative afterimages - Analogies to visual processing," Naturwissenschaften **82**, 387-389.

Wiegrebe, L., Kössl, M., and Schmidt, S. (**1996**). "Auditory enhancement at the absolute threshold of hearing and its relationship to the Zwicker tone," Hear. Res. **100**, 171-180.

Zwicker, E. (**1964**). "'Negative afterimage' in hearing," J. Acoust. Soc. Am. **36**, 2413-2415.