Published in final edited form as:

J Acoust Soc Am. 2016 October ; 140(4): 2257. doi:10.1121/1.4963865.

On Zwicker tones and musical pitch in the likely absence of phase locking corresponding to the pitch a)

Hedwig E. Gockel^{b)} and Robert P. Carlyon

MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 7EF, United Kingdom

Abstract

It was assessed whether Zwicker tones (ZTs) (an auditory afterimage produced by a band-stop noise) have a musical pitch. First (stage I), musically trained subjects adjusted the frequency, level, and decay time of an exponentially decaying diotic sinusoid to sound similar to the ZT they perceived following the presentation of diotic broadband noise, for various band-stop positions. Next (stage II), subjects adjusted a sinusoid in frequency and level so that its pitch was a specified musical interval below that of either a preceding ZT or a preceding sinusoid, and so that it was equally loud. For each subject the reference sinusoid corresponded to their adjusted sinusoid from stage I. Subjects selected appropriate frequency ratios for ZTs, although the standard deviations of the adjustments were larger for the ZTs than for the equally salient sinusoids by a factor of 1.0–2.2. Experiments with monaural stimuli led to similar results, although the pitch of the ZTs could differ for monaural and diotic presentation of the ZT-exciting noise. The results suggest that a weak musical pitch may exist in the absence of phase locking in the auditory nerve to the frequency corresponding to the pitch (or harmonics thereof) at the time of the percept.

I Introduction

Most periodic sounds produce periodic patterns of phase-locked activity in the auditory nerve (AN). It has been argued that this temporal code is the basis for our sensation of pitch, and specifically that musical pitch, i.e., pitch in its strictest sense, requires phase locking. In this vein, most current models of pitch perception rely on the precise timing of action potentials, or spikes, within the AN (Cariani and Delgutte, 1996b; Meddis and O'Mard, 1997; de Cheveigné, 1998). Furthermore, the precision of phase locking weakens at high frequencies, and it is usually assumed to be very weak for frequencies above 5 kHz (Johnson, 1980; Palmer and Russell, 1986). This has been used to explain the finding that the ability to recognize melodies, the accuracy of musical interval judgements, and frequency discrimination are all severely degraded for pure tones with frequencies above about 4–5 kHz (Ward, 1954; Attneave and Olson, 1971; Moore, 1973; Sek and Moore, 1995). According to such accounts, the reduced but nevertheless remaining ability to detect changes in the frequency of tones with frequencies above 4–5 kHz rests either on the use of

^{b)} hedwig.gockel@mrc-cbu.cam.ac.uk.

^{a)} Portions of this work were presented in “Do Zwicker tones evoke a musical pitch?,” in *Proceedings of the 17th International Symposium on Hearing. Physiology, Psychoacoustics and Cognition in Normal and Impaired Hearing*, edited by P. van Dijk, D. Baskent, E. Gaudrain, E. de Kleine, A. Wagner, C. Lanting (Springer, New York, 2016).

weak spike timing information, and/or on a separate mechanism such as changes in place of excitation (Moore and Sek, 1996). These accounts assume that, although accurate phase locking is not necessary for subjects to adjust two tones to have the same pitch, or to judge which tone is higher, it is necessary for tasks such as musical interval adjustment that require a truly musical pitch percept (Plack and Oxenham, 2005; Moore and Ernst, 2012; Moore, 2012; Oxenham *et al.*, 2011).

There exist, however, some observations that cast doubt on the generally accepted assumption that phase locking is necessary for perception of musical pitch. For example, while most subjects in the study of Ward (1954) were unable to make octave judgements when the reference frequency, f_1 , was above 2700 Hz, for two of his subjects the variability in octave judgements was essentially the same for $f_1 = 5$ kHz (with the octave match at ~ 10 kHz) as for lower frequencies. However, octave judgements were more difficult and subjects took a greater time when f_1 was 5 kHz. Ward suggested that experience might play a role, as these two subjects were the only ones with experience in judging the pitch of pure tones. Similarly, Burns and Feth (1983) asked musically trained subjects to adjust various musical intervals for reference frequencies of 1 and 10 kHz. Even for the higher frequency, all three subjects could still adjust the musical intervals. For the 10-kHz reference, the standard deviations (SDs) averaged across all musical intervals, were about 3.5–5.5 times larger than for the 1-kHz reference. This ratio was less than that observed for unison adjustments, which were taken as an estimate of the difference limen for frequency (DLF). Burns and Feth concluded that their results were not incompatible with a temporal basis for both frequency discrimination and musical interval adjustment, as phase locking information decreases with increasing frequency. More recently, Oxenham *et al.* (2011) showed that complex tones whose audible harmonics all fall above 6 kHz can evoke a robust sense of pitch and musical interval, but high-frequency pure tones did not, even though the DLFs for the latter were less than 1.5%. They concluded either that there is sufficient phase-locking present at high frequencies to derive musical pitch for complex tones, or that such a pitch can be derived in the absence of phase locking. Pertaining to this, Moore and Ernst (2012) reported DLFs for center frequencies from 2 to 14 kHz that were consistent with the idea that there is a transition from a temporal to a place mechanism at about 8 kHz, rather than at 4–5 kHz, as commonly assumed.

Here we assessed whether a musical pitch can be heard in the likely *absence* of phase locking in the AN to the frequency corresponding to that pitch (or at harmonics of that pitch) at the time of the percept. To do so we used Zwicker tones (ZTs) (Zwicker, 1964). A ZT is a faint, decaying tonal percept that can arise *following* the presentation of a notched broadband noise, i.e., after the cessation of the noise, and that can last up to 5–6 s. It is often thought of as an auditory afterimage, by analogy to visual afterimages. Its pitch is always within the frequency range of the notch of the preceding noise, and depends on the level of the noise and the width of the notch. ZTs with pitches corresponding to those of pure tones with frequencies between 0.5 and about 8 kHz have been reported (Fastl and Stoll, 1979; Krump, 1993). The tonal quality of the ZT is greatest within the 2–5 kHz frequency range and is greatest for medium-low levels of the noise. The salience and tonal quality of the ZT decrease at medium to high sound levels [60–70 dB sound pressure level (SPL)], and for a noise level of 80 dB SPL no tonal aftereffect is perceived (Krump, 1993). With optimally

chosen noise parameters, 85%–94% of listeners reported hearing a ZT (Fastl, 1989; Krump, 1993).

Although there is of course phase locking in the AN to components of the notched noise, there is no evidence that there is phase locking corresponding to the frequency of the pitch of the ZT over the interval of several seconds during which the ZT is heard. Furthermore, several findings indicate that ZTs are unlikely to be produced mechanically at the level of the cochlea. Usually, no otoacoustic emissions (OAEs) are found at the frequency corresponding to the pitch of the ZT, except in rare cases when a subject hears a spontaneous OAE which could be made temporarily audible by a preceding notched noise without increasing the physical level of the OAE (Krump, 1993; Wiegube *et al.*, 1995, 1996). Also, while low-frequency tones at moderate to high levels can affect the level of evoked OAEs, they did not affect the ZT (Wiegube *et al.*, 1995, 1996), and no beating was observed between ZTs and a soft physical tone (Krump, 1993; Wiegube *et al.*, 1996). In addition, while a physical tone can mask a ZT, the reverse is not true (Krump, 1993). Wiegube *et al.* (1996) concluded that the ZT is not based on cochlear amplification, which is probably responsible for the generation of spontaneous OAEs, but is rather a neural phenomenon involving a release from neural lateral inhibition in the cochlear nucleus or higher stations of the ascending auditory pathway. Overall, there is no evidence that mechanical activity correlated to the ZT exists in the cochlea, and therefore it is unlikely that there is phase locking to ZTs in the auditory periphery. In summary, while there is phase locking to the ZT-exciting noise, it is extremely unlikely that there is phase locking in the AN to the frequency corresponding to the perceived pitch of the ZT (or harmonics thereof) during the time that the ZT is perceived (see also Sec. VI B). In order to avoid tedious repetition, from hereon we will abbreviate this as “absence of phase locking” where appropriate.

The question addressed here was whether ZTs can evoke a musical pitch. To the authors’ knowledge, there are no previous studies on this topic, and thus it is not known whether listeners can match musical intervals with ZTs. If ZTs can evoke a musical pitch, this would show that a musical pitch can be heard in the likely *absence* of phase locking in the AN at the time of the percept. In addition, ZTs were compared for diotic and monaural presentation of ZT-exciting noise.

II General Methods

A Experimental design

In stage I, subjects adjusted the frequency, level, and decay time of an exponentially decaying sinusoid (the time constant describing the exponentially decaying amplitude of the sinusoid) so that it sounded similar to the ZT they perceived following a notched broadband noise, for various notch positions. In stage II, subjects adjusted the frequency (and level) of a sinusoid so that its pitch was a specified musical interval below that of either a preceding ZT or a preceding sinusoid (and it was equally loud). Accurate judgements in this task would require the subject to have extracted a musical pitch from the ZT. Importantly, for each subject, the reference sinusoids corresponded to those that were adjusted in stage I to sound similar to, i.e., have equal pitch, loudness, and decay time as the ZTs. The precision and degree of difficulty of the musical interval adjustments for the ZTs and the matched

sinusoids (PT condition) were compared. The number of trials taken served as a measure of degree of difficulty of the musical interval adjustments. These need to be considered, because the reliability of matches can be increased by taking more time to make a match and by decreasing the average step size of the adjustments (Ward, 1954; Cardozo, 1965). In stage II, sinusoids were used as adjustable stimuli rather than a second ZT so as to avoid (i) a lengthy retention interval (time interval between reference pitch and variable pitch) and (ii) a retention interval that contains an interfering sound (both caused by the presence of the noise for the second ZT if it had been used), because both are known to impair pitch comparisons (see, e.g., Massaro, 1970).

In experiment 1, the ZT-exciting noise and the sinusoid were presented diotically, while in experiment 2 they were presented monaurally. If listeners can make musical interval judgements in the PT but not in the ZT condition, this would show that, although the ZT gives rise to a tonal percept, the percept does not meet a strict definition of musical pitch. Conversely, to the extent that listeners can make accurate and reliable musical-interval judgements for a ZT, the percept corresponding to a ZT meets the strictest definition of musical pitch.

B Stimuli and general procedure

To evoke ZTs, 5-s (including 20-ms onset and offset hanning-shaped ramps) notched broadband noises (30–16 000 Hz) were presented at an rms level of 51 dB SPL. From one match to the next, the lower edge frequency (LEF) of the notch had one of eight values: 2000, 2144, 2297, 2460, 2633, 2818, 3014, and 3500 Hz; the spacing between the LEFs was 0.58 Cams on the ERB_N-number scale (Moore, 2012), except for the last which was 1.29 Cams. The higher edge frequency (HEF) of the notch was always 1.5 times the LEF. These parameters were chosen to evoke clear ZTs, based on previous reports (Neelen, 1967; Lummis and Guttman, 1972; Fastl, 1989).

One match consisted of several trials, and subjects could take as many trials as they liked to finish a match. In each trial, two stimuli were presented. When subjects were satisfied with a match they indicated this by a specific button press. The number of trials taken for each match was counted, and the count was visible to subjects. They received no feedback on the accuracy of their adjustments. The first session (2 h including breaks) was considered practice, and matches from this session were discarded. For the experiment proper, typically about ten matches were collected for each condition from each subject. This required about four 2-h sessions for stage I, and eight 2-h sessions for stage II. For each match, the LEF was chosen at random from the set of eight with the restriction that no LEF was repeated before an equal number of matches had been completed for all LEFs.

All stimuli were generated digitally in MATLAB (The Mathworks, Natick, MA) with a sampling rate of 44.1 Hz. The ZT-exciting noise was generated in the spectral domain with “brick-wall” spectral edges. Within a match, the ZT-exciting noise sample was fixed, while across matches it varied. The stimuli were played out with an Asus Xonar Essence ST sound card (Taipei, Taiwan) using 16-bit digital to analog conversion and a sampling rate of 44.1 kHz. The overall sound level was controlled by Tucker-Davis Technologies (Alachua, FL) PA4 attenuators. Stimuli were presented using Sennheiser HD650 headphones (Wedemark,

Germany). Subjects were seated individually in an IAC (Winchester, UK) double-walled sound-attenuating booth, which provided an attenuation of about 97 dB over the frequency range of the noise notches and was itself situated in a quiet room, thus eliminating any audible ambient noise. Weak internal noises, generated for example by blood flow, might be present but these are thought to be well below the lowest detectable levels for medium and high frequencies (Stone *et al.*, 2014). The experiment was controlled with MATLAB software, and subjects responded by using a computer mouse to click on virtual buttons displayed on a monitor.

C Subjects

Four young (three female) normal-hearing musically trained subjects took part. None of them had absolute pitch or was a professional musician. Subjects 1–3 were singers, subject 2 played the flute, and subject 4 played the violin. All were between the ages of 20 and 35 years. The four were selected from a pool of five because they confirmed hearing a clear tonal aftereffect when listening informally to several examples of the ZT-exciting noises before the experiment started; the fifth person did not. Some of the subjects (including the fifth) reported hearing a pitch-like percept during the presentation of the ZT-exciting noises, which differed in pitch from the ZT that was perceived after the noise stopped. 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.

III Experiment I (Diotic Stimuli)

The ZT is usually assumed to depend on processes occurring before binaural interaction or in monaural pathways (Krump, 1993; Norena *et al.*, 2000). Here, the ZT-exciting noise and the matching sinusoid were presented diotically, as diotic presentation was expected to increase the salience of the ZT, at least in the absence of pronounced binaural diplacusis. All four subjects participated.

A Stage I: Identity matches between ZT and sinusoid

1 Stimuli and procedure—Subjects adjusted the frequency, level, and decay time of an exponentially decaying sinusoid so that it sounded similar to the ZT they perceived following a 5-s notched broadband noise. The adjustable sinusoid followed the ZT-exciting noise with an inter-stimulus interval (ISI, time between the end of one stimulus presentation and start of the next stimulus presentation) of 5.5 s. The adjustable sinusoid had a duration of 5 s (including 20-ms onset and offset hanning-shaped ramps), although, depending on the adjusted time constants, it may not have been audible over its whole duration. Thus, in each trial the overall listening time was 15.5 s. After cessation of the sinusoid, the subject indicated by button presses the desired direction and amount of change for the frequency, level, and time constant of the sinusoid for the next trial, and/or initiated the next trial. In each trial, the subject was allowed an unlimited number of button presses before s/he initiated the next trial and thus could adjust all parameters quasi simultaneously. The starting frequency of the adjustable sinusoid was chosen randomly between 0.9 and 1.6 times the LEF. Subjects could adjust the frequency upwards or downwards via virtual button presses

with step sizes of 4, 1, 1/4, and 1/16 semitones. The starting level of the sinusoid was chosen randomly between 17 and 27 dB SPL and could be adjusted up or down with step sizes of 10, 5, 2, and 1 dB. The starting time constant defining the exponential decay of the amplitude of the adjustable sinusoid was chosen randomly between 0.5 and 10 s and could be adjusted up or down by factors of 4, 2, $2^{1/2}$, and $2^{1/4}$.

After collection of the data for the identity matches, additional measurements were made to determine the absolute hearing thresholds, for the left and right ear separately, for 0.3-s pure tones at the frequencies that were matched to the pitches of the ZTs. Thresholds were obtained using a two-interval two-alternative forced choice adaptive tracking procedure estimating the 70.7% point on the psychometric function (Levitt, 1971). The last 8 of 12 turnpoints were averaged to obtain one estimate, and the final threshold for each frequency was based on the average of three such estimates.

2 Results—Figure 1 shows, for each subject, the geometric mean frequency expressed relative to the LEF [and the corresponding standard error of the mean (SE)] of the matched sinusoid that was adjusted to have the same pitch as the ZT, as a function of the LEF. The matched frequency was always clearly above the LEF and below the upper frequency limit of the notch, which was 1.5 LEF. This indicates that subjects indeed perceived and matched a ZT, arising *after* the cessation of the noise (in agreement with subjective reports), and that subjects did not match the upper edge pitch of the lower noise band or lower edge pitch of the upper noise band or a combination thereof; such pitches may arise *during* the presentation of the noise (Fastl, 1971; Klein and Hartmann, 1981; Bilsen, 1977). In most cases, the match was roughly a constant factor of 1.1–1.2 above the LEF. Matches for two of the subjects corresponded to a notably higher ratio of 1.25–1.45 above the LEF for two or three LEFs; for subject 1 the sixth and seventh LEFs and for subject 4 the second, third, and seventh LEFs resulted in distinctly higher factors than those observed for adjacent LEFs. These two subjects reported that sometimes the perceived ZT sounded like several tones. In these cases, they matched to the most dominant one. Subject 1, who was available for a longer time, was asked to take notes on the occasions when he heard several pitches rather than one. These records showed that this occurred for the sixth and the seventh LEFs, where the alternative pitch usually was described as about a minor third (3 semitones, or ratio = 1.19) higher or lower than the matched one.

Figure 2 shows the same data as in Fig. 1 (but with SDs), with the ordinate showing the absolute matched frequency. The dashed line indicates the LEF. It can be seen that the above-mentioned irregularities led to cases where two different LEFs resulted in essentially the same pitch match (the sixth and seventh LEFs for subject 1) or to cases where two lower LEFs (second and third LEFs for subject 4) led to a higher pitch match than for a condition with a higher LEF. Across the data for all subjects, there was no significant correlation between the frequency of the pitch-matched sinusoid and the SD of the matches expressed as a proportion of the mean matched frequency (Pearson's $r = -0.05$, $p = 0.788$).

Figure 3 shows, for each subject, the mean (and the SE) initial level of the sinusoid adjusted to have the same initial loudness as the ZT, as a function of the matched frequency of the ZT (circles connected by solid lines). Note that the line connects the circles in ascending order

of the LEFs to allow a mapping between adjusted frequency and LEF condition and thus reflects the occasional non-monotonic change in matched frequency with increasing LEF. The downward- and upward-pointing triangles connected by dashed lines show the absolute thresholds for the left and right ears, respectively. Absolute thresholds were measured exactly at the matched frequencies with a few exceptions (subject 1: seventh LEF; subject 4: third, fourth, and fifth LEFs), where thresholds were measured at different frequencies in order to cover more equally the total range of matched frequencies for each subject in the time available. Hearing thresholds and absolute matched levels are shown for completeness here, but normally the sensation levels (SLs) of the ZTs are reported. For the purpose of determining the initial SL of the ZTs, the hearing threshold was estimated by interpolation between thresholds for nearby frequencies in the few cases needed.

The solid circles connected by solid lines in Fig. 4 show the estimated initial SLs of the matched sinusoids as a function of the matched frequency, plotted with respect to the left-hand axis. SLs were derived from the lower of the two threshold measurements across the two ears. SLs ranged from 5 to 24 dB, across subjects and LEFs. The open circles connected by dashed lines, plotted with respect to the right-hand axis, show the matched time constants (τ , with their associated SE) which had mean values ranging from 1.1 to 5.1 s, across subjects and LEFs. There was a weak non-significant negative correlation between SL and time constant for subjects 1–3, while for subject 4 the correlation was weakly positive and non-significant. When SL and time constant were each averaged across all LEFs for a given subject, there was a perfect negative correlation between time constant and SL (Spearman's $\rho = -1$, $p = 0.1$, two-tailed) across subjects, but this was not significant due to the small number of subjects. On average, subject 2 gave the highest SLs and the smallest time constants, while the opposite was true for subject 4.

The reason for the irregularities in the functions relating the pitch matches to the LEF is not clear; possible explanations are considered in Sec. IV. For the present purpose of determining whether the ZT could evoke a musical pitch, these cases (subject 1: sixth and seventh LEFs; subject 4: second, third, and seventh LEFs) were excluded from further analysis of the data collected in stage II.

B Stage II: Musical interval matches

1 Stimuli and procedure—Subjects adjusted the frequency of a sinusoid so that its pitch was perceived to be a specified musical interval below the pitch of the preceding reference tone. The target musical intervals were a minor third (three semitones down) and a perfect fifth (seven semitones down). In different sessions, the reference tones were either ZTs (the same as in stage I) or physically presented pure tones (PTs). The latter corresponded to the matched tones from stage I, i.e., exponentially decaying tones matched in frequency and level to the ZTs. Subjects also adjusted the level of the adjustable sinusoid so that it was perceived as equal in loudness to the reference tone. The adjustable sinusoid had a fixed duration of 2 s, with a constant amplitude envelope except for the 20-ms on/off ramps. Thus, in both conditions, subjects matched a steady tone to form a specific musical interval with a decaying reference tone and to be equally loud. The loudness matching was done to avoid loudness differences between the reference and matching sound affecting

musical interval matches but was otherwise not of interest here. From one match to the next, the reference tone could be any one of the set of eight. Typically about 12 (at least 10) musical interval matches were collected for each condition and reference tone from each subject.

The trial duration was the same for both reference tone conditions. In the PT condition, a 5-s silence replaced the ZT-exciting noise. The adjustable 2-s sinusoid followed the ZT-exciting noise with an ISI of 5.5 s and, in the PT condition, followed the reference tone with an inter-onset interval of 5.5 s. Thus, in each trial the overall listening time was 12.5 s. The general trial structure was the same for both reference tone conditions and similar to that in stage I. While in stage I, subjects could adjust the frequency upwards or downwards via virtual button presses with fixed step sizes of 4, 1, 1/4, and 1/16 semitones, in stage II, the mean step sizes associated with the four buttons were identical to those in stage I, but the actual step size for each button was randomly varied across matches; before each match, the actual step sizes of the virtual buttons were chosen randomly from the range 0.75 to 1.25 times the mean step size. This was done to discourage subjects from “calculating”—after the first sound exposure—a sequence of button presses that was deemed to give the desired musical interval, rather than actually listening to and comparing the sounds in each trial.

2 Results—The accuracy of the musical interval adjustments for each subject is shown in Fig. 5, separately for the PT conditions (solid bar on the left of each group of two bars) and for the ZT conditions (open bar on the right side of each group of two bars). The accuracy was determined as the geometric mean (across LEFs or reference frequencies) of the mean ratio (across repetitions) of the adjusted frequency of the variable pure tone to the expected frequency. Error bars show the SD of this ratio across the eight reference frequencies or LEFs (six for subject 1 and five for subject 4, see above). The expected frequency was determined on the equal temperament scale. That is, the expected frequencies for the minor third and the perfect fifth were exactly three semitones (a factor of 1/1.189) and seven semitones (a factor of 1/1.498) below the reference frequency. For all subjects, the adjusted frequencies were somewhat flat, i.e., slightly lower than expected, leading to somewhat larger musical intervals. However, this was true for both the PT and the ZT conditions. To test whether the accuracy of the interval matches differed statistically between the PT and the ZT conditions, in this and the following experiments, univariate analyses of variance (ANOVAs), with tone type (ZT, PT) and musical interval (minor third, fifth) as fixed factors and frequency of the reference tone as random factor, were conducted separately on the data for each subject. Input data were the logarithms of the ratios of geometric mean (across repetitions) adjusted frequency to expected frequency for the different LEFs and reference frequencies. For subject 1, this ratio was significantly smaller for the ZT than for the PT conditions [$F(1,5) = 14.88$, $p = 0.012$], indicating higher accuracy for the PT conditions, while for subject 3 it was significantly smaller for the PT conditions [$F(1,7) = 7.83$, $p = 0.027$], indicating higher accuracy for the ZT conditions. For the remaining two subjects there was no significant difference between the accuracy of the interval matches for the PT and the ZT conditions [subject 2: $F(1,7) = 1.16$, $p = 0.317$; subject 4: $F(1,4) = 4.90$, $p = 0.091$]. Across the data from all subjects and musical intervals, there was a significant negative correlation between the frequency of the reference tone and the ratio of adjusted

frequency to expected frequency, and this was very similar for the two conditions (ZT: Pearson's $r = -0.45$, $p < 0.001$; PT: Pearson's $r = -0.49$, $p < 0.001$). For both conditions this reflects the tendency for the adjusted interval to increase with increasing reference-tone frequency. Generally, the adjusted frequencies were within about 2% of the expected frequencies, with no systematic differences between the ZT and the PT conditions. Thus, subjects were able to match musical intervals with good accuracy for both conditions.

Figure 6 shows, for each subject, a measure of the repeatability (“SD,” left-hand group of two bars) and the degree of difficulty (“n-listen,” right-hand group of two bars) of the musical interval adjustments for the target intervals of the fifth (square-patterned bar, on the left-hand side of each group of two bars) and the minor third (diamond-patterned bar, on the right-hand side of each group of two bars). Consider first the repeatability measure. To compare the repeatability of the adjustments in the ZT conditions with those in the matched PT conditions, the ratio of the respective SDs was calculated as follows. (1) Separately for each subject, each musical interval (minor third, fifth), each reference condition (ZT, PT), and each LEF (or reference PT tone frequency), the SD of the adjusted frequencies across the 12 interval matches was calculated as a percentage of the geometric mean. (2) These SDs were geometrically averaged across the different LEFs (or PT frequencies) to obtain the representative SD. (3) The representative SD for the ZT was divided by the representative SD for the PT. This ratio is shown by the height of the two bars on the left-hand side of each panel in Fig. 6, separately for each musical interval. The number above each bar in Fig. 6 indicates the size of the representative SD for the PT condition in percent (ranging from 0.9% to 1.5%). The representative SD for the ZT condition ranged from 1.2% to 2.5%. Adjustments for the ZTs were somewhat more variable than for the PTs; the ratio of the SDs was between 1.0 and 2.2. In this and the following experiments, ANOVAs with tone type and musical interval as fixed factors and frequency of the reference tone as random factor, were conducted separately on the data for each subject, using the SD of the logarithms of the adjusted frequencies within a given LEF or reference frequency as input data. For subjects 1–3, the SDs were significantly larger for the ZT than for the PT conditions [subject 1: $F(1,5) = 22.49$, $p = 0.005$; subject 2: $F(1,7) = 13.93$, $p = 0.007$; subject 3: $F(1,7) = 8.41$, $p = 0.023$], while for subject 4 they were not [$F(1,4) = 0.38$, $p = 0.855$]. Across the data for all subjects and musical intervals, there was no significant correlation between the reference-tone frequency and the SD of the musical interval matches expressed as a percentage of the geometric mean matched frequency, for either condition (ZT: Pearson's $r = 0.08$, $p = 0.544$; PT: Pearson's $r = 0.01$, $p = 0.930$).

Consider next the measure of difficulty. The group of two bars on the right-hand side of each panel in Fig. 6 shows the ratio of the average number of listening times (n-listen) in the ZT and the PT conditions. The number above each bar indicates the mean value of n-listen for the PT condition. There were individual differences in how many trials were taken to make a match. For example, in the PT conditions, on average subject 1 took 7.6 trials while subject 3 took 5.7 trials. However, n-listen was very similar for the ZT and PT conditions, i.e., the ratio was very close to 1. In this and the following experiments, ANOVAs with tone type and musical interval as fixed factors and frequency of the reference tone as random factor were conducted separately on the data for each subject, using n-listen within a given condition as input data. For subjects 1–3, there was no significant difference between the ZT and the PT

conditions [subject 1: $F(1,5) = 0.34$, $p = 0.586$; subject 2: $F(1,7) = 0.40$, $p = 0.548$; subject 3: $F(1,7) = 0.65$, $p = 0.448$], while for subject 4 the mean value of n -listen was significantly *smaller* for the ZT than for the PT conditions [$F(1,4) = 25.01$, $p = 0.007$]. This means that the achieved accuracy and reliability in the ZT conditions did not come at the cost of listening more often in the ZT than in the PT conditions.

3 Discussion—The results showed that, on average, subjects selected similar frequencies in a musical interval adjustment task irrespective of whether the reference tone was a ZT or an equally salient pure tone. The adjusted frequencies were slightly flat in both cases. This might partly have resulted from subjects' bias towards a “just scale” with musical intervals corresponding to integer ratios of 6/5 and 3/2 for the minor third and the perfect fifth, respectively. If this were the only reason, the “bias” should have been larger for the minor third than for the perfect fifth, which was not observed. Rakowski (1990) reported that when melodic intervals were tuned without a musical context there was a tendency for small intervals to be tuned somewhat smaller and for large intervals to be tuned somewhat larger than expected from an equal-temperament scale, and he considered the commonly observed octave enlargement phenomenon (Terhardt, 1971) to be part of this general pattern. In the present results, the minor third was not adjusted to be slightly smaller than expected from an equal temperament scale, although its enlargement was on average slightly smaller than that of the perfect fifth. The SDs of the musical interval adjustments were only slightly larger (a factor of 1.0–2.2) for the ZTs than for the PTs, with no increase in n -listen. Thus, overall the results show that ZTs can indeed evoke a weak musical pitch.

IV Experiment 2 (Monaural Stimuli)

In the second experiment, all stimuli (ZT-exciting noise and sinusoids) were presented monaurally. This was done (i) to try to shed some light on the reasons for the irregularities in the identity matches to the pitch of the ZT and (ii) to assess whether musical interval adjustments were indeed improved (or at least were equally good) for diotic relative to monaural presentation of the ZT-exciting noise.

Three (two female) of the four subjects who participated in experiment 1 took part. In stage I, two of them (subjects 1 and 3) matched the pitch of the ZT for both left- and right-ear monaural presentation for all LEFs, while the third (subject 4) completed identity matches for four out of the eight LEF conditions (for the second, third, sixth, and seventh LEFs). In stage II, subjects 1 and 3 matched musical intervals with stimuli presented to the ear for which they reported hearing the ZT most clearly (subject 1: right, subject 3: left). The experimental procedure was the same as for experiment 1.

A Results of stage I: Identity matches between ZT and sinusoid

Figure 7 shows, for each subject, the geometric mean frequency of the sinusoid matched in pitch to the ZT, expressed as the ratio relative to the LEF (together with the corresponding SE), as a function of the LEF. The results for the monaural conditions are indicated by triangles (downward pointing for the left ear and upward pointing for the right ear). For comparison, the results for the diotic condition in experiment 1 are re-plotted. Also indicated are the results of a repeat measurement for some LEFs in the diotic condition that were

collected after the monaural conditions were completed, to check the stability of the ZT pitch.

As was the case for the diotic condition, the matched frequency was always within the notch region of the noise indicating that a ZT could be heard. Also, as for the diotic condition, in most cases the frequency of the matched sinusoid was roughly a constant factor above the LEF, and hence, in absolute terms, increased with increasing LEF, as expected. However, as occurred for the diotic condition, subject 1 matched the ZT for some LEFs at a ratio substantially higher than for neighboring LEFs (e.g., third and seventh LEFs for the right ear). Surprisingly, these irregularities did not always occur for the same LEFs as for the diotic condition. There were some cases where two different LEFs resulted in essentially the same pitch match in Hz (subject 1: right ear, the third and fourth LEFs). In addition, the left- and right-ear pitch matches could *both* be markedly different from the pitch matched in the diotic condition, being either both lower (subject 1: sixth LEF, subject 4: second LEF) or both higher (subject 3: 3–5 LEFs). The repeat measures obtained in the diotic conditions (empty circles) showed that these marked differences between the diotic and the monaural presentation were reliable and that the pitch of the diotic ZT was quite stable. Usually, the ZT has been assumed to depend on processes occurring before binaural interaction or in monaural pathways (Krump, 1993; Norena *et al.*, 2000). For example, Krump (1993) reported that the weak percept of a ZT after presentation of a lowpass noise to one ear was unaffected by variation of the highpass cutoff frequency of a highpass noise presented to the other ear. The current observation of marked and reliable differences between the pitches of the ZT for diotic and monaural presentation of the noise may indicate that at least some integration of information from both ears is involved in determining the pitch of the ZT. Finally, the left- and right-ear matches sometimes differed substantially from each other, such as for the third LEF for subject 1 and the highest LEF for subject 3.

In some further control measurements for the sixth LEF only (not shown), subject 1 was asked to adjust the frequency of the PT so that it matched (i) the highest pitch heard in the ZT and (ii) the lowest pitch heard in the ZT (remember that with diotic presentation the sixth LEF led to the percept of several tones for this subject). These matches were collected for ZT-exciting noise presented monaurally to the left and right ears, and diotically. The results showed that none of the higher pitch matches in the monaural conditions was as high as the higher pitch match in the diotic condition, where the latter corresponded to the pitch generally matched previously with diotic presentation. This indicates that the dominant pitch perceived in the diotic condition probably did not correspond to any pitch heard in either of the two monaural conditions.

Figure 8 shows the estimated initial SLs (solid triangles connected by solid lines plotted with respect to the left-hand axis) and the time constants with their associated SEs (empty triangles connected by dashed lines plotted with respect to the right-hand axis) of the monaurally presented matched sinusoids as a function of the matched frequency. SLs ranged from 3 to 22 dB across subjects, ears, and LEFs, while time constants had mean values ranging from 0.7 to 5.7 s. Negative correlations between SL and time constant were observed for all ears, except for the left ear of subject 1; correlations were significant only for subject 3 (left ear: Spearman's $\rho = -0.88$, $p = 0.004$; right ear: Spearman's $\rho =$

−0.81, $p = 0.015$; both two-tailed). When SL and time constant were each averaged across all LEFs for a given subject and ear, there was a significant perfect negative correlation between time constant and SL (Spearman's $\rho = -1$, $p < 0.01$, two-tailed) across ears (and subjects); on average subject 1 selected the highest SLs and the smallest time constants for stimuli presented to the left ear, while subject 4 selected the lowest SLs and the largest time constants for stimuli presented to the right ear.

Subjects 1 and 3 reported that the ZT was more salient in the diotic than in the monaural conditions, while subject 4, when questioned, thought them to be about equal. To facilitate comparison across conditions, Fig. 9 shows the adjusted SLs and time constants for the monaural conditions together with the corresponding results for the diotic condition, as a function of the frequency matched to the pitch of the ZT. Note that if the SL in the diotic and *both* monaural conditions were the same (as was the case, e.g., for subject 3, second LEF) this would not imply a lack of binaural loudness summation for the ZT because the matching sinusoid was presented diotically when the ZT-exciting noise was presented diotically and was presented monaurally when the ZT-exciting noise was presented monaurally. Rather it would imply that the binaural summation of loudness for the ZT is equal to the binaural summation of loudness for the PT. Also, the same SL in the diotic and *one* of the monaural conditions would imply that the ZT evoked by the diotic noise was somewhat louder than the ZT evoked by the monaural noise. The estimated SLs for the diotic condition were always at least as high as the lower of the two monaural SLs and were mostly although not always somewhat below the higher of the two monaural SLs. This suggests that some loudness summation across ears occurs for ZTs. It should be noted, however, that both the initial SL and the time constant are likely to affect the overall loudness and salience of the percept. For subjects 1 and 3, the time constants were greater for the diotic than for the monaural conditions, in agreement with their subjective report of a more salient ZT in the diotic condition.

The cases with strong irregularities in the function mapping the LEF to the pitch match (subject 1, right ear: third and seventh LEFs) were excluded from further analysis of the data collected in stage II of the experiment.

B Results of stage II: Musical interval matches

Figure 10 shows the accuracy of the musical interval adjustments for subjects 1 and 3 for monaural presentation (left-hand column). Their results for diotic presentation are re-plotted from Fig. 5 (right-hand column). The format is the same as for Fig. 5. For both subjects the adjusted frequency for monaural presentation was on average somewhat lower than the “correct” value. For subject 1, there was no significant difference between the ratio of adjusted to expected frequency for the ZT and the PT conditions [$F(1,5) = 1.17$, $p = 0.329$], while for subject 3 the ratio was significantly smaller for the ZT than for the PT condition [$F(1,7) = 30.28$, $p = 0.001$], indicating significantly higher accuracy for the PT condition. However, for this subject, the opposite was true for the diotic condition, where the ratio was significantly smaller for the PT than for the ZT condition. Generally, the adjusted frequencies were within about 3% of the expected frequencies. To evaluate whether monaural presentation led to different accuracy than binaural stimulation, for each subject

accuracy ANOVAs were conducted separately for the ZT and the PT conditions, with stimulation type (monaural, diotic) and musical interval as fixed factors and frequency of the reference tone as random factor. For the PT condition, there was no significant difference between monaural and diotic stimulation, for either subject [subject 1: $F(1,4) = 0.33$, $p = 0.599$; subject 3: $F(1,7) = 1.71$, $p = 0.232$]. For the ZT condition, only subject 3 showed a significant effect of stimulation type, accuracy being significantly *worse* in the monaural than in the diotic condition [subject 1: $F(1,4) = 0.16$, $p = 0.713$; subject 3: $F(1,7) = 36.65$, $p = 0.001$].

Figure 11 shows the repeatability (left-hand group of two bars) and the degree of difficulty (right-hand group of two bars) of the musical interval adjustments for subjects 1 and 3 for monaural presentation of the ZT-exciting noise to their “preferred” ear (left column). For comparison, their results for the diotic condition are re-plotted from Fig. 6 (right column). The format is the same as for Fig. 6. Consider first the repeatability measure, which was calculated as before. As observed for diotic presentation, adjustments for the ZTs were somewhat more variable than for the matched PTs; the ratio of the SDs was between 1.2 and 2.2. For subject 3, SDs were significantly larger for the ZT than for the PT conditions [$F(1,7) = 16.09$, $p = 0.005$], while for subject 1 they were not [$F(1,5) = 3.08$, $p = 0.140$]. The number above each bar in Fig. 11 indicates the size of the average SD across LEFs for the PT condition in percent (ranging from 1.1% to 1.5%). The representative SD for the ZT condition ranged from 1.4% to 3.2%.

To evaluate whether the SDs differed significantly for monaural and diotic presentation, for each subject the SDs were compared separately for the ZT and the PT conditions in two ANOVAs, with stimulation type and musical interval as fixed factors and frequency of the reference tone as random factor. For the PT condition, there was no significant difference between monaural and diotic presentation, for either subject [subject 1: $F(1,4) = 4.43$, $p = 0.103$; subject 3: $F(1,7) = 0.28$, $p = 0.873$]. For the ZT condition, only subject 3 showed a significant effect of stimulation type, SDs being significantly larger for monaural than for diotic presentation [subject 1: $F(1,4) = 0.31$, $p = 0.610$; subject 3: $F(1,7) = 11.81$, $p = 0.011$].

Consider next the degree of difficulty (n-listen) of the interval matches (two bars on the right of each panel in Fig. 11). There was no significant difference in n-listen for the ZT and the PT conditions for either subject [subject 1: $F(1,5) = 5.36$, $p = 0.069$; subject 3: $F(1,7) = 0.75$, $p = 0.415$]. Thus, as for diotic presentation, the good accuracy and reliability in the ZT conditions did not come at the cost of listening more often in the ZT than in the PT conditions.

For each subject, n-listen was compared separately for the ZT and the PT conditions in two ANOVAs, with stimulation type and musical interval as fixed factors and frequency of the reference tone as random factor. For the PT condition, the effect of stimulation type was significant only for subject 3, with significantly more trials taken in the monaural than in the diotic condition [subject 1: $F(1,4) = 2.50$, $p = 0.189$; subject 3: $F(1,7) = 18.51$, $p = 0.004$]. For the ZT condition, both subjects took significantly more trials in the monaural than in the diotic condition [subject 1: $F(1,4) = 29.57$, $p = 0.006$; subject 3: $F(1,7) = 13.28$, $p = 0.008$].

In summary, subjects were able to make musical interval adjustments for ZTs following monaural presentation of ZT-exciting noise. Both subjects took more trials for monaural (“preferred-ear”) than for diotic presentation of the ZT-exciting noise, suggesting that musical interval adjustments were harder for monaural presentation. In addition, monaural presentation led to significantly poorer accuracy and larger SDs for subject 3, probably due to the greater salience of the ZT percept in the diotic than in the monaural condition.

V Supplementary Measurements: Spontaneous Otoacoustic Emissions and Lower Noise Level

A Spontaneous otoacoustic emissions

As mentioned in Sec. I, normally no OAEs have been found at the frequency corresponding to the pitch of the ZT (Krump, 1993). However, there were cases when a subject had a spontaneous OAE (SOAE) which could be made temporarily audible by a preceding notched noise when the frequency region of the notch covered the frequency of the SOAE. This change in audibility was not accompanied by an increase in the physical level of the SOAE (Krump, 1993). It is thus conceivable that some of the irregularities in the functions relating the pitch of the ZT to the LEF, observed in experiments 1 and 2 for subjects 1 and 4, were caused by a normally non-audible SOAE that became momentarily audible after listening to a ZT-exciting noise. To investigate this possibility, SOAEs were measured for subject 1. Subject 4 was not available anymore. SOAE measurement was under-taken with the Otodynamics (Hatfield, UK) Echoport ILO 292-I system (using a UGD TE+DPOAE probe, Software Module ILO V6, Option: Spontaneous Test) by measuring the OAE after the immediate transient OAE response to a short (80 μ s) synchronizing broadband click (level in ear canal at about 81 dB SPL/Peak) had died away (time interval of 21–80 ms after the click presentation, including 20-ms cosine onset- and offset-ramps; averaged over 1000 repetitions). The measurement was later verified with a research setup similar to that described by Zhou *et al.* (2010). Briefly, custom MAC software (OSX) was used to record the ear canal signals. Three five-minute samples were recorded with an Etymotic Research ER-10C microphone/preamplifier system connected to the computer via a MOTU828 firewire interface (24 bit, 44.1 kHz) and amplified by a Stanford SR560 low-noise amplifier. Two SOAEs were observed for the right ear of subject 1 (one at about 460 Hz and one at about 1415 Hz) within the assessed frequency range (0.4–6.4 kHz), and none in the left ear. Both SOAE frequencies were far below the lowest LEF of 2000 Hz. Therefore it is unlikely that SOAEs were responsible for the observed irregularities in the pitch matches.

B Lower noise levels

In experiments 1 and 2, the ZT-exciting noises were presented at an rms level of 51 dB SPL because informal listening at various levels showed this to give a salient and clear ZT. In a further attempt to clarify the reasons for the irregularities in the identity matches to the pitch of the ZT, and also to investigate the influence of level on the ability to make musical interval judgements, subject 1 was tested with ZT-exciting noises presented monaurally to his “preferred” ear at 41 dB SPL. The results of this supplementary experiment are provided in the Appendix. Briefly, it was found that (i) ZTs had a slightly (2.3 dB) lower initial SL at

the lower noise level, (ii) the matched pitches were mostly slightly lower at the lower noise level, (iii) the irregularities in the pitch matches observed at 51 dB SPL were not present at 41 dB SPL, (iv) nevertheless, SDs were higher at the LEFs for which anomalous matches were obtained at 51 dB SPL, and (v) the subject could make reliable musical interval judgements at 41 dB SPL.

VI General Discussion

A Comparison to previous results

The main objective of this study was to assess whether ZTs have a musical pitch. The results for the musical interval adjustment tasks showed that subjects were able to make musical interval matches to ZTs, with only a slight increase in the SDs (by a factor of 1.0–2.2) relative to those obtained for PT references with matched pitch and loudness, with no increase in listening time and no decrease in accuracy. Thus, the results suggest that a weak musical pitch can exist in the absence of phase locking in the AN at the time of the percept. In addition, the comparison between diotic and monaural conditions suggests that musical interval adjustments were harder for monaural presentation of the ZT-exciting noise, as subjects needed more trials in the monaural condition and accuracy was reduced and SDs increased for one subject.

The authors are not aware of any previous reports on musical interval adjustments using ZTs. However, musical-interval identification in the presumed absence of accurate phase locking has been assessed by asking subjects to match intervals between PTs with very high frequencies, above the proposed upper limit of phase locking. As mentioned in Sec. I, Burns and Feth (1983) had musically trained subjects adjust various musical intervals for reference frequencies of 1 and 10 kHz. For the same musical intervals as used in the present study, the geometric mean of the SDs for the 70-dB SPL 1-kHz reference tone in their data (1.19%) was very similar to that observed for the PT condition in the present study (1.16%). For the 10-kHz reference, SDs in their data were 1.6–11.7 (mean of 4) times larger than for the 1-kHz reference. Here, SDs for musical interval adjustments with the ZT as reference were 1.0–2.2 (mean of 1.5) times larger than with the PT as reference. Hence, if both methods measure the decrease in the accuracy of interval adjustments when phase-locking cues are removed, it would seem that the method using high-frequency PTs produces a larger deterioration than replacing a moderate-frequency PT with a ZT of the same pitch. However, to fairly compare the increase in SDs for the ZT conditions relative to the PT conditions in the present study with the increase in SDs for the 10 kHz relative to the 1 kHz reference tone in the study of Burns and Feth, one has to take into account that in the present study only the reference tone was a ZT while the adjustable tone was a pure tone. With some reasonable assumptions,¹ one can estimate the SD expected if both tones had been ZTs (and there was no deleterious effect of prolonged retention interval and/or interfering noise in the retention

¹The following assumptions were made. The expected variance for musical interval matches with two PTs, σ_{PPM}^2 , is given by

$$\sigma_{PPM}^2 = \sigma_P^2 + \sigma_M^2, \quad (1)$$

interval) to be a factor of about 1.87 larger than for the PT condition, while the currently observed factor was 1.5. Therefore, the increase in SDs for the 10-kHz frequency found by Burns and Feth was larger than would have been observed for interval matches between two ZTs with pitches corresponding to frequencies between about 2–4.2 kHz. Burns and Feth (1983) interpreted their results solely in terms of phase locking, and a decrease thereof with increasing frequency. The comparison of the present results with theirs suggests that an additional factor influenced their results. One possible factor is lack of familiarity with high-frequency pure tones (Ward, 1954; Attneave and Olson, 1971). Note that the highest note for the grand piano (C8 is 4186 Hz) and the highest notes for orchestral instruments (such as the piccolo) fall between 4 and 5 kHz.

The results from stage I of the present study can be compared to those for previous reports of unison pitch matches to ZTs. The majority of pitch matches observed here were a factor of 1.1–1.2 above the LEF. This is similar to previous findings for ZT-exciting noises with comparable notch widths and levels (Zwicker, 1964; Fastl and Krump, 1995; Zwicker and Fastl, 1999). The slight decrease in the pitch of the ZT for the lower noise level of 41 dB SPL used in the supplementary experiment also fits well with previous reports about the dependence of the ZT pitch on the level of the ZT exciter (Zwicker, 1964; Fastl and Krump, 1995; Zwicker and Fastl, 1999). The pitch of the ZT for notched-noise exciters corresponds well to the frequency giving the minimum in the masking pattern of the ZT-exciting noise or, for wider notch widths and/or low levels, to the frequency where the upper side of the masking pattern of the lower noise band crosses the absolute threshold in quiet (Krump, 1993; Fastl and Krump, 1995). Within this framework, the increase in the pitch of the ZT

where σ_P^2 corresponds to the variance associated with the internal representations of the PTs, and σ_M^2 corresponds to the variance associated with the internal template of the musical interval and other sources of variance in the musical interval task that are independent of the internal representation of the pitch and are in common for the conditions considered here. If there were no deleterious effect of prolonged retention interval and/or interfering noise in the retention interval, the expected variance for musical interval matches with two ZTs, σ_{ZZM}^2 , would be given by

$$\sigma_{ZZM}^2 = \sigma_Z^2 + \sigma_M^2, \quad (2)$$

where σ_Z^2 corresponds to the variance associated with the internal representations of the ZTs. Then the expected variance of the interval matches in the present ZT condition, σ_{ZPM}^2 , is given by

$$\sigma_{ZPM}^2 = (\sigma_Z^2 + \sigma_P^2) / 2 + \sigma_M^2. \quad (3)$$

The results showed that

$$\sigma_{ZPM}^2 = 1.5^2 \cdot \sigma_{PPM}^2. \quad (4)$$

Substitution of Eqs. (1) and (3) into Eq. (4), solving for σ_Z^2 , and substituting the result of this into Eq. (2), σ_{ZZM}^2 is determined as $\sqrt{2 \times 1.5^2 - 1}$, which equals 1.87.

with increasing noise level is linked to an upward shift in the minimum of the excitation pattern evoked by the notched noise with increasing level (Glasberg and Moore, 1990). For example, based on excitation patterns calculated according to Moore *et al.* (1997) [using the executable file excit2005.exe (available on website <http://hearing.psychol.cam.ac.uk/Demos/excit2005.zip>)], for an LEF of 2 kHz the predicted pitch based on the minimum in the excitation pattern was a factor of 1.19 above the LEF for a level of 51 dB SPL and a factor of 1.15 above the LEF for a level of 41 dB SPL. Both values fit well within the observed ranges of unison matches reported here for the two noise levels.

For monaurally presented ZT-excitors, previous studies reported a mean matching SL of 10–15 dB (Krump, 1993), or about 10 dB with large variations across individuals (Zwicker, 1964; Fastl, 1989). The initial matching SLs observed in the present study are in general agreement with these reports. The time constants describing the exponential amplitude decay of the matched pure tone measured here were negatively correlated with the initial SLs, indicating either that the higher the matching SL the more rapid the decay of sensation or that there is a trade-off between the two parameters for the purpose of matching the loudness of the ZT. For two out of three subjects, time constants were smaller for the monaural than for the diotic condition, consistent with subjective reports of a more salient ZT percept in the diotic condition. The time constants observed here (mean between 2 and 3 s) fall well within the range of ZT durations reported by Fastl (1989).

For two out of four subjects, the function relating the ZT matching frequency to the LEF showed some irregularities, with no consistent pattern across monaural and diotic conditions. It is unclear what underlies these irregularities. It has been reported that a spontaneous OAE can become audible—without physical increase in level of the OAE—after presentation of a ZT-exciting noise with a notch around the OAE frequency (Krump, 1993; Wiegand *et al.*, 1995, 1996). However, this possible explanation for the irregularity in the ZT-pitch function observed here was rejected, at least for one subject, as no spontaneous OAE with frequency in the notch region was observed. The ZT has generally been described as a monaural phenomenon. Here, for several subjects, marked and reliable differences between the pitches of the ZT for diotic and monaural presentation of the noise to either ear were observed. This may indicate that at least some degree of integration of information from the two ears is involved in determining the pitch of the ZT.

B Absence of phase-locking in the AN to the frequency corresponding to the ZT-pitch?

The finding that listeners can adjust musical intervals to ZTs suggests that a weak musical pitch can exist in the absence of phase locking in the AN to the frequency corresponding to that pitch at the time of the pitch percept. In the following, several arguments are presented in support of this.

- (i) As described in Sec. I (not repeated here), several findings indicate that ZTs are unlikely to be produced mechanically at the level of the cochlea.
- (ii) During the presentation of the notched noise (and, due to ringing of the auditory filters, for a very brief time after the cessation of the noise), there will be phase locking in the AN to the frequencies within the noise passbands, as AN fibers phase lock to frequencies close to their best frequency in response to stimulation

with white noise (de Boer and Jongkees, 1968; Ruggero, 1973). Neurons with best frequencies in the frequency region of the notch will either phase lock to frequency components of the noise away from their best frequency or will show only spontaneous activity. Summary autocorrelation functions (Meddis and O'Mard, 1997) of the modeled neural activity in the AN may possibly be construed as corresponding to a pitch heard *during* the presentation of the noise, but the ZT arises only after the noise has ceased and lasts for several seconds, during which phase locking in the AN to the noise components has ceased.

- (iii) It is generally assumed that without mechanical activation of the basilar membrane there is no phase locking (or phase stable activity) in the mammalian AN. After the cessation of the notched noise, there will be spontaneous firing in the AN, including from fibers with best frequencies in the notch region. However, spontaneous firing is not phase stable, i.e., it is not phase locked to the CF of the fiber (Kiang *et al.*, 1965; Manley and Robertson, 1976) and the spontaneous firing patterns are uncorrelated across fibers (Johnson and Kiang, 1976). Note that while these studies used anesthetized animals, there is no evidence that Nembutal/Pentobarbital, as used extensively in auditory research and used by Manley and Robertson (1976), increases the threshold of AN neurons at their best frequency (Dodd and Capranica, 1992).
- (iv) At every step along the ascending auditory pathway, the upper frequency limit for phase locking decreases (de Ribaupierre *et al.*, 1980). Thus, at a more central level, phase-locking information present in the AN must be replaced by another code, which may or may not be a rate-place code, especially for higher signal frequencies. The ZT could then arise from, for example, adaptation of lateral inhibition at that level of representation. In this view, phase locking in the AN might be involved in “creating” the ZT, but there would be no phase locking to the ZT pitch in the AN at the time of the pitch percept. This is currently the most parsimonious explanation for the ZT (see also below).
- (v) The physiological bases of the ZT are still unclear. Up to now, a direct physiological correlate of the ZT has been reported only in the auditory cortex in a neuromagnetic study (Hoke *et al.*, 1996; Hoke *et al.*, 1998). The processes underlying the ZT have sometimes been likened to those underlying the perceptual enhancement of a probe signal within the notch region of a simultaneous notched wideband masker produced by extending the masker in time before the onset of the probe (Wiegand *et al.*, 1996; Zhou *et al.*, 2010). A direct correlate of the enhancement effect has been reported in the inferior colliculus of the marmoset monkey (Nelson and Young, 2010) and the dorsal cochlear nucleus of the Mongolian gerbil (Fontaine *et al.*, 2015), but was not found in the AN of the guinea pig, where only a distributed representation was observed (Palmer *et al.*, 1995).
- (vi) Some attempts to model the ZT involve asymmetric lateral inhibition and noise-detection neurons (Franosch *et al.*, 2003), while others involve a central gain adaptation mechanism (Parra and Pearlmutter, 2007; but see Zhou *et al.*, 2010).

Models of the enhancement effect involve interaction between adaptation of wideband lateral inhibition and more sharply tuned excitation (Nelson and Young, 2010). The details of these models are beyond the scope of the present paper. However, none of them assumes the involvement of phase locking at the frequency corresponding to the ZT pitch either in the AN, or at a later processing stage. Instead, the pitch of the ZT is given by a rate-place code. Similarly, the enhancement effect has been modeled by an increase in the firing rate of neurons with best frequencies close to the probe frequency (Nelson and Young, 2010).

Overall, psychophysical findings and current physiological knowledge suggest that the ZT is unlikely to be produced at the level of the cochlea, and is unlikely to evoke phase locking in the AN to the frequency corresponding to the pitch at the time of the pitch percept. This view is reflected in attempts to model the ZT. Thus, the present results suggest that a weak musical pitch can indeed exist in the likely “absence of phase locking.”

C Previous evidence for pitch perception in the absence of phase locking in the AN

We consider now whether previous findings may be interpreted as evidence for pitch perception in the absence of phase locking in the AN related to the perceived pitch at the time of the percept. It is well known that listeners can make “high-low” judgements between stimuli that differ only in their frequency composition and where phase locking information is likely to be very weak. For example, although DLFs for pure tones increase with increasing frequency, frequency discrimination remains possible for frequencies up to at least 10 kHz (Henning, 1966; Moore and Ernst, 2012). Furthermore, cochlear implant (CI) listeners can discriminate between electrical stimulation of individual electrodes based solely on place of excitation. Performance differs across patients, but, for the most sensitive listeners, the minimum detectable shift in excitation peak along the AN array is similar to that produced by a 3% change in acoustic frequency for a normal-hearing (NH) listener (Moore and Carlyon, 2005). Indeed, one of the checks that audiologists make to ensure that a CI is working properly is to ask the patient whether pitch increases monotonically with stimulation of increasingly basal electrodes. The important question, both for high-frequency pure tones in NH and electrical stimulation in CI is whether the reported percept corresponds to pitch in its strictest, musical sense, or whether it is a timbre change that listeners associate with pitch.

We are not aware of convincing evidence that CI listeners can consistently identify musical intervals based solely on place of excitation (see, e.g., McDermott and McKay, 1997), which may be due to several factors including the difficulty of producing focused electrical stimulation and the often patchy neural survival in CI listeners. As discussed in Sec. I, there is evidence that high-frequency tones can evoke a musical pitch for some NH listeners. Although the subjects in Ward’s (1954) study took longer to make octave adjustments at higher than at lower frequencies, two subjects could make octave adjustments between two tones even when the frequency of the higher tone was about 10 kHz. Similarly, although Burns and Feth (1983) found that musical interval judgements were less accurate for a 10-kHz than for a 1-kHz standard, this difference was smaller than that for unison adjustments. This evidence is bolstered by two studies (Oxenham *et al.*, 2011; Macherey and Carlyon,

2014) showing that judgements of the missing fundamental are possible when the frequencies of the individual frequency components are all above the limit of 4–5 kHz that has previously been proposed for the use of phase locking in musical pitch judgements of pure tones (Plack and Oxenham, 2005; Moore and Ernst, 2012; Moore, 2012; Oxenham *et al.*, 2011).

The above might be regarded as evidence for the existence of musical pitch in the absence of phase locking. However, recently the frequency at which the putative transition from a temporal to a place mechanism occurs, at least with regard to frequency discrimination of pure tones, has been revised upwards to about 8 kHz (Moore and Ernst, 2012). As noted by Moore and Ernst (2012), this transition point may depend on the specific stimuli and task used. They argued that while the weak phase locking occurring for frequencies between 4 and 8 kHz may be sufficient to achieve reasonably small frequency DLs, it may not be sufficient to give a strong sense of musical interval for pure tones at high frequencies. On the other hand, when several high-frequency resolved harmonics are presented simultaneously, the combined phase-locking information from all of them may be sufficient to evoke a robust pitch and allow musical melody discrimination (Oxenham *et al.*, 2011).

Finally, it is known that Huggins Pitch (HP) (Cramer and Huggins, 1958) can support melody recognition (Akeroyd *et al.*, 2001). As HP is derived from a stimulus that, in each ear, is a broadband noise, one might argue that it does not depend on phase locking to the frequency corresponding to the perceived pitch (or harmonics of it) at the level of the AN. However, phase locking in the AN across a wide range of frequencies *including* that of the HP at the time of the pitch percept is essential. Most accounts of dichotic pitches posit the creation of a central activity pattern (for example, binaural cross-correlation versus best frequency versus time) and the extraction of pitch from features of that pattern (Culling *et al.*, 1998a; Culling *et al.*, 1998b). Phase locking is involved in the creation of the pattern, although, at the next stage, the pattern itself does not necessarily preserve phase locking. Note, however, that Akeroyd and Summerfield (1999) presented a fully temporal model of dichotic pitch in which phase locking was preserved even after the binaural analysis stage. Thus HP could be encoded, e.g., at the level of the medial superior olive, in terms of phase locking, or features in a place-type representation, or some combination of the two. Of course there is also the well-documented percept of the low pitch or residue pitch of complex tones (Schouten, 1940; Licklider, 1956) which, in the case of resolved harmonics and in the absence of a spectral component at the fundamental frequency, can be perceived without phase locking in the AN at the frequency corresponding to the perceived pitch. However, phase locking exists at the time the pitch is heard to harmonics of the perceived pitch in the AN, at least for harmonic frequencies below about 4–5 kHz, and many models can account for how the perceived pitch could be extracted from this phase-locking information (see, e.g., Cariani and Delgutte, 1996a; Meddis and O'Mard, 1997; Patterson *et al.*, 1995; Ives and Patterson, 2008). In summary, we are not aware of previous evidence for the perception of a truly musical pitch in the absence of phase locking in the AN at the time of the percept.

D Theoretical implications and practical applications

The present results show that the ZT has a weak musical pitch, and we have argued that this is unlikely to arise from phase locking to the frequency corresponding to the pitch of the ZT at the level of the AN at the time of the pitch percept. This does not prove that phase locking is not necessary for pitch perception, as it might be used in creating a central rate-place representation that in turn may lead to the ZT percept. Hence, although the present results present a challenge to contemporary models of pitch perception, which universally rely on AN phase locking *at the time of the percept* either to the frequency corresponding to the perceived pitch or to harmonics thereof (Cariani and Delgutte, 1996b; Meddis and O'Mard, 1997; Patterson *et al.*, 1995; Bleeck *et al.*, 2004) or, for dichotic pitch, to a wide frequency range including the frequency corresponding to the dichotic pitch, the resolution of the challenge may be one of refinement rather than replacement. Such an endeavor may eventually yield practical, as well as theoretical, benefits. One potential benefit is suggested by the fact that CI and auditory brainstem implant users have a very limited ability to use phase-locking information, and this ability is typically limited to electrical pulse rates below about 300 pps. By understanding the more central representations of musical pitch, it may one day be possible to bypass this phase-locking bottleneck, by stimulating neurons at the level of the brainstem or midbrain (Lim *et al.*, 2008), thereby providing an improvement in pitch perception.

VII Summary and Conclusions

This study assessed whether ZTs have a musical pitch. Musically trained subjects adjusted the frequency of a pure tone to be a specified musical interval below the pitch of either a preceding ZT or a preceding pure tone that was matched in pitch and loudness to the ZT. Subjects were able to make musical interval matches to ZTs, with only a slight increase in the SDs (by a factor of 1.0–2.2) relative to those obtained for the PT references, and without any increase in listening time or decrease in accuracy of the mean adjustment. Thus, the results suggest that a weak musical pitch can exist in the absence of phase locking in the AN to the frequency corresponding to the perceived pitch or to harmonics thereof at the time of the pitch percept. However, phase locking in the AN to the ZT-exciting noise might be used to create a more central neural activity pattern, which may or may not be a rate-place code, and the ZT might arise from adaptation and/or lateral inhibition in that pattern. Comparison of the results for diotic and the monaural conditions suggests that musical interval adjustments were harder for monaural presentation of the ZT-exciting noise, as subjects needed more trials in the monaural condition and, for one subject, accuracy was reduced and SDs increased. In addition, identity matches to the pitch of the ZTs could differ for monaural and diotic presentation of the ZT-exciting noise indicating that information from the two ears may be integrated to at least some degree in determining the pitch of the ZT.

Acknowledgments

This work was supported by intramural funding from the MRC (MC_A060_5PQ70). We thank Richard Knight and Glenis Long for help with the setups for the measurement of spontaneous OAEs, and Brian Moore, Peter Cariani, and Bill Hartmann for helpful comments on previous versions of this paper.

Appendix

In experiments 1 and 2, the ZT-exciting noises were presented at a level of 51 dB SPL. This appendix presents supplementary results from subject 1, obtained using a lower noise level. ZT-exciting noises were presented monaurally to his “preferred” ear at a level of 41 dB SPL. This was done to (i) clarify the reasons for the irregularities in the identity matches to the pitch of the ZT and (ii) investigate the influence of level on the ability to make musical interval judgements. In 15 sessions, each lasting 2 h, both identity matches and musical interval matches were collected following the same general procedure as for the previous experiments.

1 Results of stage I: Identity matches between ZT and sinusoid

Figure 12 shows the results from stage I of the experiment. Panel (A) shows the geometric mean frequency relative to the LEF (and the corresponding SE) of the sinusoid matched in pitch to the ZT, as a function of the LEF. The results are indicated by open triangles. For comparison, the results for the noise level of 51 dB SPL are re-plotted as filled triangles. As previously, the matched frequency was always within the notch region of the noise, indicating that a ZT could be heard. In most cases the frequency of the matched sinusoid was below that for the higher level, consistent with previous reports (Zwicker, 1964; Fastl and Krump, 1995; Zwicker and Fastl, 1999). The frequency of the matched sinusoid corresponded to a roughly constant factor above the LEF. However, for the third and seventh LEFs, the SEs were markedly larger than for the remaining LEFs and the subject reported an unclear ambiguous pitch. These two LEFs were the ones that produced the irregularities in the function relating pitch to LEF for the level of 51 dB SPL. Therefore, they were excluded again from the analysis of the results of stage II of the experiment.

Panels (B) and (C) show the estimated initial SLs and time constants (with their associated SE), respectively, as a function of the matched frequency of the ZT. Across LEFs, the SLs ranged from 11 to 21 dB and the time constants ranged from 0.9 to 1.7 s. There was no correlation between SL and time constant. The subject reported that the ZT was less salient for the 41 dB SPL ZT-exciting noise than for the 51 dB SPL noise level. This is reflected in the estimated SLs, which were on average 2.3 dB lower for the 41 dB SPL than for the 51 dB SPL noise [$F(1,7) = 11.65$, $p = 0.011$]. The time constants were unaffected by the noise level [$F(1,7) = 0.21$, $p = 0.659$].

2 Results of stage II: Musical interval matches

Figure 13 compares the accuracy of the musical interval adjustments for the 41-dB SPL noise [panel (A)] and the 51-dB SPL noise [panel (B)]. As before, the adjusted frequency was on average somewhat lower than “correct.” Generally, the adjusted frequencies were within about 3% of the expected frequencies. The difference between the ratios of adjusted to expected frequency in the ZT and PT conditions was not significant [$F(1,5) = 3.39$, $p = 0.125$]. Two ANOVAs were conducted (one for the PT and one for the ZT condition) with level and musical interval as fixed factors and frequency of the reference tone as random

factor. The effect of level was not significant for either the PT condition [$F(1,5) = 0.64$, $p = 0.459$] or the ZT condition [$F(1,5) = 2.44$, $p = 0.179$].

Figure 14(A) shows the repeatability (left-hand group of two bars) and the n-listen values (right-hand group of two bars) of the interval adjustments for the level of 41 dB SPL. For comparison, the results for the level of 51 dB SPL are re-plotted [Fig. 14(B)]. As observed previously, interval adjustments for the ZTs were somewhat more variable than for the PTs; the ratio of the SDs was 1.4 and 1.7 for the fifth and the minor third, respectively. The number above each bar in Fig. 14 indicates the size of the representative SD for the PT condition in percent. The representative SD for the ZT condition ranged from 1.6% to 1.9%. The difference between the SDs for the ZT and the PT conditions just missed significance [$F(1,5) = 5.63$, $p = 0.064$]. Two ANOVAs were conducted (one for the PT and one for the ZT condition) with level and musical interval as fixed factors and frequency of the reference tone as random factor. The effect of level was not significant for either the PT condition [$F(1,5) = 0.14$, $p = 0.725$] or the ZT condition [$F(1,5) = 0.61$, $p = 0.470$].

Consider next the ratio of the values of n-listen for the ZT and the PT conditions. The ratio was slightly larger than one, but the difference between the values of n-listen for the ZT and the PT conditions was not significant [$F(1,5) = 1.48$, $p = 0.278$]. Thus, the good accuracy and reliability in the ZT conditions did not come at the cost of taking significantly more trials in the ZT than the PT conditions. The value of n-listen was compared separately for the ZT and the PT conditions in two ANOVAs, with level and musical interval as fixed factors and frequency of the reference tone as random factor. The effect of level was not significant for either the PT condition [$F(1,5) = 1.55$, $p = 0.268$] or the ZT condition [$F(1,5) = 2.12$, $p = 0.205$].

In summary, subject 1 was able to make musical interval adjustments for ZTs evoked by monaural presentation of a 41-dB SPL ZT-exciting noise. SDs were marginally larger for the ZT than for the equally salient PT. There was no significant effect of level (relative to monaural presentation at 51 dB SPL) on the accuracy, reliability or degree of difficulty of the musical interval matches. Thus, although the ZT was reported as less salient for the lower than the higher level, and this was reflected in a reduction of the levels of the pure tones matched to the initial loudness of the ZTs, the ZT percept was sufficiently loud and clear to allow musical interval matches to be unaffected.

¹The following assumptions were made. The expected variance for musical interval matches with two PTs, σ_{PPM}^2 , is given by

$$\sigma_{PPM}^2 = \sigma_P^2 + \sigma_M^2, \quad (1)$$

where σ_P^2 corresponds to the variance associated with the internal representations of the PTs, and σ_M^2 corresponds to the variance associated with the internal template of the musical interval and other sources of variance in the musical interval task that are independent of the internal representation of the pitch and are in common for the conditions considered here. If

there were no deleterious effect of prolonged retention interval and/or interfering noise in the retention interval, the expected variance for musical interval matches with two ZTs, σ_{ZZM}^2 , would be given by

$$\sigma_{ZZM}^2 = \sigma_Z^2 + \sigma_M^2, \quad (2)$$

where σ_Z^2 corresponds to the variance associated with the internal representations of the ZTs. Then the expected variance of the interval matches in the present ZT condition, σ_{ZPM}^2 , is given by

$$\sigma_{ZPM}^2 = (\sigma_Z^2 + \sigma_P^2) / 2 + \sigma_M^2. \quad (3)$$

The results showed that

$$\sigma_{ZPM}^2 = 1.5^2 \cdot \sigma_{PPM}^2. \quad (4)$$

Substitution of Eqs. (1) and (3) into Eq. (4), solving for σ_Z^2 , and substituting the result of this into Eq. (2), σ_{ZZM}^2 is determined as $\sqrt{2 \times 1.5^2 - 1}$, which equals 1.87.

References

- Akeroyd MA, Moore BCJ, Moore GA. Melody recognition using three types of dichotic-pitch stimulus. *J Acoust Soc Am*. 2001; 110:1498–1504. DOI: 10.1121/1.1390336 [PubMed: 11572360]
- Akeroyd MA, Summerfield AQ. A fully-temporal account of the perception of dichotic pitches. *Brit J Audiol*. 1999; 33:106–107.
- Attneave F, Olson RK. Pitch as a medium: A new approach to psychophysical scaling. *Am J Psychol*. 1971; 84:147–166. DOI: 10.2307/1421351 [PubMed: 5566581]
- Bilsen FA. Pitch of noise signals: Evidence for a ‘central spectrum.’ *J Acoust Soc Am*. 1977; 61:150–161. DOI: 10.1121/1.381276 [PubMed: 833366]
- Bleeck S, Ives T, Patterson RD. Aim-Mat: The auditory image model in MATLAB. *Acta Acust Acust*. 2004; 90:781–787.
- Burns, EM., Feth, LL. Pitch of sinusoids and complex tones above 10 kHz. *Hearing—Physiological Bases and Psychophysics*. Klinke, R., Hartmann, R., editors. Springer; Berlin: 1983. p. 327–333.
- Cardozo BL. Adjusting the method of adjustment: SD vs DL. *J Acoust Soc Am*. 1965; 37:786–792. DOI: 10.1121/1.1909439
- Cariani PA, Delgutte B. Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. *J Neurophysiol*. 1996a; 76:1698–1716. [PubMed: 8890286]
- Cariani PA, Delgutte B. 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*. 1996b; 76:1717–1734. [PubMed: 8890287]
- Cramer EM, Huggins WH. Creation of pitch through binaural interaction. *J Acoust Soc Am*. 1958; 30:413–417. DOI: 10.1121/1.1909628
- Culling JF, Marshall DH, Summerfield AQ. Dichotic pitches as illusions of binaural unmasking. II. The Fourcin pitch and the dichotic repetition pitch. *J Acoust Soc Am*. 1998a; 103:3527–3539. DOI: 10.1121/1.423060 [PubMed: 9637036]

- Culling JF, Summerfield AQ, Marshall DH. Dichotic pitches as illusions of binaural unmasking. I. Huggins' pitch and the 'binaural edge pitch,'. *J Acoust Soc Am*. 1998b; 103:3509–3526. DOI: 10.1121/1.423059 [PubMed: 9637035]
- de Boer E, Jongkees LBW. On cochlear sharpening and cross-correlation methods. *Acta Oto-Laryngol*. 1968; 65:97–104. DOI: 10.3109/00016486809120947
- de Cheveigné A. Cancellation model of pitch perception. *J Acoust Soc Am*. 1998; 103:1261–1271. DOI: 10.1121/1.423232 [PubMed: 9514016]
- de Ribaupierre F, Rouiller E, Toros A, de Ribaupierre Y. Transmission delay of phase-locked cells in the medial geniculate body. *Hear Res*. 1980; 3:65–77. DOI: 10.1016/0378-5955(80)90008-8 [PubMed: 6249782]
- Dodd F, Capranica RR. A comparison of anesthetic agents and their effects on the response properties of the peripheral auditory system. *Hear Res*. 1992; 62:173–180. DOI: 10.1016/0378-5955(92)90183-N [PubMed: 1429259]
- Fastl H. Über Tonhöhenempfindungen bei Rauschen ("On sensations of pitch evoked by noise"). *Acustica*. 1971; 25:350–354.
- Fastl H. Zum Zwicker-Ton bei Linienspektren mit spektralen Lücken ("On the Zwicker-tone of line spectra with a spectral gap"). *Acustica*. 1989; 67:177–186.
- Fastl H, Krump G. Pitch of the Zwicker-tone and masking patterns. In: Manley, GA, Klump, GM, Köppl, C, Fastl, H., Oeckinghaus, H., editors. *Advances in Hearing Research, Proceedings of the 10th International Symposium on Hearing*; Singapore: World Scientific; 1995.
- Fastl H, Stoll G. Scaling of pitch strength. *Hear Res*. 1979; 1:293–301. DOI: 10.1016/0378-5955(79)90002-9 [PubMed: 541278]
- Fontaine B, Franken T, Joris PX. Neural correlates of context-dependent enhancement in the dorsal cochlear nucleus. *ARO Abstracts*. 2015; 38:42.
- Franssen JMP, Kemper R, Fastl H, van Hemmen JL. Zwicker tone illusion and noise reduction in the auditory system. *Phys Rev Lett*. 2003; 90:178103. doi: 10.1103/PhysRevLett.90.178103 [PubMed: 12786108]
- Glasberg BR, Moore BCJ. Derivation of auditory filter shapes from notched-noise data. *Hear Res*. 1990; 47:103–138. DOI: 10.1016/0378-5955(90)90170-T [PubMed: 2228789]
- Henning GB. Frequency discrimination of random-amplitude tones. *J Acoust Soc Am*. 1966; 39:336–339. DOI: 10.1121/1.1909894 [PubMed: 5908437]
- Hoke ES, Hoke M, Ross B. Neurophysiological correlate of the auditory after-image ('Zwicker tone'). *Audiol Neurotol*. 1996; 1:161–174. DOI: 10.1159/000259196 [PubMed: 9390799]
- Hoke ES, Ross B, Hoke M. Auditory afterimage: Tonotopic representation in the auditory cortex. *NeuroReport*. 1998; 9:3065–3068. DOI: 10.1097/00001756-199809140-00027 [PubMed: 9804317]
- Ives DT, Patterson RD. Pitch strength decreases as F0 and harmonic resolution increase in complex tones composed exclusively of high harmonics. *J Acoust Soc Am*. 2008; 123:2670–2679. DOI: 10.1121/1.2890737 [PubMed: 18529186]
- Johnson DH. The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones. *J Acoust Soc Am*. 1980; 68:1115–1122. DOI: 10.1121/1.384982 [PubMed: 7419827]
- Johnson DH, Kiang NY. Analysis of discharges recorded simultaneously from pairs of auditory nerve fibers. *Biophys J*. 1976; 16:719–734. DOI: 10.1016/S0006-3495(76)85724-4 [PubMed: 938715]
- Kiang, NY-S., Watanabe, T., Thomas, EC., Clark, LF. *Discharge Patterns of Single Fibers in the Cat's Auditory Nerve*. MIT Press; Cambridge, MA: 1965.
- Klein MA, Hartmann WM. Binaural edge pitch. *J Acoust Soc Am*. 1981; 70:51–61. DOI: 10.1121/1.386581 [PubMed: 7264072]
- Krump, G. Beschreibung des akustischen Nachtones mit Hilfe von Mithörschwellenmustern ("Description of the auditory after-image with masked-threshold patterns"). Ph.D. thesis, Technical University München; München, Germany: 1993.
- Levitt H. Transformed up-down methods in psychoacoustics. *J Acoust Soc Am*. 1971; 49:467–477. DOI: 10.1121/1.1912375

- Licklider, JCR. Auditory frequency analysis. Information Theory. Cherry, C., editor. Academic Press; New York: 1956. p. 253-268.
- Lim HH, Lenarz T, Joseph G, Battmer RD, Patrick JF, Lenarz M. Effects of phase duration and pulse rate on loudness and pitch percepts in the first auditory midbrain implant patients: Comparison to cochlear implant and auditory brainstem implant results. *Neurosci*. 2008; 154:370–380. DOI: 10.1016/j.neuroscience.2008.02.041
- Lummis RC, Guttman N. Exploratory studies of Zwicker's 'negative afterimage' in hearing. *J Acoust Soc Am*. 1972; 51:1930–1944. DOI: 10.1121/1.1913052 [PubMed: 5045252]
- Macherey O, Carlyon RP. Re-examining the upper limit of temporal pitch. *J Acoust Soc Am*. 2014; 136:3186–3199. DOI: 10.1121/1.4900917 [PubMed: 25480066]
- Manley GA, Robertson D. Analysis of spontaneous activity of auditory neurons in spiral ganglion of guinea-pig cochlea. *J Physiol*. 1976; 258:323–336. DOI: 10.1113/jphysiol.1976.sp011422 [PubMed: 957159]
- Massaro DW. Retroactive interference in short-term recognition memory for pitch. *J Exp Psychol*. 1970; 83:32–39. DOI: 10.1037/h0028566 [PubMed: 5436482]
- McDermott HJ, McKay CM. Musical pitch perception with electrical stimulation of the cochlea. *J Acoust Soc Am*. 1997; 101:1622–1631. DOI: 10.1121/1.418177 [PubMed: 9069629]
- Meddis R, O'Mard L. A unitary model of pitch perception. *J Acoust Soc Am*. 1997; 102:1811–1820. DOI: 10.1121/1.420088 [PubMed: 9301058]
- Moore BCJ. Frequency difference limens for short-duration tones. *J Acoust Soc Am*. 1973; 54:610–619. DOI: 10.1121/1.1913640 [PubMed: 4754385]
- Moore, BCJ. *An Introduction to the Psychology of Hearing*. 6th ed. Brill; the Netherlands: 2012. p. 203-243.
- Moore, BCJ., Carlyon, RP. Perception of pitch by people with cochlear hearing loss and cochlear implant users. *Pitch: Neural Coding and Perception*. Plack, CJ, Oxenham, AJ, Fay, RR., Popper, AN., editors. Springer; New York: 2005. p. 234-277.
- Moore BCJ, Ernst SMA. Frequency difference limens at high frequencies: Evidence for a transition from a temporal to a place code. *J Acoust Soc Am*. 2012; 132:1542–1547. DOI: 10.1121/1.4739444 [PubMed: 22978883]
- Moore BCJ, Glasberg BR, Baer T. A model for the prediction of thresholds, loudness and partial loudness. *J Audio Eng Soc*. 1997; 45:224–240. Executable file excit2005.exe, <http://hearing.psychol.cam.ac.uk/Demos/excit2005.zip> (Last viewed 9/27/2016).
- Moore BCJ, Sek A. Detection of frequency modulation at low modulation rates: Evidence for a mechanism based on phase locking. *J Acoust Soc Am*. 1996; 100:2320–2331. DOI: 10.1121/1.417941 [PubMed: 8865639]
- Neelen JJM. Auditory afterimages produced by incomplete line spectra. *IPO Ann Prog Rep*. 1967; 2:37–43.
- Nelson PC, Young ED. Neural correlates of context-dependent perceptual enhancement in the inferior colliculus. *J Neurosci*. 2010; 30:6577–6587. DOI: 10.1523/JNEUROSCI.0277-10.2010 [PubMed: 20463220]
- Norena A, Micheyl C, Chery-Croze S. An auditory negative after-image as a human model of tinnitus. *Hear Res*. 2000; 149:24–32. DOI: 10.1016/S0378-5955(00)00158-1 [PubMed: 11033244]
- Oxenham AJ, Micheyl C, Keebler MV, Loper A, Santurette S. Pitch perception beyond the traditional existence region of pitch. *Proc Natl Acad Sci USA*. 2011; 108:7629–7634. DOI: 10.1073/pnas.1015291108 [PubMed: 21502495]
- Palmer AR, Russell IJ. Phase-locking in the cochlear nerve of the guinea-pig and its relation to the receptor potential of inner hair-cells. *Hear Res*. 1986; 24:1–15. DOI: 10.1016/0378-5955(86)90002-X [PubMed: 3759671]
- Palmer AR, Summerfield Q, Fantini DA. Responses of auditory-nerve fibers to stimuli producing psychophysical enhancement. *J Acoust Soc Am*. 1995; 97:1786–1799. DOI: 10.1121/1.412055 [PubMed: 7699160]
- Parra LC, Pearlmutter BA. Illusory percepts from auditory adaptation. *J Acoust Soc Am*. 2007; 121:1632–1641. DOI: 10.1121/1.2431346 [PubMed: 17407900]

- Patterson RD, Allerhand MH, Giguere C. Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform. *J Acoust Soc Am*. 1995; 98:1890–1894. DOI: 10.1121/1.414456 [PubMed: 7593913]
- Plack, CJ., Oxenham, AJ. The psychophysics of pitch. *Pitch: Neural Coding and Perception*. Plack, CJ., Oxenham, AJ., Fay, RR., Popper, AN., editors. Springer; New York: 2005. p. 8-13.
- Rakowski A. Intonation variants of musical intervals in isolation and in musical contexts. *Psychol Music*. 1990; 18:60–72. DOI: 10.1177/0305735690181005
- Ruggero MA. Response to noise of auditory nerve fibers in the squirrel monkey. *J Neurophysiol*. 1973; 36:569–587. [PubMed: 4197339]
- Schouten JF. The residue and the mechanism of hearing. *Proc Kon Akad Wetenschap*. 1940; 43:991–999.
- Sek A, Moore BCJ. Frequency discrimination as a function of frequency, measured in several ways. *J Acoust Soc Am*. 1995; 97:2479–2486. DOI: 10.1121/1.411968 [PubMed: 7714264]
- Stone MA, Paul AM, Axon P, Moore BC. A technique for estimating the occlusion effect for frequencies below 125 Hz. *Ear Hear*. 2014; 35:49–55. DOI: 10.1097/AUD.0b013e31829f2672 [PubMed: 24141593]
- Terhardt E. Die Tonhöhe harmonischer Klänge und das Oktavintervall (“The pitch harmonic clang and the octave interval”). *Acustica*. 1971; 24:126–136.
- Ward WD. Subjective musical pitch. *J Acoust Soc Am*. 1954; 26:369–380. DOI: 10.1121/1.1907344
- Wiegrefe L, Kössl M, Schmidt S. Auditory sensitization during the perception of acoustical negative afterimages: Analogies to visual processing. *Naturwissenschaften*. 1995; 82:387–389. DOI: 10.1007/BF01134569 [PubMed: 7566183]
- Wiegrefe L, Kössl M, Schmidt S. Auditory enhancement at the absolute threshold of hearing and its relationship to the Zwicker tone. *Hear Res*. 1996; 100:171–180. DOI: 10.1016/0378-5955(96)00111-6 [PubMed: 8922992]
- Zhou X, Henin S, Thompson SE, Long GR, Parra LC. Sensitization to masked tones following notched-noise correlates with estimates of cochlear function using distortion product otoacoustic emissions. *J Acoust Soc Am*. 2010; 127:970–976. DOI: 10.1121/1.3277156 [PubMed: 20136219]
- Zwicker E. ‘Negative afterimage’ in hearing. *J Acoust Soc Am*. 1964; 36:2413–2415. DOI: 10.1121/1.1919373
- Zwicker, E., Fastl, H. *Psychoacoustics—Facts and Models*. Springer; Berlin: 1999. p. 129-134.

Fig. 1.
Geometric mean frequency (and the corresponding SE) of the diotic sinusoid matched in pitch to the ZT, expressed as the ratio relative to the LEF, plotted as a function of the LEF of the notch in the diotically presented ZT-exciting noise.

Fig. 2.

As Fig. 1, but showing the absolute frequency (and the corresponding SD) of the pitch-matched sinusoid.

Fig. 3.

Initial level of exponentially decaying diotic sinusoid matched to the initial loudness of the ZT perceived after diotic presentation of a ZT-exciting noise (circles), and absolute thresholds in quiet for left (downward-pointing triangles) and right ears (upward-pointing triangles) at pure tone frequencies matched to the pitch of the ZT, as a function of the frequency of the sinusoid matched in pitch to the ZTs. Levels are equivalent diffuse-field levels.

Fig. 4.

Initial sensation level (SL) of exponentially decaying diotic sinusoid matched to the initial loudness of the ZT perceived after diotic presentation of a ZT-exciting noise (filled circles connected by solid lines) plotted with respect to left-hand y-axis, and the corresponding time constants (Tau, open circles connected by dashed lines) plotted with respect to right-hand y-axis, as a function of the matched frequency.

Fig. 5.

Results of the four subjects for experiment 1. The figure shows the geometric mean (and corresponding SD across reference frequencies or LEFs) of the ratio of the adjusted frequency to the expected frequency, with the PT and the ZT as reference (see legend), for each target musical interval (fifth and minor third).

Fig. 6.

Ratio of the geometric mean (across reference frequencies or LEFs) of the SDs (left) and of the average number of trials (n-listen, right) between ZT and PT conditions, for musical intervals of a perfect fifth and a minor third (see legend). The value of each measure for the PT condition is given by the number above the corresponding bar; for the SD this is the representative SD (see text), expressed as a percentage of the geometric mean of the adjusted frequencies.

Fig. 7.

As Fig. 1, but for experiment 2. Results for the diotic presentation used in experiment 1 are re-plotted for comparison, together with a repeat measure for the diotic presentation collected after completion of experiment 2.

Fig. 8.

As Fig. 4, but for the results of experiment 2. Left ear: left-hand column, right ear: right-hand column.

Fig. 9.

Left column: Initial SL of exponentially decaying sinusoid matched to the initial loudness of the ZT for conditions monaural left (filled downward-pointing triangles), monaural right (filled upward-pointing triangles), and diotic (filled circles), plotted with regard to left-hand axis, as a function of the matched frequency. Right column: Corresponding time constants of exponentially decaying sinusoid matched to the initial loudness of the ZT for conditions monaural left (empty downward-pointing triangles), monaural right (empty upward-pointing

triangles), and diotic (empty circles), plotted with regard to right-hand axis, as a function of the matched frequency.

Fig. 10.

As in Fig. 5, but for experiment 2 (monaural, left column). Results for diotic presentation are re-plotted for comparison (right-hand column).

Fig. 11.

As in Fig. 6, but for experiment 2 (monaural, left). Results for diotic presentation are re-plotted for comparison (right).

Fig. 12.

Identity matches obtained for the lower presentation level of the ZT-exciting noise (open triangles, subject 1 only). Results of experiment 2 (shown in Figs. 7 and 8) are re-plotted for comparison (filled triangles). (A) Geometric mean frequency (and SEs) of the monaural sinusoid matched in pitch to the ZT, expressed relative to the LEF, plotted as a function of the LEF. (B) Initial SL of exponentially decaying monaural sinusoid matched to the initial loudness of the ZT, plotted with respect to left-hand axis, as a function of the matched

frequency. (C) Corresponding time constants (and SEs), plotted with respect to right-hand axis, as a function of the matched frequency.

Fig. 13.

As Fig. 10, but for the lower noise level (monaural, 41 dB SPL noise) (A). Results from experiment 2 (monaural, 51 dB SPL noise) are re-plotted for comparison (B).

Fig. 14.

As Fig. 11, but for the lower noise level (monaural, 41 dB SPL noise) (A). Results from experiment 2 (monaural, 51 dB SPL noise) are re-plotted for comparison (B).