

# Do Zwicker Tones Evoke a Musical Pitch?

Hedwig E. Gockel and Robert P. Carlyon

**Abstract** It has been argued that musical pitch, i.e. pitch in its strictest sense, requires phase locking at the level of the auditory nerve. The aim of the present study was to assess whether a musical pitch can be heard in the absence of peripheral phase locking, using Zwicker tones (ZTs). A ZT is a faint, decaying tonal percept that arises after listening to a band-stop (notched) broadband noise. The pitch is within the frequency range of the notch. Several findings indicate that ZTs are unlikely to be produced mechanically at the level of the cochlea and, therefore, there is unlikely to be phase locking to ZTs in the auditory periphery. In stage I of the experiment, musically trained 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 broadband noise, for various notch positions. In stage II, subjects adjusted the frequency of a sinusoid so that its pitch was a specified musical interval below that of either a preceding ZT or a preceding sinusoid (as determined in 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 1.1–2.2. The results suggest that a musical pitch may exist in the absence of peripheral phase locking.

**Keywords** Auditory afterimage · Tonal percept · Band-stop noise · Notched noise · Phase locking · Pitch match · Frequency adjustment · Pitch salience · Musical interval

---

H. E. Gockel (✉) · R. P. Carlyon

MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 7EF, UK  
e-mail: hedwig.gockel@mrc-cbu.cam.ac.uk

R. P. Carlyon

e-mail: bob.carlyon@mrc-cbu.cam.ac.uk

© The Author(s) 2016

P. van Dijk et al. (eds.), *Physiology, Psychoacoustics and Cognition in Normal and Impaired Hearing*, Advances in Experimental Medicine and Biology 894, DOI 10.1007/978-3-319-25474-6\_44

## 1 Introduction

Periodic sounds produce periodic patterns of phase-locked activity in the auditory nerve. 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 auditory nerve (Cariani and Delgutte 1996; Meddis and O'Mard 1997). Furthermore, the precision of phase locking weakens at high frequencies, and the putative upper limit of phase locking at frequencies of about 4–5 kHz (Johnson 1980; Palmer and Russell 1986) 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 above about 4–5 kHz (Ward 1954; Attneave and Olson 1971; Moore 1973; Sek and Moore 1995).

There exist, however, some observations that cast doubt on the generally accepted assumption that phase locking is necessary for musical pitch. For example, while most subjects in Ward's (1954) study 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  $\sim 10$  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 kHz and 10 kHz. Even for the high frequency, all three subjects could do the task. 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 increase 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 just noticeable difference for the latter was less than 1.5%. They concluded either that there is sufficient phase-locking present at high frequencies to derive complex musical pitch, or that a complex 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 *absence* of phase locking, using Zwicker tones, ZTs (Zwicker 1964). A ZT is a faint, decaying tonal percept that can arise following the presentation of a notched broadband noise and that can last up to 5–6 s. The pitch is always within the frequency range of the

notch, and depends on the level of the noise and the width of the notch. Several findings indicate that ZTs are unlikely to be produced mechanically at the level of the cochlea. Normally, no otoacoustic emissions (OAEs) were found at the frequency corresponding to the pitch of the ZT, except in the rare case when a subject had a spontaneous OAE which could be made temporarily audible by a preceding notched noise (without increasing its physical level), and no beating was observed between ZTs and a soft physical tone (Krump 1993). Also, while low frequency tones at moderate to high levels can affect the level of evoked OAEs, they did not affect the ZT (Wiegrebe et al. 1995). Thus, no mechanical activity correlated to the ZT exists in the cochlea, and there is unlikely to be phase locking to ZTs in the auditory periphery. The question addressed here is whether ZTs can evoke a musical pitch.

## 2 Methods

### 2.1 *Experimental Design*

Four young normal-hearing musically trained subjects took part. In stage I, 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 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 to be equally loud). 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 of the musical interval adjustments for the ZTs and the matched sinusoids (PT condition) was compared.

### 2.2 *Stimuli and General Procedure*

To evoke ZTs, 5-s (including 20-ms onset and offset ramps) notched broadband noises (30-16000 Hz) were presented diotically at an rms level of 51 dB SPL. From one match to the next, the lower edge frequency (LEF) of the notch could take one of eight values: 2000, 2144, 2297, 2460, 2633, 2818, 3014, and 3500 Hz. The higher edge frequency (HEF) of the notch was always 1.5 times the LEF. The adjustable sinusoid had a maximum duration of 5 s (including 20-ms onset and offset ramps) and followed the ZT-exciting noise with an inter-stimulus interval of 5.5 s. After cessation of the sinusoid, the subject indicated by button presses the desired direction of change for the frequency, level, and (stage I only) time constant of the sinusoid for the next presentation, and/or initiated the next sound presentation, i.e. the next trial. In each trial, the subject was allowed an unlimited number of button presses before s/he initiated the next trial. For each parameter that needed adjusting, three different

step sizes were available for each direction of change. The smallest available step sizes were 1/16 semitone, 1 dB, and a factor of  $2^{1/4}$  for the frequency, level, and time constant adjustments, respectively. Subjects could take as many trials as they liked to finish a match. 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.

In stage I, in most cases, the adjusted frequency of the matched sinusoid increased with increase in LEF and was a factor of 1.1–1.2 above the LEF. Two of the subjects reliably matched higher frequencies (factor of 1.3–1.4 above the LEF) in two or three LEF conditions, leading to functions where the matched frequency sometimes did not increase with increase in LEF. For the present purposes, these cases were excluded from further analysis of stage II data, as they could be due to spontaneous OAEs that became audible after listening to the notched noise (Krump 1993; Wiegrebe et al. 1995). The initial level and time constant of the matched sinusoids had mean values that ranged from 5 to 23 dB above threshold in quiet and from 1.3 to 5.1 s, respectively, across subjects and LEFs. In the following, we consider the interval adjustments of stage II of the experiment only.

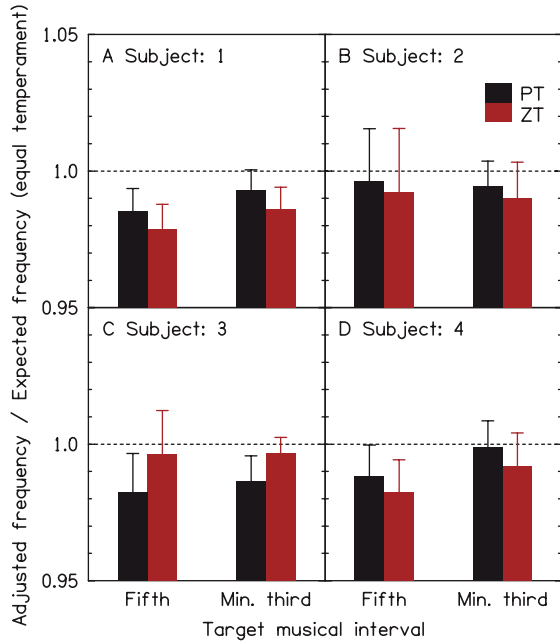
### 2.3 *Musical Interval Adjustments*

In stage II, subjects adjusted a sinusoid to be a specified musical interval below the preceding reference tone. The musical intervals that had to be adjusted were a minor third (3 semitones down) and a perfect fifth (7 semitones down). In different sessions, the reference tone was either a ZT or a physically presented pure tone (PT). The latter corresponded to the matched tones from stage I. From one match to the next, the reference tone could be any one of the set of eight. The general trial structure was the same as in stage I. Typically about 12 (at least 10) interval adjustments were collected for each condition from each subject.

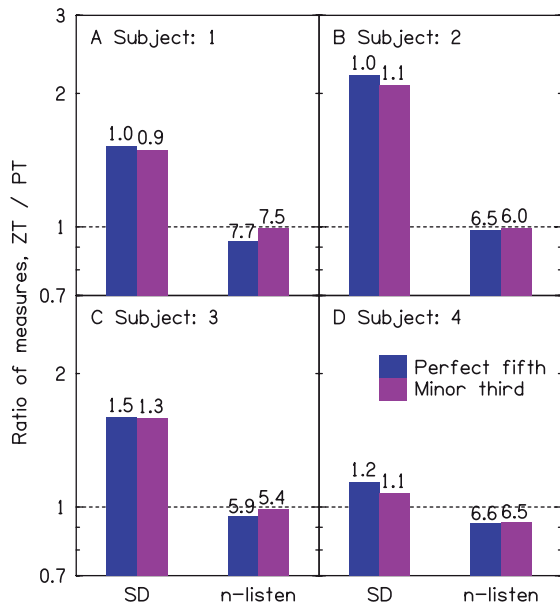
## 3 Results

Figures 1 and 2 show, for each subject, several measures of the accuracy, repeatability, and ease of the musical interval adjustments in the ZT and PT conditions. Figure 1 shows the geometric mean (and SDs across the 8 reference frequencies or LEFs) of the ratio of the adjusted frequency of the variable pure tone to the expected frequency. 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 3 semitones (a factor of 1/1.189) and 7 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 (different coloured bars). Generally,

**Fig. 1** Geometric mean (and SDs across the 8 reference frequencies or LEFs) of the ratio of the adjusted frequency of the variable pure tone to the expected frequency for musical intervals of a perfect fifth and a minor third



**Fig. 2** Ratio of the geometric mean (across the 8 reference frequencies or LEFs) of the SDs (*left-hand side*) and of the average number of trials (*right-hand side*) between ZT and PT conditions, for musical intervals of a perfect fifth and a minor third. 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



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.

To compare the reliability of the musical interval adjustments in the ZT conditions with that in the matched PT conditions, we calculated the ratio of the respective SDs. This was done as follows: (1) Separately for each subject, each musical interval (minor third, perfect fifth), each reference condition (ZT, PT) and each LEF (or reference PT tone frequency) we determined the SD of the adjusted frequencies across the 12 interval adjustments expressed as a percentage of the geometric mean; (2) These were geometrically averaged across the eight 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 in the group of two bars on the left-hand side of each panel in Fig. 2. Musical interval adjustments for the ZTs were somewhat more variable than for the matched PTs; the ratio of the SDs was between 1.1 and 2.2. Four univariate analyses of variance conducted separately on the data for each subject (using the SD of the logarithms of the individual adjusted frequencies within a given LEF or reference frequency as input data) showed that for three of the subjects the SDs were significantly larger for the ZT than for the PT conditions (with  $p < 0.01$  for subjects 1 and 2, and  $p < 0.05$  for subject 3). The number above each bar in Fig. 2 indicates the size of the representative SD for the PT condition in percent (ranging from 0.9–1.5%). The representative SD for the ZT condition ranged from 1.2–2.5%. By comparison, the SDs observed by Burns and Feth (1983) for the same musical interval adjustments were on average 1.2% and 4.2% for 1-kHz and 10-kHz 70-dB SPL reference tones, respectively.

The reliability of adjustments can be increased by listening more often to the stimuli, and thus needs to be considered as well. The group of two bars on the right-hand side of each panel in Fig. 2 shows the ratio of the average number of listening times (trials) in the ZT and the PT conditions. The number above each bar indicates the average number of trials ( $n$ -listen) for the PT condition. There were inter-individual differences in how many trials a subject needed on average to make a match. For example, in the PT conditions, subject 1 needed on average 7.6 trials while subject 3 needed on average 5.7 trials. However,  $n$ -listen was very similar for the two conditions, i.e. the ratio was very close to one. If anything,  $n$ -listen was very slightly lower for the ZTs than for the PT. 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 the PT conditions.

## 4 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 result from subjects' bias towards a "just scale" with musi-

cal intervals corresponding to integer ratios of 6/5 and 3/2 for the minor third and the perfect fifth, respectively. If this was the only underlying reason, the “bias” should have been larger for the minor third than for the perfect fifth, which was not observed. The SDs of the musical interval adjustments were only slightly larger (a factor of 1.1–2.2) for the ZTs than for the PTs, with no increase in listening time. Thus, overall the results suggest that a weak musical pitch can exist in the absence of peripheral phase locking.

Burns and Feth (1983) had three musically trained subjects adjust various musical intervals for reference frequencies of 1 kHz and 10 kHz. For the same musical intervals as used in the present study, SDs in their data were a factor of 1.6–11.7 (mean of 4) times larger for the 10-kHz reference than for the 1-kHz reference. Thus, the increase in SDs for the 10-kHz frequency was larger than that observed here for the ZTs. Burns and Feth (1983) interpreted their results in terms of phase locking, and a decrease thereof with increasing frequency. The present results suggest that their relatively large increase in SDs at 10 kHz reflects the combined effect of lack of familiarity with high-frequency pure tones and the transition from a temporal code to a place code.

**Acknowledgments** Supported by intramural funding from the MRC. Thanks to Brian Moore for helpful comments.

**Open Access** This chapter is distributed under the terms of the Creative Commons Attribution-Noncommercial 2.5 License (<http://creativecommons.org/licenses/by-nc/2.5/>) which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

The images or other third party material in this chapter are included in the work’s Creative Commons license, unless indicated otherwise in the credit line; if such material is not included in the work’s Creative Commons license and the respective action is not permitted by statutory regulation, users will need to obtain permission from the license holder to duplicate, adapt or reproduce the material.

## References

- Attneave F, Olson RK (1971) Pitch as a medium—new approach to psychophysical scaling. *Am J Psychol* 84:147–166
- Burns EM, Feth LL (1983) Pitch of sinusoids and complex tones above 10 kHz. In: Klinke R, Hartmann R (eds) *Hearing—physiological bases and psychophysics*. Springer, Berlin, pp 327–333
- Cariani PA, Delgutte B (1996) Neural correlates of the pitch of complex tones. II. Pitch shift, pitch ambiguity, phase invariance, pitch circularity, rate pitch, and the dominance region for pitch. *J Neurophysiol* 76:1717–1734
- Johnson DH (1980) The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones. *J Acoust Soc Am* 68:1115–1122
- Krump G (1993) *Beschreibung des akustischen Nachtönes mit Hilfe von Mithörschwellenmustern*. Ph.D. thesis. Technical University, München, Germany
- Meddis R, O’Mard L (1997) A unitary model of pitch perception. *J Acoust Soc Am* 102:1811–1820
- Moore BCJ (1973) Frequency difference limens for short-duration tones. *J Acoust Soc Am* 54:610–619

- Moore BCJ, Ernst SMA (2012) Frequency difference limens at high frequencies: evidence for a transition from a temporal to a place code. *J Acoust Soc Am* 132:1542–1547
- Oxenham AJ, Micheyl C, Keebler MV, Loper A, Santurette S (2011) Pitch perception beyond the traditional existence region of pitch. *Proc Natl Acad Sci USA* 108:7629–7634
- Palmer AR, Russell IJ (1986) Phase-locking in the cochlear nerve of the guinea-pig and its relation to the receptor potential of inner hair-cells. *Hear Res* 24:1–15
- Sek A, Moore BCJ (1995) Frequency discrimination as a function of frequency, measured in several ways. *J Acoust Soc Am* 97:2479–2486
- Ward WD (1954) Subjective musical pitch. *J Acoust Soc Am* 26:369–380
- Wiegrebe L, Kössl M, Schmidt S (1995) Auditory sensitization during the perception of acoustical negative afterimages—analogs to visual processing. *Naturwissenschaften* 82:387–389
- Zwicker E (1964) ‘Negative afterimage’ in hearing. *J Acoust Soc Am* 36:2413–2415