Acta Acustica united with Acustica
Vol. 104 (2018) 766 – 769

DOI 10.3813/AAA.919219

Detection of Mistuning in Harmonic Complex Tones at High Frequencies

Hedwig E. Gockel, Robert P. Carlyon
MRC-Cognition and Brain Sciences Unit, University of Cambridge, Cambridge CB2 7EF, UK.
hedwig.gockel@mrc-cbu.ac.uk

Summary

Lau *et al.* [1] showed that listeners could discriminate differences in the fundamental frequency (F0) of harmonic complex tones containing only very high frequency components, even though the discriminability of those components, when presented alone, was very poor. They attributed this finding to the operation of central pitch-sensitive neurons. We probed the characteristics of the hypothesized neurons by measuring the detection of mistuning of a single harmonic in stimuli similar to those of Lau *et al.* [1]. We mistuned the 8th harmonic in a complex tone containing harmonics 6–10 of an F0 of 1400 or 280 Hz with components presented either diotically or dichotically (odd harmonics in one ear and even harmonics in the other). Mistuning detection thresholds for both F0s were very low in the diotic condition, consistent with listeners detecting peripheral interactions ("beats") between nearby harmonics. In the dichotic condition, subjects reported that, for the low F0, the mistuned component "popped out" and this led to good performance. For the high F0, no such effect was heard and performance was close to chance. Thus, if there are pitch-sensitive neurons at very high frequencies, they do not provide a basis for perceptual segregation based on mistuning.

PACS no. 43.66.-x

1. Introduction

The low (residue) pitch of a complex tone may be derived from (1) place and/or temporal fine structure (TFS) information relating to the frequencies of individual resolved components, i.e. components with harmonic ranks below about 7 [2]; (2) TFS combined with envelope information derived from a region on the basilar membrane (BM) where harmonics with ranks between about 7 to 14 are interfering [3, 4]; (3) The envelope repetition rate of the waveform evoked on the BM by harmonics with higher ranks. Detection of envelope fluctuations is very poor for rates above about 1 kHz [5], so for high fundamental frequencies (F0s) only the first two of the above processes are likely to operate. Furthermore, phase locking to TFS becomes very weak at high frequencies and is probably completely unusable for frequencies above 5–8 kHz [6, 7, 8]. Therefore, if TFS is essential for processes (1) and (2) to operate, then no residue pitch should be perceived for a complex tone with all frequency components above about 8 kHz and with F0 above about 1 kHz.

Recently, Lau *et al.* [1] described an experiment that seems to contradict this prediction. They measured difference limens for the F0 (F0DLs) of a complex tone containing harmonics 6–10 of a nominal F0 of 1400 Hz, and frequency difference limens (FDLs) for each of the harmonics presented in isolation. All stimuli were presented in threshold equalizing noise (TEN) [9]. The FDLs were very large, ranging from about 22% for a 8.4 kHz tone to 35% for a 14 kHz tone. However, the F0DLs were much smaller (about 5%). Indeed, they were smaller than predicted from the optimal combination of information from the individual harmonics, assuming that performance is limited by peripheral coding variability. Similar results were obtained for diotic presentation and dichotic presentation, when even harmonics were presented to one ear and odd harmonics to the other. Lau *et al.* [1] suggested that the results can be explained by the existence of central, possibly cortical, pitch-sensitive neurons that respond selectively to a combination of harmonics but not to any individual harmonic, and that this activation leads to the percept of residue pitch.

Here, we first replicated measurement of the F0DLs for diotic and dichotic stimulus presentation. Second, we investigated listeners' sensitivity to mistuning of a single component in an otherwise harmonic complex sound, for complex tones resembling those used by Lau *et al.* [1]. For low F0s, mistuning of a single harmonic may be detected in two ways: (1) For low resolved harmonics (up to about the 6th) a mistuned component appears to "pop out" and is heard as a separate tone from the rest of the com-

Received 27 February 2018,
accepted 10 July 2018.

plex [10, 11], probably because it is no longer consistent with a harmonic template or no longer passes through a "harmonic sieve" [12]. (2) For higher unresolved harmonics (above about the 8th), the mistuned harmonic may be detected as temporal fluctuation or "beat", resulting from interaction between the mistuned harmonic and the other harmonics. In this case, very small amounts of mistuning can be detected if the stimuli have a long duration [10]. If, for the 1400-Hz F0, a residue pitch results from the activation of pitch sensitive neurons by a combination of higher harmonics, the pitch sensitive neurons may act like a harmonic template and the mistuned component may be segregated from the rest of the complex.

The harmonics in Lau *et al.*'s [1] stimuli would have interfered to some extent, at least for the diotic stimuli, potentially allowing mistuning to be detected via beat detection rather than via deviation of the mistuned component from a harmonic template. Therefore, we also measured mistuning detection for a dichotic condition similar to that used by Lau *et al.* This doubled the frequency separation between adjacent harmonics in each ear, thereby greatly reducing the interference between harmonics and reducing the likelihood of beat detection. The results for the dichotic condition were therefore expected to give a better indication of the ability to detect mistuning based on a harmonic- template-like mechanism.

2. Method

2.1. Subjects

Five young normal-hearing musically trained subjects between 21 and 28 years of age took part in the experiment proper. As in Lau *et al.* [1], all subjects were required to pass a 3-stage screening: (1) Pure tone thresholds in quiet at octave frequencies from 0.25–8.0 kHz had to be $<$ 20 dB HL. (2) Masked thresholds for pure tones at 10, 12, and 14 kHz in a TEN, the same TEN as in the main experiment, had to be $\leq$ 45 dB SPL. The TEN extended from 20 Hz–22 kHz and, at 1 kHz, had a level of 45 dB SPL/ERB$_N$, where ERB$_N$ stands for the average value of the equivalent rectangular bandwidth of the auditory filter for young NH listeners tested at low sound levels [13]. (3) F0DLs and FDLs for the same stimuli as in the main experiment but without the TEN and without level randomization had to be $<$ 6% and $<$20% in the low and the high spectral frequency regions, respectively (see below). Fourteen young subjects were tested, five of whom passed all screening stages.

2.2. Stimuli and procedure

The stimuli were very similar to those used by Lau *et al.* [1]. The complex tones consisted of harmonics 6–10. In the low-frequency condition, the F0 was 280 Hz. In the high-frequency condition, the F0 was 1400 Hz. Harmonics were presented either diotically (all harmonics to both ears) or dichotically (odd harmonics to the left and even harmonics to the right ear). The harmonics had random starting phases for each presentation. The presentation level was 55 ±3 dB SPL for the inner components (harmonics 7–9) and 49 ±3 dB SPL for the edge components; the level randomization was independent across components and presentations. The tones were presented in a continuous TEN background with a level of 45 dB SPL /ERB$_N$, (see above) via Sennheiser HD 650 headphones. The TEN was presented diotically in diotic tone conditions, while independent TENs were presented to each ear in the dichotic tone conditions. These stimulus parameters apply to all experiments described here and are identical to those used by Lau *et al.* [1], except that here we used continuous rather than gated TEN. Subjects were seated in a double-walled sound-insulated booth.

Experiment 1 measured F0DLs for diotic and dichotic 210-ms harmonic complexes, using a two-alternative forced-choice (2AFC) adaptive procedure with a 3-down 1-up rule tracking 79% correct. Observation intervals were marked by boxes on a screen. Feedback was provided after each trial. The starting value for the F0 difference was 20%, with the F0 of the two complexes centered geometrically on the nominal F0. The initial stepsize, a factor of 2, was decreased to a factor of 1.41 after two reversals, and to a factor of 1.2 after four reversals. The threshold was determined as the geometric mean of the frequency difference at eight more reversals collected at the smallest stepsize. For each subject and condition, the final threshold corresponds to the geometric mean of five such threshold estimates.

Experiment 2A measured thresholds for detecting mistuning of the 8th harmonic for complex tones of 210, 1000, and 2000-ms duration. One random interval contained a harmonic complex, and the other interval contained the mistuned complex where the 8th harmonic was randomly shifted upwards or downwards from its nominal frequency by a certain amount. This amount was specified as a percentage of the F0, and was limited to a maximum of 50%. The same adaptive procedure was used as in Experiment 1, with the following exceptions. After some preliminary testing indicating that subjects seemed to be able to detect the mistuning on the basis of beats, the starting value for the mistuning was set to 0.7% and 0.1% of the F0 for the low and the high F0s, respectively, in both the diotic and the dichotic conditions. This was done to shift the 8th harmonic by similar amounts in Hz for both F0s (and thus give similar beat rates), and because slow beats are salient [14, 15].

An adaptive run was terminated if the procedure required a mistuning larger than 50% of the F0 more than six times in a row. Because this happened mainly in the dichotic condition for the high F0, Experiment 2B measured percent correct for detection of mistuning in the dichotic condition for a 1-s stimulus duration. The amounts of mistuning were fixed at 50% and at 0.356% or 1.78% for the high and the low F0, respectively. The larger amounts were chosen so as to maximize the chance of perceptual segregation of the mistuned component from the remainder of the complex, while the smaller amounts (5 Hz as a percentage of the high and low F0s) were chosen so as to maxi-

Table I. Mean F0DLs in percentage (range).

	Diotic	Dichotic
F0 = 280 Hz	0.43 (0.32-0.71)	0.38 (0.20-0.66)
F0 = 1400 Hz	2.18 (0.97-4.49)	2.05 (0.64-4.12)

mize the salience of a possible beat cue. In Experiment 2A, at least 6 adaptive runs were collected for each subject and condition, and the reported threshold is the median value across these runs. In Experiment 2B, at least 200 trials were collected for each subject and condition.

3. Results and discussion

3.1. Experiment 1

Table I shows the mean F0DLs (and overall range) averaged across all subjects. Overall, thresholds were about a factor of 2 smaller than those reported by Lau et al. [1], possibly because of more musical training in our listeners or because the present study used continuous TEN while Lau *et al.* [1] used gated TEN. Lau *et al.* [1] argued that if performance was based on temporal envelope cues, then the wider component spacing in the dichotic condition should result in larger F0DLs due to the attenuation of interaction between adjacent components and the increase in envelope rate. F0DLs were not significantly different between diotic and dichotic conditions for either the low or the high F0 ($p = 0.43$ and $p = 0.70$, paired-samples t tests). This replicates Lau *et al.*'s finding and is consistent with the idea that F0 discrimination was not based on temporal envelope fluctuations.

3.2. Experiment 2A

Figure 1 shows mistuning detection thresholds as a function of stimulus duration.

In the diotic condition (triangles), thresholds increased with decreasing stimulus duration for both frequency regions. For the 1s and 2s durations, all subjects reported using a beat cue to select the complex with the mistuned component. Low amplitude modulation rates around 4–5 Hz have the highest fluctuation strength [14], and thus beats with rates of 4–5 Hz would be expected to be most salient. For the long durations, thresholds were below 5 Hz for both F0s; the horizontal dashed and solid lines indicate 5 Hz as a percentage of 1400 Hz and 280 Hz, respectively. For the short duration, only subjects 1 and 5 were able to consistently make use of beat cues. For short durations, beat cues become unavailable, as fewer periods of the envelope modulation are present in the short stimulus duration. The results show that in the diotic condition, adjacent components clearly interact in the periphery, even in the high frequency region where it has been proposed that, at least at low levels, auditory filters are much narrower than commonly assumed [16].

In the dichotic condition, thresholds did not depend on duration and subjects did not report using beats. For the low F0, all except one subject (subject 4) could detect the mistuning on the basis of perceived segregation. The data of subject 4 were excluded from the calculation of the mean for the low-F0 dichotic condition. For the high F0, adaptive runs often hit the upper limit and terminated, for all subjects. There was no segregation cue available, and performance seemed close to chance.

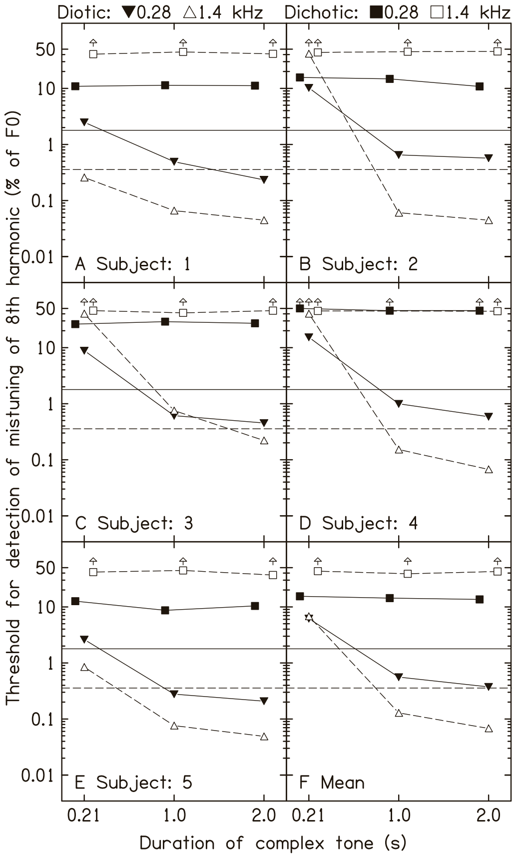

Figure 1. Mistuning detection thresholds for each subject (A-E) and geometric means across subjects (F). Upward pointing arrows indicate conditions where adaptive runs hit the upper limit. The horizontal dashed and solid lines indicate 5 Hz as a percentage of 1400 Hz and 280 Hz, respectively.

3.3. Experiment 2B

Figure 2 shows percent correct for mistuning detection in the dichotic condition.

For the 5-Hz mistuning, performance was close to chance (50%) for both F0s, indicating that a beat cue was not available in the dichotic condition. When the 8th harmonic was mistuned by 50% of the F0, performance was well above chance level for the low F0 for all except one subject (subject 4, treated as an outlier and not included in the average), and was at ceiling for two out of the five subjects. Subjects reported using a segregation cue. In contrast, for the high F0 performance was only slightly above chance level for the mistuning of 50%. Subjects reported no segregation of the mistuned component from the remainder of the complex. Instead they tried to use small differences in timbre between the harmonic and the mistuned complex.

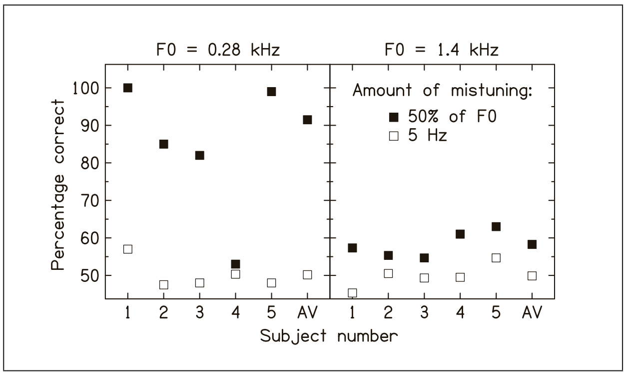

Figure 2. Percentage correct for detection of mistuning of the 8th harmonic in a dichotic complex tone consisting of harmonics 6–10. AV means average.

4. General discussion

The results of the first experiment replicated the finding of Lau *et al.* [1] of similar F0DLs for a diotically presented harmonic complex tone and for one where the odd and even harmonics were presented to opposite ears. This is consistent with the assumption that performance in this task was not based on temporal envelope cues. The results of Experiment 2A suggested that, in the diotic condition, mistuning detection was based on beating cues for both spectral regions. Thus, useful temporal envelope information was available in both spectral regions at lower rates than F0. This implies interaction of harmonics in the auditory periphery, which in turn might impair the ability of a harmonic-template-like mechanism to extract residue pitch. The slow salient beat cues most probably arose from interaction of the mistuned component with more than just one adjacent component, as (i) the beat rate between the mistuned component and the closest adjacent harmonic, corresponding to the frequency difference between them, would have been too high to be usable, and (ii) beats between the mistuned component and a weak distortion component were unlikely to be audible, given that the level of the distortion product would have been 15–20 dB below the level of the mistuned component.

Thresholds for mistuning detection were much higher in the dichotic than in the diotic condition, at least for longer durations. In contrast to the diotic condition, dichotic presentation did not provide temporal-envelope-based cues for the detection of mistuning. For the low spectral region, where phase locking and TFS information would be available, the mistuned component perceptually segregated from the remainder of the complex tone and this provided a cue for detecting mistuning. In contrast, for the high spectral region the mistuned component was not perceived as a separate auditory object, and performance was much worse. If residue pitch perception for very high spectral components is based on central pitch-sensitive neurons that respond selectively to a combination of harmonics, as suggested by Lau *et al.* [1], the harmonic template associated with these neurons either has wider slots than for low F0s, or it has comparable slots to those for low F0s but, unlike at low F0s, there is no mechanism that gives rise to a salient percept of the rejected component. The results are compatible with the idea that TFS information is necessary in order for perceptual segregation of the mistuned component to occur.

Acknowledgement

This work was supported by intramural funding from the MRC (MC_UU_00005/3). We thank Brian Moore for helpful comments on this manuscript.

References

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

[2] B. C. J. Moore, H. E. Gockel: Resolvability of components in complex tones and implications for theories of pitch perception. Hear Res **276** (2011) 88–97.

[3] J. F. Schouten: The residue and the mechanism of hearing. Proceedings of the Koninklijke Akademie van Wetenschap **43** (1940) 991–999.

[4] G. A. Moore, B. C. J. Moore: Perception of the low pitch of frequency-shifted complexes. J Acoust Soc Am **113** (2003) 977–985.

[5] E. M. Burns, N. F. Viemeister: Nonspectral pitch. J Acoust Soc Am **60** (1976) 863–869.

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

[7] A. Recio-Spinoso, A. N. Temchin, P. van Dijk, Y. H. Fan, M. A. Ruggero: Wiener-kernel analysis of responses to noise of chinchilla auditory-nerve fibers. J Neurophysiol **93** (2005) 3615–3634.

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

[9] B. C. J. Moore, M. Huss, D. A. Vickers, B. R. Glasberg, J. I. Alcántara: A test for the diagnosis of dead regions in the cochlea. Brit J Audiol **34** (2000) 205–224.

[10] B. C. J. Moore, R. W. Peters, B. R. Glasberg: Thresholds for the detection of inharmonicity in complex tones. J Acoust Soc Am **77** (1985) 1861–1867.

[11] B. C. J. Moore, B. R. Glasberg, R. W. Peters: Thresholds for hearing mistuned partials as separate tones in harmonic complexes. J Acoust Soc Am **80** (1986) 479–483.

[12] H. Duifhuis, L. F. Willems, R. J. Sluyter: Measurement of pitch in speech: An implementation of Goldstein's theory of pitch perception. J Acoust Soc Am **71** (1982) 1568–1580.

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

[14] H. Fastl: Fluctuation strength of modulated tones and broadband noise. – In: Hearing – Physiological Bases and Psychophysics. R. Klinke, R. Hartmann (eds.). Springer, Stuttgart, 1983, 282–288.

[15] A. Kohlrausch, R. Fassel, T. Dau: The influence of carrier level and frequency on modulation and beat-detection thresholds for sinusoidal carriers. J Acoust Soc Am **108** (2000) 723–734.

[16] C. A. Shera, J. J. Guinan, Jr., A. J. Oxenham: Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements. Proc Natl Acad Sci USA **99** (2002) 3318–3323.