Pitch perception: Frequency selectivity and temporal coding

ANDREW J. OXENHAM[1] AND MICHAEL V. KEEBLER[1]

[1] *Department of Psychology, University of Minnesota, Minneapolis, MN 55455, USA*

A number of studies have shown that the ability to accurately discriminate small changes in fundamental frequency (F0) varies with the lowest harmonic present in the complex. When low-order harmonics are present, F0 difference limens (DLs) are generally small, indicating good performance. When only high-order harmonics (greater than the 10th) are present, performance can be worse by an order of magnitude. Poor performance when only high-number harmonics are present has been ascribed to a lack of peripherally resolved harmonics. Here we tested this notion by measuring F0DLs for complexes consisting of twelve consecutive harmonics over a wide range of F0s (30 – 2000 Hz) as a function of the lowest harmonic number present. For F0s between 100 and 200 Hz, performance went from good to poor as the lowest harmonic present increased from 9 to 12. In contrast, at lower and higher F0s, the transition occurred at lower harmonics in ways that would not be necessarily predicted simply by harmonic resolvability and frequency selectivity. At high F0s, good performance was often observed even when all the harmonics were above 6 kHz, and some harmonics were peripherally resolved, suggesting that temporal fine-structure coding of individual harmonics may not be a prerequisite for complex pitch perception.

INTRODUCTION

Pitch is a fundamental auditory percept that, for most natural stimuli, co-varies with the fundamental frequency (F0) of the sound. We ascribe a pitch corresponding to the F0 of the harmonic complex (such as a note on an instrument or a voiced speech sound), even the complex has no energy at the actual F0 itself. This phenomenon is known as the pitch of the missing fundamental, or periodicity pitch (for a recent review, see Plack *et al.*, 2005). Pitch plays a crucial role in music and speech, be it as prosody in non-tone languages, such as English, or as lexical information in tone languages, such as Cantonese or Mandarin. Despite numerous systematic studies of pitch over several decades, the neural mechanisms underlying pitch remain unclear. In particular, even on the most basic questions, such as whether pitch is coded via place or timing mechanisms in the cochlea, there is surprisingly little consensus.

A number of studies have investigated the effects of presenting only a selection of all possible harmonics, with the aim of discovering which harmonics contribute most to the overall pitch (Plomp, 1967; Moore *et al.*, 1985; Dai, 2000), or of discovering how accuracy in pitch coding depends on which harmonics are present (Hoekstra, 1979; Houtsma and Smurzynski, 1990; Shackleton and Carlyon, 1994; Bernstein and Oxenham, 2003). Such studies have typically found that low-order harmonics are dominant in determining the overall pitch of the complex. Furthermore, the ability to discrim-

Auditory signal processing in hearing-impaired listeners. 1st International Symposium on Auditory and Audiological Research (ISAAR 2007). T. Dau, J. M. Buchholz, J. M. Harte, T. U. Christiansen (Eds.). ISBN: 87-990013-1-4. Print: Centertryk A/S.

inate small changes in F0, as estimated by the F0 difference limen (F0DL), degrades dramatically, as the lowest harmonic present in a complex increases beyond about the 10^{th} (Houtsma and Smurzynski, 1990). A link has been drawn repeatedly between the transition between good and poor F0DLs and the transition between resolved and unresolved harmonics, respectively, as determined by the frequency selectivity of the peripheral auditory system (Hoekstra, 1979; Houtsma and Smurzynski, 1990; Shackleton and Carlyon, 1994; Bernstein and Oxenham, 2006a). Interestingly, however, most of the relevant data have been collected using F0s of between 100 and 200 Hz. The purpose of this study was to provide a more stringent test of the relationship between F0DLs and frequency selectivity, by testing performance over a much wider range of F0s. Some earlier studies have examined F0DLs over ranges wider than just 100-200 Hz; however, they have used bandpass filtering to limit the harmonics present, and were not able to rule out the possibility that listeners could complete the task based on the frequency shifts of the lowest audible harmonic, as opposed to the F0 of the whole complex (Hoekstra, 1979; Krumbholz *et al.*, 2000). It is important to understand the role of frequency selectivity in pitch perception because a loss of frequency selectivity is one of the most common symptoms of sensorineural hearing loss. Understanding how hearing impairment affects frequency selectivity and pitch perception, and determining whether the two are related, should help us in our quest to better alleviate the symptoms of hearing loss (e.g., Bernstein and Oxenham, 2006b).

F0DLs AS A FUNCTION OF F0 AND HARMONIC NUMBER

Methods

All the stimuli were presented in a background of noise to prevent the detection of distortion products and to ensure that all the components were presented at approximately equal sensation level up to frequencies of 16 kHz. To confirm audibility, detection thresholds for single pure tones at frequencies between 60 and 20000 Hz were measured using an adaptive (2-down 1-up) 3-interval 3-alternative forced-choice procedure in quiet and in a background of threshold equalizing noise (TEN) (Moore *et al.*, 2000) set to a level of 45 dB SPL per ERB. The threshold for each condition was measured twice and the mean of the two runs was calculated. To qualify for participation in the study, a subject's detection thresholds in noise had to be 50 dB SPL or lower at all frequencies up to and including 16 kHz. In fact, almost all thresholds fell below 45 dB SPL. Six subjects, all with extensive musical training, took part in this study. They had audiometric thresholds of 15 dB HL or less at octave frequencies between 125 and 8000 Hz. Figure 1 shows the thresholds in quiet and in noise, averaged across the six listeners, and indicates that the TEN was successful at producing roughly equal masked thresholds over a wide frequency range. Thresholds at 20 kHz are not shown (and were not analyzed further) because they were quite variable within and between subjects, and were not always measurable. A repeated-measures analysis of variance (RM-ANOVA), including a Greenhouse-Geisser sphericity correction, failed to show a significant effect of frequency ($F_{2.4,12.1} = 2.96$, $p = 0.083$) and displayed no linear or quadratic trends ($p > 0.2$). The average threshold in noise, pooled across subjects and frequencies between 60 and 16000 Hz, was 39.9 dB SPL.

In the F0DL experiment, listeners were presented with two successive complex tones, separated by an inter-stimulus interval of 500 ms, and were asked to judge which had the higher pitch or F0. The duration of the tones was 500 ms, including 30-ms raised-cosine onset and offset ramps. The tones were presented in a TEN background noise at a level of 45 dB SPL per ERB, as in the detection experiment. The complex tones were comprised of twelve consecutive harmonics, with their levels set to 55 dB SPL per component. Thus each component was approximately 15 dB above its detection threshold in the noise. If the frequency of a component was higher than 20 kHz, it was omitted from the complex. The components in each complex were added in random phase, with new starting phases selected for each presentation. The nominal lowest harmonic number present, N, was 3, 6, 9, 12 or 15. To discourage listeners using the spectral edge of the stimulus to solve the task, as opposed to the F0, the lowest harmonic present was roved across trials, such that the actual lowest harmonic number was *N*-1, N, or *N*+1. In each trial, the lowest harmonic numbers for the two intervals were selected without replacement from these three possibilities (see also Houtsma and Smurzynski, 1990).

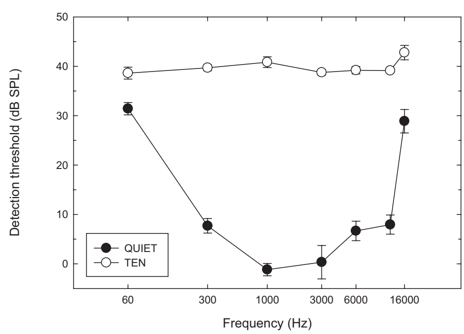

Fig. 1: Detection thresholds for pure tones in quiet (filled symbols) and in threshold-equalizing noise (TEN) with a level of 45 dB SPL per ERB. Error bars represent ±1 s.e. across six listeners.

The F0 difference (ΔF0) between the two intervals was initially 10%. This value was increased or decreased by a factor of 1.41 after one incorrect or two consecutive correct responses, respectively (2-down 1-up adaptive procedure). After the first four reversals in the adaptive procedure, the step size was reduced to a factor of 1.2 and the run continued for another six reversals. Threshold was defined as the geometric mean value of ΔF0 at the last six reversals. At least four such runs were completed by each subject in each condition. The order of conditions (F0s and N) was selected randomly with the constraint that each condition should be run before any condition was repeated.

Results

The pattern of results was quite similar between subjects, and so only the mean data from the six subjects are shown. Figure 2 shows how F0DLs vary as a function of lowest harmonic number (N) for various reference F0s. Consider first the results using

F0s of 100 and 200 Hz (Fig. 2, upper right panel). These are conditions that have been tested most in earlier studies (e.g., Houtsma and Smurzynski, 1990; Bernstein and Oxenham, 2003). In line with previous data, thresholds are low (good) for low values of N and increase at higher values of N, reaching what appears to be a plateau at an N of 12, at least for an F0 of 200 Hz. If one takes an F0DL of 2% to be the cut-off between good and poor pitch perception, the transition occurs between N=9 and N=12 for both F0s.

At the lower F0s of 30 and 50 Hz (Fig. 2, upper left panel), the pattern is rather different. At 30 Hz, performance is consistently poor for all values of N, in line with having reached the lower limits of pitch (Krumbholz *et al.*, 2000). At 50 Hz, performance at low values of N is reasonable, but thresholds seem to reach a plateau at a slightly lower value of N than was seen at 100 and 200 Hz. At higher values of F0 (Fig. 2, lower panels), performance is more clearly divided between good and poor performance as a function of F0. However, the transition between good and poor performance occurs consistently between N=6 and N=9, rather than between N=9 and N=12, as was found at 100 and 200 Hz.

Discussion

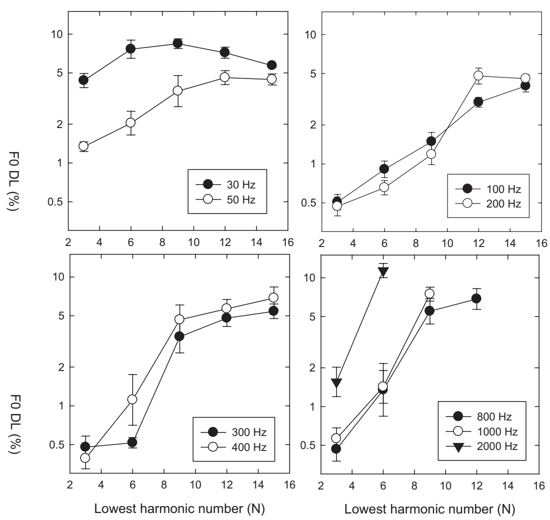

Fig. 2: F0DLs as a function of average lowest harmonic number (N) for F0s ranging from 30 Hz to 2000 Hz.

The results are not consistent with the predictions of a theory based solely on frequency selectivity. The bandwidths of the normal auditory filters can be approximated as a constant proportion of the filter's centre frequency at least for frequencies above about 1000 Hz (Glasberg and Moore, 1990). Therefore, if peripheral frequency selectivity determined performance, then one would expect the transition point between good and poor performance to remain roughly constant, when expressed in terms of N. This is not the case; in fact, the only conditions that conform to the earlier work, suggesting that the transition between good and poor performance occurs for values of N between 9 and 12, are the 100- and 200-Hz F0s, which were the only F0s used in the most comparable earlier studies (Houtsma and Smurzynski, 1990; Bernstein and Oxenham, 2003).

The change in the transition point does not seem to be simply an effect of absolute frequency, with poorer coding at higher frequencies: for F0s of 100 Hz or higher (with the exception of the 2000-Hz F0) all F0DLs for N=3 are around 0.5% and show no deterioration with increasing F0. Overall, it does not seem possible to account satisfactorily for these data with current models. The data do not at first sight appear consistent with earlier data that measured F0DLs of filtered click trains as a function of the lower cutoff frequency of the filter (Ritsma and Hoekstra, 1974; Cullen and Long, 1986), which found a roughly constant pattern of F0DLs across different F0s, when the results were plotted as a function of N. One reason for this discrepancy may be that the earlier studies did not rove the lowest harmonic present, leaving open the possibility that listeners were able to perform the task based on the frequency of the lowest harmonic present, and not on the F0 itself.

Another noteworthy aspect of the data is that there are instances, such as F0 = 1000 Hz, N = 6 and F0 = 2000 Hz, N = 3 (Fig. 2, lower right panel), in which all the harmonics are above 6 kHz, but good F0 discrimination still appears to be possible. This contradicts Ritsma's (1962) "existence region", whereby only components below about 6 kHz elicit a strong pitch percept. However, Ritsma only used three consecutive harmonics (as opposed to our 12), and it is known that the pitch elicited by so few harmonics is generally weak. Also, Ritsma did not use a background noise; it is possible that the presence of the noise enhanced the tendency to hear the underlying F0, through means of some form of spectral completion, whereby the noise induces the percept of some of the missing harmonics, possibly including the F0 (Houtgast, 1976; Hall and Peters, 1981; Grose *et al.*, 2002). The implication here is that complex pitch, based on low-order resolved harmonics may be possible, even if all the harmonics are above the putative limits of temporal phase locking. This provides some indication that a temporal phase-locked code in the auditory nerve may not be necessary for complex pitch perception.

CONCLUSIONS

F0DLs were estimated as a function of the lowest harmonic number present (N) for wide range of F0s, from 30 to 2000 Hz. Several features of the results were noteworthy. First, in contrast to predictions based on frequency selectivity and harmonic resolvability, the relationship between F0DLs and N was not constant as a function of F0; the earlier finding of a transition between the 9th and 12th harmonic seems to be the exception rather than the rule. Second, accurate complex pitch perception was found in some cases for situations where all the harmonics were above 6 kHz. Because 6 kHz is above the frequency at which phase locking in the auditory nerve is believed to break down, this suggests that phase locking in the auditory nerve may not be necessary for complex pitch perception.

ACKNOWLEDGEMENTS

This work was supported by NIH grant R01 DC 05216.

REFERENCES

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

Bernstein, J. G., and Oxenham, A. J. (**2006a**). "The relationship between frequency selectivity and pitch discrimination: Effects of stimulus level," J. Acoust. Soc. Am. **120**, 3916-3928.

Bernstein, J. G., and Oxenham, A. J. (**2006b**). "The relationship between frequency selectivity and pitch discrimination: Sensorineural hearing loss," J. Acoust. Soc. Am. **120**, 3929-3945.

Cullen, J. K., and Long, G. R. (**1986**). "Rate discrimination of high-pass-filtered pulse trains," J. Acoust. Soc. Am. **79**, 114-119.

Dai, H. (**2000**). "On the relative influence of individual harmonics on pitch judgment," J. Acoust. Soc. Am. **107**, 953-959.

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

Grose, J. H., Hall, J. W., and Buss, E. (**2002**). "Virtual pitch integration for asynchronous harmonics," J. Acoust. Soc. Am. **112**, 2956-2961.

Hall, J. W., and Peters, R. W. (**1981**). "Pitch from nonsimultaneous successive harmonics in quiet and noise," J. Acoust. Soc. Am. **69**, 509-513.

Hoekstra, A. (**1979**). "Frequency discrimination and frequency analysis in hearing," (Institute of Audiology, University Hospital, Groningen, Netherlands).

Houtgast, T. (**1976**). "Subharmonic pitches of a pure tone at low S/N ratio," J. Acoust. Soc. Am. **60**, 405-409.

Houtsma, A. J. M., and Smurzynski, J. (**1990**). "Pitch identification and discrimination for complex tones with many harmonics," J. Acoust. Soc. Am. **87**, 304-310.

Krumbholz, K., Patterson, R. D., and Pressnitzer, D. (**2000**). "The lower limit of pitch as determined by rate discrimination," J. Acoust. Soc. Am. **108**, 1170-1180.

Moore, B. C., Huss, M., Vickers, D. A., Glasberg, B. R., and Alcantara, J. I. (**2000**). "A test for the diagnosis of dead regions in the cochlea," Br. J. Audiol. **34**, 205-224.

Moore, B. C. J., Glasberg, B. R., and Peters, R. W. (**1985**). "Relative dominance of individual partials in determining the pitch of complex tones," J. Acoust. Soc. Am. **77**, 1853-1860.

Plack, C. J., Oxenham, A. J., Popper, A. N., and Fay, R. (eds). (**2005**). Pitch: Neural Coding and Perception Springer Verlag, New York.

Plomp, R. (**1967**). "Pitch of complex tones," J. Acoust. Soc. Am. **41**, 1526-1533.

Ritsma, R. J. (**1962**). "Existence region of the tonal residue. I," J. Acoust. Soc. Am. **34**, 1224-1229.

Ritsma, R. J., and Hoekstra, A. (**1974**). "Frequency selectivity and the tonal residue," in Facts and Models in Hearing, edited by E. Zwicker, and E. Terhardt Springer, Berlin, 156-163.

Shackleton, T. M., and Carlyon, R. P. (**1994**). "The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination," J. Acoust. Soc. Am. **95**, 3529-3540.