UNIVERSITY OF CAMBRIDGE

Modelling the strength of Melodic pitch using time-interval information

D. Timothy Ives, Steven M. Bailey and Roy D. Patterson
dti20@cam.ac.uk steve.bailey@cantab.net rdp1@cam.ac.uk
Centre for the Neural Basis of Hearing, Physiology, Development and Neuroscience Department, University of Cambridge, Downing Street, Cambridge, CB2 3EG United Kingdom.
http://www.pdn.cam.ac.uk/groups/cnbh/

P37

1. Introduction

Residue pitch underlies our ability to hear the pitch of sounds, such as speech and music.
Schouten (1940) argued that residue pitch arose from the interaction of the higher unresolved components, resulting in a complex waveform with a period of the fundamental frequency (F0). This **temporal** theory can be consistent with a peripheral origin of residue pitch.
Houtsma and Goldstein (1972) showed that melodies constructed from two-component harmonic complexes are recognized equally well when presented dichotically as when presented monotically. Several researchers (e.g. Goldstein, 1973) proposed that the F0 might be determined by **spectral** cues using a 'central pattern recognition' model that identifies the common spacing between resolved harmonics.
Spectral models increase in resolution with increasing F0 whereas temporal model decrease in resolution with increasing F0. We present stimuli that show an effect of F0 to test both models.

2. Experimental procedure

Listeners were presented with two successive melodies, the second melody was a repetition of the first but had one of the notes changed. The task was to identify which note had changed in the second melody. Melodies were four notes from the diatonic major scale. The structure of the notes was such that only the residue pitch was consistent with the musical scale and the sinusoidal pitch could not provide consistent cues. To hear the intended melody, the listener was required to attend to the residue pitch and ignore the sinusoidal pitch. A melody task was used rather than a pitch discrimination task as it is a better measure of pitch strength.

A. Stimuli
The notes in the melodies were synthesized from a harmonic series with missing lower harmonics. The pitch of the note corresponded to the F0 of the harmonic series. Applied a low-pass filter of 6 dB / octave relative to the lowest component. Low-pass noise was used to mask distortion products. Note duration was 500 ms, including a 100 ms raised half-cosine onset and a 333 ms raised half-cosine offset. Stimuli were presented diotically at approximately 60 dB SPL.

Performance for the melody task was measured for four note synthesis parameters:
*(1).*fundamental frequency (**F0**) 75 and 300 Hz; *(2).*average lowest component (**ALC**) 3, 7, 11, or 15; *(3).*number of harmonic components (**NHC**) 2, 4, and 8; *(4).*lowest component rove (**LCR**) 1 and 3;

B. Procedure
For each pair of melodies, the fundamental frequency of the tonic was randomly selected from a half-octave range centered at either 75 Hz or 300 Hz depending on which F0 was being run. The intervals are therefore musical but, due to the roving fundamental, the notes generated rarely matched actual notes found on a keyboard. The four-note melody itself was generated by drawing the fundamental frequency for each note randomly, with replacement, from a range of five notes, made up of the tonic (***doh***) and the four notes above (***ray, me, fah, soh***), see Figure 1.

Four listeners were used in each of the F0 conditions three were common to both conditions. The tonic, which defines the scale and remained constant throughout the trial, was presented twice before each presentation of the melody. After the presentation of the second melody, there was an indefinite response interval, which is terminated by the subject's response. Feedback was then given as to which note actually changed, and another trial was begun.

With the exception of the F0, the note synthesis parameters, were combined to give 12 conditions (3×NHC 2×LCR 4×ALC). The order of these 12 conditions was randomized and constituted one repetition. Subjects performed three or four repetitions in a 20 minute block and about four or five blocks in a two hour session together with breaks. Data was collected on each subject for between 40 to 50 repetitions. The F0 conditions were run independently on separate days.

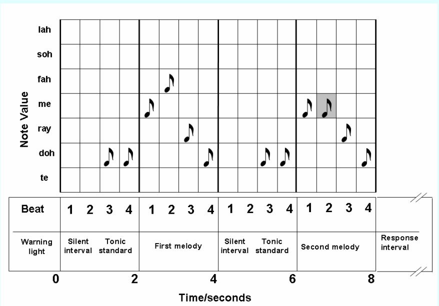

Figure 1. Melody task. One note changes (grey square) in the second melody by a single diatonic interval and must be identified by the listener.

3. Results

Figures 2(a) and (b) show that performance decreases as ALC is increased.

This is more apparent for the higher fundamental frequency of 300 Hz where the decrease in performance is sudden as the ALC is increased beyond 7. This decrease in performance for stimuli containing the higher unresolved components was markedly less for the lower F0 condition. This is significant as it highlights a difference in the prediction by a spectral and temporal model. A spectral model would predict that there should be no reduction in listener performance for unresolved harmonics when the fundamental frequency F0 is increased. Indeed, performance should decrease slightly for the lower F0 as the auditory filters become a little wider at the lower frequencies thereby decreasing spectral resolution. The performance decrease for unresolved harmonics with a higher F0 would be in accordance with a temporal model which predicts a decrease in temporal resolution due to a gradual reduction in the phase locking of nerve fibers.

As the NHC is decreased performance also decreases, the effect is more apparent for the lower fundamental frequency of 75 Hz. A temporal model would predict better listener performance for more harmonic components as the period of the fundamental would be defined better. A spectral model would predict an improvement in listener performance for resolved components, however the stimuli uses components in the unresolved region. Additional components in the unresolved region would lead to a reduction in spectral resolution and consequently listener performance.

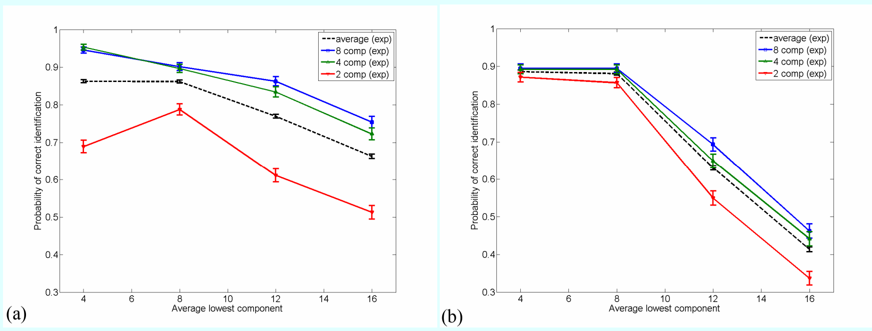

Figure 2. Results for the melody task with the **(a)** 75 Hz fundamental and **(b)** 300 Hz fundamental.
The abscissa shows the ALC and the ordinate shows the probability of the listener correctly identifying the note which had changed. Performance is plotted for each NHC condition as a function of ALC. The solid lines are the average of four listeners for each of the NHC conditions. The dotted line is the average of the three solid lines, plotted as a function of ALC. The errors bars are one standard deviation.

4. Analysis

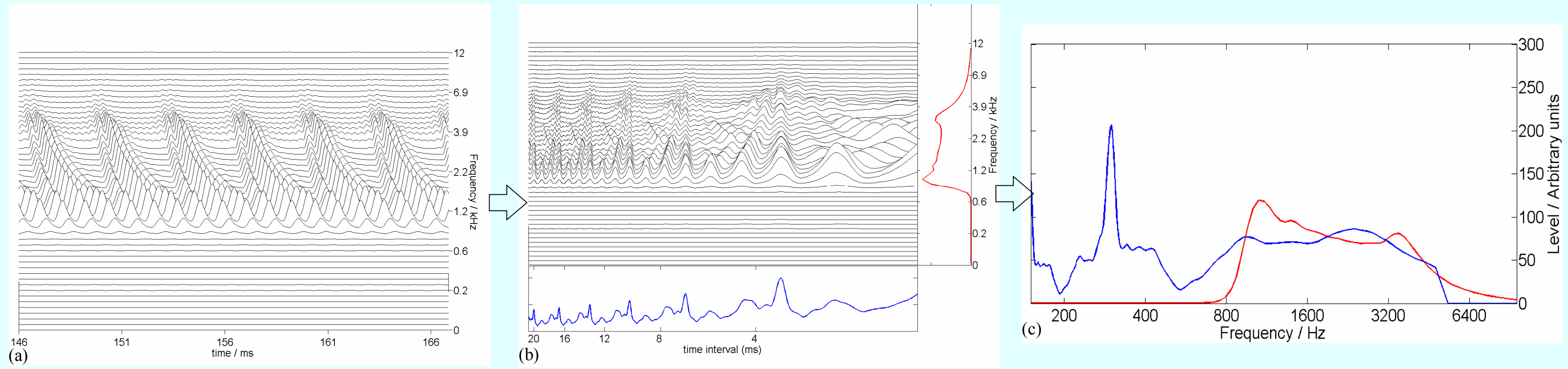

Figure 3. Generation of Dual profile from initial input. **(a)** A neural activity pattern (NAP) for a harmonic complex sound composed of the third to tenth harmonic of a 300 Hz fundamental. The horizontal ridges every 3 ms show the period of the fundamental (300 Hz).
(b) Stabilised Auditory Image (SAI) of preceding NAP. A neural unit for each NAP channel monitors its activity and locates peaks like those produced by the horizontal ridges in (a). The peaks cause the unit to 'strobe' the temporal integration process which a) measures the time intervals from the strobe time to succeeding peaks in the decaying resonance, and b) enters the time intervals into an interval histogram as they are generated (Patterson, 1994). The vertical and horizontal side panels in (b) are a summation across one of the axis. Summation across the frequency axis produces the **spectral profile** and summation across the time-interval axis produces the **temporal profile**.
(c) Dual Profile. The **spectral** and **temporal** profiles are plotted on the same axis, the time interval of the temporal profile is converted to the equivalent frequency.

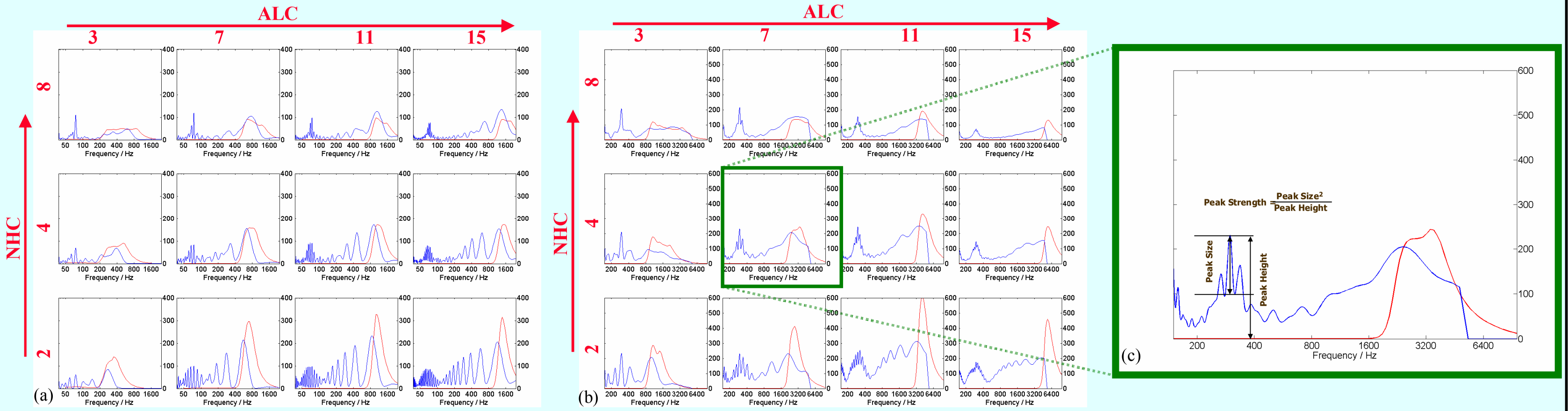

Figure 4. Dual Profiles for all stimuli. **(a)** F0 = 75 Hz, **(b)** F0 = 300 Hz. **(c)** measurement of Peak Strength from dual profile. The size and the height of the peak are used to determine the peak strength. The height is defined as the actual value of the peak and the size is defined as the height minus the average of each trough either side of the peak. The strength of the peak in the temporal profile is characterised as the square of the peak size divided by the height.

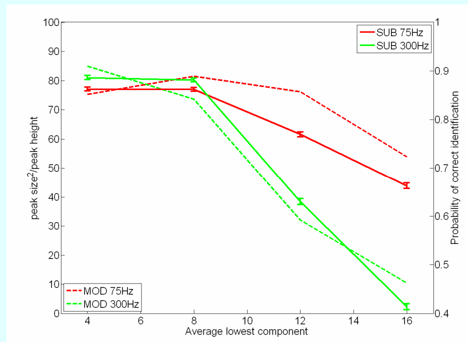

Figure 5. Comparison of experimental results with model. Solid lines are average subject data for an F0 of 75 Hz (**red** line) and for an F0 of 300 Hz (**green** line). Dotted lines are the model predictions for an F0 of 75 Hz (**red** line) and for an F0 of 300 Hz (**green** line).

5. Conclusions

Performance was worse for unresolved components with a high F0 but not for unresolved components with a low F0.

Temporal models predict this reduction in performance, due to an increasing loss of phase locking.

Spectral models incorrectly predict a performance increase due to wider auditory filters (increase in resolution) with increasing F0.

Acknowledgements

Research supported by the U.K. Medical Research Council (G0500221, G9900369).

References

Goldstein, J.L. (1973). "An optimum processor theory for the central formation of the pitch of complex tones," J. Acoust. Soc. Am. 54, 1496-1516.
Houtsma, A.J.M., and Goldstein, J.L. (1972). "The central origin of the pitch of complex tones: Evidence from musical interval recognition," J. Acoust. Soc. Am. 51, 520-529
Moore, B.C.J. and Rosen, S.M. (1979). "Tune recognition with reduced pitch and interval information," J. Exp. Psychol. 31, 229-240.
Moore, B.C.J., Glasberg, B.R., Flanagan, H.J. and Adams, J. (2006). "Frequency discrimination of complex tones; assessing the role of component resolvability and temporal fine structure," J. Acoust. Soc. Am. 119, 480-490.
Patterson, R.D. (1994). "The sound of a sinusoid: Time-interval models." J. Acoust. Soc. Am. 96, 1419-1428.
Schouten, J.F. (1940) "The residue and the mechanisms of hearing," Proc. Kon. Acad. Wentensch. (Neth) 43, 991-999.