Patterson, R.D., Peters, R. W. and Milroy, R. (1983). Threshold duration for melodic pitch. In R. Klinke and R. Hartmann (Eds.), Hearing - Physiological Bases and Psychophysics, Proceedings of the 6th International Symposium on Hearing. Berlin: Springer-Verlag, 321-326.

THRESHOLD DURATION FOR MELODIC PITCH

Roy D. Patterson, Robert W. Peters and Robert Milroy

MRC Applied Psychology Unit
15 Chaucer Road, Cambridge CB2 2EF

Broadly speaking, the pitch of sinusoidal stimuli becomes indistinct to the point of unusable when the duration of the stimulus is less than five complete cycles. The pitch of the voices of many men and some women is at and even below 100 Hz. At first sight this suggests that it takes more than 50 ms to acquire the pitch of each vowel in the speech of these people. Whereas, in point of fact, listeners can follow changes in the pitch of speech almost on a cycle-to-cycle basis. Of course, the explanation for this apparent paradox is that the pitch of the voice is mediated by the residue pitch mechanism which extracts the low pitch of a sound from its higher harmonics - harmonics that complete 25 or more cycles in 50 ms. This suggested to us that one of the important advantages of the residue pitch mechanism is that it enables the listener to extract the low pitch associated with the fundamental of a sound much faster than would be possible if the information had to be extracted from the fundamental alone. This paper presents our attempt to determine the relative advantage of the residue-pitch mechanism for acquiring low melodic pitch.

Ritzma and Cardozo (1963) reported that no residue pitch is perceived when the stimulus duration is below about four complete cycles. The data of Metters and Williams (1973) indicate a lower limit for three-tone complexes of about three cycles. Pollack (1967) found that his listeners could detect a minimal pitch with two to three cycles of a pulse train. Thus, for residue stimuli with pitches in the region of 100 Hz, these studies suggest a threshold for pitch of 25-40 ms. None of the studies, however, employed a melodic pitch task.

1. MELODIC PITCH THRESHOLD

Since our primary interest was in stimuli that give rise to a definite pitch perception, we chose to use a melodic pitch paradigm rather than a frequency discrimination paradigm. Initially, we assembled a set of ten familiar melodies that had no rhythmic information (e.g. Yankee Doodle) and performed a melody recognition experiment in which the listeners simply matched the melody names to the tunes as they were played. The average note value for each melody (or its mean pitch in Hz) was set to 100 Hz. (There is no ambiguity about the pitch of the residue stimuli used in this experiment, and since the paper is concerned with musical pitch, we have adopted the musical convention and describe the pitch in terms of the frequency, in Hz, of the fundamental of the set of components.) The 16 notes of the melody were played over 6 sec. For residue stimuli with five adjacent harmonics, we found that listeners could identify the individual melodies when the duration of the individual notes was less than 10 ms -- at which point the notes sounded like little more than "ticks" and "tocks". The listeners said that although the notes had no real pitch, they could determine the melody by just listening to the overall contour and comparing the vague impression it conveyed with the contours of the melodies of the set. Dowling (1978) has argued that memory for melodies involves two components; memory for a specific musical scale and memory for a contour presented via that scale. The melody recognition task emphasises the contour aspect of this memory, and de-emphasises the musical scale component. Listeners' ability to use melodic contours in the absence of musical scales has also been demonstrated by Moore and Rosen (1979) who used multi-component residue stimuli, 200 ms in duration. They compressed or expanded the musical scale so that the octave bore a ratio of 1:1.3 or 1:4 instead of 1:2. This preserves the contour of the melody but changes the scale to a non-standard size. The compression resulted in slightly worse performance than the expansion but in both cases the listeners were still able to identify the melodies.

To avoid reliance on contours alone, we have used a random melody task in which the listener hears a short melody and then identifies which note has changed when the melody is replayed. To make the task as musical as possible we (a) chose the notes from the diatonic scale, (b) presented the tonic of the scale at the start of each trial, and (c) presented all of the notes in a strict rhythm. A diagram of the sequence of events in each trial appears in Fig. 1. Following a warning light, the listener was presented with what can be thought of as four bars of music, and then a response interval. The basic time interval, or beat, was 2/3 of a second; each interval had either one or no notes and the note started at the beginning of the interval independent of its duration. In the first bar the listener was presented with two beats of the tonic (doh) followed by two silent beats. The second interval contained a four note melody; each note was chosen at random from the notes doh, ray, me, fah, soh of the diatonic scale. The third bar was entirely silent. The fourth bar contained a repeat of the melody, except that one of the notes (chosen at random) was moved up or down (at random) by one step on the diatonic scale. In the example, the second note is moved down from soh to fah. The listeners had a four-button response panel on which to indicate which note had changed, and threshold was taken to be the duration that supported 62.5% correct identification in this four-alternative, forced-choice task. We used four-note melodies as it is a common bar length and minimises the short-term memory load. We have, then, defined melodic pitch threshold as the stimulus duration required to know which member of a four-note melody has been changed by one note on the second presentation of the melody.

2. GENERAL METHOD

The melody was played using either a sinusoidal stimulus or a multi-harmonic stimulus that produced a low residue pitch. The fundamental of the harmonic series determines the note value, that is, the pitch of the note in Hz. The stimuli were computer generated and the output of the digital-to-analogue converter was lowpass filtered at 1.5 kHz. For the note me of the scale, the fundamental was 100 Hz and the sampling rate was 10 kHz. The fundamental was varied to obtain the other notes by adjusting the sampling rate.

Performance with these residue stimuli was compared with that obtained using sinusoids; the me of the scale was set to one of four separate frequencies ranging from 100 to 900 Hz. The stimulus duration was varied from 10 to 80 ms. The total energy of the stimuli was held constant; thus stimulus power was halved when stimulus duration was doubled, and the sinusoidal stimuli had n times the power of each component of a residue stimulus with n components.

The stimuli were presented binaurally over headphones and there was feedback as to which note had changed at the end of every trial. Blocks of 50 trials were run with fixed signal type and duration, and all of the durations associated with one signal type were run before proceeding to another signal type. The basic measure was the percent correct in a block of 50 trials. Each condition was replicated 4-8 days after it was first run. Beyond the first day, which was given over to practice, there were no obvious learning effects.

There were 4 listeners (HM, EL, RM, WW) ranging in age from 24 to 44 years, all of whom had normal hearing in the range 0.25 to 8 kHz. All four participated in the second experiment; HM was not in the first experiment. One of the listeners was author RM. The youngest and oldest listeners had no musical training; the remaining listeners were amateur musicians. The pattern of results was the same for all of the listeners, although the musically trained produced slightly higher scores.

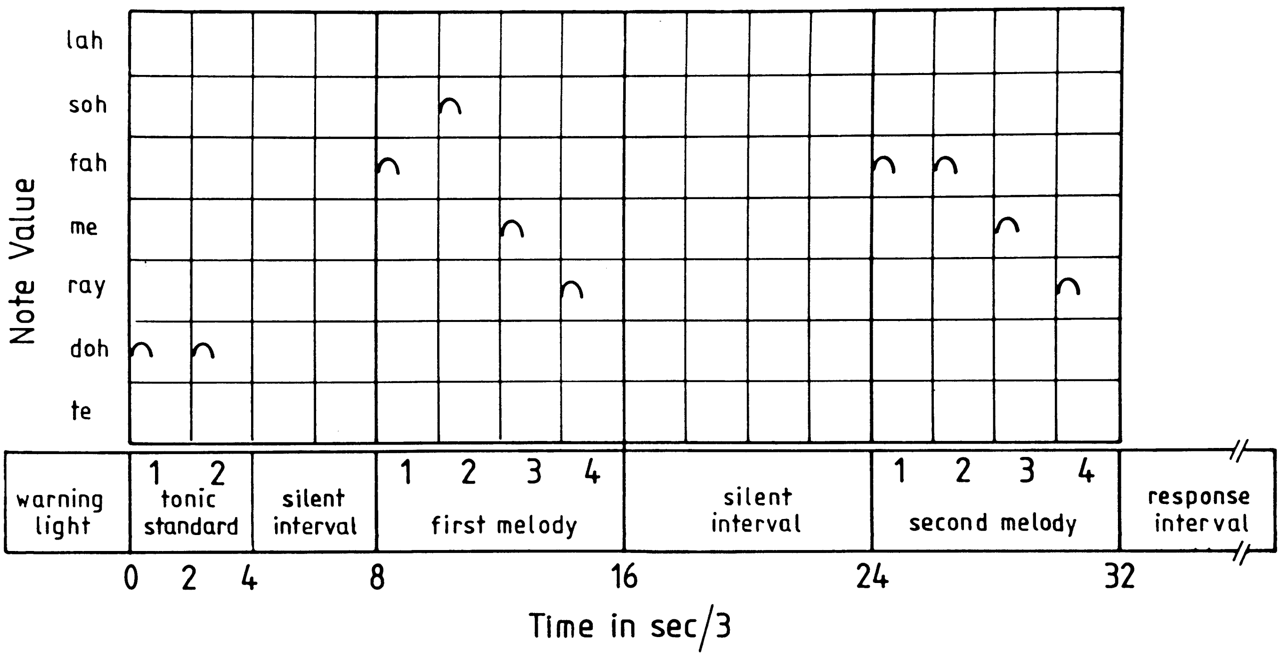

Fig. 1. Trial Sequence for the Melodic Pitch Experiment: The strip at the bottom shows the order and timing of events in a trial. The time unit is 0.66 sec. The upper portion shows the possible note values and a typical trial.

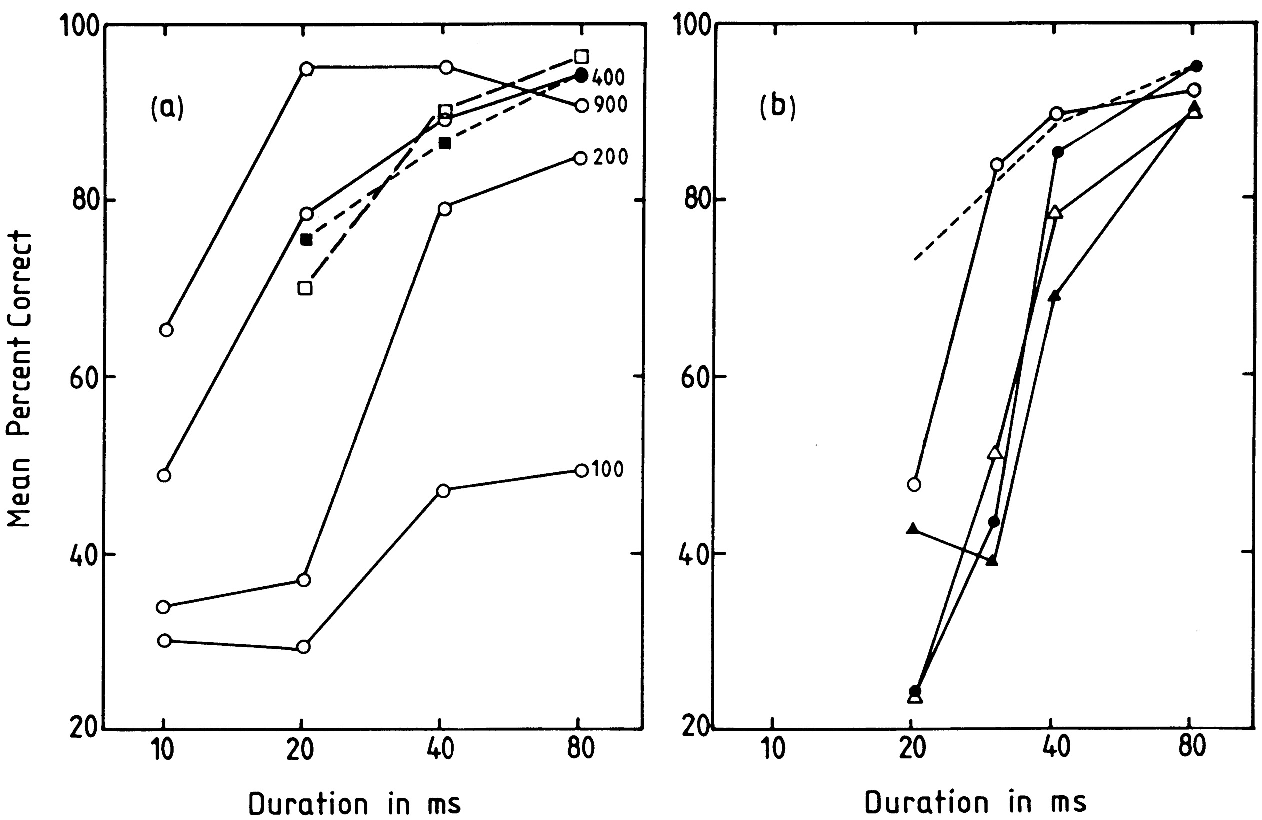

Fig. 2. Performance versus duration in two melodic pitch experiments: a) Residue stimuli with five harmonics (squares) versus sinusoidal stimuli (circles) with frequencies of 100, 200, 400 or 900 Hz, b) Residue stimuli presented in a notched noise; circles and triangles for split-cosine and complete-cosine on/off ramps, filled and open symbols for ramps starting on or between waveform peaks.

3. EXPERIMENT I

a) Specific Methodological Details

In this experiment the residue stimulus was a set of six adjacent harmonics, either harmonics 2-6 or harmonics 7-11. The individual components were presented at about 50 dB SPL. Both the residue and sinusoidal stimuli were gated on and off with 5 ms raised-cosine ramps; the 10 ms stimuli, then, had no steady-state section. The stimuli were presented over a broadband noise, lowpass filtered at 2.0 kHz. Its level was set approximately 25 dB below the point where it would mask the stimuli.

b) Results

The data from Experiment I are shown in Fig. 2(a) for the sinusoidal and residue stimuli by circles and squares respectively. The circles show that melodic pitch is extracted from high frequency sinusoids much faster than from low frequency sinusoids. If threshold is taken to be the duration that supports 62.5% correct, then it takes about 7 cycles of a sinusoidal stimulus to support melodic pitch - a value that is not much above that required for minimal pitch perception. The two residue producing stimuli (open and filled squares) produce performance which is roughly comparable to that which can be obtained with a 400 Hz sinusoid. The 100 Hz pitch associated with the fundamental is acquired in under 20 ms, whereas that of the 100 Hz sinusoid takes in excess of 80 ms. The low pitch of the residue stimuli is not, however, acquired as quickly as the high pitch associated with a high frequency tone, even though both of the residue stimuli contain high frequency tones. If one could listen separately to the highest component of the lower residue stimulus or the lowest component of the higher residue stimulus, performance should have been better than was achieved with either of these stimuli.

This experiment, then, suggests that low pitch in the region of 100 Hz can be extracted from a set of harmonics about 4 times as fast as from the corresponding low-frequency sinusoid. Furthermore, harmonics 7-11 which are only poorly resolved by the auditory system transmit the musical pitch as well as harmonics 2-6.

4. EXPERIMENT II

In the previous experiment all of the components of the residue stimulus moved together when the note value changed. Thus it is not clear whether the listeners were using temporal or spectral information. Specifically, one might use the edges of the spectrum to perform the task. In the second experiment a residue producing stimulus was masked by a broadband noise that had a notch in the region 0.2 to 1.4 kHz. This prevents the listeners from using the edges of the signal spectrum to determine the pitch of the stimulus and forces them to use the central section of the stimulus which is quite flat for brief stimuli.

a) Specific Methodological Details

The second experiment employed the same procedure for measuring melodic pitch. The only differences were in the masking noise and the residue-producing stimuli; there were no sinusoidal stimuli in this second experiment.

The residue stimulus was composed of 20 harmonics of a fundamental in the region of 100 Hz, and it was lowpass filtered at 1.5 kHz. All of the components started in cosine phase, so producing a peaked waveform. The gating function was either a split cosine as before, wherein the onset and offset ramps were each 5 ms independent of the stimulus duration, or the gate was a complete cosine wherein the onset and offset ramps were half the duration of the stimulus. In addition, the phase of the gate relative to the signal was varied: in one condition waveform peaks occurred at the beginning and end of the interval; in the other condition the gate began and finished mid-way between two peaks. These two variables, gate shape and gate phase, produce small differences in the resolution of the harmonics in the spectrum of the stimulus when the signal duration is short (under 40 ms). The split-cosine gate starting between peaks produces the best resolution; at 30 ms the peak-to-valley ratio for components in the centre of the spectrum is about 20 dB. At 30 ms the remaining gate and phase combinations all produce approximately the same peak-to-valley ratios (7 dB). Signal duration was varied from 20 to 80 ms.

The masker was a broadband noise which had a notch in the region 0.2 to 1.4 kHz. Below the lower cutoff the spectrum level of the noise was 48 dB, above the upper edge the spectrum level was 46 dB; the level of the signal components fell from 56 to 54 dB across the width of the notch. The floor of the notch was 30 dB down and so the residue stimulus was clearly audible.

b) Results

The psychometric functions for this experiment are shown in Fig. 2(b); the circles and triangles show the data associated with the split-cosine and complete-cosine conditions respectively. The filled and open symbols show conditions in which the gate began on a pulse or mid-way between two pulses, respectively. In every case, melodic pitch is acquired from these harmonic stimuli in under 40 ms, less than half the time taken to acquire the corresponding low pitch from a sinusoid. And when the gate is a split cosine and begins midway between two pulses, performance is as good as it would be in the absence of the notched-noise masker, down to a duration of 30 ms. Thus, low melodic pitch can be extracted from this type of stimulus even when the listener cannot make use of the edges of the signal spectrum.

The spectra of the residue stimuli are very flat when the duration is 20 ms and in this condition performance is above chance (25%) for several of the stimuli. This may indicate that timing information can be used to extract melodic pitch. However, by the time performance rises convincingly above chance at 30 ms, the stimuli already show peak-to-valley ratios around the harmonics of 6 or more dB and so we cannot rule out the possibility that spectral cues are used to derive the low melodic pitch.

Acknowledgement: The authors are grateful to D.M. Green for setting out the effects of cosine gates on short-term power spectra, and to B.C.J. Moore for comments on an earlier version of the paper.

REFERENCES

Dowling, W.J. (1978). Scale and Contour: two components of a theory of memory for melodies. *Psychological Review* 85, 341-354.

Metters, P.J. and Williams, R.P. (1973). Experiments on tonal residues of short duration. *Journal of Sound and Vibration* 26, 432-436.

Moore, B.C.J. and Rosen, S.M. (1979). Tune recognition with reduced pitch and interval information. *Quarterly Journal of Experimental Psychology* 31, 229-240.

Pollack, I. (1967). Number of pulses required for minimal pitch. *Journal of the Acoustical Society of America* 42, 895.

Ritsman, R.J. and Cardozo, B.L. (1963). The perception of pitch. *Philips Technical Review* 25, 37-43.