Proceedings of The Institute of Acoustics

SCALE AND CONTOUR IN THE DISCRIMINATION OF APPROXIMATIONS TO MELODY: EFFECTS OF COMPONENT FREQUENCY

ANTHONY J. WATKINS

DEPARTMENT OF PSYCHOLOGY, READING UNIVERSITY

Background

Well known melodies such as nursery rhymes or folk tunes may be played in a variety of note ranges, but are readily named by sophisticated and unsophisticated listeners regardless of the particular transposition heard. In other words, the "form" of the melody is governed by the relationships between the note frequencies, and melodic form is perceptually invariant over a wide range of the pitch continuum. This morphophoric or form-bearing property of pitch was demonstrated by Attneave and Olson (1971). They asked listeners to adjust the frequencies of three successive tones so that they played the notes G E C of the NBC chimes. These musically unsophisticated listeners could produce this tune quite accurately in a wide range of frequency regions, even though the same three note frequencies are always used when the chimes are broadcast.

One engaging aspect of these data is that adjustment accuracy decreases quite markedly for note frequencies above 5 kHz. Attneave and Olsen offered two possible explanations. An "empiricist" view: note frequencies above this value are not heard in music. And a "nativist" view: the ability of primary auditory neurones to "follow" the temporal course of the waveform decreases markedly at these high frequencies (Rose, Brugge, Anderson and Hind, 1968). The latter notion does not deny that frequency is also coded by the frequency-to-place mapping seen in the activity patterns of the basilar membrane and in the "tonotopic" organisation of primary auditory neurones (Kiang, 1965). However, it does imply that such "place-coding" is not sufficient for the perception of "melodic" pitch. A similar distinction has recently been made by van Noorden (1982). He reviews evidence for a distinction between two perceptual dimensions for pitch. In order to illustrate this, consider what is heard when the frequency of a note is doubled. Something is heard to change, while something else is heard as the same. That which changes is called "pitch height", and that which stays the same is called "pitch chroma". Chroma therefore provides a perceptual complement of the "octave equivalence" found in music. van Noorden goes on to speculate that pitch chroma might derive from the temporal coding of frequency, whereas pitch height is derived from the tonotopic code. According to this view, melody perception changes qualitatively as the frequency rises above 5 kHz. This is because the absence of neural timing information precludes the perception of chroma.

By contrast, Goldstein and Srulovicz (1977) and Srulovicz and Goldstein (1983) maintain that temporal and tonotopic information are combined at an early stage of auditory processing: the time course of neural activity is band-pass filtered at the characteristic frequency of each auditory neurone. In this way only a very narrow range of neurones will be maximally active in response to a single sinusoid. The predictions of this model are in good agreement with psychoacoustic results on the discrimination of single tones: discrimination is best in the region of 1 to 2 kHz, rather poorer at lower frequencies, and falls off steeply at higher frequencies. This model has implications for the perception of tone sequences: the accuracy of frequency representation might reflect the accuracy with which the intervals of a melody are coded, so one might expect performance on a pure-tone melody perception task to reflect this accuracy, and to vary with the frequency region of the notes accordingly. Such a view would imply that there should be quantitative differences in melody discrimination as the note frequencies are varied, and that performance should approach chance levels as the size of the minimum discriminable frequency difference approaches the size of musical intervals. According to the model of Goldstein and Srulovicz, this should occur for note frequencies above 10 kHz. This prediction is not entirely congruent with the data of Burns and Feth (1983). They played well known melodies to listeners using note frequencies which were all higher than 10 kHz. The overall performance of the group of 14 listeners was rather poor, as one might expect. However, 3 of the listeners were able to extract sufficient information from these sounds for them to achieve identification scores which cannot be attributed to pure guess-work (8, 9, & 10 correct out of 12). The present experiments were designed in order to clarify these issues.

Experiments

The discrimination of sequences of pure tones was measured using different centre frequencies. We ask whether performance varies with centre frequency, and whether the form of any variation reflects the accuracy of frequency representation predicted by Goldstein and Srulovicz. We will be particularly interested in the performance accuracy at higher frequencies.

Method

In this study groups of 20 listeners heard pure-tone sequences with notes centred at either 0.130, 0.523, 2.073 or 7.071 kHz. Each note was selected according to 12-tone equal temperament tuning from the range between 7 semitones above and 6 semitones below centre frequency. Each trial consisted of a pair of 14-note tone sequences separated by a two second gap. The second sequence was either the same as the first or not, but was always transposed by one or two semitones up or down. After each pair of sequences the listeners were given as long as they wished to indicate whether the tone sequences were same-and-transposed or not. Feedback was then provided before the next pair of tone sequences was presented. The first tone sequence in a trial was held constant over a block of 10 trials. This allows the listener to become familiar with the sequence, and there is a consequent improvement in discrimination. Each listener completed 12 of these blocks. Non-parametric assessments of the areas under ROCs were calculated for each listener giving p(A) values. These run from 0.5 for chance discrimination, up to a perfect score of 1.0. An arcsin transform of these values (McNicol, 1972) preceded further data reduction.

The tone sequences used were novel approximations to melody generated by a computational technique described elsewhere (Watkins, 1981). The first step is to select an initial note for the sequence: this is either the note at the centre of the frequency range, or one semitone above or below. The remaining notes are then selected using a random process, with probabilistic biases in favour of certain notes and intervals. In all sequences used there was a bias in favour of intervals smaller than 6 semitones. In some sequences there was a bias in favour of the 7 notes of the diatonic scale (i.e., "do" "re" "me" etc.). In others there was a bias in favour of seven "non-diatonic" notes, i.e., a set of notes which cannot all belong to the same diatonic scale. Previous experiments using complex tones have shown that the former are easier to learn and discriminate (Watkins, 1982). This experiment asks whether this discrimination advantage holds for pure tones, and whether there is any variation with centre frequency.

When the second tone sequence of a pair was different from the first, differences were produced in one of two ways: the "melodic contour" of the melody was either preserved or violated. Contour refers to a representation of the intervals of a melody which indicates whether adjacent notes are higher or lower than one another. Thus, a melody such as "Three Blind Mice" would be reduced to a representation such as: down, down, up, down, down. Dowling and Fujitani (1971) have shown that exact transpositions of different novel melodies which share the same contour tend to sound like transpositions of the same melody. In this experiment we ask whether contour differences aid discrimination and whether any such discrimination advantages vary with centre frequency.

Melodies were played at an even tempo of four notes per second with each note lasting for its full quarter of a second. Onsets and offsets were the appropriate half of a 40 ms Hanning window. The notes were produced by varying the frequency of a digital sinusoid, followed by a 12-bit digital-to-analogue conversion at 32,000 samples per second. The signals were low-pass filtered by a Kemo VBF/8 which was 3 dB down at 13.1 kHz and rolled off at 48 dB per octave. The sounds were heard monaurally (right ear) via Sennheiser HD-424 earphones. The level was attenuated to roughly 30 dB above the threshold of the listener, measured in the quiet of an IAC 1201 booth. (The listeners heard all 14 notes of the range in a repeating sequence, and the level was reduced until one or more notes became inaudible. The session was then run with the attenuator set at 30 dB above this threshold value.) System hum, noise and distortion was inaudible. Probe-microphone measurements near the ear-canal entrance, and third-octave band analysis, confirmed that all off-signal-frequency noise and distortion was more than 40 dB below the level of the signal.

The initial note of the first tone sequence, and the transposition of the second, were randomly selected from the stated ranges on a trial-by-trial basis. Listeners were sampled from a group of undergraduate students. Some were educated in music but generally not beyond secondary school level. They did not describe themselves as musically sophisticated. None of them reported any hearing problems, and their measured thresholds were all in the normal range. Listeners were run individually, allowing different randomisations of tone sequences, tone-sequence orders and trial orders to be used for each of them. A computing system based on a PDP-8 synthesised the melodies, administered the experiment and analysed the data.

SCALE AND CONTOUR IN THE DISCRIMINATION OF APPROXIMATIONS TO MELODY: EFFECTS OF COMPONENT FREQUENCY

Results

Overall performance varied with centre frequency. Performance was best at 0.523 kHz and systematically poorer at 0.130 kHz and at 2.093 kHz. Performance was worst of all at 7.071 kHz, although still better than chance level. A tentative extrapolation of the high frequency roll-off predicts chance performance at a centre frequency of around 10 kHz. Differences in contour between the pairs of tone sequences aided discrimination. This performance difference, between same-contour and different-contour trials, was not affected by the centre frequency. By contrast, discrimination advantages for the diatonic melodies over the non-diatonic melodies were generally weaker than those previously observed using complex tones. Furthermore, the advantage was clearest at 2.093 kHz, less pronounced at 0.523 kHz, and completely absent at 0.130 kHz and at 7.071 kHz.

Discussion

The findings support the idea that the perception of sequences of pure tones reflects the accuracy of the sensory representation of frequency, and that this accuracy is well described by the model of Goldstein and Srulovicz. This is evidenced by the variation of performance with frequency, and the implication of chance performance where the sensory representation fails to preserve differences of the order of a semitone. The findings might also explain why some of Burns and Feth's listeners were able to identify melodies at frequencies above 10 kHz. These listeners would have some melodic contour information available, because the present results show that the salience of contour does not vary with frequency. They may also have been able to derive a certain amount of information about the relative sizes of the larger intervals. The combination of these factors can give rise to quite accurate melody identification, even when exact information about interval sizes is removed by note-frequency compression (Idson and Massaro, 1978).

The findings of advantages for diatonic melodies in just two frequency regions is rather more puzzling. One possibility is that scale perception requires the most accurate representation of intervals, and this is found at the middle two frequency regions, where overall performance is higher. However, the biggest difference between diatonic and non-diatonic melodies does not occur in the frequency region giving the best overall performance. The frequency region centred at 0.523 kHz gives the best overall performance, whereas the biggest advantage for diatonic melodies is seen at 2.093 kHz. I am tentatively proposing an alternative possibility. This is that the advantages observed for diatonic melodies are mediated by the auditory processes responsible for "virtual pitch" as described by Terhardt (1974). This requires that pure tones in the region of 2 kHz give more salient virtual pitch "cues" than tones in the region of 500 Hz, and that tones as low as 100 Hz, and as high as 7 kHz, will not give sufficient cues. It also requires that a single frequency component is sufficient for some form of virtual pitch. Some support for this notion comes from an experiment subsequently performed. Listeners were given the same task as described above. The centre frequency of the note range was 0.130 kHz, and the notes were complex tones consisting of the first 8 harmonics. The results contrast sharply with the pure tone data: a clear advantage for the diatonic melodies emerged. Furthermore, the performance level for the diatonic melodies was comparable with that for the "best" pure-tone frequency region.

References

1. F. ATTNEAVE and R.K. OLSEN 1971 American Journal of Psychology 34, 147-166. Pitch as a medium: a new approach to psychophysical scaling.

2. E.M. BURNS and L.L. FETH 1983 in R. Klinke and R. Hartman (eds.) Hearing: physiological bases and psychophysics. Springer-Verlag, Berlin. Pitch of sinusoids and complex tones above 10 kHz.

3. W.J. DOWLING and D.S. FUJITANI 1971 J. Acoust. Soc. Am. 49, 524-531. Contour, interval and pitch recognition in memory for melodies.

4. J.L. GOLDSTEIN and P. SRULOVICZ 1977 in E.F. Evans and J.P. Wilson (eds.) Psychophysics and physiology of hearing. Academic Press, New York. Auditory-nerve spike intervals as an adequate basis for aural spectrum analysis.

5. W.L. IDSON and D.W. MASSARO 1978 Perception and Psychophysics 24, 551-556. A bidimensional model of pitch in the recognition of melodies.

6. N.Y.-S. KIANG 1965 Acta Otolar. 59, 186-200. Stimulus coding in the auditory nerve and cochlear nucleus.

7. D. McNICOL 1972 Allen and Unwin, London. A primer of signal detection theory.

8. L.P.A.S. van NOORDEN, 1982 in M. Clynes (ed.) Music mind and brain. New York, Plenum. Two channel pitch perception.

9. J.E. ROSE, J.F. BRUGGE, D.J. ANDERSON and J.E. HIND, 1971 in A.V.S. deReuck and J. Knight (eds.) Hearing mechanisms in vertebrates. Churchill, London. Patterns of activity of single auditory nerve fibers of the squirrel monkey.

10. P. SRULOVICZ and J.L. GOLDSTEIN, 1983 J. Acoust. Soc. Am. 73, 1266-1276. A central spectrum model: a synthesis of auditory-nerve timing and place cues in monaural communication of frequency spectrum.

11. E. TERHARDT, 1974 J. Acoust. Soc. Am. 55, 1061-1069. Pitch, consonance, and harmony.

12. A.J. WATKINS, 1981 Proceedings of the Institute of Acoustics, Spring Conference, 347-350. Tonal sequences, melody and music.

13. A.J. WATKINS, 1982 J. Acoust. Soc. Am. 72, S45. Figural and tonal aspects of melody discrimination.