Contemporary Music Review,
1993, Vol. 9, Parts 1 & 2, pp. 69-81
Photocopying permitted by license only

What is the octave of a harmonically rich note?

Roy D. Patterson, Robert Milroy and Michael Allerhand

M.RC Applied Psychology Unit, University of Cambridge, UK

A sound composed of the harmonics of 131 Hz produces the note C3 with a buzzy timbre if the components have equal amplitude and are in cosine phase. Notes with the same tone chroma and a tone height between C3 and C4 can be produced by 1) attenuating the lower harmonics of the sound, 2) attenuating the odd harmonics, or 3) shifting the phase of the odd harmonics. The effects of the manipulations were measured in an octave experiment: a note was chosen at random and used to play a brief melody on the notes around C; the octave varied from Cl to C6 and the listeners judged the octave of each melody on a scale from CO to C7. The results show that waves with the same period can lead to average octave judgments that differ consistently by more than half an octave, and that a substantial component of many timbre differences (e.g., that between a piano and a harpsichord) is actually a tone-height difference. The effects of manipulations 2) and 3) are difficult to explain with traditional hearing theories because the manipulations do not affect the centre of gravity of the spectrum of the sound. The effects can be explained by the "spiral" model of pitch (Patterson, 1987) because the spokes of the multi-channel spiral contain both a spectral dimension (within circuits) and a temporal dimension (across circuits). It appears that octave judgments are closely related to the position of the centre of gravity of activity on the main spoke of the spiral.

KEY WORDS: Pitch perception, timbre perception.

Introduction

There is a serious discrepancy between the psychological representation of pitch, the mel scale, and the musical representation of pitch (the pitch helix). The mel scale is a monotonic, unidimensional mapping of the frequency of a pure tone (a sine wave); the helix is a cyclic, bi-dimensional mapping of the repetition rate of multi-harmonic tones (musical notes). The circular dimension of the pitch helix is tone chroma and the longitudinal dimension is tone height (see Ueda and Oghushi, 1988, for a review). Recently, Patterson (1989, 1990) has emphasized the bi-dimensionality of pitch by demonstrating that one can construct a sequence of notes in which tone height rises an octave while tone chroma remains fixed. Consider a sound composed of 20 harmonics of 100 Hz, and the perceptual change that occurs as the odd harmonics (100, 300, 500, ...) are attenuated, as a group, by an ever increasing amount. The tone height rises smoothly from 100 to 200 Hz without any change in tone chroma. In retrospect, this is not surprising; when the attenuation is greater than about 20 dB, the odd harmonics are effectively removed, leaving a harmonic series that is the octave of the original note (200, 400, 600. ...). What it demonstrates, however, is that the mel scale is completely inadequate as a representation of pitch; it cannot explain how we move continuously from a note to its octave without going through all the intervening tone chromas. The pitch helix has a separate dimension for tone height and, in this case, the new data can be accommodated simply by assuming that it is possible to move continuously along a line from a note to its octave, as well as around the chroma circle. The fact that notes exist between the circuits of the helix means that it is actually a helical cylinder rather than a helical wire, as suggested previously. But the helical cylinder is an obvious extension of the traditional representation.

Patterson (1990) showed that there are several ways to alter the tone height of a multi-harmonic tone without changing its tone chroma. One can attenuate the even harmonics rather than the odd harmonics, and one can phase shift either the even or the odd harmonics. In general, however, he employed only the extremities of the stimulus dimensions; that is, he used only complete attenuation or 90-degree phase shifts. In the first part of the current paper, we report data from an experiment designed to measure intermediate points on the attenuation and phase-shift functions. The effect of phase on tone height is particularly difficult to explain within a spectral theory of pitch perception; the longterm spectrum is not affected by the phase shift. In the latter part of the paper we describe a spectro-temporal model of hearing that explains the tone-height changes and how one might measure tone height.

I. The Experiment

Method

The control stimulus was a set of 28 harmonics of a fundamental, fo. The amplitude of the harmonics was reduced at the rate of 1.5 dB per octave from harmonic 1 to 24; beyond that the amplitude fell 12 dB per component. The fundamental ranged from 31.25 to 1000 Hz in octave steps; for convenience, the notes are designated C1 - C6, although they are about 5 below the corresponding keyboard frequencies. All the components started at their maximum value and so these control stimuli are referred to as 'cosine-phase' or CPH sounds. In one experimental condition, the amplitude of all the odd harmonics was reduced by 9, 18 or 27dB. These stimuli are referred to as 'alternating-amplitude' or AAMP sounds. They have the same spectral centre of gravity as the corresponding control sound. In another condition, the starting phase of the odd harmonics was shifted by either 60 or 90 degrees. These stimuli are referred to as 'alternating-phase' or APH sounds. They and the CPH sound have the identical longterm spectra. In the final condition, the experimental manipulations were combined; the odd harmonics were attenuated by a fixed 9 dB and then these same harmonics were phase shifted either 60 or 90 degrees. They are referred to as AAMP/APH sounds. A detailed study of listeners' abilities to detect the phase manipulation is presented in Patterson (1987b) along with a review of previous studies of APH sounds.

On each trial of the experiment, one of the sounds was chosen at random (without replacement) and used to construct a short melody that converged on three identical half notes. The pitch of the half notes was one of the notes C1 - C6; the duration of the half note was 500 ms. The listener's task was to judge the octave of the half notes. The primary concern in these studies is the musical perception of sounds rather than the audibility of individual harmonics. Presenting the sound as a melody promotes synthetic listening over analytic listening. Details of the procedure and rationale are presented in Patterson(1990).

The musical designation of the C's on the keyboard (CO - C8) was explained to the listeners (C4 is middle C). Although the notes in the experiment ranged from C1 - C6, the listeners were told to use the range CO-C7 to ensure that their responses were not artificially restricted by the response scale. We told them that some of the notes would have intermediate octave values and instructed them to use two digit responses with the second digit indicating the position within the octave. There were five listeners and all of them found the task easy to perform. There were two training runs in which all of the stimuli were presented once, and then 10 replications of the complete experiment. Although the average response did vary across listeners, they all produced the same pattern of results, and so, for brevity, the discussion is limited to the average data. Thus, there are 50 judgments per point for the average data presented in the next section (5 listeners by 10 replications).

Results

The data are presented as 'confusion-matrix' plots in Figure 1; the physical octave (that is, the true period of the acoustic waveform measured in octaves of 31.25 Hz) is the abscissa, the average octave response is the ordinate. If the listeners invariably gave the physical octave as their response, the data would fall along the central dashed diagonal running from (1,1) to (6,6) in each subfigure. The upper and lower diagonals show where responses an octave above and below the physical octave would fall, respectively.

The AAMP results are presented by the bold lines in Figure 1a. The bottom data line (solid line, no symbols) presents the data for the control stimulus (the CPH sound). These data show that listeners can readily identify the octave of the basic sound and label it as requested with octave numbers. The average response bias is only 0.16 octaves and the largest average deviation is 0.49 octaves. The top line (filled diamonds) presents the average response when the attenuation is 27 dB. When compared with the upper dashed diagonal, it shows that the average response is about an octave above the response to the corresponding CPH sound, indicating that 27 dB is effectively complete attenuation. The average bias in this case is I.17, or 0.17 above the upper octave. The remaining lines with open squares and open diamonds present the data for the conditions where the attenuation is 9 and 18 dB, respectively. The 18-dB data fall above the 9-dB data and both fall in the range bounded by the CPH and 27-dB data. The average biases for the 9- and 18-dB data are 0.45 and 0.93, respectively, which when measured relative to the average biases for the CPH and 27-dB conditions, shows them to be 29 and 77 of the way from the initial to the final octave, on average. These percentages are reasonably representative for the upper three octaves (4, 5, & 6), but as the physical octave decreases from 3 to 1, a given attenuation value has progressively more effect on the octave response. In summary, the AAMP data show that a) octave judgments are highly regular, b) the attenuation required to raise the response a full octave is surprisingly large, 27-dB, and c) it is possible to measure tone-height as a function of the attenuation of the odd harmonics.

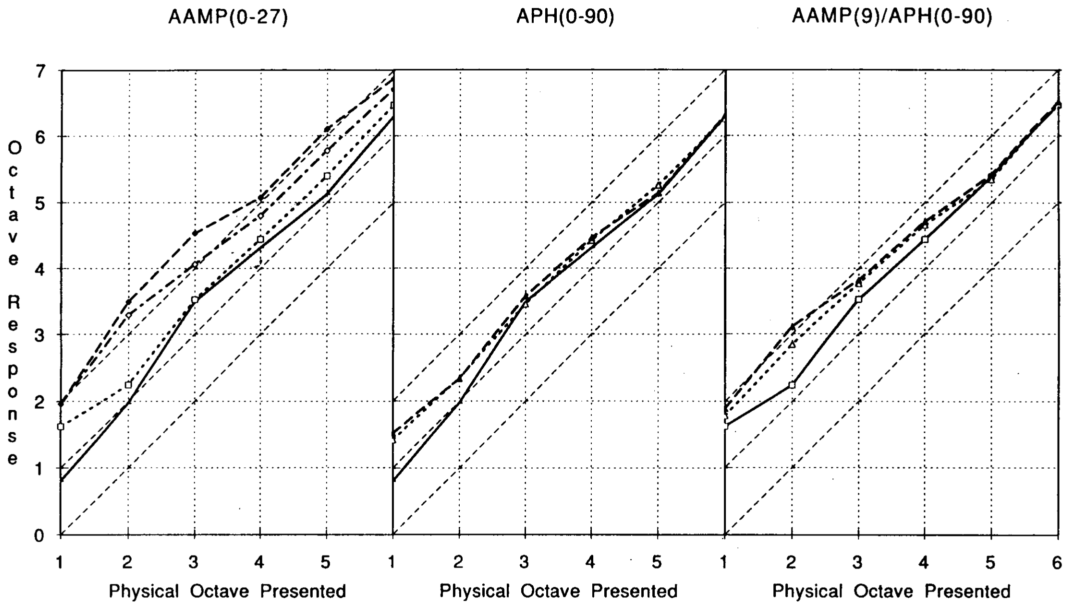


Figure 1. Average octave responses for five listners as a function of physical octave presented, for sounds in which the odd harmonics are (a) attenuated 0-27 dB, (b) phase shifted 0-90 degrees, or (c) attenuated 9 dB and phase shifted 0-90 degrees. The central dashed diagonal in each subfigure shows the position of responses at the physical octave.

The APH results are presented in Figure 1b. The solid line (no symbols) presents the same CPH data as in Figure 1a. The phase shift elevates the octave response for the lowest two octaves but not at the higher octaves. The 60- and 90-degree APH data give the same results in this case. Patterson (1987b) showed that the detectability of the phase shift was limited to sounds with repetition rates less than 400 Hz which means that phase shifts were not expected for the upper two octaves (500 and 1000 Hz). A difference might have been expected, however, for the middle two octaves (125 and 250 Hz).

The AAMP/APH results are presented in Figure 1c. The solid line (open squares) shows 9-dB AAMP data replotted from Figure 1a as the appropriate comparison. The upper two curves show the data for sounds where the odd harmonics are attenuated 9 dB and the same components are phase shifted either 60 degrees (open triangles) or 90 degrees (filled triangles). In this case we see the expected effects. Both phase shifts raise tone height for octaves below 400 Hz and not above, and the 90 degree phase shift has more effect than the 60 degree shift. It seems likely that the elevation of the CPH responses to sounds at octaves three and four limits our ability to measure a phase effect when there is no attenuation of the odd harmonics.

In summary, attenuating the odd harmonics and phase shifting the odd harmonics both raise tone height, and the manipulations combine to produce additional rises in tone height. These results are difficult to explain with spectral models of pitch perception.

II. Modelling Octave Perception

The waveform for the CPH sound, with fundamental 125 Hz, is presented in Figure 2, along with examples of the 125-Hz waveforms produced by attenuating or phase shifting the odd harmonics. The CPH wave (Figure 2a) is a modified pulse train; the small oscillations leading away from the pulses simply indicate the absence of full-size harmonics in the region above the 24th. A comparison of the CPH wave with the AAMP(9) and AAMP(18) waves in Figures 2b and 2c shows that attenuating the odd harmonics produces a secondary pulse half way through the period of the CPH wave. As the attenuation increases, the secondary pulses grow and the primary pulses shrink, and when the attenuation is carried to completion, the pulses are equal in size and the fundamental becomes 250 Hz.

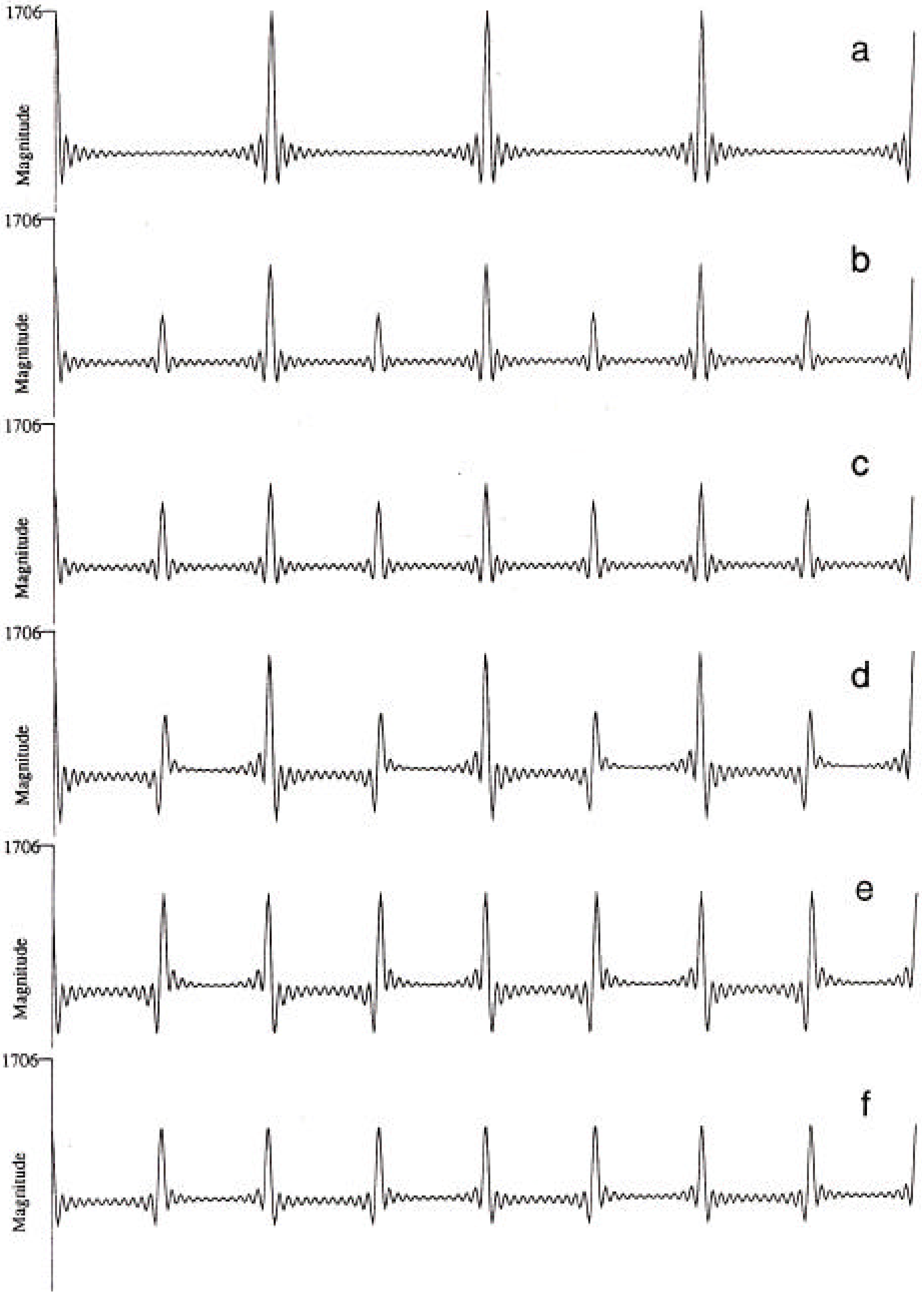


Figure 2. Four cycle segments of selected stimuli with 8-ms periods (fo=125 Hz). The sounds are (a) CPH, (b) AAMP(9), (c) AAMP(18), (d) APH(60), (e) APH(90), and (f) AAMP(9)/APH(90). The attenuation and phase-shift manipulations introduce secondary peaks in the central portion of the period of the CPH wave.

The APH(60) wave in Figure 2d shows that phase shifting the odd harmonics also introduces a secondary pulse in the waveform, but in this case, the peak of the pulse is just after the mid-point of the period. Both the primary and secondary pulses have negative excursions and they are on opposite sides of the pulses. As the phase shift increases to 90 degrees (Figure 2e) the primary and secondary pulses converge on the same height but the asymmetries in pulse position and pulse shape remain. As the odd harmonics of the APH waves are attenuated, the asymmetries in pulse position and pulse shape diminish as shown by the AAMP(9)/APH(90) wave in Figure 2f.

Auditory Sensation Processing and Auditory Images

The APH sounds were originally introduced in a study of monaural phase perception (Patterson, 1987a,b) designed to support a spectro-temporal model of hearing – the Pulse Ribbon Model. The work has now been extended, with the addition of a temporal integration mechanism, to the point where it can simulate the Auditory Sensation Processing necessary to convert acoustic waves into a reasonable representation of the auditory images we hear when presented with musical notes (Patterson and Holdsworth, 1990). The first stage in the model is an auditory filterbank which performs a spectral analysis much like that of the basilar membrane (Patterson and Holdsworth, 1991). The second stage is a bank of two-dimensional, adaptive-threshold generators that perform compression, rectification, adaptation and suppression on the outputs of the individual channels of the filterbank (Holdsworth, 1990). In so doing the generators simulate the function of the hair cells in the cochlea and convert the output of the filterbank into a simulation of the neural activity pattern flowing from the cochlea.

The final stage of the model converts the fast flowing neural activity patterns of periodic sounds into stabilized auditory images through a process referred to as triggered, quantized, temporal integration (Patterson and Holdsworth, 1990, 1991). In essence, the larger peaks in each channel of the neural activity pattern are used as strobe pulses for the integration process. When they occur, a portion of the neural activity pattern in that channel is transferred as a unit to the corresponding channel of the auditory image and added point for point to what is already there. When a sound is periodic, the strobe pulses are synchronized to the period of the wave and the sections of the neural pattern transferred to the image are all very similar. They are also aligned and, as a result, they accumulate to form an auditory image that is stationary even though the neural activity pattern is streaming past at a rapid rate.

The auditory model has been implemented as a computer program that converts waves into auditory images; the images, of the waves in Figure 2 are presented in Figure 3. The filterbank has 49 channels with the lowest and highest filters centered at 100 and 2500 Hz, respectively; the filter spacing is quasi-logarithmic. Both the neural activity pattern and the auditory image have the same number of channels as the filterbank. The channels are indicated by the horizontal lines in each subsection of Figure 3. The abscissa is 'time since the last strobe pulse'.

The auditory image of the CPH sound is presented in Figure 3a. The period of the wave is 8ms and, in the auditory image, the channels with activity show a peak at this time. In the upper channels there is no activity in the centre portion of the period. Lower channels containing resolved harmonics have peaks with reduced amplitude in the centre portion of the period. The harmonic number of a resolved harmonic can be identified by the number of peaks in one period of the wave. For example, the second harmonic which appears in channels 7-10 has two peaks per period. The effect of attenuating the odd harmonics is shown in the auditory images of the AAMP(9) and AAMP(18) sounds (Figures 3b and 3c). The initial attenuation (9 dB) causes a thinning of the ridge of activity at 8 ms, and an increase in activity in the central portion of the period. As the attenuation continues (Figure 3c), the patterns of activity along the ridge and in the central section become more and more similar, and for attenuations in excess of 25 dB, the pattern is essentially the same; that is the CPH 8-ms sound has become a CPH 4-ms sound. Thus, in the auditory image model, the perception of a continuous progression from a note to its octave (Figure 1a) occurs through continuous change in the relative strength of peaks in the image that the note and its octave have in common.

The effect of phase shifting the odd harmonics is shown in the auditory images of the APH(60) and APH(90) sounds (Figures 3d and 3e). The phase shift does not cause a thinning of activity on the 8-ms ridge because it does not reduce the level of the odd harmonics. Rather, it reduces the suppression of peaks in the centre of the period for all harmonics. As a result, the even resolved harmonics and all of the unresolved harmonics are a little stronger in the central portion of the APH images than they are in the CPH image. When the odd harmonics are reduced by 9 dB, the phase shifting has a greater effect. The activity in the central portion of the AAMP(9)/APH(90) sound (Figure 3f) is greater than it is in the corresponding portion of either the AAMP(9) sound (Figure 3b) or the APH(90) sound (Figure 3e). Indeed, phase shifting the attenuated components increases activity in the central portion of the image to levels like those in the AAMP(18) image. A comparison of the data in Figures 1c and 1a shows that the combination of manipulations also increases the octave responses for AAMP(9)/APH(90) to near the level of those for AAMP(18).

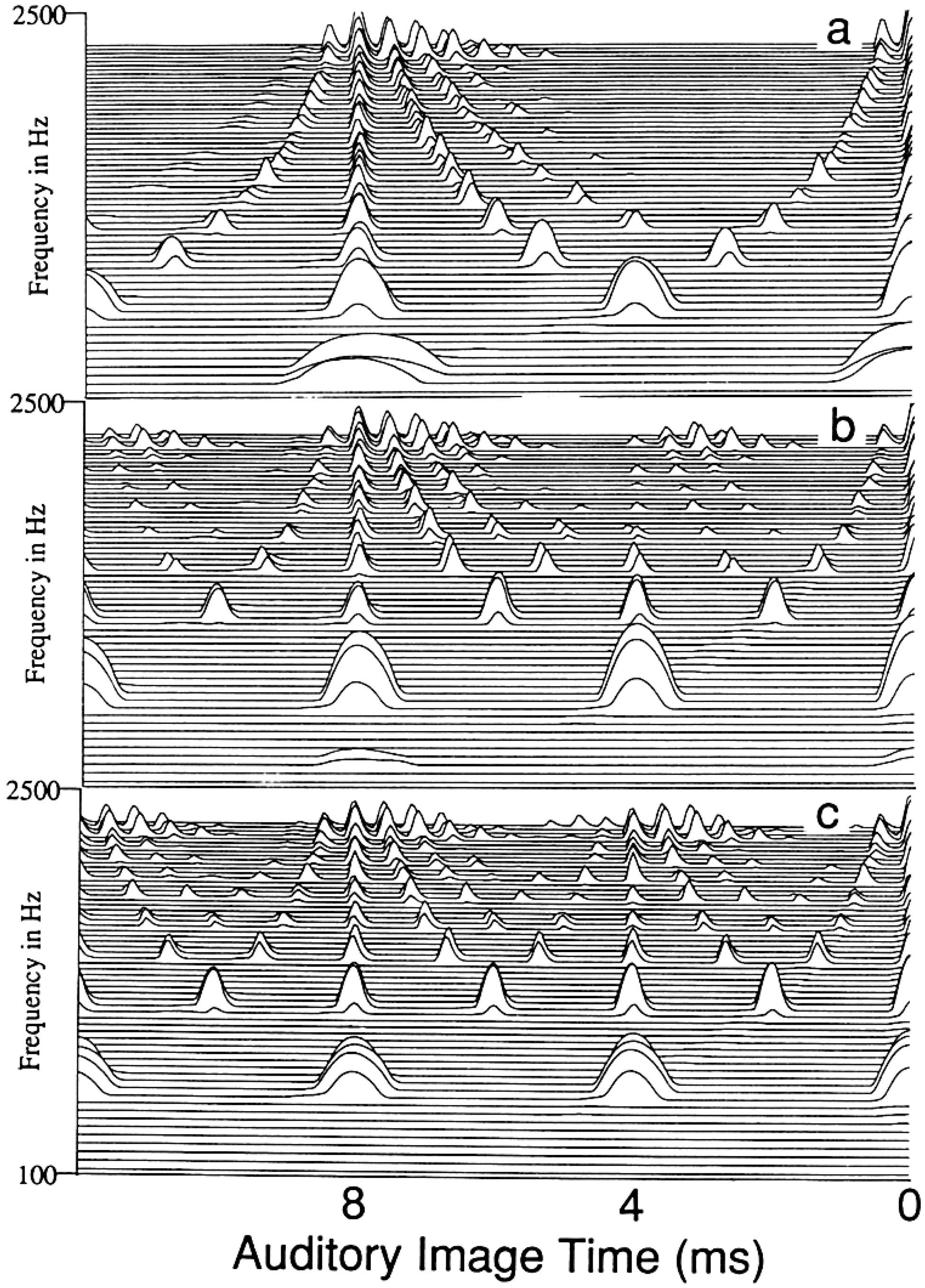
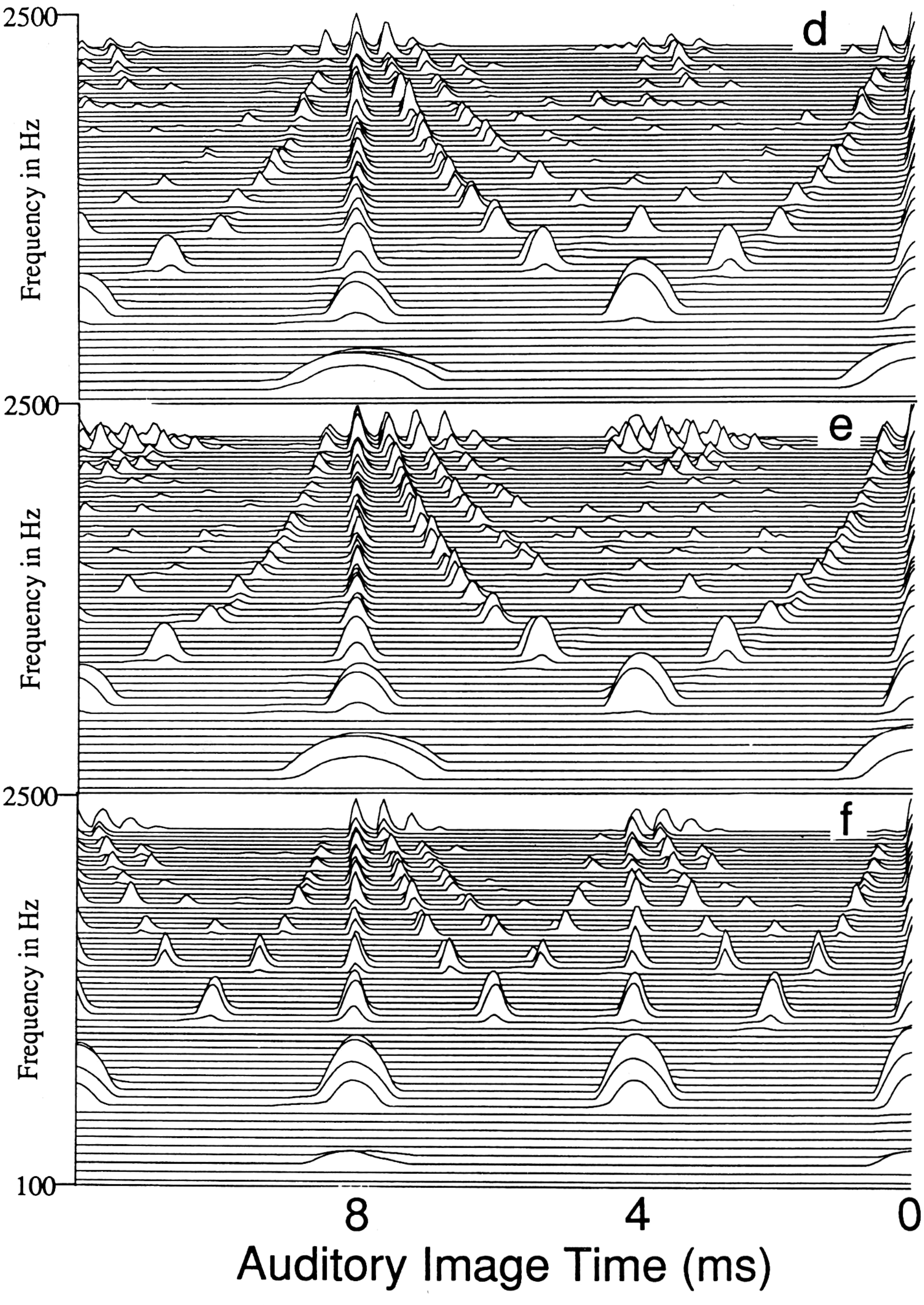


Figure 3. Auditory images of the sounds presented in Figure 2: (a) CPH, (b) AAMP(9), (c) AAMP(18), (d) APH(60), (e) APH(90), and (f) AAMP(9)/APH(90). The manipulations introduce activity into the central portion of the auditory image where the CPH image has no activity.



Spiral Excitation Patterns

In the auditory image model, the tone chroma of the sound is determined with the aid of a "spiral processor" (Patterson. 1986; Patterson and Holdsworth, 1990). In essence, the processor looks at the activity on sets of vertical slices of the auditory image that are separated by doublings in time to see if one such set has more activity than the others (Patterson and Nimmo-Smith, 1986). For the sounds in the current experiment, the sequence of vertical slices at 1, 2, 4, 8, 16, 32 and 64 ms is correctly identified as the tone chroma in every case. The tone height of the notes is not specified by the spiral processor in its original form. But the fact that the octave-response data are closely related to the activity on the 4- and 8-ms slices of the auditory images suggests that the spiral algorithm might be extended to include a measure of tone height, and so become a complete model of musical pitch. Accordingly, the model was used to extract slices through the auditory images in Figure 3 at 1 ms and its successive doublings. The slices at the 8-ms point and and the 4-ms point are presented in the left and right columns of Figure 4, respectively. The abscissa is channel number; the spectral resolution was increased to 99 channels for these plots. For convenince, all of these activity patterns will be referred to as 'spiral excitation patterns', or simply 'excitation patterns' when there is no ambiguity.

In the left-hand column of Figure 4a is the spiral excitation pattern at the physical period of the CPH sound (8 ms). It is much like the traditional excitation pattern that any spectral model of pitch would produce for the CPH sound, with five largely resolved harmonics in the lower channels (0-50) and a band of largely unresolved harmonics in the higher channels. The spiral excitation patterns on slices farther along the sequence (e.g. at 16, 32 and 64ms) are all very similar to that at 8 ms. The same is not true, however, at shorter periods in the sequence. The right-hand portion of Figure 4a shows the spiral excitation pattern at 4ms for the CPH sound. Only the channels associated with the second and fourth

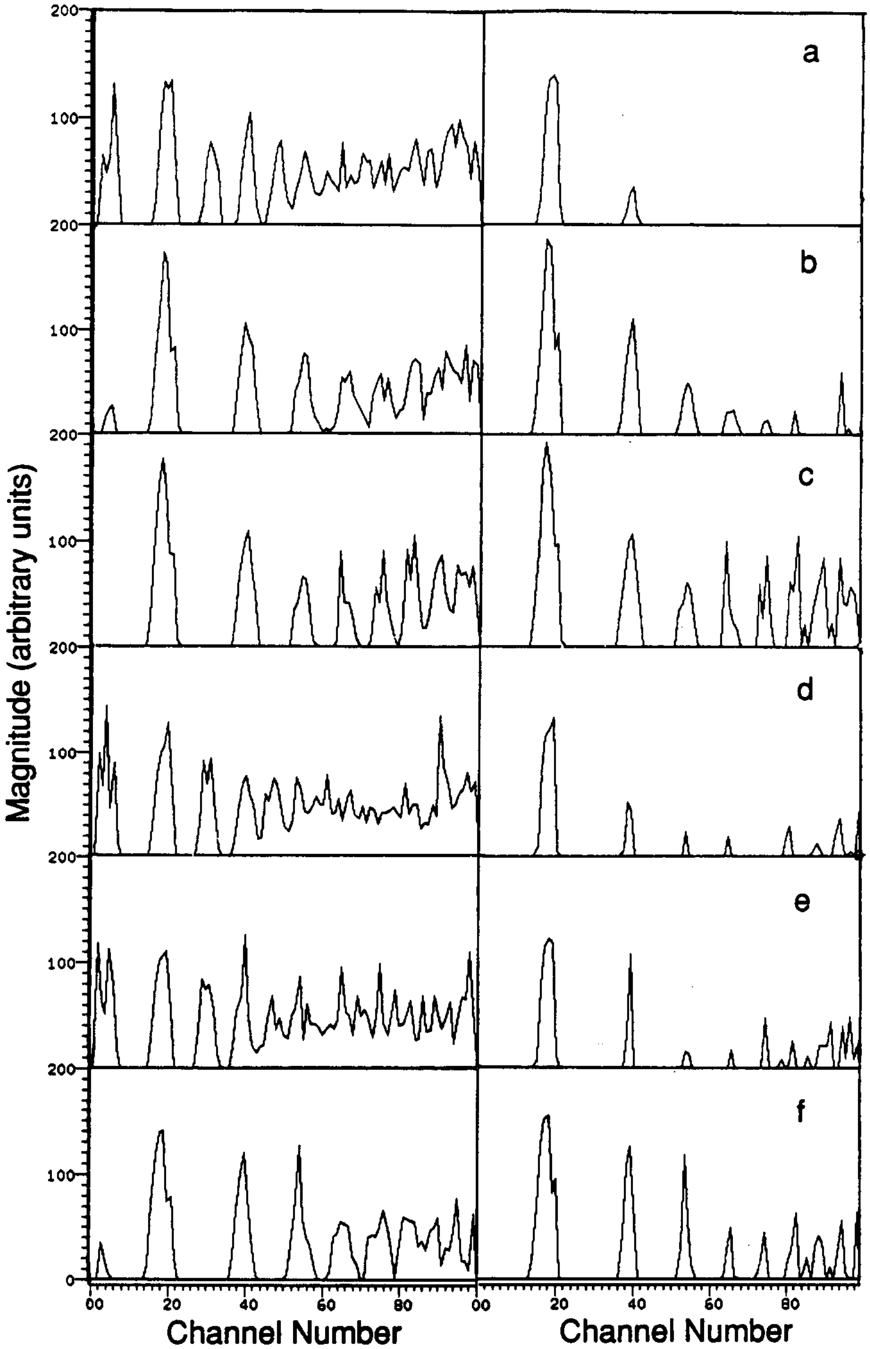


Figure 4. Spiral excitation patterns extracted at 8 ms (left-hand column) and 4 ms (right-hand column) from the auditory images in Figure 3: (a) CPH, (b) AAMP(9), (c) AAMP(18), (d) APH(60), (e) APH(90), and (f) AAMP(9)/APH(90). Tone height increases when the similarity between the 8- and 4-ms patterns increases.

harmonics appear in this excitation pattern, since only these components have peaks half way through the period of the sound. This abrupt change in the shape of the spiral excitation pattern signals the position of the octave. Excitation patterns at 2 and 1 ms show even less activity than that at 4 ms.

One procedure for locating the position of rapid change would be as follows: Chose a low fundamental of the given tone chroma, say 31.25 Hz, and for each harmonic of that fundamental, calculate the difference between the level of the harmonic in one excitation pattern and the next. The point in the sequence where the level difference is greatest is the tone-height estimate associated with that harmonic. The overall tone-height value for the sound is a weighted average of the individual estimates. Consider the case of the CPH sound. The spiral excitation patterns beyond 8 ms in the sequence are all very similar to that at 8ms, and so the level differences would be small for each and every harmonic. For most harmonics, there is one large level difference which occurs between the 8- and 4ms excitation patterns, as shown in Figure 4a. The patterns at 2 and 1 ms have no activity for most of the harmonics. The only exception is harmonic 2. Its strong presence in the 4-ms excitation pattern means that its tone-height estimate will occur between the 4- and 2-ms patterns - an octave higher than the rest. Thus, the overall tone-height value will be a little over octave 3.

Now consider how the tone-height values would change as the odd harmonics are attenuated, or phase shifted, or both. The attenuation of the odd harmonics rapidly removes them from the 8-ms excitation pattern as shown by the patterns for AAMP(9) and AAMP(18) on the left-hand sides of Figures 4b and 4c. At the same time, however, it removes the odd harmonics from later patterns in the sequence and so the odd harmonics rapidly drop out of the calculation altogether. The attenuation increases the activity of the even harmonics in the 4-ms excitation patterns as shown on the right-hand sides of the same subfigures. This gradually shifts the maximum-level-difference of more and more harmonics from the 8- ms/4-ms pair of patterns to the 4-ms/2-ms pair of patterns, and so gradually raises the overall tone-height value from octave 3 to octave 4.

Shifting the phase of the odd harmonics has essentially no affect on the 8-ms excitation pattern of the sound and those further along in the sequence (compare the pattern in the left-hand column of Figure 4a with those in Figures 4d and 4e). It does, however, affect the 4-ms pattern because it reduces suppression of peaks in the even harmonics in the centre portion of the period (compare the right-hand column of Figure 4a with those in Figures 4d and 4e). Thus, for the even harmonics, it decreases the level differences associated with the 8-ms/4-ms pair of excitation patterns and increases the level differences for the 4-ms/2-ms pair of patterns. The effect on the overall tone-height value is rather small in this case, but it is also a small effect in the data (Figure 1b). The combination of 9-dB of attenuation and a 90-degree phase shift raises the even harmonics in the 4-ms excitation pattern to near the level of those in the 4-ms pattern of the AAMP(18) sound. Thus, it moves the position of the large level differences to the 4-ms/2-ms pair of patterns for many of the even harmonics. At the same time, the attenuation reduces the influence of the odd harmonics, and so there is a stronger rise in the overall tone-height value - a rise that is mirrored in the octave-response data.

In summary, then, it is possible to produce a tone-height measure that will track the rise in tone height produced by attenuating and/or phase shifting the odd harmonics of the CPH sound. In traditional spectral models, the spiral excitation patterns do not exist as separable entities and so this type of measure is precluded from the outset.

Acknowledgements

The authors would like to thank John Holdsworth for programming the cochlea simulation and the platform for the auditory image model. The work was supported by the MRC and two grants: MOD PE XTR/2239 and Esprit BRA 3207.

References

Holdsworth, J. (1990). Two-dimensional adaptive thresholding. Annex 4 of APU AAM-HAP Report 1

Patterson, R. D. (1986) Spiral detection of periodicity and the spiral form of musical scales. Psychology of Music **14**. 44-61.

Patterson, R. D. (1987a) A pulse ribbon model of peripheral auditory processing. In Auditory Processing of Complex Sounds, edited by W. A. Yost and C. S. Watson pp 167-179, New Jersey, Earlbaum.

Patterson, R. D. (1987b) A pulse ribbon model of monaural phase perception. Journal of the Acoustical Society of America **82**. 1560-1586.

Patterson, R. D. (1989) The tone height of multi-harmonic tones. In Proceedings of the First International Conference on Music Perception and Cognition. Kyoto, Japan

Patterson, R. D. (1990) The tone height of multiharmonic sounds. Music Perception **8**. 201-21 1 .

Patterson, R. D. and Holdsworth, J. (1990) An introduction to auditory sensation processing. Annex 1 of APU AAM-HAP Report 1

Patterson, R. D. and Holdsworth, J. (1991) A functional model of neural activity patterns and auditory images. In Advances in Speech. Hearing and Language Processing Vol 3. edited by W.A. Ainsworth, JAI Press, London.

Patterson, R.D. and Nimmo-Smith, I. (1986) Thinning Periodicity Detectors for Modulated Pulse Streams. In Auditory Frequency Selectivitrv. edited by B.C.J. Moore and R.D. Patterson, pp 299-307, New. York, Plenum.

Ueda, K. and Oghushi, K. (1987) Perceptual components of pitch: Spatial representation using a multidimensional scaling technique. Journal of the Acoustical Society of America **82**, 1193-1200.