

Circularity in Judgments of Relative Pitch

ROGER N. SHEPARD

Bell Telephone Laboratories, Inc., Murray Hill, New Jersey
(Received 23 July 1964)

A special set of computer-generated complex tones is shown to lead to a complete breakdown of transitivity in judgments of relative pitch. Indeed, the tones can be represented as equally spaced points around a circle in such a way that the clockwise neighbor of each tone is judged higher in pitch while the counterclockwise neighbor is judged lower in pitch. Diametrically opposed tones—though clearly different in pitch—are quite ambiguous as to the direction of the difference. The results demonstrate the operation of a “proximity principle” for the continuum of frequency and suggest that perceived pitch cannot be adequately represented by a purely rectilinear scale.

INTRODUCTION

THE construction of one-dimensional psychological scales of pitch corresponding to the one-dimensional physical scale of frequency for tones has been accomplished by Stevens, Volkman, and Newman¹ and others.² Still others, however, have suggested that the perception of pitch may be too complex to be adequately represented by a single, linear scale. They have pointed out, for example, that the resemblance in pitch between two notes may actually be greater if they are exactly an octave apart than if they are somewhat less than an octave apart.³ Indeed, there is some evidence that even the rat hears an increase in similarity at the interval of an octave.⁴ Moreover, I have observed that, at least for harmonically rich tones, a sequence that increases in intervals of a major seventh (i.e., intervals that are just one half-tone short of an octave) has a peculiarly

ambiguous character. Although it tends, naturally, to be heard as increasing in pitch, it can in some sense be heard, also, as a chromatically *decreasing* sequence. Clearly, such effects are difficult to reconcile with a rectilinear, unidirectional scale of pitch.

As early as 1846, Drobisch attempted to accommodate some of these effects by distorting the continuum of pitch into a helical curve in such a way that tones just an octave apart would be represented by corresponding points on successive turns of the helix.⁵ Such a representation has the advantage of bringing tones an octave apart into closer spatial proximity. At the same time, it permits an explanation of other seemingly anomalous phenomena by providing for the analysis of pitch into two distinct dimensions: namely, “height” (or overall pitch level), represented by the vertical axis of the helix and “tonality” (“tonal quality” or “tone chroma”), represented by the circular scale at the base of the helix. The ambiguous character of the sequence of notes increasing in major sevenths, for example, can then be explained by noting that, while these notes are indeed increasing in height, they are simultaneously moving in a contrary direction with respect to tonality (as reflected in the inverse alphabetical order of the letters by which these notes are designated).

The fact that one of the two components of pitch is circular in the helical model raises the remarkable possibility that, by appropriately exaggerating that component (viz., tonality), one might be able to bring about

¹ S. S. Stevens, J. Volkman, and E. B. Newman, “A Scale for the Measurement of the Psychological Magnitude of Pitch,” *J. Acoust. Soc. Am.* 8, 185–190 (1937); see also, S. S. Stevens and J. Volkman, “The Relation of Pitch to Frequency; A Revised Scale,” *Am. J. Psychol.* 53, 329–353 (1940).

² E.g., J. Beck and W. A. Shaw, “The Scaling of Pitch by the Method of Magnitude Estimation,” *Am. J. Psychol.* 74, 242–251 (1961).

³ See E. G. Boring, *Sensation and Perception in the History of Experimental Psychology* (Appleton-Century, New York, 1942), particularly pp. 376, 380; or J. C. R. Licklider, “Basic Correlates of the Auditory Stimulus,” in *Handbook of Experimental Psychology*, S. S. Stevens, Ed. (John Wiley & Sons, Inc., New York, 1951), pp. 985–1039, particularly pp. 1003–1004.

⁴ H. R. Blackwell and H. Schlosberg, “Octave Generalization, Pitch Discrimination, and Loudness Thresholds in the White Rat,” *J. Exptl. Psychol.* 33, 407–419 (1943).

⁵ See C. A. Ruckmick, “A New Classification of Tonal Qualities,” *Psychol. Rev.* 36, 172–180 (1929).

a breakdown of transitivity in judgments of relative pitch. In the extreme case, if the dimension of height could somehow be suppressed altogether, all tones an octave apart would be mapped into the same tone; that is, the tonal helix would be collapsed into a tonal circle. Judgments of relative pitch should then become completely circular in the sense that there would be no highest or lowest tone in the set but only an isotropic ring in which every tone has both a clockwise neighbor that is judged higher in pitch and a counterclockwise neighbor that is judged lower in pitch.

In the experiments described here, this curious situation was realized by means of a specially contrived set of complex tones. The generation of these tones was made possible by a computer program developed by M. V. Mathews for the synthesis of musical sounds in an extremely flexible and precisely controlled manner.

I. GENERATION OF THE TONES

Each tone consisted of many sinusoidal components locked at successive intervals of an octave and sounded simultaneously. Thus the frequency of each component above the lowest was exactly twice the frequency of the one just below. The amplitudes were large for the components of intermediate frequency only, however, and tapered off gradually to subthreshold levels for the components at the highest and lowest extremes of frequency. The sound-pressure level (in decibels) contributed by each component of one of these multi-component tones is represented, graphically, by the height of the heavy vertical line corresponding to that component in Fig. 1. At the beginning of each tone, all components were started in the same phase relation (upward from a zero crossing).

The most important aspect of the scheme adopted for the generation of these tones was that the spectral "envelope" of the sound levels, as represented by the light curve in the Figure, was identical for all tones in the set. Consider, for example, a second tone in which all components are shifted up (in log frequency from the corresponding components of the first tone) the same fraction of the way toward the next higher octave. The spectral composition of this second tone would differ from that for the original tone in the manner in which the dashed vertical lines in Fig. 1 differ from the original solid vertical lines. The essential point to notice here is that the upward shift in frequency has been offset, in some measure at least, by increasing the contributions of the lower components while decreasing the contributions of the higher components. Indeed, if the second tone is shifted up one whole octave, it becomes identical to the original tone. For, at this point, the highest component (which has already faded below threshold) is dropped out and a new component (which will also be below threshold) is introduced one octave below the previously lowest component.

The tones generated according to this cyclic scheme

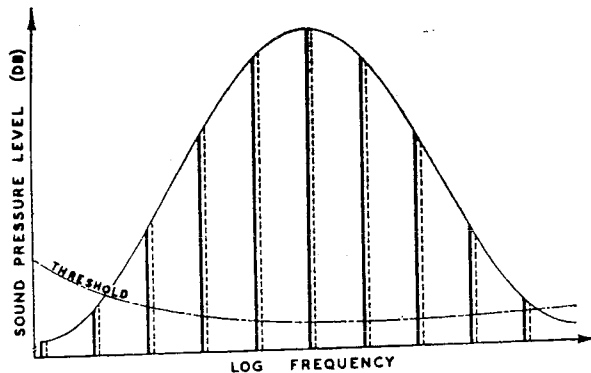


FIG. 1. Sound-pressure levels (in dB) of 10 simultaneously sounded sinusoidal components spaced at octave intervals. (The dotted lines correspond to an upward shift in the frequencies of all components.)

were confined to a discrete set of positions equally spaced (in log frequency) between the original tone and the return (one octave "above") to that same tone. In a set consisting of t_{max} different tones, the frequency of the c th component of the t th tone is

$$f(t,c) = f_{min} \cdot 2^{[(c-1) \cdot t_{max} + t - 1] / t_{max}}, \quad (1)$$

where f_{min} is the frequency of the lowest component of the first tone. The sound-pressure level of any one component (in decibels) is

$$L(t,c) = L_{min} + [L_{max} - L_{min}] \cdot [1 - \cos\theta(t,c)] / 2, \quad (2)$$

where L_{min} and L_{max} fix the range of sound levels for the individual components and where the function θ is defined by

$$\theta(t,c) = 2\pi[(c-1) \cdot t_{max} + t - 1] / [t_{max} \cdot c_{max}]. \quad (3)$$

Here, c_{max} denotes the total number of components in each tone.

The trigonometric "envelope" defined by Eq. 2 (and illustrated in Fig. 1) was chosen as the most convenient function of the general shape desired that is "smooth" and, at the same time, satisfies the relation

$$\sum_{c=1}^{c_{max}} L(t,c) = \text{const} \quad (\text{over all } t). \quad (4)$$

This relation ensured that the different tones in the set did not vary appreciably in over-all loudness. For, if the components of a complex tone are widely separated (as here), the over-all loudness of the tone is approximated by the sum of the loudnesses of its separate components.⁶⁻⁷

⁶ H. Fletcher and W. A. Munson, "Loudness, Its Definition, Measurement, and Calculation," *J. Acoust. Soc. Am.* 5, 82-108 (1933); D. H. Howes, "The Loudness of Multicomponent Tones," *Am. J. Psychol.* 63, 1-30 (1950).

⁷ The invariance of loudness was of course only approximate since the threshold curve is not flat (Fig. 1) and since subjective loudness is not strictly linear with sound-pressure level. In fact, however, there was no noticeable variation in loudness among the different tones.

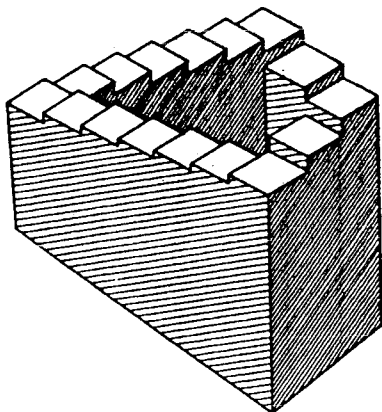


FIG. 2. "Circular" staircase illusion.

The waveforms of these specially tailored multi-component tones were readily constructed on an IBM-7094 computer by means of Mathews' program for synthesizing quite arbitrarily specified musical sounds.⁸ The only addition that had to be made to Mathews' program was a short FORTRAN subroutine for the computation of the parameters of frequency and sound level as prescribed, above, in Eqs. (1)–(3). For each successive tone, the output consisted of a long sequence of digitally coded numbers on magnetic tape. These numbers specified displacement as a function of time at a rate of 10 000 samples/sec of subsequent playing time. (This sampling rate corresponds to a bandwidth of 5000 cps.) An analog tape suitable for playback (at $7\frac{1}{2}$ ips) on a standard audio tape player was then obtained from the digital tape by off-line equipment for digital-to-analog conversion.

Although each of the complex tones generated in this way was really made up of many sinusoidal tones, most listeners did not hear these sinusoids as subjectively separate components—i.e., as in a chord. Rather, they tended to describe the total complex as a single tone with a sonorous, rather organlike timbre. This is perhaps not surprising, since all higher components occurred at harmonic intervals of the lowest audible component. Moreover, the spectral-amplitude envelope modified these components much as a broad-band resonance would give rise to a formant in the spectrum of a musical instrument.

In addition to these complex tones, pure sinusoidal tones and bursts of wide-band noise were also generated by means of Mathews' computer program. The pure tones were used in some initial tests of each subject's ability to discriminate differences in pitch, and the noise bursts were used for certain signaling purposes in some of the experiments described here. The sequences of tones, noise bursts, delays, as well as other parameters of the sounds themselves, were communicated to the computer program by a deck of punched cards (called

⁸ M. V. Mathews, "The Digital Computer as a Musical Instrument," *Science* 142, 553–557 (1963).

the "score"). In order to reduce transients, attack and decay envelopes (lasting about 0.1 sec each) were imposed at the beginning and end of each tone.

II. PRELIMINARY DEMONSTRATION OF CIRCULARITY

Owing to the cyclic nature of the complex tones considered here, they can conveniently be represented as regularly spaced points around a circle. The convention followed is that a clockwise displacement between neighboring tones represents an upward shift in the frequencies of all corresponding components. For the purposes of preliminary tests, the tones were generated in the order in which they are encountered in passing around this circle in one direction. Since the steps between adjacent tones were produced by half-tone shifts in all corresponding components, the sequence resembled a chromatic scale (of equal temperament) and 12 steps were required to complete exactly one revolution about the circle.

Subjective circularity was clearly demonstrated in a sequence of this kind that made many of these revolutions. Each tone in this sequence was sounded for 0.1 sec and successive tones were always separated by a 0.84-sec period of silence. The components of each tone spanned 10 octaves and the specific values of the parameters were as follows: $t_{\max}=12$, $c_{\max}=10$, $f_{\min}=4.863$ cps, $L_{\min}=22$ dB, and $L_{\max}=56$ dB. (The overall sound-pressure level of each of the complex tones was about 66 dB, but this measure is meaningful only in a relative sense; the audio gain of the tape player was adjusted only to achieve a comfortable listening level.)⁹

The resulting audio tape, or minor modifications of it have now been played to over 60 listeners and, so far all have described the sequence as progressing monotonically in pitch (upward or downward, depending upon whether the tape is played forward or backward). With the forward order, that is, each tone evidently wa

⁹ The main interest here is in the (psychological) problem of demonstrating circularity of judgments rather than in the (psychophysical) problem of systematically exploring the effect of physical parameters on auditory perception. Since the desired demonstration of circularity can be accomplished by an analysis of the pattern of judgments alone, it does not really depend upon any physical measurements of the stimuli. For this reason, determination of the absolute energies of the various components of a tone (as well as of the tone as a whole) at the ear was not required. Also, owing to the relative insensitivity of the ear to tones of low frequency, the lowest three or more components were probably entirely inaudible with the particular values chosen for the parameters (especially since the lowest components range down to only 4.86 cps). Thus, the symmetry of the distribution of sound-pressure levels of the components with respect to low frequency (Fig. 1) did not entail a strict symmetry of the distribution of psychological loudnesses. However, a strict symmetry turned out to be unnecessary for the phenomenon and, probably, almost any smooth distribution that tapers off to subthreshold levels at low and high frequencies would have done as well as the cosine curve actually employed.

always heard as higher in pitch than the preceding. Toward the end of the sequence, some of the listeners became puzzled by the fact that the tones (which clearly had been going up for so long) did not really seem to be getting much "higher." Other listeners, however, did not notice this stationarity of height. Indeed, these latter subjects were astonished to learn that the sequence was cyclic rather than monotonic and that it in fact repeatedly returned to precisely the tone with which it had begun. Several listeners likened the auditory effect to the "stairway to heaven" visual illusion reproduced in Fig. 2. *Note added in proof.* At the time of writing, I was unaware of the origin of the visual "staircase" illusion shown in Fig. 2. J. F. Schouten subsequently called my attention to the fact that it was presented (in slightly different form) by L. S. Penrose and R. Penrose, "Impossible Objects: A Special Type of Visual Illusion," *Brit. J. Psychol.* 49, 31-33 (1958).

The 0.84-sec delays in the sequence just described (or perhaps merely the retardation of the progression that resulted from these delays) seem to be essential to the phenomenon. In any case, the illusion of a monotonic progression failed to emerge in an earlier attempt, in which the sequence proceeded from tone to tone in rapid succession without any delays. That sequence was heard to drop abruptly down an octave at one point during each revolution about the "circle" of 12 tones. Curiously, though, the point at which this downward jump was heard was not completely determined by the physical stimulus but depended, in part, upon the point at which the cyclic sequence was begun. Some subjects with musical training indicated that the perceived drop occurred when the sequence was heard to return to the tonic. Thus it may be an attentional phenomenon governed by the key to which the subject has been set (perhaps by the first note) to relate the ensuing tones.

III. EXPERIMENTAL PROCEDURE

In order to examine circularity of pitch judgments in a more extensive and objective manner, it seemed desirable to randomize the order of isolated pairs of tones so that a subject would have to make a separate judgment (of "up" or "down") with respect to each pair separately. Moreover, such pairs would not have to be confined to adjacent tones in the circular representation. Pairs of diametrically opposed tones, for example, should provide an interesting case of ambiguity, for the second tone could then be regarded equally well as clockwise or counterclockwise from the first.

Accordingly, four sequences of tone pairs were generated for the purposes of a more formal experiment of this kind. The duration of each tone was $\frac{1}{4}$ sec and the two tones within the same pair were also separated by a $\frac{1}{4}$ -sec gap. A relatively soft burst of broad-band noise (54 dB) preceded and followed each pair of tones. This served to alert the subject that a pair was about to be presented, whether the tape was played forward or

backward. These noise bursts lasted $\frac{1}{4}$ sec each and were both separated from the intervening tones by gaps of that same duration. A single presentation thus took about 1 sec for each pair. Between such presentations there was a 3-sec silence during which the subject was to write down his response to the immediately preceding pair—a "U" (if the second tone in the pair seemed to go up from the first) or a "D" (if the second tone seemed to go down from the first).

In order to reduce the number of pairs to be judged, the tonal circle was divided into 10 rather than 12 intervals. Hence, adjacent tones were slightly more than a tempered half-tone apart, and only the diametrically opposed tones were separated by a conventional musical interval (the diminished fifth or "tritone"). A total of 90 distinct ordered pairs could therefore be formed from this set of 10 tones.

The first two sequences of pairs that were presented to each subject were included only as a test of the subject's ability at discriminating differences in pitch. For this purpose, the pairs were made up from 10 pure sinusoidal tones rather than from the 10 complex tones described above. The frequency spacing between these pure tones was the same as that between the complex tones, however. That is, the 10 tones divided one octave into 10 equal intervals on a scale of log frequency. In each of these two sequences, only adjacent tones were presented for comparison and each of the 18 ordered pairs of these appeared just once (in a randomly determined position) within each sequence. In the first sequence, which was intended to be relatively easy, the tones ranged in frequency from 311.2 to 580.8 cps at 62 dB. The second sequence was made more difficult by lowering all tones two octaves (so that the lowest frequency was 77.8 cps) and by reducing the sound level to 56 dB.

The third sequence was designed to provide conclusive evidence of circularity in pitch judgments with the 10 complex tones. For this purpose, only adjacent tones (in the circular representation) were presented in each pair and each of the 20 such ordered pairs was presented twice within the total resulting sequence of 40 pairs. The positions in which the ordered pairs were presented within this sequence were chosen at random with the restrictions that (a) the same tone never appeared in both of two consecutive pairs and (b) each combination to two tones appeared together as a pair just once within each successive block of 10 pairs. The parameters of these tones were as follows: $f_{\max} = 10$, $c_{\max} = 10$, $f_{\min} = 4.863$ cps, $L_{\min} = 22$ dB, and $L_{\max} = 56$ dB.

The fourth sequence that was presented to each subject included all 90 ordered pairs of the 10 complex tones. The purpose of this final sequence was to investigate the possible ambiguity of pairs in which the second tone is on the opposite side of the circle from the first. The parameters of the individual tones were the same as given above, for the third sequence. (In both of these last two sequences, the over-all sound-pressure level for