

Hearing with the Third Ear: Dichotic Perception of a Melody without Monaural Familiarity Cues

Michael Kubovy, James E. Cutting and Roderick McI. McGuire

Hearing with the Third Ear: Dichotic Perception of a Melody without Monaural Familiarity Cues

Abstract. *Julesz has shown that cross-correlations between two patterns that appear random to either eye alone can give rise to the perception of form and depth when viewed stereoscopically. We produced auditory analogs by presenting eight simultaneous and continuous sine waves to both ears and by either phase-shifting or frequency-shifting one of them relative to its counterpart in the opposite ear. Particular tones were shifted in sequence such that a melody was heard which was undetectable by either ear alone.*

Julesz (1) has shown that if one presents a field of random dots to one eye and the same field to the other eye, but with a small portion shifted horizontally, a certain area of the percept appears to stand out in depth. Its contour is the boundary of the shifted portion of dots, and the shift is logically impossible to detect by one eye alone. Julesz named this phenomenon *cyclopean* perception, after the mythical beast that looked out at the world through a single eye in mid-forehead. With random-dot stereograms it is possible to bypass, as it were, the peripheral visual apparatus and project information to the cyclopean eye and onto "the 'mind's retina"—that is, at a place where the left and right visual pathways combine in the visual cortex" (1, p. 3). Our goal was to devise an auditory analog to the cyclopean percept, one which for etymological reasons we call *cyclotean* (2), in which the peripheral auditory apparatus is bypassed and information is projected onto the "mind's cochlea."

We were provoked into seeking this goal, in part, by Julesz (1, p. 51) and Julesz and Hirsh (3), who claim that analogies between visual and auditory perception are not "very deep." The basis for their view is that visual perception is primarily concerned with spatial *objects* whereas auditory perception is primarily concerned with temporal *events*. The distinction between ob-

jects and events appears to be primarily founded on the potential richness of percept in each modality: two spatial dimensions are possible for a visual percept, whereas only the one temporal dimension is available for an auditory percept.

In pre-cyclopean days, Huggins and later Cramer and Huggins (4) demonstrated that if one presents white noise to one ear and the same white noise to the other, but with a narrow band of frequencies time-delayed, a faint pitch quality is heard. It sounds like narrow-band noise, and is logically impossible to detect with one ear alone. Nevertheless, it fails to meet Julesz' criterion for an auditory object: It is "not a truly cyclopean phenomenon since the input variable is a single time delay, while the perceived variable is a single pitch" (1, p. 51). Julesz elaborates by stating that to create a true analog to the visual phenomenon with such pitches one would need to generate a melody. Since a melody is a pattern of pitches and varies in both time and frequency, it is multidimensional and hence an auditory object. Thus, we decided to produce a cyclopean melody. In order to probe the generality of auditory analogs we chose to generate it by two conceptually distinct methods. The first is a methodological offshoot of Cramer and Huggins and is analogous to existing visual

work; the second method, on the other hand, is wholly new.

In both demonstrations the basic stimulus consists of eight, simultaneous, continuous, computer-generated sine waves whose frequencies were chosen from the even-tempered scale in the key of G (5). In both, the tune "Daisy" was embedded in the tonal arrays as a cyclotean melody. All stimuli were computed numerically, generated by the PCM system (6) at Haskins Laboratories, and recorded at the same time on two channels of audio tape.

Our first demonstration begins with the presentation of the basic stimulus to both ears, by earphones, with a lag of 1 msec between the onsets of the two inputs (input A leading input B). A constant discord of eight tones is heard for 1500 msec. At that point the first note, D₅, is introduced by advancing the phase of the D₅ component of input B by 1 msec and by delaying the phase of its counterpart in input A by the same amount. The phase-shifting process is not instantaneous, but occurs over a 45-msec duration. The first note is sustained by maintaining the new phase relation until 900 msec have elapsed after the initiation of the phase shift, at which time the phase-shifting process is reversed (again taking 45 msec) until the two corresponding sine waves resume their original phase relationship. The offset phase-shifting of D₅ completely overlaps with the onset of the subsequent note, B₄. Subsequent notes in the tune are introduced and removed in the same fashion. The duration of each note is from 112.5 msec for an eighth note to 1800 msec for a double-whole note. The duration of the entire sequence is approximately 24 seconds. A spectral segment of it is represented in the top panel of Fig. 1.

Subjectively, the melody is perceived to occur inside the head but displaced to one side of the midline, while a background noise is localized to the

opposite side (7). The notes are perceived to have a sparkling onset, much like that of a chime. Offsets, however, are not particularly striking in any way. When input A or input B is listened to alone, only occasional faint perturbations are audible in an otherwise continuous noise-like signal.

All these observations were confirmed by 24 subjects in two experimental conditions. In the dichotic condition, where input A was presented to one ear and input B to the other, all subjects reported hearing a clear sequence of tones embedded in noise. Only a few failed to recognize the tune (8). In the diotic condition, on the other hand, where one input was presented alone to both ears, no subject reported hearing individual tones, let alone the melody.

The analogies between the preceding demonstration and Julesz' stereograms are fairly straightforward: interaural phase differences correspond to interocular disparities, auditory localization corresponds to visual depth perception, and melodic contour corresponds to visual form.

In the second demonstration, however, the analogies are less obvious: this cyclotean "Daisy" is generated by binaural beats (9). Here, the demonstration begins with the presentation of the basic stimulus to both ears, but with all components of input A 180 degrees out of phase with the corresponding components of input B. A discord of eight tones is heard for 1200 msec. Then D_5 is introduced by increasing the frequency of the D_5 component in input A by 5 hertz and decreasing the frequency of its counterpart in input B by the same amount. Frequency shifts are instantaneous (within a range of $\pm 250 \mu\text{sec}$ in accuracy). The first note is perceived to have the pitch of the original D_5 around which binaural beats of 10 hertz occur. D_5 is sustained for 800 msec, at which point the frequency shift is instantaneously reversed and the two corresponding sine waves resume their original phase and frequency relationships. The second note, B_4 , is introduced by the same process at the offset of D_5 , and all subsequent notes follow the same pattern. Each note is between 100 and 1600 msec in duration, thus embracing 1 to 16 beats. The duration of this second demonstration is approximately 22 seconds. A spectral segment of it is represented in the lower panel of Fig. 1.

Subjectively, the melody is again perceived to occur inside the head. It is not perceived to be localized differently from the background, and yet it does stand out as figure against ground. Beats are not always perceptible, and when they are, they do not appear to be part of the figure or ground but rather a disembodied roughness. Aside from the beats and localization, the new "Daisy" sounds similar to the first, but perceptually more prominent. Again, inputs A and B are not sufficient by themselves to yield the percept. Occasional beats are audible in a single stimulus but these are due to monaural beats where adjacent sine waves are frequency shifted to within 15 hertz of one another.

These observations were confirmed

by a group of 12 subjects. Dichotically, all subjects heard and identified the melody, whereas diotically, none heard any notes. After listening to the first and second demonstrations, most agreed that the second rendition was perceptually more compelling.

Although from the point of view of auditory theory the first demonstration is intimately related to the domain of masking level difference (MLD), it deals with a qualitatively different phenomenon. Whereas MLD is investigated with stimuli that are barely detectable monaurally and better detected binaurally, this cyclotean effect deals with the binaural segregation of auditory figure from ground, where no figure can be perceived monaurally. We do not know, however, what the relation

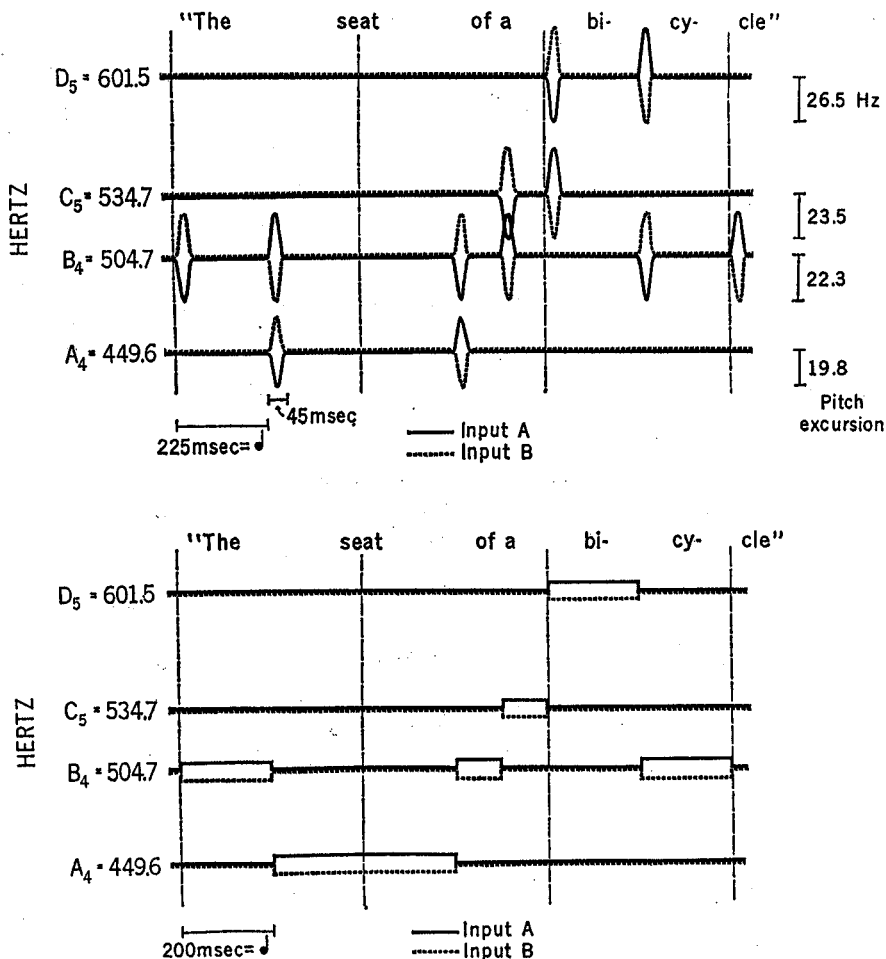


Fig. 1. Schematic spectrograms of the four highest-frequency sine waves in the basic stimuli during that portion of "Daisy" corresponding to the lyric "upon the seat of a bicycle built for two." The frequencies of opposite-ear stimuli are superimposed. The top panel shows that in the first demonstration phase-shifting of opposite-ear stimuli was realized in terms of frequency changes. Since it is not possible to effect a change in phase without also temporarily changing pitch, pitch excursions were necessary. They were realized in raised and lowered cosine functions with maximal increase and decrease of 4.4 percent. The lower panel shows how frequency shifts were realized in the second demonstration for the same segment.

is between MLD's and our second demonstration, since MLD's have not been studied with relatively fast-beating stimuli and it is not known whether they can produce unmasking. Egan (10) has used slowly beating stimuli to demonstrate MLD's, exploiting the principle that stimuli which differ slightly in frequency can be conceived of as stimuli with identical frequencies but with constantly changing phase relations. In his demonstration a signal is heard to fade in and out of noise with the beat frequency. No such fading occurs in our second demonstration. In addition, it should be noted that no cyclopean (visual) analog has yet been generated (11).

It can be argued that both demonstrations are "weakly cyclotean" according to the Julesz definition (1, pp. 17-21): Changes are present in a single stimulus that an instrument could detect to determine which of the simultaneous eight tones underwent sudden phase or frequency shift. The ear, however, is not such an instrument. Moreover, to generate the strongly cyclotean stimulus one need only shift more than one component (preferably all eight) in phase or in frequency, but making such changes identical in both ears except for the crucial melodic component.

MICHAEL KUBOVY

Department of Psychology,
Yale University,
New Haven, Connecticut 06520

JAMES E. CUTTING*

RODERICK McI. McGUIRE

Haskins Laboratories, New Haven,
Connecticut 06510, and Yale University

References and Notes

1. B. Julesz, *Foundations of Cyclopean Perception* (Univ. of Chicago Press, Chicago, 1971).
2. This neologism is constructed from Greek roots to be analogous to the term *cyclopean*.

- Since *ops* is the Greek root for eye and *oto* the root for ear, we feel that *cyclotean* is the proper term for this phenomenon. As yet we have been unable to find reference to such a mythical beast. One reason for this may be that the *cyclo* was a much less truculent, and hence much less memorable, creature.
3. B. Julesz and I. J. Hirsh, in *Human Communication: A Unified View*, E. E. David and P. Denes, Eds. (McGraw-Hill, New York, 1973), pp. 283-340.
 4. First reported by J. C. R. Licklider, in *Information Theory, Third London Symposium*, C. Cherty, Ed. (Butterworth, London, 1956), pp. 253-368, and investigated by E. M. Cramer and W. H. Huggins, *J. Acoust. Soc. Am.* 30, 413 (1958). Subsequent reports have been made by A. J. Fourcin, *Proceedings of the 4th International Congress of Phonetic Sciences* (Mouton, The Hague, 1962), pp. 355-359, and by N. Guttman, *J. Acoust. Soc. Am.* 34, 1996(A) (1962).
 5. The lowest note, D_e , was 300 hz, whereas the highest note, D_g , was adjusted to 601.5 hz to avoid harmonic relationships with D_e . The other six tones were left at even-tempered frequencies since their pairwise frequency ratios are all irrational.
 6. F. S. Cooper and I. G. Mattingly, *J. Acoust. Soc. Am.* 46, 115(A) (1969); also in *Haskins Laboratories Status Report on Speech Research*, SR 17/18 (1969), pp. 17-21.
 7. With phase shifts of 1 msec for each of the eight sine waves, the degree to which opposite-ear sine waves are out of phase is not optimal, but varies with frequency. The amount of interaural phase differences at different frequencies should therefore result in different perceived spatial locations. Critical listening can give rise to a spatially varying cyclotean percept, but few listeners can detect it without being instructed what to listen for and without having previously listened to the second demonstration.
 8. One subject, for example, was a New Zealander who reported hearing "Waltzing Matilda," not "Daisy." We attribute this misperception to his ethnocentricity and general musical inability, not to the subjective strength of the percept.
 9. For discussions of binaural beats see J. C. R. Licklider, J. C. Webster, and J. M. Hedlun, *J. Acoust. Soc. Am.* 22, 468 (1950); D. R. Perrot and M. A. Nelson, *ibid.* 46, 1477 (1969); G. Oster, *Sci. Am.* 229, 94 (Oct. 1973).
 10. J. P. Egan, *J. Acoust. Soc. Am.* 37, 1143 (1965).
 11. Any attempt to generate such a pattern would probably rely on previous work by C. Blakemore [*Vision Res.* 10, 1181 (1970)], who studied binocular combinations of gratings.
 12. Supported by NICHD grant HD-01994 to the Haskins Laboratories. Parts of this report were read at the 87th meeting of the Acoustical Society of America, New York, 23-26 April 1974.
- * Present address: Wesleyan University, Middletown, Conn., and Haskins Laboratories.

12 June 1974