

**CENTER FOR COMPUTER RESEARCH IN MUSIC AND ACOUSTICS
NOVEMBER 1991**

**Department of Music
Report No. STAN-M-74**

**TEMPORAL CONSTRAINTS ON APPARENT MOTION
IN AUDITORY SPACE**

Stephen Lakatos

**CCRMA
DEPARTMENT OF MUSIC
Stanford University
Stanford, California 94305**

Abstract

The hypothesis that the extent of spatial separation between successive sound events directly affects the perception of time intervals between these events was tested using an apparent motion paradigm. Subjects listened to four-note pitch patterns whose individual tones were sounded alternately at one of two loudspeaker positions, and adjusted the alternation rate until they could no longer distinguish the four-tone ordering of the pattern; four horizontal and two vertical loudspeaker separations were tested. Results indicate a direct relationship between horizontal separation and the critical stimulus onset asynchrony (SOA) between successive tones within a pattern. At the critical SOA, subjects reported hearing not a four-tone pattern, but two pairs of two-note groups overlapping in time. The findings are discussed in the context of auditory spatial processing mechanisms and possible sensory-specific representational constraints.

Temporal Constraints on
Apparent Motion in Auditory Space

Investigators have made extensive use of the phenomenon of apparent motion in the study of mental representations and transformations. In the illusion of apparent motion, successive presentation of objects in different locations under suitable conditions leads to the compelling experience of a single object moving continuously between locations. Shepard (1981) has argued that the tendency to interpret successive glimpses of two discrete entities as a single moving object stems from our exposure on an evolutionary scale to transformations of semirigid objects in a three-dimensional world. Studies of visual apparent motion suggest ecological constraints governing depth perception (Corbin, 1942), spatial trajectory (Shepard & Zare, 1983), and transformations in shape and orientation (Farrell & Shepard, 1981), among others. There is usually a clear and systematic relation between the minimum onset asynchrony between presentations that yields good apparent motion and the nature and extent of the implied transformation between them. In general, mental transformations induced by apparent motion consistently reflect constraints found in our three-dimensional, locally Euclidean, isotropic world (Shepard & Cooper, 1982)

If the most enduring transformational constraints of three-dimensional space have indeed been internalized to form part of our innate perceptual mechanisms, logic would dictate that internalized rules of spatial transformation should be amodal. However, evidence for the same rules governing spatial apparent motion in vision and

hearing has not been compelling. For example, the direct relationship between object separation and stimulus onset asynchrony (SOA) first described by Korte (1915) and subsequently replicated in numerous different contexts (see Anstis, 1978, for an overview of the visual apparent motion literature) does not appear to hold for spatial hearing (Strybel et al., 1989). Vision and hearing may simply differ in terms of their spatial resolution. The spatial acuity of the visual system is vastly superior to that of the auditory system; the maximum spatial resolution of the auditory system under ideal conditions is only about 1° (Mills, 1958). Perrott and Musicant (1977) have found the minimum audible angle for a sound source moving at 360° per second to be only about 21° . The auditory system may be incapable of discriminating angular separations wherein apparent motion effects can be induced. Kubovy (1981) has argued that the proper auditory analogue of visual space is auditory pitch. In fact, studies of streaming effects in auditory pitch (Bregman & Campbell, 1971) show increases in SOA with increasing separation in pitch between two alternately presented tones. Further evidence for consistent cross-modal influences on apparent movement in vision and auditory pitch (O'Leary & Rhodes, 1984) strengthens the theory that auditory pitch is the correct analogue of visual space.

Is auditory space unconstrained by representational and transformational principles that apply to visual space? Although auditory localization does not have the spatial resolution of vision, it would seem premature to rule out the existence of internalized spatial constraints on this basis alone, without considering more fully

how the auditory system processes spatial information. Auditory localization not only has limited spatial resolution, but also encodes spatial information indirectly during its early processing stages, in terms of interaural time and intensity differences. Since spatial information is encoded topographically only at higher brain levels (see Knudsen, 1984, for physiological evidence in owls), analogical constraints for spatial hearing may not appear until relatively advanced cognitive stages. Rhodes (1987), for example, has found evidence that auditory attention shifts analogically in space.

In this study, the hypothesis that initial temporal processing of auditory spatial information would preclude the existence of rapidly applicable spatial constraints was tested using an experimental apparatus designed to examine apparent motion in auditory space. Specifically, I conjectured that auditory localization would not internalize analogical constraints relating physical distance and motion during its early stages; rather, given the initial temporal representation of spatial information, one might expect a relationship between distance and time. Instead of an interpolation between discrete physical objects induced by temporal factors (the case with visual apparent motion), one could expect the opposite: An impletion of temporal relations between sound sources, as it were, based upon their physical separation. The auditory system might perceive a temporal continuity between rapidly changing events based on their spatial proximity, rather than the opposite.

To this end, I designed an experiment to separate the individual components of an auditory pattern in space and to determine whether

spatial separation directly affected perception of temporal order within the pattern. If we tend to hear rapidly alternating sounds as occurring closer together in time when they are physically near as opposed to when they are far apart, then a pattern comprising such sounds should undergo changes in its temporal structure depending on the relative spatial separation of its component sounds. Specifically, one might expect a decrease in the apparent time interval between sounds that are physically near. A four-tone pitch pattern (see Figure 1, Pattern 1) was tested in a pilot study to determine whether its temporal order changed as a function of varying horizontal and vertical separation at various alternation rates. At a specific separation, each of the four tones sounded alternately from one of two loudspeakers. Results indicated that the rate at which a four-tone order could be followed decreased with increasing angular separation; beyond a certain threshold rate, four-tone grouping consistently deteriorated into two increasingly overlapping two-tone patterns grouped by loudspeaker location. In light of these observations, I predicted an inverse relationship between the extent of angular separation and the critical rate for maintaining four-tone pattern integrity.

Three different pitch patterns were selected to test the additional prediction that apparent motion in auditory pitch between pairs of tones within the patterns would modulate to some degree any spatiotemporal motion effects. Studies of apparent motion in auditory pitch (e.g., Bregman & Campbell, 1971) have indicated that optimal rates of apparent motion are inversely related to differences

in pitch height. Based on these results, it was predicted that pitch relationships between individual tones would interact with spatial separation in determining critical rates for temporal order. The three pitch patterns in this study were selected to offer minimal to substantial resistance to alternate groupings induced by spatial separation. For example, if the perceptual streaming by auditory pitch coincided with streaming effects induced by spatial separation, breakdown of the four-tone pattern would be expected at fairly slow alternation rates; on the other hand, if pitch and spatial streaming were placed in opposition, it would probably be quite difficult to segregate four-tone patterns, even at fast rates of alternation. Therefore, pitch relations within patterns were expected to influence but not to mask temporal changes in pattern structures induced by spatial separation.

Method

Subjects

Twelve researchers at Stanford's Center for Computer Research in Music and Acoustics (CCRMA) participated in the experiment. All subjects had normal hearing and extensive musical training.

Stimuli and Apparatus

The experiment was conducted in an isolated room lined with fiberglass panels and cylindrical sound traps to yield a listening environment relatively free from reverberation. An array of 8 horizontal and 4 vertical loudspeakers (Boston Acoustic) was positioned 4.5 ft from the subject's head. Four different horizontal

(20°, 50°, 80°, 110°) and two vertical (25°, 67°) angular separations were tested.

The stimuli were three sets of continuously cycled four-tone patterns (Figure 1) whose presentation rate could be varied continuously by means of a lap-held slider. On a given trial, each tone of a given pattern was sounded alternately at one of the two speaker positions (horizontal or vertical separation) so that the pattern was spatially distributed between the two speakers on a tone-by-tone basis. Patterns were distinguished by their relative

Insert Figure 1 about here

intervallic structures: Pitches from the same speakers in Pattern 1 were clustered more closely (Speaker A: 392, 440 Hz; Speaker B: 554, 622 Hz) than they were in Pattern 3 (Speaker A: 392, 554 Hz; Speaker B: 440, 622 Hz), with Pattern 2 as an intermediate case (Speaker A: 392, 466 Hz; Speaker A: 523, 622 Hz). Tones used in the patterns were complex signals whose musical timbre can best be described subjectively as that of a sharply plucked string. The duration of each sound in the pattern was always a fixed percentage (18.75%) of the duration of the pattern as a whole. Individual tones were separated by a duration 6.25% of the overall pattern duration at any given moment for setting of the slider.

Stimulus presentation was controlled by a Macintosh microcomputer operating in conjunction with two digital synthesizers (Yamaha TX802 FM Tone Generators) by means of a MIDI link

(Opcode Systems). The synthesizers provided 12 independent audio channels, one for each speaker. Subject responses were recorded using a MIDI-compatible control board (Yamaha MCS2 MIDI Control Station) connected to the microcomputer.

Procedure

At the beginning of the experiment, the subject was shown the speaker array and instructed to increase the rate of alternation on a given trial until the four-tone ordering of the pattern became ambiguous; in other words, the tones of the pattern were to alternate between two speakers just fast enough so that subjects could not distinguish their ordering. Subjects were told that they might hear alternate groupings of pattern tones at such rates, such as two temporally overlapping pairs. A short practice session was administered before the main body of the experiment to familiarize the subject with the task.

There were 4 blocks of 18 trials each. Each of the six loudspeaker separations appeared with each of the three pitch patterns once per block. All four blocks were administered during a single session that lasted approximately 1 h.

Results

One subject consistently selected nearly identical rates of alternation for all horizontal and vertical separations. This subject's data were excluded from statistical calculations.

The effects of different horizontal speaker separations on temporal coherence of each pitch pattern were above the chance level for all pairs of separations, except for the two largest separations, 80°

and 110° . The main differences among the horizontal separations are shown for each pitch pattern in Figures 2a-c, respectively (standard error bars are shown). Means are pooled across subjects and are

Insert Figure 2 about here

expressed as critical SOAs between consecutive tones within a pattern. Each bar is based on 33 trials. There was a significant increase in threshold for identification of temporal order of all three pitch patterns as horizontal angular separation increased from 20° to 80° . Two-way analysis of variance of horizontal-angle/pitch-pattern means revealed highly significant differences among angular separations ($F(3, 120) = 17.90, p < .001$) and pitch patterns ($F(2, 120) = 18.30, p < .001$), but no significant interactions between these variables.

Vertical speaker separation had no significant effect on critical SOA (Figure 3a-c). Despite a difference in angular separation of 42° ,

Insert Figure 3 about here

subject responses tended to remain consistently at SOAs of between approximately 100 msec and 120 msec, well below those of all horizontal separations except one (Figure 2c, 20° separation). Significant F-ratios were found for pitch-pattern means ($F(2, 60) = 16.36, p < .001$), but not for vertical separation means ($F(1, 60) = 0.83, p > .25$), indicating that individual pitch patterns differed in their

overall temporal salience, while exhibiting no statistically reliable changes as a function of vertical location.

Post-experimental questioning indicated that subjects had little trouble locating a distinct point at which the temporal integrity of the three pitch patterns broke down. As rates increased beyond this threshold, subjects tended to hear two pairs of alternating tones whose temporal boundaries began to overlap. At extremely high rates of alternation (approx. 50 msec SOA) and wide horizontal separations (80° and 110°), subjects reported hearing two pairs of simultaneous tones alternating between speakers; tones from the same speaker appeared to overlap completely. Vertical separations yielded less clear-cut threshold points; nonetheless, subjects did not indicate substantial difficulties in making judgements for this dimension.

Discussion

The results indicate a direct relation between temporal and spatial constraints in the experience of auditory apparent motion. A distortion of temporal events, in which tones from the same speaker appeared to occur closer in time, was induced by distributing the individual tones of rapidly alternating temporal patterns horizontally. Subjects consistently required longer SOAs for the identification of temporal order as the horizontal distribution of pattern components increased. The lack of significant difference in threshold rates between 80° and 110° separations may simply reflect the inability of the auditory system to resolve localization cues at these positions; thresholds for detecting angular differences

in location have been shown to increase markedly in this azimuthal range (Mills, 1958). The frequency distribution of tones within patterns had the predicted effect of modulating the overall threshold for identification of temporal order across all horizontal separations. The pattern of the rate at which four-tone order could be followed with increasing horizontal spacing was essentially the same for each of the pitch sequences.

Vertical spatial separation of pattern components appeared to have no significant temporal consequences. This may stem from the different processing mechanisms for horizontal and vertical localization. Acoustical information about source elevation is derived from pinna-induced spectral filtering (Butler & Belendiuk, 1977), while positions having nonzero azimuth angles are detected primarily through interaural time and intensity differences (Mills, 1972), and to a lesser extent through directional filtering by the external ears (Musicant & Butler, 1985). It may be that the complex filtering mechanisms involved in vertical localization operate on time scales that would preclude their use in rapidly changing localization situations. Some subjects noted a loss or compression of vertical sense as pattern presentation rates increased, a possible indication that vertical localization mechanisms were unable to processing spatial input beyond a certain rate.

The direct relationship between time and spatial location presented here confirms that early temporal processing of spatial information by the auditory system leads to internalized constraints governing temporal continuity between spatially discrete events,

rather than spatial continuity between temporally discrete objects, as one observes in visual apparent motion. It would seem inherently logical for the auditory system, whose principal function is the interpretation of temporal information, to attempt to resolve ambiguous and rapidly changing spatial information (i.e., stroboscopic motion) assuming that sounds from the same location are closer in time than they actually are. In this sense, the results indicate a closer interaction between temporal and spatial factors in spatial hearing than one might expect by distinguishing between auditory parameters on the basis of dispensable vs. indispensable attributes (Kubovy, 1981) or Gestalt grouping mechanisms (Deutsch, 1982), for instance.

In more general terms, the results reported here indicate that it may be more fruitful to think of representational constraints on rapidly changing sensory input as products of processing mechanisms at specific levels in perceptual systems, rather than as amodal internalizations. From an ecological standpoint, if an organism must act quickly on sensory input, there may simply not be enough time to allow for the processing of high-level representations; even if time constraints are not critical, spatial information may be too ambiguous to reconcile with existing representational formats. When faced with either of these situations, perceptual systems may force an interpretation based on their principal sensory dimension. The visual system, primarily a receptor of spatial information, reconciles rapidly alternating objects by interpolating an intervening spatial trajectory; audition, on the other hand, shifts temporal relations between events to compensate for spatial uncertainty. If

there is sufficient time, sensory systems may gain access to more global representational formats. The faster a judgement about ambiguous spatial input must be rendered, the more it would seem each sensory system must rely on its own resources.

References

- Anstis, S. M. (1978). Apparent movement. In R. Held, H. W. Leibowitz, & H.-L. Teuber (Eds.), *Handbook of sensory physiology VII*. Berlin, West Germany: Springer-Verlag.
- Bregman, A. S., & Campbell, J. (1971). Primary auditory stream segregation and the perception of order in rapid sequences of tones. *Journal of Experimental Psychology*, **89**, 244-249.
- Butler, R. A., & Belendiuk, K. (1977). Spectral cues utilized in the localization of sound in the median sagittal plane. *J. Acoust. Soc. Am.*, **61**, 1264-1269.
- Corbin, H. H. (1942). The perception of grouping and apparent movement in visual depth. *Archives of Psychology*, **38**[Series No. 273], 5-50.
- Deutsch, D. (1982) Grouping mechanisms in music. In D. Deutsch (Ed.), *The psychology of music*. San Fransisco: Academic Press.
- Farrell, J. E., and Shepard, R. N. (1981). Shape, orientation, and apparent rotational motion. *Journal of Experimental Psychology: Human Perception and Performance*, **7**, 477-486.
- Knudsen, E. I. (1984). Synthesis of a neural map of auditory space in the owl. *The Journal of Neuroscience*, **2**, 1177-1194.
- Korte, A. (1915). Kinematoskopische Untersuchungen. *Zeitschrift fur Psychologie*, **72**, 193-206.
- Kubovy, M. (1981). Concurrent-pitch segregation and the theory of indispensable attributes. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization*. Hillsdale, NJ: Erlbaum.

- Mills, A. W. (1958). On the minimum audible angle. *J. Acoust. Soc. Am.*, **30**, 237-246.
- Mills, A. W. (1972). Auditory localization. In J. V. Tobias (Ed.), *Foundations of modern auditory theory*, Vol. 2. New York: Academic Press.
- Musicant, A. D., & Butler, R. A. (1984). Influence of monaural spectral cues on binaural localization. *J. Acoust. Soc. Am.*, **77**, 202-208.
- O'Leary, A., & Rhodes, G. (1984). Cross-modal effects on visual and auditory object perception. *Perception & Psychophysics*, **35**, 565-69.
- Perrott, D. R., & Musicant, A. D. (1977). Minimum auditory movement angle: Binaural localization of moving sound sources. *J. Acoust. Soc. Am.*, **62**, 1463-1466.
- Rhodes, G. (1987). Auditory attention and the representation of spatial information. *Perception & Psychophysics*, **42**, 1-14.
- Shepard, R. N. (1981). Psychophysical complementarity. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization*. Hillsdale, NJ: Erlbaum.
- Shepard, R. N., & Cooper, L. A. (1982). *Mental images and their transformations*. Cambridge, MA: MIT Press.
- Shepard, R. N., & Zare, S. L. (1983). Path-guided apparent motion. *Science*, **220**, 632-634.
- Strybel, T. Z., Manligas, C. L., & Perrott, D. R. (1989). Auditory apparent motion under binaural and monaural listening conditions. *Perception & Psychophysics*, **45**, 371-377.

Author Notes

The author is grateful to Earl Schubert and Roger Shepard for their thoughtful comments on a draft of this article.

Requests for reprints should be sent to Stephen Lakatos, Center for Computer Research in Music and Acoustics (CCRMA), Stanford University, Stanford, CA 94305.

Legends

Figure 1. Pitch patterns used in the experiment. Speaker order indicates the sequential presentation of pattern tones at a speaker pair A-B separated by one of six horizontal or vertical angles. In the experiment, a given pattern cycled continuously until the subject selected a rate at which temporal order became ambiguous.

Figure 2. The effect of horizontal separation of pattern components on the critical SOA between consecutive components is shown for pitch patterns 1-3 (a-c). Standard error bars are shown.

Figure 3. The effect of vertical separation of pattern components on the critical SOA between consecutive components is shown for pitch patterns 1-3 (a-c).

Figure 1



