

Should We Resist the Seductiveness of the Space:Time::Vision:Audition Analogy?

Michael Kubovy
University of Virginia

According to Handel (1988), the analogy *space:time::vision:audition* is seductive but misleading: We cannot "imagine a visual or auditory event that is nonspatial or atemporal" (Handel, 1988, p. 315); this statement either begs the question or is false. He claims that people mentally transform space and time in both modalities; I argue that we never mentally transform auditory space but often transpose in pitch and translate in time. Handel claims that all other "modality translations," including my theory of indispensable attributes (*space-time::pitch-time::vision:audition*), "are possible and relevant" (p. 316) and that theoretical commitment on this matter is unnecessary; I reject this tolerance. I show that the analogy rejected by Handel is partly true: (a) Space is the province of vision. (b) Vision is not inherently temporal. (c) Audition is intimately tied to time. (d) Audition is not inherently spatial. These statements are implied by my theory of indispensable attributes.

People who are not cognitive psychologists have theories about cognitive phenomena; the corpus of such theories constitutes *folk cognitive psychology*. (For the analogous idea of folk taxonomy, see Lakoff, 1987, pp. 32-38). Some scientific cognitive theories are rational reconstructions (or systematic reformulations) of folk cognitive theories, and some contain implicit assumptions drawn from folk cognitive theories. It was to the latter that Garner (1981) was referring when he wrote about "experiments we know better than to do" (p. 137). Handel's article "Space is to Time as Vision is to Audition: Seductive but Misleading" (Handel, 1988), makes explicit a fragment of folk cognitive psychology (*space:time::vision:audition*), rejects it, and then proceeds to lay out alternative views, one of which is my theory of indispensable attributes (Kubovy, 1981, 1987), in which I propose the analogy *space-time::pitch-time::vision:audition*. Handel concludes that "all of the modality translations are possible and relevant" (except, presumably, the one he rejected), and that "there is no need to make a theoretical commitment" (p. 316).

The analogy is not easy to dismiss. Handel wishes to persuade us that vision and audition are both spatial and temporal by asserting that "it is impossible to imagine a visual or auditory event which is nonspatial or atemporal" (p. 315). But we should not be persuaded: This statement either begs the question, or it is false. If, on the one hand, the statement referred to a visual *event*, then of course it would be true, but trivially so, because by definition events involve change. If, on the other hand, the statement referred to a visual object or scene, then it would be false. For unless it were possible to imagine an atemporal visual experience, Handel would not have admonished, a few lines earlier, that "it is misleading to

conceptualize the visual world as a snapshot, a static picture so that space is the framework or the lattice." Others, such as Schopenhauer (1859/1969), have insisted that audition is nonspatial:

Perceptions through *hearing* are exclusively in *time*; hence the whole nature of music consists in the measure of time, and on this depends not only the quality or pitch of tones by means of vibrations, but also their quantity or duration by means of the beat or time. The perceptions of *sight*, on the other hand, are primarily and predominantly in *space*; but secondarily, through their duration, they are in time also. (Vol. 2, p. 28)

Having disagreed with Handel's argument against the analogy, I wish to delimit the scope of my disagreement: First, I am not endorsing the visual-world-as-snapshot or the hearing-is-time views that Handel decries. I am claiming, however, that (a) perceptual theory must account for the seductiveness of this folk analogy, and (b) this folk analogy is no more misleading than any of the alternatives Handel asks us to consider. Second, my objections to Handel's argument do not imply a belief in the existence of objects outside a spatiotemporal framework, which is the framework of physics. But the framework of physics should not be imposed on intermodality analogies.

Even if Handel had convincingly undermined the *space:time::vision:audition* analogy, his tolerant conclusions that "all modality translations are possible and relevant" (p. 316) and that "there is no need to make a theoretical commitment" (p. 316) would not lead to progress because, as Conant (1961) remarked,

A conceptual scheme is never discarded merely because of a few stubborn facts with which it cannot be reconciled; a conceptual scheme is either modified or replaced by a better one, never abandoned with nothing left to take its place. (p. 173)

There are four aspects to the *space:time::vision:audition* analogy. First, a positive claim about vision: Space is the

This research was supported by National Science Foundation Grant BNS-82-10578.

Correspondence concerning this article should be addressed to Michael Kubovy, Department of Psychology, Gilmer Hall, University of Virginia, Charlottesville, Virginia 22903-2477.

provision of vision. The claim probably springs from the same source as the metaphors we use to characterize the visual field (Lakoff & Johnson, 1980):

We conceptualize our visual field as a container and conceptualize what we see as being inside it. Even the term "visual *field*" suggests this. . . . Given that a bounded physical space is a CONTAINER and that our field of vision correlates with that bounded physical space, the metaphorical concept VISUAL FIELDS ARE CONTAINERS emerges naturally. Thus we can say:

The ship is *coming into view*.
I *have* him *in sight*.
I can't see him—the tree is *in* the way.
He's *out* of sight now.
That's *in* the *center* of my *field* of vision.
There's *nothing in sight*.
I can't get *all* of the ships *in sight* at once. (p. 30)

Similarly, Miller and Johnson-Laird (1976) observe that "Looking is a directional activity and presumes a space in which entities are located . . . In looking for something you are apt to scan the scene, to move around in search of the desired object . . ." (p. 617).

The second aspect of the analogy is a negative claim about vision: Vision is not inherently temporal. Looking presupposes objects located in space, but it does not presuppose time. Seeing presupposes objects, but it does not presuppose events. Events *may* take place in the visual field, but they are not necessary. Looking is an activity that involves scanning and searching, but the actions it consists of occur in the looking organism, not in the environment. Of course, any activity we engage in presupposes time, but looking does not presuppose time any more than any other activity does.

The third aspect of the analogy is a positive claim about audition: Audition is intimately tied to time. Schopenhauer explains this in the passage I quoted above: Both vibration and the musical beat are inherently temporal, albeit in different ways.

The fourth aspect of the analogy, and the most difficult to explain, is a negative claim about audition: Audition is *not* spatial. As Miller and Johnson-Laird (1976) show, although the distinction between "listen" and "hear" seems to parallel the distinction between "look" and "see," the parallel is imperfect (Miller, 1974; Rogers, 1971):

If you look at something you do not necessarily see it, but if you listen to something you hear it. We can construe listening as doing something that causes you to hear. Looking is a directional activity and presumes a space in which entities are located, whereas listening is only indirectly spatial. For example, in looking for something you are apt to scan the scene, to move around in search of the desired object, but in listening you are apt to wait quietly with your attention concentrated on what you hear. (p. 617)

Let us pursue the question of the spatiality of audition a bit further. Even though Handel admits that "it seems more difficult to think of audition as being spatial than vision as being temporal," he claims that "auditory events are invariably localized at particular places within an extended space (p. 315). In my theory of indispensable attributes¹ (Kubovy,

1981, 1987), I offer a rational basis for downplaying the role of space in audition as an analogue of space in vision. First, I argue that interesting analogies between vision and audition should be sought in analogous Gestalt phenomena, not analogous sensory phenomena. Then I claim that perceptual organization in audition is not confined to organization in time, but exists in pitch as well; however, auditory perceptual organization does not appear in space. Finally, I conclude that the most general analogy between vision and audition is space-time:pitch-time::vision:audition. I have found the following thought experiment useful in giving an intuitive interpretation to the dominance of pitch over spatial location in determining the perceptual organization of a sound. Consider two rooms in each of which are six hidden loudspeakers: In one they form a pentagon with one loudspeaker at its center; in the other they form two concentric triangles. Over these loudspeakers we play a sextet. Will a listener who enters these rooms *spontaneously* form a spatial image of the disposition of the loudspeakers? Listeners who are asked to describe the spatial configuration, will try to analyze the sound into six streams, and then try (serially) to assign a location to each source. From this information they will try to infer the shape defined by the loudspeakers. The auditory system was not designed to resolve spatial configurations of sound. For instance, the minimal audible angle (MAA) for concurrent sounds is much higher than the MAA for successive sounds (Perrot, 1984). This is not to deny that source localization, extensity perception, and spaciousness perception are important auditory functions (which Handel takes pains to establish), but to deny that it is used for auditory *spatial form* perception.² Indeed, there is some reason to believe (Auerbach & Sperling, 1974; Platt & Warren, 1972; Warren, 1970) that localization in auditory system is connected to orienting in visual system, as has been demonstrated in the owl (Konishi, Sullivan, & Takahashi, 1985), but not yet in other species (Brugge, 1985), and may have evolved in service of orienting in the visual system.

The final point of Handel's that I wish to address concerns the active perceiver:

The active perceiver can "move" in the spatial lattice to achieve a different perspective. This "movement" is illustrated by experiments on mental rotation (Shepard & Cooper, 1982), expansion and inversion (Kosslyn, 1980), the sense of reorienting a scene or a picture to obtain a frontal view (Kubovy, 1986), and the ability to translate sounds in space (Blauert, 1974/1983). By the same token, there are many cases in which perceivers move in the temporal lattice, again to perceive a different perspective. For example, when listening to music, listeners will try to perceive the characteristic repetitive structure—they will slide their perceived starting point (e.g., Garner, 1974) and will change the size of the temporal unit (the beat or meter) to match the frequency and timing pattern (Povel & Essens, 1985). (p. 316)

Handel wants to persuade us that people engage in active mental transformations of their perspective in space and time,

¹ The following comments are adapted from Kubovy (1987).

² Analogously, I have argued that color, albeit an important visual function, is not used for visual form perception.

both in audition and in vision. I agree with Handel's emphasis on mental transformations, but I believe that people perform mental transformations only in space and time for visual images, in pitch and time for auditory images. It is telling that Handel does not cite any auditory examples of spatial transformations. I believe that he does not because they do not exist. We *never* mentally change our auditory spatial perspective, whereas we easily transpose in pitch, which I claim is the auditory analogue of visual space or the visual field. Mach (1959) formulated the idea thus:

A tonal series occurs in something which is an analogue of space, but is a space of one dimension limited in both directions and exhibiting no symmetry like that, for instance, of a straight line running from right to left in a direction perpendicular to the median plane. It more resembles a vertical line. . . . That the province of tone-sensation offers an analogy to space, and to a space having no symmetry, is unconsciously expressed in language. We speak of high tones and deep tones, not of right tones and left tones, although our musical instruments suggest the latter designation as a very natural one. (p. 278)³

Following von Ehrenfels (1890), who was the first to claim that musical chords and visual forms share the property of being transposable without losing their identity, Attneave and Olson (1971) have argued:

a melodic phase. . . . is *transposable*: the pattern as a whole may be raised or lowered in pitch without destroying its perceptual identity. The corollary of this statement is that pitch is a *medium*, in which the same pattern may have different locations. (p. 148)

In conclusion, the space:time::vision:audition analogy is seductive, but not full-grown. Nevertheless, it is better to yield to its charms than to allow the other analogies to corrupt us. At least the space:time::vision:audition analogy can develop into the analogy we can marry: space-time:pitch-time::vision:audition.

³ Further observations on this point can be found in Mach, 1959, pp. 282-284. I also recommend the deep phenomenological discussion in chapters 7 through 10, and chapters 14 through 19 of Zuckerkandl (1969).

References

- Attneave, F., & Olson, R. K. (1971). Pitch as a medium: A new approach to psychophysical scaling. *Journal of Experimental Psychology*, 81, 147-166.
- Auerbach, C., & Sperling, P. A. (1974). A common visual-auditory space: Evidence for its reality. *Perception & Psychophysics*, 16, 129-135.
- Blauret, J. (1983). *Spatial hearing*. (J. S. Allen, Trans.). Cambridge, MA: MIT Press. (Original work published 1974).
- Brugge, J. F. (1985). Patterns of organization in auditory cortex. *Journal of the Acoustical Society of America*, 78, 353-359.
- Conant, J. B. (1961). *Science and common sense* (2nd ed.). New Haven: Yale University Press.
- Ehrenfels, C. von. (1890). Über "Gestaltqualitäten" [On Gestalt properties]. *Vierteljahrsschrift*, 14, 249-292.
- Garner, W. R. (1974). *The processing of information and structure*. Hillsdale, NJ: Erlbaum.
- Garner, W. R. (1981). The analysis of unanalyzed perceptions. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization* (pp. 119-139). Hillsdale, NJ: Erlbaum.
- Handel, S. (1988). Space is to time as vision is to audition: Seductive but misleading. *Journal of Experimental Psychology: Human Perception and Performance*, 14, 315-317.
- Konishi, M., Sullivan, W. E., & Takahashi, T. (1985). The owl's cochlear nuclei process different sound localization cues. *Journal of the Acoustical Society of America*, 78, 360-364.
- Kosslyn, S. M. (1980). *Image and mind*. Cambridge, MA: Harvard University Press.
- Kubovy, M. (1981). Concurrent pitch-segregation and the theory of indispensable attributes. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization* (pp. 55-98). Hillsdale, NJ: Erlbaum.
- Kubovy, M. (1986). *The psychology of perspective and Renaissance art*. Cambridge, England: Cambridge University Press.
- Kubovy, M. (1987). Concurrent pitch segregation. In W. A. Yost & C. S. Watson (Eds.), *Auditory processing of complex sounds* (pp. 299-314). Hillsdale, NJ: Erlbaum.
- Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.
- Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.
- Mach, E. (1959). *The analysis of sensations*. New York: Dover.
- Miller, G. A. (1974). Toward a third metaphor for psycholinguistics. In W. B. Weimer & D. S. Palermo (Eds.), *Cognition and the symbolic processes* (pp. 397-413). Hillsdale, NJ: Erlbaum.
- Miller, G. A., & Johnson-Laird, P. N. (1976). *Language and perception*. Cambridge, MA: The Belknap Press.
- Perrott, D. R. (1984). Concurrent minimum audible angle: A reexamination of the concept of auditory spatial acuity. *Journal of the Acoustical Society of America*, 75, 1201-1206.
- Platt, B. B., & Warren, D. H. (1972). Auditory localization: The importance of eye movements and a textured visual environment. *Perception & Psychophysics*, 12, 245-248.
- Povel, D. J., & Essens, P. (1985). Perception of temporal patterns. *Music Perception*, 2, 411-440.
- Rogers, A. (1971). *Three kinds of physical perception verbs*. Chicago: Chicago Linguistic Society.
- Schopenhauer, A. (1969). *The world as will and representation*. New York: Dover. (Translation of 3rd. ed. published in 1859).
- Shepard, R. N., & Cooper, L. A. (1982). *Mental images and their transformations*. Cambridge, MA: MIT Press.
- Warren, D. H. (1970). Intermodality interactions in spatial localization. *Cognitive Psychology*, 1, 114-133.
- Zuckerkandl, V. (1969). *Sound and symbol: Music and the external world* (W. R. Trask, Trans.). Princeton: Princeton University Press. (Original work published 1956).

Received July 13, 1987

Accepted August 10, 1987 ■