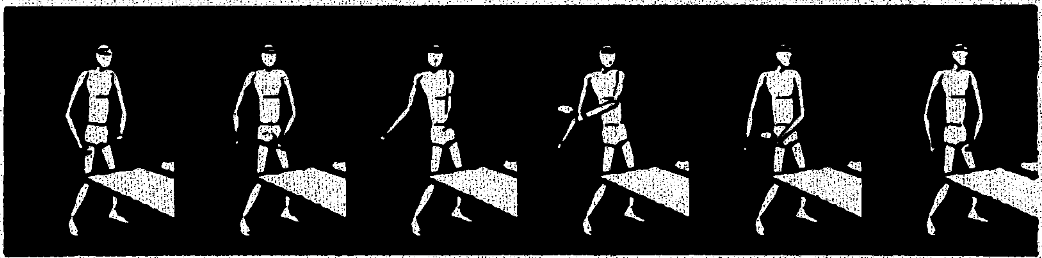
Moving Image Theory



Ecological Considerations

Edited by
Joseph D. Anderson
and Barbara Fisher Anderson

Foreword by
David Bordwell

Moving Image Theory
Ecological Considerations
Edited by: Joseph D. Anderson &
Barbara Fisher Anderson
Southern Illinois Univ. (2005)

6 Acoustic Specification of Object Properties

Claudia Carello, Jeffrey B. Wagman, and Michael T. Turvey

A BOTTLE SITS unseen on the kitchen counter. Your elbow clips it inadvertently and sends it hurtling towards the floor. As you cringe, waiting for the crash, what you hear is not a shattering mess but a harmless bounce. The impact of glass on linoleum has set the materials into vibration, generating compression waves in the air. Somehow, from this sound structure, you know that the bottle did not break. Moreover, the people listening in the dining room heard something bounce as well. Let's focus on those listeners whose only contact with the event is from sound. What else do they know about what happened? Can they hear, for example, that the fallen object was made of glass, that it was a bottle, whether it was large or small, full or empty?

The preceding questions concern what listeners perceive about events happening around them. On the basis of its inattention to such questions, the science of perception can be considered skeptical that audition makes us aware of our surroundings with anything approaching the level of precision that vision allows (Jenkins, 1985). Hearing's specialties are thought to lie in perceiving speech and music. Beyond orienting the listener to the direction of a crash, hearing is not considered to be of much use in obtaining information about geometric properties, such as letting us know the sizes and shapes of objects. At least this has been the bias of orthodox approaches to perception. Shape perception and space perception are the traditional province of vision; pitch perception and loudness perception are the traditional domain of audition. Whereas vision is about awareness of environmental properties, audition seems to be largely about the awareness of sound as such.

Orthodoxy, of course, is likely to inspire heterodoxy. For perception, a contrary treatment can be found in the ecological approach of James J. Gibson. A major innovation is its focus on perceiving object properties rather than sound properties—hearing a small, hollow, glass object falling onto a hard surface rather than a loud, low-pitched, brief sound. In what follows, we describe this approach in general terms. From this description, it will become apparent that limiting the perception of geometric properties to the domain of vision is more tradition than necessity. Subsequent sections will address the lawfulness of acoustic structure that ought to permit auditory perception of geometric properties. Finally, we summarize research that documents listeners' success.

The Ecological Approach to Perception

The ecological approach to perception is a metatheory. As such, it describes a particular conceptualization of how we can be aware of our surroundings. In general, metatheories endorse a particular style of framing questions, promote certain strategies for addressing them, and seek particular kinds of answers.[1] The prevailing metatheory is that perception begins with inadequate input, which becomes meaningful by virtue of internal computations. These computations permit awareness of the world indirectly in the form of a mental representation. In contrast, the ecological approach asserts that input, once properly construed, is rich and lawful and specific to its source directly without elaboration by internal mechanisms.

The properness of the construal is an important issue in revealing lawfulness, and this construal is what motivates the epithet ecological. Consider the form that a so-called ecological law takes:

$$\text{event} \quad \underset{\textit{specifies}}{\overset{\textit{generates}}{\rightleftarrows}} \quad \text{structured energy distribution}$$

The inclusion of something as coarse-grained as an event highlights the ecological commitment to discovering the appropriate level of description of an animal-environment setting. The object undergoing some style of change is what structures the energy patterns that reach our perceptual systems. Significantly, the object undergoing some style of change is also what is perceived. In the place of the kinds of isolated sound properties that are the focus of traditional psychophysical approaches, the source event and its properties are the focus of the ecological approach.

In the opening example, the bouncing bottle generates a particular acoustic pattern that is specific to the event that gave rise to that pattern: a bouncing bottle. A breaking bottle or a bouncing ball would generate different acoustic structure specific to those different events just as those same events would generate distinct optical structure. In the domain of ecological acoustics, our task is twofold: (1) to document the capabilities of perception on the basis of sound and (2) to identify what in the acoustic structure supports those capabilities. We can expect to exploit successes in other perceptual systems—most notably, vision and touch—in guiding our search for the relevant description. We are seeking *invariants* of energy distributions that are generated lawfully by a particular event. These reliable patternings may well be indifferent to the medium.[2]

Protocol Studies and Perceiving Based on Sound

It is not uncommon for experimenters in a traditional psychoacoustics experiment to ask their listeners to match two tones on some low-level dimension such as loudness or pitch. Of course, loudness is affected by the frequency of the tone as well as its amplitude; pitch is affected by the amplitude as well as the frequency. From these carefully controlled studies, one generates the classic equal-loudness contours that show how perception is faulty in its ability to faithfully register physical properties. Listeners in these experiments adopt what Gaver (1993b) has dubbed a "musical" attitude. Listeners are paying attention to the sounds as such with no concern for the source event that produced them. Given that the source event is typically a tone generator and that the sounds tend to be harmonic with a relatively simple development over time, listeners really have no choice. But, of course, this bears little resemblance to how we are guided by hearing under ordinary circumstance. Everyday sounds tend to be inharmonic, vary complexly over time, and, more than likely, vary along dimensions of pragmatic utility. "Everyday listeners" care about the source of the sound, not the sound itself. They want to know what happened and what it means for them (Gaver, 1993b; Schubert, 1975).

More in keeping with the style of everyday listening are so-called protocol studies. Listeners are presented with a variety of recorded sounds and asked to identify what they hear. They are generally quite successful at identifying footsteps, clapping, hammering, filing, tearing paper, jangling keys, and so on (Gaver, 1988; Vanderveer, 1979). Protocol studies are, of course, only a preliminary step in ecological acoustics because they permit no systematic control of individual events, and they make no demands of metrical precision on the part of the listener (Gaver, 1993b). Nonetheless, two nice theoretical points are apparent. First, it is only when listeners are unable to identify the source of a sound that they resort to reporting its sensory aspects, in effect converting from everyday to musical listening. Second, listeners' confusions tend to be of events that have similar temporal profiles (e.g., hammering and stepping; filing and scratching), pointing at the relevance of higher-order structure.[3] Moreover, even without an avenue for demonstrating metrical precision, some of the distinctions are, in fact, quite subtle (e.g., listeners can distinguish ascending versus descending stairs; Gaver, 1988). All of these suggest that systematic physical analyses are possible and are needed.

Consider mechanical events. The vibration pattern in the air is influenced by the materials involved in the event and the types of interactions those materials undergo (Gaver, 1993a). The type of interaction, whether it is scraping or splashing or slapping, affects the time-varying amplitude and the spectrum of vibration. The tension and elasticity of the materials are restoring forces. The size, shape, and density of material determine its inertia. Both sets of factors determine how quickly an object returns from the deformed state brought about by its interaction with other objects or surfaces. This, in turn, determines the frequency of its vibrations. In brief, sound is structured reliably by interacting materials. That is the rightward sequence of the ecological law statement. The leftward sequence—whether this sound structure informs about those materials and interactions—is the focus of experiments in ecological acoustics. What is the object? What is it doing? Where is it doing it?

Categorical Distinctions

Our choice of the example that begins this chapter was not accidental. The first experiment to be labeled ecological acoustics was an examination of listeners' ability to distinguish breaking and bouncing events (Warren & Verbrugge, 1984). The first part of the study involved categorizing natural sounds. Listeners were presented with tape recordings of various instances of different-sized bottles and jars falling to the ground and either breaking or bouncing. Listeners simply had to identify which event they heard (or whether they could not tell). The success rate was 99%. The next step was to try to identify what in the sound structure might allow the distinction to be made. The breaking event has an initial burst of noise. Even when this burst was eliminated from the recording, however, listeners still achieved 96% accuracy in the identification of breaking. What remained, of course, was the timing of the subsequent pulses. Artificial sound tokens were constructed, therefore, by means of manipulating the timing of these pulses while leaving their spectral characteristics constant.

The construction of synthetic tokens began with pieces of a broken bottle. Four of the largest pieces were dropped individually and recorded as they bounced. In one type of synthetic token, the onsets of every impact of each of the four pieces were synchronized by inserting appropriate amounts of silence between subsequent bounces so as to preserve the single, damped, quasi-periodic pulse train that characterized bouncing.

(Despite the presence of four pieces, the synchrony supported the perception of a single object.) Sound events synthesized on this basis were successfully identified as bouncing with 92% accuracy. Synthetic tokens characterized only by a decline in the amplitude of the impacts without concomitant damping of the temporal pattern (i.e., with the intervals equated between each of the bounces) were not heard as bouncing. In a second type of synthetic token, the onsets of the four pieces were synchronized, but then each followed a different damped, quasi-periodic pulse train. These were identified as breaking with 87% accuracy (whether augmented by a noise burst or not).

The foregoing is an exemplary study in ecological acoustics. Although such a pedigree makes it more likely that a researcher will be interested in source properties, a handful of more traditionally oriented psychoacousticians have contributed to this literature. Two studies of interest concern protocol events that listeners identified successfully: footsteps and clapping.

We already noted one subtlety of footsteps that listeners can pick up on: whether those footsteps are going up the stairs or down. Given the different contributions of heel and toe strikes to those two directions, this success is understandable. A possibly more subtle distinction, the sex of the walker, can also be discerned simply on the basis of the sound of the walker's footsteps (Li, Logan, & Pastore, 1991). Sixteen different walkers, eight of each sex wearing the same style shoe (low, solid synthetic heel), walked at their normal pace on a hardwood stage, taking eight steps directly towards a microphone. Each of these strolls was recorded, and a four-step sequence from the middle was presented to listeners over headphones. The task was a simple categorization as male or female. Males were identified correctly on 69% of the trials; females were identified correctly on 75% of the trials. But there were substantial differences within each gender category. Half of the males and half of the females were identified correctly at least 85% of the time. Two more of the females were identified correctly at least 70% of the time. But the remaining walkers were identified at no better than a chance level. A statistical evaluation of various anthropometric measures indicated that walker height accounted for 70% of the variance in the judgments of gender. The investigators conjectured that height was really standing proxy for height of the center of mass, CM, which is known to differ for males and females. Although they did not determine individuals' CMs, we estimated them from anthropometric standards and, indeed, these estimated CMs account for 76% of the variance in judgments of gender. Spectral analyses suggested that male judgments were more likely when the acoustics were characterized by more energy in the low-frequency range with rapid spectral rising and falling. Conversely, female judgments were more likely when the acoustics were characterized by more energy in a higher frequency range with slow spectral rising and falling.

Clapping is another sound event that could be identified easily in a protocol study, but does it allow listeners to discern anything about the clappers themselves? Twenty different clappers, ten of each sex, were instructed to clap for ten seconds at their normal rate, as they "would normally clap after an average concert or theater performance" (Repp, 1987, p. 1101). Recordings were presented over headphones to listeners with the entire sequence being presented once for familiarization before it was presented a second time for response collection. Because the individuals were all known to one another, the task was to label each bout of clapping as belonging to one of the twenty people, whose names appeared on an alphabetic list. Performance (11% correct) was better than chance (5% correct), but it was not impressive. Self-recognition was considerably better (46% correct) but, of course, that could have benefited from factors over and above the acoustic specification of source properties.[4] Even when rescored simply for whether a clapping bout was classified as having been produced by a male or female (even if the specific identity was wrong), performance was only 54% (compared to a chance level of 50%; Repp, 1987; see also Tousman & Pastore, 1989). Nonetheless, listeners' performances were systematic. A search of acoustic characteristics that encouraged the choice of a male or female label revealed four factors that together accounted for 85% of the variance in judgments. Most important was the contribution from the inter-clap interval: The slower the applause, the more likely it was to be labeled male. Next in importance was the amplitude of the clapping: The louder the clapping the more likely it was to be labeled male. The final factors are both spectral, with the most intuitive interpretation being that low-frequency resonances were considered male. None of these factors actually distinguished male and female clappers, which accounts for the poor performance. But in a different sequence of claps, produced by a single person with different hand configurations (palm-to-palm, finger-to-palm, or something in between), a new group of listeners was successful in judging the configuration of the hands that produced the clapping sequences.

Quantifying the mechanical properties of human effectors—the legs of a walker, the hands of a clapper—that give rise to particular acoustic properties is not straightforward. Moreover, the sample of sound-producers may be idiosyncratic, either in physical dimensions (height and weight combinations, hand sizes and shapes) or personal style of walking or clapping, that may introduce variation that masks the dimension of interest. Finally, judgments of characteristics such as the sex of the sound source may simply reflect general (and often erroneous) sex stereotypes. Mechanical events, in contrast to biological events, allow more control over the source characteristics with a consequent increase in the resolution that can be asked of perceivers. But, in the modern idiom, if you don't ask, they don't tell. We turn next to experiments that ask for modest performance on the part of listeners, rank-ordering sounds without indicating their appreciation of relative differences.

Perceiving Mechanical Events on the Basis of Sound

We know that listeners can appreciate that a collision event did not entail breaking. Is anything else known about the objects in the event? A variety of questions could be asked: What were the materials, how large were the objects, were they solid or hollow? One such investigation focused on percussive sound events—a single impact of a mallet on a hollow receptacle (Freed, 1990). Six mallets that varied in hardness were used to strike four receptacles (cast-aluminum cooking pans) that varied in size. The twenty-four percussive events were recorded in an anechoic chamber and presented over a speaker. The guiding force was explicitly ecological, focusing listeners on physical properties of the objects—the mallets' hardness—rather than abstract properties of the sound—whether it is bright or dull, thin or full (cf. Lichte, 1941). Even with this ecological goal, psychophysical timidity was nonetheless in evidence in the study. The sound-pressure level was roughly equalized across tokens, thereby restricting differences to spectral parameters. Each session began with a demonstration of the hardest mallet hitting first the smallest pan and then the largest pan, followed by the softest mallet hitting the same two pans. Listeners were told that these were instances of a hard and a soft mallet. The entire sequence of sounds, which included four repetitions each of six mallets striking four different receptacles, was then played once to familiarize listeners with the range of sounds. (The untutored informativeness of the sounds was not strictly assessed.)

Perceived mallet hardness was indicated by continuously adjusting the length of a visible line segment on a scale marked from one to nine for softest to hardest. Harder mallets generally elicited higher ratings of hardness and seemed unaffected by the pan that was struck. More germane for the interests of the investigator was the identification of acoustic parameters that would be useful as predictors of timbre. A combination of four spectral parameters characterizing the initial 325 ms of each event accounted for 75% of the variance in hardness ratings. One is essentially a loudness measure (the log of the area under the spectrum). Its slope represents the softening over time. A third reflects the brightness (the mean of the centroid of the spectrum), and the fourth (the time-weighted average of the centroid of the spectrum) represents the darkening over time.[5]

An inevitable implication of source-oriented auditory perception is that information about an event is infused throughout the acoustic signal. Mechanical events involve the movements of masses that are mutually constraining. How an event began influences how it can unfold; how it is unfolding is informative about how it may have started. In principle, therefore, later-coming acoustic structure can influence perception of an earlier event. This was the premise of an investigation of the perception of steepness based on the sound of a ball rolling down a ramp (Fowler, 1990). A steel ball was recorded rolling down five different ramps (10, 20, 30, 40, and 50 deg) onto either a flat track or an upward-sloping track. During such events, the time that the ball spends on the track is determined by the slant of the initial ramp but in opposite ways for the two tracks. On a flat track, a longer duration means that the early part of the event was shallower; on an up-slope, a longer duration means that the early part of the event was steeper (fig. 6.1). The experimental sequences were hybrid stimuli constructed by splicing acoustic signals from the extremes of the track portions onto acoustic signals from each ramp slope. That is, for a given ramp, the track portion was replaced by four options: the flat slope and the upward slope that had accompanied the 10° and the 50° ramps. If the later-coming acoustic structure influences perception of an earlier event, then listeners judging the slant of the ramp ought to be influenced by the track portions they hear and in specific ways. For the up-sloped tracks, the shorter-duration track sounds ought to yield flatter-perceived ramp slopes than longer-duration track sounds. For the flat tracks, shorter track portions ought to yield steeper perceived ramp slopes than longer track portions. In other words, duration per se is not the critical acoustic structure.

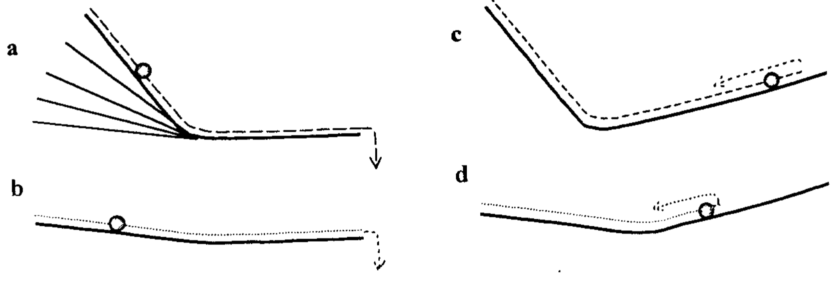


Fig. 6.1. *(a)* Initial ramps of different slopes will influence the length of time a ball bearing spends on a flat track before falling off. *(b)* The time on the track is longer the flatter the initial ramp. *(c)* When the track slopes upward, the ball slows before rolling back down. *(d)* The time on the track is shorter the flatter the initial slope.

Prior to the experimental test, listeners were presented with recordings of just the ramp portions from the 10° and 50° ramps five times in alternation and told which was which. They were instructed that in the test, they would hear the entire rolling event and they were to judge whether the ramp portion was steep or flat. Ten listeners heard the ramps spliced onto the flat tracks and another ten listeners heard the ramps spliced onto the up-slopes. Ramps were generally ordered appropriately in all conditions, even though the track portions did not help this specific discrimination. More importantly, the track portions did produce the expected interaction. Listeners in the flat-track condition heard all ramps as steeper if they had been spliced to a short duration track portion; listeners in the sloped-track condition heard all ramps as steeper if they had been spliced to a long duration track portion. A subsequent experiment, limited to the flat-track condition, addressed the duration issue directly. A track portion from the 10° ramp was cut at regular intervals until it was equal in duration to the 50° portion (it was essentially halved). Ramps appended to this shortened track portion were indeed heard as steeper than those with the original long, 10° flat track but not as steep as those with the original 50° track. In other words, duration is not the only dimension that is informative about ramp slope. Still audible, for example, was the ball's revolution speed.

The upshot of this research is the emphasis on perceiving the source event. Had conditions been limited to the flat tracks, one might infer that listeners are simply influenced by a general auditory cue, durational contrast. The opposite durational pattern from the upward sloping track instead supports the notion that listeners recover physical event properties from the information available in acoustic structure.

Both the mallet hardness and the ramp slant studies show that listeners perceive properties of environmental objects and surfaces with some degree of gradation: mallets are more or less hard, ramps are more or less slanted. However, neither the mallet nor the slope investigations exploited a specific advantage of mechanical events. Metrical precision of perception, not just rank-ordering, can be assessed. That is to say, perceived increments in a dimension of interest can be evaluated relative to physical increments in that dimension. This has been the focus of recent investigations of shape and size perception on the basis of sound.

Toward Metrical Precision in Perceiving Mechanical Events

The in-principle argument for specification is straightforward. To the extent that an event structures sound reliably, that structure ought to specify the source event. But what provides the metric? Characteristic modes of vibration are one possibility. Objects with a high degree of symmetry, for example, have three orthogonal modes (fig. 6.2) that are determined by the physical dimensions (size, shape) and material properties (mass, density, elasticity) of the source. An object set into vibration will conform to one of its characteristic modes despite variation in pitch and timbre (Lakatos et al., 1997). The relative contribution of each mode, however, depends on how an object is struck. This so-called exciter-resonator relationship is often nonlinear (Fletcher & Rossing, 1991). Consequently, one methodological strategy is to keep the strike position constant while varying spatial dimensions of the resonating object.

In one such investigation (Lakatos et al., 1997), the objects were long bars made of steel or wood, suspended lengthwise and struck at the center with a mallet (steel for the steel bars, resin for the wooden bars). The bars were of a fixed length but varied in thickness and width; bar "girth" is what listeners had to discern (the specific dimensions dif-

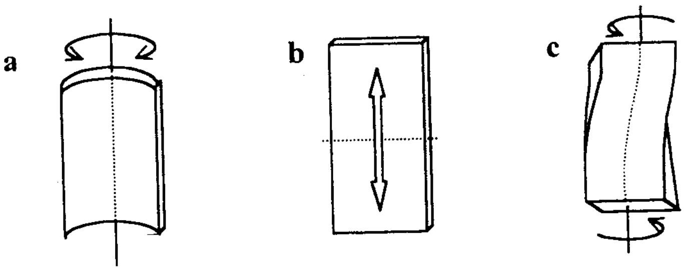


Fig. 6.2. A metal bar tends to vibrate *(a)* transversely, *(b)* longitudinally, and *(c)* torsionally. The modes of vibration of are orthogonal.

fered for the two materials). Sound recordings were made in an anechoic chamber and matched for loudness. Multiple recordings were made so that only clear samples were selected. During the experiment, sounds were presented over a speaker in a sound-isolation booth. A trial consisted of a pair of sounds in succession accompanied by two response alternatives presented visually. These alternatives represented the actual cross-sectional proportions of the bars (scaled to the computer screen) but in both the correct and the incorrect order. Listeners had to choose which order corresponded to the order of the sounds they heard (twenty practice trials without feedback familiarized them with this so-called two-alternative forced-choice, or 2AFC, procedure). Prior to the practice trials, listeners were allowed to strike five sample bars (not from the experiment set) with the appropriate mallet to familiarize themselves with the type of interaction they would be hearing. Finally, they were told to "use any available timbral cues" to make their decision.

Listeners who failed to attain a 75% performance criterion were excluded from the analysis (out of the original sixty listeners, five did not reach criterion for the steel bars, and ten did not reach criterion for the wooden bars). With these exclusions, listeners' performance did indeed vary with differences in the width-to-height ratios and improved the greater those differences were. For metal bars, the more block-like cross sections tended to cluster together as did the more plate-like cross sections. This pattern was not evident for the wooden bars, which listeners found harder to discriminate (perhaps because the signals were so much shorter). Acoustic analyses indicated that matching performance correlated strongly with the frequencies of the vibration modes. In particular, for metal bars, the frequencies of the torsional modes, $F_{T,}$ accounted for 86% of the variance in listener responses. As an alternative, the ratio of the transverse bending modes dependent on width to those dependent on height, $F_W/F_{H,}$ accounted for 88% of the variance. These relationships were less secure for the wooden bars, not only because the variance accounted for was less (58% by $F_{T;}$ 67% by F_W/F_H) but also because those components are sometimes quite weak or absent in those bars. No attempt was made to provide a comprehensive account of performance with both sets of bars (Lakatos et al., 1997). One problem is that wood is an *orthotropic* material (Rossing & Fletcher, 1995); its mechanical properties, unlike metals', tend to vary along three perpendicular axes.

At some level, the preceding illustrates a finer degree of resolution on the part of listeners. They were asked to discriminate the cross-sectional shapes of struck bars whose width-height ratios ranged from .13 to 1.00. Given the 2AFC procedure, however, we only know that listeners could discriminate different-shaped rods and that discrimination was easier the more different the rods were. But we have no appreciation for listeners' accuracy in perceiving the rods. In the absence of the visual matches, could they have indicated that one rod was block-like and another was plate-like? They reached a performance criterion of 75%, but that was after poor listeners were eliminated. Given the precision in the sound recordings—anechoic chambers eliminate reverberation, loudness is equalized, and only clear recordings are used—should we be impressed with listener achievement? How would listeners fare if the sounds came from real objects in an ordinary room? What if listeners were provided with some leeway in responding? Do they actually need all of the practice and restrictions that characterize the research considered so far?

Metrical Precision in Perceiving Struck Plates

We have been building a case, at least implicitly, that by the very nature of the research questions asked in auditory experiments, the science of perception doubts the fine-grained spatial capabilities of hearing. Consider this opinion from Sir Arthur Schuster in 1882, who was illustrating the challenge to be faced by the then-new science of spectroscopy:

> To find out the different tunes sent out by a vibrating system is a problem which may or may not be solvable in certain special cases, but it would baffle the most skillful mathematician to solve the inverse problem and to find out the shape of a bell by means of the sounds which it is capable of sending out. (qtd. in Gordon & Webb, 1996, p. 46)

Mathematicians generally endorse this skepticism. Kac (1966) posed the question explicitly: "Can one hear the shape of a drum?" and it has been answered explicitly, "You can't hear the shape of a drum" (Gordon & Webb, 1996). Isospectral companions—identical spectra produced by two manifolds that differ geometrically—are to blame. But, of course, mathematicians operate in idealized space. Just because isospectral companions are possible need not mean that they are representative or problematic. First, they seem to be the exception (Gordon & Webb, 1996). Second, trying to implement such manifolds with real physical objects would likely introduce differences, however small, in their physical parameters (Kunkler-Peck & Turvey, 2000). What happens if we ask the shape question of perceivers rather than mathematicians?

The drums were, in fact, flat steel plates (circle, square, and triangle) struck by a steel pendulum bob released from a fixed location so that the same amount of energy would always be imparted (Kunkler-Peck & Turvey, 2000; see fig. 6.3a). The plates, which had the same mass and surface area, were simply suspended by fishing line to provide stability with minimal damping and without eliciting vibration in the support structure. A listener sat on one side of a screen that hid the shape that was being struck (about 1 m away on the other side). On each trial, the bob was released from the starting location and caught right after it bounced off the plate. This was repeated three times so that a trial was defined by three strikes. Each of the three objects was presented three times in random order. Listeners had to verbally indicate which plate had been struck, with no prior demonstrations of the sounds and no practice trials. Nonetheless, the correct shape was chosen 58% of the time (where chance performance would have been 33%).

To make the task a little harder, a further experiment included the same three shapes in wood and in Plexiglas as well as steel, again with dimensions chosen to provide the same surface area (mass differed across material but was the same for the three shapes of a given material). Because listeners were asked about the material as well, the verbal

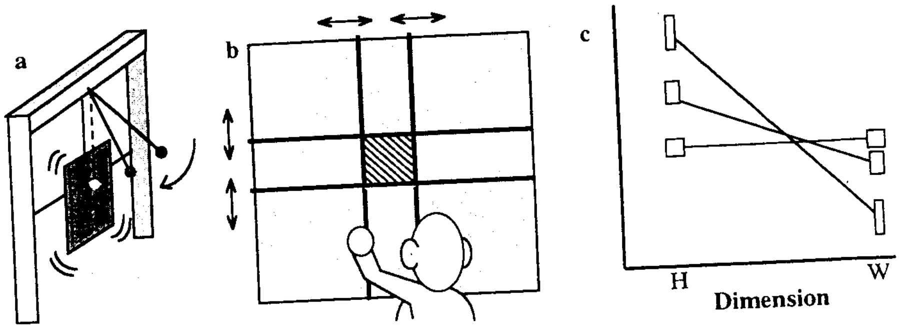


Fig. 6.3. *(a)* A simply suspended plate is struck by a pendulum bob. *(b)* A listener whose view is occluded adjusts dowels to indicate the height or width. *(c)* Proportional shape is distinguished.

response was replaced by having them point at replicas of the nine objects mounted on the occlusion screen. Once again, there were no demonstrations or practice trials. Perceiving material was absolutely straightforward; one of the seven listeners made one misidentification. Shape performance was comparable to the previous experiment. The correct shape was chosen 56% of the time (where chance performance on shape again would have been 33%). Moreover, there were no systematic confusions in the errors. This suggests that circles, squares, and triangles really sounded like circles, squares, and triangles rather than simply being labeled by a strategy that could be consistently wrong.

The shape experiment demonstrates a degree of geometric sensitivity. With real sounds in an ordinary room, listeners identified shape reliably. Their responses were constrained, of course, by the shape categories provided, thereby limiting how wrong they could be. A truly metrical response would provide a stronger test. In order to remain in the realm of perceiving shape, this time the steel plates were all rectangles but of different proportions: a square (48.2 × 48.2 cm), a medium rectangle (38.1 × 61.0 cm), and a long rectangle (25.4 × 91.4 cm). Again, dimensions were chosen so that the plates were equal in mass and surface area. The occlusion screen was augmented with a response apparatus that allowed a listener to provide a visual match for a plate's height independently of its width (fig. 6.3b). The width indicator ranged from 0 to 2.5 m; the height indicator ranged from 0 to 1.5 m. As before, there were no demonstrations or practice, and this time there was no information about the number of objects. Additionally, there was no indication of the sizes of the objects (other than the 2.5 × 1.5 m maximum allowed by the apparatus). On a given trial, the listener was told which dimension to report before the plate was struck. He or she adjusted the appropriate indicator to provide a visual match for the heard height or width. Each of the three rectangles was presented six times (three for width, three for height).

The actual linear dimensions accounted for 98% of the variance in listeners' responses. Although perceived dimensions were underestimates of actual dimensions (ranging from 25.2 cm to 44.5 cm for an actual range from 25.4 cm to 91.4 cm), they were in the approximate range. We refer to this as *definite scaling* rather than as *relative scaling* (Bingham, 1993; Turvey & Carello, 1995) because responses are more than simply ordered, arbitrary magnitudes. Listeners do not use the entire range, nor do they use either extreme of the range. They appear to have a definite impression of size (indeed, the average reliability of their responses was 6%, which compares favorably to the visual impression of size; cf. Norman, Todd, Perotti, & Tittle, 1996). Most impressive is their sensitivity to shape (fig. 6.3c). When they attended to the height of the rectangles, their responses were larger than when they attended to the widths of those same rectangles. For the squares, reports of height were the same as reports of width.

Before addressing the acoustic support for this performance, let's consider one more demand placed on the listeners. The preceding experiment was replicated with the three rectangular shapes cut from wood and from Plexiglas as well as from steel. As before, all plates had the same surface area; plates of the same material had the same mass. As before, listeners positioned the report apparatus to indicate the dimension requested on a given trial. As before, perceived dimensions (23.5 cm to 51.4 cm) were in the approximate range of actual dimensions (25.4 cm to 91.6 cm) with a mean reliability of 5.5%.

The simple support of the plates means that their vibrational dynamics are captured by the two-dimensional wave equation (Rossing & Fletcher, 1995). The frequencies associated with the solutions to that equation are given by

$$f_{mn} = 0.453\, h \sqrt{\frac{E}{\rho(1-\sigma^2)}} \left[\left(\frac{m+1}{L_x}\right)^2 + \left(\frac{n+1}{L_y}\right)^2 \right]$$

where h is the thickness of the plate, E is Young's modulus, ρ is the mass density, σ is Poisson's ratio, m and n are integers indexing the vibratory modes, L_x is width, and L_y is height. Obviously, the modal frequencies are influenced by more than the plate's linear dimensions. The physical parameters constrain how the plate bends (fig. 6.2). This is the hypothesized informational support for perceived dimension. It accounted for 87% of the variance in perceived dimension in the preceding experiment. Note that asking a listener to report height or width is asking that listener to perceive *selectively*. Out of the flux of stimulation, they are to extract just that structure relevant to the requested dimension. Analytically, this selectivity was captured by calculating f_{mn} with $m = 0$ for one dimension and $n = 0$ for the other. While this works quite well as a way of summarizing data, we really don't understand what it would mean for the listener to do the same thing functionally.

Metrical Precision in Perceiving the Lengths of Dropped Rods

The clean strike of a supported object (or a sequence of such strikes) allows sound structure that is simpler than that normally encountered in everyday listening. The experimental rationale is that the scientist needs to know what's in the sound in order to determine what listeners can respond to. A different strategy—inspired by the distinction between everyday listening and musical listening—is to give listeners as much structure as they ordinarily encounter in everyday events with acoustic consequences. Apart from the dinner gong, not many sounds are simple strikes. Objects fall to the floor, they clatter and bounce and roll to a stop, with the sound reverberating in the room. Listeners have access to all of that in ordinary experience so let's give them access to all of that in the experiment. Once we assess how metrically precise listeners can be, then we can worry about how to quantify the available structure.

This was the premise of experiments in which listeners were asked to indicate the lengths of cylindrical rods that fell to the floor (Carello, Anderson, & Kunkler-Peck, 1998). In one experiment, wooden dowels 1.25 cm in diameter were cut in lengths from 30 cm to 120 cm (in 15 cm increments). In a second experiment (with different listeners), the diameters were .32 cm and the lengths ranged from 10 cm to 40 cm (in 5 cm increments). On a trial, a rod was dropped five times in succession. To standardize the drop as much as possible, a rod was balanced at its center of mass on a support 72 cm above a hard linoleum surface. The turn of a handle allowed the rod to fall from a fixed height. Listeners sat at a student desk on the other side of an occlusion screen in front of a response apparatus that allowed them to position a marker anywhere from 0 to 2 m to coincide with how far they could reach with the rod (fig. 6.4a). They were provided no practice and no information about the number or sizes of the rods; they simply listened.

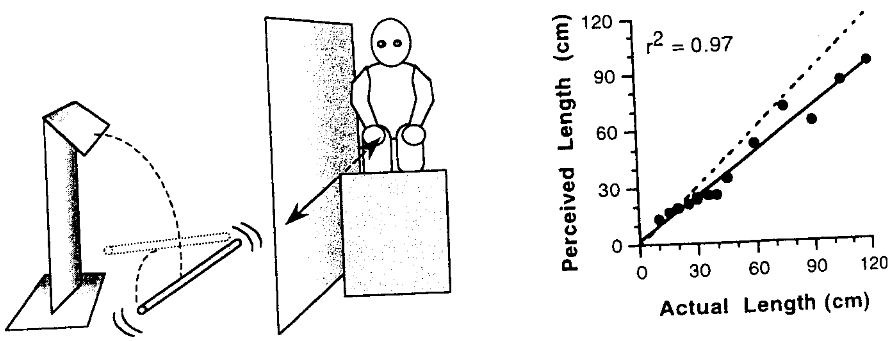


Fig. 6.4. *(a)* A listener adjusted the position of a report board to indicate the length of a rod that had dropped on the other side of an occlusion screen. The rod was released from a support stand by the turn of a small lever. *(b)* Length was perceived with surprising accuracy (perfect performance is indicated by the dashed line) even though listeners were unaware of the number of rods or their possible lengths.

For the large rods, perceived lengths ranged from 24 cm to 95 cm. Actual length accounted for 95% of the variance in perceived length. For the small rods, perceived lengths ranged from 14 to 27 cm. Actual length accounted for 95% of the variance in perceived length. The slopes of the perceived-actual functions were different for the two rod sizes, with the discrimination being sharper for large rods (slope = .78) than for small rods (slope = .44). Nonetheless, the two data sets in combination provided a strong dependence of perceived length on actual length (an overall slope of .77 with 97% of the variance accounted for; fig. 6.4b).[6]

For the acoustic analyses, the sounds were recorded from the listeners' position under the same conditions as described in the foregoing with the exception that a given rod was dropped only three times. One might think that a simple acoustic variable, something that relates to length straightforwardly, ought to account for performance. But neither the duration of the signal, its average amplitude, nor its frequency centroid approached the success of actual length in constraining performance. Average amplitude fared best, accounting for 70% of the overall variance, but it fared less well on the set of large rods (21%). Frequency centroid accounted for 66% of the overall variance but less for the individual sets (59% and 37% for the large and small rods, respectively). Signal duration accommodated the large rods better than the small (65% versus 12% of the variance) but overall, it accounted for only 9% of the variance in listener responses.

Although actual length predicts perceived length successfully, it cannot be the constraining variable. Length is a geometric property, not the kind of mechanical property that can affect acoustic structure. Its success in the preceding experiments was because the rods of different diameters were also in non-overlapping length ranges. When this is no longer so, actual length is not a good predictor of perceived length (Anderson, Carello, & Kunkler-Peck, 1996). Using the same experimental procedure as in the foregoing, listeners judged five rod lengths cut from dowels of three different diameters. Perceived length increased with increases in actual length but the latter accounted for only 39% of the variance in overall responses. Similarly, when rod lengths are fashioned from different materials (steel, Plexiglas, and wood), actual length accounts for only 18% of the variance in perceived length.

Manipulations of length, diameter, and material density are simply different ways of manipulating a higher-order property, an object's mass distribution. This property is quantified through the *inertia tensor,* essentially, the resistances of an object to being rotated in different directions. As noted earlier, inertia influences the vibratory pattern in an impact event and, as such, provides a sensible candidate for the mechanical constraint on perceived length by hearing. Not surprisingly, the rods' inertia tensors account for nearly all of the variance in perceived length in the initial experiments with large and small rods and 91% of the variance in the diameter experiments. Manipulations of material provide an interesting case, however. Material density has consequences for the inertia tensor, to be sure, but it also is related to elasticity and stiffness that, in turn, have consequences for how a body returns to an equilibrium state after being displaced by an external forcing function (such as accompanies dropping to a surface). Indeed, the rods' inertia tensors alone do not predict perceived length of rods of different densities. But when augmented by Young's modulus of elasticity, the variance accounted for is 96%.[7]

As a matter of pedagogical convenience, we have been characterizing the achievements of listeners with respect to perceiving particular properties of objects (e.g., length, shape, slant) or types of interactions (e.g., jangling, bouncing, dropping). In its focus on awareness of the source rather than awareness of the sensations, this emphasis has illustrated a central concern of the ecological approach. But it is incomplete. Perception's raison d'être is guiding activity. Perceivers need to be aware of objects and events because of the consequences for what perceivers-as-actors can do. From the perspective of ecological psychology, perception is not awareness of objects and events per se but awareness of their behavioral relevance. Such opportunities for behavior are what Gibson (1979, 1983) termed *affordances.* Behavioral possibilities are central to a good deal of acoustic research to which we now turn.

Affordances: Perceiving Behavioral Possibilities

We have already introduced the notion of affordance indirectly in the rod-dropping experiments. Listeners were not asked to provide a report of length in units of inches or centimeters. Instead, they were asked to position the visible report surface to coincide with how far they could reach with the rod they had heard. Reconceptualizing absolute geometric properties such as length in terms of activity-relevant properties such as "reach-with-able" is a goal of ecological theory. An affordance is a legitimate occupant of the left slot in the ecological law statement. Such a reconceptualization is ultimately a goal of ecological research as well. But fulfilling that goal requires that we formally fill the right slot in the ecological law statement as it pertains to a particular affordance. What pattern of structure in an ambient energy array is specific to an affordance such as reachable? *Reachable* is a very different kind of thing from *length.* We know how to measure length; we have to discover how to measure reachable.

Ecological psychologists are faced with the very real possibility that the currently available scientific "toolkit" is not up to the job of quantifying information specific to affordances. In the meantime, a good deal of the discovery process involves identifying the variety of affordances that are perceived and the kinds of manipulations that affect them. Oftentimes, the targeted affordances have already been investigated in the visual domain, and the hope is that finding commonalities will give us a foothold on identifying the information. And here's why. We are seeking *invariants* of structured energy distributions, patterns that are always produced by a given affordance regardless of incidental details that can give rise to dramatically different sensations. An object's shape and material surely matter to the intensities and wavelengths of reflected light that reach the eye, but they do not matter to whether you can reach that object. Instead, we seek some invariant of structured light specific to whether something is reachable or not. And, on extension, whether an object structures light or compression waves is also irrelevant to whether you can reach it. We might, therefore, seek a higher-order pattern common to both ambient energy arrays that is lawfully related to what is reachable (see endnote 1).

A key issue is one of prospective control (e.g., Turvey, 1992; E. J. Gibson, 1994). Your actions are organized in such a way that you can effect certain outcomes. For example, picking up a pencil from the desk does not begin with your throwing an arm out in its general direction and hoping for the best. You can see that the pencil is within reach with an outstretched arm or that it will require an additional bend at the hip. Indeed, examinations of this visual ability have shown that sensitivity to the boundary of reach is *body-scaled,* that is, it is the same for tall and short reachers (with correspondingly long and short arms), once the boundary is scaled to the appropriate effector (Carello, Grosofsky, Reichel, Solomon, & Turvey, 1989). But if that same pencil rolled off the desk out of view, could we hear whether it dropped within reach?

Developmental data clearly support the salience of hearing what is reachable in that infants reach more often for a sound source in the dark when that sound source is within reach than when it is out of reach (Clifton, Perris, & Bullinger, 1991; Perris & Clifton, 1988). Formal auditory reaching experiments were modeled after those in vision. Listeners were selected to be tall or short. The sound-emitting target, a kind of rattle, was placed at different distances from the listeners (fig. 6.5a). When the reach was to be with an arm outstretched from the shoulder, distances ranged from 38 cm to 110 cm, in 8 cm increments. When the reach was to be with the outstretched arm augmented by a bend at the hip, distances ranged from 75 cm to 150 cm, also in 8 cm increments. There were three repetitions of each distance within each type of reach, and each block began with ten practice trials. During a trial, listeners simply judged whether the rattle was within reach. Actual maximum reaches in the two conditions were obtained only after all judgments had been collected. In all respects, the results rivaled or exceeded those from vision. Whereas visual judgments tended to be overestimates (Carello et al., 1989), auditory judgments of maximum reach did not differ from actual maximum reach

(Rosenblum, Wuestefeld, & Anderson, 1996). As would be expected, tall reachers had a farther reaching boundary than short reachers (fig. 6.5b), but these boundaries were the same when scaled intrinsically (fig. 6.5c).

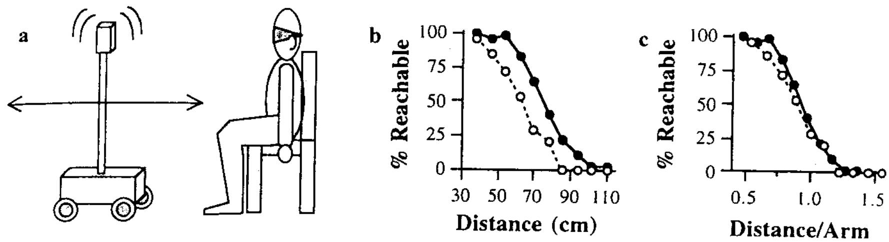


Fig. 6.5. *(a)* A blindfolded listener judges whether the rattle is within reach. *(b)* Short listeners (open circles) cannot reach as far as tall listeners (filled circles). *(c)* When target distance is scaled by the appropriate effector, judgments by tall and short listeners do not differ.

Attempts to characterize the informational support for auditory perception of reachability implicate a higher-order combination of sound intensity and the ratio of direct-to-indirect sound reaching the listener (Wightman & Jenison, 1995). But these simply address distance-relevant structure. The key to perceiving what is reachable must address the body-scaling. How do I know that rattle is reachable by me? One intriguing conjecture received modest experimental support. Judgments of what is reachable were less consistent (though no less accurate) when acoustic structure was restricted to one ear (with the other ear being substantially muffled). Given the allometric relationship between interaural distance (i.e., cranium size) and arm length (Snyder et al., 1974), binaural superiority suggests the possibility of an intrinsic metric in acoustic structure that, in effect, scales what is reachable for the listener (Rosenblum et al., 1996). An analogous intrinsic metric in visual structure, interocular distance, has been promoted as serving a similar function for optically specified reaches in the praying mantis (Michaels, Prindle, & Turvey, 1985) and distance perception, in general, for humans (Michaels, 1986).

Just as questions of perceiving what is reachable ecologize the problem of distance perception, other issues of space perception can be made similarly functional. Consider the question of localizing a sound source, one of the few routinely source-oriented classical questions. In classical hands, it is simply a question of the distance and direction to a sound-emitting object. But a sound-source does not sit in isolation. It is found, for example, among surfaces that reflect its sound.[8] Its relation to these other surfaces may have consequences for behavior. For affordance-minded researchers, one consequence that has been addressed experimentally is whether the gap between the sound source and a vertical surface would allow the listener-as-walker to pass through. A recording of a complex sound, the assembly call of a mallard duck, was played through a loudspeaker that faced a side wall in a long room. The speaker, which was at ear-height, was placed at twelve distances ranging from 20 cm to 75 cm from the wall. The listener stood with eyes closed, one shoulder next to the wall 2 m from the aperture and, on a given trial, judged whether he or she could walk through the gap between the sound and the wall. For all eight listeners, narrow gaps were heard as not allowing passage, wide gaps were heard as allowing passage (fig. 6.6a) and, as is standard in these kinds of experiments, variability in judgments was greatest around the transition from passable to not passable (Russell & Turvey, 1999; cf. Fitzpatrick, Carello, Schmidt, & Corey, 1994). This transition occurred at 1.11 times the listeners' shoulder width, which compares favorably with the value of 1.16 obtained in similar work with visual apertures (Warren & Whang, 1987).

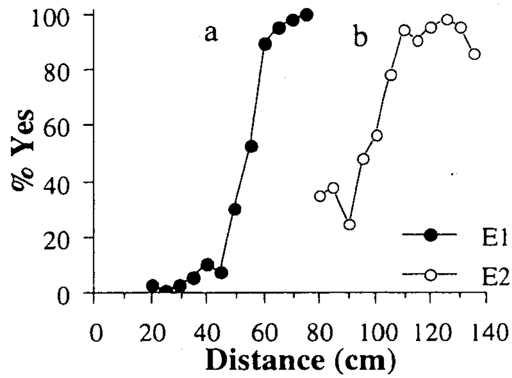


Fig. 6.6. *(a)* Judgments of whether an aperture could be walked through are categorical. *(b)* When the sound source and listener are shifted so that all gaps are passable, the percentage of passable responses increases although the pattern is still somewhat categorical.

Because the right shoulder was close to the wall, listeners may have been aware of the wall's location not by reflected sound but by touching the wall with that shoulder during ordinary postural sway. Judging what allowed passage, then, would simply be a matter of locating the sound source to the left of the left shoulder, perhaps constrained by a certain magnitude of azimuth. To address this possibility, the sound source and listener were positioned 60 cm away from the wall. The relationship between the speaker and the sound source remained the same, but all apertures were, in fact, passable. Judgments based on body-scaled information about gap size should have been 100% passable; judgments based on the location of the speaker relative to the left shoulder should simply have shifted the category boundary rightward. Figure 6.6b shows that performance by a new group of eight listeners was reasonably categorical with the boundary shifted to the right. However, listeners were much more likely to judge gaps as passable in experiment 2 (70% of the time compared to 40% in experiment 1), suggesting some body-scaling. The remaining categorical pattern may reflect a methodological quirk: Listeners may have found it difficult to say "yes" on every trial. Clearly, unlike experiment 1, there were no apertures that were judged not passable all of the time.

A more straightforward test of the azimuthal hypothesis was provided by locating the aperture at three distances from the listener. If a gap is judged passable when the sound source exceeds a certain magnitude of azimuth, then closer gaps ought to be judged more passable than farther gaps. However, performance was equivalent with the aperture at 1 m and at 2 m; it deteriorated completely with the aperture at 3 m. Minimally, this result suggests that listeners were not using the azimuthal direction of the sound source. Interestingly, listeners who had seen the spatial layout of the room before beginning the experiment produced a cleaner category boundary than those who were led into the room with their eyes already closed. That is, they saw the size of the room and where they would be standing (but not where the speaker would be). This suggests that the scaling of perception on the basis of sound is very abstract indeed.

Although obviously ecological in context, these kinds of affordance experiments may encourage participants to be too analytic (Heft, 1993). If listeners dwell on what the experimenter wants them to do, they may try to elect a conscious strategy to satisfy those demands (recall the problem of stereotypes in perceiving the sex of a walker or clapper). The analytic attitude, in some sense, undermines their naturalness in detecting the appropriate information. When participants are hurried or less focused on explicit judgments they are, in fact, more accurate (Heft, 1993). Making the perceptual task part of an on-going activity would be the ideal situation from a theoretical as well as methodological perspective. This is easier to achieve in the realm of interceptive behaviors, to which we now turn.

Acoustic Information about Impending Collisions

Perceiving *time to contact* in the visual domain has been a topic of much investigation (for reviews, see Lee, 1980, 1990; Tresilian, 1993). The basic premise is that, given a collision course between a perceiver and a surface, the perceiver needs to see when the surface will be reached before the collision actually occurs. This is so whether the perceiver is moving towards a surface or a surface is moving towards the perceiver. A single event illustrates both of these. A base runner-as-perceiver churns towards the catcher-as-upcoming surface; the catcher-as-perceiver prepares to make the tag on that runner-as-upcoming-surface. In the simplest case of rectilinear forward motion with constant velocity, the perceiver who would control that collision prospectively needs to know if current conditions persist, when will contact occur?

The optical variable that specifies time to contact (under conditions of constant velocity and rectilinear motion) is a quantity, termed τ (Lee, 1976, 1980), which is given by the inverse of the relative rate of dilation of an optical angle. In the preceding example, an optical angle defined with the catcher as its base and the runner's point of observation as its apex expands as the runner gets closer to the catcher (fig. 6.7a). That angle increases gradually while the runner is still far away but virtually explodes when contact is imminent (fig. 6.7b). Lee (1990) has argued that τ is quite general and applies to any time-varying array variable. This includes other aspects of optical structure (e.g., the relative rate of constriction of an optical gap might specify contact between two objects; Bootsma & Oudejans, 1993) as well as any of a variety of acoustic-array variables that change over time as a sound source nears a listener. For example, as a sound source approaches from the right, its loudness increases, its pitch rises in a pattern known as the Doppler shift, and the onset of sound at the right ear precedes the onset at the left ear.

The amplitude changes are most like the traditional understanding of optical τ. Indeed, a formal derivation of acoustic τ for rectilinear approach verifies an intensity-based structure specific to time to contact (Shaw, McGowan, & Turvey, 1991):

$$\tau_{acoustic} = 2I/\mathrm{d}I\mathrm{d}t$$

where I is given by the inverse-square law, and velocity is constant. Plots of intensity changes during approach to a sound source look very much like figure 6.7b. A variation on this theme has also been derived for pass-by: the negative of the intensity relative to the time derivative of intensity (Erwin, 1995).

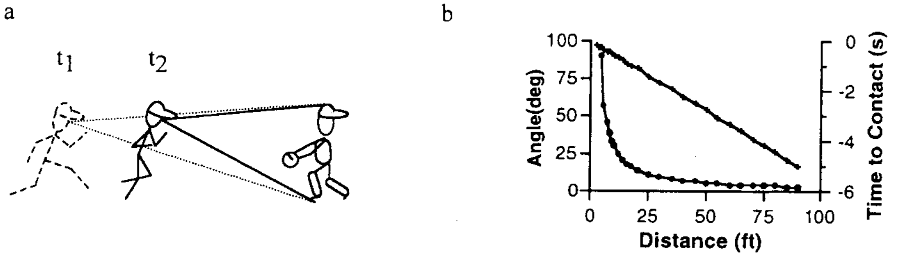


Fig. 6.7. *(a)* As a base runner approaches the catcher, the optical angle subtended by the catcher expands. *(b)* The expansion (circles) changes most dramatically right before contact. For constant velocity, time to contact changes at a constant rate (pluses).

This means that acoustic structure specific to time to contact exists. But can listeners use it? Some listeners, at least, can perceive time to contact by ear as well as observers judge it by eye. Blind adults listened to the sound tracks of filmed events that sighted adults viewed: approach trajectories (either direct or near-misses less than 5° from midline) of twelve vehicles and two people filmed at varying speeds (Schiff & Oldak, 1990). Vehicle sounds included engine and tire noises; people sounds were primarily continuous talking. Film clips lasted from 4 sec to 6 sec after which the approaching sound source "vanished" 1.5 sec to 6.5 sec before it would have reached the plane of the participant's shoulders. Participants pressed a key to indicate when the vehicle or walker would have reached them had it continued at the same speed. Six congenitally or early-blind adult listeners performed at least as well on the audio-alone sequences as did twenty sighted observers on the video-alone portions (the small sample size of blind participants precluded a statistical evaluation). Performance by both groups deteriorated for long times-to-arrival.

Two more groups of sighted observers (twenty in each condition) listened to the audio alone or were presented with the complete audio + video sequences. During those first 3 to 4 sec, their performance did not differ from their video-alone counterparts, nor did the combination of audio and video improve performance (and all were underestimates). These groups also experienced more difficulty for long times-to-arrival, with the deterioration being especially dramatic for the audio-alone, sighted listeners. Not surprisingly, kinematic analyses of acoustic structure (and, apparently, optical structure as well) for events in this velocity range indicate little differentiation in intensity, Doppler, or inter-aural patterns beyond 3 sec (Wightman & Jenison, 1995). In other words, listeners and observers have difficulty in a range where the available structure is not informative. It could be argued that this range is beyond the need for fine-grained prospective control (e.g., Shaw et al., 1991). Alternatively, one could take the perspective that the acoustical structure in question, while not appropriate for the timing of interceptive behavior, is appropriate for orienting the perceiver toward the approaching object (Guski, 1992). When appropriately oriented, the optical τ can then be adaptively exploited. The preceding orienting hypothesis highlights that audition and vision, like all perceptual systems, are typically used together. In the setting under discussion, vision and audition coordinate in localizing an object and determining time to arrival (Guski, 1992).

The preceding results suggest that when contact is less than 4 sec away, acoustic structure is equivalent to optical structure in specifying time to contact. A finer scaling might be provided, however, by an auditory analog of a standard visual experiment that asks observers to decide which of two approaching objects would hit them first. When the objects are computer simulations of approaching squares, viewers achieve a 75% success criterion as long as the arrival times of the two objects differ by at least 50 ms (Todd, 1981). When the simulations are of two sounds approaching a listener, one from the left and one from the right, preliminary data indicate that the 75% criterion is achieved as long as the arrival time differs by at least 300 ms (Wightman & Jenison, 1995).

The kinematic analyses by Wightman and Jenison (1995) show less dramatic changes for a transverse trajectory compared to oblique trajectories. We might expect listeners to have even more difficulties in such situations. The contribution of intensity, Doppler, and interaural patterns to perceiving pass-by on a transverse trajectory has been examined experimentally with a simulated European ambulance siren presented over earphones (Rosenblum, Carello, & Pastore, 1987). These variables change coherently in naturally occurring events. In simulations, however, they can be provided singly or jointly and, if the latter, they can be made consistent or be put in competition. Listeners were to indicate, by pressing a key, when the simulated siren passed right in front of them. Each type of change on its own was sufficient to support perception of the moment of pass-by, but listeners were most accurate when all three types of change were available and consistent. When placed in competition, that is, when the three variables simulated different times of passage, listeners were biased in favor of amplitude change. Doppler shift fared least well in competition, perhaps because this variable is least general—it is not detectable for sounds that lack pitch change (clicks, squeaks, and slow velocities).

Unlike the rectilinear approaches, the transverse events included the moment of pass-by. The importance of this structure has been assessed directly (Rosenblum, Wuestefeld, & Saldaña, 1993). Recordings of an approaching car at two speeds, 15 mph and 25 mph, were divided into thirds, and combinations of these thirds were presented to six listeners who were provided with practice trials but no feedback. Judgments of time to arrival were as accurate when the acoustic signal contained the actual time of passage as when it did not, showing the importance of acoustic structure *before* the moment of arrival. However, systematic "occlusion" or removal of portions of the acoustic signal interferes with listeners' ability to judge time to arrival (Rosenblum, 1993). Thus, if the time-varying aspect of the acoustic signal provides information about time to arrival, performance declines as this information is degraded or impoverished (Wuestefeld & Rosenblum, 1996).

A direct evaluation of one of the derived quantities, the intensity-based τ for pass-by, found mixed results (Erwin, 1995). Simulated pass-by events used intensity modulated sine waves (1000 Hz) to vary speed while keeping offset distance from the listener constant or to vary offset distance while keeping speed constant. Events were truncated before the object passed in front of the listener who pressed a key to indicate when that pass-by would have occurred. The constant error of those judgments varied as a function of the moment of inflection of the τ-functions but with opposite signs for the two manipulations.

What do these results tell us about the information for imminent contact? In many treatments of similar phenomena, each of the acoustic variables and portions of the signal might be labeled a "source" of information (e.g., Cutting, 1986), and the event would be said to be multiply specified. To our way of thinking, however, there is only one source, and that is the moving, sound-emitting object. Amplitude changes, Doppler shift, and interaural differences are simply variables that accompany that event. They probably combine in a nonlinear fashion to form a *higher order variable* that is specific to time to arrival (cf. Michaels & Carello, 1981).

Controlling collisions is actually a rather general phenomenon (Kugler, Turvey, Carello, & Shaw, 1985). An obvious example includes braking before crashing. Less obvious examples include muscle activation preparatory to landing after jumping from a height (Sidaway, McNitt-Gray, & Davis, 1989) or making postural adjustments to stay within the region of reversibility for balance (Carello, Turvey, & Kugler, 1985; Riccio, 1993). Guiding the filling of a vessel to the brim with liquid also qualifies as a controlled collision, and this one has been examined acoustically (Cabe & Pittenger, 2000). For a closed cylindrical tube, the fundamental resonant frequency f is influenced by the height of the air column in the tube, the radius of the tube, and the speed of sound in air. As the tube is filled, the column of air shortens, and f increases, but as the tube is drained, the column of air lengthens, and f decreases. If liquid is released at the same rate it is introduced, f should not change. Acoustic structure, therefore, distinguishes these three events. Monophonic tape recordings of three filling events, three emptying events, and three maintenance events were created with water flowing from a spigot into a plastic tube (30 cm long with a radius of 5 cm). Randomized blocks of the nine events were presented nine times to nine listeners over a loudspeaker without practice or feedback. Given three possible events, chance performance was 33%. Listeners exceeded this in all cases: 67% correct for filling, 87% for emptying, and 67% for maintaining.

Subsequent experiments had more of the flavor of visual experiments in that the listeners actively controlled the flow of water through the spigot and arrested the stream to effect one of two outcomes: full to the brim or to preferred drinking level. Their level of accuracy was established with full information—holding a 17.5 cm long tube while watching and hearing it fill. For audition alone, their eyes were closed, and they did not hold the tube. Ten listeners in each information condition were instructed to turn the water on and off only once during each of 30 trials. The tubes were filled to 96% capacity under full information and to 88% for auditory information alone. For preferred drinking level, they were filled to 86% and 70%, respectively. Although the particular levels achieved under auditory control differed from the particular levels achieved under full-information control, auditory control nonetheless allowed listeners to distinguish the two levels of fullness.

More variation in the acoustic signal was introduced by the use of three different vessel volumes and two different flow rates (with ten trials of each combination). Ten blindfolded listeners were instructed to fill each vessel to the brim with one opening and one closing of the spigot. They were not told of the variations in vessel size or flow rate. Although there was a tendency to underfill vessels (at the fast rate, small vessels were filled to 95% capacity, medium to 86%, and large to 77%; at the slow rate, small vessels were filled to 94% capacity, medium to 85%, and large to 74%), generally the larger the vessel, the higher the fill level. Moreover, ten blind listeners who performed the same task produced the same pattern of results. Finally, a counterpart to the truncated approach and pass-by events described earlier (i.e., indicate when the object would have hit you had it continued) required listeners to anticipate the end point of the filling event in the absence of continuing acoustic structure. Nine listeners heard filling events controlled by the experimenter. Three flow rates were used to fill one vessel to one-quarter, one-half, or three-quarters full. Once the spigot was closed, listeners waited to respond until they thought the vessel would have been full had that flow rate continued. Estimated time to fill tracked actual time to fill very well (with 80–90% of the variance being accounted for).

The Problem of Synthesis

The Warren and Verbrugge (1984) study provides a model for doing ecological acoustics research: Establish the ability, identify candidate sound variables, and produce synthetic events on this basis. The process of synthesis is a tricky one, however. It can be motivated by the structure of natural events, or it can be motivated by the idealized structure expected from the physical equations for the events. One problem with the latter is that the events are simple and idealized, with putatively messy structure eliminated, quite likely at the cost of informativeness. Our survey of experimental investigations of auditory source perception strongly suggests that listeners are good when they have a lot of acoustical structure, even if that acoustical structure is not readily quantified by the scientists. In contrast, it seems to be the case that when events are constrained so that the scientist has a better understanding of the sound structure, perceiving the source becomes harder for the listener.

The issues are readily illustrated by the methods and arguments presented in an investigation of auditory discrimination of the material composition of struck, clamped bars (Lutfi & Oh, 1997). The sounds were synthesized according to principles of theoretical acoustics, with material composition as the only difference among bars. Each material was uniquely identified by nominal values of frequency, amplitude, and decay (which were chosen to be in a range that typically allowed discrimination performance in the range of 70–85% correct). The sounds were presented over headphones in a sound attenuation chamber. Their six listeners were musically trained, with extensive practice and feedback in each condition (e.g., they were told that the sounds would differ in pitch, loudness, and decay). The task was a 2AFC: which tone was the iron bar (where the alternatives were silver, steel, or copper) or which tone was the glass bar (where the alternatives were crystal, quartz, or aluminum). There were four thousand trials per listener for the iron bar as target and four thousand trials per listener with the glass bar as target. The basic result was that all listeners depended excessively on frequency, with a reduction in performance efficiency that sometimes approached 80%.

The investigators, Robert Lufti and Eunmi Oh, characterized their setting as a "best case scenario" (p. 3647)—they had made the task for the listener as straightforward as possible—yet performance was less than ideal. Consequently, they suggested that the "optimistic view" that sources are perceived on the basis of available information (Gibson, 1966/1983; Fowler, 1990) was not supported because listeners did not "optimize decisions based on appropriate combinations of frequency, amplitude, and decay" (Lufti & Oh, 1997, p. 3655). Instead, listeners seemed to make inferences based on only a single cue (cf. Wildes & Richards, 1988). While Lufti and Oh acknowledged the constraints on their experimental setting, these investigators thought that inexperienced listeners and multiple varying sound sources ought to make the task even harder. As already discussed, this expectation is belied by recent findings: Listeners exposed to actual impact events that included variation of shape were nearly perfect in the identification of material (Kunkler-Peck & Turvey, 2000). Far from causing listeners difficulties, the added natural variation seems to have made the relevant time-varying structure more readily apparent.

Synthesizing events is not an inherently bad thing for ecologically oriented psychologists. The issue is simply one of what is being synthesized. Gibson himself put the issue in terms of synthesizing information (see also Gaver, 1993a). And this is the crucial point for an application to film sound. Because an object's physical parameters and spatial dimensions determine the acoustic structure it generates when mechanically disturbed, such structure is potentially informative about object properties and therefore about the object's (or event's) affordances. The Foleyed effects in a film's soundtrack (e.g., alternately slapping coconut halves to simulate a galloping horse) are really more in the spirit of synthesizing information than the technically more sophisticated simulations permitted by theoretical acoustics. The properties shared by a horse's hooves hitting the ground during a gallop and the sound made by coconut halves slapping against a hard surface are just that—shared properties—that in conjunction with the *image* of a galloping horse call attention to those properties of a galloping horse that the filmmaker wishes to emphasize.[9]

In those instances when filmmakers are dealing with nonexistent characters or situations, as is often the case in science fiction or fantasy films, the corresponding sound must be entirely synthetic. It must of necessity be synthetic, but if it is to be convincing, it must not be arbitrary. The sound must share with the character/event being portrayed basic properties that convey the affordances appropriate to the filmic scene. The sounds emitted by the dinosaurs in *Jurassic Park,* for example, are sounds that communicate information about the size and disposition of the creatures, not their specific height or length or weight.

Film sound is designed to enhance the particular properties of an object or event that is of central concern to the filmmaker at that point in the narrative, to facilitate the pickup of the information the filmmaker wishes to convey about the characters or events on the screen. Like the musical scoring that accompanies lurking madmen or sea battles or madcap exploits, it points to the abstractness of the information that is unlikely to be found in nominal values of physical parameters of frequency, amplitude, and decay.

Notes

Preparation of this manuscript was supported by grants from the National Science Foundation (BCS00-04097 awarded to M. T. Turvey and Claudia Carello) and the National Institutes of Health (HD 01994 awarded to Haskins Laboratories).

1. Theories are consistent with these frameworks but aim at a more specific level, such as a theory of space perception or a theory of movement perception.

2. We are endorsing the notion of amodal invariants: The same abstract invariant ought to characterize all of the energy distributions structured by a given event. Some ecological psychologists consider invariants to be mode specific, with different invariants in each type of energy distribution specific to the same event. Others argue for intermodal invariants, a position that an event is specified by one emergent invariant that is defined over all energy distributions. At the present time, the last of these is very difficult to assess experimentally.

3. Gaver (1988, 1993b) identified three broad categories of interactions—vibrating solids, aerodynamic sounds, and liquid sounds. In his protocol studies, confusion remained within category. As he noted, they need not if the temporal patterns of the events are similar (e.g., as illustrated in the use of tubes, pegs, and beads to simulate the sound of rain in the so-called rain sticks of Mexico).

4. When the record of a movement is a visual trace (e.g., a trajectory of a figure eight on a computer screen), people are also better at identifying their own movements than identifying the movements of another (Prinz, 1997). Some researchers would like to understand that superiority with respect to a common code for perception and action.

5. Because these parameters are not invariant over resonators, they are unlikely candidates as the definitive constraints on perceived mallet hardness, given that listeners' mallet ratings were indifferent to what was struck (Lakatos, McAdams, & Caussé, 1997).

6. With somewhat less success, Gaver (1988) asked listeners to indicate the lengths of bars, made of iron or oak, that were struck once at their centers while sitting on a carpeted floor. The sounds were recorded and played to listeners over headphones. Not all listeners discriminated length, and they were differentially affected by material. The fidelity of the recordings may be at fault, along with the less informative impact events.

7. The relevance of the inertia tensor to the perception of length has long been appreciated for dynamic touch (cf. Turvey, 1996), work which inspired the dropped-rods experiments. And the inspiration goes both ways. The inertia tensor completely constrains perceived length when wielded objects vary in diameter and density (Fitzpatrick, Carello, & Turvey, 1994), but elasticity is a constraint when the objects are nonrigid (GrandPré & Carello, 2001).

8. Our discussion is limited to situations involving direct sounds. A considerable literature addresses the perception of surface layout on the basis of *echolocation,* in which the listeners use reflections of self-produced sounds (e.g., footfalls, cane taps, spoken syllables). An ecological treatment can be found in Stoffregen & Pittenger, 1995).

9. Similarly, when scientists investigating dynamic touch manipulate the inertial characteristics of an experimental object, they are synthesizing information about its movableness not about its mass or length or diameter (Shockley, Grocki, Carello, & Turvey, 2001; Turvey, Shockley, & Carello, 1999).

References

Anderson, K., Carello, C., & Kunkler-Peck, A. J. (1996, May). *Rotational inertia constrains perception of object length by sound.* Poster presented at the North American Society for Ecological Psychology, Hartford, CT.

Bingham, G. P. (1993). Perceiving the size of trees: Forms as information about scale. *Journal of Experimental Psychology: Human Perception & Performance, 19,* 1139–161.

Bootsma, R. J., & Oudejans, R. R. D. (1993). Visual information about time-to-collision between two objects. *Journal of Experimental Psychology: Human Perception & Performance, 19,* 1041–52.

Cabe, P., & Pittenger, J. B. (2000). Human sensitivity to acoustic information from vessel filling. *Journal of Experimental Psychology: Human Perception & Performance, 26,* 313–24.

Carello, C., Anderson, K. L., & Peck, A. (1998). Perception of object length by sound. *Psychological Science, 9,* 211–14.

Carello, C., Grosofsky, A., Reichel, F. D., Solomon, H. Y., & Turvey, M. T. (1989). Visually perceiving what is reachable. *Ecological Psychology, 1,* 27–54.

Carello, C., Turvey, M. T., & Kugler, P. N. (1985). The informational support for upright stance. *Behavioral and Brain Sciences, 8,* 151–52.

Clifton, R., Perris, E., & Bullinger, A. (1991). Infants' perception of auditory space. *Developmental Psychology, 27,* 187–97.

Cutting, J. E. (1986). *Perception with an eye for motion.* Cambridge, MA: MIT Press.

Erwin, T. (1995). Prospective acoustic information for object pass by. In B. G. Bardy, R. J. Bootsma, & Y. Guiard (Eds.), *Studies in perception and action III* (pp. 325–28). Mahwah, NJ: Erlbaum.

Fitzpatrick, P., Carello, C., Schmidt, R. C., & Corey, D. (1994). Haptic and visual perception of an affordance for upright posture. *Ecological Psychology, 6,* 265–87.

Fitzpatrick, P., Carello, C., & Turvey, M. T. (1994). Eigenvalues of the inertia tensor and exteroception by the "muscular sense." *Neuroscience, 60,* 551–68.

Fletcher, N. H., & Rossing, T. D. (1991). *The physics of musical instruments.* New York: Springer-Verlag.

Fowler, C. A. (1990). Sound-producing sources as object of perception: Rate normalization and nonspeech perception. *Journal of the Acoustical Society of America, 88,* 1236–49.

Freed, D. J. (1990). Auditory correlates of perceived mallet hardness for a set of recorded percussive events. *Journal of the Acoustical Society of America, 87,* 311–22.

Gaver, W. W. (1988). *Everyday listening and auditory icons.* Unpublished doctoral dissertation, University of California, San Diego.

———. (1993a). How do we hear in the world? Explorations in ecological acoustics. *Ecological Psychology, 5,* 285–313.

———. (1993b). What in the world do we hear? An ecological approach to auditory event perception. *Ecological Psychology, 5,* 1–29.

Gibson, E. J. (1994). Has psychology a future? *Psychological Science, 5,* 69–76.

Gibson, J. J. (1979). *The ecological approach to visual perception.* Boston: Houghton Mifflin.

———. (1983). *The senses considered as perceptual systems.* 1966. Prospect Heights, IL: Wavelan.

Gordon, C., & Webb, D. (1996). You can't hear the shape of a drum. *American Scientist, 84,* 46–55.

GrandPré, D., & Carello, C. (2001). Elasticity and length perception by dynamic touch. In G. Burton & R. C. Schmidt (Eds.), *Studies in perception and action, Vol. VI* (pp. 21–24). Mahwah, NJ: Erlbaum.

Guski, R. (1992). Acoustic tau: an easy analog to visual tau? *Ecological Psychology, 4,* 189–97.

Heft, H. (1993). A methodological note on overestimates of reaching distance: Distinguishing between perceptual and analytic judgments. *Ecological Psychology, 5,* 255–71.

Jenkins, J. J. (1985). Acoustic information for places, objects, and events. In W. H. Warren & R. E. Shaw (Eds.), *Persistence and change: Proceedings of the first international conference on event perception* (pp. 115–38). Hillsdale, NJ: Erlbaum.

Kac, M. (1966). Can one hear the shape of a drum? *Mathematics Monthly, 13,* 21–23.

Kugler, P. N., Turvey, M. T., Carello, C., & Shaw, R. E. (1985). The physics of controlled collisions: A reverie about locomotion. In W. H. Warren & Shaw (Eds.), *Persistence and Change: Proceedings of the First International Conference on Event Perception* (pp. 195–222). Hillsdale, NJ: Erlbaum.

Kunkler-Peck, A., & Turvey, M. T. (2000). Hearing shape. *Journal of Experimental Psychology: Human Perception & Performance, 1,* 279–94.

Lakatos, S., McAdams, S., & Caussé, R. (1997). The representation of auditory source characteristics: Simple geometric form. *Perception & Psychophysics, 59,* 1180–90.

Lee, D. N. (1976). A theory of visual control of braking based on information about time-to-collision. *Perception, 5,* 437–59.

———. (1980). Visuo-motor coordination in space-time. In G. Stelmach & J. Requin (Eds.), *Tutorials in motor behavior* (pp. 281–95). Amsterdam: North Holland Press.

———. (1990). Getting around with light or sound. In R. Warren & A. Wertheim, (Eds.). *Perception and control of self-motion* (pp. 487–505). Hillsdale, NJ: Erlbaum.

Li, X., Logan, R. J., & Pastore, R. E. (1991). Perception of acoustic source characteristics: Walking sounds. *Journal of the Acoustical Society of America, 90,* 3036–49.

Lichte, W. H. (1941). Attributes of complex tones. *Journal of Experimental Psychology, 28,* 455–80.

Lutfi, R. A., & Oh, E. (1997). Auditory discrimination of material changes in a struck-clamped bar. *Journal of the Acoustical Society of America, 102,* 3647–56.

Michaels, C. F. (1986). An ecological analysis of binocular vision. *Psychological Research, 48,* 1–22.

Michaels, C. F., & Carello, C. (1981). *Direct perception.* New York: Prentice Hall.

Michaels, C. F., Prindle, S., & Turvey, M. T. (1985). A note on the natural basis of action categories: The catching distance of mantids. *Journal of Motor Behavior,* 17, 255–64.

Norman, J. F., Todd, J. T., Perotti, V. J., & Tittle, J. S. (1996). The visual perception of three-dimensional length. *Journal of Experimental Psychology: Human Perception & Performance, 22,* 173–86.

Perris, E., & Clifton, R. K. (1988). Reaching in the dark toward sound as a measure of auditory localization in infants. *Infant Behavior and Development, 11,* 473–91.

Prinz, W. (1997). Perception-action planning. *European Journal of Cognitive Psychology, 9,* 129–54.

Repp, B. H. (1987). The sound of two hands clapping: An exploratory study. *Journal of the Acoustical Society of America, 81,* 1100–10.

Riccio, G. E. (1993). Information in movement variability about the qualitative dynamics of posture and orientation. In K. M. Newell & D. M. Corcos (Eds.), *Variability and motor control* (pp. 317–58). Champaign, IL: Human Kinetics.

Rosenblum, L. D. (1993). Acoustical information for controlled collisions. In A. Schick (Ed.), *Contributions to psychological acoustics* (pp. 303–22). Oldenberg, Germany: Bibliotheks.

Rosenblum, L. D., Carello, C., & Pastore, R. E. (1987). Relative effectiveness of three stimulus variables for locating a moving sound source. *Perception, 16,* 175–86.

Rosenblum, L. D., Wuestefeld, A., & Anderson, K. (1996). Auditory reachability: an affordance approach to the perception of distance. *Ecological Psychology, 8,* 1–24.

Rosenblum, L. D., Wuestefeld, A., & Saldaña, H. (1993). Auditory looming perception: influences on anticipator judgments. *Perception, 22,* 1467–82.

Rossing, T. D., & Fletcher, N. H. (1995). *Principles of vibrations and sound.* New York: Springer-Verlag.

Russell, M., & Turvey, M. T. (1999). Auditory perception of unimpeded passage. *Ecological Psychology, 11,* 175–88.

Schiff, W., & Oldak, R. (1990). Accuracy of judging time to arrival: Effects of modality, trajectory, and gender. *Journal of Experimental Psychology: Human Perception & Performance, 16,* 303–16.

Schubert, E. D. (1975). The role of auditory perception in language processing. In M. Rawson & D. D. Duane (Eds.), *Language, perception, and reading* (pp. 97–130). Baltimore: York.

Shaw, B. K., McGowan, R., & Turvey, M. T. (1991). An acoustic variable specifying time to contact. *Ecological Psychology, 3,* 253–61.

Shockley, K., Grocki, M., Carello, C., & Turvey, M. T. (2001). Heaviness perception implicates an ecological form of the rigid body laws. *Experimental Brain Research, 136,* 133–37.

Sidaway, B., McNitt-Gray, J., & Davis, G. (1989). Visual timing of muscle preactivation in preparation for landing. *Ecological Psychology, 1,* 253–64.

Snyder, W. S., Cook, M. J., Nasset, E. S., Karhausen, L. R., Howells, G. P., & Tipton, I. H. (1974). *Report of the task group on reference man.* Oxford, England: Pergamon.

Stoffregen, T. A., & Pittenger, J. B. (1995). Human echolocation as a basic form of perception and action. *Ecological Psychology, 7,* 181–216.

Todd, J. T. (1981). Visual information about moving objects. *Journal of Experimental Psychology: Human Perception & Performance, 7,* 795–810.

Tousman, S. A., & Pastore, R. E. (1989). Source characteristics: A study of hand clapping. *Journal of the Acoustical Society of America Supplement 1*(85), S53.

Tresilian, J. R. (1993). Four questions of time to contact: A critical examination of research on interceptive timing. *Perception, 22,* 653–80.

Turvey, M. T. (1992). Affordances and prospective control: An outline of the ontology. *Ecological Psychology, 4,* 173–87.

———. (1996). Dynamic touch. *American Psychologist, 51,* 1134–52.

Turvey, M. T., & Carello, C. (1995). Dynamic touch. In W. Epstein & S. Rogers (Eds.), *Handbook of perception and cognition, Vol. 5. Perception of space and motion* (pp. 401–90). New York: Academic Press.

Turvey, M. T., Shockley, K., & Carello, C. (1999). Affordance, proper function, and the physical basis of perceived heaviness. *Cognition, 17,* B17–B26.

Vanderveer, N. J. (1979). Ecological acoustics: Human perception of environmental sounds. *Dissertation Abstracts International, 40*(09B), 4543. (University Microfilms No. 8004002).

Warren, W. H., & Verbrugge, R. R. (1984). Auditory perception of breaking and bouncing events: A case study in ecological acoustics. *Journal of Experimental Psychology: Human Perception & Performance, 10,* 704–12.

Warren, W. H., & Whang, S. (1987). Visual guidance of walking through apertures: Body-scaled information for affordances. *Journal of Experimental Psychology: Human Perception & Performance, 13,* 371–83.

Wightman, F. L., & Jenison, R. (1995). Auditory spatial layout. In W. Epstein & S. J. Rogers (Eds.), *Perception of space and motion* (pp. 365–400). San Diego: Academic Press.

Wildes, R., & Richards, W. (1988). Recovering material properties from sound. In W. Richards (Ed.), *Natural computation* (pp. 356–63). Cambridge, MA: MIT Press.

Wuestefeld, A., & Rosenblum, L. (1996). *Perceptual learning in an auditory time to arrival task.* Technical Report UCR.96.011. University of California, Riverside.