CellPress
Current Biology
Magazine

Primer

Physics, ecological acoustics and the auditory system

Josh H. McDermott[1,2,*],
Vinayak Agarwal[1,3], and James Traer[4]

Sound occurs when vibrations travel through a medium. These vibrations, their propagation, and their interaction with the environment are dictated by laws of physics, and indirectly reflect underlying physical properties of things in the world. The sense of audition exists to measure sound and infer its causes in the world, so as to help organisms interact with the world around them. Audition is thus indirectly but intimately shaped by the physics of sound.

The way in which physical laws, objects, and environments collectively result in sound signals is known as ecological acoustics. Implicit knowledge of ecological acoustics is integral to real-world perception. At a minimum, we must have knowledge of the associations between acoustic signals and types of events in the world, for instance knowing that heavy objects sound one way and light objects sound a different way. In addition, most of our perceptual inferences are ill-posed, meaning that the inference is not possible without other constraints. This is often because the problem effectively requires solving an equation that contains multiple unknown variables. The best-known example is the problem of inferring individual concurrent sound events. The ear receives a sound signal that is a mixture of the sounds that would be produced by the individual events on their own, and there are many combinations of events that are consistent with a received sound mixture.

Such problems can be solved by leveraging the fact that the variables to be estimated, and the ways they interact to generate sound, are not random. They are non-random by virtue of the fact that they result from lawful processes in the world. As a result, explaining how we hear in the world requires first understanding ecological acoustics, then assessing the extent to which organisms have internalized ecological acoustics, and then ultimately developing theories of perceptual inferences rooted in the aspects of ecological acoustics that matter to an organism.

This primer will review how physical principles influence sound, how these influences encode information about the world into sound signals, and how this information is used by the auditory system to infer the structure of the world.

The physics of sound in the natural environment

The vibrations that constitute sound originate in physical interactions between objects or in gasses or liquids. When a rigid object is struck or rubbed, for instance, vibrations are set up inside the object, some of which are preferentially maintained due to resonances that depend on the object material, size and shape.

Some of the underlying principles can be seen in the classic example of a string under tension. The string's geometry causes it to vibrate much more readily at some frequencies than others. If plucked, the string initially vibrates at all frequencies. However, frequencies whose wavelengths do not divide evenly into the string length interfere with each other and dissipate almost immediately, leaving only the frequencies whose wavelengths are integer fractions of the string length (Figure 1A). The frequencies corresponding to these wavelengths are determined by the tension and string density (because these determine the speed of wave propagation). These are the resonant frequencies, also known as modes, which for a string are the fundamental frequency (one cycle of this frequency is the time it takes for vibration to traverse the length of the string) and its harmonics (integer multiples of the fundamental).

The same principles give rise to resonances in three-dimensional objects. In three dimensions, however, the resonant frequencies are typically inharmonic, as the geometry supports multiple paths of different lengths for vibrations to traverse (Figure 1B). Akin to the effect of tension and density in strings, the speed of wave propagation is different in different object materials (being faster for harder and denser materials), yielding different resonant frequencies for different materials. As in strings, larger and softer objects have lower-frequency resonances. The sound that radiates from the object thus encodes physical properties of the object.

Similarly, turbulent liquids and gasses generate vibrations, and are subject to resonances whose frequency depends upon the dimensions of the turbulent region. For instance, consider how the sound of an open car window can change from a dull roar (containing energy at all frequencies) to a high-frequency hiss (containing a narrower range of frequencies) as the window closes. When resonances are sufficiently pronounced, they create whistling sounds that are dominated by a single frequency (as in a tea kettle when water reaches a boil). Bubbles in liquids serve as resonators, such that the acoustic properties of many liquid sounds are determined by the number and size of their bubbles.

Once vibrations radiate from a source, they propagate through a medium, which in turn alters the sound. Sounds can often be approximated as spherical waves, emanating in all directions from the radiating object. As a consequence, the sound energy at a point some distance away from a radiating object decreases roughly with the square of the distance from the object (the 'inverse square law', whereby sound intensity decreases by 6 decibels with every doubling of distance; Figure 1C). Sound energy is also absorbed by the medium through which it travels, with the absorption proportional to frequency. All waves lose some fraction of energy per oscillation cycle, and high-frequency sounds have more cycles per second, causing sound to be low-pass filtered — with low frequencies attenuated more than high frequencies — to an extent that depends on the distance from the source.

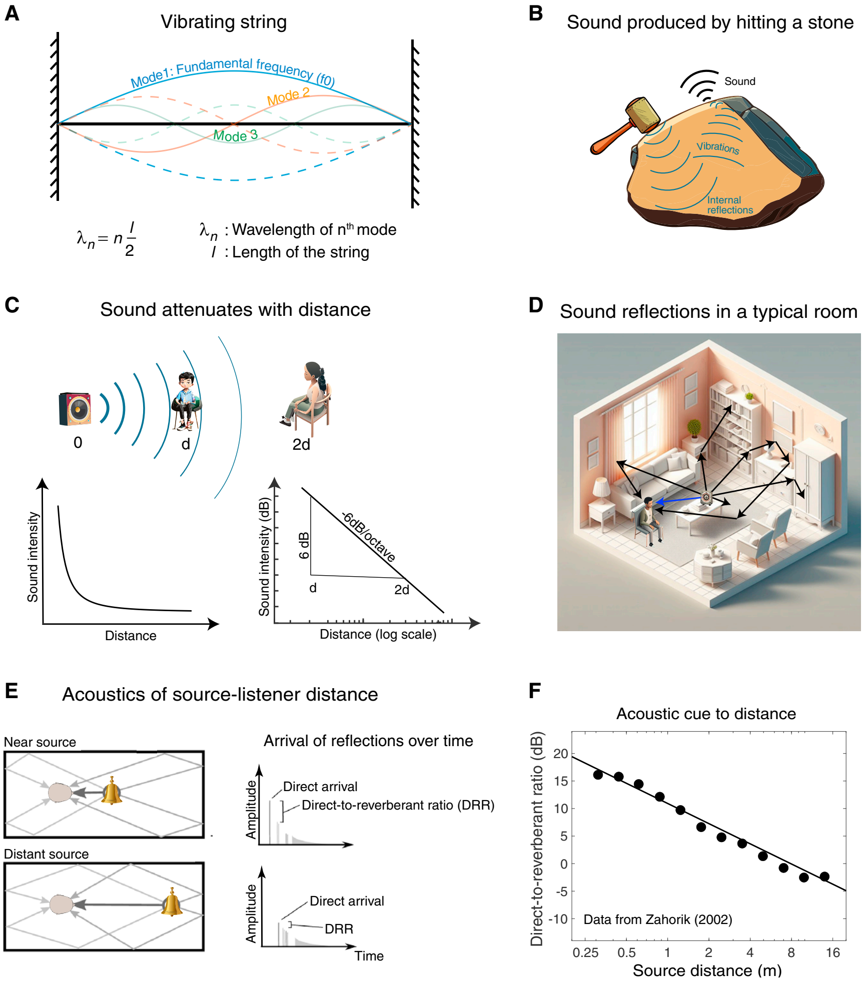
When the radiating sound encounters a surface, some sound energy is absorbed, and some is reflected. The amount that is absorbed versus reflected varies with frequency, and this dependence, as well as the overall reflectivity, depends on the surface material — specifically, on the difference between the medium (typically air) and the surface material — as well as on the surface finish. For instance, ceramic tiles reflect a large proportion of sound whereas carpet reflects less.

The propagation of sound within an environmental space can create resonances via the same principles that underlie resonances in strings and objects, with resonances again determined by the wavelengths that are an integer fraction of one of the room dimensions. Because wave propagation is much slower in air compared to in solid objects, however, the resonances are usually much weaker compared to those in objects (or strings), and are a less salient aspect of room acoustics compared to object acoustics.

One consequence of the reflections in a space is that an organism listening to a sound source receives sound via indirect paths that involve reflections (Figure 1D). Because these indirect paths are longer than the direct path the sound may also take, and because each reflection absorbs some fraction of the incoming sound energy, the sound received via the indirect paths is delayed, attenuated, and filtered relative to the direct sound. The reflections add together in the air, and their collective effect is known as reverberation (Figures 1D and 2A).

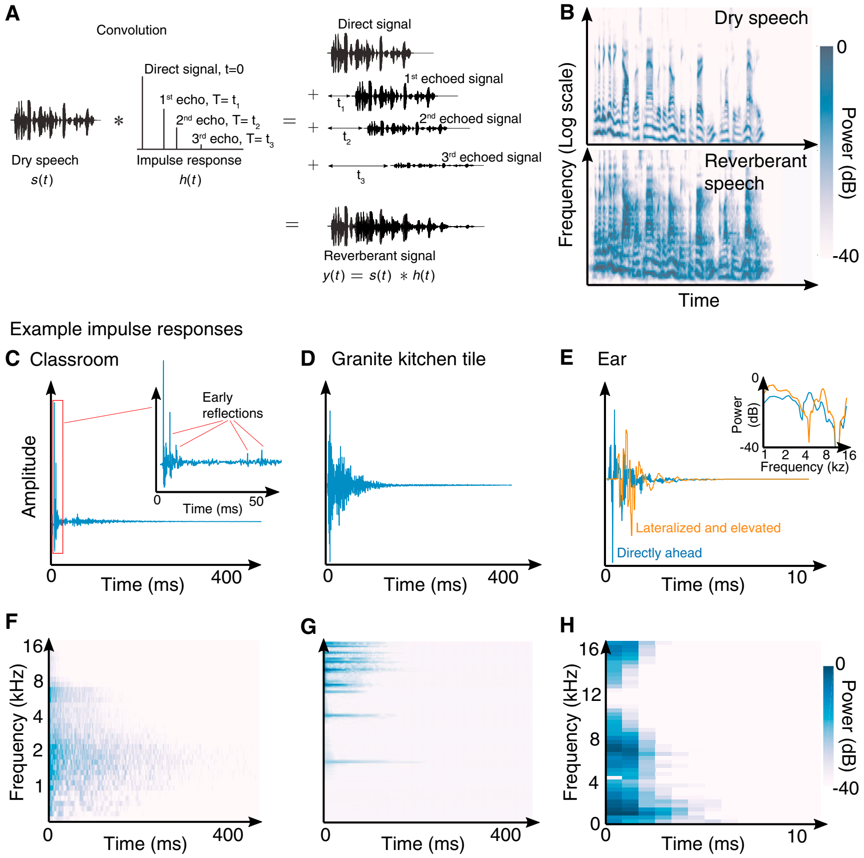
Reverberant sound contains information about the sound source's position. By the inverse square law, the sound coming to a listener directly from the source decreases in intensity with the source distance. However, in many spaces the reverberant sound energy is fairly constant with distance, because no matter the positions of the source and listener, the average indirect path length, and the average number of reflections per path, tends to be fairly constant (Figure 1E). As a result, the ratio of the direct sound energy to the reverberant sound energy provides information about the source distance (Figure 1F). The energy ratio cue is not necessarily straightforward for the listener to estimate, because the listener receives the sum of the direct sound and its reflections, and would have to somehow separate them in order to compute the ratio (another ill-posed problem). But because the direct sound reaches the listener first, one can imagine that this cue could be extracted fairly well in many situations.

Reverberation also conveys information about the environment. If a space is large, the length of time



Current Biology

Figure 1. Resonances and reverberation convey information about the physical world.
(A) Resonances in a string. The resonances, also known as modes, occur for wavelengths that divide evenly into the string length. The frequencies of the resonances are determined by the wavelength in conjunction with the string tension and density. (B) Sound propagation within an object. Object resonances are determined by the same physical principles governing strings but are more complicated because there are multiple paths that sound can take through three dimensions. (C) Sound attenuation with distance. Because sound waves can often be well approximated as spherical, the power measured in a unit area decreases with the square of the distance from the source (the 'inverse square law'). This means that sound intensity decreases by 6 dB with every doubling of distance. (D) Environmental reflections. The sound from the source on the table reaches the listener via a direct path but also after reflecting off surfaces in the surrounding environment. The collective effect of the reflections is known as reverberation. (E) The distance of a sound source from a listener is partially conveyed via the ratio of direct to reverberant sound. Left panel depicts a sound source at two different distances from the listener. The overall sound energy from reflections is similar for the two configurations because the number of reflections, and their average path length to the listener, is similar irrespective of the source position. But the amplitude of the direct sound decreases with distance due to the inverse square law, such that the ratio of direct to reverberant energy provides a cue to distance. Right panel shows an idealized depiction of the amplitude of the direct sound and reflections for the two source positions. The timing of the initial reflections (as well as their path direction) varies with the position of the source, but the overall reverberant energy is fairly consistent. (F) The direct-to-reverberant energy ratio measured from the right ear of five human listeners. Data are replotted from Zahorik (2002). Note that the energy ratio cue is not necessarily straightforward for the listener to estimate because the listener does not receive the direct sound and its reflections as separate quantities.



Current Biology

Figure 2. Many acoustic effects can be summarized with impulse responses.

(A) The effect of environmental reflections on the sound received by a listener from a source can be summarized with an impulse response. The sound generated by a source signal is the convolution of the source signal with the impulse response, which scales and sums delayed copies of the source signal. The word "dry" describes a source signal without reverberation. (B) Spectrograms of a speech signal, and of its convolution with the impulse response from a reverberant room. Reverberation blurs the source signal over time. (C) Impulse response of a classroom. Inset shows the initial 50 ms of the impulse response, in which individual early reflections are apparent. (D) Impulse response of an object (a granite tile). (E) Impulse response of a human outer ear, measured with sources at two directions relative to the listener (blue and orange). Note that the time scale is much shorter than in (C) and (D). Inset shows transfer functions for the two impulse responses. Note that the peaks and valleys are at different frequencies for the two sound directions, providing a localization cue. Data replotted from the KEMAR recordings of Bill Gardner and Keith Martin (https://sound.media.mit.edu/resources/KEMAR.html). (F) Time-frequency representation of room impulse response from (C). Note that the middle frequencies decay more slowly than higher or lower frequencies. (G) Time-frequency representation of object impulse response from (D). Note that the object impulse response contains relatively discrete modes (these are the resonant frequencies of the object). (H) Time-frequency representation of the directly ahead (blue) ear impulse response from (E). A linear frequency scale is used to better show the spectral structure. Note the peaks and valleys in the spectrum, which provide signatures of the incoming sound's direction.

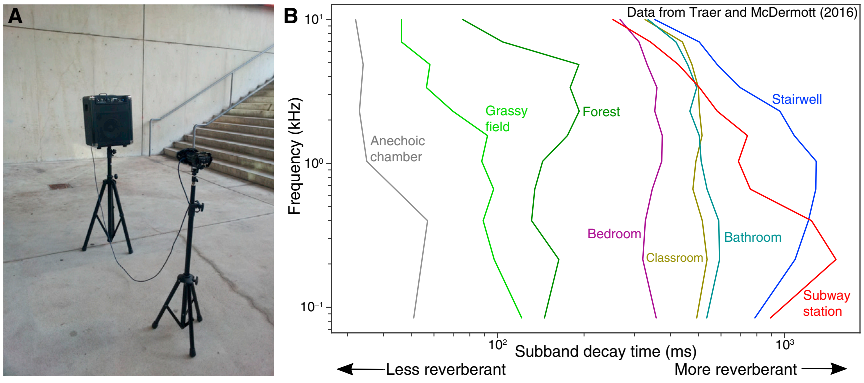
between reflections is longer, and thus the reverberation decays more slowly than in a smaller space. This cue is not perfectly reliable because the decay time is also influenced by the material of environmental surfaces, which can absorb sound to different degrees. For instance, the reverberation in a bathroom could decay at a similar rate to that of a larger room with less reflective walls (see Figure 3 for an example). However, reverberation is nonetheless correlated with room size (also evident in Figure 3).

Linear systems approximations

Many of the physical effects that influence sound are well approximated as linear systems. Most obviously, if there are multiple sound sources that concurrently produce sound, the pressure waveform that arrives at the ear is simply the sum of the sounds that would have been produced by the sources individually. This situation is much simpler than what occurs for image formation, in which two objects in the same scene can occlude one another, such that the image of the scene is a nonlinear function of the individual objects.

The filtering of sound that is induced by objects and spaces is also approximately linear most of the time. One consequence is that, for both objects and spaces, the sound they produce can be predicted from an impulse response. The impulse response, as the term implies, is the vibrational response of the system to an impulse (a brief 'click'). The sound produced by a linear system in response to a vibration is well approximated as the convolution of the impulse response and the source vibration (Figure 2A,B). For objects, the source is the time-varying force that creates vibrations in the object, for instance if it is struck by another object or rubbed against a surface. The sound radiated from another point on the object is approximately equal to the convolution of the time-varying force and the impulse response of the object between the point of contact and the point of radiation. For a room, the source would be the sound signal emitted by a sound source in the room (such as a person talking, or a struck object), and the impulse response is defined between the source and listener positions.

The impulse response is a powerful descriptor because it fully defines the dynamics of a linear system, thus allowing the estimation of the system output for any input. It thus provides a summary of the acoustic properties of an object, or an environmental space. Each pair of points on an object's surface in principle has a distinct impulse response, but impulse responses for different points on an object are often fairly consistent, being constrained by the object shape and material. Similarly, each pair of source–listener positions in a space has a distinct impulse response, but there is usually some consistency across positions that reflects the dimensions of the space and surface materials.



Current Biology

Figure 3. Impulse response properties vary across everyday environments.
(A) Measurement of an environmental impulse response. A noise signal is played from a speaker (left) and recorded at a nearby location (right). The recorded sound is the convolution of the environment's impulse response with the noise signal. The impulse response can be estimated given the recorded audio and the original noise signal. (B) Decay time in different frequency bands of impulse responses for a set of natural environments. Impulse responses were measured in everyday environments, along with an anechoic chamber, and analyzed using a set of bandpass filters. Decay time was quantified as the reverberation time (RT60; the time it takes sound energy to decay by 60 decibels) within individual frequency subbands. The results were then averaged across multiple impulse response measurements for a particular type of environment. Longer decay times result in more pronounced smearing of sound over time.

Another important type of linear filtering is that induced by an organism's body and outer ear. Sound entering the ear interacts with the head and torso, and reflects around the outer ear (pinna) in ways that depend on the direction of arrival. As a result, the sound is filtered differently depending on where it comes from, providing an important source of information about a sound source's location. The filtering can again be summarized with an impulse response for each incoming sound direction and ear. These impulse responses can be convolved with a recorded sound, and then played over earphones, recreating the acoustics, and perception, of a sound at an external location in space.

Filtering also happens in vocal tracts. The vocal tract has resonances that are determined by the dimensions of the internal cavities, and that are imposed on the sound produced by vocal cords. Organisms, most notably humans, manipulate the cavity dimensions to change these resonances, providing a communication signal.

The physics that underlies the filtering of pinnae, vocal tracts, objects and rooms is largely the same set of simple principles, but the effect of these different types of filters can be evident in different ways due to the timescale of the filtering. In Figure 2C and 2F, note that the impulse response of the example room has a timescale of hundreds of milliseconds. When convolved with a source signal (Figure 2B), this has the effect of blurring the source structure in time (compare top and bottom spectrograms of Figure 2B). By comparison, the impulse response from the torso and outer ear (Figure 2E,H) is much shorter in time. This latter filtering effect is thus primarily evident in the frequency domain, where it induces changes in the incoming sound's spectrum. Vocal tract filtering has a similarly fast time-scale, and is accordingly also mostly evident in the frequency domain.

There are admittedly common cases where the underlying vibrational mechanics are not linear. For instance, when objects are nonrigid — for example, something thin and sheet-like that flexes when struck — the impulse response between two points is not fixed, but rather varies with the object configuration. Another example is that if forces are sufficiently high, an object's vibration typically ceases to be a linear function of the force. But linearity is a good approximation much of the time.

The prevalence of approximately linear interactions in acoustics has at least two important consequences. One is that it facilitates the rendering of realistic synthetic scenes, by allowing many scenes to be synthesized from a set of impulse response measurements. A second consequence of linearity is that it facilitates the characterization of real-world acoustics, in that the impulse response is a natural quantity to analyze to reveal physics-induced regularities that might be important for perception.

Physics-based regularities in environmental acoustics

The main goal of audition is to estimate the causes of sound in the environment. Part of the challenge is that there are typically multiple such causes for any observed sound. For instance, the sound received from a source in a room is a function of both the signal that radiates from the source, and the environment — specifically, the impulse response for the source and receiver positions in a room. The problem of estimating the source signal, and/or the impulse response that summarizes the effects of the environment, is ill-posed: the problem lacks a unique solution because there are an infinite number of combinations of source signals and impulse responses that combine to yield the same convolved signal. A similar statement could be made about the force that generates an object sound and the impulse response of the object, and about the sounds from multiple objects in the environment. Such ill-posed inferences can only be made successfully if the system performing the inference has knowledge of the structure that makes the generative process non-random. Regularities based on physics are thus of potentially great perceptual importance.

Impulse responses provide a natural focus for the examination of acoustic regularities, because they summarize the acoustic effect of spaces and objects. When examined at scale, impulse responses of environmental spaces tend to have a fairly stereotyped form. Some of the stereotyped structure is evident when they are plotted as waveforms (Figure 2C): there is usually an initial impulse corresponding to the sound that reaches the receiver directly from the source, followed by some 'early' reflections that reach the receiver after a small number of surface reflections. After that, the reflections tend to blur together to form a dense 'tail' that decays exponentially over time. The impulse response can also be plotted as a spectrogram that shows

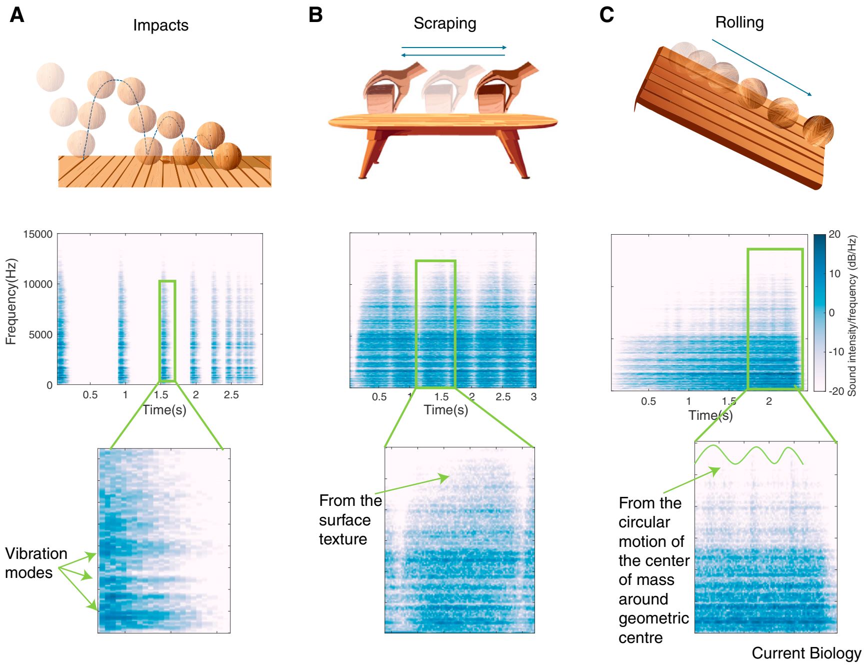


Figure 4. Sound produced by object impacts, scrapes and rolls.
In each case the sound reflects the combined influences of the time-varying force between the object and surface, and the impulse responses of the object and surface. Here the object is small relative to the surface, and so the radiated sound is dominated by that from the wooden surface. (A) Each impact produces a brief sound that is the convolution of a momentary force between the ball and the floor and the impulse response of the wooden floor. The modes of the impulse response are evident as dominant frequencies in the individual impact sounds. (B) The scraping sound reflects the motion of the hand, the texture of the surface, and the resonances of the surface (evident as the horizontal bands of energy in the spectrogram). (C) The surface texture and resonances are also evident in the rolling sound. In addition, the sound contains periodic structure that speeds up over time (evident as vertical streaks in the spectrogram). The periodicities are caused by the ball's deviations from perfect sphericity, and speed up as the ball accelerates down the ramp.

how energy at different frequencies decays over time (Figure 2E). In general, for impulse responses from real-world environments, middle frequencies tend to decay more slowly than lower or higher frequencies, due to the combined effects of typical surface absorption properties and the absorption properties of air.

Object impulse responses also have structure when viewed in the time-frequency domain. Like room impulse responses, their energy decays over time (Figure 2D). However, the spectral structure of object impulse responses usually consists of relatively discrete modes that resemble decaying sinusoids (Figure 2F). Although rooms and other spaces often have modes, in most real-world situations the modes are quite coarse. This difference between objects and rooms may underlie the ability to distinguish the effects of rooms and objects, as when an object is struck in a reverberant environment.

Room and object impulse responses also vary somewhat systematically with the type of room or object. For example, Figure 3 shows how sound at different frequencies decays over time in different environmental spaces. There is variation in frequency, as discussed above, but also pronounced differences between different spaces. Reverberation persists for longer in larger spaces (such as subway stations), and in spaces with reflective walls (such as bathrooms), and for indoor compared to outdoor spaces. In all cases the reverberation in natural environments exceeds that in an anechoic chamber, which is designed to eliminate most reflections (with sound absorbing foam on all surfaces).

There is similarly systematic variation present in objects. Though it is less thoroughly characterized at present, harder materials tend to have longer impulse responses, and higher frequency modes.

Audition has internalized physical dependencies in sound

The prevalence of physics-based regularities in real-world sound raises the possibility that the brain has internalized these regularities in the service of auditory perception. One way to test whether humans have implicit knowledge of the acoustic regularities imbued by physics is to violate these regularities and assess whether perception is qualitatively altered. For instance, if reverberation is synthesized with impulse responses that deviate from real-world reverberation — either by decaying linearly rather than exponentially, or by having an unnatural dependence of energy decay on frequency — humans typically do not interpret it as reverberation. If synthesized with high frequencies that decay more slowly than middle frequencies, the resulting impulse response tends to sound like the hiss of air rather than reverberation from a space.

Humans also appear to be less able to separate impulse responses from source signals when the impulse responses are unnatural. Specifically, they are impaired at discriminating source signals convolved with impulse responses when the impulse responses violate the characteristics of natural reverberation. Such results suggest that humans have internalized both the compositional structure of the causes of sound and the distribution of the constituent causes in the world (in this case environmental impulse responses and sound sources), and that they use implicit knowledge of this structure to disentangle the causes of a sound.

Humans also have implicit knowledge of the acoustic signatures of specific physical causes in the world. In particular, we have some ability to infer room size from reverberation, and some ability to infer distance from reverberation (via the ratio of direct to reverberant energy; Figure 1E). Distance cues are important for telling whether sound sources are near or far away, but also constrain sound recognition. One diagnostic feature of sound sources is the energy produced by the source. For instance, the sound of someone combing their hair is low in intensity, because the forces involved are small. By contrast, scraping a rake over a driveway produces considerably more sound energy due to the larger forces generating the sound. Humans use sound intensity to help determine a sound's cause, such that if the sound of combing hair is played at an unnaturally high intensity, it tends to be misrecognized as a high-intensity source such as raking. However, the intensity of a sound at the ear is also a function of the source's distance from the listener, and humans have internalized this relationship, appearing to use distance cues to infer the intensity of a sound at its source. One consequence is that low-intensity sources, such as the combing of hair, become more poorly recognized when reverberation is added that suggests the source is far away, even if the overall intensity is preserved. This result and others indicate that implicit knowledge of the relationship between intensity and distance is woven into the way we derive auditory information about the world.

Physical influences on sounds from object interactions

Sounds from object interactions are particularly informative of physical variables in the world, in part because different types of physical interactions produce different sounds. If an object is struck by another object, a momentary force is created as the two objects come into contact and then move apart, which then resonates within the objects as can be predicted by their impulse responses (Figure 4A). The force reflects the mass and stiffness of the colliding objects, being briefer for stiffer objects and smaller masses and longer for softer objects and larger masses. If two objects scrape against each other, the force that is created depends on the object surface textures because variations in force occur as the fine-grained peaks in the object surfaces come in and out of contact (Figure 4B). This force is also a function of the object motion and the vertical pressure on the object. The rolling of spherical objects also produces sound that is influenced by the texture of the surface on which the rolling occurs, with the rolling object successively impacting the fine-grained peaks on the surface. One key difference with scraping is that rolling sounds contain signatures of the approximate sphericity of the rolling object. The small deviations from sphericity that are inevitably present in a real-world sphere create periodicities in the resulting sound that we detect as rolling (Figure 4C).

Most physical interactions have a temporal evolution that is determined by physics. For instance, a dropped object might bounce a few times before settling on a surface, with the bounce pattern determined in part by the object's shape and the surface geometry. Similarly, an object pushed across a surface will slide to a halt once released, with the trajectory determined by the surface friction. And as a glass is filled with water, the water volume increases and air volume decreases, changing the resonances in the glass and causing the dominant frequency to rise. By listening, we can often apprehend the physical interactions that gave rise to the sound.

Some of the physical properties evident in sound are more evident acoustically than via other sensory modalities. For instance, an object's mass or material may not be very evident from sight alone, but is often obvious from the sound an object makes when dropped. Similarly, sound is often uniquely informative as to whether two objects are in contact when one of them moves (because they make sound if they are in contact).

Many of the forces exerted on objects originate from animate organisms and come from familiar actions. It is clear from everyday experience that we use sound to infer such actions, and it is plausible that these inferences leverage implicit knowledge of physics, as when we recognize the sound of walking, scratching, or screwing on a jar lid, each of which is produced by particular types of impact and scraping sounds. Animate organisms also have the ability to inject energy into an object, which produces sequences of sounds that are distinct from those produced by the passive dynamics of inanimate objects (imagine the difference between a ball that falls off a table and settles on the ground and one that is dribbled by hand and that continues bouncing for an extended period of time). When we hear these dynamics we can thus infer much about what happened to cause the sound.

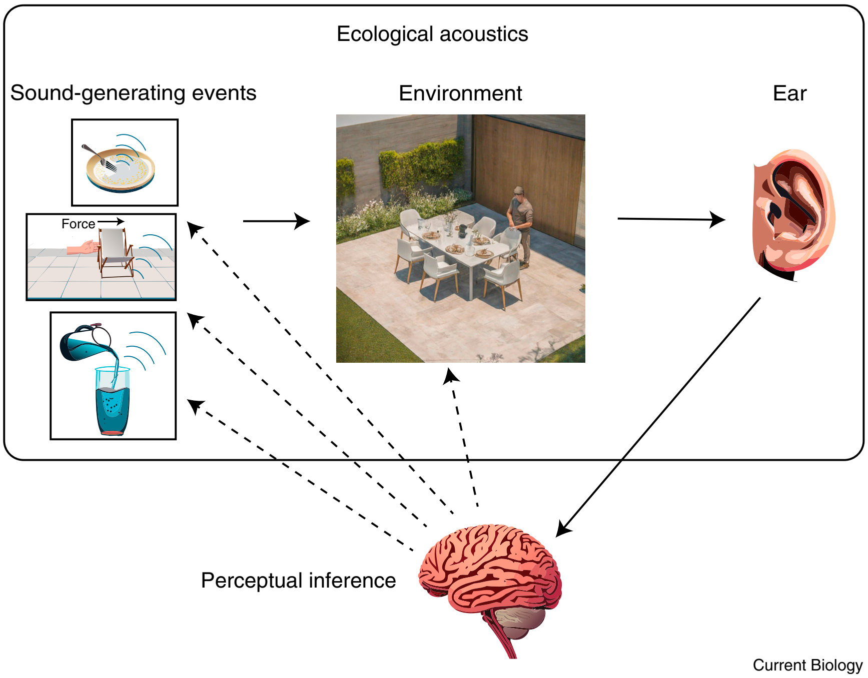
There is a small perceptual literature characterizing human perception in these domains, and some work in sound synthesis that has facilitated controlled perceptual experiments, but much more work is needed to document and understand human abilities, particularly in realistic settings.

Role in music and film production

Physics-driven acoustic effects, and their role in audition, are critical to sound production for music and film. Synthetic reverberation is perhaps the most common effect in sound production, being used to imply space, distance, and energy by reproducing the acoustic signatures of sounds in different environmental spaces and different distances from the listener. Such effects were so valued by early music producers that they built dedicated rooms (echo chambers) in which sound was played and then re-recorded to capture the sound of a reverberant space before it was possible to do so electronically. Musical instrument playing itself also leverages physical acoustic signatures – the sound of a drum played with great force is very different from one played gently, and each might be used to create arousing and relaxing music, respectively.

Production effects also employ violations of physical principles to expand the artistic palette. For instance, a reverberant impulse response can be time-reversed and then convolved with musical instrument sounds, creating a psychedelic effect common to 'shoegaze' rock music (most famously embodied by My Bloody Valentine). Alternatively, reverberation can be truncated, creating 'gated' reverb. This effect was popular in early 1980s rock and pop production (think of the drum sounds from "In the Air Tonight" by Phil Collins).

Physical influences on sound and audition are particularly evident in sound effects generation for film. Such effects are often physically generated through a process known as 'foley', in which sounds needed for a film are approximated with convenient everyday objects and materials. The process depends on the fact that distinct types of physical events can approximate the same acoustic signal. For instance, the sound of footsteps in snow is often simulated by squeezing a bag of corn starch. Effective foley tends to reproduce the type of object interaction that the sound aims to evoke, but with



Current Biology

Figure 5. Ecological acoustics and perceptual inference.
Real-world auditory scenes consist of sounds generated by physical events that each have multiple latent causes. For instance, the sound made by a fork hitting a plate is determined by the force of the impact, the materials of the fork and plate, and the sizes of the fork and plate. Similarly, the sound of a chair scraping over a patio is determined by the surface texture of the patio and chair feet, the material of the patio and chair, and the downward force exerted by the chair, and the motion with which the chair is translated. The sounds from these events then interact with the environment and receive spatial cues from their interaction with the ear. The brain must use implicit knowledge of ecological acoustics to infer the physical causes of the sound in the world.

different objects. This practice indicates that these interactions are physically distinctive and perceptually salient to humans.

Open questions

Although much of the physics underlying the effects discussed here is well established, and has been for centuries, some aspects remain incompletely understood. For instance, the physics that dictates some everyday object interactions — for example, those related to friction, or to an object settling to a halt once set in motion — is not yet understood well enough to be simulated accurately. As the understanding of the underlying physics improves, we will likely be better able to synthesize the resulting sounds, and to study their perception.

The fact that humans can correctly interpret the physical causes of sequences of sounds suggests that we have an internal model of the world that incorporates physical principles, but the nature of any such internal model remains unclear. One open question is whether physics-related representations are shared across sensory modalities. Some of the physics-related phenomena discussed here are predominantly relevant to audition (for example, those related to sound propagation and reverberation), but others are equally relevant to other senses (for example, those related to object interactions, which we often see or feel as well as hear). It is thus plausible that the senses interact substantially to make physical inferences, but such interactions remain largely uncharacterized.

Another open question concerns the role of implicit physical knowledge in perceptual inferences of the constituent causes of sound, often known as auditory scene analysis. Auditory scene analysis has tended to be associated with the problem of inferring individual sound sources from mixtures of sounds. The study of this problem has largely focused on communication signals such as speech or music, at the expense of other environmental sounds, such that we know little about how the sounds made by inanimate objects are analyzed in auditory scenes. As a result, the problem has historically not been associated with physical principles, but the regularities imposed by physics on sound could provide a powerful constraint. For instance, there is extensive evidence that harmonically related frequencies, as produced by voices and musical instruments, are grouped together by the brain. The sounds made by objects are typically inharmonic, and thus should not benefit from harmonic grouping, but are nonetheless highly nonrandom. The auditory system may well contain distinct processes for grouping such sounds and the sequences that result from physical dynamics in the world. Experiments assessing whether we tend to infer sound sources that are physically plausible could provide insight into this issue.

Consideration of the physical causes of sound also suggests that the problem of auditory scene analysis should be more broadly construed (Figure 5). Even a simple real-world scene that in traditional terms contains a single source is typically the result of many distinct causal factors in the world. The sound of an object dropped on the floor reflects the force with which the object hits the floor, the material of the object and floor, the shape and size of the object, and the reverberation imposed by the surrounding environment. Organisms appear to infer these distinct causal factors, presumably using implicit knowledge of the world to constrain what would otherwise be an ill-posed inference, analogous to how we are thought to separate individual sound sources from mixtures. Auditory scene analysis may thus be best conceived as a larger and more diverse domain of causal inference, and should be studied in the context of ecological acoustics and the physical causes of sound.

DECLARATION OF INTERESTS

The authors declare no competing interests.

FURTHER READING

Agarwal, V., Cusimano, M., Traer, J., and McDermott, J.H. (2021). Object-based synthesis of scraping and rolling sounds based on non-linear physical constraints (Proceedings of the 24th International Conference on Digital Audio Effects DAFx-20in21). http://dx.doi.org/10.23919/DAFx51585.2021.9768225.

Conan, S., Thoret, E., Aramaki, M., Derrien, O., Gondre, C., Ystad, S., and Kronland-Martinet, R. (2014). Intuitive synthesizer of continuous-interaction sounds: Rubbing, scratching, and rolling. Comput. Music J. *38*, 24–37.

Gaver, W.W. (1993). What in the world do we hear? An ecological approach to auditory source perception. Ecol. Psychol. *5,* 1–29.

Giordano, B.L., and McAdams, S. (2006). Material identification of real impact sounds: Effects of size variation in steel, glass, wood, and plexiglass plates. J. Acoust. Soc. Am. *119*, 1171–1181.

Grassi, M., Pastore, M., and Lemaitre, G. (2013). Looking at the world with your ears: How do we get the size of an object from its sound? Acta Psychol. *143*, 96–104.

Ivanov, A.Z., King, A.J., Willmore, B.D., Walker, K.M., and Harper, N.S. (2022). Cortical adaptation to sound reverberation. eLife *11*, e75090.

Klatzky, R.L., Pai, D.K., and Krotkov, E.P. (2000). Perception of material from contact sounds. Presence *9*, 399–410.

Koumura, T., and Furukawa, S. (2017). Context-dependent effect of reverberation on material perception from impact sound. Sci. Rep. *7*, 16455.

Lemaitre, G., and Heller, L.M. (2012). Auditory perception of material is fragile while action is strikingly robust. J. Acoust. Soc. Am. *131*, 1337–1348.

Lutfi, R.A., and Oh, E.L. (1997). Auditory discrimination of material changes in a struck-clamped bar. J. Acoust. Soc. Am. *102*, 3647–3656.

Rath, M., and Rocchesso, D. (2005). Continuous sonic feedback from a rolling ball. IEEE MultiMedia *12*, 60–69.

Selfridge, R., Moffat, D., Avital, E.J., and Reiss, J.D. (2018). Creating real-time aeroacoustic sound effects using physically informed models. J. Audio. Eng. Soc. *66*, 594–607.

Traer, J. and McDermott, J.H. (2016). Statistics of natural reverberation enable perceptual separation of sound and space. Proc. Natl. Acad. Sci. USA *113*, E7856–E7865.

Traer, J., Cusimano, M., and McDermott, J.H. (2019). A perceptually inspired generative model of rigid-body contact sounds (Proceedings of the 22nd International Conference on Digital Audio Effects DAFx-19).

Traer, J., Norman-Haignere, S.V., and McDermott, J.H. (2021). Causal inference in environmental sound recognition. Cognition *214*, 104627.

Warren, W.H., and Verbrugge, R.R. (1984). Auditory perception of breaking and bouncing events: A case study in ecological acoustics. J. Exp. Psychol. Hum. Percept. Perform. *10*, 704.

Zahorik, P. (2002). Direct-to-reverberant energy ratio sensitivity. J. Acoust. Soc. Am. *112*, 2110–2117.

Zheng, C., and James, D.L. (2011). Toward high-quality modal contact sound. ACM Trans. Graph. *30*, 1–12.

[1]Department of Brain and Cognitive Sciences, McGovern Institute for Brain Research, and Center for Brains, Minds and Machines, MIT, Cambridge, MA 02139, USA. [2]Program in Speech and Hearing Biosciences and Technology, Harvard University, Cambridge, MA 02138, USA. [3]Department of Mechanical Engineering, MIT, Cambridge, MA 02139, USA. [4]Department of Psychological and Brain Sciences and Iowa Neuroscience Institute, University of Iowa, Iowa City, IA 52242, USA.
*E-mail: jhm@mit.edu

Primer

Olfactory navigation in fluctuating environments

Venkatesh N. Murthy

We humans move around in the world, guided largely by light and sound. Many cohabitants of our planet, however, predominantly use their chemical senses to navigate a rich landscape. Light and sound propagate with predictable geometric precision, and animals in particular have evolved ways to exploit these physical principles. Odors, on the other hand, are at the mercy of the carrier medium, air or water, for long distance transport, which can quickly become turbulent and unpredictable. Nevertheless, animals have found ways to navigate these fickle features to chase mates, find food or return home. Understanding the physics of odor transport can help rationalize the strategies animals use for navigation and guide studies of how the corresponding algorithms are implemented by their brains and bodies.

Animals use chemical signals to search for food or avoid dangerous situations. Chemical cues can be used as landmarks, such that traveling animals can associate smells with specific locations. In this case, olfactory cues are not necessarily sensed from afar and tracked towards as a target; instead their presence marks a waypoint.

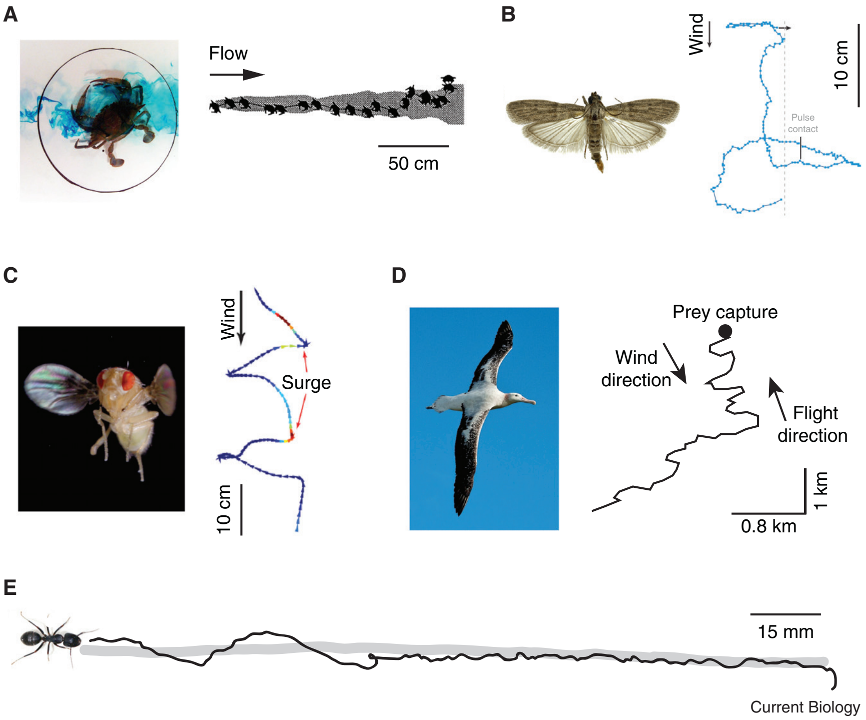


Figure 1. Trajectories of animals navigating through fluctuating chemical landscapes.
(A) Track of blue crab navigating prey odors, visualized with fluorescent dye (gray), in turbulent flow. Crab symbols shown every second (left: © Dr. Donald R. Webster and Dr. Marc Weissburg, Georgia Institute of Technology; right: reused with permission of University of Chicago Press — Journals, from Zimmer-Faust *et al*. (1995) *188*, 111–116). (B) Almond moth and a flight track in a wind tunnel with a brief pheromone pulse released at the indicated time, with an upwind surge occurring 200 ms later (left: almond moth © Birgit E. Rhode/Wikimedia Commons (CC BY 4.0); right: reused from Cardé (2021) *66*, 317–336 with permission of Annual Reviews Inc.). (C) *Drosophila* flying upwind in a wind tunnel with a plume of ethanol released at the top location (photograph courtesy of Floris van Breugel, from van Breugel *et al*. (2014) *24*, 274–286). (D) Albatross flight tracked with a global positioning system (GPS) and prey capture inferred from stomach temperature measurements (left: © Marc Guyt/www.agami.nl; right: from Nevitt *et al*. (2008) *105*, 4576–4581 © National Academy of Sciences, USA). (E) Carpenter ant following a pheromone trail (gray) drawn in a laboratory setting (used with permission of The Company of Biologists Ltd., from Draft *et al*. (2018) *221*, jeb185124).