3 Sources and Images

Sounds are caused by mechanical movements that disturb the surrounding air. The sound may be caused by a brief burst of energy (such as an object falling on the ground) or a more sustained energy source (such as two objects rubbing against each other).

What we call "a sound" is actually pretty complicated. Even a simple sound, like a single tone played on a flute, is really a collection of component frequencies called *partials*. A partial is an elementary "part" of a sound that cannot be further broken apart. The discovery that nearly all sounds consist of such "subsounds" is the single most important acoustical discovery in history. Yet the discovery was made over twenty-five hundred years ago in both ancient China and ancient Greece. Each partial has a unique frequency (or rate) of vibration. Partials exist because most vibrating objects can vibrate in more than one way at the same time. If you strike a cube of gelatin with a spoon, for example, it won't simply rock back and forth; rather, it will wobble and jiggle in a complicated dance. That is, the gelatin will vibrate in several different ways simultaneously—left to right, back and forth, up and down, twisting clockwise then counterclockwise, and so on. Technically, each manner of vibration is referred to as a *mode*. Each mode has its own distinct frequency (or rate) of vibration for a given vibrating object.

The unit of frequency is the *hertz* (*Hz*). One hertz is one cycle of movement per second. When you walk, your head bobs up and down about two times each second: its frequency of movement is about 2 Hz.[1] Your head bobbing generates a sound, but its frequency is too low for humans to hear. A wobbling cube of gelatin also produces frequencies that are too low for humans (although elephants can probably hear it). When you knock on a door, the energy you impart will cause the door to wobble much like the cube of gelatin, although the movement will be much smaller and much faster than the gelatin cube. Because the rate of vibration is so much faster, knocking on the door will generate a sound humans can hear. A given vibrating door might produce a sound consisting of six main partials whose frequencies are (say) 48 Hz, 119 Hz, 142 Hz, 703 Hz, 972 Hz, and 1,488 Hz. Each of these frequencies arises from a different mode of vibration. That is, each partial arises from a different way in which the door bends and flexes in response to being struck. When a sound source vibrates in only one mode it produces a single partial and the resulting sound is referred to as a *pure tone*. When a source vibrates in two or more modes it produces two or more partials and the resulting sound is referred to as a *complex tone*. Nearly all sounds heard in the real world are complex. For most complex sounds, all of the partials will appear at the same time; however, some partials are slower to get started, and some may die away sooner than others.

Depending on the sound source, the partials may be simple multiples of some basic frequency: that is, they may form a series of frequencies that are two, three, four … times some starting value. An example of such a series is 101 Hz, 202 Hz, 303 Hz, and 404 Hz. Each successive frequency is a multiple of 101, which is called the *fundamental* frequency for this set of partials. When partials have frequencies that are related by simple integer multiples, they are referred to as *harmonics*. All harmonics are partials, but not all partials are harmonics. Knocking on a door produces many partials but no harmonics.

Acousticians have shown that there are different classes of vibrating objects. For example, membranes (like balloons, bubbles, and drums) have distinctive modes of vibration that differ from the class of solid or semisolid objects (like doors, wood blocks, and gelatin cubes). An important class of acoustic vibrators happens to include both vibrating strings and vibrating air-filled tubes. A distinctive property of these vibrators is that their partials tend to have frequencies that form a series of simple multiples—that is, their modes of vibration generate a *harmonic series*. Incidentally, the human voice belongs to this class of vibrators.

Most of the common sound sources used to make music produce harmonic partials. Along with the human voice, this includes strings, brass, woodwinds, pianos, organs, guitars, harmonicas, bagpipes, didjeridoos, and many other instruments from around the world. Instruments that produce *inharmonic* partials include drums, timpani, bells, gongs, cymbals, marimbas, glockenspiels, and vibraphones. As a general rule of thumb, if the instrument is played with a mallet or stick, the sound produced probably does not contain harmonic partials (although there are exceptions).

Hearing

Sounds in the world impinge on the eardrum, causing it to vibrate in a way that mimics the vibration of the sound source. The vibrations of the eardrum are then conveyed to a series of three small bones called the *ossicles*, the smallest bones in the human body. The smallest of these, the *stapes*, is just half the size of a grain of rice. The ossicles conduct the sound from the eardrum to the organ of hearing, the *cochlea*.[2] The cochlea is a fluid-filled tube that is curled up so that it resembles a small snail. In fact, *cochlea* is the Latin word for snail. The fluid inside the cochlea moves back and forth in response to the movements of the eardrum.

If we were to uncurl the cochlea, we would end up with a straight tube about 4 centimeters (1.5 inches) long whose diameter is tapered from one end to the other. Sound vibrations enter at the end with the large diameter and travel down the tube toward the smaller end. The small end is called the *apex* and the large end is called the *base*. The tube is divided along its length by a partition that creates two parallel channels. Within this partition is the *basilar membrane*, which contains a dense collection of sensory neurons that connect to the brain via the auditory nerve. The basilar membrane is the most important part of the hearing organ.

Through a tedious series of experiments, the Hungarian physiologist Georg von Békésy discovered that different frequencies stimulate different places along the basilar membrane.[3] Low frequencies are propagated nearly the whole length of the tube and maximally stimulate the most distant points (near the apex of the cochlea). High frequencies, by contrast, do not travel very far and so stimulate points near the base of the cochlea. In effect, there is a mapping between frequency and place of stimulation along the membrane. For this discovery, von Békésy was awarded the Nobel Prize in 1961.

A note played on a piano may generate twenty to thirty partials. These sounds are transmitted through the air to the eardrum, then relayed by the ossicles to the cochlea. Inside the cochlea, the different partials generate different hot spots of stimulation along the basilar membrane. By spreading the frequencies across the membrane, the cochlea attempts to separate out each individual component partial.

If partials are too close together, it will be impossible to detect their separate presence. Two partials need to be at least 1 millimeter apart on the membrane for them to be reliably detected as separate sounds. When the basilar membrane successfully isolates a particular partial, we say that the partial has been *resolved*. Most partials aren't resolved, however. For most sounds, the uppermost partials are too close together for the cochlea to tease them apart. For a typical isolated sound (like a trumpet tone), the basilar membrane may resolve between about five and ten partials. The unresolved partials still contribute to the perception of the sound, but only as a group. For a complicated chord played by an orchestra, hundreds of partials may be present, but the basilar membrane may resolve only twenty-five to thirty of them.

The Packaging of Partials

Oddly, when we listen to a piano tone containing (say) thirty partials, there is nothing in our conscious experience to suggest that there are actually thirty component sounds present. Similarly, there is nothing in our conscious experience to suggest that our cochlea has resolved (say) eight of these partials. We don't hear thirty things or even eight things. Instead, we hear a single thing: a piano tone.

If we open the lid of the piano, we can see two or three strings vibrating for each note. That is, two or three physical sound sources are active when we strike a key on the piano. Still, we don't hear two or three things: we hear one thing, a single sound. The reason is that the brain takes all of the different frequency components and repackages them into something else. Hearing scientists like to call this repackaged thing an *auditory image*. That is, the thirty frequencies generated by three physically vibrating strings (and resolved by the cochlea into perhaps eight partials) ultimately lead to the subjective impression of a single piano tone.

Our experience of the sound is so natural and compelling to us that we mistakenly think we are hearing "the real thing"—what the piano really sounds like. A visiting Martian might find our experience puzzling. The Martian might find it odd that we can't hear that there are actually three piano strings vibrating. Even if we accept the fact that humans can't distinguish the individual vibrating strings, our Martian visitor might still wonder why we can't hear each of the thirty harmonics present in the sound. By talking to a hearing scientist, the Martian might understand that the human basilar membrane is capable of resolving only eight partials for this particular tone. However, if the hearing organ has isolated eight partials in the piano tone, why, the Martian might ask, don't we at least hear the eight resolved harmonics? Eight resolved harmonics are conveyed along the auditory nerve up to the brain, yet our conscious experience is of a single tone rather than eight tones. What is the human brain doing? And why?

There is a good reason for repackaging partials into auditory images. The reason is most obvious when we consider what happens when two unrelated sources generate sound at the same time. Instead of our isolated piano tone, suppose that we heard a piano playing middle C (C4) and a flute playing E5. Together, the two sounds might produce fifty or sixty partials. Perhaps ten or fifteen of these partials will be individually resolved by the listener's basilar membrane. These ten or fifteen partials will be transmitted along the auditory nerve up to the brain. But what is the brain supposed to do with ten or fifteen resolved frequencies?

Brains are practical: they exist to solve practical problems in the world. When interpreting sensations, brains cut to the important stuff: Is that food? Is that my baby crying? Is there any danger here? Where is that sound coming from? What caused that sound?

In the case of sound, it is much more useful for us to perceive "objects" rather than "frequencies." It is much more useful for us to hear "a train whistle" or "a meowing cat" than to identify specific partials. Said another way, the auditory system is designed to recognize sound-producing objects rather than individual modes of vibration. The biological goal here is to create auditory images that represent the actual sound-producing objects in the world. Instead of hearing ten or fifteen individual frequencies, we hear "a piano playing C" and "a flute playing E."

In carrying out this task, the brain has to solve two conceptually separate problems: (1) assembling collections of frequencies into plausible sonic objects and (2) naming or recognizing the physical source for each assembled object. A listener can form an auditory image without necessarily being able to recognize or label the sound. But most of the time we can identify all the sounds we hear.

Brains aren't always successful at this task. That is, they don't always generate auditory images that have a one-to-one relationship with the actual sound sources in the world. For example, twelve violinists playing in unison will not evoke twelve auditory images. Of course, most of us would think that this is not much of a problem. But our Martian visitor might consider this a serious defect in human hearing. Striking an octave on the piano might fool many listeners into thinking that they are hearing a single piano tone. However, many musicians would think that this indicates a deficiency in the listener's skill.

Why do these failures occur? That is, why is the brain not always successful in assembling an accurate picture of the sound-generating objects in the world? Suppose you are the part of the brain responsible for processing sounds. You are looking at ten resolved input frequencies. This might represent ten separate sound sources, each generating a single partial. Or it might be a single sound containing all ten resolved partials. Or it might be five sound-producing objects, each generating two resolved partials. In trying to decipher what might be causing these ten input frequencies, the number of possible combinations turns out to be extraordinarily large.

Hearing scientists refer to the general problem of assembling auditory images from a gaggle of partials as *auditory scene analysis*.[4] In auditory scene analysis, the various partials are grouped together to form plausible auditory images. A common metaphor is to say that the acoustic input is "parsed." In the same way that we parse a sentence into subject, verb, object, and other grammatical elements, the auditory system assigns the partials to different auditory sources. For example, a complex sound scene might be parsed into three images: "slow footsteps over to the right, a telephone ringing behind me, and an aircraft flying overhead." Sometimes identifying labels may not be available to a listener. Where a musician might say, "I hear an oboe, a clarinet, and a bassoon," a nonmusician unfamiliar with instrument names might say, "that nasal sound, that other woody sound, and that third buzzy sound." Both listeners may have successfully parsed the auditory scene, although only the musician has correctly named the sound sources.

The process of auditory scene analysis might be considered successful when each auditory image corresponds to only one actual sound source in the world. An error can occur when several actual sound sources are mistakenly grouped together as a single auditory image or when phantom images are generated that correspond to no actual sound source in the environment.

Notice, by the way, that the biological goal of accurately parsing the auditory scene differs from the musical goal. In music, we may want twelve violins to fuse into a single auditory image, or the pitch sequence sung by a yodeler to break apart and sound like more than one singer.

How does the brain carry out auditory scene analysis? Since the 1970s, researchers have made extraordinary strides in figuring out how the brain does this. The leading researcher in this field has been McGill University psychologist Albert Bregman.[5] Bregman's work suggests that the brain performs scene analysis by relying on a set of heuristics, or rules of thumb. The heuristics are generally successful, but they are also fallible. There is no surefire way to parse an auditory scene into a set of accurate auditory images.

Localization Cues

The flavor of these heuristics can be conveyed by looking at a very simple rule related to *source location*. In general, sounds that come from the same location in space are more likely to come from the same physical sound source. Conversely, sounds that come from different spatial locations are more likely to originate from separate physical sound sources.

Since we are equipped with two ears, the brain has access to some useful localization information. The two principal cues are *time delay* and *intensity*. If a sound source is located to your right, the sound it produces will strike your right ear slightly before your left ear and is likely to be a little louder in your right ear. The brain uses these *interaural* (between-ear) cues to construct a sense of the sound's location.

Localization cues, however, do not provide infallible information for separating sound sources. A simple illustration of how we can be fooled is evident in an ordinary stereo system. When you listen to a stereo, all of the sounds come from the left and right loudspeakers. If you remove the cloth covering the speaker cabinets, you will typically discover that each cabinet contains two physically distinct sound sources: the *woofer* (specialized for low frequencies) and the *tweeter* (specialized for high frequencies). That is, all of the sounds produced by your stereo may be generated by four physical sound sources. Yet when you listen to your stereo, you don't have any sense that there are four physical sound sources.

Stereo technology is intentionally deceptive. The technology has been designed to deceive the human auditory system about the locations of the sounds. However, not all animals will be fooled by a modern stereo system. Bats, for example, have superb localization abilities, so a bat may very well be able to distinguish sounds that originate in the woofer from those generated by the tweeter. Once again, our hypothetical Martian might be surprised that we humans can't tell that all of the sounds are coming from four physical sources. When we listen to a recording, we might hear the singer front and center, the piano left of center, and the acoustic bass right of center. But these are three "virtual" sound sources that don't exist in the real world. In the real world, there are actually four sound sources.

Of course, all this doesn't matter. We don't really want listeners to perceive the real acoustical world of the stereo. We want the stereo's deception to be successful. Ideally, the stereo should act like a sort of transparent sonic window opening onto a fictitious space where the musical sounds are placed. We want people to experience the four parts of a string quartet—first violin, second violin, viola, and cello—not the four parts of the left woofer, right woofer, left tweeter, and right tweeter.

When looking at a mirror, people can often apprehend both the mirror and the virtual images at the same time. That is, a mirror can be perceived as a flat shiny object at the same time that the images "in" the mirror can be seen. In the case of the stereo system, we can sometimes perceive both that sounds are being emitted by a pair of loudspeakers, as well as the virtual sound images that are conveyed. The best sound systems, however, simply evaporate—leaving us with the feeling that the virtual images are the only real ones. Stereo sound is most successful when the loudspeakers themselves disappear from our awareness.

The Pleasures of Stereo

One of the most interesting facts about stereo reproduction is how much nicer it sounds than mono reproduction. When you switch from mono to stereo (or from stereo to mono) there is a direct effect on the experience of pleasure: listeners all over the world prefer stereophonic sound over monophonic sound. Why?

In order to understand why stereo sounds better than mono, we must first consider why pleasure exists. Pleasure—and its opposite, pain—is a powerful motivator. When you accidentally bite your tongue, the experience can be agonizing. But the pain serves a useful purpose: it teaches you to be very careful in coordinating the movements of teeth and tongue. If we didn't experience pain, our tongues would soon be reduced to tatters from constant self-injury.

Like pain, pleasure also serves as an important motivator. Pleasure encourages us to do all sorts of things that are biologically important, like eating, nurturing children, and falling in love. Pleasure and pain provide the carrots and sticks through which biology encourages adaptive behaviors and discourages maladaptive behaviors.

Pain and pleasure are used not simply to punish or reward particular actions but also to punish or reward particular mental behaviors. When you forget a word, you often experience a feeling of mild annoyance or frustration. When you finally retrieve the right word, there is a sense of mild relief or gratification. When you solve a difficult problem, you might have an "aha!" experience that can be deeply satisfying. In each case, the brain provides rewards and punishments to encourage useful mental activity. Without these incentives, our thoughts would simply wander aimlessly.

Apart from encouraging adaptive actions and adaptive thoughts, biological carrots and sticks are also at work in the process of perception. When perceiving the world, the brain is constantly attempting to assemble a coherent scene populated by distinct objects. That is, given all of the sensory elements, the brain is attempting to assemble these elements into a picture where all of the elements are accounted for and all the assembled objects make sense. When this is achieved, we experience a mild sense of pleasure. When we fail at this task, we experience a mild sense of irritation.

When the auditory system is presented with a large number of resolved partials, it faces a considerable challenge in determining how to assign each partial to a particular acoustic source. Localization cues provide one of the helping hands. If frequencies *W* and *X* seem to come from the left and frequencies *Y* and *Z* seem to come from the right, this is evidence suggesting that *W* and *X* may be produced by a single sound source that differs from the sound source producing *Y* and *Z*.

When we listen to a mono recording, all of these localization cues are lost. If there is just one loudspeaker, then partials *W*, *X*, *Y*, and *Z* will all come from the same point in space.[6] The brain must attempt to assign frequency components to different sources without the benefit of localization cues. Does this partial belong to the first violin or to the cello? When the first violin was located off to the right and the cello was located left of center, it was easier for the brain to correctly assign each partial to its true source.

In effect, stereophonic reproduction provides the listener with additional cues that aid in auditory scene analysis. If the brain rewards itself for forming coherent auditory images and if stereophonic localization cues contribute to the listener's ability to form vivid and coherent auditory images, then we can understand why listeners might experience stereophonic sound as sounding more pleasant than monophonic sound.

Now consider the voice-leading rule that forbids parallel octaves. When a solo violin plays an ascending scale, lots of parallel octaves occur as all of the harmonics move upward together in pitch. In fact, research shows that these parallel frequency movements help bind the component partials together so that listeners form a single auditory image of the violin. (We'll discuss this phenomenon more in chapter 6.) But what happens when both the first and second violins play in ascending octaves? At this point, the brain is likely to amalgamate all of the partials from the first and second violins into a single "super-instrument," or *virtual instrument*. By itself, this is not bad. You could have two violins playing a sustained passage in octaves with little perceptual consequence: the brain would simply hear the two instruments as a single auditory image. Like the three strings of the piano tone, most listeners wouldn't be aware of the true number of sound sources. Problems arise, however, when the brain is confused about how to parse the auditory scene. At one moment, the brain may segregate the two instruments as separate auditory images, but at the next moment, the two instruments fuse into a single auditory image. In a passage of otherwise good part-writing, a single parallel octave may sow seeds of doubt in the auditory system about the correct assignment of partials to images. As in the case of a poorly tuned radio that switches from stereo to mono, the momentary parallel octave hampers the efforts of the auditory system to carry out auditory scene analysis. In both cases, there is a loss of pleasure.

The Theater of the Mind

With this background in place, we can now introduce the single most important concept in the field of auditory perception: the distinction between *acoustic* phenomena and *auditory* phenomena. Something is said to be *acoustic* when it pertains to the physical world apart from human beings. For example, when a violin produces a sound, both the vibrating string and the propagated sound are acoustic phenomena. The harmonics or partials that make up the sound are also acoustic phenomena. Something is said to be *auditory* when it pertains to the brain's interpretation of the sonic world. When resolved partials are relayed from the cochlea to the brain, the brain assembles these partials into one or more images, and it is these images that are accessible to consciousness. The violin "out there" is acoustical; the violin in your head is auditory. It should now be clear that you have never heard the true (acoustic) sound of a violin. All you have ever heard are the (auditory) images constructed by your auditory system. In effect, the brain takes the sensory information it receives and creates a sort of theater of the mind where a simulation of the outside world is reenacted on a purely mental stage.[7] If a dolphin or a bat were exposed to the same acoustic violin sound, their brains would create somewhat different theatrical representations of the outside world. The dolphin and bat would have different auditory experiences from us.

Auditory and acoustic phenomena are constantly being confused with each other. In ordinary conversation, the word *sound* is typically used in an ambiguous fashion. We could be speaking of the physical sound generated by a violin or the subjective experience of the auditory image (the "sound") of a violin. When I say, "I hear a violin," I am necessarily speaking of an auditory phenomenon. I might discover that the sound was generated by a computer program simulating the physics of a violin. In this case, I would understand that the sound was not actually produced by a violin. To reduce confusion, it is often helpful to use the modifiers *generated* (for acoustic phenomena) and *evoked* or *perceived* (for auditory phenomena). The violin generates a physical sound. But the physical scraping of a bow on a string evokes the sound of a violin in a person's brain. That is, we perceive a violin. Both the computer program and the actual violin evoke the image of a violin even though the acoustic phenomena are different.

When I say, "I hear Janice playing a violin," what is happening is something along the following lines: Janice is drawing a bow across the string of a physical instrument that people call a violin. The sound is stimulating my auditory system and evoking an image in my brain that I would normally call "the sound of a violin."

This way of speaking might seem ridiculous, but it is musically important. Normally when a (physical) violin is playing, you hear (that is, the sound evokes the experience of) a violin. But there will be times when a (physical) violin is playing yet you will not hear a violin. For example, if a cello and a violin are playing a sustained octave, your brain might misinterpret the partials generated by the violin as belonging to the cello. In other words, your (auditory) experience may be of an especially rich-sounding cello playing a single tone rather than a violin and cello playing two different tones. In one case, the theater of your mind constructs separate images of a cello and a violin. But in the other case, the theater of your mind constructs the image of a sort of cello on steroids. In both cases, however, the acoustic sounds are identical. If we want to understand how people experience sound, it is essential to distinguish acoustic phenomena from auditory phenomena.

Auditory Streams

To this point, I have been speaking as though auditory images are static entities. In vision, objects might move around, but they rarely fade in or out. Sound is different. Unlike physical objects, sounds require a constant infusion of energy in order to maintain their presence. Unlike physical objects, sounds are volatile: they appear and disappear. The sound of dripping water provides a good example of an intermittent sound. Each "drip" sound can evoke an auditory image, but there is also the sense that the drips are connected together in time: they are perceived as a single continuous sound source. As long as the drips aren't separated too much in time, we experience "something dripping," not just the individual drip sounds. In short, sounds don't just form images; they can also form a sense of continuity across time.

Apart from connecting sounds isolated in time, a sense of continuity may also emerge from physically separate sound sources. When you play a one-octave scale on a harp, you are successively activating eight different sound generators, that is, eight independent strings. But the brain doesn't interpret this as eight separate sound sources that happen to follow one after the other. Instead, the brain interprets the sounds as forming a single line.

In 1971, Jeffrey Campbell coined the useful term *auditory stream* to refer to the subjective image of a sound activity continuing over time.[8] An auditory stream emerges when the brain "connects the dots" and forms a continuous image that persists from sound to sound. As in dripping water, that connection may arise from a single physical source that produces intermittent sounds or, as in the harp scale, that connection may arise from interpreting independent sound sources as though they were produced by a single source. In the perception of music, the phenomenon of auditory streams is critically important. Without this psychological phenomenon, there would be no musical lines, just a succession of disconnected sound events with no feeling of connection. Notice that streams are *auditory* phenomena: the sense of connectedness may be entirely an auditory illusion with no acoustic reality.

As we will see in later chapters, stringent conditions determine whether we hear discrete sounds as connected. That is, we will discover that lines of sound (such as melodies) form only when certain criteria are met. We will also discover that when several concurrent lines of sound are present (as in part-writing), the potential for auditory confusion increases dramatically. We will see that the practice of voice leading provides a sonic toolkit that helps us keep control over auditory confusion.

The Pleasures of Puzzle Solving

Throughout the remainder of this book, I will often compare auditory scene analysis to working on a jigsaw puzzle. The puzzle pieces are the resolved partials that must be assembled into a coherent picture. Ideally, this process will result in a set of auditory images where each image corresponds to an actual acoustic source in the environment. So how does the brain know when it has successfully solved the puzzle? One clue is that there are no left-over puzzle pieces—there are no orphan partials that haven't been assigned to an auditory image. Another clue is that all the puzzle pieces fit snugly with no forced connections: that is, each partial combines easily with the other partials that make up a given auditory image. Yet another clue is that there is no indecision about whether a particular puzzle piece is in its correct location: that is, there is no flip-flopping about assigning a partial first to one image and then to another. A final clue is that the overall picture makes sense. For example, images that existed a moment ago don't inexplicably disappear. If a person is able to *see* the sound sources, then the visual and auditory perceptions can be compared: the auditory scene analysis might be deemed a success if the number and apparent locations of the auditory images do not contradict what a person is seeing.

Reprise

When someone says, "I hear the sound of a squeaking chair," this simple statement masks a remarkably sophisticated mental achievement of packaging and labeling. In this chapter, we have noted that acoustic sources typically vibrate in several *modes* concurrently. Each mode produces a distinctive component sound, called a *partial*—with its own distinctive *frequency*. When a sound source produces two or more partials, the result is referred to as a *complex tone*. The sound energy is communicated through the air to the eardrums, and the flapping eardrums in turn relay this energy through the *ossicles* to the *cochlea*. Inside the cochlea, different frequencies cause different points of maximum stimulation along the *basilar membrane*. If two frequencies activate points that are sufficiently far apart, the corresponding partials will be resolved as separate signals. *Resolved partials* are transmitted up the *auditory nerve* to the brain, where they provide the main puzzle pieces used to assemble *auditory images*.[9] Ideally, this process will result in a set of auditory images where each image corresponds to an actual acoustic source in the environment. Acoustic sources (that exist in the physical world) are translated into auditory images (that exist only in the mental world); *acoustic scenes* (potentially containing several physical sources) are reformulated in the theater of the mind as *auditory scenes*. When an auditory image is maintained over the course of time, the sound events are heard as connected, forming a sense of sonic line or *auditory stream*. Successful *scene analysis* is thought to evoke a sense of pleasure: a mental reward for using all of the puzzle pieces and assembling coherent auditory objects. The success (or lack of success) of this process can lead to positive (or negative) feelings. We will return to discuss the aesthetics of image formation in chapter 16.