

1

The Elements of Sound and Audio Recording

Audio recording is the recording of sound. It is the act of capturing the physical dimensions of sound, then reproducing those dimensions either immediately or from a storage medium (magnetic, solid, electronic, digital), and thereby returning those dimensions to their physical, acoustic state. The process moves from physical sound, through the recording/reproduction chain, and back to physical sound.

The “art” in recording centers around the artistically sensitive application of the recording process. The recording process is being used to shape or create sound as an artistic statement (piece of music), or supporting artistic material. To be in control of crafting the artistic product, one must be in control of the recording process, be in control of the ways in which the recording process modifies sound, and be in control of communicating well-defined creative ideas.

These areas of control of the artistic process all closely involve a human interaction with sound. Inconsistencies between the various states of sound are present throughout the audio recording process. Many of these inconsistencies are the result of the human factor: the ways in which humans perceive sound and interpret or formulate its meanings. In order for material to be under their control, the artist (audio professional) must understand the substance of their material: sound, in all its inconsistencies.

The States of Sound

In audio recording, sound is encountered in three different states. Each of these three states directly influences the recording process, and the creation (or capturing) of a piece of art. These three states are:

1. Sound as it exists physically (having physical dimensions);

2. Sound as it exists in human perception (psychoacoustic conception); sound being perceived by humans after being transformed by the ear and interpreted by the mind (the perceived parameters of sound being human perceptions of the physical dimensions); and
3. Sound as idea; sound as it exists as an aural representation of an abstract or a tangible concept, as an emotion or feeling, or representing a physical object or activity (this is how the mind finds meaning from its attention to the perceived parameters of sound); sounds as meaningful events, capable of communication, provide a medium for artistic expression; sounds hereby communicate, have meaning.

The audio recording process ends with sound reproduced over loudspeakers, as sound existing in its physical state, in air. Often the audio recording process will begin with sound in this physical state, to be captured by a microphone.

Humans are directly involved in all facets of the audio recording process through listening to sound. They evaluate the audio signal at all stages while the recording is being made (including the *recordist*—the person making the recording—and all others involved in the industry), and what is heard by the end listener is the reason for making the recording. Humans translate the physical dimensions of sound into the perceived parameters of sound through the listening process (aural perception).

This translation process involves the hearing mechanism functioning on the physical dimensions of sound, and the transmission of neural signals to the brain. The process is nonlinear and alters the information; the hearing mechanism does not produce nerve impulses that are exact replicas of the applied acoustic energy.

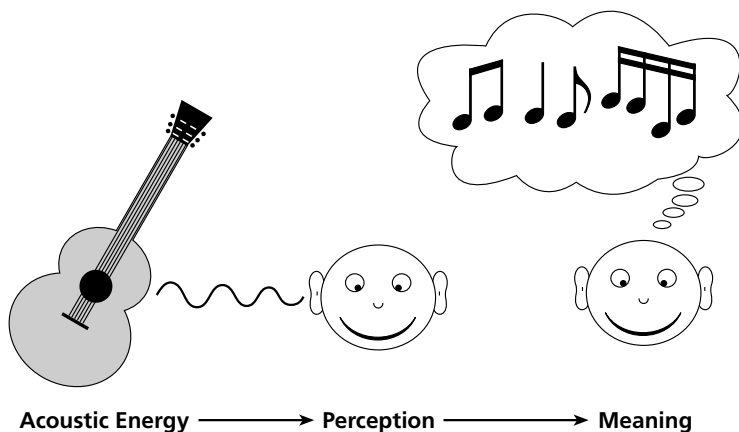


Figure 1-1 Three states of sound: in air, in perception, as message.

Certain aspects of the distortion caused by the translation process are, in general, consistent between listeners and between hearings; they are related to the physical workings of the inner ear or the transfer of the perceived sounds to the mind/brain. Other aspects are not consistent between listeners and between hearings; they relate to the listener's unique hearing characteristics and their experience and intelligence.

The final function occurs at the brain. At a certain area of the cortex, the neural information is processed, identified, consciously perceived, and stored in short-term memory; the neural signals are transferred to other centers of the brain for long-term memory. At this point, the knowledge, experience, attentiveness, and intelligence of the listener become factors in the understanding and perception of sound's artistic elements (or the meanings or message of the sound). The individual is not always sensitive or attentive to the material or to the listening activity, and the individual is not always able to match the sound to their previous experiences or known circumstances.

The physical dimensions (1) are interpreted as perceived parameters of the sound (2). The perceived parameters of sound (2) provide a resource of elements that allow for the communication and understanding of the meaning of sound (and artistic expression) (3).

The audio recording process communicates ideas, and can express feelings and emotions. Audio might take the forms of music, dialog, motion picture action sounds, whale songs, or others. Whatever its form, audio is sound that has some type of meaning to the listener. The perceived sound provides a medium of variables that are recognizable and have meaning, when presented in certain orders or patterns. Sound, as perceived and understood by the human mind, becomes the resource for creative and artistic expression. The artist uses the perceived parameters of sound as the artistic elements of sound, to create and ensure the communication of meaningful (musical) messages.

The individual states of sound as physical dimensions and as perceived parameters will be discussed individually, in the next section. The interaction of the perceived parameters of sound will follow the discussion of the individual parameters. These discussions provide critical information to understanding the breadth of the "artistic elements of sound" in audio recording, presented in the next chapter.

Physical Dimensions of Sound

Five physical dimensions of sound are central to the audio recording process. These physical dimensions are: the characteristics of the sound waveform as (1) *frequency* and (2) *amplitude* displacements, occurring within the continuum of (3) *time*; the fusion of the many frequency and amplitude anomalies of the single sound to create a global, complex

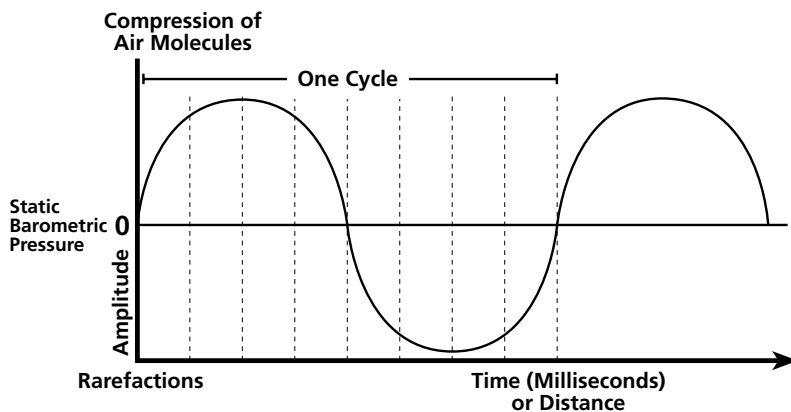


Figure 1-2 Dimensions of the waveform.

waveform as (4) *timbre*; and, the interaction of the sound source (timbre) and the environment in which it exists, create alterations to the waveform according to variables of (5) *space*.

Frequency is the number of similar, cyclical displacements in the medium, air, per time unit (measured in cycles of the waveform per second, or Hz). Each similar compression/rarefaction combination creates a single cycle of the waveform. Amplitude is the amount of displacement of the medium at any moment, within each cycle of the waveform (measured as the magnitude of displacement in relation to a reference level, or decibels).

Timbre

Timbre is a composite of a multitude of functions of frequency and amplitude displacements; it is the global result of all the amplitude and frequency components that create the individual sound. Timbre is the overall quality of a sound. Its primary component parts are the dynamic envelope, spectrum, and spectral envelope.

The *dynamic envelope* of a sound is the contour of the changes in the overall dynamic level of the sound throughout its existence. Dynamic envelopes of individual acoustic instruments and voices vary greatly in content and contours. The dynamic envelope is often thought of as being divided into a number of component parts. These component parts may or may not be present in any individual sound. The widely accepted components of the dynamic envelope are: attack (time), initial decay (time), initial sustain level, secondary decay (time), primary sustain level, and final decay (release time).

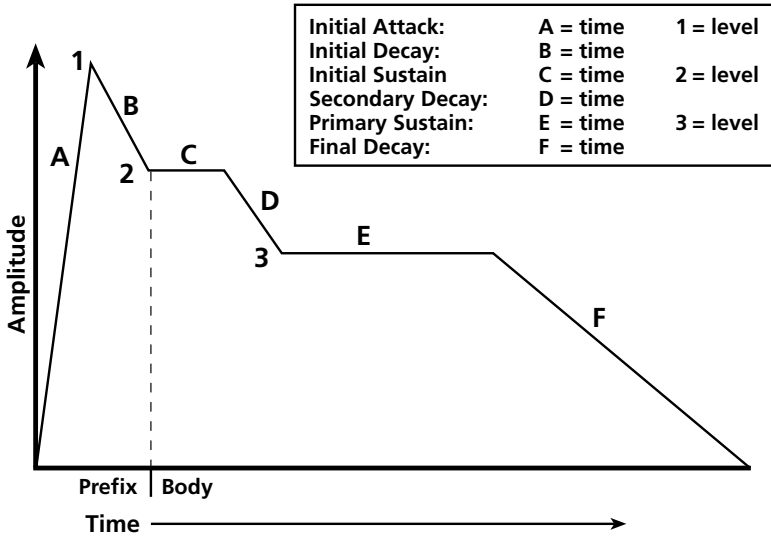


Figure 1-3 Dynamic envelope.

Dynamic envelope shapes other than those created by the above outline are common. Many musical instruments have more or fewer parts to their characteristic dynamic envelope. Further, vocalists and the performers of many instruments have great control over the sustaining portions of the envelope, providing internal dynamic changes to sounds. Musical sounds that do not have some variation of level during the sustain portion of the envelope are rare; the organ is one such exception.

The *spectrum* of a sound is the composite of all of the frequency components of the sound. It is comprised of the fundamental frequency, harmonics, and overtones.

The periodic vibration of the waveform produces the sensation of a dominant frequency. The number of periodic vibrations, or cycles of the waveform is the *fundamental frequency*. The fundamental frequency is also that frequency at which the sounding body resonates along its entire length. The fundamental frequency is often the most prominent frequency in the spectrum, and will often have the greatest amplitude of any component of the spectrum.

In all sounds except the pure sine wave, frequencies other than the fundamental are present in the spectrum. These frequencies are usually higher than the fundamental frequency. They may or may not be in a whole-number relationship to the fundamental. Frequency components of the spectrum that are whole-number multiples of the fundamental are *harmonics*; these frequencies reinforce the prominence of the fundamental

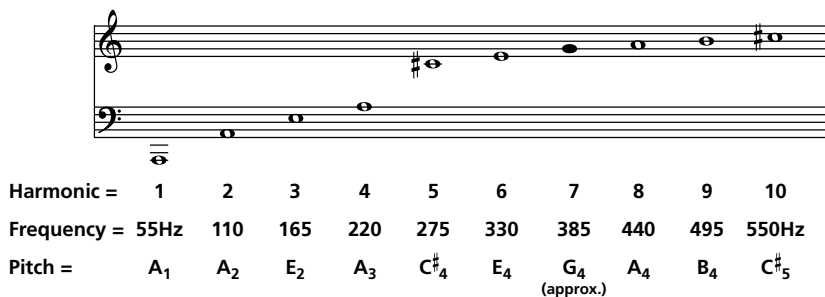


Figure 1-4 Harmonic series.

frequency (and the pitched quality of the sound). Those components of the spectrum that are not proportionally related to the fundamental we will refer to as *overtones*. Traditional musical acoustics studies define overtones as being proportional to the fundamental, but with a different sequence than harmonics (first overtone = second harmonic, etc.); this traditional definition is herein replaced by a differentiation between *partials* that are proportional to the fundamental (harmonics), and those that are not (overtones). This distinction will prove important in the evaluation of timbre and sound quality in later chapters. All of the individual components of the spectrum are partials. Partial (overtone and harmonic) can exist below the fundamental frequency as well as above; they are accordingly referred to as subharmonics and subtones.

For each individual instrument or voice, certain ranges of frequencies within the spectrum will be emphasized consistently, no matter the fundamental frequency. Instruments and voices will have resonances that will strengthen those spectral components that fall within these definable frequency ranges. These areas are called *formants*, *formant regions*, or *resonance peaks*. Formants remain largely constant, and modify the same frequency areas no matter the fundamental frequency. Spectral modifications will be present in all occurrences of the sound source with harmonics or overtones in the formant regions. Formants can appear as increases in the amplitudes of partials that appear in certain frequency bands, or as spectral components in themselves (such as noise transients caused by a hammer striking a string). They can also be associated with resonances of the particular mechanism that produced the source sound. Formants are largely responsible for shaping the characteristic sounds of specific instruments; they allow us to differentiate between the instruments of different manufacturers, or even to tell the difference between two instruments of the same model/make.

A sound's spectrum is comprised primarily of partials that create a characteristic pattern, which is recognizable as being characteristic of a particular instrument or voice. This pattern of spectrum will transpose

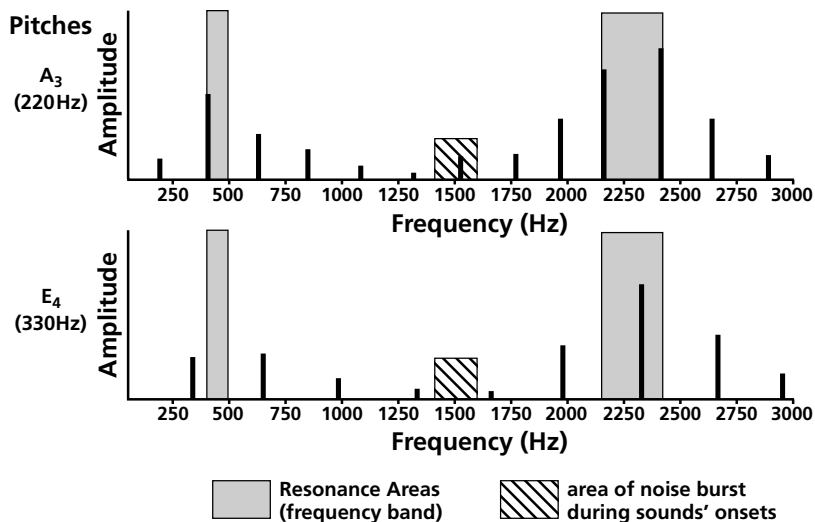


Figure 1-5 Formant regions of two pitches from a hypothetical instrument. Vertical lines represent the partials of the two pitches, placed at specific frequencies and at specific amplitudes.

(change level) with every new fundamental frequency of the same sound source and remain mostly unchanged. This consistent pattern will form a similar timbre at different pitch levels. Formants establish frequency areas that will be emphasized for a particular instrument or voice. These areas will not change with varied fundamental frequency, as they are fixed characteristics (such as resonant frequencies) of the device that created the sound. Formants may also take the form of spectral information that is present in all sounds produced by the instrument or voice.

The frequencies that comprise the spectrum (fundamental frequency, harmonics, overtones, subharmonics and subtones) all have different amplitudes that change independently over the sound's duration. Thus, each partial has a different dynamic envelope. Altogether these dynamic envelopes of all the partials make up the *spectral envelope*. The spectral envelope is the composite of each individual dynamic level and dynamic envelope of all of the components of the spectrum.

The component parts of timbre (dynamic envelope, spectrum, and spectral envelope) display strikingly different characteristics during different parts of the duration of the sound. The duration of a sound is commonly divided into two time units: the *prefix* or *onset*, and the *body*. The initial portion of the sound is the prefix or onset; it is markedly different from the remainder of the sound, the body. The time length of the prefix is usually determined by the way a sound is initiated, and is often the same time unit as the initial attack. The actual time increment of the prefix may be anywhere from a few microseconds to 20–30 milliseconds.

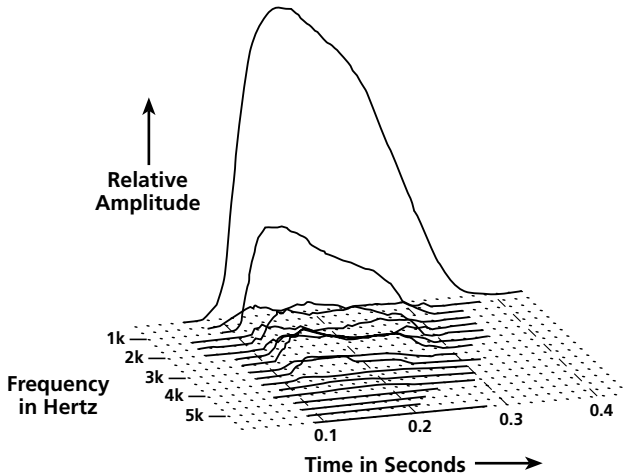


Figure 1-6 Spectral envelope.

The prefix is defined as the initial portion of the sound that has markedly different characteristics of dynamic envelope, spectrum, and spectral envelope, than the remainder of the sound. The body of the sound is usually much longer in duration than the prefix. (See Figure 1-3.)

Space

The interaction of the sound source (timbre) and the environment, in which it is produced, will create alterations to sound. These changes to the sound source's sound quality are created by the acoustic space. The nature of these alterations are directly related to (1) the characteristics of the acoustic space in which the sound is produced and (2) the location of the sound source within the environment.

Space-related sound measurements must be performed at a specific, physical location. The measurements are calculated from the point in space where a receptor (perhaps a microphone or a listener) will capture the composite sound (the sound source within the acoustic space). The location of the listener (or other receptor) becomes a reference in the measurement of the acoustic properties of space.

The aspects of space that influence sound in audio recording are: (1) the *distance* of the sound source to the listener, (2) the *angle* of the sound source to the listener, (3) the *geometry of the environment* in which the sound source is sounding, and (4) the *location* of the sound source *within the host environment*.

The environment in which the sound source is sounding is often referred to as the *host environment*. Within the host environment, sound

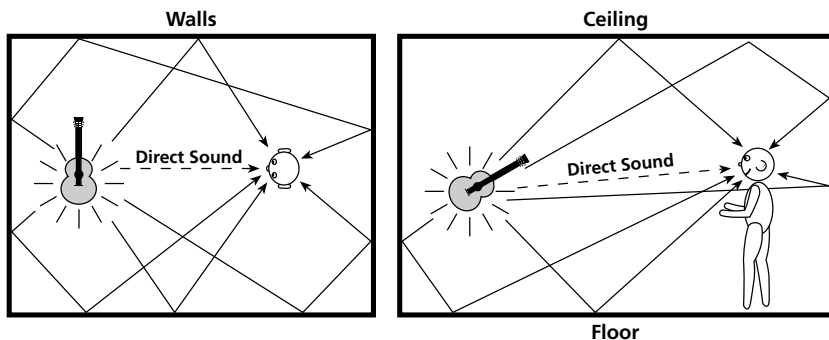


Figure 1-7 Paths of reflected sound within an enclosed space.

will travel on a direct path to the listener (as *direct sound*) and sound will bounce off reflective surfaces before arriving at the listener (as *reflected sound*).

Reverberant sound is a composite of many reflections of the sound arriving at the listener (or microphone) in close succession. The many reflections that comprise the reverberant sound are spaced so closely that the individual reflections cannot be measured; these many reflections are therefore considered as a single entity. As time progresses, these closely spaced reflections become more closely spaced and of diminishing amplitude, until they are no longer of consequence. *Reverberation time* (often referred to as RT60) is the length of time required for the reflections to

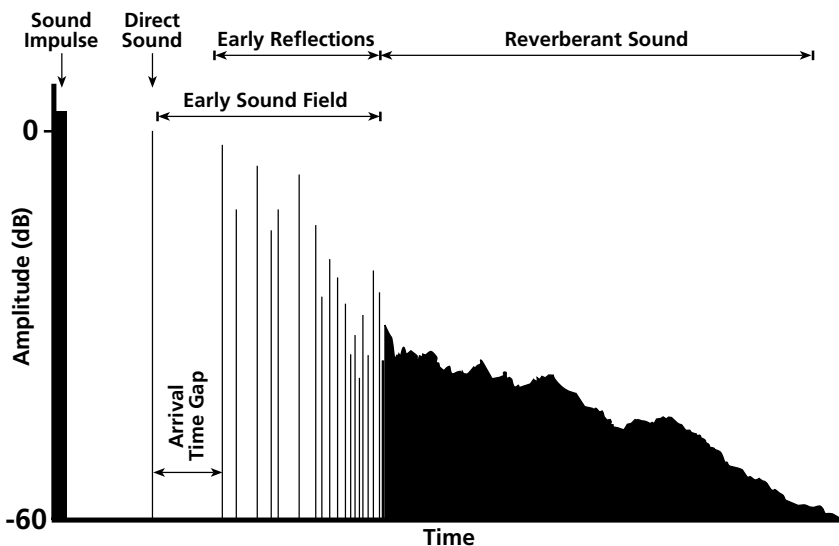


Figure 1-8 Reflected sound.

reach an amplitude level of 60 dB lower than that of the original sound source.

Early reflections are those reflections that arrive at the ear or microphone within around 50 milliseconds of the direct sound. As a collection, the reflections that arrive at the receptor within the first 50 milliseconds after the arrival of the direct sound comprise the *early sound field*.

Varying the *distance* of the sound source from the receptor (ear or microphone) alters the sound at the receptor. The sound at the receptor will be a composite of the direct sound and the reflected sounds (reverberation and early sound field). The composite sound at the receptor is affected by the distance of the sound source from the receptor in two ways: (1) low amplitude portions of the sound's spectrum (usually high frequencies) are lost with increasing distance of the sound source to the receptor and (2) reflected sound increases in prominence to the direct sound as distance increases. Figure 1-9 illustrates the loss of *timbral detail* (the subtle aspects and changes in the content of a sound's timbre, also called *definition of timbre*) with increasing distance as well as the change of the proportion of direct to reflected sound.

The characteristic changes to the composite sound caused by the geometry of the host environment and by the location of the sound source within the host environment, are also influenced by the changes caused by distance.

These two dimensions of the relationship of the sound source to its acoustic space may alter the composite sound in four, additional ways: (1) timbre differences between the direct and reflected sounds; (2) time differences between the arrivals of the direct sound, the initial reflections, and the reverberant sound; (3) spacing in time of the early reflections; and (4) amplitude differences between direct and reflected sounds.

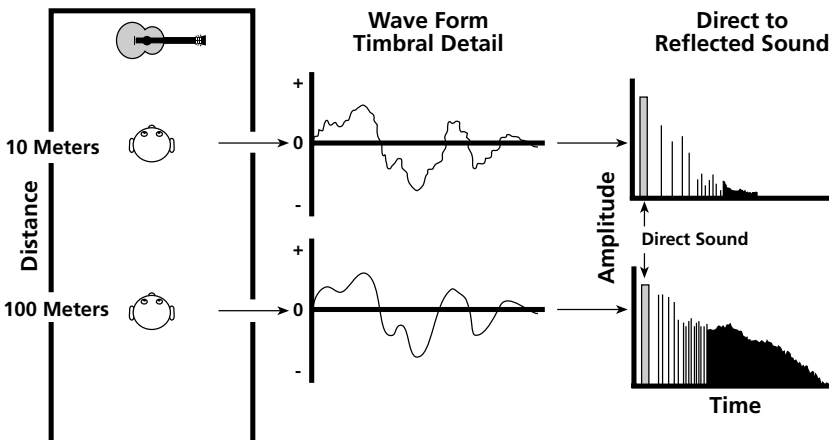


Figure 1-9 Changes in sound with distance.

The geometry of the host environment greatly influences the content of the composite sound. The dimensions and volume of the space, the angles of boundaries (walls, floors, ceilings), materials of construction, and the presence of openings (such as windows) and large objects within the space will all alter the composite sound. Host environments cover the gamut of all the physical spaces and open areas that create our reality (from small room, to a large concert hall; from the corridor of a city street, to an open field, etc.).

Unique sequences of reflected sound are created when a sound is produced within an environment, and sequences are shaped by the location of the sound source within the host environment. These unique sequences contain patterns of reflections that are defined by the spacing of reflections over time and the amplitudes of the reflections. A “rhythm of reflections” exists, and will form the basis of important observations in later chapters. By altering the early time field and reverberant sound, the location of the sound source within the host environment may cause significant alterations to the composite sound at the receptor (ear or microphone).

The location of the sound source within the host environment may strongly influence the composite sound. The amount will be directly related to the proximity of the sound source to the walls, ceiling, floor, openings (such as windows and doors), and large objects reflecting sound within the host environment.

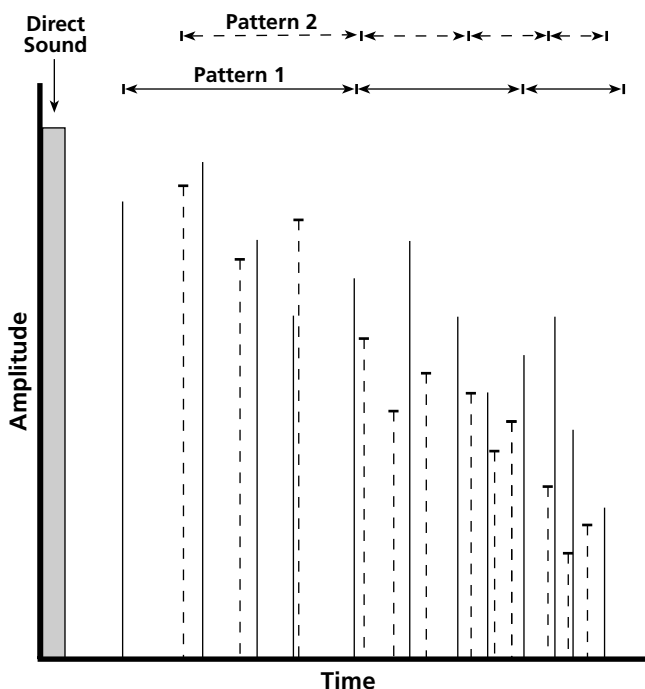


Figure 1-10 Patterns of reflections.

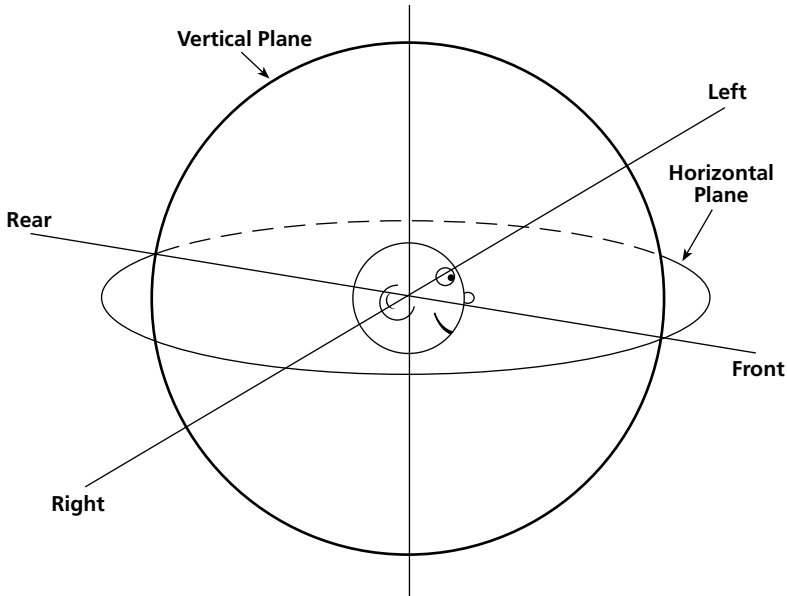


Figure 1-11 Horizontal and vertical planes.

In audio production, the spatial properties of host environments and the location of a sound source within the environment can be generated artificially. It is common to use reverberation units and delays to create environmental cues. These cues may be very realistic representations of natural spaces or they may be environmental characteristics that cannot occur in our physical reality.

The angle of the sound source to the receptor is an important influence in audio recording. The sound source may be at any angle from the receptor (listener or omnidirectional microphone) and be detected. The sound source may be present at any location in the sphere surrounding the receptor. The location is calculated with reference to the 360° vertical and horizontal planes that encompass the receptor.

The angle of the sound source to the receptor may be calculated against the horizontal plane (parallel to the floor), the vertical plane (height), or by combining the two (in a way very similar to positioning locations on a globe). Defining elevation (vertical plane) and direction (left, right, front, rear) can determine the precise location of the sound source within our three-dimensional space by precise increments of degrees.

Angles of source locations on the horizontal plane are captured or generated in audio recording to provide stereo and surround sound. To date, the vertical plane has received little attention in audio because of

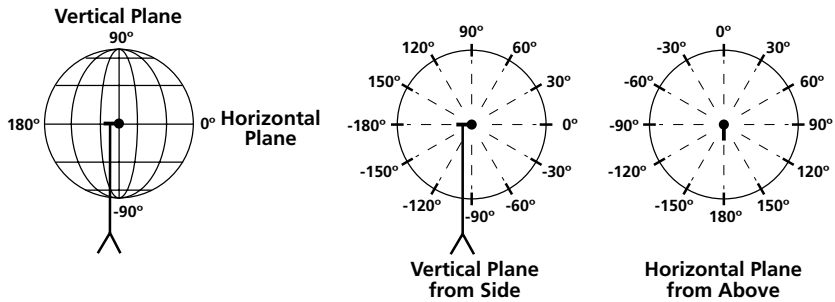


Figure 1-12 Defining sound source angle from a microphone.

playback format difficulties. Recent surround sound advances have produced formats that provide these cues in ways that can strikingly enhance programs.

Perceived Parameters of Sound

The five physical dimensions of sound translate into respective perceived parameters of sound. Sound as it exists in human perception is quite different from sound in its physical state, in air. Our perception of sound is a result of the physical dimensions being transformed by the ear and interpreted by the mind. The perceived parameters of sound are human perceptions of the physical dimensions of sound.

This translation process from the physical dimensions to the perceived parameters is nonlinear, and differs between individuals. The hearing mechanism does not directly transfer acoustic energy into equivalent nerve impulses. The human ear is not equally sensitive in all frequency ranges, nor is it equally sensitive to sound at all amplitude levels. This nonlinearity in transferring acoustic energy to neural impulses causes sound to be in a different state in our perception, than what exists in air. Thus, the physical states of sound captured by recording equipment will be heard by the recordist in ways that may be unexpected, without knowledge of these differences.

Complicating this further, there is no reason to believe any two people actually hear the characteristics of sound in exactly the same way. If it were possible for all conditions for two sounds to be identically sent to two listeners, the two people likely would hear slightly (or strikingly) different characteristics. We only need notice the different ear shapes around us to recognize no two people will pick up acoustic energy in precisely the same way.

Table 1-1

The Physical Dimensions and the Perceived Parameters of Sound

<i>Physical Dimensions</i>	<i>Perceived Parameters</i>
Frequency	Pitch
Amplitude	Loudness
Time	Duration
Timbre (physical components)	Timbre (perceived overall quality)
Space (physical components)	Space (perceived characteristics)

Pitch

Pitch is the perception of the frequency of the waveform. The frequency area most widely accepted as encompassing the hearing range of the normal human spans the boundaries of 20–20,000 Hz (20 kHz), though recent research has proven that humans are sensitive to (if not actually hear) frequencies below and above this range.

Most humans cannot identify specific pitch-levels. Some people have been blessed with, or have developed the ability to recognize specific pitch levels (in relation to specific tuning systems). These people are said to have “absolute” or “perfect pitch.” The ability to accurately recognize pitch-levels is not common even among well-trained musicians.

It is commonly within human ability, however, to determine the relative placement of a pitch within the hearing range. A *register* is a specific portion of the *range*. It is entirely possible to determine, within certain consistent limits of accuracy, the relative register of a perceived pitch-level. This skill can be developed, and accuracy improved significantly.

Humans are able to consistently perform the estimation of the approximate level of a pitch, associating pitch-level with register. With practice, this consistency can be accurate to within a minor third (within three semi-tones). This skill in the “estimation of pitch-level” will be an important part of the method for evaluating sound presented later.

Humans perceive pitch most accurately as relationships. We perceive pitch as the relationship between two or more soundings of the same or related sound sources. We do not perceive pitch as identifiable, discrete increments; we do not listen to pitch material to define the letter-names (increments) of pitches. Instead, humans calculate the distance (or interval) between pitches by gauging the distance between the perceived levels of the two (or more) pitches.

The interval between pitches becomes the basis for all judgments that define and relate the sounds. Thus, melody is the perception of successively sounded pitches (creating linear intervals), and chords are the perceptions of simultaneously sounded pitches (creating harmonic intervals). We often perceive pitch in relation to a reference level (one

predominating pitch that acts as the key or pitch-center of a piece of music), or to a system of organization to which pitches can be related (a tonal system, such as major or minor).

Our ability to recognize the interval between two pitches is not consistent throughout the hearing range. Most listeners have the ability to accurately judge the size of the semi-tone (or minor second, the smallest musical interval of the equal tempered system) within the range of 60 Hz and 4 kHz. As pitch material moves below 60 Hz, a typical listener will have increased difficulty in accurately judging interval size. As pitch material moves above 4 kHz, the typical listener will also experience increased difficulty in accurately judging interval size.

The smallest interval humans can accurately perceive changes with the register and placement of the two pitches creating the interval. The size of the minimum audible interval varies from about 1/12 of a semi-tone between 1–4 kHz, to about 1/2 of a semi-tone (a quarter-tone) at approximately 65 Hz. These figures are dependent upon optimum duration and loudness levels of the pitches; sudden changes of pitch-level are up to 30 times easier to detect than gradual changes. It is possible for humans to distinguish up to 1,500 individual pitch levels by spacing out the appropriate minimum-audible intervals, throughout the hearing range.

With all factors being equal, the perception of harmonic intervals (simultaneously sounding pitches) is more accurate than the perception of melodic intervals (successively sounded pitches). Up to approximately 500 Hz, melodic and harmonic intervals are perceived equally well. Above 1 kHz, humans begin to be able to judge harmonic intervals with greater accuracy than melodic intervals; above 3,500 Hz, this difference becomes pronounced.

Loudness

Loudness is the perception of the overall excursion of the waveform (amplitude). Amplitude can be physically measured as a sound pressure level. In perception, loudness level cannot be accurately perceived in discrete levels.

Loudness is referred to in relative values, not as having separate and distinct levels of value. Traditionally loudness levels have been described by analogy (“louder than,” “softer than,” etc.) or by relative values (“soft,” “medium loud,” “very soft,” “extremely loud,” etc.). Humans compare loudness levels and conceive loudness levels as being “louder than” or “softer than” the previous, succeeding, or remembered loudness level(s).

A great difference exists between loudness as perceived by humans and the physical amplitude of the sound wave. This difference can be quite large at certain frequencies. In order for a sound of 20 Hz to be audible, a sound pressure level of 75 dB must be present. At 1 kHz, the human ear

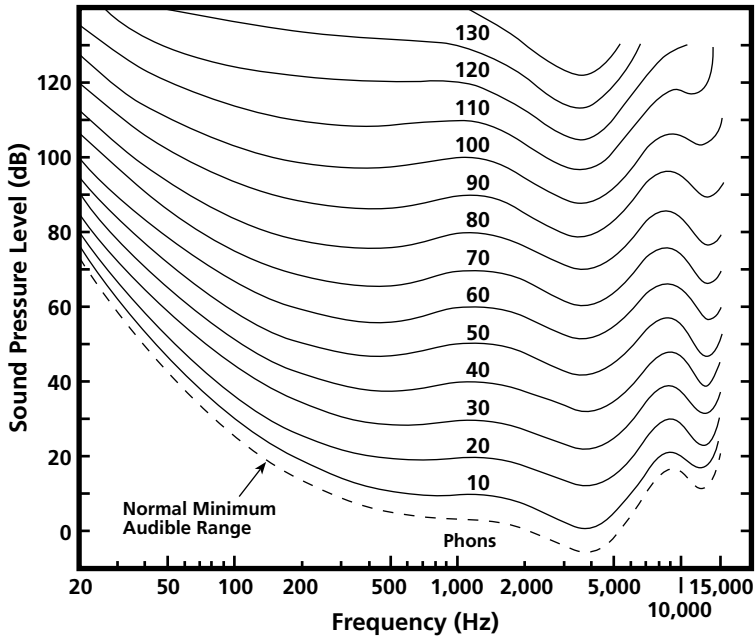


Figure 1-13 Equal loudness contour.

will perceive the sound with a minute amount of sound pressure level, and at 10 kHz a sound pressure level of approximately 18 dB is required for audibility. The unit *phon* is the measure of perceived loudness established at 1 kHz, based on subjective listening tests.

The nonlinear frequency response of the ear and the fatigue of the hearing mechanism over time play an important role in further inaccuracies of the human perception of loudness. With sounds of long durations and steady loudness level, loudness will be perceived as increasing with the progression of the sound until approximately 0.2 seconds of duration. At that time, the gradual fatigue of the ear (and possibly shifts of attention by the listener) will cause perceived loudness to diminish.

As loudness level of the sound is increased, the ear requires increasingly more time between soundings, before it can accurately judge the loudness level of a succeeding sound. We are unable to accurately judge the individual loudness levels of a sequence of high-intensity sounds as accurately as we can judge the individual loudness levels of mid- to low-intensity sounds; the inner ear itself requires time to reestablish a state of normalcy, from which it can accurately track the next sound level.

As a sound of long duration is being sustained, its perceived loudness level will gradually diminish. This is especially true for sounds with high sound pressure levels. The ear gradually becomes desensitized to the

loudness level. The physical masking (covering) of softer sounds and an inability to accurately judge changes in loudness levels will result from the fatigue. When the listener is hearing under listening fatigue, slight changes of loudness may be judged as being large. Listening fatigue may desensitize the ear's ability to detect new sounds at frequencies within the frequency band (frequency area) where the high sound pressure level was formerly present.

Duration

Humans perceive time as *duration*. Sound durations are not perceived individually. We cannot accurately judge time increments without a reference time unit. Regular reference time units are found in musical contexts, and rarely in other types of human experiences. Even the human heartbeat is rarely consistent enough to act as a reliable reference. The underlying metric pulse of a piece of music does, however, allow for accurate duration perception. This accuracy cannot be achieved in any other context of the human experience.

In music, the listener remembers the relative duration values of successive sounds, in a similar process to that of perceiving melodic pitch intervals. These successive durations create musical rhythm. The listener calculates the length of time between when a sound starts and when it ends, in relation to what precedes it, what follows it, what occurs simultaneously with it, and what is known (what has been remembered). Instead of calculating an interval of pitch, the listener proceeds to calculate a span of time, as a durational value.

Metric Grid

A *metric grid*, or an underlying pulse, is quickly established in the perception of the listener, as a piece of music unfolds. This creates a reference pulse against which all durations can be defined. The listener is thereby able to make rhythmic judgments in a precise and consistent manner. The equal divisions of the grid allow the listener to compare all durations, and to calculate the pulse-related values of the perceived sounds. Durations are calculated as being in proportion to the underlying pulse: at the pulse, half pulses, quarter of the pulse, double the pulse, etc.

In the absence of the metric grid, durational values cannot be accurately perceived as proportional ratios. Humans will not be able to perceive slight differences in duration when a metric grid is not established.

The listener is only able to establish a metric grid within certain limits. Humans will be able to accurately utilize the metric grid between 30 to 260 pulses per minute. Beyond these boundaries, the pulse is not perceived as the primary underlying division of the grid. The human mind

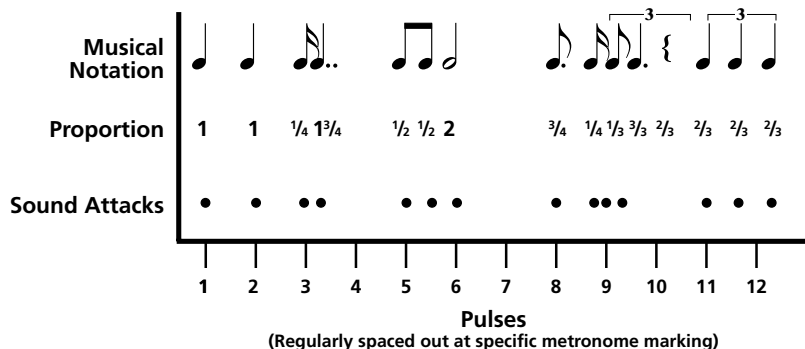


Figure 1-14 Metric grid.

will replace the pulse with a duration of either one-half or twice the value, or the listener might become confused and unable to make sense of the rhythmic activity.

The metric grid is the dominant factor in our perception of tempo, as well as musical rhythm. In most instances, the metric grid itself represents the steady pulsation of the tempo of a piece of music.

Time

The listener's *perception of time* plays a peripheral role in the perception of rhythm. Time perception is significant to the perception of the global qualities of a piece of music, and to the estimation of durations when a metric grid is not present in the music. The global qualities of aesthetic, communicative, and extra-musical ideas within a piece of music are largely dependent on the living experience of music; on the passage of musical materials across the listener's time perception of their existence.

Time perception is distinctly different from duration perception. The human mind makes judgments of elapsed time based on the perceived length of the present. The length of time humans perceive to be "the present" is normally two to three seconds, but might be extended to as much as five seconds and beyond.

The "present" is our window of consciousness, through which we perceive the world, and listen to sound. We are at once experiencing the moment of our existence, evaluating the immediate past of what has just happened and anticipating the future (projecting what will follow the present moment, given our experiences of the recently passed moments, and our knowledge of previous, similar events).

Human time judgments are imprecise. The speed at which events take place and the amount of information that takes place within the "present" greatly influences time judgments. The amount of time perceived to have passed will change to conform to the number of events

experienced within the present; the listener will estimate the amount of time passed in relation to the number of experiences during the present, and make time judgments accordingly.

Time judgments are greatly influenced by the individual listener's attentiveness and interest in what is being heard. If the material stimulates thought within the listener, the event will seem shorter; if the listener finds the listening activity desirable in some way, the experience will seem to occupy less time than would an undesirable experience of the same (or even shorter) length. Expectations caused by, boredom with, interest in, contemplation of, and even pleasure caused by music, alter the listener's sense of elapsed time.

The time length of a piece of music (or any time-based art form, such as a motion picture) is separate and distinct from clock time. A lifetime can pass in a moment, through the experience of a work of art. A brief moment of sound might elevate the listener to extend the experience to an infinity of existence.

Timbre in Perception

The overall quality of a sound, its *timbre*, is the perception of the mixture of all of the physical aspects that comprise a sound. Timbre is the global form, or the overall character of a sound, which we can recognize as being unique.

The overall form (timbre) is perceived as the states and interactions of its component parts. The physical dimensions of sound discussed above, are perceived as dynamic envelope, spectral content, and spectral envelope (perceived values, not physical values). The perceived dimensions are interpreted, and shape an overall quality, or conception of the sound.

Humans remember timbres as entities, as single objects having an overall quality (that is comprised of many unique characteristics), and sometimes as having meaning in themselves (as a timbre can bring with it associations in the mind of the listener). We recognize the sounds of hundreds of human voices because we remember their timbres. We remember the timbres of a multitude of sounds from our living experiences. We remember the timbres of many musical instruments and their different timbres as they are performed in many different ways.

The global quality that is timbre allows us to remember and recognize specific timbres as unique and identifiable objects.

Humans have the ability to recognize and remember a large number of timbres. Further, listeners have the ability to scan timbres and relate unknown sounds to sounds stored in the listener's long-term memory. The listener is then able to make meaningful comparisons of the states and values of the component parts of those timbres. These skills will serve as meaningful points of departure for the method for evaluating timbre in Part 2.

Sufficient time is required for the mind to process the many characteristics of a sound in order to recognize and understand its overall image. The time required to perceive the component parts of timbre vary significantly with the complexity of the sound, and the listener's previous knowledge of the sound. For rather simple sounds, the time required for accurate perception is approximately 60 milliseconds. As the complexity of the sound is increased, the time needed to perceive the sound's component parts will also increase. All sounds lasting less than 50 milliseconds are perceived as noise-like, since a specific timbre cannot be identified at that short a duration; exceptions occur when the listener is well acquainted with the sound, and the timbre can be recognized from this small bit of information.

The partials of the timbre's spectrum fuse to create the impression of a single sound. Although many frequencies are present, the tendency of our perception is to combine them into one overall texture. We fuse partials that are harmonically related to the fundamental frequency, as well as overtones that are distantly related to the fundamental, into a single impression.

It is especially important for the recording professional to note, fusion can also occur between two separate timbres (two individual sound sources) if the proper conditions are present. Timbres that are attacked simultaneously, or are of a close harmonic relationship to each other, are most likely to fuse into the perception of a single sound. The more complex the individual sound, the more likely that fusion will not occur. Furthermore, if the listener recognizes one of the timbres, fusion will be far less likely to occur. Also related to recognizing timbre, synthesized sounds are more likely to fuse with other sounds than are known sounds of an acoustic origin.

Spatial Characteristics

The perception of the *spatial characteristics* of sound is the impression of the physical location of a sound source in an environment, together with the modifications the environment itself places on the sound source's timbre.

The perception of *space* in audio recording (reproduction) is not the same as the perception of space of an acoustic source in a physical environment. In an acoustic space, listeners perceive the location of sound in relation to the three-dimensional space around them: distance, vertical plane, and horizontal plane. Sound is perceived at any possible angle from the listener, and sound is perceived at a distance from the listener; both of these perceptions involve an evaluation of the characteristics of the sound source's host environment.

In audio recording, illusions of space are created. Sound sources are given spatial characteristics through the recording process and/or through signal processing. This spatial information is intended to complement the timbre of the sound source. The spatial characteristics may simulate particular known, physical environments or activities, or be intended to provide spatial cues that have no relation to our reality. In theory, all of the interactions of the sound with its host environment are captured with, or can be simulated and applied to, the sound source; upon playback through two or more loudspeakers, the spatial cues are reproduced.

Playback Environment

Recordings and their sound sources (combined with their spatial characteristics) are heard through two (or more) loudspeakers. The loudspeakers themselves are placed in and interact with, a playback environment—such as a living room or automobile. The playback environment is nearly always quite unrelated to the spaces on the recording. Thus, spatial characteristics applied to the sound source are ultimately perceived by the listener after they have been altered by the characteristics of loudspeakers, altered by the interaction of the loudspeakers and the environment (caused by placement of loudspeakers within the playback environment), and altered by the playback environment itself. The listener perceives the reproduced spatial characteristics of the sound source within the three-dimensional space of their listening environment (headphone monitoring is not a solution, as will be later discussed).

To accurately perceive the spatial information of an audio recording, the listening environment must be acoustically neutral, and the listener must be carefully positioned within the environment and in relation to the loudspeakers. The listening environment (including the loudspeakers) should not place additional spatial cues onto the reproduced sound.

Perceived Spatial Relationships and Current Sound Reproduction

Humans perceive spatial relationships (1) as the location of the sound source being at an angle to the listener (above, below, behind, to the left, to the right, in front, etc.), (2) as the location of the sound source being at distance from the listener, (3) as an impression of the type, size, and properties of the host environment, and at times (4) as the location of the sound source being within an environment.

These perceptions are transferred into the recording medium, to provide a realistic illusion of space, with one major exception. The angular location is severely restricted in audio reproduction, as compared to human perceptual abilities. Currently used audio playback formats can only accurately and consistently reproduce localization cues on the horizontal plane, and then only slightly beyond the loudspeaker array in

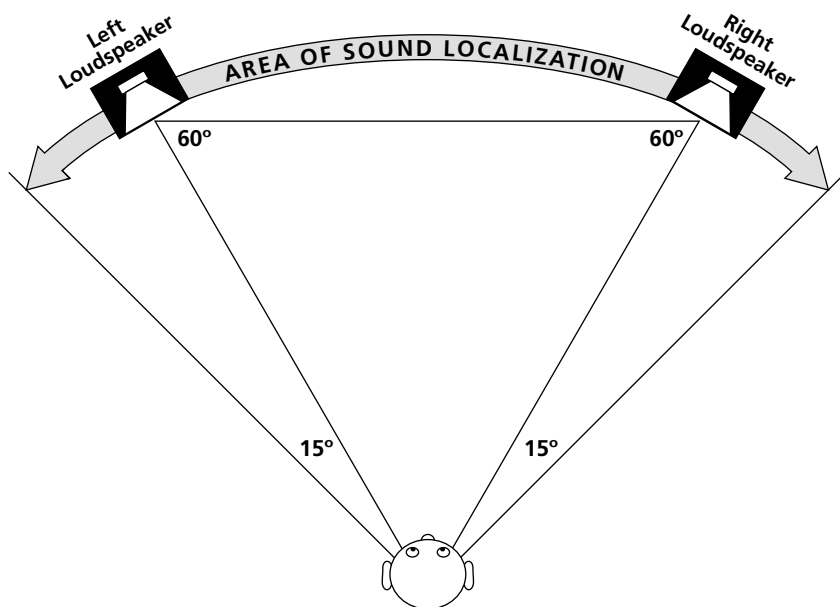


Figure 1-15 Area of sound localization in two-channel audio playback.

stereo recordings. The three-dimensional space of our reality is simulated (with dubious accuracy) in the two dimensions of audio recording.

Much research and development is taking place attempting to extend the localization of sound sources to behind the listener and to the vertical plane, as well as to provide a more realistic reproduction of distance and environmental cues. Significant advances are being made. Surround sound is now widely accepted and provides reproduction around the listener that can be mostly stable and accurate. Control of sound localization on the vertical plane and a more complete simulation of environmental characteristics are not, however, presently feasible, though the technology of the future will likely address these areas as well.

The following discussions of the perception of the spatial dimensions of reproduced sound refer to common two-channel, stereo and to surround sound systems. These concepts will transfer to other systems such as binaural recordings, holophonics, and others; but with different boundaries of the limits of perception inherent with each particular format.

Localization of Direction

The ability for humans to localize sounds is one of the survival mechanisms we have retained from our ancient past. This ability has been developed throughout our evolution; we have learned to perceive the

direction of even those sounds that our hearing mechanism has difficulty processing.

Humans use differences in the same sound wave appearing at the two ears for the accurate *localization of direction*. Interaural time differences (ITD) are the result of the sound arriving at each ear at a different time. A sound that is not precisely in front or in back of the listener will arrive at the ear closest to the source before it reaches the furthest ear. These time differences are sometimes referred to as phase differences. The sounds arriving at each ear are almost identical during the initial moments of the sound, except the sound at each ear is at a different point in the waveform's cycle (and might contain minute spectral differences).

Interaural amplitude differences (IAD) work in conjunction with ITD in the localization of the direction of the sound source. IAD are also referred to as interaural spectral differences. IAD is the result of sound pressure level differences at high frequencies present at the two ears. The head of the listener, which blocks certain frequencies from the furthest ear (when the sound is not centered), causes the interaural spectral differences. This occurrence has been termed the "shadow effect." Interaural amplitude differences (IAD) will at times consist solely of amplitude differences between the two ears, with the spectral content of the waveform being the same at both ears.

The sound wave is almost always different at each ear. The differences between the sound waves may be time/phase-related, amplitude/spectrum-related, and/or spectrum differences. These differences in the waveforms are essential to the perception of the direction of the sound source. In addition, these same cues play a major role in perceiving the characteristics of host environments.

Up to approximately 800 Hz, humans rely on ITD for localization cues. Phase differences are utilized for localization perception up to about 800 Hz, as amplitude appears to be the same at both ears.

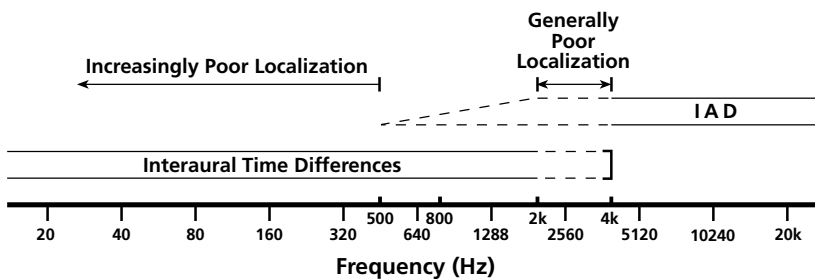


Figure 1-16 Frequency ranges of localization cues.

Between 800 Hz and about 2 kHz both phase and amplitude differences are present between the two ears. IAD and ITD are both used for the perception of direction in this frequency range, with amplitude differences becoming significant around 1,250 Hz.

In general, time/phase differences seem to dominate the perception of direction up to about 4 kHz. Although IAD are present, ITD dominates the perception of direction between 2 kHz and 4 kHz. Humans have poor localization ability for sounds in this frequency band.

Above 4 kHz, interaural amplitude differences (IAD) are the cues that determine the perception of location. Localization ability improves at 4 kHz and is quite accurate throughout the upper registers of our hearing range.

Recent studies have revealed the human body also generates physical cues for localization. The chest, head, shoulders, and outer ears all affect sounds of various wavelengths in different ways. Our body parts' different sizes and their different angles to the hearing canal create a very complex source of reflected and diffracted sound waves. These waves all lead to important interaural spectrum differences between the two ears. These differences are created by a comb-filter effect, comprised of minute cancellations and reinforcements of frequencies. The brain processes these subtle differences to aid in identifying a sound source's direction.

As we have seen, humans do not perceive direction accurately at all frequencies. Below approximately 500 Hz, our perception of the angle of the sound source becomes increasingly inaccurate, to the point where sounds seem to have no apparent, focused location. An area exists around 3 kHz where localization is also poor; wavelength similarities between the distance between the two ears and those of the frequencies around 3 kHz cause interaural time/phase differences to be unstable.

Humans have a well-refined ability to localize sounds in the approximate frequency areas: 500 Hz–2 kHz, and 4 kHz to upper threshold of hearing (whatever that might be). Within these areas, the minimum discernible angle is approximately one to two degrees, with less accuracy at the sides and back than in the front. Sounds that have fundamental frequencies outside of these frequency areas, but that have considerable spectral content within these bands, will also be localized quite accurately.

Interaural spectral differences occur throughout the frequency spectrum. While they may be subtle, it appears they are important for the localization of objects in frequency ranges where IAD and ITD are ineffective. The makeup of these interaural spectral differences will inherently be unique to individuals, as (obviously) no two people are the same size and shape, and no two outer ears are alike.

Pinna is the name of the outer ear. This part of our anatomy (an elaborately shaped piece of cartilage) plays several important roles in our perception of direction. The pinna gathers sound and funnels it into the ear canal. As the ridges of the outer ear reflect sound into the ear, the ridges

introduce small time delays between their reflections and the direct sound that travels directly to the ear canal. These small time delays vary according to sound source location, and are important components of the interaural spectral differences described above. The pinna and the delays it generates aid us in differentiating between sounds arriving from the front and those arriving from the rear.

Resonances also appear to be excited in the outer ear. These also alter the frequency response of the sound source in predictable ways that vary between individuals. The brain learns these patterns of spectral changes to assist in localization. Pinna cues play significant roles in direction perception even though each individual has a unique ear-shape, and the resultant spectrum changes are equally unique. Location cues based on spectrum are thus not universal, but unique to the individual and are learned.

Pinnae serve a critical function in front to back localization. When sound arrives at the head from the rear, ridge reflections are not generated. The pinna actually blocks the direct sound from reaching the hearing canal and its ridges when sounds are generated beyond 130° from the front center. The pinna allows us to perceive the sound source as being generated to our rear because of the absence of spectral differences.

It is interesting to note, our distance and location judgments are not as accurate to the sides and the rear. The absence of this spectral information generated and collected by the outer ear may well play a role.

We actually move our heads involuntarily to assist in locating sounds—especially those sounds that are not in front. In moving our heads to remove location confusion we bring the source into our front listening field, and thus reintroduce the IAD, ITD and spectral differences of the pinnae. We also instinctively seek to bring the source into visual view, which eliminates all ambiguity for acoustic sources—but not for phantom images.

Front-back hearing is only partially understood. Little relevant research in spatial perception of sounds arriving from the rear and the sides is available. Certainly more will need to be accomplished before we are able to more thoroughly understand this area. It is increasingly important, however, that we understand how we perceive sound arrival from the rear and the sides, and the different qualities of those sounds, if we are to fully understand and control the differences between surround sound and stereo.

Distance Perception

Distance perception has not been studied thoroughly. The following information is well documented, and it is likely that numerous subtleties will be discovered in the future.

Two impressions lead to the perception of the distance of a sound source from the listener: (1) the ratio of the amount of direct sound to reverberant sound, and (2) the primary determinant, the loss of low amplitude (usually high frequency) partials from the sound's spectrum with increasing distance (*definition of timbre* or *timbral detail*). Both of these functions rely on the listener's knowledge of the sound's timbre for accurate perception of distance-location. While sound pressure decreases with distance, loudness itself does not factor into distance location perception.

Low energy spectral information is lost with the compressions and rarefactions of the waveform over distance. Some information is simply absorbed by the atmosphere due to air friction. This leads to the listener's determination of the level of timbral detail (*definition of timbre*) that is the major factor in distance perception.

Some timbre-related distance information results from waveform travel and the speed of sound. As high frequencies travel slightly faster than low frequencies, the spectrum of the sound is altered with increasing distance. The partials of complex sounds will become increasingly out of phase with the fundamental frequency, and between themselves, the longer the propagation of the sound.

The percentage of direct sound decreases while the percentage of reflected sound increases, as the source moves from the listener. This pertains to enclosed spaces only. This ratio of direct to reflected sound may play a significant role in distance perception when timbral detail cues are diminished or are unknown.

The listener must know the timbre of a sound in order to recognize missing timbral detail. If the sound is unknown or not recognized, the listener cannot identify the potential loss of low energy components from its spectrum. With knowledge of the sound source, the listener will be able to calculate how much low energy information is missing and thus be able to determine the general amount of distance between them and the object.

Knowledge of the timbre of the sound source will assist the listener in recognizing the absence of spectral information and/or perceiving the reiterations of the direct sound and the reverberant sound. These perceptions will provide the listener with the required information to judge distance. The previous experiences and listening skills of the listener will play a major role in the accuracy of judgments made.

Without prior knowledge of the timbre of a sound, perception of distance location is considerably less accurate, if not impossible.

Related to the ratio of direct to reflected sound, the time difference between the ceasing of the direct sound and the ceasing of the reverberant energy will increase with distance. Through *temporal fusion* we perceive the reverberant sound as being a part of the direct sound. This creates a single impression of the sound in its environment (referred to as the

composite sound, above). As distance increases, temporal fusion begins to diminish and the ending of the direct sound and the continuance of the reverberant energy become more prominent.

Perception of Environmental Characteristics

The perceptions of the *characteristics of the host environment* and the *placement of the sound source within the host environment* are also dependent upon the ratio of direct to reflected sound and the loss of low-level spectral components with increasing distance. In addition, the characteristics of the host environment are perceived through (1) the time difference between the arrival of the direct sound and the arrival of the initial reflections, (2) the spacing in time of the early reflections, (3) amplitude differences between the direct sound and all reflected sound (the individual initial reflections and the reverberant sound), and (4) timbre differences between the direct sound, the initial reflections, and the reverberant sound.

The time delay between the direct and the reflected sounds is directly related to (1) the distance between the sound source and the listener, (2) the distance between the sound source and the reflective surfaces (which send the reflected sound to the listener), and (3) the distance of the reflective surfaces from the listener. These three physical distances also create the patterns of time relationships (the rhythms) of the early reflections.

Early reflections arrive at the listener within 50 milliseconds of the direct sound. These early reflections comprise the *early sound field*. The early sound field is composed of the first few reflections that reach the listener before the beginning of the diffused, reverberant sound (see Figure 1-8). Many of the characteristics of a host environment are disclosed during this initial portion of the sound. The early sound field contains information that provides clues as to the size of the environment, the type and angles of the reflective surfaces, even the construction materials and surface coverings of the space.

Humans have the ability to accurately judge the size and characteristics of the host environments of sound sources. This is accomplished by evaluating the sound qualities of the environment. Humans experience and remember the sound qualities of a great many natural environments in much the same way as we recognize and remember timbres. Further, we have the ability to compare the sound qualities of new environments we encounter to our memories of environments we have previously experienced.

The listening skill needed to evaluate and recognize environmental characteristics can be developed to a highly refined level. Some people who work regularly with acoustical environments develop these listening skills to a point where many can perceive the dimensions and volume of an environment, its surface coverings, or even openings within the space (doors, windows, etc.).

Interaction of the Perceived Parameters

The perception of any parameter of sound is always dependent upon the current states of the other parameters. Altering any of the perceived parameters of sound will cause a change in the perceived state of at least one other parameter.

The parameters of sound interact, causing the perception of the state of one parameter to be altered by the state of another. Certain occurrences of these interactions were noted under individual perceived parameters. The following are additional examples of note and are separated for clarity.

Duration for Pitch Perception

Sufficient duration is required for the ear to perceive pitch. If the duration is too short, the sound will be perceived as having indefinite pitch, as being noise-like. The time necessary for the mind to determine the pitch of a sound is dependent on the frequency of the sound. Sounds lower than 500 Hz and sounds higher than 4 kHz require more time to establish pitch quality, than sounds pitched between 2 kHz and 4 kHz where pitch perception is most acute. At the extremes of the hearing range, pitch quality may require as much as 60 milliseconds to become established.

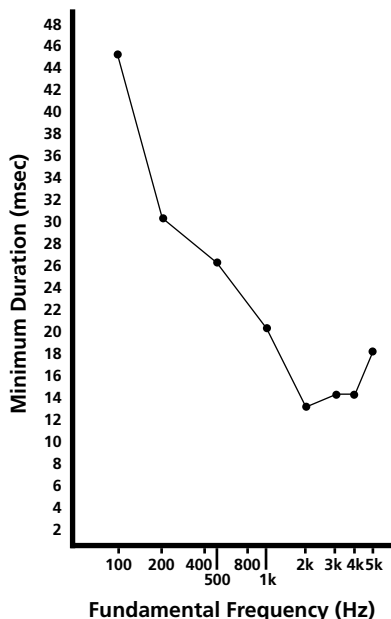


Figure 1-17 Minimum durations for pitch perception at select frequencies.

The length of time required to establish a perception of pitch will also depend on the sound's attack characteristics and its spectral content (its timbre). Sounds with complex (but mostly harmonic) spectra and sounds with short attack times will establish a perception of pitch sooner than other sounds.

Loudness and Pitch Perception

Loudness will influence the perception of pitch, as humans will perceive a change of pitch with a change of loudness (dynamic) level. The level 60 dB SPL (or about the loudness level of normal conversation) is considered to be a threshold where increases or decreases in loudness affect pitch perception oppositely. Above 60 dB, for sounds below 2 kHz a substantial increase in loudness level will cause an apparent lowering of the pitch level; the sound will appear to go flat, although no actual change of pitch level has occurred. Similarly, a substantial increase in the dynamic (loudness) level of a pitch above 2 kHz will cause the sound to appear to go sharp; an impression of the raising of the pitch level is created, although the actual pitch level of the sound has remained unaltered. Below 60 dB an increase in loudness will cause sounds below 2 kHz to be perceived as getting sharp and sounds above 2 kHz perceived as going flat.

Loudness and Time Perception

Loudness level can influence perceived time relationships. When two sounds begin simultaneously, they will appear to have staggered entrances if one of the two sounds is significantly louder than the other. The louder sound will be perceived as having been started first.

Perceived loudness level is often distorted by the speed at which information is processed. When a large number of sounds occur in a short period of time, the listener will perceive those sounds as having a higher loudness level than sounds of the same sound pressure level, but that are distributed over a longer period of time. This distortion of loudness level is caused by the amount of information being processed within a specific period of time (the time period is related to the perceived length of the present).

Loudness Perception Altered by Duration and Timbre

Duration can distort the perception of loudness. Humans tend to average loudness levels over a time period of about 2/10 of a second. Sounds of shorter durations will appear to be louder than sounds (of the same intensity) with durations longer than 2/10 of a second.

Timbre can also influence perceived loudness. Sounds with a complex spectrum will be perceived as being louder than sounds that contain fewer partials. Similarly, sounds with more complex spectra (including many overtones) will be perceived as louder than those that contain a

greater number of proportionally-related partials (harmonics) when both are at the same sound pressure level. Following this principle, a change of timbre during the sustained duration of a sound will result in a perceived change in loudness.

Pitch Perception and Spectrum

As a product of the interaction of the harmonics and closely related overtones of a sound's spectrum, a timbre can create a perception of pitch where a fundamental frequency is not physically present. The harmonics of a sound reinforce its fundamental frequency to enhance the perception of pitch. This phenomenon is so capable of producing the perception of the fundamental frequency that a harmonic spectrum can provide the perception of pitch when the fundamental frequency is not physically present (missing fundamental). A perception of the periodicity of the fundamental frequency is created by the spectrum of the sound, although that frequency itself may not be present.

Amplitude, Time, and Location

The amplitude of two reiterated sounds (separated in time) can influence location perception. The precedence or Haas effect results when two loudspeakers reproduce the same sound in close succession. The effect works against the principle that when two loudspeakers reproduce a sound simultaneously, and at the same amplitude, the sound appears to be centered between the two loudspeakers.

When two loudspeakers reproduce the same sound source in close succession, normal perception would seem to be to localize a sound source at the earliest sounding loudspeaker, then to shift the image to center when the second loudspeaker is sounded. The Haas effect functions to continue the localization of the sound source at the first speaker location, while adding the second loudspeaker (to reinforce the sound intensity of the first speaker) without losing the localization of the sound source at the location of the first loudspeaker. The time difference between the sounding of the loudspeakers must be at least 3 milliseconds to keep the sound at the leading speaker, with 5 milliseconds being a more effective minimum; a maximum delay of approximately 25–30 milliseconds may be used before the delayed signal is perceived as an echo (echoes will be perceived at all frequencies at a delay of 50 milliseconds). If the leading channel is lowered by 10 dB, or the following channel increased by 10 dB, the sound source will again be centered.

Masking

Masking occurs when a sound (or a portion of a sound) is not perceived because of the qualities of another sound. The simultaneous sounding of two or more sounds can cause a sound of lower loudness

level, or a sound of more simple spectral content, to be masked or hidden from the perception of the listener. The masking of sounds is a common problem for people beginning their studies or work in audio recording.

When two simultaneous sounds of relatively simple spectral content have close fundamental frequencies, the sounds will tend to mask each other and blend into a single, perceived sound. As the two sounds become separated in frequency, the masking will become less pronounced until both sounds are clearly distinguishable.

Sounds of relatively simple spectral content tend to mask sounds that are at higher frequencies. This masking becomes more pronounced as the loudness of the lower sound is increased, and is more likely to occur when a large interval separates the two pitch-levels of the sounds. This masking is especially prominent if the two pitch-levels are in a simple harmonic relationship (especially 2:1, 3:1, and 5:1). A higher pitched sound can mask a lower pitched sound if the higher sound is significantly higher in loudness level, and given the same conditions as above; the higher the loudness level, the broader the range of frequencies a sound can mask.

Masking can occur between successive sounds. With sounds separated in time by up to 20–30 milliseconds, the second sound may not be perceived if the initial sound is of sufficient loudness level to draw and retain listener attention, or to fatigue the ear. In a similar way, a sound may not be perceived if it is followed by another sound of great intensity within 10 milliseconds.

Audio equipment can produce “white” and other broadband noise that can mask sounds at all or many frequencies. An entire program might be masked by the noise of the sound system itself, should the loudness level of the noise be sufficiently higher than that of the program. This type of masking problem will first be noticed in the high frequencies, where low loudness levels exist in the upper components of the sound’s spectrum.

Summary

The three states of sound that concern audio recording are sound as it exists in air (the physical dimensions of sound), sound as it exists in human perception (the perceived parameters of sound), the understanding of the meaning of a sound (sound as a resource for artistic expression). The physical dimensions of sound in air are transformed into neural impulses as the perceived parameters of the sound by the ear and brain. The perceived parameters of sound become understood as a resource of elements that allow for the communication and understanding of the meaning of sound (and artistic expression).

The two physical dimensions of the waveform are frequency and amplitude. They function in time, and form the basis for our under-

standing of timbre and spatial properties. As frequency becomes perceived as pitch, amplitude as loudness, time as duration, timbre as timbral characteristics, and space as perceived locations and environmental characteristics—the anomalies of human hearing transform acoustic energy into our perception with marked changes. What these changes actually are, and how these changes take place, are of great concern to the recording professional as they work in the many ways sound is captured, created, modified, and perceived.

Sound as it is perceived and understood by the human mind, becomes the resource for creative and artistic expression in sound. The perceived parameters of sound become the artistic elements of sound in creating musical material and in communicating other meaningful messages. The aesthetic and artistic elements of sound in audio recording are presented in the next chapter.