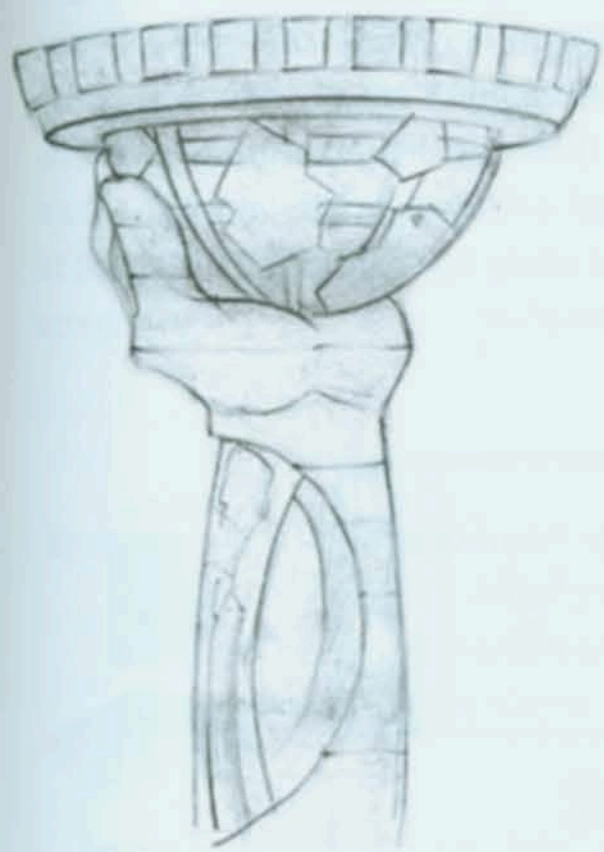
CHAPTER 1

FOUNDATIONS OF AUDIO FOR IMAGE



OVERVIEW

The concepts presented in this chapter are intended to develop conceptual common ground and a working vocabulary that facilitates communication between the filmmaker, sound designer, and composer. Where possible, these concepts are presented in the context of the narrative film.

PERCEPTION OF SOUND

SOUND

There are three basic requirements for sound to exist in the physical world. First, there must be a *sound source*, such as a gunshot, which generates acoustic energy. The acoustic energy must then be transmitted through a *medium* such as air. Finally, a *receiver*, such as a listener's ears, must perceive and interpret this acoustic energy as *sound*. In Film, the animator creates the first two conditions, the sound designer represents these conditions with sound, and the audience processes the sound to derive meaning. Sound can also be experienced as a part of our thoughts, in a psychological process known as *audiation*. As you are silently reading this book, the words are sounding in your head. Just as animators visualize their creations, composers and sound designers conceptualize elements of the soundtrack through the process of audiation. Voice over (in the first person) allows us to hear the interior thoughts of a character, an example of scripted audiation.

> Audio without image is called radio, video without audio is called surveillance.
>
> Anonymous

> To listen is an effort, and just to hear is no merit. A duck hears also.
>
> Igor Stravinsky

HEARING VERSUS LISTENING

When acoustic energy arrives at our ears, it excites the hearing apparatus and causes a physiological sensation, interpreted by the brain as sound. This physiological process is called *hearing*. However, if we are to derive meaning from sound, we must first perceive and respond to the sound through active *listening*. Audiences can actively listen to a limited number of sounds present in the soundtrack. Fortunately, they can also filter extraneous sounds while focusing on selected sounds; this phenomenon is known as the *cocktail effect*. One shared goal of sound design and mixing is to focus the audience's attention on specific sounds critical to the narrative.

LOCALIZATION

In most theaters (and an increasing number of homes), sound extends beyond the screen to include the sides and back of the room. The physical space implied by this speaker configuration is referred to as the *sound field*. Our ability to perceive specific sound placements within this field is known as *localization* (Figure 1.1).

The panners (pan pots) on a mixing board facilitate the movement of sound from left to right by adjusting the relative levels presented in each speaker. Using this approach, individual sounds can be placed (*panned*) within the sound field to accurately match on-screen visuals even as they move.

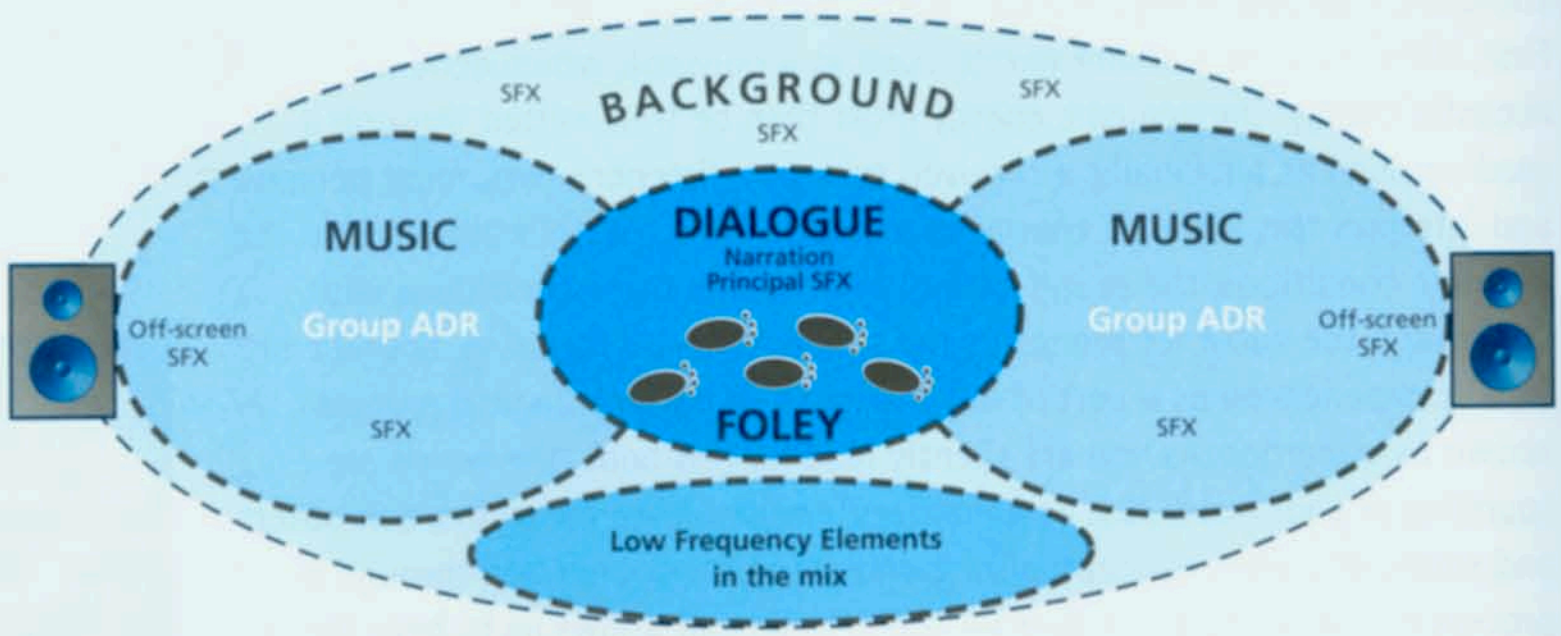


Figure 1.1 The Soundtrack in the Sound Field

Off-screen action can be implied by panning sound to the far left or right in the stereo field (Figure 1.2).

Independent film submission requirements typically call for a stereo mix but there is an increasing acceptance of multi-channel mixes. *Multi-channel* sound extends the sound field behind the audience (surrounds). The surrounds have been used primarily to deliver ambient sounds but this role is expanding with the popularity of stereoscopic (3D) films. Walt Disney pioneered multi-channel mixing with the Fanta-sound presentations of "Fantasia" in the late 1930s, adding height perspective for the sixtieth anniversary screening.

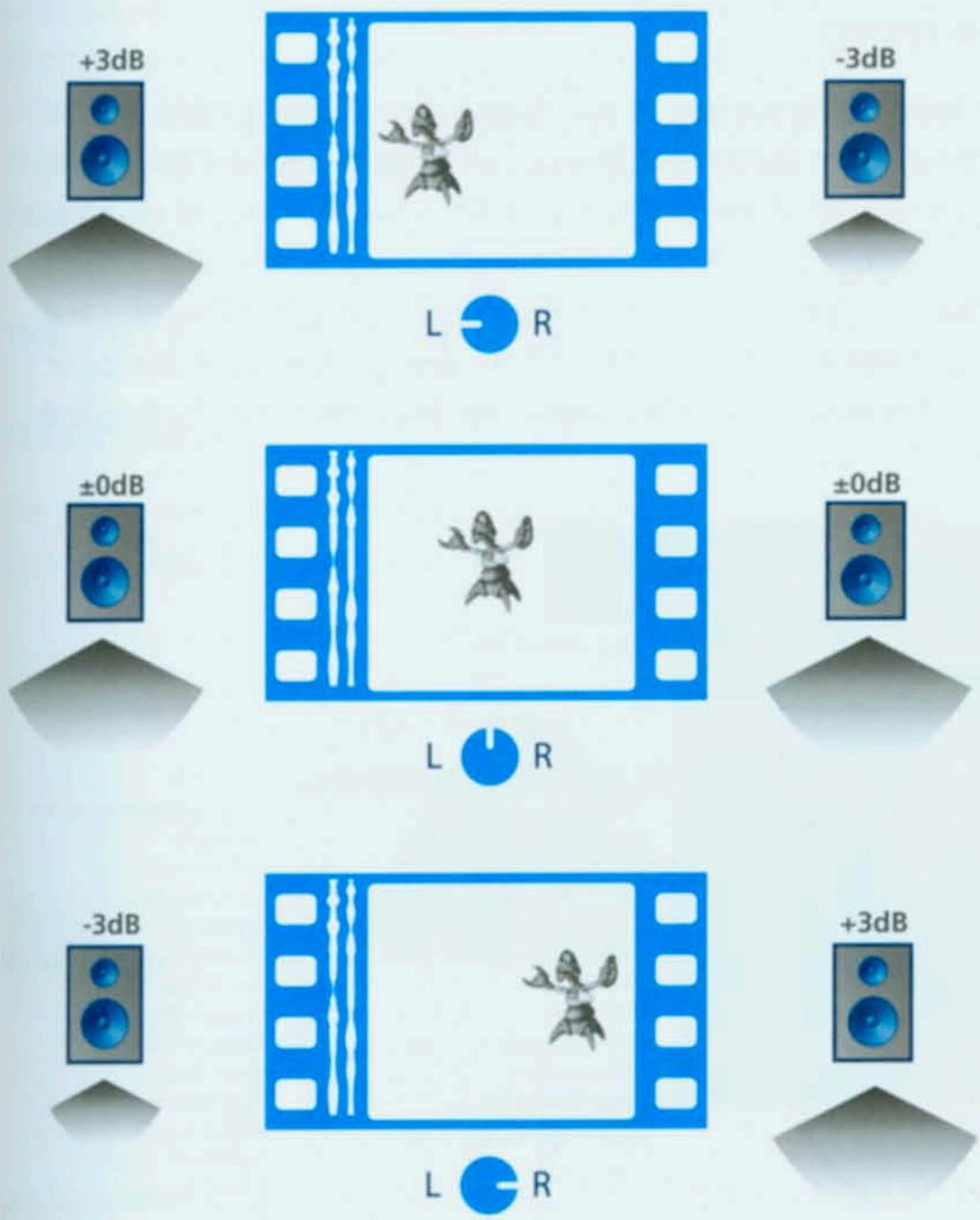


Figure 1.2 Volume Panning

ACOUSTICS

Acoustics is a term associated with the characteristics of sound interacting in a given space. Film mixers apply reverb and delay to the soundtrack to establish and reinforce the space implied onscreen. The controls or parameters of these processors can be used to establish the size and physical properties of a space as well as relative distances between objects. One of the most basic presets on a reverb plug-in is the reverb type or room (Figure 1.3).

The terms *dry* and *wet* are subjective terms denoting direct and reflected sound. Re-recording mixers frequently adjust the reverb settings to support transitions from environment to environment.

RHYTHM AND TEMPO

Rhythm is the identifiable pattern of sound and silence. The speed of these patterns is referred to as the *tempo*. Tempo can remain constant to provide continuity, or accelerate/decelerate to match the visual timings of on-screen images. Footsteps, clocks, and heartbeats are all examples of sound objects that typically have recognizable rhythm and tempo. Vehicles, weapons, and dialogue are often varied in this respect. Many sounds such as footsteps or individual lines of dialogue derive additional meaning from the rhythm and

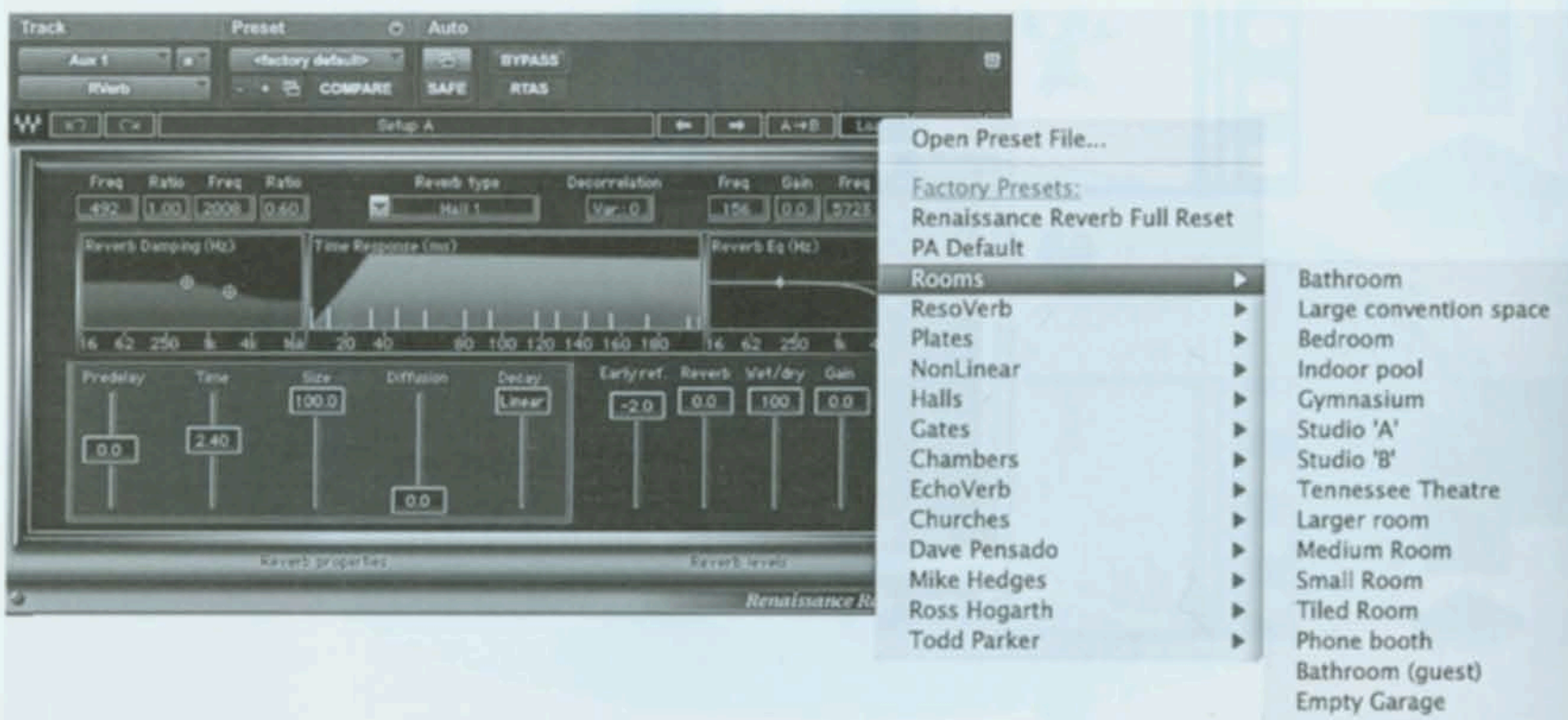


Figure 1.3 Reverb Room Presets

tempo of their delivery. This is an important point to consider when editing sound to picture. Composers often seek to identify the rhythm or pacing of a scene when developing the rhythmic character of their cues.

NOISE AND SILENCE

The aesthetic definition of *noise* includes any unwanted sound found in the soundtrack. Noise always exists to some degree and sound editors and mixers have many tools and techniques to minimize noise. Backgrounds are sometimes mistakenly referred to as noise. However, backgrounds are carefully constructed to add depth to a scene whereas noise is carefully managed as not to detract from the narrative. *Silence* is perhaps the least understood component of sound design. Silence can be an effective means of creating tension, release, or contrast. However, complete silence is unnatural and can pull the audience out of the narrative. Silence before an explosion creates contrast, effectively making the explosion perceptually louder.

> It's the space between the notes that give them meaning.
>
> Nathan East

THE PHYSICS OF SOUND

SOUND WAVES

The sine wave is the most basic component of sound. The horizontal line shown in Figure 1.4 represents the *null or zero point*, the point at which no

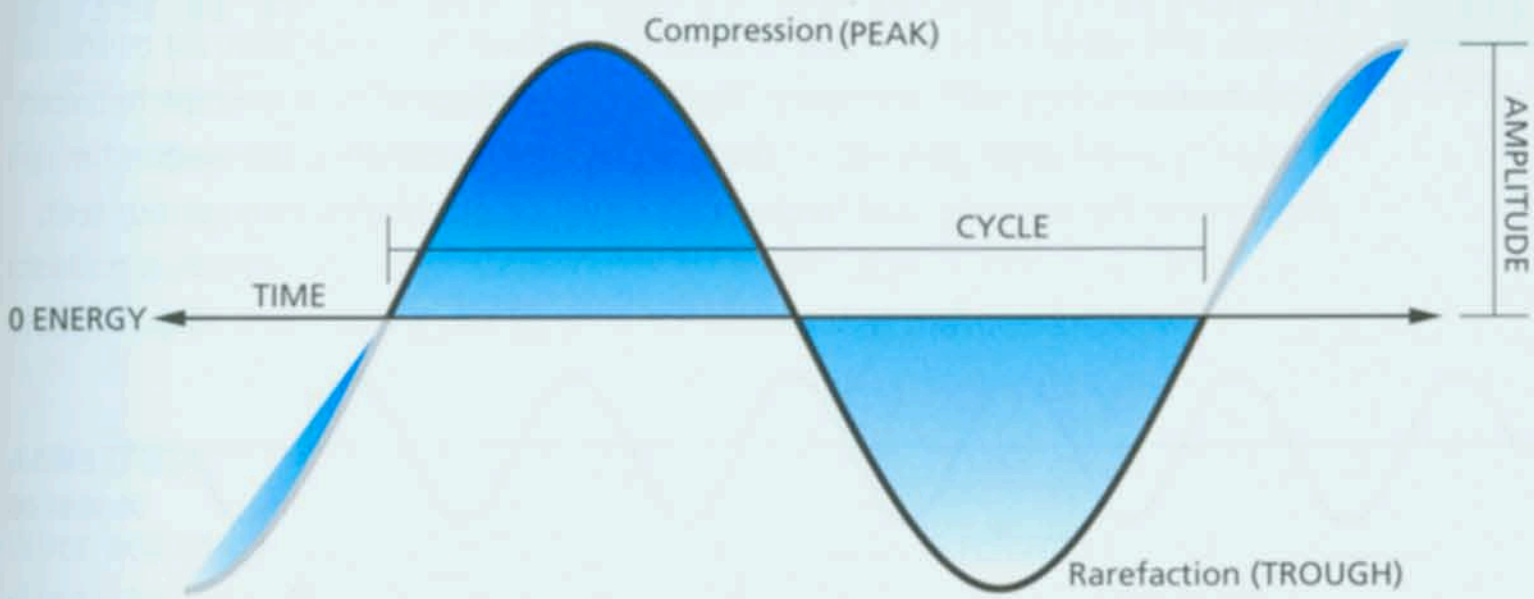


Figure 1.4 The Sound Wave

energy exists. The space above the line represents high pressure (*compression*) that pushes inward on our hearing mechanisms. The higher the wave ascends, the greater the sound pressure, the more volume we perceive. The highest point in the excursion above the line is the *peak*. The space below the line represents low pressure (*rarefaction*). As the wave descends, a vacuum is created which pulls outward on our hearing mechanism. The lowest point in the downward excursion is the *trough*. A single, 360° excursion of a wave (over time) is a *cycle*.

FREQUENCY

Frequency, determined by counting the number of *cycles* per second, is expressed in units called *hertz* (Hz); one cycle per second is equivalent to 1 hertz (Figure 1.5).

Pitch is our subjective interpretation of *frequency* such as the tuning note for an orchestra being A=440 Hertz. The *frequency* range for humans begins on average at 20 Hz and extends upwards of 20,000 Hz (20 kHz). *Frequency response* refers to the range of fundamental frequencies that an object can produce. Frequency response is a critical factor in the selection of microphones, recording devices, headphones, speakers, and commercial SFX/Music. It is also an important qualitative feature relating to audio compression codecs such as Mp3 and AAC. In musical terms, the frequency range of human hearing is 10 octaves, eight of which are present in a grand piano (Figure 1.6).

Frequency can be used to re-enforce many narrative principles such as age, size, gender, and speed. SFX are often pitched up or down to create contrast or to work harmoniously with the score. There are physiological relationships between frequency and human perception that can be used to enhance the cinematic experience. For example, low frequencies travel up our bodies through our feet,

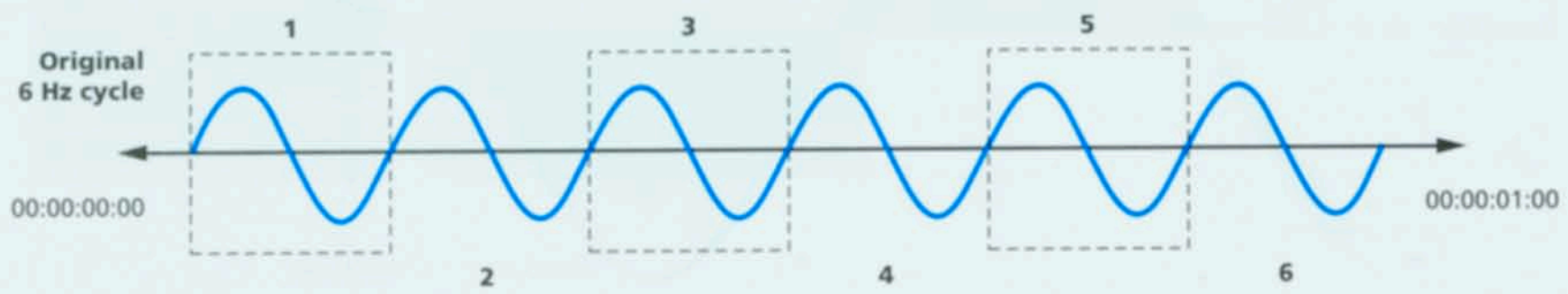


Figure 1.5 Six Cycles Per Second

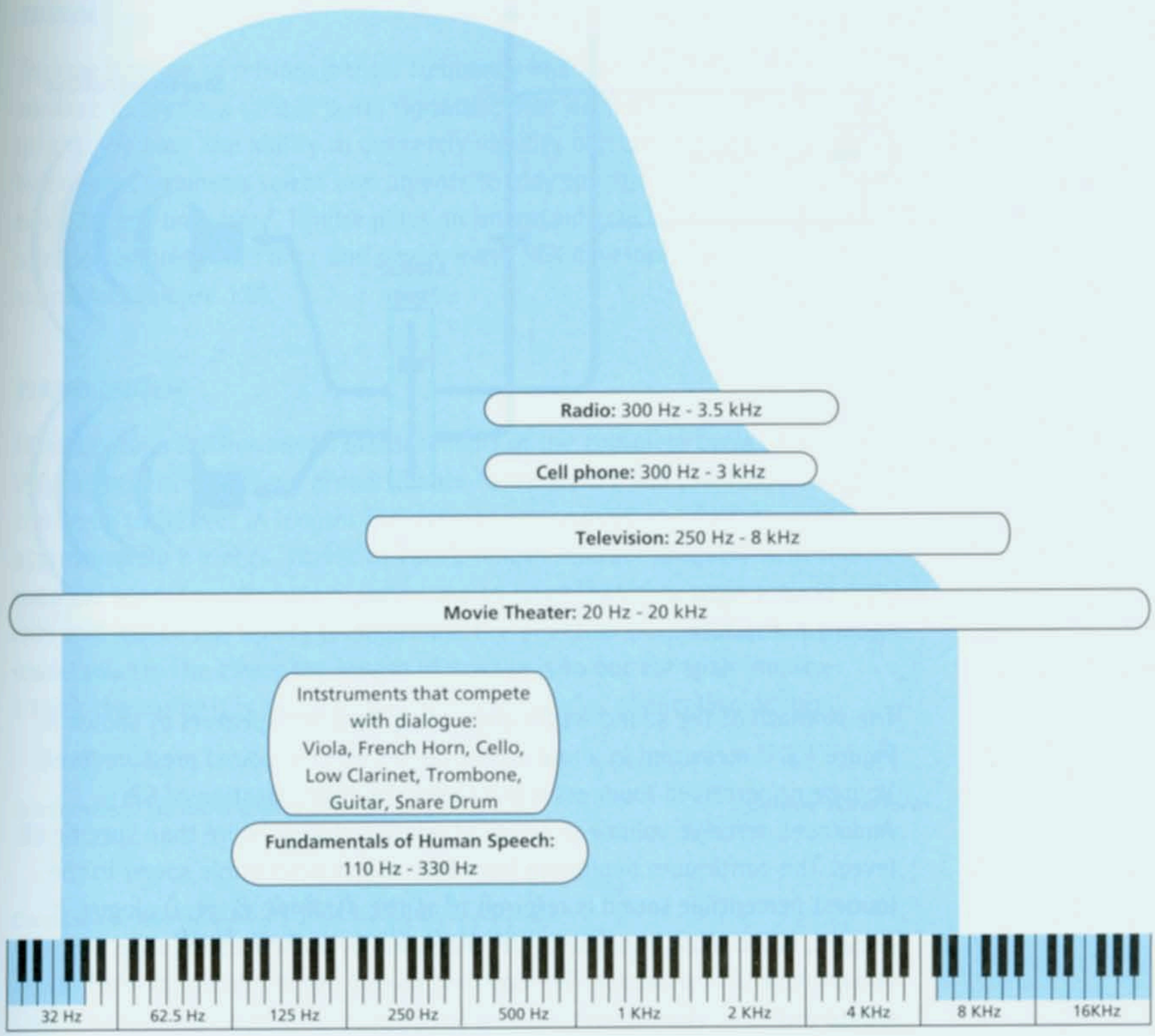


Figure 1.6 Frequency in Context

creating a vertical sensation. Consequently low frequency content can be brought up or down in the mix to enhance visuals that move vertically.

AMPLITUDE

When acoustic energy is digitized, the resultant wave is referred to as a *signal*. *Amplitude* is used to describe the amount of energy (voltage) present in the signal (Figure 1.7).

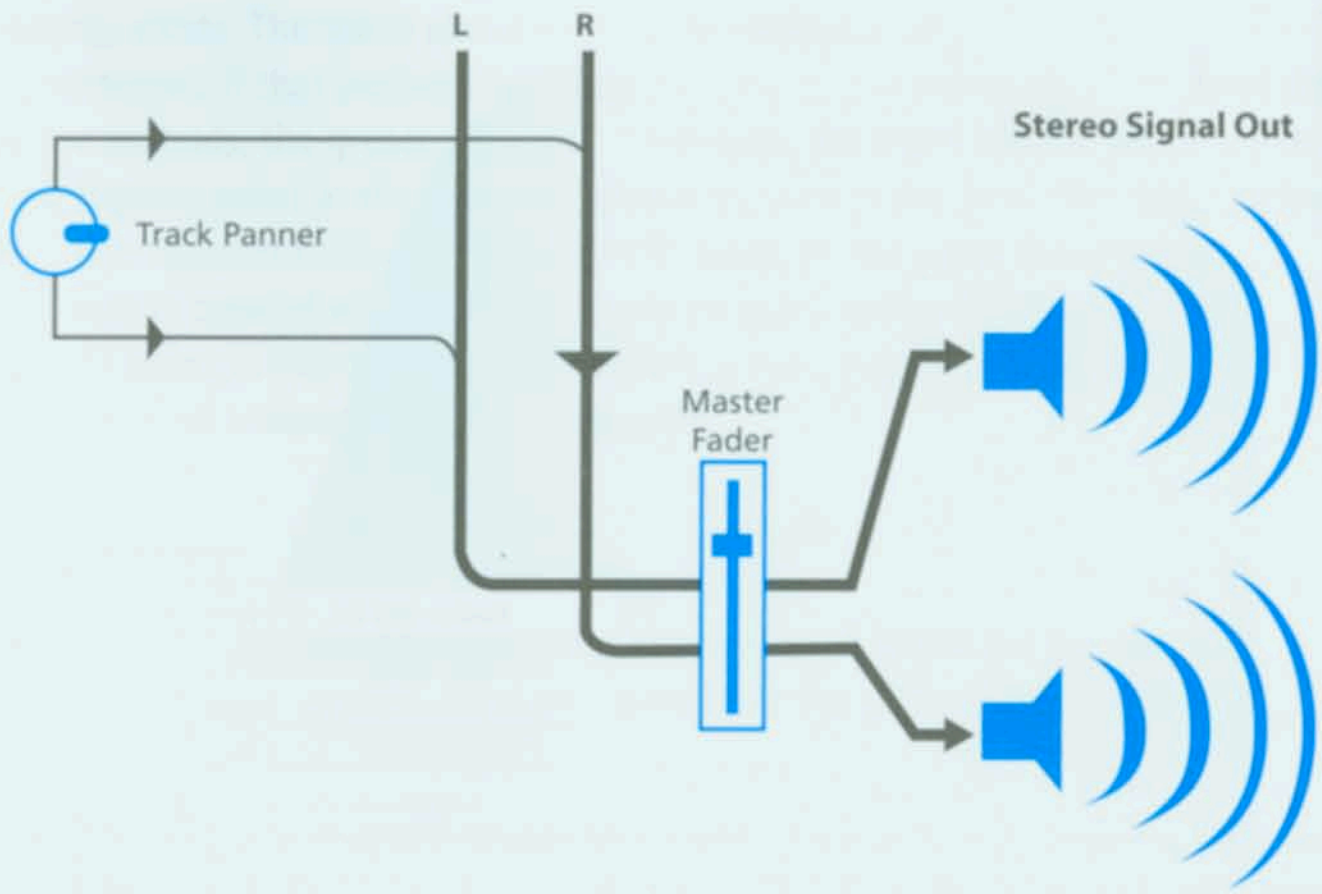


Figure 1.7 Amplitude

The strength of the sound waves emanating from the speakers as shown in Figure 1.6 is measured in a unit called *dB SPL (decibel sound pressure level). Volume* or perceived *loudness* is our subjective interpretation of SPL. Audiences perceive volume in terms of relative change more than specific dB levels. The continuum beginning from the softest perceptible sound to the loudest perceptible sound is referred to as the *dynamic range*. Dialogue, if present, is the most important reference used by audiences to judge the volume levels for playback. Re-recording mixers use amplitude or level to create intensity, provide emphasis, promote scale or size, and establish proximity of sound objects. An audience's perception of volume is also linked with frequency. Human hearing is most sensitive in the mid-range frequencies. More amplitude is required for low and high frequencies to match the apparent loudness of the mid-range frequencies. This perceptual curve is known as the *equal loudness curve*. This curve influences how we mix and why we seek to match the playback levels on the mix stage with the playback levels in theaters *(reference level)*.

TIMBRE

The combination of a fundamental frequency and associated harmonics combine to create a unique aural signature that we refer to as *timbre* or *tone quality*. We have the ability to discretely identify objects or characters solely by timbre. Composers select instruments to play specific components of a cue based largely on timbre. Timbre plays an important role in voice casting, the selection of props for Foley, and nearly every SFX developed for the soundtrack (Figure 1.8).

WAVELENGTH

Wavelength is the horizontal measurement of the complete cycle of a wave. Wavelengths are inversely proportionate to frequency and low frequencies can be up to 56 feet in length. The average ear span of an adult is approximately 7 inches. Waves of varied length interact uniquely with this fixed ear span. Low-frequency waves are so large that they wrap around our ears and reduce our ability to determine the direction (*localization*) of the sound source. The closer the length of a wave is to our ear span (approx 2 kHz), the easier it is for us to localize. One popular application of this

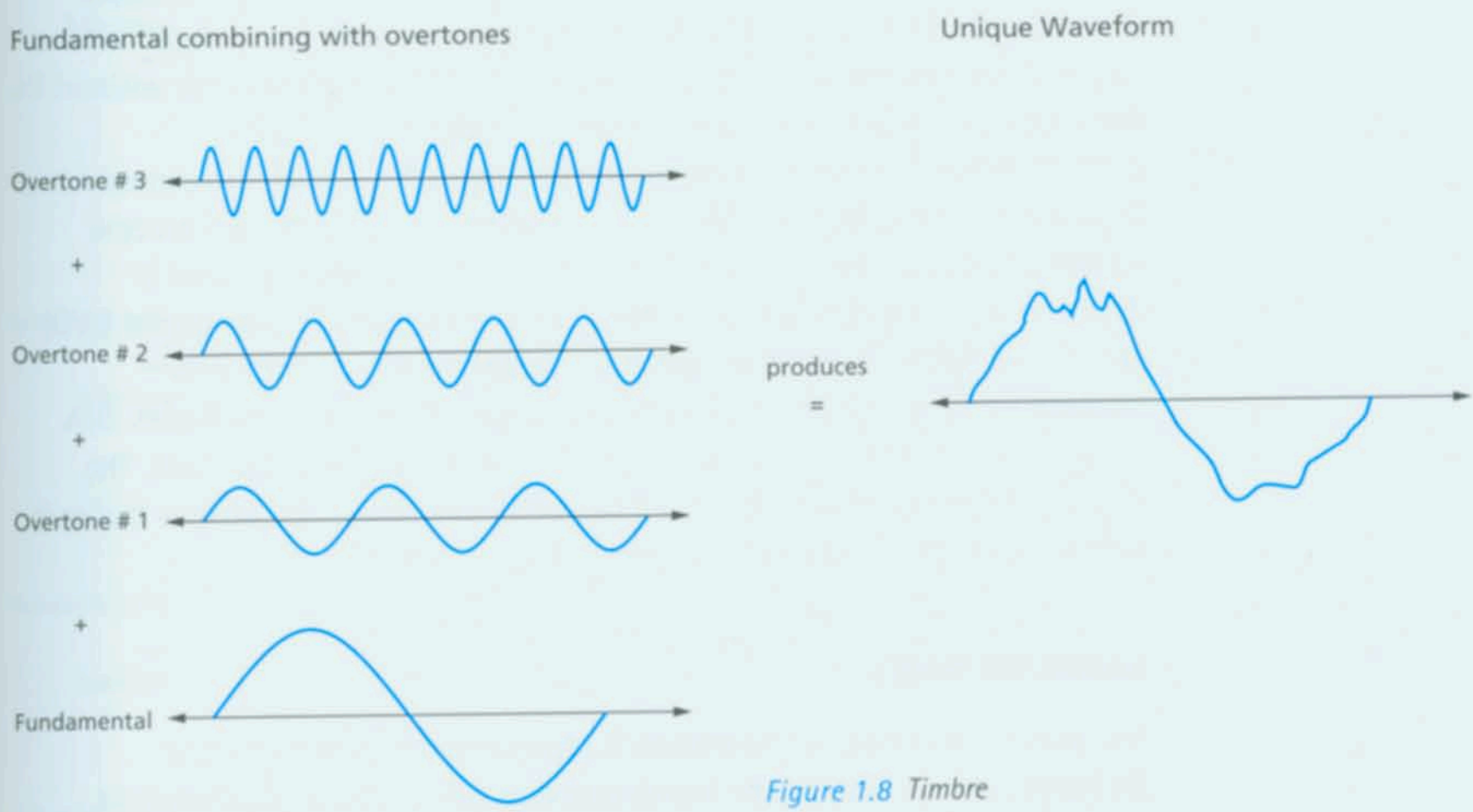


Figure 1.8 Timbre

concept is the use of low-frequency sound effects to rouse our "fight or flight" response. When audiences are presented with visual and sonic stimuli determined to be dangerous, the body instinctively prepares to fight or escape. If the audience cannot determine the origin or direction of the sound because the sounds are low in frequency, their sense of fear and vulnerability is heightened.

SPEED OF SOUND

Sound travels at roughly 1130 feet per second through air at a temperature of 70°F; the more dense the medium (steel *versus* air), the faster the sound travels. The speed of sound is equal at all frequencies. In reality, light travels significantly faster than sound. Therefore, we see an action or object before we hear it; however, the cinematic practice for sound editing is to sync the audio on or slightly after the action; this is referred to as *hard sync*.

DIGITAL AUDIO

DIGITIZING AUDIO

Digital audio has impacted nearly every aspect of soundtrack production. When handled properly, digital audio can be copied and manipulated with minimal degradation to the original sound file. Digital audio is not without its drawbacks, however, and it is important to understand its basic characteristics in order to preserve the quality of the signal. The conversion of acoustic energy to digital audio is most commonly achieved in a process known as LPCM or Linear Pulse Code Modulation. An analog signal is digitized using specialized computer chips known as analog-to-digital (A/D) converters. A/D converters are designed to deliver a range of audio resolutions by sampling a range of frequencies at discrete amplitudes. D/A converters reverse this process to deliver analog signals for playback. The quality of A/D converters vary from device to device and are often a factor in the cost of higher end audio equipment (Figure 1.9).

SAMPLING RATES

The visual component of animation is represented by a minimum of 24 frames per second. As the frame rate dips below this visual threshold,

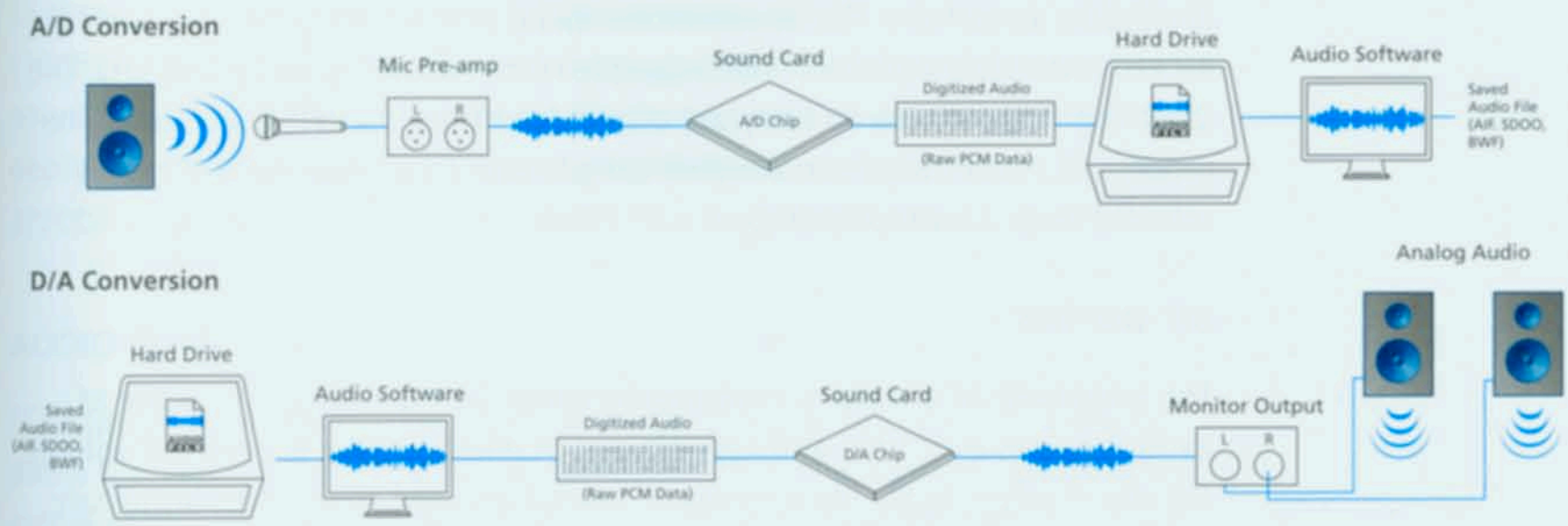


Figure 1.9 *A/D Conversion*

the image begins to flicker (persistence of vision). Similar thresholds exist for digital audio as well. Frequency is captured digitally by sampling at more than twice the rate of the highest frequency present, referred to as the *Nyquist frequency* (Figure 1.10).

If the sampling rate falls below this frequency, the resultant audio will become filled with frequencies that were not present in the original

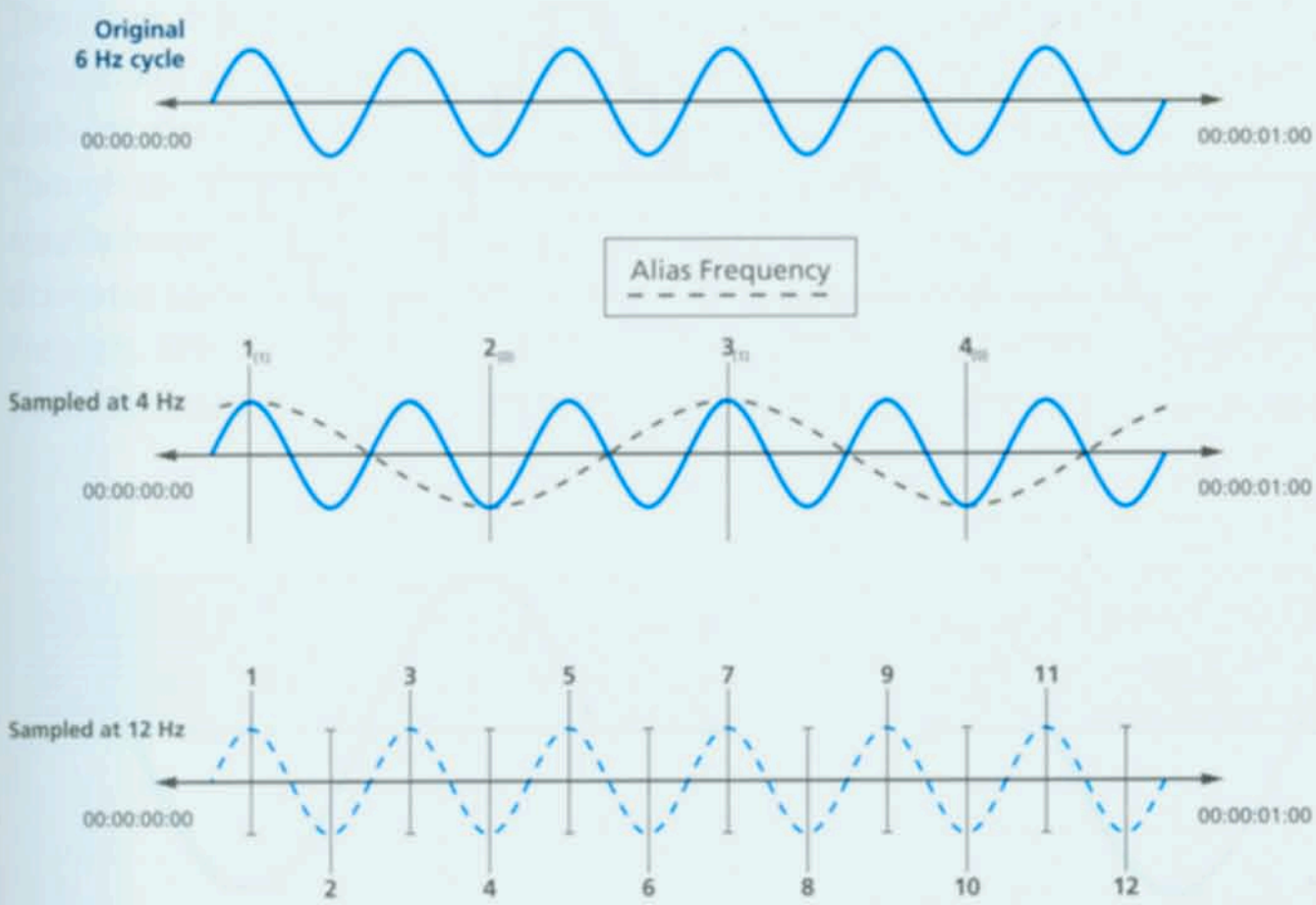


Figure 1.10 *Nyquist Frequency*

(harmonic distortion). The minimum sampling rate for film and television is 48 kHz with high definition video sample rates extending up to 192 kHz. The sampling rates of web delivery are often lower than the NTSC standard. The process of converting from a higher sample rate to a lower sample rate is referred to as *down sampling.*

BIT-DEPTHS

The amplitude of a wave is captured by sampling the energy of a wave at various points over time and assigning each point a value in terms of voltage. *Bit-depth* refers to the increments or resolution used to describe amplitude. At a bit-depth of two, the energy of a wave is sampled in four equal increments (Figure 1.11).

Notice that all portions of the wave between the sampling increments are rounded up or down (quantized) to the nearest value. *Quantization* produces errors (noise) that are sonically equivalent to visual pixilation. As the bit-depth is increased, the resolution improves and the resulting signal looks and sounds more analogous to the original (Figure 1.12).

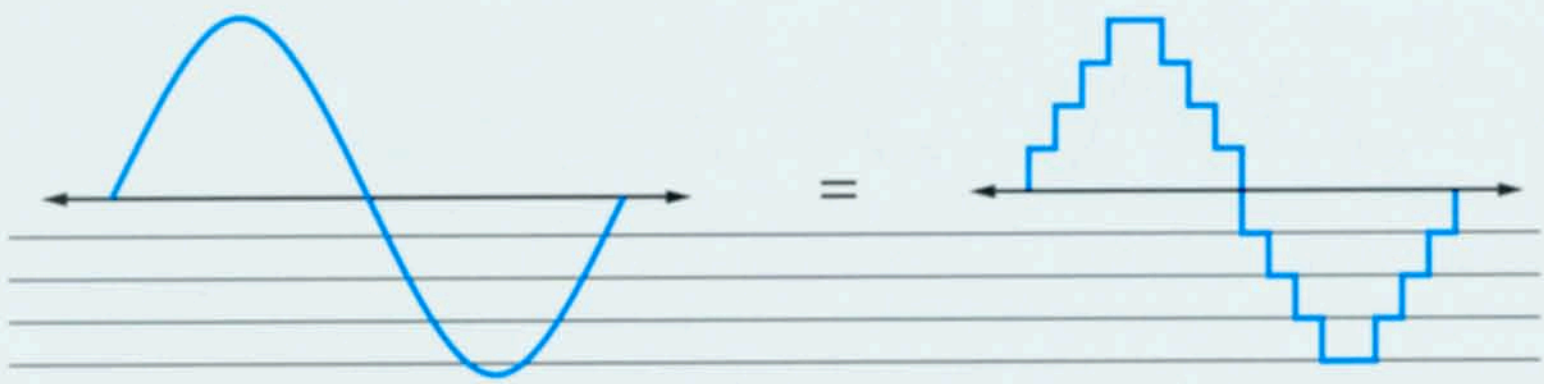


Figure 1.11 Low Bit-Depth Resolution

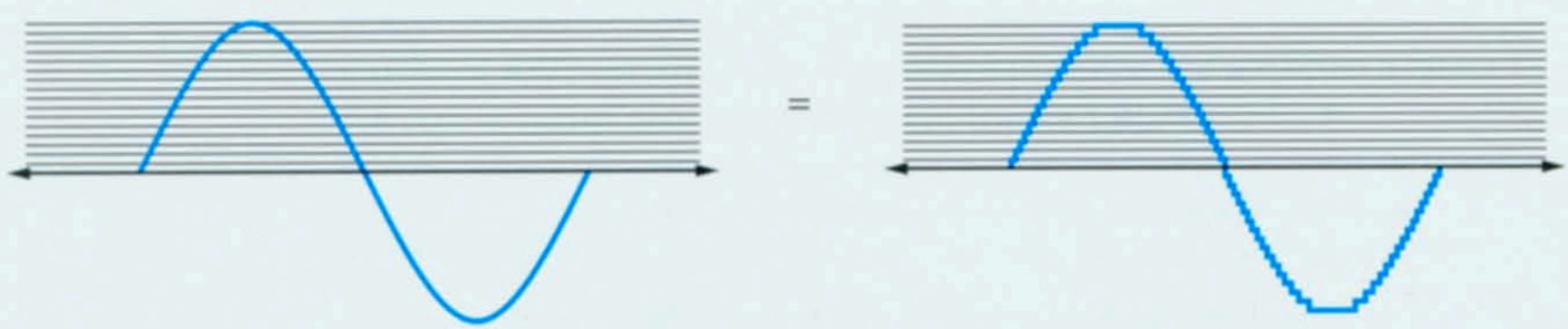


Figure 1.12 High Bit-Depth Resolution

In theoretical terms, each added bit increases the dynamic range by 6.01 dB (16 bit = 96 dB). The professional delivery standard for film and television soundtracks is currently 16 to 24 bit. However, sound editors often work at bit depths up to 32 bit when performing amplitude based signal processing. The process of going from a higher bit depth to a lower bit depth is called *dither*.

AUDIO COMPRESSION

PCM (pulse code modulation) is a non-compressed or full resolution digital audio file. Non-compressed audio represents the highest quality available for media production. File extensions such as aiff, .wav, and BWF are added to a PCM file to make them readable by computers. In a perfect world, all audio for media would be full resolution. However, due to file storage and transfer limitations for Internet, package media, and broadcast formats, audio compression continues to be a practical reality. Audio codecs (compression/decompression) have been developed to meet delivery requirements for various release formats. Of these, the most common are mp3, AAC, AC-3, and DTS. The available settings on a codec have a direct impact on the frequency response of the resultant files. For example, higher transfer rates preserve more of the high frequency content in the original file. Therefore, it is important to know the allowable transfer rate and file size limitations for specific release formats when encoding the audio. For Internet delivery, streaming options are typically lower than download options. Though the differences between high and low resolution audio may not be readily heard on consumer monitors or head phones, these differences are dramatic when played through high quality speakers like those found in theaters. Therefore it is important to understand what type of monitors your potential audiences will be listening to and compress accordingly.

CHAPTER 2

SOUND DESIGN THEORY

OVERVIEW

The pairing of audio and visual media is part of our everyday experiences. Armed with an ipod and earbuds, we intuitively underscore our daily commutes and recreation time through carefully selected song-lists. We visualize while listening to the radio and we audiate while reading. In the absence of either sound or picture, audiences will create that which is not present, potentially redefining the original intention of any given work. Therefore, in narrative animation, we must carefully select the content used in this pairing as we guide the audience through the story. To accomplish this, we must first understand the unique relationship of sound paired with image. As early as 1928 with the release of *Steamboat Willie*, directors, editors, and composers have explored this relationship as they develop the aesthetic for creative sound design.

> No theory is good except on condition that one use it to go beyond.
>
> Andre Gide

SOUND CLASSIFICATIONS

CHION CLASSIFICATIONS

Michel Chion (a noted sound theorist) classifies listening in three modes: causal, semantic, and reduced. Chion's typology provides a useful design framework for discussing sounds function in a soundtrack. *Causal* sound is aptly named, as it reinforces cause and effect (see a cow, hear a cow). Sound editors often refer to these as *hard effects*, especially when they are used in a literal fashion. The saying "we don't have to see everything we hear, but we need to hear most of what we see" describes the practical role that causal sounds play in the soundtrack. The term *semantic* is used to categorize sound in which literal meaning is the primary emphasis. Speech, whether in a native or foreign tongue, is a form of semantic sound. Morse code is also semantic and, like foreign languages, requires a mechanism for translation (e.g.

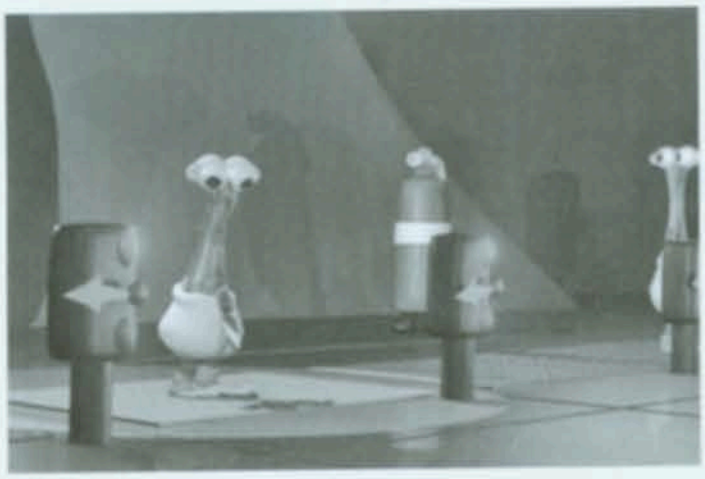


Figure 2.1 The Footsteps for the Fire Extinguisher Are Covered with the Non-Literal Sound of a Spray Paint Can Being Shaken in the Film Mercury Inspection *(2006), Directed by Leslie Wisley Padien*

subtitles) when used to deliver story points. Even the beeps placed over curse words imply a semantic message. The term *reduced* refers to sounds that are broken down, or reduced, to their fundamental elements and paired with new objects to create new meanings. Animation provides many opportunities for this non-literal use of sound.

DIEGETIC AND NON-DIEGETIC SOUND

The term *diegetic* denotes sound that is perceived by the characters in the film. Though unbeknownst to the characters, diegetic sound is also perceived by the audience. This use of sound establishes a voyeuristic relationship between the characters and their audience. *Non-diegetic* sound is heard exclusively by the audience, often providing them with more information than is provided to the characters. When discussing music cues, the terms *source* and *underscore* are used in place of diegetic and non-diegetic. Diegetic sound promotes the implied reality of a scene. Principle dialogue, hard effects, and source music are all examples of diegetic sound. In contrast, non-diegetic components of a soundtrack promote a sense of fantasy. Examples of non-diegetic sound include narration, laugh tracks, and underscore. The cinematic boundaries of diegesis are often broken or combined to meet the narrative intent at any given moment.

> They forgot to turn off the ambience again!–
>
> Alex (*Madagascar*)

NARRATIVE FUNCTIONS

GUIDED PERCEPTION

The images presented within a scene are often ambiguous in nature, inviting the audience to search for non-visual clues as they attempt to clarify both meaning and intensity. Within these shots are creative opportunities to guide the audience perceptually from frame to frame. The editors and composers for the early Warner and MGM shorts understood this and became masters at guiding the audience's perception. They accomplished this by using non-literal sound for subjective scenes. Disney also uses this design approach when producing films for younger audiences. For example, in *Beauty and the*

> If that squeak sounds kind of sad and you never meant it to sound sad, it will throw it that way, so you can't ignore it.–
>
> Andrew Stanton

Beast, the final conflict between the villagers and the castle's servants is underscored with a playful rendition of "Be our Guest". Here, the score guides the audience to a non-literal interpretation of the visuals, promoting a "softer" perception of the content, thus making the film appropriate to a wider audience.

DRAWING THE AUDIENCE INTO THE NARRATIVE

The title sequence is typically the first opportunity for soundtrack to contribute to the storytelling process. Like a hypnotist, the sound design team and composer develop elements for the title sequence to draw the audience out of their present reality and into the cinematic experience. Sound designers and re-recording mixers take great care to keep the audience in the narrative. At the conclusion of the film, the soundtrack can provide closure while transitioning the audience back to their own reality.

DIRECTING THE EYE

Re-recording mixers create sonic foregrounds, mid-grounds, and backgrounds using volume, panning, delay, and reverb. They choose what the audience hears and the perspective for which it is heard. Subtle visuals can be brought into focus with carefully placed audio that directs the eye (Figure 2.2).

Cocktail Tray	Synchronized Sound	Guided Perception
	Sonar Ping	Directing our attention to the submerged olive.
	Song "Auld Lang Syne"	Establishing the season and emotional tone.
	Ambulance Siren	Forshadowing cause and effect.

Figure 2.2 The Picture Displays a Martini Glass, an Ashtray, and a Wine Bottle. By Pairing a Seemingly Unrelated Sonar Ping with this Frame, Audiences Are Challenged to Find a Relationship or Gestalt between the Two, the Desired Effect Being that the Eyes are Directed to the Submerged Olive

ESTABLISHING OR CLARIFYING POINT OF VIEW

In a scene from *Ratatouille* (2007), Remy and Emile (rats) are being chased by an elderly French woman in her kitchen. When we experience the scene through their perspective, the principle dialogue is spoken in English. However, when the perspective shifts to the elderly women, their dialogue is reduced to rat-like squeaks. This is an example of dialogue effectively transitioning with the point of view (POV) within a scene. In *Toy Story* (1995), Gary Rydstrom (sound designer) and Randy Newman (composer) contrast diegetic sound and underscore in a street-crossing scene. The world as heard from the perspective of the toys is primarily musical. In contrast, distance shots of the toys are covered with hard effects and BGs. Cinema has borrowed a technique from Opera where specific thematic music is assigned to individual characters. This approach to scoring is referred to as *Leitmotif* and is often used to establish or clarify the POV of a scene, especially where more than one character is present. Sound designers often pitch shift or harmonize props in an attempt to create a sonic distinction between characters.

CLARIFYING THE SUBTEXT

> You have to find some way of saying it without saying it.
>
> Duke Ellington

During the silent film era, actors were prone to exaggerated expressions or gestures to convey the emotion of a scene. With the arrival of synchronized sound, the actor was liberated from exaggerated acting techniques. When a character is animated with a neutral or ambiguous expression, an opportunity for the soundtrack to reveal the characters' inner thoughts or feelings is created. When a music cue is written specifically to clarify the emotion of the scene, it is referred to as *subtext scoring*. Subtext scoring is more immediate and specific than dialogue at conveying the emotion. The subtext can also be clarified through specific design elements or BGs.

CONTRASTING REALITY AND SUBJECTIVITY

Though animation is subjective by nature, there exists an implied reality that can be contrasted with subjective moments such as slow motion scenes, montage sequences, and dream sequences. Subjective moments are often reinforced sonically by means of contrast. One common design approach to subjective moments is to minimize or remove all diegetic sound. This approach is often used in montage sequences where music is often the

exclusive sound driving the sequence. In slow motion sequences, diegetic sound is typically slowed down and lowered in pitch to simulate analog tape effects. Though clichéd in approach, dream sequences are often reinforced with additional reverb to distance the sound from reality.

EXTENDING THE FIELD OF VISION

There is a saying that a picture is worth a thousand words. In radio (theater of the mind), sound is worth a thousand pictures. When objects producing sound are visible, the sound is referred to as *sync sound* or *on-screen*. *Off-screen* sound can introduce and follow objects as they move on and off-screen, thereby re-enforcing the visual line established through a series of shots. Our field of vision is limited to roughly 180°, requiring visual cuts and camera movements to overcome this limitation. Multi-channel soundtracks, on the other hand, are omni-directional and can be used to extend the boundaries of a shot beyond the reach of the eye. The classic *camera shake* used by Hanna-Barbera is a visual approach to storytelling where off-screen sound is used to imply that which is not shown. This approach serves to minimize the perceived intensity of the impacts while avoiding costly and time-consuming animation. Multi-channel mixing has the potential to mirror front-to-back movements while extending the depth of a scene to include the space behind the audience. With the growing use of stereoscopic projection, there is increased potential for multi-channel mixing to mirror the movement of 3-dimensional images.

> The ears are the guardians of our sleep.
>
> Randy Thom

TENSION AND RELEASE

Tension and release is a primary force in all art forms. Film composers have many means of creating tension, such as the shock chords (dissonant clusters of notes) featured prominently in the early animations of Warner Brother and MGM studios. Another approach common to this period was the harmonization of a melody with minor 2nds. The suspension or resolution of a tonal chord progression or the lack of predictability in atonal harmonic motion can produce wide ranges of tension and release. A *build* is a cue that combines harmonic tension and volume to create an expectation for a given event. Some sounds have inherent qualities that generate tension or release and are often used metaphorically. For example, air-raid sirens, crying babies, emergency vehicles, growling animals, and snake rattles all evoke tension.

Conversely, the sound of crickets, lapping waves, and gentle rain produce a calming effect. Either type of sound can be used directly or blended (*sweetened)* with literal sounds to create a subliminal effect on the audience.

CONTINUITY

> I was looking for a special sound and it was kind of keyed into the design of the movie.
>
> Brad Bird

Continuity has always been an important aesthetic in filmmaking, regardless of length. In a narrative animation, continuity is not a given; it must be constructed. For this reason, directors hire a supervising sound editor to insure that the dialogue and SFX are also cohesive. In Disney's *Treasure Planet* (2002), director Ron Clemons developed a 70/30 rule to guide the design for the production. The result was a visual look that featured a blending of eighteenth-century elements (70 percent) with modern elements (30 percent). Dane Davis, sound designer for *Treasure Planet*, applied this rule when designing SFX and dialogue to produce an "antique future" for this soundtrack. Continuity is not limited to the physical and temporal realms,

Figure 2.3 In the Film Painting Day *(2003) Directed by Daesup Chang, Elements of the Soundtrack Were Intentionally Limited to Those that Existed in the Old West*

emotional continuity is equally important. The score is often the key to emotional continuity. Perhaps this is why few scores are written by multiple composers. It takes great insight to get to the core of a character's feelings and motivations and even greater talent to express those aspects across the many cuts and scenes that make up a film. When developing cues, continuity can be promoted through instrumentation. For example, in *The Many Adventures of Winnie the Pooh* (1977), composer Buddy Baker assigned specific instruments to each of the characters.

PROMOTING CHARACTER DEVELOPMENT

An effective soundtrack helps us identify with the characters and their story. Animators create characters, not actors. Much of the acting occurs in the dialogue while SFX reinforce the character's physical traits and movements. The score is perhaps the most direct means of infusing a character with emotional qualities. Character development is not limited to people. In animation, virtually any object can be personified (*anthropomorphism)* through dialogue, music, and effects. Conversely, humans are often *caricatured* in animation. For example, the plodding footsteps of Fred Flintstone are often covered with a tuba or bassoon to emphasize his oversized physique and personality.

THEORETICAL CONCEPTS SPECIFIC TO DIALOGUE

Dialogue is processed in a different manner than SFX or music in that audiences are primarily focused on deriving meaning, both overt and concealed. Consequently, dialogue tracks do not require literal panning and are most often panned to the center. One notable exception to this practice can be heard in the 1995 film *Casper*, where the dialogue of the ghost characters pans around the room as only ghosts can. Narration, like underscore, is non-diegetic, promoting the narrative fantasy and requiring a greater willing suspension of disbelief. Narration is sometimes processed with additional reverb and panned in stereo, placing it in a different narrative space than principle dialogue. When Fred Flintstone speaks *direct to the camera,* he is breaking the sonic boundary (diegesis) that normally exists between the characters and the audience. This style of delivery is referred to as *breaking the 4th wall.* One clever application of both concepts is the introduction of principle dialogue off-screen, suggesting narration, and then

through a series of shots, revealing the character speaking direct to camera. This allows dialogue to deliver story points while also minimizing animation for speech.

THEORETICAL CONCEPTS SPECIFIC TO SCORE

> You're trying to create a sonic landscape that these drawings can sit on . . . and these characters can live within.
>
> Hans Zimmer
> *(The Lion King)*

In addition to establishing the POV of a scene, an established *leitmotif* can allow audiences to experience the presence of a specific character, even when that character is off-screen. The abstract nature of animation often calls for a *metaphoric sound* treatment. For example, in *A Bug's Life* (1998), the back and forth movement of a bowed violin is used to exaggerate the scratching motion of a bug in a drunken state. The use of musical cues to mimic movement is known as *isomorphism.* Ascending/descending scales and pitch bending add directionality to isomorphism. The use of musical instruments as a substitution for hard effects gained favor in the early years of film as it facilitated coverage of complex visual action without requiring cumbersome sound editing. Percussion instruments in particular became a fast and effective means of representing complex rhythmic visuals. The substitution of musical intervals for Foley footsteps *(sneak steps)* began as early as 1929 with Disney's *The Skeleton Dance.* The melodic intervals and synchronized rhythm help to exaggerate the movement while specifying the emotional feel. *Anempathetic* cues are deliberately scored to contrast the implied emotion of a scene. This technique is typically used to increases an audience's empathy for the protagonists or to mock a character. *Misdirection* is a classic technique where the score builds toward a false conclusion. This technique is often used for dream sequences or as a setup for visual gags. Acoustic or synthetic instrumentation can be used to contrast organic and synthetic characters or environments. The size of instrumentation or orchestration establishes the scale of the film or the intimacy of a scene. Conversely, a full orchestra can give a scene an epic feel. Score can function transitionally as well. For example, a cue can morph from *source-to-underscore* as a scene transitions to a more subjective nature.

> Sound effects are a huge part of making us believe in these characters and connect to them emotionally.
>
> Ethan Van Der Ryn

THEORETICAL CONCEPTS SPECIFIC TO SFX

Sound effects are an effective means of establishing narrative elements such as time period, location, and character development. For example, seagulls imply the ocean, traffic implies urban settings, and machinery implies

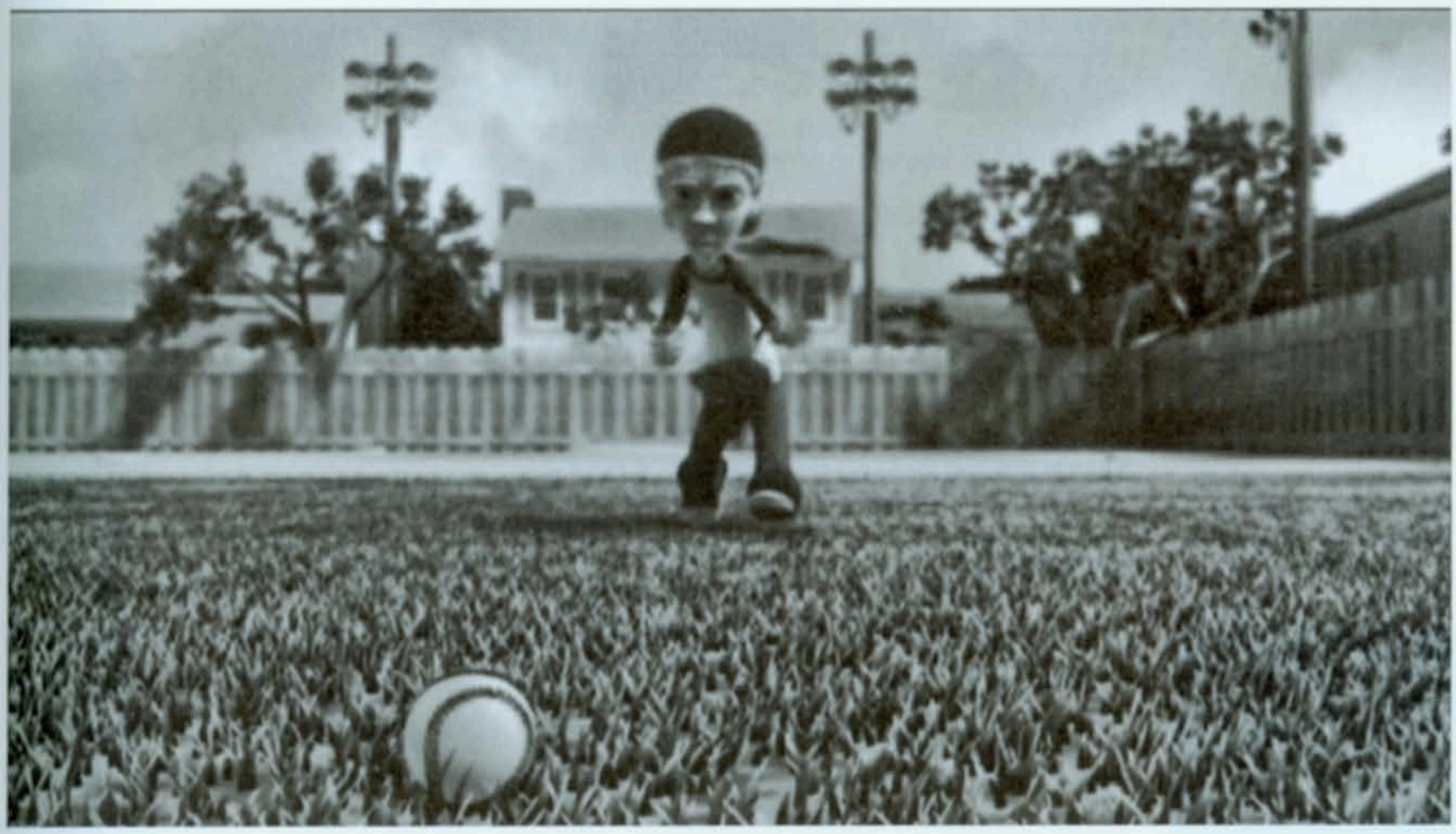


Figure 2.4 An Acoustic Bass Covers the Individual Footsteps (Sneak Steps) in the Sandlot Fantasy Pasttime *(2004) Directed by Todd Hill*

industry. World War II air-raid sirens and steam-powered trains are historical icons that help to clarify the time period. Characters are often developed through associated props, such as a typewriter (journalist) or a whistle blast (traffic cop). Many sounds elicit emotional responses from an audience due to their *associative nature*. Sounds have the potential of revealing or clarifying the underlying meaning or subtext of a scene. For example, the pairing of a gavel with a cash register *(sweetening)* implies that justice is for sale. Visual similarities between two objects such as a ceiling fan and a helicopter blade can be re-enforced by morphing their respective sounds as the images transition, this is referred to as a *form edit* (Table 2.1). Sound effects are often used to de-emphasize the intensity of the visuals, especially when used to represent off-screen events and objects.

Table 2.1 Form Edits

Woodpecker (country)	Jackhammer (urban)
Flock of geese (organized activity)	Traffic jam (breakdown)
Alarm clock (internal)	Garbage truck backing up (external)
Telegraph (antique)	Fax machine (modern)
Typewriter (documentation)	Gunshot (flashback to event)

INTERPRETING PICTURE EDITS

OVERVIEW

> In Hollywood there are no rules, but break them at your own peril.
>
> Peter Guber

In the early years of film, picture editors handled many of the tasks now handled by sound editors. They intuitively understood how the picture editor directed the flow of images, established a rhythm, maintained continuity, and facilitated the narrative. They were also sensitive to the implications of camera angles, camera movement, and framing. As a result, their sound edits were visually informed. As sound design evolves into a more specialized field, it is important for sound editors to maintain this perspective. This can be a challenging proposition for when properly executed, picture editing is nearly transparent to the audience. Sound designers and composers must develop an awareness of visual cues and their implications for sound design. The purpose of this section is to establish a common language for both picture and sound editors. There is a danger when discussing this broad a topic in a shortened format. While reading the following, be mindful that the craft and creativity of editing cannot be reduced to a set of rigid rules. As you work on each film, allow that film to evolve by never allowing methodology to be a substitute for personal logic and feelings. No filmmaker ever hoped his audience would leave the theater saying "boy that was a well-edited film."

SHOTS

The most basic element of the film is the *frame*. A series of uninterrupted frames constitute a *shot* or, in editor language, a clip. A sequence of related shots are assembled to create a *scene*. Some shots are *cut* and digitally spliced while others are given a transitional treatment such as a *wipe* or *dissolve*. In most cases, the goal of picture and sound editing is to move from

shot to shot with the greatest transparency. Storyboards, shot lists, and spotting sessions are all important means of communicating the vision of the film and the nuance associated with each frame, shot, and edit.

Framing

During the storyboard stage, the animator designs a series of shots to develop a scene. There is a multitude of approaches to individual shots and a wide range of verbiage to describe each. One aspect of the shot is *framing*, which establishes the distance of the audience from the subject. From close-up to extreme long shot, framing has its counterpart in the foreground and background of the soundtrack. For example, close-up shots (CU) isolate an action or expression, eliminating peripheral distractions. Therefore the soundtrack supports this perspective by minimizing off-screen sounds. Long shots, on the other hand, provide a more global perspective, and are better suited for sonic treatments of a more ambient nature. When designing sound for animation, it is important to consider individual shots in the larger context of a sequence. For example, in an action sequence, it is common to frame the same subject from many different perspectives. As a result, the movements of these objects appear to move faster or slower depending on the framing. To promote continuity, this type of sequence is often scored at a continuous tempo, creating a unified sense of pacing when played through the scene.

Camera Placement

Camera placements offer a variety of ways to portray a character. The most common camera placement is at the eye level. This placement implies a neutral or equal relationship with the subject. When the camera angle points up to a character, the intent often indicates a position of power. The reverse might also be true but it is important to clarify intent before designing sound based on camera placement alone. The *over-the-shoulder shot* blocks our view of the character's mouth, creating an opportunity to add or alter lines of dialogue in post-production. This placement also shifts the perspective to another character, often revealing that character's reaction to what is being said.

Camera Movement

Camera movements can be horizontal *(pan)*, vertical *(tilt)*, or move inward and outward *(zoom)*. Both pan and vertical shots are often motivated by the need to reveal more information in an extended environment. Pan shots are often underscored and/or designed with static backgrounds that provide continuity for the camera move. Tilt and zoom shots are more common in animation as they often promote an exaggerated feel. Isomorphic cues are commonly used to score these types of camera movements. A *pull back shot* gradually moves away from a subject to reveal a new environment. This type of shot can be supported by morphing from one set of BGs to another or by opening up the sound field in the mix. Though not a camera move, the shift of camera focus from the foreground to the background, known as *pulling focus*, is a natural transition within a scene. The re-recording mixer can mirror this transition by shifting elements of the soundtrack from the sonic background to the sonic foreground.

Movement of Objects

Fly-bys and *fly-throughs* refer to objects moving in relationship to a fixed camera position to create the illusion that an object is passing left to right or front to back. Fly-bys are re-enforced by panning sound left or right as indicated by the on-screen movement. Fly-throughs can be realized by panning to and from the screen channels and the surround channels as indicated by the on-screen movement. When panning bys of either type, it is still important to observe the line. Processing the sound object with Doppler and/or adding *whoosh effects* can provide additional motion and energy to the visual by. Doppler is the dynamic change of pitch and volume of an object as it approaches and passes by the audience.

Perspective Shot (POV)

In a *perspective (POV) shot* we experience the action subjectively through the eyes of a specific character. Just as we are seeing with their eyes, so too are we hearing through their ears. A good example of POV can be heard in *The Ant Bully* (2006), where the audience views the action through the mask of a bug exterminator. This perspective was achieved by contrasting the exterminator's breathing (interior) with the muffled sound of the environment (exterior).

Insert Shots and Cutaways

An *insert shot* cuts from a shot framed at a greater distance to close-up shot. Insert shots are used to reveal detailed information like the time on a watch or a message written on a note. The insert shot can be further exaggerated by briefly suspending the time of the reveal by *freezing the frame*. One effective design approach to freeze-frames is to cut the SFX and dialogue (diegetic elements) but play the underscore (non-diegetic and linear). A *cutaway shot* moves to a framing of greater distance, providing information from larger objects like a grandfather clock or a billboard sign. Sound elements that hit with a reveal are typically timed a few frames late to suggest the character's reaction time. If the soundtrack does not hit when the information is revealed, it is said to *play through the scene*. In the early years of animation, music cues hit on much of the action, a technique referred to as *Mickey Mousing*. This technique is used more sparingly today and many directors view its use in a negative context. However, in the hands of a skilled composer, cues can hit a significant portion of the action without calling attention to the technique.

CUTS

A *cut* is the most common type of edit and consists of two shots joined without any transitional treatment. In Walter Murch's book *In the Blink of an Eye* (2001), Murch discusses the cut as a cinematic experience unique to visual media and dreams. As part of our "willing suspension of disbelief," audiences have come to accept cuts as a normal part of the film experience. Sound editors often use the terms *audio clip* and *audio region* interchangeably. When discussing sound edits, they refer to the beginning of a clip as the *header*, the end of the clip as the *tail*, and a specific point within the clip as a *sync point*. The point where two clips are joined is called the *edit seam* (Figure 2.5).

The question of where to cut involves timing and pacing. The question of why to cut involves form and function. When a visual edit interrupts a linear motion such as a walk cycle, the edit jumps out at the audience. *Jump cuts* present a unique challenge for the sound editor in that any attempt to hard sync sound to the linear motion will further advertise the edit. If sound is not required for that specific motion, it is better to leave it out altogether. However, if sound is essential, the sound editor must find a way to *cheat* the

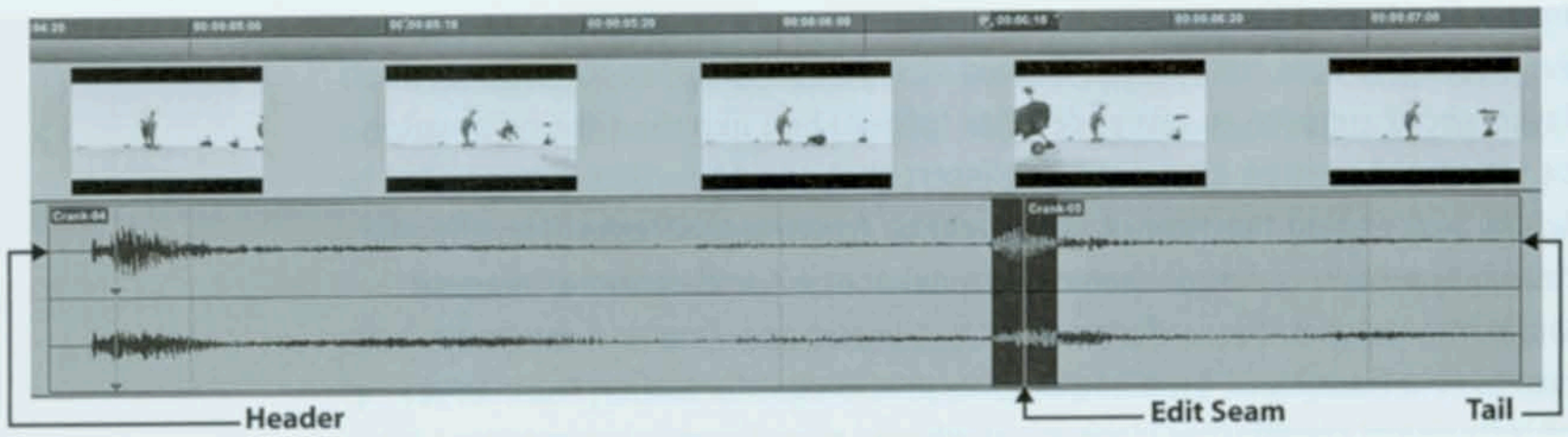


Figure 2.5 Audio Clip Header, Edit Seam, and Tail

sound in an effort to minimize continuity issues. This is a situation where linear music can be used effectively to mask or smooth the visual edit.

TRANSITIONS

Dissolves

A visual *dissolve* is a gradual scene transition using overlap similar to an *audio cross-fade*. Dissolves are used to indicate the passage of time or a change in location. The sonic equivalent of a visual dissolve is an audio cross-fade, a technique commonly used in connection to visual dissolves. In *Antz* (1998), as the male character Z waits excitedly for his love interest Princess Bala to join him on a date, the instrumentation of the underscore gradually thinned through the many dissolves, finally ending on a solo instrument to play his isolation and disappointment for getting stood up.

Wipes

A *wipe* is a transitional device that involves pushing one shot off and pulling the next shot into place. Unlike the dissolve, there is no overlap in a wipe. Blue Sky uses wipes cleverly in the *Ice Age* films, integrating seasonal objects like leaves and snow to re-enforce the passage of time. This approach lends itself to hard effects, BGs, and score. In the *Flintstones* television series, wipes are frequently used and often scored with cues timed specifically to the transition.

Fades

A *fade* uses black to transition the audience in and out of scenes. Fade-outs indicate closure and are timed to allow the audience time to process what has recently transpired. Closure can be supported through music cues that harmonically resolve or cadence. In many episodes of the *Flintstones*, Hoyt Curtain's music cues would half cadence prior to the first commercial break and fully cadence prior to the second commercial break indicating the resolution of the first act. Fade-ins invite the audience into a new scene. BGs, underscore, and dialogue are often introduced off-screen as the film fades in from black. The mixer typically smoothes fade transitions by gradually increasing or decreasing levels in the soundtrack.

Sound Transitions

Not all sound edits are motivated by the literal interpretation of picture edits. As discussed earlier, pre- and post-lap techniques are useful transitional devices. Pre-lapping sound or score to a subsequent scene can be an effective smoothing element when moving between shots, whereas a hard cut of a SFX or BG can advertise the transition and quickly place the audience in a new space. Post-laps allow the audience to continue processing the narrative content or emotion of the preceding scene. In many cases, audio simply *overlaps* from cut to cut. This technique is used for dialogue, SFX, and source music. A *ring-out* on a line of dialogue, music cue, or SFX can be a very subtle way to overlap visual edits. Ring-outs are created by adding reverb at the end of an audio clip, causing the sound to sustain well after the clip has played. This editing technique will be covered more extensively in Chapter 6. Dialogue is sometimes used to *bridge* edits. For example, in *Finding Nemo* (2003), the backstory is delivered through a series of characters handing off the dialogue in mid sentence from cut to cut. With each of these transitional treatments, there are no hard fast rules as to guide the editor. This is why editing is both an art and a craft.

SCENES

Parallel Edit

Parallel editing or *cross cutting* is designed to present two separate but related characters or actions through a series of alternating shots. This editing approach often implies that the actions are occurring simultaneously.

Though the audience is made aware of this relationship, the characters are often unaware; this is an important design consideration. When a music cue plays through the scene, it typically indicates that characters on both sides of the edit are aware of each other and the presenting conflict. If the sound design and score are made to contrast with each cut, this typically indicates that the characters are unaware of their shared experience.

Montage Sequence

A *montage sequence* consists of rapid visual edits designed to compress the narrative. Montage sequences typically feature songs or thematic score with little or no dialogue or SFX. The songs used in montage sequences often contain lyrics that relate to the narrative *(narrative lyrics)*. Songs or thematic score contain melodic lines that create linear motion across the non-linear video edits. In a scene from *Bee Movie* (2007), Barry B Benson falls asleep in the pool of honey. During the dream sequence that follows, Barry's relationship with love interest Vanessa Bloome develops through montage and is scored with the song "Sugar Sugar," a clever play on words.

Time-Lapse and Flashback Sequences

Short form animation often seeks to compress the narrative through the use of temporal devices such as time-lapse and flashback sequences. Time-lapse is an effective means of compressing the narrative. It differs from the montage largely in the way individual shots are transitioned and in its use of diegetic sound. If the sequence does not feature diegetic elements, than underscore is an effective means of promoting continuity. When diegetic elements are the focus of each transition (e.g. a television or radio), then sound edits are often deliberately apparent and designed to exaggerate the passage of time. Flashback sequences are an effective means of delivering backstory. Typically, a flashback sequence has a more subjective look and feel than scenes in the present time. This contrast can be achieved by adding more reverb to the mix or by allowing the underscore to drive the scene (Figure 2.6).

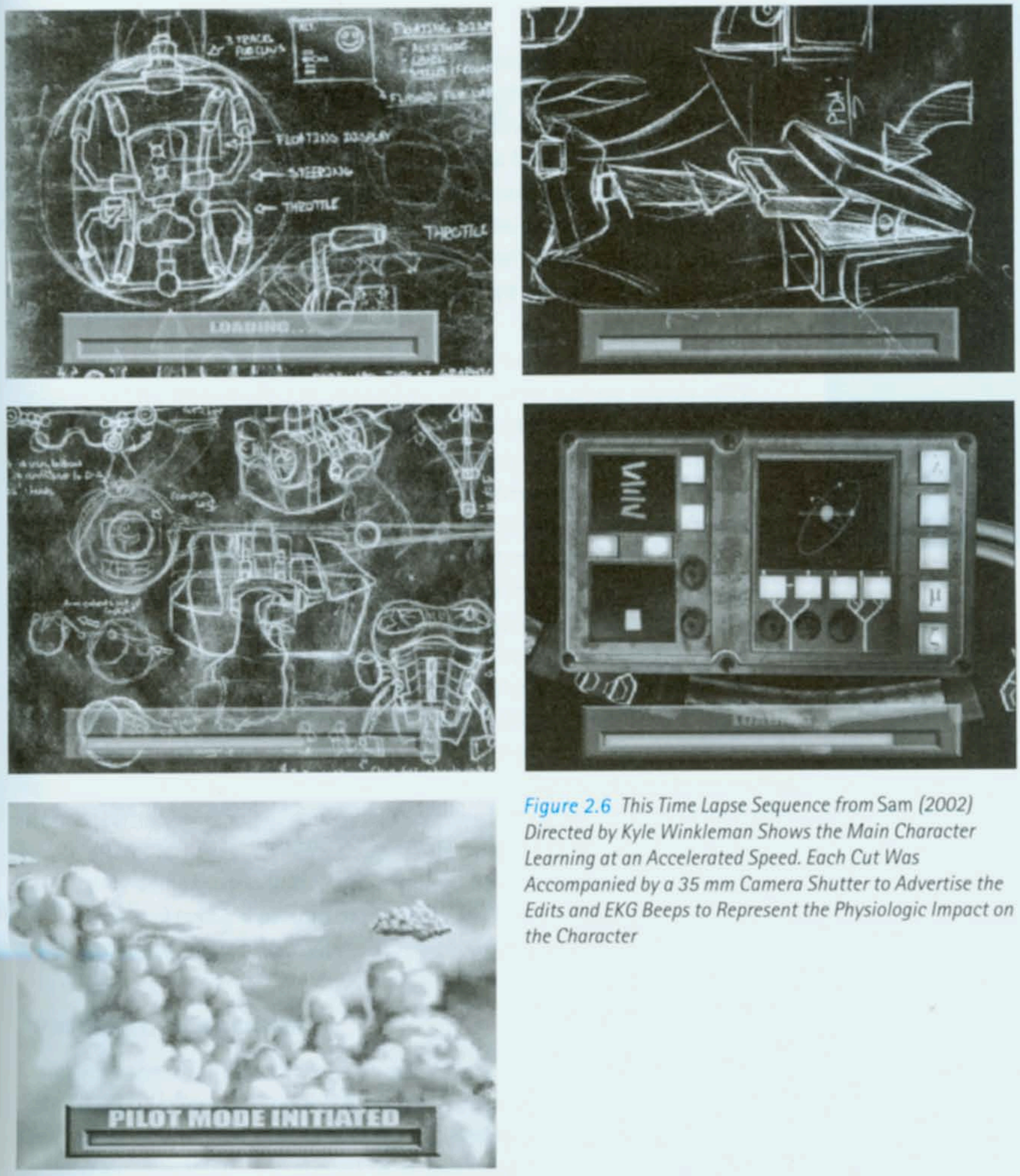


Figure 2.6 *This Time Lapse Sequence from* Sam *(2002) Directed by Kyle Winkleman Shows the Main Character Learning at an Accelerated Speed. Each Cut Was Accompanied by a 35 mm Camera Shutter to Advertise the Edits and EKG Beeps to Represent the Physiologic Impact on the Character*

If something works when it "shouldn't" that's when we have to pay attention and try to figure out why. And that's when real learning about editing results.

B.J. Sears

CONCLUSION

The term sound design implies the thought process that begins at the storyboarding stage and continues through the final mix. It is an inclusive approach to soundtrack development that facilitates both technical and creative uses. In film, sound and image are symbiotic. Consequently, sound designers and composers must integrate visual thought into their creative process as animators must integrate aural thought into theirs. The ability to design and execute a film from this mutual perspective will be seen, heard, and, most importantly, felt by the audience.