Digital Audio Production

A Trip to the Studio

Our client wants a half-minute narration over music to play when a visitor hits his web site. Tomorrow. No problem, we just booked an hour of time at the audio studio this morning. Here's how the trip to the studio went.

8:03 A.M.

When we arrived, the receptionist immediately offered us coffee, soda, or juice. My partner and I both thought juice sounded good, so the receptionist told us to go on back and raid the fridge. The engineer met us there, poured himself a cup of coffee, and led us to a comfortable listening room. The walls were lined with CDs. "I cued up five cuts from the production music library. They all have the sort of groove you described. Here's number one." We listened to them all and picked number three. It was perfect. "That's $135 for a needle-drop," he said, grabbing the disc and dashing out. We followed him down the hall to Studio A.

8:14 A.M.

We sat behind the console, and the engineer popped the CD into the computer. He ripped the cut we selected from the disc and converted it into a WAV file. Then he opened ProTools and placed the file in the first track. The voice talent had entered the booth, and the engineer went in with a glass of water and the script my partner had given him. He seemed pleased that the script was short and printed in a 20-point font. Then he adjusted the microphone, pointing it down towards her mouth about six inches away. He told her to put on the cans (headphones), and came back into the studio to set the levels.

8:22 A.M.

She read through the script perfectly the first time. The engineer pushed the talkback button on the console, and asked her to move into the mic slightly and to speak a bit faster. He had noticed that the duration of the take would be a few seconds too long, and it had to fit in 30 seconds. The third take had the right feel, and we thanked the talent. The engineer noted that we could have been there a long time with an amateur voice, at $175 an hour for Studio A and $300 an hour for the talent.

8:34 A.M.

The next step was to drop in a few sound effects. The engineer opened a database with thousands of samples, neatly categorized and indexed. We auditioned several and chose three that would enhance the mix. He laid them into open channels of ProTools and quickly performed his magic at mixing levels. Then he compressed the dynamic range and played back the finished mix for our approval. It sounded great, so we had him burn a CD-R with the high resolution original along with compressed versions in MP3, QuickTime, and RealAudio formats.

8:55 A.M.

We left the studio smiling because we had a professional product under budget and on time. "Same billing address?" asked the receptionist. I replied, "Yes, thanks. You guys saved our lives again!"

Audio Parameters

The engineer that did our recording had years of experience with recording gear and digital systems. He also knew a lot about sound. Anyone who undertakes an audio project needs to have a grasp of the physical properties of sound and concepts of acoustics. A sound consists of waves of varying pressure, or vibrations, in the atmosphere. Two important features of every sound are frequency and amplitude. These are factors we can measure and edit digitally.

Frequency is pitch. Higher pitched sound waves move faster than lower pitched sound waves. The length of a waveform determines its frequency. Frequency is measured in *hertz* (Hz) or cycles per second. A single cycle includes both the peak and the trough of the wave. The harmonic series consists of combined waves, each moving at a multiple of the fundamental lowest frequency. This series of overtones, referred to as partials, is the genesis of the tones found in musical chords. The range of human hearing is generally between 20Hz and 20KHz. As a point of reference, middle C on a keyboard is about 256Hz.

Amplitude is power, the volume, intensity, or loudness of a sound. The height of a waveform determines the amplitude. Sound pressure is measured in *decibels*. An electronic signal representing a sound is also measured in decibels, but in this case, the decibel is a reference voltage indicating the relative strength of the signal.

Dynamic range is the difference between the loudest and the softest levels in a sound track. *Stereo* sound consists of two separate channels of audio. Phase relationships exist between these two channels, and there is a possibility that sounds coming from different channels may cancel one another if they are out of phase. In this context, phase refers to the point in time that one wave begins compared to another.

Acoustics play a role in how we deal with sound when recording or playing it back. Hard surfaces reflect sound and create natural reverb in a room. Acoustic treatment, such as curtains and foam, is often applied to the walls of a recording studio. *Dry* sounds are devoid of reverb or other extraneous content, and *wet* sounds are heavily processed with effects.

Production Tools

In order to define a sound digitally, we must convert it from its natural condition, which is analog. Anything that is analog, such as sound pressure, can vary over an infinite range without finite gradations or discrete levels. A knob is an analog controller and can be smoothly turned to control the volume of an amplifier. A digital control over the amplifier gives the user discrete levels to select. The sweeping second hand of a clock is an analog readout while a digital watch shows 60 discrete increments.

Pieces of analog equipment used for sound production and recording are microphones, mixers, and signal processors. If a processor is not integrated within a mixing board, it is referred to as an *outboard* processor. *Signal processing* is the term used to describe how a signal is manipulated as it moves in a path from a microphone to a mixer, through processors, and then to an amplifier and speakers. Processors include equalizers (EQ), amplifiers, filters, compressors, limiters, gates, and reverb units, to name a few.

Cables and Connectors

Signals are moved electronically between devices by means of cables and connectors. Cables vary in several ways. There are usually two wires and a shield or ground wire in an audio cable. In stereo applications, one is for the left channel, the other is for the right channel, and the third is for grounding. Cables should be well-shielded from external noise so that they don't behave like an antenna. The diameter of the wire in the cable can vary, typically from a narrow gauge of 20 to 24 to a wider gauge of 14 to 18. Generally, larger-gauge wires offer less resistance and the current flows more freely through them. Speaker wire, which resembles lamp cord, typically has two conductors and is not shielded.

A distinction is often made in audio connections between balanced and unbalanced lines or cables. Unbalanced cables have one conductor with a shield and carry a single signal. Balanced lines have two conductors and a shield. The two conductors carry the same signal, but the polarity is reversed in one. This means that they are 180 degrees out of phase with each other, reducing the possibility of interference.

Connectors attached to the ends of a cable may be male or female in design. A male connector, with pins, is typically called a plug. A female connector is called a jack, with receptacles rather than pins, may either be attached to a cable or mounted on a piece of equipment. The common types of audio connectors are the male and female XLR, ¼-inch phone, RCA (phono), and ⅛-inch mini plug. The phone plug or the mini plug may be stereo with a tip, sleeve, and ring (TSR) or mono with just a tip and sleeve.

Microphones

The microphone is a critical component in the recording chain, and a high quality mic is essential for a faithful reproduction of sound. Microphones may be broadly categorized by how they function, as either *condenser* or *dynamic*.

Condenser and dynamic microphones employ different techniques to reproduce sounds. Condenser microphones have an electrically charged diaphragm that moves in response to the varying pressure of sound waves. As the diaphragm moves, the capacitance value of the diaphragm changes. The capacitance change is converted into a low impedance electrical signal that is transmitted by the microphone. The circuitry converts the signal into a form accepted by the input of standard audio equipment. Condenser microphones require external power, provided by a battery, or “phantom” power, provided by a mixing board or pre-amp.

An electrical charge must be maintained on the diaphragm for a condenser microphone to function. The diaphragm of an electret microphone is permanently charged. Other condenser microphones use external power to keep the diaphragm charged.

Dynamic microphones require no external power to operate. They operate on the same principle as a loudspeaker. A dynamic microphone has a fine wire coil attached to the back of the diaphragm. The coil is surrounded by a magnetic field created by a permanent magnet in the microphone. As the diaphragm moves, the coil moves. The movement of the coil in the magnetic field generates the electrical signal.

Dynamic microphones are durable and relatively inexpensive because their construction is less complex, but they can produce a high-quality sound. Dynamic microphones have a broad dynamic range with minimal distortion at high sound levels. They are predominantly used in sound reinforcement applications. These high impedance mics use an unbalanced cable with a ¼-inch phone connector.

Condenser microphones are more complex, and therefore usually more costly, than dynamic microphones. They typically provide higher signal levels (volume) and broader and flatter frequency response (particularly at higher frequencies) and can be made extremely small without affecting performance. Condenser microphones provide the most realistic sound quality and are used for most studio applications. They use a balanced line with XLR connectors.

High-quality mics are available from many manufacturers. Some of the best-known names in the field are Shure, AKG, AudioTechnica, Neumann, Sennheiser, and Crown.

Each microphone has a pickup pattern that determines in what direction it is most sensitive to sound. Directional mics typically exhibit “off-axis rejection,” which means that sounds coming from somewhere other than the axis of sensitivity are not reproduced well. Common patterns are the cardioid, the hypercardioid, and the unidirectional. A shotgun mic is used to capture sound from a distance because it has a very narrow pattern, similar to the beam of a spotlight. A lapel mic, or lavaliere, is attached to the clothing of a presenter.

Signal Sources

The first step in creating a digital audio file is to convert a signal that consists of a series of voltages into a series of digits that faithfully represents the original signal with an analog-to-digital (A/D) converter. The converter performs sampling, quantizing, and smoothing. An assortment of analog equipment, such as microphones, mixers, recording decks, and CD-Audio players can provide the signal source.

The two most common signal sources are a microphone and the output of a tape deck or a CD-Audio player. These two sources have different signal levels. It is important to match the output level of the source with the level that the input is designed to receive. Digital signal levels and decibels are discussed in depth later. Sounds are usually sampled by connecting the source to a sound card on the computer.

Digital Levels

In acoustics, the decibel (dB) is used to measure variation in air pressure. In audio engineering, the decibel expresses the difference in intensity between two signals, or the ratio between the two powers. To double the power of a signal is to increase its level by 3 dB. To double the voltage of a signal increases its power four times, which results in a 6-dB increase. The reference value of a 0-dBm signal has been standardized as 1 milliwatt at 600 Hz in a 600-ohm line. This represents a level of 0.7746 volts.

When recording to magnetic tape, it is common practice to keep the level meters close to 0 dB, which fully saturates the tape. The level meters read in VU (volume units). This reading is based on the strength of the electrical current. Doing so ensures a high signal-to-noise ratio and allows some "headroom" to avoid distortion. Recording a few peaks in the "red" that rise above 0 usually doesn't cause any problems since the tape saturation point is not an absolute.

In the digital realm, where amplitudes are stored as discrete numbers instead of continuous variables, the saturation point is an absolute value. Instead of having a flexible and forgiving recording ceiling, the absolute maximum amplitudes are -32,768 and +32,767 in 16-bit audio. No signal can be stored with a value that exceeds these numbers. The input signal gets chopped down to these values and wave peaks are clipped off, resulting in audible distortion. Digital audio has absolutely no headroom. When you hit the "red" zone on the meter, the signal is clipped.

To determine the level at which a signal should be recorded digitally, the maximum possible sample amplitude is used as a reference point. This value (32,768) is referred to as 0 decibels or 0 dB. Decibels represent fractions logarithmically. The equation used to convert to decibels is dB = 20 log (amplitude/32,768).

Start with a sine wave with peak amplitude of 50 percent of full scale. Applying the equation, the result is 20 log (0.50) or -6.0 dB. When the amplitude of a signal is cut in half, 6 dB is subtracted from its dB value. Doubling the amplitude of a signal increases its dB value by 6 dB. The lowest peak dB possible is -90.3 dB. Decibels are used for convenience. It is easier to express a value as -90 dB than as 0.000030 (1/32,768).

A peak meter shows the maximum amplitudes reached during a recording in dB. It is a useful tool to determine whether a recorded signal has clipped. Peak meters are not as precise as RMS (root mean square) power readings when measuring loudness. The peak amplitude of a square wave is much higher than that of a sine wave using the RMS method of measurement. On an RMS meter, a maximum amplitude square wave reaches 0 dB. A maximum amplitude sine wave reaches -3 dB.

If the loudest section of an audio track can be determined in advance, recording levels can be set so that the peaks are close to 0 dB and the dynamic range of the digital medium is maximized. In most cases, the loudest level is unknown, so it is safe to allow at least 3 to 6 dB of headroom for unexpected peaks. Headroom can be defined as the amount of additional saturation a recording device will tolerate after its meters read 0 dB and before distortion occurs. A digital recording system has no headroom, so it is best to begin recording below 0 dB to allow for unexpected peaks. Some DAT recorders show a reading of -18 dB as the nominal level. This would be equivalent to 0 dB on analog tape recorders with that much headroom.

Digital Recording

Computer Sound Cards

There are many sound cards available for capturing audio on the personal computer. They all contain analog to digital converters, and the quality of these converters determines the quality of the sound and the signal-to-noise (S/N) ratio. An audio card typically has two inputs and two outputs. One input is for a microphone, which reads a low level, and the other is for line-level input, which is a higher voltage. Since the signal from a microphone is so low, the sound card has a built-in pre-amp that boosts it. On most sound cards, these pre-amps are not of high quality, and a cleaner signal can be achieved from an external pre-amp. Of the two outputs, one is a lower-level line output for recording from the sound card, and the other is amplified slightly for headphones. The Macintosh has traditionally had built-in sound recording and playback capabilities, but the iMac, G4, and other recent models require an external A/D converter. Most PC audio cards, such as the Sound Blaster from Creative Labs, as well as Turtle Beach cards and Yamaha cards, can all record audio files with decent quality. DigiDesign is a manufacturer of professional audio processing equipment, such as the AudioMedia III card. Most professional studios use an outboard dedicated A/D converter.

Sampling

Digitizing sound is called sampling. Two important variables that may be controlled when sampling are the bit rate and the sample rate. Common bit rates are 8-bit, 16-bit, and 24-bit. An 8-bit sample has very poor audio quality. A 16-bit sample is the standard for CD audio and delivers high fidelity. Professional quality audio is sampled at 24 bits or higher. The sample rate determines how many times per second a wave is analyzed and recorded. The sample rate must be twice as fast as the highest frequency that appears in the sound track sampled. This principle is referred to as the "Nyquist theorem." The most common file types for digital audio are .wav for Windows, .aiff for Macintosh, and .au for Unix.

Each increase in bit rate doubles the size of the data file.

- 8-bit = 256 available integers to define a sound parameter
- 16-bit = 65,536 available integers (256 times better!)
- 24-bit = 16.7 million available integers

In converting from one bit rate to another, dithering noise may be introduced by software as it attempts to redefine the wave with less data. This noise is similar to the anomalies that occur in a dithered graphic that has been reduced from 16 bits to 8 bits.

Each increase in the sample rate also doubles the size of the file. The most common sample rates used in digital audio are 44,100; 22,050; and 11,025 samples per second. The Red Book CD-Audio specifies 16-bit, 44.1K samples. In stereo, one minute of data requires about 10 megabytes of storage space. That is why CD-Recordable blanks are defined as having "73 minutes" of storage space. Professional equipment, such as DAT recorders and high-end sampling cards, also sample at 48K and 96K. As a convenience, many DAT recorders will record in a "long play" mode at 32K.

The following table shows the amount of data in megabytes or kilobytes required for one minute of uncompressed audio at common sample rates and bit rates. The highest frequency found in a sample is half the sample rate, which means that an 11.050K sample has a maximum frequency of approximately 5.5 KHz. Because of the low quality of the 8-bit format, it is more useful for voice tracks than for music or complex mixes.

Sample Rate	Bit Rate	One Minute Stereo	One Minute Mono
48K	16-bit	11.346 MB	5.673 MB
44.1K	16-bit	10.350 MB	5.175 MB
32K	16-bit	7.564 MB	3.782 MB
22.050K	16-bit	5.178 MB	2.589 MB
22.050K	8-bit	2.592 MB	1.296 MB
11.025K	8-bit	1.296 MB	648 KB

Processing Sound with Software

Capturing the Sample

Several types of software packages are available for working with audio. Most of these perform both the capture and editing functions. When sampling audio, it is critical to monitor the input levels on a meter at all times. Audio capture programs provide a level meter or some way of viewing the input level in real time. If the input level is too low, the recording will be of poor quality and very noisy. If input levels are too high, the result will be distorted and peaks will be clipped off. Unacceptable distortion is introduced when input levels are too high.

Sound Forge, developed by Sonic Foundry, is a popular professional quality software package for capturing and editing audio in Windows. In addition to controlling the digitizing process, it offers the capability to perform a wide variety of processes and effects, to translate a file into a number of different formats, and to compress the file into many commonly used formats. A comparable program for the Macintosh is SoundEdit 16 from Macromedia. This application is also used to capture, edit, and manipulate sound files. For those who are on a tight budget or experimenting with audio production, there are many shareware packages available for download from the Internet for audio production. Professional audio engineers use programs such as DigiDesign ProTools for multitrack recording and mixing.

Applications for Editing and Processing

Once a sound has been digitally recorded, the first step is to evaluate the waveform on the screen while listening to it critically. It may be best to record it again if there are major imperfections. Listen for noise in the background, for pops, and for hiss. A "60-cycle hum" may be present, caused by faulty grounding of the AC circuit. There may be detectable "RF noise" that sounds like static. Look at the levels of the waveform on the monitor. If peaks are clipped off because they exceeded the maximum input level, the track will be distorted. It may be best to re-sample in this case.

A series of "takes" is pretty common for many reasons. The producer has choices between different versions of the content, and the engineer has choices between various signal levels. Once a usable waveform has been captured with the appropriate content, the next step is to trim

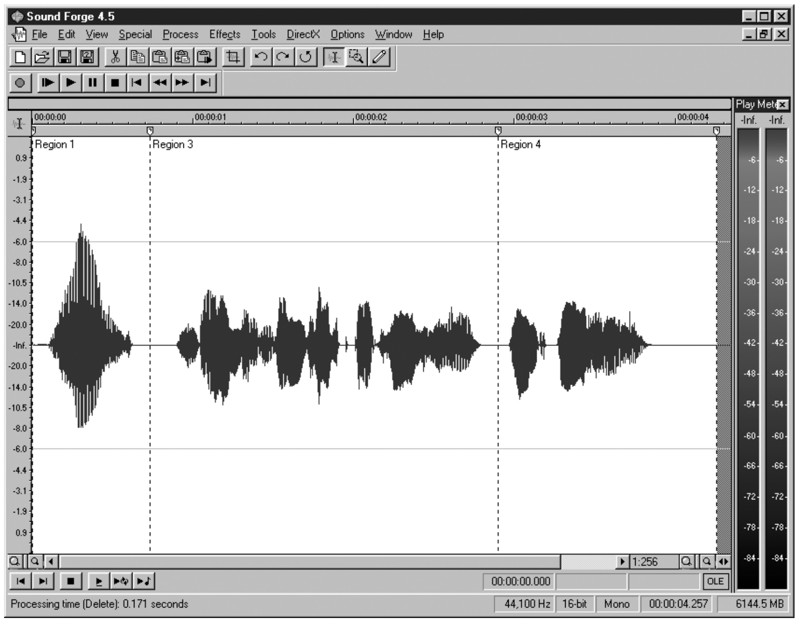

Figure 1—Sound Forge interface

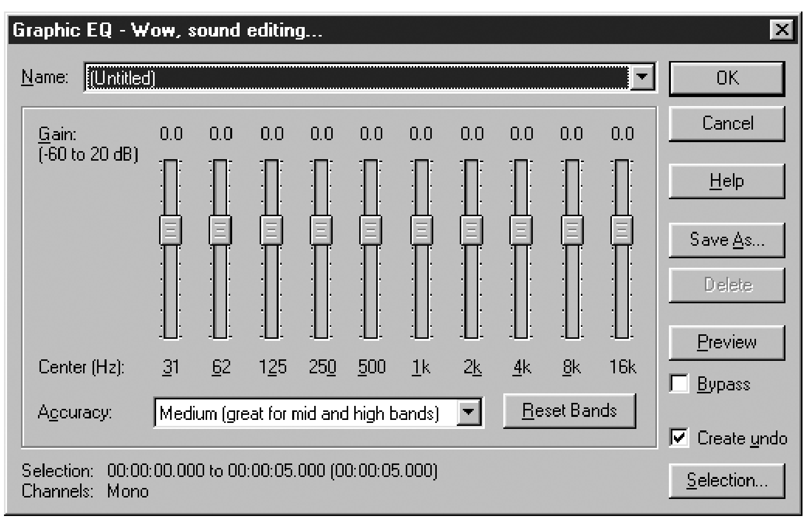

Figure 2—Sound Forge graphic equalizer

off dead space at the beginning and at the end. Cut as close to the program material as possible. After trimming the wave, determine which effects could be applied to improve the product, and audition them.

Most software allows the user to “undo” a process or effect that has been applied. This is referred to as “non-destructive” editing. Some editing software offers only one level of undo, which means that only the last change that was applied can be reversed. It is wise to save each version of the processed wave, if the software does not allow you to reverse a series of processes.

Undesirable pops and spikes can be eliminated by selecting the offensive portion of the waveform and reducing the level in that region. Is also possible to increase the amplitude in any region of the track.

When viewing a waveform in a software editor, the line in the center of the waveform represents silence or -90 dB. A full spectrum sound, which reaches the top of the window, is a strong waveform with a peak amplitude of 0 dB. Short waves that do not rise far above the center line are weak and may need to be boosted. The distance between the soft sounds and the loud sounds is referred to as the “dynamic range” of the track. If the track is “hot,” signal levels are uniformly high. This is a very desirable condition for sound used in multimedia and the web, particularly if it will be compressed later.

Signal Processing and Special Effects

The following are some typical processes that are performed on sounds to manipulate them or to improve their quality.

- **Normalize**—To normalize is to increase the loudest sound to a peak value or to a percentage of full spectrum and to proportionally increase the amplitude of all the sounds throughout the sample. Applying this effect can improve the signal level and the general presence of a track. It may be wise to select and reduce the strongest peaks in a track first so that the whole track can be increased by a greater percentage.
- **Equalization (EQ)**—Equalization affects the relative strength of a signal in a region of specific frequencies or “bands.” A parametric equalizer allows the user to identify a narrow frequency range and amplify or attenuate just the sounds in that frequency range. The width of the band that is treated is referred to as the “Q” factor. For example, most tape hiss lies between 8 kHz and 12 kHz, so attenuating that frequency range may reduce the noise. Unfortunately, this can make the program sound dull, since all the other desirable high frequency content in that range is also reduced.
- **Filters**—These are applied to remove or reduce sound in a specific band of frequencies. Filters may allow sounds to pass, or they may reject sounds above or below a preset frequency.
- **Compressor**—This processor reduces the difference between loud and soft sounds, making the dynamic range smaller. It usually boosts soft sounds more than loud sounds.
- **Limiter**—A limiter prevents a signal from passing through the circuit above a specified level or limit. Limiting the level sent to a sound card when recording can eliminate clipped sounds, which result from peaks in the input level that are too high.
- **Pan**—To pan a sound is to move it between the left and right stereo channels. If a sound is panned equally to both channels, it sounds as if it is in the center of the aural panorama.

- **Gate**—A gate establishes a level below which quiet sounds are eliminated. It is used to remove tape hiss or quiet background noise on a track. Setting the gate level just above the "noise floor" will effectively silence the most quiet portions of a track, but aggressive application of this process can produce "pumping" and "breathing" as the gate cuts in and out.
- **Delay**—A basic delay line continuously creates a copy of the original sound then mixes it with the sound file to create an echo effect. The duration between the original sound and the echo is user-defined. A "slap-back" effect results from a setting of about half a second. Multiple echoes can be produced with a delay processor.
- **Reverb**—This effect simulates an acoustic space, such as a concert hall. Often the settings on a reverb processor are chosen by selecting the size and type of room that would dictate the reflections of sound.
- **Chorus**—This effect occurs naturally when two or more voices or instruments play the same note at the same time. Variations in pitch and intensity create a "shimmering" sound, such as that produced by a violin section playing in unison.
- **Vibrato**—This effect introduces small periodic changes, or modulations, in the pitch.

Storage Formats

Reel-to-reel magnetic tapes have a limited shelf life. Recordings in the DAT format last longer because the information stored on them is of a different type. A CD-Recordable typically has a life expectancy of 100 years. When performing frequent recording sessions directly to hard disk, it is a good practice to back up the data on the hard drive and defragment it often. A fragmented disk can lead to problems, such as storing parts of the same file in discontiguous sectors. A CD-Recordable is a cost-effective medium for backing up data, and the same blanks can be formatted for CD-Audio players. Optical media, such as Zip or Jaz cartridges, may also be used.

Some of the more common audio storage formats are:

- **Tape**—Magnetic tape was the first widespread storage format for audio. Tape is an analog format, and works by arranging particles on the surface of the tape that are analogous to the shape of sound waves. The quality of magnetic tape recording is dependent upon the speed with which the tape moves across the heads and the width of the track that is recorded. Cassette tapes are very low quality because the tape moves slowly and the width of recording tracks is extremely narrow. Considerable loss of quality occurs in each successive generation when "dubbing," or copying, from one tape to another.
- **Digital Audio Tape (DAT)**—Sound may be digitized by a Digital Audio Tape (DAT) recorder and stored as digits (1s and 0s) on the tape. A DAT recorder performs the analog to digital conversion when recording and digital to analog conversion on playback. The digital audio file can be transferred from one DAT to another or to a computer hard drive with no loss of quality using an AES/EBU or an S/PDIF interface.
- **CD-Audio**—Digital audio may be sampled by the sound card in a computer and stored directly on a hard disk drive. It may then be formatted for CD-Audio in the Red Book format. This format allows 73 minutes of 16-bit, 44.1K audio to be burned onto a standard blank CD-Recordable disc. When properly formatted, the product will play back on any CD-Audio player.

- **MiniDisc (MD)**—Another popular format is the Sony MiniDisc. The MiniDisc records very good quality audio, but it is not quite as pristine as a professional DAT recorder. The discs are smaller in diameter than an audio CD, and the recorders are durable and portable. Blank media may be reused many times.

Compressing Audio Files

Audio files are relatively large. The standard .wav, aiff, and .au formats are not compressed. When audio files are compressed, some quality is lost, and the degree of compression determines how much data is thrown out. When preparing an audio file for compression, there are some processing techniques that will greatly improve the quality of the result. Begin with a file that is normalized to full spectrum with a very high signal-to-noise ratio, preferably compressed within a narrow dynamic range. It may also be helpful to cut the highest and lowest frequencies before compressing since this data is usually lost.

There are several commonly used *codecs* (compression/decompression algorithms) to make a file smaller. To compress a file is to encode it into a different file type, and users need to have the application or plug-in resident on their machines to decode the file. The most widely used formats are listed below.

- **MP3**—MPEG-1, Layer Three, is the compressed audio format developed by the Moving Picture Expert Group for use with MPEG-1 video. The suffix .mpg is used to identify any type of MPEG-1 file. It has become popular as a method of sharing music on the web. Applying this codec can reduce the file size to a small fraction of its original size without significantly reducing quality. This file type has received much attention from record labels concerned with illegal piracy of copyrighted music. Among the various algorithms available for encoding MP3 files, the Fraunhoffer is a recognized standard. Many MP3 players are available as shareware, such as M-Player and Sonique. The Windows Media Player and QuickTime can both decode MP3 files.
- **RealAudio**—This codec has been used since 1994 for delivering sound files over the web at narrow bandwidths, with the extension .ra. Developed by Progressive Networks, it is among the family of audio and video codecs in the evolving RealMedia (.rm) family. The player and encoder are both freely available from prognet.com. It is necessary to have the RealMedia server installed in order to stream audio or video files to a client, but the player can be used as a stand-alone device. Several general-purpose audio players are able to decode RealAudio files.
- **QuickTime**—This venerable architecture supports all types of streaming media, and there are numerous codecs that may be applied to audio files. The extension for all types of QuickTime files is .mov. The version of the QuickTime player installed on users' machines determines whether they will be able to decode the file. Two popular QuickTime version 4.0 codecs are QDesign Music and Qualcomm PureVoice.
- **Windows Media**—These codecs are capable of high-quality compression in either audio or video formats, applying the .asf suffix to the file. Windows Media Audio V2 compression creates smaller files than the Fraunhoffer MP3 compressor, with equivalent or higher

fidelity. A current version of the Windows Media Player should be installed on the client machine for best results. A full installation of Internet Explorer includes the plug-in.
- **Macromedia Shockwave and/or Flash**—Both of these compressors are similar to MP3, yet they are capable of streaming audio files embedded in a Shockwave or Flash movie. The latest plug-in for Shockwave/Flash is recommended for best results. A special server is not required to stream these audio files in real time.
- **Liquid Audio**—This proprietary codec is used to compress music. It is used by a number of independent labels and songwriters to distribute their compositions. It renders good sound quality, and the player is able to decode several other compressed file types.

Delivery of Sound Files on the World Wide Web

The Internet is a packet-driven network, not designed for streaming media. Media that must stream, such as audio, is time sensitive. For smooth continuous playback, all the bits need to be lined up and ready for decoding when the file begins to play. Otherwise, they must be buffered on the client computer at a fast enough rate to allow continuous playback while downloading progresses. RealMedia servers address these challenges in streaming data, as do the Windows Media Server running under Windows NT and the Macintosh QuickTime server. In all cases, the browser on the client computer must have the exact same version of the codec for decoding that was used to encode the file. It is wise to make users aware of any audio codecs they will need and provide a link for them to follow to download and install the proper codec.

Recording a Voice Session

Noise and distortion are the major concerns, and good hardware can make a dramatic difference in the resulting recording.

The first and most critical choice is the microphone. Each microphone will imprint different characteristics on a recording. Microphones supplied with sound cards are generally of very poor quality. For best results, use a condenser mic and a pre-amp connected to the line input of the sound card. Microphone pre-amplifiers built into sound cards are also of poor quality, so using the line input instead of the mic input will improve the results.

Sound cards vary widely in their quality of sound recording. Recordings made with inexpensive cards have excessive background noise, especially when the built-in microphone input is used. Professional cards are of higher quality and considerably more expensive. When evaluating a high-end sound card, make sure that it is MPC compatible. If the sound card has only DirectX compatibility, it may not work with your recording software.

Windows Volume Control

The Volume Control program built into Microsoft Windows controls all of the features of the sound card. The Volume Control program can be started by double-clicking the speaker icon in the "system tray."

When the Volume Control program opens, it displays the default playback volume controls. This display can be modified by making selections from the menu bar Options/Properties. You have the following controls:

- **Mixer device**—Your computer can have more than one sound card. Most recording software only works with the first card installed in the system (this should be the first card displayed in this selection box). You should not normally need to change this setting.
- **Adjust volume**—You can select Playback (the default), Recording, or Other. To control the playback volume, make sure Playback is selected. To set recording sources and levels, select Recording.
- **Show volume controls**—You can display all of the controls at once or select just the controls you are interested in using.

You may need to start your recording software and the Volume Control programs together. Changes made in the recording controls for microphone and line sources will also be displayed on the corresponding controls of the recording software. Changes made in the recording software will display in the Volume Control program as well. Some sound cards are supplied with a custom volume control program that displays the controls differently from the standard Windows Volume Control program.

Recording Techniques

Getting the best possible sound quality requires experimentation with a few recording parameters to find what works best in your situation. Voices should be recorded in a quiet room. You will need to isolate the mic from computer fans and other sources of noise. Use a "unidirectional" mic that picks up sound in a single direction instead of an "omnidirectional" microphone. Use the best microphone available and a microphone stand to eliminate the problem of handling noise added to the recording.

Position the microphone close to the person speaking, about six inches in front of his or her mouth. The exact distance will depend on how loudly the person speaks, the type of microphone, and the desired sound. Experiment with placing the microphone in different positions, such as directly in front of the mouth, above the mouth pointing down, or to one side of the mouth. Try to pick up as little environmental noise as possible.

Microphones increase low frequencies when placed closer to the mouth. Close positioning also increases detailed vocal sounds, such as wind noises from the popping of "p" sounds and the sibilance of "s" sounds. Placing the mic below the lips often tends to accentuate undesirable sounds. Changing the microphone position can help control these problems. Positioning a screen of nylon mesh between the mouth and the microphone can greatly reduce explosive sounds.

Monitor the Levels

With analog tape recorders, recording at the highest possible level before distortion will usually provide the best results. With digital recording in general, this is not necessarily true. *Watch the level meters at all times*. Adjust the input level as needed to keep it in a medium to high range on the level meters for most of the recording. If the clipping indicators light up, this may indicate that an overload of the digital signal has occurred. This can cause undesirable distortion to be added to the recording.

Audition the loudest sections of the material to be recorded with the recording software's level monitoring enabled. Monitor the volume level display and clipping indicators before making the actual recording. Speak with the same intensity during the level-setting process that you will use in the actual recording.

Maintain Consistent Levels

The presentation will be easier for the listener to hear if all of the words are spoken in a strong, consistent fashion, especially if the track is later compressed. The dynamic range of Internet audio and other highly compressed files is very limited. Words that are very soft in a sentence may be lost. Listen to the results of your recording session before quitting since it may be difficult to simulate the exact conditions again.

Sound Design

There are a number of reasons for embedding sound in media production and web sites. Sound serves the following basic functions:

- Ambient Sound: Background audio establishes an environment. Examples are the sound of birds in a forest, traffic on a busy street, crowd noise, factory equipment, or waves crashing on the beach.
- Underscore: This is a type of background sound dominated by music, which sets a mood. It may include sound effects in the mix. A logo theme is a more dramatic example of music used in a consistent fashion in a production. Different music tracks may be used to distinguish different segments of a production.
- Voice: Narration is one of the most common ways of communicating with a user. The voice might be a host, offering assistance with the operation of the program. In many cases, voice tracks are a significant element of the user interface, giving instructions or feedback to the user. When synchronized with video, a story is often told that clarifies the action on the screen.
- Sound Effects (SFX): This category includes short sounds, such as button clicks and transition sounds that lend interest to a program. Often, sound effects are used with animation to emphasize movements on the screen.

Recent Developments in Pro-audio

Dolby Digital and 5.1-channel Surround Sound are current standards for digital audio. Surround Sound provides two front channels, two surround channels, and a front center channel. The ".1" channel is for bass frequencies only, which are sent to a subwoofer. DVD sound tracks typically are encoded in the Dolby AC-3 format, which also delivers 5.1-channel audio. Expect more audio production and delivery to take advantage of these multiple channels. The specification for an audio-only DVD format to replace the audio CD has spent many years in the draft mode, due to rapid advances and different business models. It will provide higher fidelity and many hours of music on a single disc.