# Audio Pitfalls and How to Avoid Them

by Jeff Essex

Sponsored by Apple Computer, Inc.
Apple Developer Connection

media

# Audio Pitfalls and How to Avoid Them

by Jeff Essex

As multimedia producers, we wear a lot of different hats during our careers. Sometimes we wear them all during the course of creating a single presentation: graphic artist, writer, programmer, video editor and audio engineer. In my role as an audio producer for multimedia projects, working closely with other members of a production team, I'm often asked for advice on little audio-oriented tasks that are being tackled by other team members – small editing tricks, recording rough voice tracks for placeholders, etc. I also find myself in the position of having to fix or clean up the results of efforts by well-meaning folks who didn't quite grasp the consequences of their actions. For this reason, I thought it might be helpful to draft up a quick list of common audio pitfalls, and, more importantly, their solutions.

In my book, "Multimedia Sound and Music Studio," I laid the seven basic steps of multimedia audio production: Planning, Recording, Capture, Editing, Output, Integration and Playback. I'll use those steps as a guide for pointing out some of the spots you should watch for.

## Planning

Before you begin a sizeable production process, it's important to lay out guidelines for creating audio assets. If you're just creating a piece of background music and a few button clicks, this may not be necessary. But if you're working on a complex entertainment or training application, it's crucial. Planning can be divided into two main areas: production management and technical specifications.

On the production management side, take some time to think about overall design. Will the presentation be fast-paced and snazzy, or low-key and factual? The type of audio assets you choose should help support the message, complement the graphic design, and have the most impact on your specific intended audience. Create a schedule and a list of deliverables, and talk with other members of the production team to establish who's responsible for approving audio assets and distributing them to the programmers and artists.

On the technical side, talk with producers and programmer to specify file formats, including type (i.e. AIFF, WAV, MIDI, QuickTime, RealAudio, Shockwave, etc.) rate and resolution (16-bit 22 kHz for CD-ROM , 16 kBps for the Web, etc.) and file size. File size guidelines should include storage (i.e. space on a disk), memory (how much RAM will be required) and throughput (how fast the media can be delivered from disc or modem) .

Above all, agree on a base computer platform to be the "slowest common denominator" for playback. Designing within the constraints of this platform will help assure decent performance, particularly in the area of synchronization.

**Recording**

If you place yourself in the hands of an audio professional, you'll rarely have to worry about this stage of the process. But if you do it yourself, you might run into trouble in a few places: quality of equipment, background noise and acoustics, and microphone technique.

Here's a standard scenario: you're doing a quick presentation with a few pages worth of voiceover. Someone in marketing who "has a nice voice on the phone/ wants to break into voiceover work / DJ's weddings on weekends" would be perfect for the job. You invite them over to your desk, show them the script, plug a microphone into the back of your computer and you're in business, right? Well, maybe not. Keep in mind that your audience probably hears several hours of professionally produced voiceover every day, just by listening to the radio or watching TV. You need to get reasonably close to that level of quality.

*Equipment*

First, let's look at the equipment issue. The microphone that ships with your computer won't really provide the level of quality you need. It's okay for adding voice annotation to e-mail, chatting through an internet phone application or doing a rough voiceover for a proof of concept. But standard computer microphones don't provide high enough quality for creating a finished product. If you plan to do a fair amount of voice recording, consider investing at least $100 (and if possible, $400) in a microphone. Once you step up to a new microphone, you'll also need mixer to adjust volume levels before the signal gets to the audio input jack on your computer. You can't just use adapters to plug the mic directly into your computer, because the two devices have radically different voltage requirements. A microphone puts out a low voltage signal (in the range of -40dBm to -60dBm) while the audio input on your computer wants to see a "line level" signal of -10dBv (this is the same signal level used in consumer stereo equipment, camcorders, video decks, etc.). Inserting a mixer between the microphone and computer will help boost the power of the signal to the proper level. As a bonus, you may also get control over equalization, to help adjust the tonal quality of the voice while you're recording.

*Acoustics*

Now that you have a great microphone setup, you'll be in a better position to notice the next potential issue: background noise and poor acoustics. Audio engineers take great pains to minimize the amount of unwanted noise in a recording. The best time to reduce noise is before it ever gets recorded. This is why professional recording studios invest in isolated rooms with expensive acoustical treatments. Before recording a voiceover, try to find a place that's really quiet (conference rooms can be ideal). Then take a moment to really listen to the sounds in the room. You'll probably hear noise coming from computers and their peripherals, the ventilation system, and from outside the building. Try to identify the sources of the noise and see if you can eliminate them by closing doors and windows or temporarily turning off fans. If you're recording straight into a computer, try putting the computer outside the room and running a microphone cable back into the room. It might also be best to wait until after your co-workers have gone home for the day and the phones have stopped ringing.

Try to reduce reflected noise in the recording environment. Echoes and reverberation are like motion blur applied to sound. They reduce the clarity and intelligibility of speech. Sound echoing from hard surfaces like windows, tile floors, whiteboards or large tabletops can be lessened by draping blankets or cloth tarps. Carpeted rooms and drapes are great for minimizing reflections.

*Microphone Technique*
Next, experiment with the placement of the microphone in the room and its position relative to the person speaking. If the person is too close to the mic, it may pick up an undesireable amount of mouth noise (smacking lips, etc.) or be subjected to “popping” during plosive sounds like “p” and “b”. If they’re too far away, there will be a greater proportion of background ambient noise relative to the voice.

Finally, keep in mind that the easiest solution may be to use a professional recording facility. These folks will solve the production issues for you. Today it’s common for studios to have digital audio production tools like hard disk recording systems and CD burners, and many have the ability to deliver edited files on CD-ROM. If you use a professional studio and record to DAT (Digital Audio Tape), tell the engineer to record at a rate of 44.1 kHz (the CD standard) rather than 48 kHz (the DAT standard). This can save you a step of sample rate conversion later in the production process.

## Capture

Capture is the process of capturing audio into the computer from an external source. This can be done by recording directly into the audio input on the back of the computer through a microphone and mixer (as described above) or a VCR, camcorder, CD player, cassette deck or other similar consumer audio product. It can also be done by transfering digital audio data from a DAT recorder to a digital audio interface card in the computer. Finally, audio can be captured directly from audio CDs using QuickTime and MoviePlayer. Here are a few things to watch out for when capturing.

First, and most important, always be sure to capture at the highest possible sample rate and resolution (usually 16-bit, 44.1 kHz). This gives you the best quality audio for mixing and editing tasks. You can always downsample or apply data compression as a final step.

*Capture via the Audio Input*
If you’re recording into the computer’s audio input, beware of noise that can be added to the signal as it moves through audio cables on its way to the computer. Computers and their peripherals generate electromagnetic fields that can create buzz and hum in audio cables. The solution is to use well shielded, high quality cables. Keep the audio cables away from power cables by at least a foot. If you must cross over power and audio cables, do it at right angles to minimize the area exposed to interference.

Another source of noise is distortion caused by excessive volume levels. This is often called “clipping” because the peaks of waveforms are clipped off when they exceed the maximum threshold.

Once clipping and distortion occur, the noise is almost impossible to remove. If your audio source has wide variations in volume and you suspect clipping might occur, keep in mind that it's better to have the record levels a bit low, then amplify later. Don't take this advice to extremes, however. Your recording levels should average around 80-90% of maximum possible volume. Recording at extremely low levels (i.e. the maximum volume is never above 30% of the total possible) leaves your signal much more susceptible to unwanted noise.

*Capture from DAT*

DAT is a very popular medium for recording and transfering digital audio. It combines the convenience of tape and the quality of CD. With additional digital audio hardware (like Digidesign's Audiomedia III) in your computer, you can capture digital audio straight into the computer without going through the extra analog-digital conversions that can add noise to the signal. When transferring DATs across a digital connection, be careful when you adjust settings on the capture card. First, be sure the card is set to the same sample rate as the DAT (usually 44.1 kHz instead of 48 kHz). Also, be sure the card is synchronizing to the digital signal coming from the DAT rather its internal clock. This assures that each sample sent from the DAT is recorded with the correct time stamp. If either the sample rate or synchronization are off, you'll notice that the audio you transferred plays at a different pitch than the original audio coming from the DAT.

*Capture From CD*

Finally, capturing audio from CDs is a powerful feature embedded into QuickTime. Always check the capture options to be sure you're getting the data at the desired rate. Click the Options ... button (Figure 1), then check the Import Options dialog (Figure 2) to verify that your sounds will be captured and saved as 16-bit 44.1 kHz data, rather than letting QuickTime downsample the audio during capture.

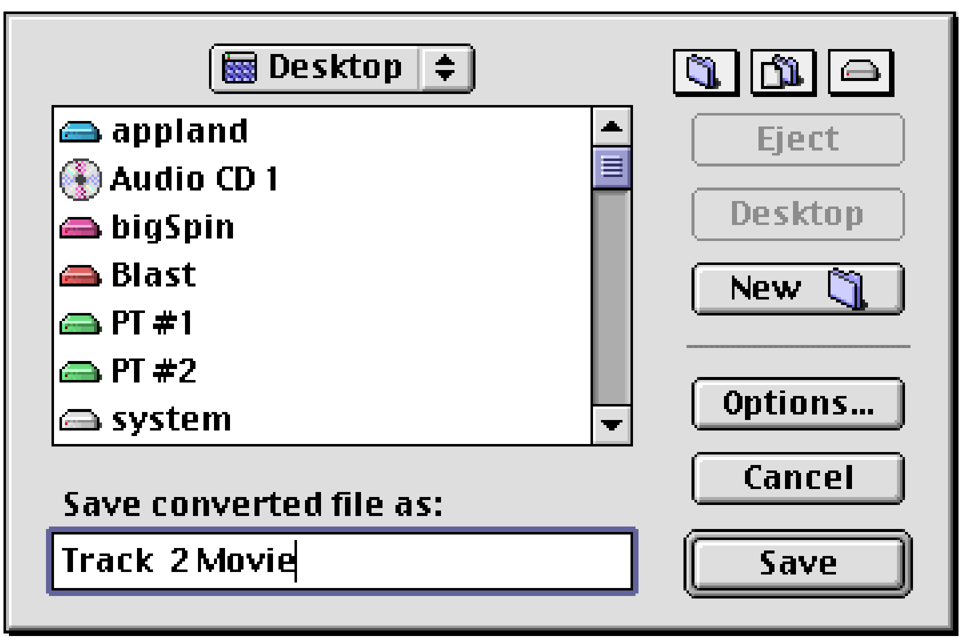


Figure 1. Note the Options...button when capturing audio from CDs

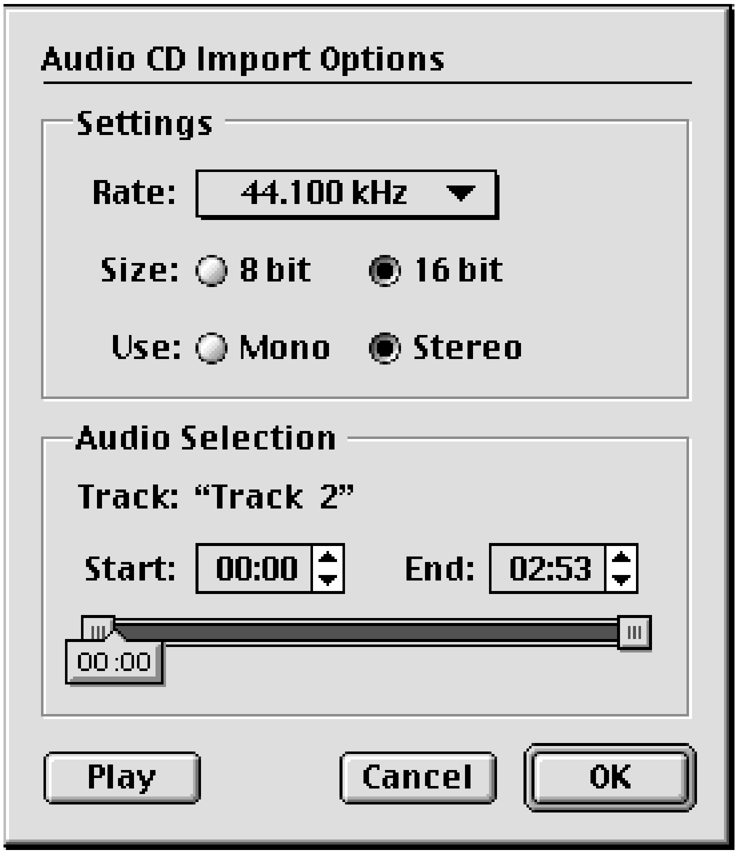


Figure 2. The Import Options Dialog lets you select sample rate, sample resolution and start/stop times for CD audio capture.

## Editing

As mentioned above, the cardinal rule is to do all your editing with the source material at the highest possible sample rate and resolution (usually 16-bit, 44.1 kHz). Here are a few other tricks to keep in mind while slicing a bunch of sounds into easily digested little files.

*Start and End with Silence*
Files should start and end with a tiny bit of silence. Don't overdo it: a few milliseconds will do, just enough to be sure that the audio data is at the zero axis at each end of the file. Otherwise you'll probably hear pops and clicks as the file starts or stops.

*Zap the Popping P's*
If a popping "P" sneaks into your voiceover, you can still fix it during editing, using tools like Macromedia's SoundEdit 16 or BIAS Peak. Open your audio editor of choice and locate the offending syllable; it should appear as a big spike in the signal. Select just the spike portion (audition it to be sure you've located the culprit) and use the "Amplify" or "Gain Change" effect to reduce the volume by about 50%. If that still isn't enough, apply EQ to the same section, reducing the low frequency content (below 250 Hz) to get rid of the "thump".

*Balance Volume Levels*

When you're creating lots of files, it's important that they all have the same relative volume. The best way to achieve this (particularly when editing a long voiceover session into shorter takes) is normalize the entire file before dicing it into smaller files. Normalization is a process that raises the volume of a file to the highest possible value short of clipping. It's much easier to normalize one big file than it is to normalize 100 little files. You'll get more consistent results, too.

This is also the best time to compress the dynamic range of your audio, especially if your voiceover wasn't recorded using professional voice talent in a professional studio (they will already have applied dynamic compression as the signal was being recorded to tape). Dynamics compression reduces the high peaks of audio spikes so that the overall volume of the file can be increased without clipping. While it's best to apply dynamics compression during recording (using a dedicated piece of audio hardware called a compressor), you can take advantage of several software tools that offer similar processing functions, particularly Waves' AudioTrack. Careful use of dynamics compression helps you get the best signal-to-noise ratio before downsampling or applying an audio file compression algorithm (like RealAudio or the QDesign codec).

(Note: Don't confuse the two types of compression. Dynamics compression reduces the dynamic range, lowering peaks in the amplitude of a signal. Audio file compression reduces the amount of data in a digital audio file, reducing file size to facilitate delivery over a modem.)

## Output

Output is the stage where final processing is applied to files as they are exported for use in a presentation. By this point your files should have had dynamics compression and normalization applied to them, to minimize the background noise and distortion created by downsampling and audio file compression. Thanks to advances in computer hardware (including CPU speed, available RAM and CD-ROM drive speeds), most multimedia presentations can support 16-bit 22 kHz audio. This is a tremendous improvement from the days when 8-bit 22 kHz or 8-bit 11 kHz was the norm. Be sure to avoid the old Mac standard sampling rate of 22.254 kHz when exporting finished files. Every PC on the market today prefers the 22.050 kHz rate.

If you're using an audio data compression tool like the QDesign music codec, RealAudio, Shockwave audio, or IMA 4:1, you should also keep your audio assets at 16-bit for the best results. These algorithms expect to see 16-bit input. Reducing your files to 8-bit won't save any space. Instead, you'll just add lots of quantization noise before the algorithm upsamples it back to 16 bits on the way to performing its processing tasks.

## Integration

Now you're ready to begin plugging your audio into a finished presentation. Whether it's a CD-ROM or Web site, there are several factors to keep in mind. In my experience, the most frequently overlooked issue is defining a playback strategy, and dealing with the ramifications of that strategy. In the CD-ROM arena, this boils down to choosing whether to play audio from RAM vs. spooling from the disc. On the Web, it's a question of downloading the audio vs. streaming.

This distinction is critical to the performance of your application, and should be determined by the style and pacing of the presentation. Here are some examples of how the playback of audio can be handled so that it offers the best experience to the end user:

1) A Shockwave game could be structured such that the game itself downloads and begins playing, then music begins streaming in the background during gameplay. Since the modem bandwidth is no longer needed for downloading other content, it can be dedicated to streaming audio.

2) A MIDI file could be downloaded to the viewer's computer, leaving the modem bandwidth available for streaming animations.

3) In a CD-ROM project, long segments of audio can be streamed from the CD, while animation and graphics are resident in RAM or are cached on a local hard drive.

4) In another style of CD-ROM project, audio is loaded into RAM before playback, leaving the disc free for random access of animation, graphics or streaming media like QuickTime video.

In short, it's a risky strategy to assume that both graphics and audio can stream simultaneously from a CD-ROM or through a modem without one element suffering at the expense of the other. Before you import assets into an authoring tool and integrate them into the presentation, consider how you'll juggle the assets during playback. In Director, for example, sounds imported directly into the cast are always loaded into RAM and played from memory. On the other hand, linked external sounds and sounds played with Lingo's `sound playFile` command will always stream from the disc. In this case, you need to choose your playback strategy before you begin importing the sounds.

This "preload vs. stream" limitation doesn't always apply, especially if you have a very high-perfomance base platform, or a controlled audience where you can specify the performance of the components in the system. But it's likely to be the case if your presentation is being distributed to a broad audience. One obvious exception is when the media are merged into a single linear stream as with QuickTime or RealFlash. While merged data streams are workable for linear presentations, they're not always useful for interactive presentations that may need to access graphics independently from the audio.

## Playback

Now you're ready to hear the final results of your work. How good will it sound on the broad range of audio playback hardware that's available today? Most Macintoshes still ship with one internal speaker. Many PCs ship with speakers bundled right in the box, and the output of these can range from what could charitably called "marginal" to sub-woofer equipped systems that rival the performance of a home stereo system. How well your audio translates across these systems will depend in large measure on the quality of your monitoring system. Audio monitors are like the lenses through which you "view" your sounds. If your monitors can't reproduce bass frequencies very well, you may be surprised by what comes out of systems with subwoofers. On the other hand, if your monitors have lots of power in the midrange frequencies but not enough on top, you may overcompensate when boosting the EQ on high frequencies, creating sounds that are harsh and overly bright on other speakers.

If you're going to be doing a lot of audio production, invest in the best monitors you can afford. In the past few years, there has been an explosion in the number of high-quality, self-powered monitors in the $1000-$1500 range. If most of your work takes place in a shared office space, you may want to use headphones instead. Don't trust this job to the little foam-padded things that came with your portable cassette player. Be prepared to spend at least $100 to get a good pair of "closed ear" headphones, that isolate your ears from background noise in your workspace. When shopping for monitors, look for models that offer flat frequency response: that is, the monitors themselves shouldn't raise or lower any areas of the frequency spectrum when a signal is passed through the monitor. Very few home stereo speakers stick to this approach: most will boost low bass frequencies and high frequencies to make music sound "warm" and "sparkling." Look for studio reference monitors at your local professional recording equipment store rather than a traditional stereo retailer. You should also look for shielded monitors – the extra shielding helps prevent the magnets in the monitors from creating distortion on your computer screen when the speakers are placed on either side of the computer screen. Again, thanks to the boom in personal studios, digital recording and computer technology, there are a number of good products on the market.

## Conclusion

It's a daunting task to keep up with the pace of changing technologies and tools that are available to digital content creators. Keeping up with one area is hard enough, but multimedia developers have to work overtime to stay current in multiple technologies. When you're in the position of trading in one of your many hats for a pair of headphones, remember that your work could either make life a lot easier or a lot harder for the next person in the process. While it's impossible to hope for perfection in the development cycle, if this article helps you avoid at least one potentail pothole on the way to shipping a product, the effort will not have been in vain.

## Resources

Digidesign: http://www.digidesign.com
(Audiomedia III digital audio capture card)

Event Electronics: http://www.event1.com
(digital audio capture, microphone and reference monitor manufacturers)

BIAS: http://www.bias-inc.com
(Peak audio editing software)

Macromedia: http://www.macromedia.com
(SoundEdit 16 audio editing software)

Waves: http://www.waves.com
(AudioTrack audio processing plug-in for Peak and SoundEdit 16)

Sweetwater Sound: http://www.sweetwater.com
(professional recording equipment by mail order)

Mackie Designs: http://www.mackie.com
(1202 mixing console and HR824 self-powered reference monitors)

Mix Magazine: http://www.mixonline.com
(a good source for product reviews and production tecthniques)

## About the Author

Jeff Essex, Creative Director of audiosyncrasy, specializes in the use of digital audio and MIDI to create music, sound effects and voiceover for multimedia. In addition to his creative and audio engineering skills, he is intimately familiar with multimedia tools and technology. A veteran of MacroMind and Macromedia, he served two years as Technical Support Lead for Macromedia Director and SoundEdit. He is credited on over 40 CD-ROM titles, including products from Virgin Sound and Vision, Corbis Productions, Mindscape, 3DO, and Disney Interactive. For the past year he has worked as the primary composer and audio consultant for the leading children's interactive web site. His book, Multimedia Sound and Music Studio, (Random House/Apple New Media Library,1996) is the definitive guide to multimedia audio production. It won the Computer Press Award for Best Advanced How-To Book of 1996.