

Chapter 7: Working with Digital Sounds

Chapter contributed by Erik Larson

While the world of multimedia has long been enriched with sound effects and music, the World Wide Web has lagged behind. This is not because web developers deem the use of sound frivolous, but because of the combination of large file sizes and inadequate network speeds. Today, however, tools for delivering audio have come a long way, and now millions of users have the software tools and network connections to listen to audio over the Internet. Moreover, there are inexpensive and easy-to-use tools for capturing, editing, and serving digital audio over the Internet. No community networking team—no matter how small—should shy away from digital audio if they have a project that demands it.

Bandwidth is the term used to describe the amount of data a network connection can receive and interpret (such as through a modem and a phone line). A few years ago, a user usually had to wait several minutes or even up to an hour to receive a high quality sound file. With the use of compression, and, more recently, streaming techniques—downloading only a portion of a sound file at a time and playing only that portion—we are now better able to include music and speech as a part of a web site without causing frustration for the end user.

This document describes the tools and techniques for adding high quality speech and/or music to a Web site. Part I discusses digital audio and typical computer formats, Part II deals with basic recording techniques, and Part III explains how to include these files into a site on the World Wide Web.

Table 1. Approximate file sizes of a one-minute sound file

Sound Quality	Stereo 16 bit	Stereo 8 bit	Mono 16 bit	Mono 8 bit
CD (44.1 kHz)	10 Megabytes	5 Megabytes	5 Megabytes	2.5 Megabytes
FM Radio (22.05 kHz)	5 Megabytes	2.5 Megabytes	2.5 Megabytes	1.25 Megabytes
AM Radio (8 kHz)	1.8 Megabytes	900 Kilobytes	900 Kilobytes	450 Kilobytes

Fundamentals of Digital Audio

When you enter the realm of digital multimedia, capturing (digitizing) content can require an enormous amount of hard disk space and processing power. Fortunately, the demands of digital audio are considerably less than those of digital video. To explain the problem with the amount of hard disk space, Table 1 shows various sound formats and the corresponding file sizes. The sound quality column above uses the sampling rate of the sound file. When digitizing a sound file (digitizing is the term used to convert an analog signal into a digital

signal), there are certain items to consider. Instead of capturing the sound waves of an acoustic signal (such as a human voice) through grooves on a record or impulses on a magnetic tape, an analog to digital converter must convert the fluctuations in air into numbers in this case, zeroes and ones, hence the term digital. The 16 bit and 8 bit columns in Table 1 simply describe how many bits of information are used to accurately depict a waveform and are called the "resolution" of the sound file. In other words, the higher the number of bits the more numbers the better quality of the sound file. This is analogous to the visual world, where the higher the resolution of a monitor (8 bit, 16 bit, 24 bit), the more colors it can display. In sound, the higher resolution translates as a better representation of the waveform, resulting in better sound.

Sampling rate is a bit more complicated. In 1938, Henry Nyquist proposed a theorem for digital signals that says that the sampling rate of a file (the number of snapshots of a waveform a computer calculates each second) must be at least twice the highest frequency or pitch. Nyquist's Theorem tells us that higher sampling rates yield better sound quality. There are limits, however. Human beings can at most hear frequencies in the ranges of 20 Hz (hertz, or cycles per second) to 20,000 Hz. The standard sampling rate of compact discs is 44,100 Hz more than twice the highest frequency we can hear.

If you intend to record music, your choice of sampling rate is crucial since musical instruments have the ability to produce a wide range of frequencies, from the low notes of a string bass to the high notes of a piccolo. For critical listening to music, you will want to choose a sampling rate as close to the CD rate as possible. If recording speech, however, a lower sampling rate will suffice since the range of frequencies in a human voice is closer to the lower end of the frequency spectrum.

After a sound file is "captured" by a computer, the file needs to be saved into a standard format. Two uncompressed digital audio formats, WAV and AIF, are the most ubiquitous uncompressed sound formats; they can be used in Web pages without any special server. Keep in mind, however, that the sizes of the files can be quite large. For example, a mono 8 bit sound file at AM radio quality (roughly 8 kHz) will take up approximately 450 Kbytes per minute. A 10-minute speech would be about 4.5 Megabytes, which would take approximately 21 minutes to download using a 28.8 kilobit/second modem. To eliminate this long download time, the technology of streaming was developed. (See chapter 8 for more information on streaming audio and video.)

The tools needed to digitize and/or record speech and music are relatively inexpensive. The main items to purchase are a microphone, a sound card, and audio editing software. Many computer systems come bundled with one or all of the above items so check your system. You might not need to purchase anything. If you do, here are a few suggestions. Many of the microphones that are available for computer systems are adequate for speech, but not for music. If your quest is to put up a weekly speech by the mayor of your town, this will work fine.

If you want to be sure to capture the best quality, you might want to opt for a higher quality microphone. One type of microphone, called a PZM (pressure zone microphone), is ideal for speech and can be easily placed on a podium or table. Radio Shack has a decent quality PZM that is available for about \$60 and runs on a 1.5 volt AA battery. Whether you record

into a cassette deck, a MiniDisc recorder, or directly into your computer, you will probably need some adapters, which are also available at your local Radio Shack. If you wish to record speech only, you'll only need one microphone with the appropriate adapters since stereo is not necessary for speech.

If you're interested in recording music, your work will be much more complicated. Music recording is a real art and will not be discussed in great detail here. What you basically need to know is that music always sounds best when recorded at the highest possible rate (both bit resolution and sampling rate) and in stereo. This usually means a larger resulting sound file and the need for compression and streaming.

Sound cards, which have been standard on almost every Macintosh computers, are available for most PCs. Due to the ubiquity of the Creative Labs SoundBlaster it's a standard consider either getting the actual brand (www.creativelabs.com) or a sound card that says it is Sound Blaster compatible. The more you spend on a sound card they range from \$25 to over \$1000 the more features and fidelity. For most speech applications, you can get by with a relatively inexpensive SoundBlaster compatible card. For music, you should consider spending a bit more.

The last part of the recording package is the software. While there are plenty of shareware products for both Macintosh and Windows, it's usually worth the money to invest in a commercial product such as SoundForge for Windows (www.sonicfoundry.com) which retails for \$495. If this exceeds your budget you can opt for SoundForge XP 4.0 which is a stripped down model that retails for \$49.95 but still offers most of the features needed to create and edit audio for the Web. If you are using a Macintosh, Macromedia's SoundEdit 16 (www.macromedia.com) is a good choice and retails for \$419. Both packages are easy to install and use and have the ability to save files in many formats as well as alter sampling rate and resolution.

As mentioned earlier, two common sound formats are the Windows WAV file and the Macintosh AIF files. Both these formats are available on each platform so the fact that I mentioned one corresponding to the other isn't as crucial as it was a few years ago. If you want a quick and dirty way to put up a small sound file on your site, I would recommend using a WAV file since the majority of the computers on the web should be able to play them. Although not a true compression format, you can reduce the size of a WAV file by reducing the sampling rate, converting from stereo to mono, or reducing the resolution from 16 down to 8, or all of the above. Keep in mind, however, that every reduction in the above parameters will result in a lower quality sound.

When the Web was young, the only way to deliver audio in an acceptable time was to follow this compression by reduction scheme. Not long after, however, groups began to create highly complex algorithms that allowed for reduction in the size of the file while retaining the highest quality possible. An early effort was by the Moving Picture Experts Group or MPEG (www.mpeg.org). Intended to develop a compression scheme for video, their efforts in audio compression combine the best sound quality to file size ratio. Although it has been slow to catch on, it has been in the news a lot lately and is worth considering. The MP3 format is great for musicians to deliver near-CD quality audio files on the World Wide Web,

but may be too complicated for adding simple sound files to the web. For more information, check out the MP3 web site at www.mp3.com.

Then came Real Networks. In 1995 they simultaneously creating an acceptable compression scheme with (here's the real kicker) streaming technology. For the first time, hour-long sound files could be played on a computer with a 28.8 modem. This was accomplished by allowing a Real Audio player to download only the appropriate amount it needed to continue to play a sound file. The higher the bandwidth available to you, the better quality the sound you heard. If you were on a high-speed network LAN with high-speed Internet connectivity, you could hear FM radio quality sound files either live or on demand (a previously recorded sound file). This opened a large door in that people with 28.8 modems could listen to live news and sports information with decent quality sound without having to wait for files to download.

An overview of the steps involved in embracing streaming media is provided in Chapter 8. Note that you can take advantage of Real Media's compression schemes without getting involved with the server side. If you download the Real Media encoder from their site (www.real.com) you can compress files in the Real Audio format and place them on your Web page. The user then only has to wait for the relatively small file to download and use the Real Audio player to listen to the file. Table 2 shows typical file sizes for Real Audio files in various formats. It is also possible to do what's called http streaming which will stream audio files using the standard Web protocol of HTTP.

Table 2. Approximate file sizes of a one-minute Real Audio sound file.

Sound Quality (Sampling Rate)	Approximate file size	Approximate time to download on 28.8 Kbps modem
Original CD Quality (44.1 kHz)	10 Megabytes	45 minutes
Broadcast Quality (32 kHz)	600 Kilobytes	2.7 minutes
AM Radio Stereo (8 kHz)	150 Kilobytes	42 seconds
AM Radio Mono (8 kHz) suitable for voice	64 Kilobytes	18 seconds

It now has become customary to speak of compression rates in Kilobits per second, meaning how much information can be downloaded. In other words, the compression scheme should match the intended bandwidth of your audience. While this is necessary when planning for using streaming audio files, you only need to know that it, just as before, affects the size of the file and the overall fidelity.

Basic Recording Techniques

Now that we have discussed elements of digital audio, we turn our attention to good recording techniques so that you are able to start with the best quality sound. Most of the techniques that were developed in the analog world (i.e. tape recorders) are still applicable. The most important aspect of recording that affects the quality of the original sound file is the recording level. Too high a level will result in distortion, while too low a level results in high background noise levels.

The first decision to make is whether you want to do your master recording in an analog format such as cassette or conventional reel-to-reel tape, and then convert to digital, or record digitally to begin with. Popular digital formats include Digital Audio Tape (DAT) or MiniDisc; another digital alternative is to record digitally directly into the computer.

All are perfectly acceptable methods. Using a laptop or desktop computer for direct digital recording is perhaps the least common approach, but it is perfectly viable. You may find it more convenient to stick with conventional recording, whether analog cassette, DAT, or MiniDisc, for several reasons:

- You may not want to use your expensive laptop computer for location taping.
- You may want the comfort of having a backup master tape to revert to should your computer's hard disk fail.
- Your computer itself may make noise that is hard to eliminate from the recording process.

The following two methods are for recording a person speaking, the first using some remote recording unit, the second recording directly to a computer.

Method 1

If you have access to a digital recording format such as a DAT recorder or a MiniDisc, that would be your first choice. If not, a good quality cassette deck such as the Sony Professional Walkman will suffice. It's important to realize that a recording is only as good as its weakest link. Make sure all the equipment is in good working order. Using the aforementioned Radio Shack PZM microphone with a decent cassette deck will result in an adequate sound quality suitable for most speech applications.

Before starting recording, have the speaker say a few words into the microphone so that you can set a level. Ideally, the microphone should be placed no more than a foot or two from the sound source. If you've ever played around with any type of tape recorder, you'll probably remember the effects of too much sound going onto a tape. In the analog domain, sound impulses are recorded onto a magnetic tape. The more the volume of the sound, the more information appears on the tape. At some point, the waveform of the sound will go beyond the capacity of the tape, causing a phenomenon known as "clipping." The sound of

"clipping" is what we commonly refer to as distortion. Distortion is basically the phenomenon of an inaccurately recorded waveform. In this case, the inaccuracy appears since the tape can not capture all of the information.

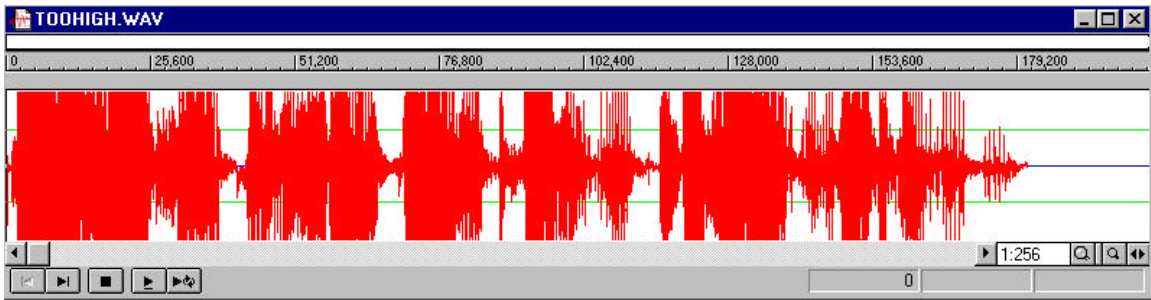
Conversely, if the sound level is too low, the phenomenon of tape hiss occurs. This is caused by the fact that magnetic tape creates a noise when pulled across the tape head. If the signal on the tape is ideal, the tape hiss isn't noticeable. If the signal is too low, however, we still hear the sound of the blank tape. Since the end user also has to turn up the volume on the other end, this tape hiss becomes more pronounced. Although the method of digital recording is different in that it doesn't use analog waveforms but rather bits to represent the waveform, the effects of too high or too low a signal are still present.

If your cassette deck has a VU meter—a digital or analog gauge that reacts to someone speaking into the microphone—you'll be able to monitor the recording levels without too much effort.

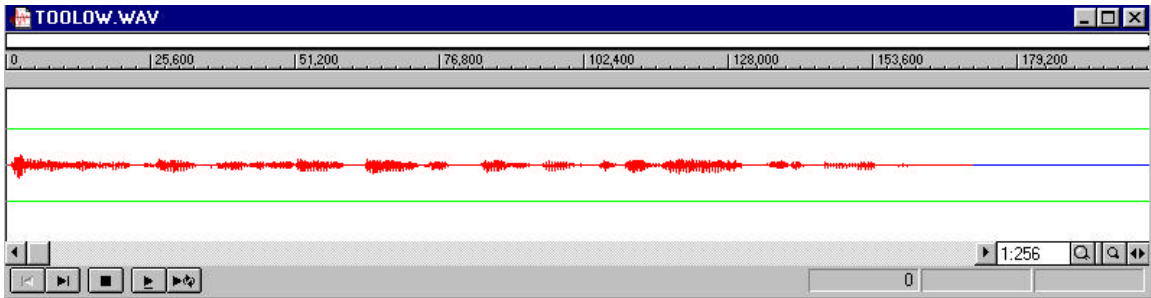
As you have the speaker say some words, look at the meters. If they seem to be in the red a lot of the time, you need to cut back on the level. On the other hand, if the levels never get close to the red, you'll need to bring up the level. As a rule of thumb, it's okay for the level to go into the red once in a while. If you can set the volume so that this is the case, the resulting recording should be fine. If your cassette deck doesn't have meters, it probably will automatically set the levels, which should work fine. Figure 2 displays windows from SoundForge XP showing the waveforms from various recording signals.

Figure 2. Recording windows of three sound files.

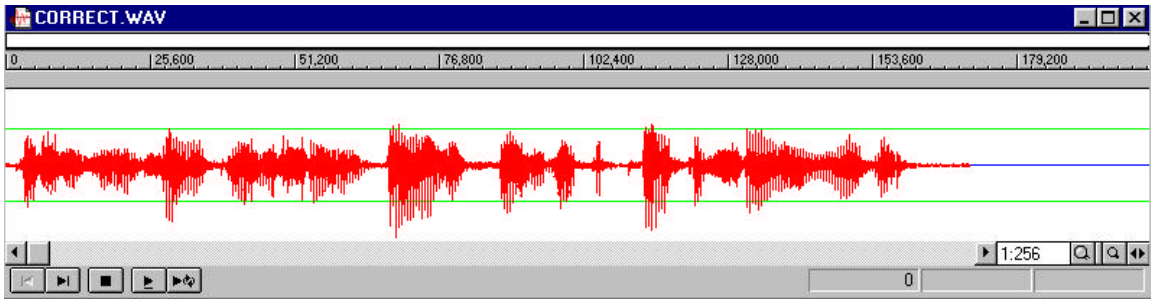
A. Too much signal or clipping.



B. Too low of a level which creates noise



C. Proper sound level



Method 2

This method assumes you want to record directly into a desktop or laptop computer. Most of the techniques explained above are applicable to recording directly into a computer. Keep in mind that computers tend to be noisy contraptions, especially the hard drive. If you want to record this way, make sure you have a microphone cable long enough to put some distance between the microphone and the computer. A common mistake is to use the built in microphone in a laptop; built-in mikes tend to be of quite low quality, and have a tendency to pick up the whirring of the computer's hard drive.

If you choose to use a computer for direct digital recording, be sure to get a microphone that is matched to the input level your sound card expects on its microphone jack. Unfortunately, the mini-plugs used for microphones to be used with devices such as portable cassette deck fit are the same size as the mini-plugs used for microphones intended for computer applications. *They are not compatible.*

After the microphone has been plugged into the "microphone" port on your sound card, you'll need to go into the control panels to set the recording source. In Windows machines, this is found in the Control Panel in the "multimedia" section. It is here where you get to set the recording quality by way of the sampling rate. As mentioned before, it is always advisable to start with the best quality sound file and then edit it and compress it. That way, if you make a mistake you can always go back to the original sound file and start over. It's a good idea to always set the recording level to "CD quality," which means 44.1 kHz sampling rate and 16 bit resolution. If you're recording speech only, however, you could turn it down to 22.05 kHz to save hard disk space by fifty percent. Keep in mind that the file sizes in Table 1 apply when you're recording. Both SoundForge and SoundEdit 16 display the available hard disk space.

One misconception is that digital recording frees you from the burden of worrying about recording levels. Although digital formats may have more leeway or headroom than some analog formats, setting an appropriate level is still important. Thus, you still need to keep an eye on the recording level, whether you're recording on an external digital deck or directly into your computer in digital format. In SoundForge there is a way to monitor the input from the microphone before you actually record. Again, make sure the meters don't register in the red too often or you'll get distortion.

Whatever sound recording package you use, you'll have to get used to what settings work well. At the beginning of sessions, you should make it a practice of recording about 10 seconds of audio and then listening to the result and observing the sound file window. This way you can be sure that the audio levels you've set will work. See Figure 2 for a visual representation of proper signal level.

In both methods, you'll want to do some clean up after you have recorded. This usually means taking out the dead air at the beginning and end of the sound file (and this also saves file size). It's also a good idea to fade in at the beginning and fade out at the end to make insure a smooth start and finish. It's also wise to implement one other adjustment, called normalization, at this point. Normalization uses a mathematical functional that finds the

highest amplitude and shifts the entire sound file to the highest level without clipping or distortion. This is useful if you're creating several sound files and you want them all to be the same relative volume.

When you have the sound file the way you want you need to save it in the highest possible format. Since we've set our computer to record in the highest possible quality (44.1 KHz, 16bit), we should save it in that format. In both SoundEdit and SoundForge you'll have the option of saving the document in various formats. Since we want to save the original file in the highest settings, save it as .WAV on a Windows machine and .AIF on a Macintosh.

Including the File in a Web Page

Now we have some decisions to make. We have saved our original content in a high quality and therefore large sized format. We now need to determine how to compress or reduce the sound file so that a user can easily download it. If the content is only speech, a simple solution is to downsample the file so that it is reduced in size. If you refer back to the Table 1, you'll see that you can greatly reduce the size of the sound file by changing the sampling rate to 8 kHz which should work for speech. If you're adding music, you'll want to use a higher sampling rate and possibly 16 bit resolution for clarity. Try experimenting with the rates and listen to the outcome to determine the best combination of file size and fidelity. After you have changed these properties, you can save the file in this reduced rate and resolution. On either a Macintosh or a PC, you should save the file as a WAV file and add the .wav extension to the file name.

If you want to make the file sizes even smaller, you should convert your files to the RealMedia format. This can be easily accomplished using the software from Real Networks (www.real.com) which you can download from their site. Taking a look at Table 2 you will see the significant decrease in file size. Keep in mind again that the lower the bit rate (the smaller the sound file size), the lower the fidelity. Again, experiment with various bit rates to determine the best combination of file size and fidelity. The converted file should have the .rm extension denoting RealMedia (earlier versions of the Real system used .ra file extensions, denoting RealAudio).

Now comes the easy part. If you're already familiar with HTML, this will be consistent with your knowledge. To add a sound file, use the A HREF tag just like you would with any link. Therefore, if you want to make a link that will point to the sound file and subsequently allow the file to be heard, use the following syntax:

```
<a href="soundfile.wav">Listen to our Mayor speak</a>
```

or

```
<a href="soundfile.ra">Listen to our Mayor speak</a>
```

and make sure the sound file you created is in the same directory as your HTML document. The above examples will create the clickable text Listen to our Mayor speak. If the browser is configured properly (most newer versions of Netscape and Internet

Explorer are), your audience should be able to listen to the words of your leader. That's all there is to it!

After experimentation, you will be able to quickly create high quality audio files that will enrich your web site and deliver your message in a format that augments the text and images on the screen. If you decide that you want to eventually include lots and lots of these sound files, you might want to consider installing a streaming audio server, which will be discussed in Chapter 8.