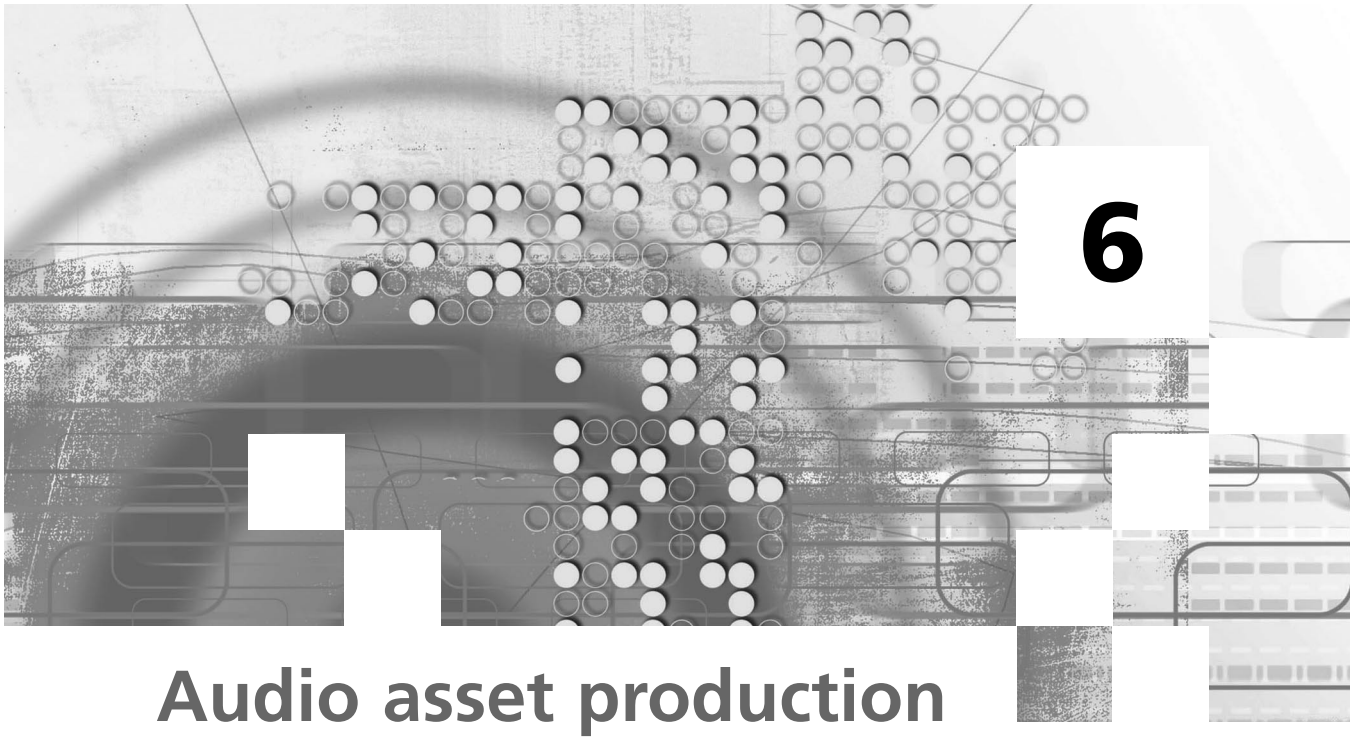


6

Audio asset production

Project manager's responsibilities

- To book the studio and the artiste
- To make sure scripts are ready
- To make sure that the recording session is successful
- To ensure that the material is prepared to the correct specification and encoded with an appropriate codec
- To understand the processes involved in producing this kind of asset

Managing audio

During the course of making an offline multimedia production there will be, by definition, assets to be created or manipulated that make use of time-based media such as sound and video. Web pages are also incorporating such things as the bandwidth of Internet connections steadily goes up. Unlike text and graphics, you may make use of specialist facilities with personnel that you hire to work on your assets, but even if there is a studio in your office which you can use for the recording, the basic principles are the same. Video will be discussed in the next chapter; this chapter describes the processes involved in dealing with audio from a practical point of view.

The basic idea behind this chapter is to provide you, as a producer or project manager, with enough background on the technical processes to enable you to hold your own in discussions with experts. It will also make your use of specialized facilities more interesting and rewarding. Of course, depending on your background, you might already be an expert in one of these fields. It may also be the case that, in a small development company, you will have the opportunity to 'be' the expert and carry out some of the audiovisual manipulation yourself. So treat our use of the word 'you' lightly, since if you hire an external facility to record, mix or edit, it will be an engineer that actually carries out these tasks.

The aim in writing this chapter will be achieved if, next time you go into a sound suite and the engineer asks if you want the sound limited, or what sample rate you want, you can tell him or her, or even discuss it, with confidence. If you hire a facility and an engineer, make use of their knowledge and do not be self-conscious about asking advice.

One final point: working with sound on a video is much like working with sound by itself. The main extra difficulty is in keeping the sound in time with the video: in synch(ronism) with it.

Before the session

It is most likely that your first use of a sound studio will be to record a narration voice-over for your project. This might be for the soundtrack of a movie or to accompany an animation or group of stills. Music and drama are other possible kinds of material you might record, but this chapter will concentrate on recording a single voice. The basics are the same, but music and drama have the bells, the whistles and the fairy dust.

When you decide to record a voice you need three things: the script, the voice, and the studio. You will also have thought about how you want the voice to sound, and this will have influenced your choice of voice-over talent.

Incidentally, do you have to use a studio for this, or could you record in an ordinary office or at home? If you or a member of your team have a background in sound recording and have the equipment then of course you have

that option. But it can become very frustrating when you begin to notice all the extraneous sounds, such as aircraft and motor cars and sometimes even birdsong, that are so much part of our background sounds that we often forget they are there. Being in a studio can actually be more relaxing since the voice-over artiste can concentrate on the performance rather than worry about yet another passing motorcycle.

When choosing a voice-over artiste you might have decided to use a famous actor or actress, and would like to include their name and photograph on the cover of the CD or on the home page of the website. Alternatively your voice-over artiste might be a person who specializes in being a voice. In some circumstances you might do it yourself, or use a friend or someone in your company who has experience, perhaps as a radio presenter. For the first stages of a project it is not unusual to make a guide voice track yourself, which will be replaced with a professional one later. You might even record this guide track at home or in the office.

Unless you know someone already, your route to your voice will be through an agency. There are many who specialize in providing voices, usually for radio advertising, and they will have both famous actors and professional voices on their books. Many of them now have websites or produce CDs where you can listen to their clients and make your choice.

The voice-over artiste will like to have a clean script, probably double-spaced so that changes can be made clearly. The artiste will often mark in the emphasis to be used when reading. You should send the script to them a few days before the recording if you can. The script should be printed out so as to avoid paragraphs going over a page boundary. The paper should be stiff so that it does not rustle. You should check pronunciations of any unusual words, especially proper names, and if you are producing the session you should be sure about every word in the script. Be prepared to make changes to make the script easier to read. Often the voice-over artiste will make very useful suggestions about this. Besides the possible direct benefit, you will be helping to create a good working atmosphere, and that will help the artiste to perform better.

With any luck you will find that your voice-over talent can read virtually anything you put before him or her. Many of these people spend the whole day reading one kind of script or another and can cope with most things. Interactive media may be so different from their usual work of advertising and corporate videos that telling them a little about the project will pay dividends or alternatively they may have worked on more websites than you have.

You can work out the timings for the speech yourself before you go into the studio. All you need to do is read it at about the right pace, and time yourself. A rough guide is about 200 words a minute.

You can find your studio either by asking a professional studio body (in the UK this is the Association of Professional Recording Studios) or from a yearbook or even the *Yellow Pages*. Word-of-mouth recommendation or a studio you saw credited on another product is also a useful guide. When you

book the studio, as with any outside facility, you will need to agree the rate, how overruns (needing more time on the day) are charged, what happens if you underrun, and the arrangements for paying. You need to tell them the format that you want to take the recording away on (for example do you want to take away a DAT tape or a WAV file on a CD), and whether you will edit the recording yourself (if you have the technology to do so) or will ask the facility to do it.

The background

The processes for recording sound date back over a century. Sound is the result of fluctuations in air pressure, which cause our eardrums to vibrate. If the frequency is right these vibrations get passed to the brain and are heard as sounds. In the earliest kinds of recorders you spoke, or rather shouted, into a horn, and the power of your voice caused a diaphragm to vibrate. A stylus was connected to the diaphragm, and this distorted a metal or wax surface over which it moved. To play the sound back you reversed the process and listened to sound coming out of the horn.

The microphone (usually just called a mic or mike and always pronounced 'mike') and loudspeaker (usually just called speakers or monitors) are still the mechanical components of sound reproduction. They work by detecting or creating the movement of air.

In the studio

You will find that a recording studio will almost always be in two parts: the control room and the studio itself. The studio may be called the booth if it is small and used only for recording voices. There will usually be a glass window between the two rooms.

Recording studios are strange places. You might find that no two surfaces are parallel because this stops sounds bouncing between the walls and setting up resonances and standing waves (where the room acts like a big organ pipe). Legend has it that some enthusiastic builders once thought that the plans for a sound studio were wrong because the corners were not right angles. So they kindly corrected the error.

The windows, while not being parallel either, will have double or even triple glazing, and the walls, doors and even the furniture will look as if they are either carpeted or designed by someone who likes to hang carpets and boxes on the wall. This is to reduce the reflections (for which read echo or reverberation) in the room. The difference between echo and reverberation (reverb) is simply in the time between the echoes. Reverb sounds smooth and continuous because the echoes are too close together for us to distin-

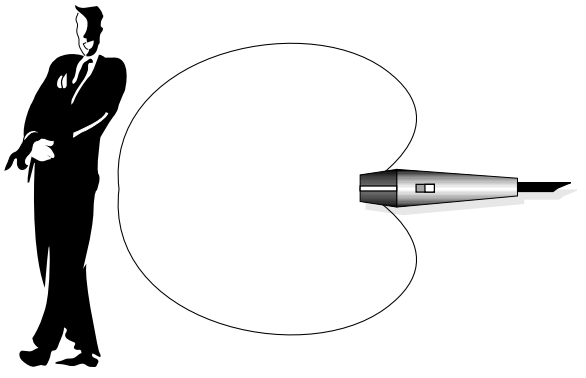
guish them. Unless you want to remake 'Heartbreak Hotel' you are unlikely to use echo as such. If there is no reverberation around a sound then we say the recording (or the room) is 'dead' as opposed to 'live'. In fact a room with absolutely no reverberation (called an anechoic chamber) is very uncomfortable to talk in, because we need something of the sound of our own voice to help us speak and some room reverberation around a recording to make it seem natural. With music, especially popular music, this reverb is normally added afterwards using reverb units (the modern electronic version of the echo chamber – which was simply a room with hard walls like a bathroom). Sometime a short delay is added before the reverberation starts, to mimic the sound getting to the walls, and this can help make reverb sound more natural and less obtrusive. This reverb should not favour some frequencies of sound over others. Reverb that does is called coloured and sounds unnatural. The natural small amount of reverberation in a room, together with any other background sounds, is sometimes called ambience or ambient noise.

To record a voice the microphone is usually placed about 18 inches to 2 feet (45–60 cm) in front of the speaker's mouth in a reasonably dead room. If the mic is too close it will pick up lip smacks and other bodily noises. If the mic is too far away the sound will be too 'live', which means there will be too much reverberation.

When you get close to some kinds of mic you suffer from a phenomenon called bass tip-up or proximity effect. This is, as the name suggests, an increase in the bass sounds in the voice, and it is caused by cancellation of high frequencies when the source of the sound (your mouth) is too close to the diaphragm of the microphone.

In general, the positioning of the mic in front of the speaker is crucial in getting a good sound, and an experienced engineer will know where to move the mic to avoid popping, breath noises and sibilance, which is an unnatural whistle in any S sounds. Popping, as the name suggests, is the effect caused by blasts of air from the mouth hitting the microphone. In fact it is sometimes also called blasting. This is at its worst with the letter P, and a good test is the old tongue twister 'Peter Piper'. If the mic is very close, just breathing out may cause a noise. Most studios will put a wind shield in front of mics to stop this. To a large extent these problems can be reduced by having the mic slightly to one side rather than straight in front, and this is called being off-axis. Sibilance is more difficult to control.

The mic will probably be on a stand with a gallows arm or boom suspending it over the table – assuming your speaker is sitting at a table. Some people will sound better, and project more, if they stand up. If the speaker is using a table then be careful about where he or she puts the script. It will probably be under the mic and so you could hear the paper rustling. Less obviously, the relatively hard surface of the paper will affect the acoustic around the voice. The movement of paper, and the movement of the speaker's head as he or she reads, can change the high-frequency component of the voice.



A cardioid microphone.

You can use most kinds of microphone for a voice recording, but the best kind is what is called a large-diaphragm condenser. Although these mics are very expensive they produce a smooth sound that is very easy on the ear.

There are basically three types of microphone, and their names come from the shape of their sensitivity, or polar response curves, as the following diagrams show. The further the curve is from the mic, the more sensitive the mic is in that direction. The reasons for the names will soon be apparent. In most cases you can assume that this curve is the same in three dimensions: a shape turned on the axis of the microphone. The exception is the bi-directional mic.

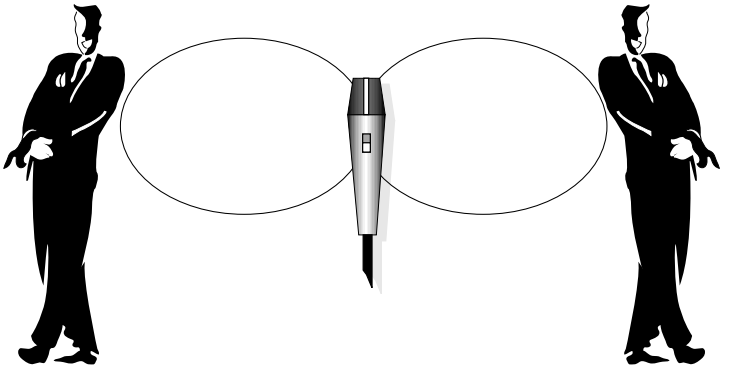
A cardioid mic will pick up more sound from in front than anywhere else. It is called cardioid because the polar response curve is shaped like a heart.

A bi-directional mic is sensitive on two sides, and this is also known as a figure of eight.

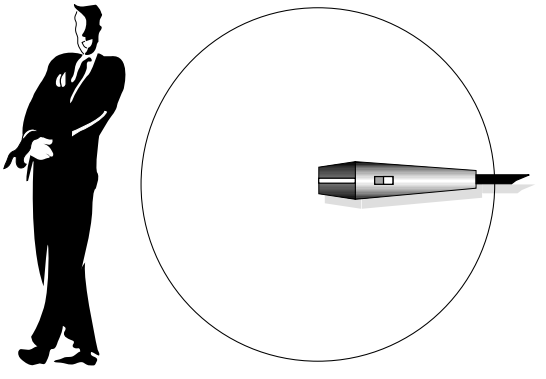
Finally there is the omnidirectional mic, which is equally sensitive all around. You should be able to speak closer to an omni mic than you would with a cardioid without bass tip-up and risk of popping. Some microphones can be switched between all these response types. To record a voice any of these microphone types can be used. A cardioid will have less pick-up from the room but may suffer from popping.

During the recording you, as producer, will be in the control room. The speaker may, or may not, be listening to the sound of his or her own voice in the headphones. Different people will want to hear themselves at different volumes (also called levels), and this can be critical to their ability to read well if they are inexperienced. Giving the speaker a volume control for their headphones is a good idea.

Most people's speaking voices will have quite a wide range between the quietest and loudest sounds. This is called the dynamic range. The engineer can control this dynamic range manually, by adjusting the volume control



A figure of eight microphone.



An omnidirectional microphone.

fader in the control room, or by using electronics to automatically compress or limit the signal.

Although the same word is used, audio compression is totally different from and unconnected with data compression. To avoid confusion, the term bit rate reduction is often used instead of data compression, especially in broadcasting. Compressors and limiters reduce the dynamic range of a sound. They are a sophisticated version of an automatic volume control in a tape recorder. A limiter is set to tightly control the volume of sounds that exceed a certain value, whereas a compressor operates over a wider range of levels but more gently. It is difficult to describe the effect of using these devices, but you will find some examples on the website.

You would choose to limit or compress a voice, and therefore reduce its dynamic range, for a number of reasons. You are most likely to choose to make the master recording with compression or limiting that is inaudible to

the listener, and most often limiting is used to catch and reduce a few bursts of the loudest moments in the speech. In this way you can bring up the level of the whole speech (that is, make it all sound louder) without the few high points causing problems by overloading the electronics. With a digital signal, for which overloading causes more distortion than with analogue, limiting is useful as a back-stop to prevent accidental overload, especially if you can do only one take.

Both devices, but especially compression, can be used for effect. Because our ear's response to sound is not linear we tend to hear loud sounds with less dynamic range than quiet ones. You can fool the ear into thinking something is louder than it really is by compressing the sound. This is also useful if you want to put speech over some music and it is important that the speech is heard all the time without sounding shouted. However, remember to keep (and archive) a 'clean' copy of your master recording before you start to process it for delivery on your website or CD. This is an obvious thing to do if you are using tape, but if you are working entirely digitally in a computer it needs remembering.

If your sound is finally going to be played on a system with substantial bit rate reduction, as you might on a website, then processing it to reduce the dynamic range of the recording will make the playback sound better because there will be fewer quiet parts to disappear into the noise. For similar reasons, anything that will be played in a noisy environment, such as a point-of-information kiosk at a trade show, should be compressed to make it easier to hear.

If you are in doubt about using limiters or compressors, a basic recommendation is to use a little limiting to catch the loudest peaks. If the recording is quiet, with no background hiss or reverb, you can always compress it afterwards. However, you should remember that compression tends to exaggerate reverberation, and you have to be wary of this. If there is a high amount of background noise then compression and limiting will noticeably affect this by making it seem to get louder and softer depending on the foreground sound, leading to an effect called pumping. An example of this is also on the website. If the response time of the compressor/limiter is too slow, you will hear it pushing and pulling the sound as the amplification goes up and down. If it is too fast, it will distort the waveform of the sound.

Volume, otherwise known as level, has been mentioned already, but how do you measure it? There are two kinds of meters in use: VUs and PPMs. VU (pronounced vee-you) stands for volume unit, and PPM stands for peak programme meter. You will find them with a variety of displays such as dials, bars and sets of LED lights. Actually these kinds of meters are supposed to have a standardized response so you could say that there is third kind of meter which just gives you a 'general idea of levels' rather than truly following VU or PPM measuring.

VUs are the most common although they do not really tell you anything exact about the signal. However, an experienced engineer can judge the level of something very well with a VU, and it arguably gives a good rep-

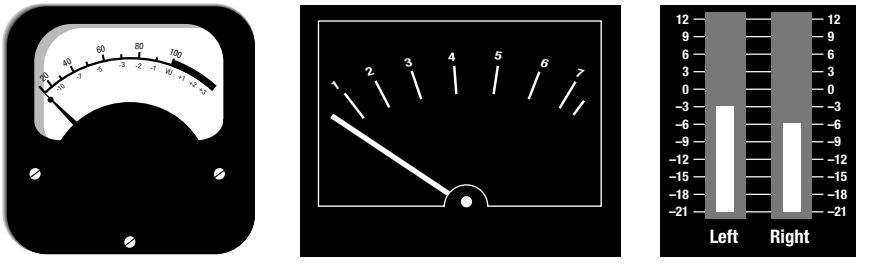
resentation of the loudness of the sound. A VU meter display will move around very quickly. A VU has a red band at the top (especially if it is a moving-pointer type of meter) and you will quite often see it running into the red. This is not necessarily an indication of overload, especially in an analogue system, even though it is supposed to be.

A PPM is a more exact kind of meter, and was developed by the BBC to control levels being fed into transmitters, although it is also used in recording. What a PPM does tell you is the actual peak signal going onto the tape or into the transmitter. PPMs are designed to have a fast rise time and a slower release. This helps you to read them. BBC PPMs also have an integration time (in other words they are not measuring the instantaneous level but an average over a small fraction of a second), which means that they will not detect very short peaks. This could be a problem for digital systems, but in practice, even though a short peak may be distorted if you look at the waveform, it will often sound fine because the ear may not hear a very short burst of distortion.

It is more usual now to find VU meters and PPMs that use lines of lights to show levels (sometimes called a bar-graph), but some would argue that a meter with a needle is easier to read. The meter may not be a real one, but could be an image of one on a computer or television screen. A particularly useful type of PPM is the dual stereo PPM, which has two meters side by side, each with two needles. One shows the left and right channels on two coincident needles while the other shows the sum and difference signals for stereo.

Measuring of levels for surround sound is more complex and while individual meters can be allocated to individual channels you can also produce a screen display that shows the levels around the sound stage with zero at the centre of a circle and full volume moving the display out to the circumference of a circle. This is called a 'Jelly Fish' display.

Besides level meters and a compressor/limiter, there will be equalization (usually just called EQ) in the channel of the desk through which the voice passes. This is a glorified bass and treble control, and will help the engineer



VU meter PPM meter Bar-graph

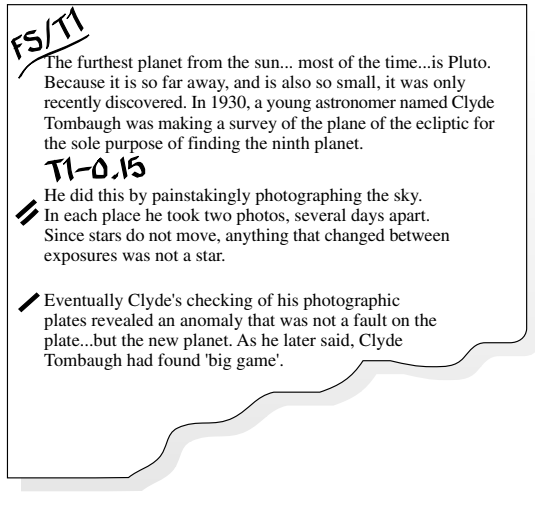
VU, PPM and bar-graph level meters.

to get a suitable sound out of the voice. Because of the ear's sensitivity you can often make a voice sound closer – have more 'presence' – by slightly boosting frequencies around 2 kHz.

Working with your voice-over artiste

Your recording might be used against a video, a sequence of still images, or in isolation. If you are recording a voice track that goes with a sequence of stills or even over a single image then there will be no problem in recording the voice without any reference to the sequence. This is sometimes called recording wild. In this case you can work your way through the script, one discrete section at a time, rehearsing and then recording. If the artiste makes a mistake (fluffs), all you need do is ask him or her to go back a sentence and read it again. This is best done as you go along rather than at the end. Anything that will be heard in a continuous sequence should be recorded as nearly as possible in that sequence so as to avoid subtle changes in tone or speed, which would be very apparent over an edit. You can mark up your copy of the script with timings for paragraphs, and you should also note where any fluffs occur and how many takes it took to get things right.

You can mark takes by using a diagonal line like this /, which you mark in at the point you will probably use for an edit, or the beginning of the sentence. If there are two takes you can put in two lines //, and you may also need to note timings by the lines. You can get timings from a stopwatch, or better still from the timer on the tape machine or digital recording system.



Marking a script.

Mini-disks, DAT machines and hard-disk recorders have a built-in time code that is recorded on the tape or disk (optionally with professional machines), and this is very useful for finding takes. A take that gets through only a few words and then falters is known as a false start (FS) and is not usually marked as a take. To assist editing, you should always record a few words in advance of where you know you will want to edit, to allow the speaker to get up to speed and give you a few choices of where to edit.

There is another way of dealing with takes, and that is to roll the tape back, play the preceding sentence and switch into record as the artiste speaks the lines again. This method avoids editing afterwards.

You might be recording a narration that has to be timed to a moving picture. In these cases you could bring in the computer and run the movie or animation or whatever, but an alternative is to record the movie onto a videocassette. In this way you can make use of a facility that has recording and mix-to-picture capability for TV or films so that the narrator can watch the movie as the recording is made.

Other things to remember about the session: check the spelling of the artiste's names for the credits and make sure that you have agreed the appropriate rights. If you have to go back to the artiste later to sort out rights you are at a disadvantage, and you cannot assume that because they came to the recording they have granted you the use of the material you require.

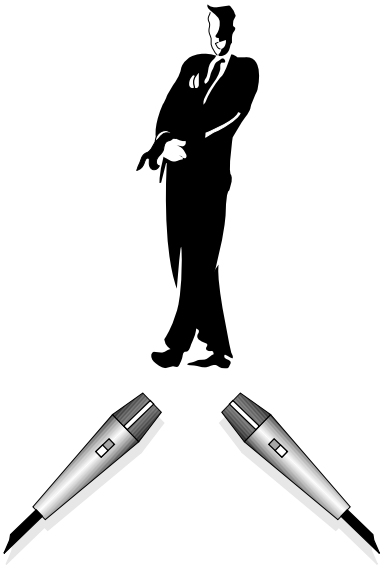
What can I ask the studio to do for me?

A recording studio will be able to do far more than just record your voice-over. If you want them to do so they can edit the takes together under your direction (and sometimes without) to produce a finished master. They can take your voice, script and music tracks and mix them together to produce a finished track for your application. Some facilities can even digitally compress the track into RealAudio or MPEG audio for you, but remember not to confuse audio dynamic range compression with data bit rate reduction. It is your choice as to how much you ask the facility to do, and how much you do yourself. This will depend on factors such as the tools you have available to you, your ability and your budget, because the facility will need to be paid, and this is an above-the-line cost.

Mono, stereo and surround sound

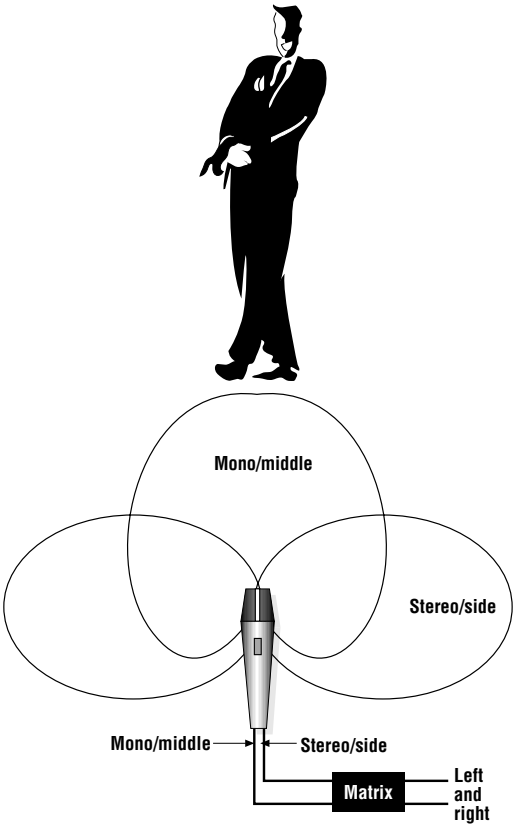
A lot of sound in new media is mono, which means single channel. This is because mono sound, by definition, takes up half the space of stereo sound. If you have the capability, stereo will be useful to you because, rather like moving images, it helps you build up the effect you want to convey in your application. If you decide to make your application with stereo sound you will need to know whether it might also be used in mono: if in doubt assume

A spaced pair.



A crossed pair.

it will. This is because a little care needs to be taken to make sure that stereo recordings are mono compatible and still sound correct when the left and right channels are added together to make mono. Similarly, if you want to avoid doing several mixes, you should ensure that surround sound recordings are stereo compatible (if not mono).



The M and S method.

Stereophonic sound means two speakers. We allocate a position to a sound by a combination of time delay – the sound reaching one ear before the other – and level. Using level to give positioning is by far the most common way, and in mixing (which we will come on to later) the balance engineer will position sounds between the speakers by changing the amount of the sound fed to the left and right channels. This is known as panning, and the control on a mixing desk that does it is called a pan pot (for panning potentiometer). Most of the time this simple panning of sounds will work perfectly. Incidentally, we get our directional information at high frequencies rather than at low ones.

When recording a real sound in stereo things are a little different. There are three main ways of recording in stereo. They are called the spaced pair, the crossed pair and the M and S pair. The pair refers to a pair of microphones.

A spaced pair will be two omnidirectional microphones positioned several feet apart. For distant sounds, such as a crowd at a sports event, this

will be fine, but the sound will tend to have a hole in the middle since, confusingly, a sound close to the microphones but between them is probably not close enough to either. (This kind of set-up sounds particularly good if you listen using headphones, but strictly speaking this then makes it a binaural sound system, not a stereo one.)

A crossed pair is also known as a Blumlein pair, after the English engineer who invented the technique. You take two cardioid directional microphones and place them as close together as you can but pointing 90° apart. This gives a good stereo image and no hole in the middle. For the best results the microphones need to have identical frequency responses.

You can do a simple calculation to show that M (which is mono, left + right) and S (stereo, left – right) can be combined, or matrixed, back to left and right. Left is M plus S and right is M minus S: you can ignore the factor of two here. This method combines a cardioid microphone pointing forwards and a figure of eight microphone pointing sideways. This system has several advantages: mono compatibility is assured; good stereo can be obtained at a distance using a rifle mic (a very directional hyper-cardioid mic) for the M channel; and the stereo image is very good, especially at the centre. The microphones do not have to have an exactly matched response. You can in fact buy single microphones that use this technique and yet give left and right outputs as in the diagram.

One additional factor to consider is that, strictly speaking, a sound in the centre will sound twice as loud as one at the extreme left or right when the channels are combined into mono. This should be a factor in the mixing, and most panning controls compensate for this. If you want to position a voice off to one side for effect, or to have two voices discussing something, then the best position in the stereo sound field is half-way between the centre and a loudspeaker. This will also give you a good mono signal.

Although quadraphonic sound on record, with four channels, came and went in the mid-1970s, multichannel sound has made a comeback thanks to surround sound movies and DVDs. This is because surround sound systems are being used on the soundtracks of movies to enhance the cinematic experience. More than four speakers are used. Originally, a speaker at the centre-front position was added to improve the stereo positioning for a cinema audience, since very few of them sit in the optimal place for good stereo sound when only two speakers are used, i.e. the middle. Basically a cinema audience is too big and spread out for conventional stereo. Having started with three speakers at the front and adding two at the back left and right to give a surround sound field, you can add what is called a sub-woofer, which gives out only very low-frequency sounds, and possibly two extra front speakers at the mid-left and right positions. This gives either a 5 + 1 or 7 + 1 surround sound field (usually referred to as 5.1 and 7.1) with the first number referring to the number of main speakers and the +1 referring to the sub-woofer, as used in the cinema and DVD. It is also possible to encode a version of surround sound into two channels, and this legacy of

quadraphonic sound is still found in surround sound on TV and video and is used for the Dolby Surround format.

Although in theory mixing for one 5.1 sound field should be like mixing for another, it always makes sense if you can to at least check the mix using the delivery format. Since this could be Dolby Digital, Dolby Surround, TDS, MPEG multi-channel or plain old stereo or even mono, this is no trivial task. There is also a certification system for surround sound playback devised by Lucasfilm and called THX and since many movie theatres and home cinema systems are set up for this you might want to take it into account as well.

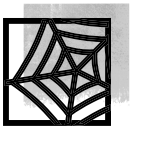
Recording surround sound is best done using a double-crossed pair with four cardioid mics facing in four directions 90° apart: double stereo. Otherwise a pan pot can be made that moves a sound source around a surround sound field, and some of the digital audio editing and mixing systems allow you to do this. It is possible to record two-channel sound for listening on headphones which gives a very realistic impression of a surround sound field. This is known as binaural recording and in principle, is recorded using a pair of omnidirectional microphones placed either side of a solid object like a cushion or even a plastic disk. This emulates your ears on either side of you head. In fact one famous binaural microphone set-up consisted of a realistic-sized dummy head with microphones in the ear channels. Audio recorded this way relies a lot on the minute phase relationships between the left and right channels and is not really mono compatible. It also sounds flat if you listen using loudspeakers instead of headphones but it is possible to process binaural sound using a complex arrangement of mixing, filtering and time delays to produce a very realistic sound field on speakers as well.

Tricks with sounds

There is a wide range of what are called psycho-acoustical effects which affect the way we hear sound, but it might be useful to describe a few potentially useful, or awkward, ones.

The ear's sensitivity to high and low frequencies diminishes at low volume levels. The 'loudness' button on your hi-fi amplifier takes account of this and lifts treble and bass to make it more pleasant to listen at low levels. Another effect of this phenomenon is that if you take a recording and then attenuate (drop the volume of) it, it will seem to lose top, or high, frequencies.

If you want to give a cheap imitation of a sound coming from behind the listener when you are working in stereo then it can be done by making the sound to the two loudspeakers out of phase. It might be that because we cannot detect a position for an out-of-phase sound, many people think it is coming from behind them. This is nothing like as good as real surround sound, of course. 'Out of phase' is the exact electrical opposite of mono, and is actually difficult to achieve except in a professional sound system. The website audio resources include an example of in- and out-of-phase sound so that you can compare them.



You may be tempted to work with headphones rather than speakers in order to cut down noise leakage and avoid irritating other people, especially if you are carrying out editing yourself in the office. Beware, however, that when it comes to judging sound quality and the balance of a mix, headphones are notoriously unreliable. They should be fine for editing.

One way in which we detect the loudness of a sound in the real world is by feeling the pressure of air on our bodies. For this reason it is dangerous to listen to sound loudly in headphones. The air pressure clue to loudness will not be there, and there is a tendency to turn up the volume to the ears to compensate.

A detailed knowledge of psycho-acoustics has led to high-quality and efficient ways of reducing the bit rate of digital audio files in systems such as MPEG audio. Sounds that we do not hear because they are, for example, masked by other sounds can be removed from a recording without noticeable effect.

The most-efficient layer of MPEG audio, layer 3, has become a popular standard for compressing audio and is known as MP3. This reduces the data and bit rate required for a sound recording by analysing the audio and doing its best to remove the parts of the sound that are not actually heard. How well it does this depends on the data rate and other factors in the compression. There is an example of an MP3 file on the website together with a recording of what the MP3 compression removed.

Digital basics

Digital technology has entered most aspects of sound recording and editing. The basics of digitization are that the continuously varying sound waveform (the electrical representation of the vibrating microphone diaphragm) is sampled. This means that many times a second the instantaneous voltage of the waveform is measured. Audio is usually sampled at 44.1 kHz, which is the sample rate for compact disk (and Minidisk), which means that 44,100 times per second the instantaneous value of the waveform is measured and stored as a 16-bit number. This is a technique known as pulse code modulation (PCM) and it has been a basis of digital audio since its inception. For reference, middle A on a piano is currently standardized at 440 Hz, and when you double the frequency of a sound its pitch goes up an octave. (This standard has changed gradually: in Mozart's time, middle A was 430 Hz.) So a single cycle of a 440 Hz sound will be sampled in about one hundred places. PCM isn't the only game in town as we'll see in a moment.

The highest frequency of sound that can be faithfully reproduced by a particular sample rate is just under half that sample rate (as discovered by a Swedish scientist named Nyquist and known as the Nyquist Theorem), so the range of frequencies, called the bandwidth, of a compact disk is 22 kHz.

The 22 kHz bandwidth of a compact disk should be enough to reproduce all the frequencies you could hear but there are higher quality audio formats now used professionally and even reaching the consumer market. Incidentally, the digital tape cassette format DAT uses 48 kHz as its standard sample rate, but most DAT machines can also record at 44.1 (after a while you stop saying kHz every time), and, if you have a choice, 44.1 is the better sample rate because any compact disks or digital transfers from MiniDisk that you include in your mixes will have to be at 44.1 and ought to stay that way to help keep the quality up.

This relatively simple sample rate picture has become more complex as more digital formats enter the professional market. Some professional audio is sampled at 88.2 and 96 (twice 44.1 and 48) and even higher to give a cleaner sound for reasons too esoteric to debate here and there are emerging super-CD formats using DVD disks that use these formats. Video formats with digital audio may use sample rates of 48 or even 32.

CDs and DAT share a bit depth of 16 bits. This means that the sound can, in theory, be digitized with a precision of 16 bits, or 65,536 levels.

So far we've been discussing PCM, whereby the actual value of the waveform is measured thousands of times a second and stored. One of the new DVD-based audio formats, SACD (Super Audio CD), takes a different approach and measures changes in the waveform's absolute value rather than the values themselves (sometimes called delta or difference coding or pulse density modulation) and in the case of SACD this is done over 2.8 million times a second. (The actual value was chosen so that it could easily be down-converted to PCM sample rates such as 32, 44.1, 48 and their multiples.) What is measured is a simple up or down value for the waveform. The system claims a frequency response from DC (a frequency of zero) to 100 KHz and unparalleled quality. By sampling at such a high frequency it is possible to move sampling artefacts and noise way out of the audible range without having to actually filter them out using imperfect real-world filters. They also use the DVD trick of dual layers to put both a CD audio and SACD layer on a disk, making them backwards compatible. SACD's rival in the Hi-Q audio stakes is DVD audio which in two-channel mode is 192 kHz sampling at 24-bits. Six-channel is 96 kHz/24-bit, which is still higher than a two-channel compact disc.

Until the advent of sophisticated audio compression for the Web, from the likes of MP3, Liquid Audio and RealAudio, one common way to get a smaller file size for audio on the desktop was to reduce to 8 bits and 22 kHz. These days, 8-bit has become less common although you may still find it used as a format for system alert sounds or WAV files on a PC.

You can work out the background (for which read 'error') noise of a PCM digital signal from the bit depth, since the maximum error between the 'real' sound and the digitized version of it is half the minimum step in the digitization. Since 16-bit has 65,536 steps and 8-bit has only 255 you can see that 8-bit will be 256 times as noisy as 16-bit. You can hear examples of the dif-

ference that is caused by some different sample rates and bit depth in a file on the website.

Fortunately our ears do not respond to sound levels in a linear fashion, which is why a logarithmic measurement, the decibel or dB, is used to measure it. This means that we do not actually hear 256 times more noise. In fact the signal-to-noise ratio of a 16-bit system is 98 dB (which basically means you will never usually hear it) whereas for an 8-bit system it is 50 dB. Since every 6 dB increment makes a sound twice as loud this means that 8-bit is eight times as noisy as 16-bit. Also, the noise only occurs in the sound, not in the silences (unlike analogue hiss which is usually relatively constant and caused mostly by random background electrical impulses in the amplifiers), but it will be very noticeable on slight noises like rustles, so these should be removed from any recording destined for 8-bit.

With digital audio recording a balance has to be struck between recording at so high a level that you risk overloading – which means clipping the waveform and producing distortion – and recording so quietly that noise becomes noticeable. As a result, the reference level at which audio is recorded on professional digital audio and video tape machines is set so as to give plenty of leeway for loud sounds. This is called 'headroom' and usually means that the peak audio levels may be as much as 10 dB below the maximum possible. Increasing the audio level in post-production to maximize the level, so that the loudest peak fills all the bits of the sample, is called normalizing. Professional systems usually record sound with 20 or even 24 bits (24 bits means that the noise is 150 dB down which is the difference between a jet engine close-up and a silent room), which gives the engineers the freedom to record at a safe level without noise. For distribution the sound will, in most instances, still be converted back to 16-bit.

To convert from, say, 16 to 8 bits the procedure is simply to divide each sample value by 256 and round the errors to the nearest whole number. To convert from 8 to 16 you multiply the sample value by 256. This will, unfortunately, also multiply the errors that cause noise, so your 16-bit version will not sound any better than the 8-bit original.

To get the best level out of a digital audio recording you should normalize it. This means finding the loudest part and adjusting the level of the whole file so that the loudest part fills all 16 bits of the sample (or something close to that). Most audio CDs have been normalized. However, professional audio source tapes may not have been. The standard peak level for professional videotape is about 12 dB below the normalized maximum. If your other audio has been normalized then this will result in a level difference between different files.

In principle, you should make sure that digital audio is normalized to full volume before encoding, although there are some exceptions. Some Web compression systems for audio will distort with a normalized file and so may some playback systems. To help avoid this you can normalize to less than 100%. RealNetworks, for example, suggest setting the maximum a little

lower, say 95% of full level. This is equivalent to a level reduction of 0.5 dB. Also, files with the same levels can sound louder or softer since loudness does not completely depend on level. There is no hard-and-fast rule about this, especially when a user can choose the order in which things are heard. Perhaps the best approach is to normalize to 3 dB below peak, which will give a little leeway to boost quieter sounding files by normalizing them higher. DVD video disks usually have their average sound levels set lower so that the apparent loudness of dialogue is similar between movies and to allow for sound effects and music which will often be at a much higher level: explosions for example. The peak levels on a movie will be higher than the average level to a greater extent than would be the case on a rock album, so it will sound quieter most of the time.

The sample rate of a recording can be changed by recalculating the samples, and most audio editing and processing software will allow you to do this. When you take a sound file and process it with software such as Cleaner and RealProducer, any change in the sample rate will be recalculated. There are professional boxes that will do this as well and work from the standard digital audio interface connections. Reducing the sample rate of a file will reduce its frequency response and if the sound is not filtered first, following Nyquist's theorem, you will generate aliasing artefacts.

For reference, here are bit depths and sample rates that you might come across and where you might find them. This list is not exhaustive.

- 44.1 kHz 16-bit – CD, DAT and digital audio editing systems. As a rule of thumb this kind of digitized audio takes up 10 megabytes per minute. It is also known as Red Book after the name of the standard for compact disks and as PCM (Pulse Code Modulation).
- 48 kHz 16-bit – digital videotape formats, DAT, digital tracks on LaserDisks and some digital audio editing systems.
- 22 kHz 8-bit (or lower) – older personal computer sound and some streamed audio on the Internet.
- 44.1, 48, 88.2, 96, 176.4 or 192 kHz 20-bit (or more) – some professional audio systems and digital videotape recorders.
- 32 kHz 12-bit – long-play DAT and consumer DV.

There are what are termed DASH systems (digital audio stationary head), which you might come across in a recording studio, but these also use either 32, 44.1 or 48 for their sampling rate. Variants of DAT and hard-disk systems are now also using sample rates of 88.2 or 96. Sony used to sell a digital audio add-on for their Betamax video recorders called the F1. This was a 44.1 kHz 12-bit system. Other digital audio formats you may come across include:

- μ-law (used in telecommunications);
- NICAM (a 14-bit system used for stereo television sound);

- MPEG/ISO layers 1, 2 and 3 (layer 1 is very similar to PASC, which was used in Digital Compact Cassette; layer 2 is also called MUSICAM, and is used with MPEG-1 video and digital broadcasting and as an option for DVD; layer 3 – better known as MP3 – is used for telecommunications such as audio files on the Internet and broadcast radio contribution links down ISDN lines);
- ATRAC (used in MiniDisk);
- Dolby AC-3 (used on DVDs);
- MPEG-4 has an audio component and this includes more efficient compression but can also take an object-oriented approach to sound with separate parts of the 'mix' sent as separate objects and combined at the receiver. This way you could listen to an orchestra minus one instrument, for example, in case you want to play along.

Audio files can be downloaded (i.e. copied) across the Internet of course, but it is possible to stream audio on the Web using such formats as RealAudio and MP3. Streamed audio and streamed video are played in real time as the data arrives over the Internet. It is more like broadcasting than file transfer, and has the advantage (for rights owners) that a copy of the file is not usually left on the listener's computer.

The encoding for streamed audio over a modem is very efficient but results in an audio file with a very small audio bandwidth – as low as 4 kHz in the case of RealAudio for a 14.4 kBit modem – and high background noise levels. However, if you have two ISDN lines or a cable or DSL connection you can receive streamed audio of very high quality.

With the advent of DVD you will come across the two main standards for audio on the new digital video disks. Dolby AC-3 is the surround-sound format used for most DVDs, but a surround-sound version of MPEG audio (layer 2) is also possible as is a third format called DTS.

Finally, there are two common types of connection for digital audio: SPDIF and AES/EBU. SPDIF is the semi-professional system and is found on many consumer digital devices like CD players. It can use RCA/synch/phono connectors or, occasionally, miniature jack plugs for the electrical version but there is also a popular optical fibre implementation. AES/EBU is the professional version and uses XLR connectors like those used on microphones.

Aliasing

The Nyquist theorem says that in order to accurately digitize a sound (or indeed any waveform) of frequency n you must sample at a frequency of at least $2n$. If this rule is not followed strange results can occur. The phenomenon is called aliasing. (The effect on pictures will be discussed in Chapter 8 on graphics.) The result of aliasing is that the digital signal does not

accurately represent the analogue original. If there are frequencies higher than n in the signal that you digitize with frequency $2n$ there will be spurious samples in the result, and in audio this will usually sound like squeaking, and the signal should be filtered to remove such high frequencies before digitizing. Aliasing is a common problem in digital systems but it can occur in analogue systems as well. A popular example of the effect is seen on film and video when the wheels of a wagon appear to move backwards, and this also underlies the principle of using a stroboscope (regular flashing light) to 'freeze' fast regular motion.

This aliasing is a particular problem if you down-sample, which is where you take a sound digitized at, say, 22 kHz and shift it down to 11 kHz. You will think you can hear sounds of higher frequency than 5.5 kHz in the result, but they are not genuine sounds from the input but aliasing artefacts. For this reason you need to filter before you down-sample. Not all software does this, and you will hear the distortion that results. Down-sampling and aliasing can also introduce other artefacts. One instance is if the original recording contains a small amount of television line whistle at around 15 kHz from an NTSC or PAL signal. This will be almost inaudible, but if it is shifted down by aliasing to around 7 kHz, which can happen in a 22 kHz down-sampled sound, it will suddenly become audible because although the actual volume of the sound is the same, the ear is more sensitive at 7 kHz than at 15 kHz.

MIDI

A slight diversion takes us into MIDI – Musical Instrument Digital Interface. MIDI is an alternative way of encoding music which works completely differently to digital recording. A MIDI file stores information about the music in much the same way as sheet music. It stores information like pitch, duration and the instrument that should be playing the note. The actual sound is not stored in a MIDI file, it is the responsibility of the playback system to provide the sounds.

MIDI is the core of modern music making and there are numerous packages which allow a composer to work on a computer to produce music. In many cases you can work with MIDI and real audio side by side in the same package and along the same time line. Arguably without MIDI there would be no Pet Shop Boys and no Moby, so central is it to contemporary music making.

Professional musicians using MIDI will have their own hardware to produce the sounds of the instruments. These could be 'real' instruments such as an electronic piano or they could be produced by a sampler (playing short recordings of the real thing) or even a synthesizer.

Desktop PCs, if they include a sound card, will also include some form of MIDI playback and in many cases the instruments provided are surprisingly

natural. QuickTime now contains a set of instruments and, besides hardware solutions, you can also buy sets of instruments in software. There is a set called General MIDI which is defined as a minimum requirement of MIDI systems, so that a MIDI file using this set will play back sounding similar everywhere. The piano will be a piano even if you don't know if it will sound like a Steinway or a broken-down one from the corner of a bar on any particular system.

MIDI files do not take up much data since they are descriptive rather than literal and you can look on MIDI as the equivalent of text in its relationship with a sound recording. Web browsers will usually be able to play a MIDI file so one option for including music in a web page is to use MIDI. It won't take long to download and is an alternative to streamed audio.

Doing it on hard disk

Tapeless recording and editing systems are now commonplace in audio, and inexpensive (or even free) systems are available for desktop computers and lap tops. It is possible to record straight onto hard disk, and some audio facilities will do just that for you. These same facilities will edit your audio and prepare the tracks for use in a computer system by compressing them to RealAudio or MPEG audio (or whatever is appropriate).

Where the hard-disk systems come into their own is for editing because a digital hard disk audio package allows for more versatile editing than tape ever did.

Editing

Tape editing used to be done by physically cutting the tape with scissors or a razor blade. If you were editing yourself then you rocked and rolled the tape backwards and forwards across the playback head in order to locate exactly the point at which you wished to cut. This was sometimes called scrubbing. Then you marked the back of the tape with a soft chinagraph pencil and sliced the tape with a blade or scissors. You joined the bits you wanted together with adhesive tape. The adhesive tape was slightly narrower than the recording tape and had a non-leaking adhesive. A specially machined block was used to align the tape over the join. In many places you will still find analogue tape-based audio editing.

In a hard-disk-based digital system the sound is usually manipulated by working with a representation of the audio waveform on the computer screen. The sound is cut and pasted in much the same way as text in a word processor. You can still scrub to find the place to cut but now you are

manipulating sound in a file rather than on a tape. One useful feature of professional digital systems is the ability to do a mix across the joins to 'soften' the cut, which can be used to make an otherwise impossible edit work.

Some places are easier to cut than others. You can fool the ear by cutting into a sharp sound, a transient such as a bang or a drum sound. By doing this you do not usually notice any cutting off of the preceding sound. In fact the incoming sound is more critical than the outgoing in most edits. For speech some sounds make for better cuts than others; 'p', 'k' and 't' work well, whereas vowels and the 's' sound are quite difficult. Vowels are especially difficult because they carry most of the intonation in the voice and so they can sound completely different each time they are said. Consonants are more consistent and often you can splice them around to help to clean up word endings.

You should listen to speech patterns to help with your editing. Many people miss out letters from their speech. If you were to say 'next time' you would probably not pronounce both 't' sounds, but would actually say 'nextime'. You can take advantage of these truncations to find places to join speech together. You will even find that just looking at the waveform will help you to find places because you can easily identify pauses and consonants by the shape of the waveform.

Rhythm is important in speech, and your edits should respect that rhythm and not cut across it. Although people do shift the rhythm of their speech, most of the time an edit will feel more natural if a speech rhythm is preserved. Rhythm obviously is very important in editing music, and some of the same rules apply to both music and speech. A few milliseconds can make all the difference to a music edit. You can hear an example of 'good' and 'bad' editing in the resources on the website.

It is not true to say that a good edit is a joy to hear – because you will not hear a good edit.

Judging quality

Sometimes you will be required to produce assets to a certain quality. This might be specifically mentioned in a contract. It is difficult to define the quality of audio in objective terms. You could say that the frequency response will be one thing and the signal-to-noise ratio another, but these facts will not cover how well a presenter reads a script or how well mixed is a piece of music. The best course, should this issue arise, is to say that you will apply 'appropriate' standards or even 'broadcast' standards. You, and your clients or customers, will be able to compare with what you hear on radio and on CDs. It is important for clients to realize, however, that sound heard on a computer and on the Internet will often be of poorer quality than broadcast simply because the computer or the method of distribution is not capable of that quality of reproduction, no matter how well the material is prepared. On the bright side, you will usually find that the quality issue can be handled by comparing your results with other similar applications.

Choosing a codec

In order to prepare a sound file for use on the Web or in a kiosk or CD-ROM, you need to decide what quality you will be able to use and then choose a format. The format will usually be determined by the codec (coder–decoder) that you choose. Quality is generally dependent on the amount of data you can use but also some codecs give better quality results than others. Apart from on the Web, sound can usually be left in an uncompressed format such as WAV or AIFF. For a small amount of compression there is a format called ADPCM (adaptive delta – or difference – pulse code modulation) which can compress a few-fold with very little effect on the sound. It is supported on PC and Macs under the name IMA 4:1. For the Internet, a fourfold reduction in data is usually not enough.

To compress audio for the Web, whether along with video or not, there is a bewildering range of choices. We have already discussed MP3 and RealAudio but there are other options. RealAudio, like some other codecs for web audio, has settings attuned to speech or to music. It helps to use different techniques if you want to get best results for speech on its own whereas a music codec basically does its best with the whole sound. The production guides for each codec will tell you how to judge the best settings depending on whether you want stereo or not, whether it's just speech, the frequency range of the result and the likely speed of the final connection. This last factor is especially important if the sound is going to be streamed although (with the notable exception of a streamed radio station) many audio files are small enough for users to download them. MP3 does not have such a wide range of options but you still need to choose sample rates and bit rates for the final file.

One practical advantage of MP3 is that it is treated as 'native' on PCs and Macs with the same raw MP3 file playing using both Windows Media and QuickTime.

Compared with the uncompressed bit rate of one megabit for stereo sound, codecs like MP3 and RealAudio will achieve more than 12 : 1 reduction in data.

A final reminder, as with all asset preparation: keep an archive copy of your audio at a high uncompressed quality such as the audio CD standard Red Book or better. You may need to recompress the audio again sometime using a different format.

THEORY INTO PRACTICE 6

Take the editing practice-recording on the website (download it in its MP3 format and use an audio utility to convert for editing) and, in whatever editing tool you have, try to make the speaker say the opposite of what he originally said. Listen to see how natural this sounds.

Listen to the examples of audio with different sample rates and bit depths and see if you can recognize any undesirable effects that result from them. This includes loss of high frequencies and addition of noise.

Search out some freely available encoders for MPEG audio and RealAudio and whatever you can find. Take a sound file and encode it with the different systems and compare the results.

Summary

- This chapter has looked at the background to sound recording, and has explained what you should expect when using a professional audio facility to record voices for your website or multimedia application. It has outlined the preparation you need to make to prepare for the session.
- The kind of microphone used, the way it is positioned and how the sound is treated will affect the way your recording sounds.
- Stereo and multichannel positioning is usually achieved by adjusting loudness between the channels.
- During recording, scripts should be marked up for later editing.
- Keep an archive master copy of any audio that you process for inclusion on a website or CD-ROM.
- In digitizing, the highest frequency that can be digitized with a sample rate of $2n$ is n, otherwise odd-sounding artefacts are likely to appear in the recording.

Recommended reading

Moore B.C.J. (1997). *Introduction to the Psychology of Hearing*. 4th edn London: Academic Press

Pohlmann K.C. (2000). *Principles of Digital Audio*. 4th edn, Maidenhead: McGraw-Hill

Watkinson J. (2001). *Art of Digital Audio*. 3rd edn, London: Focal Press

Information on RealAudio encoding is available at their website: www.real.com, the audio advice is at http://service.real.com/help/library/guides/production8/htmfiles/audio.htm

The MIDI Manufacturers Association are at http://www.midi.org/

MPEG Audio information (including FAQs), is available at http://www.tnt.uni-hannover.de/js/project/mpeg/audio/

A document describing SADC can be found on Sony's website at http://www.sel.sony.com/SEL/consumer/dsd/dsd.pdf

The DVD FAQ, including information on DVD audio, is at http://dvddemystified.com/dvdfaq.html