Recording and Displaying Speech

Mark Tatham
Katherine Morton

Reproduced from: Tatham, M. and Morton, K. (1997) Recording and Displaying Speech. In Ball, M.J. and Code, C. (eds), *Instrumental Clinical Phonetics*. London: Whurr Publishers Ltd., pp. 1-21

Note: Like the chapter of the same name in the earlier *Experimental Clinical Phonetics* (Code, C. and Ball, M.J. (eds), London: Croom Helm, 1984) this work is probably mostly of historical interest today. Techniques for recording and displaying speech signals are constantly changing, but a few of the basic principles (e.g. microphone and recording techniques) have not changed much and remain useful. Direct to disk recording is probably the norm today, and computer screens and display technology have evolved to enable much higher resolutions than were available in 1997.

Introduction

The experimental investigation of speech is a very broad field. The titles of the chapters in this book show that the data we might want to examine come from a variety of sources and take a variety of forms. The sound wave itself is just the beginning, for there are several other aspects of speech that can reveal to us the nature of both normal and pathological speaking. Thus, we might want to inspect data from the neuromotor system, the aerodynamic system, the vocal tract anatomy or its configuration, as well as the final acoustic signal that results from the behaviour of these ‘underlying’ systems. This hierarchical approach to modelling speech production enables us to get some idea of how the final acoustic signal is derived, and, if there are errors, might help us to pinpoint their sources.

In each case, investigation of these layers in the system can involve quite different techniques and call for different approaches to the data: examining the electromyographic signals associated with muscle contraction is not the same as determining the formant structure of the acoustic signal of vowel sounds. There are, however, some principles of investigatory technique that are common to the entire field.

All the different areas of experimental work in speech involve using some instrumental technique to convert or transduce information about speech behaviour into electrical signals. Furthermore, as any scientific investigation requires careful control and interpretation of the data there can be no question that it becomes very important to have some kind of permanent record of the phenomena under investigation. This is required because it may become necessary to repeat the experiment or check the validity of any inferences that we might make.

In the laboratory study of speech we generally make two kinds of permanent record of our data:

1. a recording of the actual or raw data, obtained as closely as possible to the original conversion of the information into electrical signals;
2. a visual recording of the final output of any electrical or other processing of the data for inspection and measurement by the investigator.

We make a record of the raw data as close as possible in the investigatory chain to the point at which it was transduced into an electrical signal so as to minimize any distortion effects that the experimental equipment itself might introduce. This raw data recording can be rerun over and over again exactly as if the experiment itself were being run repeatedly. We need a visual record of what the researcher examines at the very end of the investigatory chain so that we can go back later and see why a particular inference was made without having to rerun the entire experiment. The raw data recording starts the chain and the visual data record ends it; in between the data has been processed and manipulated in various ways as part of the experimental procedure. The procedure itself will vary depending on the nature and source of the raw data, but the need for permanent records of the initial and final stages remains the same. Figure 1.1 illustrates the chain of events.

As we shall see, the raw data record is usually a binary representation of the electrical signals transduced from the point in the speech production system under investigation. The final record is usually an image of some kind on a computer screen (soft copy) or on paper (hard copy). There are various ways in which the records at both these levels are obtained and stored, and it is important that the choices available should not be made randomly or on some basis such as cost. The wrong choice could easily result in distorted or destroyed data, or even incorrect results for the investigation.

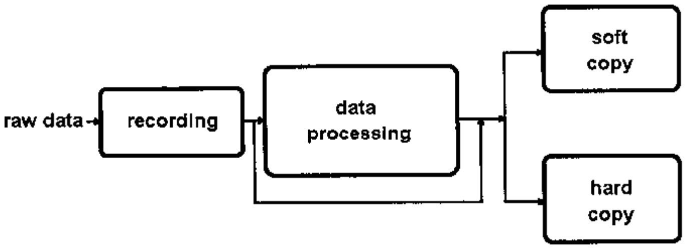

Figure 1.1. Raw data is permanently recorded before being processed and displayed – either as soft copy on a computer monitor or as hard copy on paper.

Choice among methods of recording and displaying speech signals has to be based on two major considerations:

1. the characteristics of the signal to be recorded or displayed;
2. whether or not the characteristics of the recording and displaying techniques match the signal's own characteristics without introducing distortion.

To make the necessary choices and to make the fullest use of the available equipment any researcher must have sufficient knowledge of the nature of both the data and the equipment to avoid the random or incorrect use of equipment.

Recording Speech Data Signals

Speech signals are inherently dynamic in nature and it is often the case that snapshot glimpses of speech can be very misleading. In the early days of laboratory experiments on speech, the techniques available placed severe constraints on dealing with the time-varying nature of speech. Thus, we could have a still X-ray picture of a cross-section of the vocal tract, but not a moving picture running for several seconds; or we could have a representation of a brief section of an acoustic signal from which we could measure, for example the formant frequencies of this or that vowel sound, but not a moving image showing how in running speech formant frequencies change dynamically according to the segmental context of vowels.

As the theory of speech production has shifted in the past 20 or 30 years from a static to a dynamic focus so the focus of attention in experimental work, and therefore in the equipment used in the laboratory, has become centred on the investigation of the dynamics of speech. The two, of course, go hand-in-hand, and it would be difficult to establish whether the changes in theoretical focus have influenced the introduction of new experimental techniques or whether the availability of modern, technology and the data it provides have driven the change in emphasis in the theory. What is certain, however, is that the need for permanent records at both ends of the experimental chain has become more critical and more demanding.

Analogue tape recording

The acoustic signals of speech have been recorded on tape for many years. The signal is analogue in form and the minimum of processing is needed to create a magnetic recording on tape which can be repeatedly replayed. Analogue tape recording is reliable in the sense that any distortions introduced are predictable and well understood, but it is often not appreciated that the recordings produced are not permanent. The physical and magnetic properties of the tape deteriorate with time. Magnetic deterioration is particularly troublesome in the case of the phenomenon of ‘print’. This is the tendency for the magnetic image to spread into successive layers of the tape winding, producing a faint and repeated echo synchronized with the varying time occupied by each layer of tape.

For periods longer than about a year, archiving data as an analogue magnetic tape recording is not recommended. However, in the short term, analogue tape recording remains the cheapest and most easily accessible method of recording speech data. It will be several years yet before other forms of recording become as common. This is partly because of the compatibility problem: although it is easy to make a case on technical grounds for replacing analogue by digital recording there is still a vast quantity of material in analogue form that can be used for the moment only on analogue machines. Digital recording is also, for the moment, comparatively expensive.

Realistically, therefore, the first choice of a medium for recording data signals for performing experiments (rather than archiving them) must be analogue magnetic tape, and we shall devote quite a lot of space here to examining the technique and its shortcomings. Understanding the general principles of recording is important whether the medium is analogue or digital.

In analogue tape recording the medium that actually holds the data is the magnetic oxide on one side of the tape, and this tape forms part of a mechanical system at the heart of the machine. It is important for the understanding of various types of tape recorder to realize that basically the machine consists of three parts:

1. the recording electronics;
2. the mechanical system including the tape itself;
3. the replay electronics.

In considering different types of analogue tape recorder we distinguish below between open-reel and cassette machines, and explain why for experimental purposes in the investigation of speech the open-reel machines are very much better. They are, however, available only as professional quality (as opposed to domestic quality) machines, and are consequently expensive. They have been almost completely displaced in the professional music recording business by digital machines which achieve higher quality and provide recordings that are much more stable in the long term. Having said that, many speech laboratories still use them, and the lower quality cassette machines are still the norm for domestic purposes.

Direct recording

As far as we are concerned in speech research, the limiting factor in the characteristics of tape recording rests with the mechanical tape system rather than with the electronics. The characteristics of the latter are generally sufficiently sophisticated to accommodate any speech signal. The actual process of getting the signal on to the tape, and keeping it there for replaying, however, is subject to some quite severe limitations, which will have an effect on our instrumental methods.

The first parameter of tape recorders we want to consider is that of signal-to-noise ratio. This is a way of expressing the difference between the amplitudes of the highest and lowest recordable signals. As the dynamic range of speech signals of whatever kind rarely exceeds 50 dB, we might specify that our minimum requirement is for a signal-to-noise ratio of 50 dB; the decibel is a unit of intensity, related to amplitude. That is, if our highest amplitude signal is recorded just below a level that would introduce an unacceptable amount of distortion into the signal (thereby influencing any subsequent investigation of that signal), then the noise ‘floor’ inherent in any tape recording should be at least 50 dB below that highest level.

The second important parameter of a tape recording system is its frequency response. This refers to the machine’s ability to record and replay a particular frequency range without distorting the amplitude relationships within that range. Thus, three tones of, say, 400 Hz, 1 kHz and 8 kHz of equal amplitude before recording must be reproduced after recording with their original equal amplitudes preserved. This is why frequency response specifications must be stated with reference to the recorder's ability to maintain this amplitude relationship. Generally, a typical specification might be: 45 Hz to 18 kHz plus or minus 2 dB meaning that over the frequency range stated amplitude relationships will be held on replay within a band 2 dB greater or 2 dB less than the amplitude of a reference tone at 1 kHz. Modern tape recorders easily achieve this level of amplitude integrity provided they are well maintained.

However, it is important to note also that the ability of a tape recorder to maintain amplitude integrity depends very much on the overall amplitude of the signal being recorded. A cassette tape recorder using a good quality tape would maintain the amplitude relationship in our example within 2 dB of the reference amplitude probably only if that reference amplitude were 20 dB lower than the maximum the machine could record at 1 kHz without more than the minimum of distortion. But raise the reference to that minimum distortion level (or 0 dB) and the same machine/tape combination might show a frequency response within 2 dB only over a range of 45 Hz to 8 or 9 kHz. This is insufficient for recording, say, the audio waveform of speech for the purpose of subsequent instrumental investigation. An open-reel tape recorder on the other hand would have no difficulty holding amplitude integrity to its maximum recording level for this given frequency range.

This illustrates a major difference between open-reel and cassette tape recorders. Their published frequency response specifications may often look identical, but usually for the cassette machine the reference level is 20 dB below the maximum level at which we will probably want to record. This means that high frequencies will play back with artificially reduced amplitude, making nonsense of any attempt to relate amplitude and frequency in a recording of the original signal. Or it means that you have to keep down the level of the recording, greatly reducing the usable signal-to-noise ratio of the recorder probably to a figure too narrow for our purposes.

With the cassette recorder, because of its miniature dimensions, the position quickly worsens as the machine ages or if it is not scrupulously maintained in a good and clean condition, so that, although it is true to say that high-frequency components of speech are generally low in amplitude anyway, an element of doubt is introduced when using a tape recorder that can achieve the required frequency response only at low amplitude settings.

Even with open-reel machines there is a general rule: better signal-to-noise ratios and better frequency response will be achieved with the widest tapes moving at the highest speeds. Consider that on a normal two channel cassette recorder the width of each track is one quarter (two tracks in each direction) of one eighth of an inch (the tape width) moving at 1.875 in/s, compared with a normal two-channel open-reel machine where the tracks are one half (two tracks in one direction only) of one quarter of an inch (the tape width) moving at 7.5 in/s (usually), or better at 15 in/s. The area of tape passing across the recording and replay heads in a given time is critical: the more the better. Less than one thirty second of the area of tape passes under a recording head per track on a cassette machine in a given time than on an open-reel machine running at 15 in/s. There are a few cassette recorders available that run at a speed of 3.75 in/s, which might just make acceptable recordings for instrumental analysis, but these are rare and problems of compatibility with other recorders arise.

Distortion in tape recording is another parameter that must be taken into consideration. In general, the most disturbing form of distortion occurs when the oxide on the tape becomes magnetically saturated. This happens if we attempt to record a signal of too great an amplitude (see below). Most tape recorders are satisfactory from this point of view provided no attempt is made to record a signal above the 0 dB reference point indicated on the machine's recording meters. Such a level should give a distortion level of less than 1% which should not bother us unduly in subsequent analysis of the replayed signal.

FM recording

The above description of the characteristics of analogue tape recorders refers to ordinary or direct recording machines. They are referred to as direct recording machines because the raw signal does not undergo any special transformation as part of the recording or replaying process.

The usual lowest frequency that can normally be recorded accurately is seldom below about 35Hz. Many of the signals that we need to record for instrumental analysis, however, contain components below this frequency, and indeed may contain frequencies as low as 0 Hz; that is, the analysis may contain periods where there is no change in signal. Such steady state signals are rather like the signal you would get by connecting two wires to a battery: a constant (not changing) amplitude of about 1.5 volts. Speech signals that come into this category include the aerodynamic signals of air pressure and airflow (see Chapter 4), glottograph signals (Chapter 5), and some components of electromyography signals (Chapter 3).

Clearly, an ordinary tape recorder is going to be unsuitable for recording signals of this kind: it will simply fail to record the low frequency components of the signal or will hopelessly distort them. This is an area of data recording where digital techniques have, in speech research, already completely displaced the older analogue techniques. A few laboratories still use an analogue technique known as FM (frequency modulation) recording when the signal has a predominance of low frequency components. The technique was described fully in the first edition of this book (Code and Ball, 1984). All we need take note of here is that there are many data signals derived from speech production, other than the acoustic ones, which cannot be successfully recorded on to analogue tape. Under these conditions a move must be made to digital techniques.

Noise reduction

Many analogue tape recorders, especially cassette machines, incorporate a noise reduction system. All of these alter amplitude relationships in the incoming signal in order to compress a signal’s dynamic range to make it easier for the tape to accommodate it. They are primarily designed for music recording where the dynamic range of the signal may well exceed 90 dB (much wider than the dynamic range of speech). On replay, the compression is reversed to expand the recorded signal back to its original dynamic range. Provided that the expansion is a perfect mirror image of the compression, then in theory what comes out of the machine will be identical to what went in.

In practice, such an ideal situation is never achieved, and, depending on which noise reduction system is being used, amplitude/frequency integrity is more or less disturbed and several intrusive forms of distortion are introduced. For instrumental analysis of any speech signal (as opposed to just listening to a recording) the only advice possible is: do not use any noise reduction system. If your tape recorder cannot achieve a better signal-to-noise ratio than, say, 50 dB (and many cassette machines, especially the portable ones, cannot) without the help of noise reduction then the machine is unsuitable for any of the research techniques described in this book.

Digital recording

Digital recorders work by converting the analogue signal of the speech waveform into a binary representation. It is this binary representation that is recorded on to the recorder’s magnetic tape. On playback the binary representation is read from the tape and converted back into analogue form before being amplified and sent to loudspeakers or earphones. Alternatively, the binary representation can be transferred directly to a computer.

Analogue to digital conversion

All signals connected with speech (with the exception of some components that originate from neural signals) are analogue in form. That is, amplitude variations in the signals exhibit transitions that are smooth and continuous in nature. Digital tape recorders, by contrast, expect to record signals by sampling the signal's amplitude level at particular discrete moments in time. For each time interval a number is recorded that is a rounded measurement of the average amplitude during the time interval sampled. The analogue signal's smooth (or continuous) amplitude changes in time are therefore converted into discrete (or discontinuous) amplitude measurements.

Figure 1.2 shows the relationship between an original analogue waveform (in this case a simple sine wave) and the quantized version resulting from sampling amplitude levels during discrete periods or slices of time. Notice the loss of the smoothness that is characteristic of analogue signals, and how the digital signal exhibits jagged discontinuities of amplitude. The process of changing from smooth to discontinuous representation of amplitude is called analogue to digital conversion, and we speak of digitizing or sampling the original analogue signal. There are two parameters to the analogue to digital conversion process: frequency response and dynamic range. Frequency response is determined by the sampling rate and dynamic range (the range of amplitudes that can be faithfully represented) is determined by the number of different discrete levels of amplitude to which the converter is sensitive.

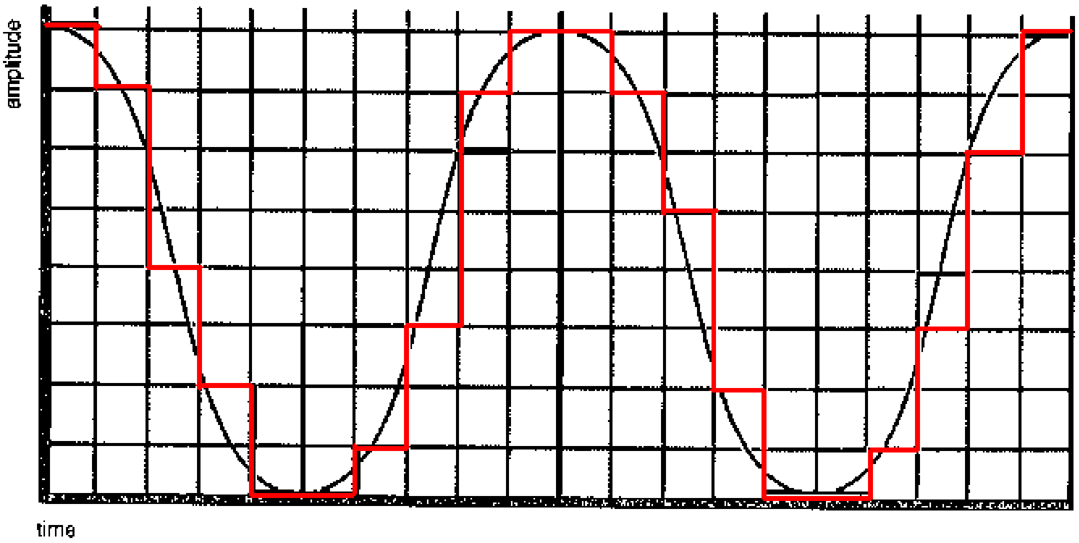

Figure 1.2. Analogue to digital conversion. The relationship between an original analogue waveform (a simple sine wave) and its quantized version.

Analogue to digital converters enable us to sample signals at different time rates, but clearly the more often the incoming signal is sampled the less obtrusive will be any discontinuities. In fact, it is possible to imagine sampling, which occurs so frequently that the original smoothness of the analogue representation is almost completely preserved; indeed, it would be preserved completely in theory at an infinitely high sampling rate. It helps to understand sampling theory if we imagine analogue representation to be simply a special case of a digitized representation: one where the sampling rate has been infinite.

In general, to capture a given frequency range a digital device needs to sample the input signal at a rate somewhat more than twice as often as the highest frequency in the incoming analogue signal; this is known as the Nyqvist rate. So, if we expect speech to have its highest frequency component around 12 kHz, then it must be sampled at least 25,000 times per second. Fortunately, no digital tape recorder's analogue to digital converter samples this slowly. They all accommodate input frequencies up to 20 kHz because all are basically designed to cover the entire range of human hearing (20 Hz to 20 kHz) and music signals. So, at least with digital tape recorders, we need not worry about frequency response. There are some established standards of sampling rate for digital audio. Thus, for example, for a CD (compact disc) the sampling rate is 44,100 times per second – or 44.1 kHz – enabling an audio signal up to about 22 kHz to be represented. Digital Audio Tape (DAT) recorders (see below) sample at rates of 48 kHz, 44.1 kHz and 32 kHz, enabling signals up to about 24 kHz, 22 kHz and 16 kHz respectively to be represented.

The dynamic range available in the analogue to digital conversion process IS expressed in terms of bits (binary digits). CD audio has a 16-bit dynamic range covering 65,536 possible discrete levels, meaning that for the CD standard the audio signal is sampled 44,100 times each second and its amplitude in any one sample is expressed as a rounded number out of 65,536 possible levels. The 95 dB dynamic range that can be achieved with 16-bit systems is easily reached on DAT recorders, though some portable versions may encode to only 14 bits, achieving an 85 dB dynamic range: still more than adequate for speech work.

Digital Audio Tape (DAT) recorders

There are several standards current in digital tape recorders, but for use in speech research one standard is pre-eminent: Digital Audio Tape recording (DAT). Laboratory and portable versions of these machines are available from several manufacturers. DAT machines use narrow tape housed in cassettes. The digital signal is laid down on the tape using a helical scan principle similar to that used in video tape recorders; this enables an effective tape speed much higher than the actual speed at which the tape is pulled through the mechanism. Many DAT machines have a so-called long play mode which reduces the tape speed to achieve longer recording times. In this mode the sampling rate is reduced from a standard 44.1 kHz to 32 kHz with a consequent drop in upper frequency response to about 15 kHz. Long play mode should generally be avoided, not because of the restricted frequency response, but because it cannot be regarded as a true standard and interchangeability of cassettes between machines is made more difficult.

A DAT recorder can be regarded as a multi-stage system. The various stages, following the progress of a signal through the system, are:

1. analogue to digital conversion;
2. signal conditioning, the digital signal suitable for recording on the tape;
3. the actual recording;
4. replaying the recording;
5. digital to analogue conversion.

It is at stage 5 that the digital signals being replayed from the tape are converted back into analogue signals. Thus, a DAT recorder accepts analogue signals as input and outputs analogue signals. This is entirely suitable for music purposes, particularly in a domestic environment, but there are problems in using DAT in a speech laboratory.

Sometimes we may want simply to copy a recording. In this case we would not wish to take an original digital recording, convert it to analogue, and then reconvert it to digital form before making the copy. Check to make sure that you can directly input and output a digital signal; if you can, then use this direct digital connection. This facility is useful for more than copying between recorders. We can record a signal that is already in digital format (say, straight from a computer) or take a digital signal directly from a DAT recording into a computer. Playing the DAT recording directly into a computer is the normal situation in a speech laboratory; it is the commonest route to processing and subsequently displaying recorded speech data. Remember: when setting up such a system it is crucial to ensure that the sampling rates and bit levels are matched for the sending and receiving equipment; failure to do so will result in pitch and amplitude anomalies.

From the point of view of actual operation the digital machine is the same as an analogue machine, with one exception: when recording it is extremely important not to try to record a signal of too high an amplitude – overload distortion on a digital tape recorder is far, far worse than on an analogue machine. However, as the dynamic range of digital machines is greater than that for analogue machines, the recording levels can be kept down to make sure that this does not happen.

One thing it is necessary to be aware of with DAT recorders is that editing (see below) on digital machines is quite different from editing on analogue machines. Physical editing in the form of tape splicing is not possible at all because of the helical scanning system adopted for digital recording. Electronic editing during copying from one DAT recorder to another, while possible and extremely accurate, is difficult and requires expensive special equipment. Therefore, editing is best done once the signal has been passed to a computer for processing and subsequent display.

Direct to disk recording

As virtually all processing and experimental work with speech signals, whether audio or of some other type, is now carried out by computers, it may sometimes be useful to set up a system where signals from transducers such as microphones, electro-glottographs, etc., are taken directly to the computer without recording on to tape. Although it is good practice to record the data, it is now possible to bypass tape and make recordings directly to the computer's hard disk. For lengthy quantities of data, though, a tape recorder is needed because, although hard disk drives of capacities of one or more gigabytes are commonplace, it is very easy to fill up the available space. Data exchange with other laboratories is still easier using a format such as DAT, although for this purpose CD-ROM (compact disk – read only memory) should also be considered (see below).

Direct to disk recording requires a sound card to be installed in the host computer. For our purposes the term 'sound card' here is some what of a misnomer because we can use the card to enable us to input and output speech signals other than the audio. All of the signals used for the experimental work described in this book can be accepted by these cards.

There are many sound cards available and there is a variety of different standards in use, but we recommend cards with variable sample rates for the analogue-to-digital and digital-to-analogue converters, and an amplitude resolution of 16 bits. Many such cards also provide an interface for a CD-ROM drive, at least one of which should be available in speech laboratories as many databases are being made available to the speech research community in this format. Sound cards have on-board means of accepting analogue signals direct from audio sources and other forms of data sources in speech investigation. Their function is to condition the signal to make it suitable for storage as data files on the machine's hard disk. They perform the additional function of enabling hard disk files to be read and converted back to analogue form for listening or other purposes. Cards are available with both analogue and digital input and output channels.

Sound cards often have on board a digital signal processing (dsp) chip. This is a processor that has been optimized for the kind of processing needed for manipulating audio (and similar) signals at high speed. Much of the commercially available software for laboratory computers uses dsp facilities for transforming signals for display, for example, as spectrograms, which require high speed and powerful processing if they are to be made available in real time; that is, while the utterance is actually being made.

CD-ROM recordings

CD-ROMs store audio and textual data on compact disks. If the speech data is accompanied by a video signal this can also be stored on CD-ROM, although in this case a video card with sound will be needed for processing. Making your own CD-ROMs from data on your hard disk is a possibility, although the drives necessary for this are still comparatively expensive. Most speech laboratories currently have read-only CD-ROM drives, and rely on obtaining from elsewhere the material already recorded on CDs. Make sure your CD-ROM drives have software enabling the material to be transferred to hard disks and at the same time converted if necessary from one standard to another. For example, it might be necessary to change the sampling rate so that the material can be further processed by additional software. The standard capacity of a CD-ROM is around 640 Mbytes, although increased capacities are available.

A word of caution if you feel you would like to use the soundtrack of a video available on CD as data in an experiment. The audio that accompanies these videos is compressed so that it takes up less space on the CD. There are standards for video compression, which include a specification of how the audio also is to be compressed by as much as 11:1; that is, such that it occupies less than 10% of the CD’s capacity than it would occupy if uncompressed. The protocols for compression and subsequent expansion are called MPEG-1 and MPEG-2 (after the Motion Picture Experts Group, which worked them out) and these protocols are incorporated in the processing carried out by the video card installed in the computer or video player.

The problem from the point of view of the researcher is that the three levels of audio compression specified by the protocols all rely on a coding scheme which is described as ‘perceptual’, and this term is not well defined. The audio signal is transformed from the time domain (where it looks like waveforms on a screen) to the frequency domain (where if displayed it looks like spectrograms). It is then manipulated to remove what MPEG consider to be perceptually redundant components of the signal. At this point, it is converted back to a time domain representation, recoded and recorded. On playback the final analogue signal fed to the loudspeakers is said to be perceptually satisfactory. However, no one really knows exactly which elements of a speech signal are or are not perceptually redundant. In any case, compression of such a severe nature where a considerable portion of the signal is irrecoverably removed (whether perceptually relevant or not) renders such signals completely useless for serious speech research. It is worth noting that some analogue broadcast TV and radio signals, especially those from satellites, and all digital broadcast TV and radio signals are similarly compressed and thus present similar problems.

Many speech laboratories can now access the Internet system for linking computers worldwide and exchanging data. Most of the data available on the Internet is text and for that reason the capacity of the system for transmitting data is strictly limited. But as users increasingly want to transmit data that is not text – including binary coded audio data – standards are being set up to enable the compression of these wider bit rate data types so that they can be sent around the Internet as easily as text. The same warning holds here as in the case of video CDs. The compression systems for audio on the Internet will often follow the MPEG standards, with similar distortion of the signal. If you intend to exchange audio data with other researchers on the Internet you should make sure that you understand how the various compression systems alter your data.

The audio recording session

In this section, we discuss some of the essential techniques and equipment for making audio recordings. In most cases, simple precautions will make the difference between a recording that is unusable for instrumental analysis purposes and one which is entirely suitable for detailed analysis. You should treat data as a valuable re-usable resource: it makes sense to have your recordings of the highest possible quality and made with care.

Making the recording

The majority of recordings that the speech pathologist will make during the course of his or her duties or research will be of the audio waveform of patients and of normal speech for comparison purposes. There are good ways and bad ways of making a recording, especially when the material is to be analysed instrumentally. Listening to a recording will not normally provide a good judgement as to its quality; this is part of the problem we noted earlier with digital perceptual coding compression techniques. The reason for this is quite simple: a subjective impression will tend to overlook the imperfections in any recording, unless they are very gross, but these imperfections will show up in any instrumental analysis that might follow. This may lead to difficulties and inaccuracies in measurements. The only way to ensure a good recording is to know what factors influence the quality of recordings, and try to make sure you have obtained the best conditions possible.

Location

Echo is one of the biggest problems likely to be encountered. Ideally, recording should be made in a studio especially designed for audio. Unfortunately this is going to be available to very few clinicians. The next best thing is to select the quietest, most heavily furnished room possible, and preferably one certainly no larger than the average-sized living room. The idea is that heavy furnishings (particularly soft chairs, carpets and curtains) absorb unwanted reflections that bare walls, floor, windows and ceiling would normally produce, and in addition provide some kind of insulation against noises coming into the room from outside. Listen carefully for such unwanted noises. Normally we tend not to notice them ourselves, but the microphone will mercilessly pick them up. Listen out especially for the noise of people walking along corridors, aircraft and traffic noise, and particularly for the drone of air conditioning systems. These have become so much a part of our lives that, on hearing a recording of ‘silence’ made in a normal room, the amount of background noise is striking.

Microphones

Having chosen the quietest, least reverberant (or deadest) room you can find further echoes and outside noises can be minimized by carefully selecting microphones and using them properly. Omni-directional microphones (which pick up sound from all around them) are usually not suitable. After all, the signal is usually coming from just one direction: from the lips of the subject. Choose a directional microphone and make sure it is pointing roughly at the subject, although not in such a way that it gets blown on to directly; the subject should be talking across the microphone. There is no need to go into the vast array of types of microphone available; most these days, except the very cheapest, are good enough.

Select a microphone that has a reasonably flat frequency response over the speech range (say, 75 Hz to 12 kHz ±3 dB). There is one kind of reliable and excellent microphone that satisfies almost all the conditions-for our recordings; this is the battery-powered electret microphone mounted with a lapel clip or slung on a cord around the neck. Choose the directional type. Such a microphone has the additional advantage that almost automatically you are likely to mount it in precisely the right place, about 40 cm from the subject’s mouth, not immediately in front of him or her (to avoid breath noises), and not on some reverberant surface like a table. In fact the only disadvantage with this kind of microphone is the possible pickup of the rustle of clothes, so check on this.

One further point on microphones: it is often necessary to record a conversation between two or more people, say between the clinician and patient. In this case, you must use two microphones connected preferably to the two separate channels of a stereo recorder. In this way you will find that on playback it is very easy to keep the two signals almost separate with just enough ‘breakthrough’ for you to hear what is going on by listening to just one channel. If more than two people are to be recorded, then the best practice is to have a microphone for each, with the signals mixed electronically using a microphone mixer before recording on one or two channels. We do not recommend placing a single omni-directional microphone on a table in the middle of a group of subjects. The mix of signals is difficult to decode because directionality is missing and this will become apparent when you come to analyse the recording. Good microphone technique cannot be overemphasized. Try making several recordings with different microphone positions; analyse them, say by making spectrograms, and see how different the signals are.

Stereo recorders are readily available these days to provide two-track recording as described above in either the analogue cassette format or in the higher quality DAT format. If you are going to do a great deal of recording and your experiments are particularly important it is well worth investing in a DAT recorder.

One word of warning: compatibility between recorders is not guaranteed, and you should be careful to ensure you can play back recordings either on the machine used for making them or on another machine you have previously tested for compatibility. Do not rely on specification sheets to indicate the compatibility of two apparently identical tape recorders, particularly the cassette type. There can be a slightly different alignment of the record and playback heads, which will make tapes recorded on the one machine reproduce badly on the other. A good diagnostic with an analogue machine, which you should listen out for, is loss of high frequencies on playback on the second machine when the tape played back perfectly satisfactorily on the original machine.

Using the gain control

Avoid using automatic gain controls for recording. These come labelled in several different ways, so if in doubt consult a competent engineer to ask whether automatic gain control (often cued AGC) is used on the machine, and if so then how to switch it off. The trouble with automatic gain control is that, although such a system takes much of the work out of making a recording, it will considerably distort the amplitude relationships of the recordings you make and will faithfully record the background noise you have gone to such pains to remove. The reason for this is that when there is no intended speech signal the AGC increases sensitivity in an attempt to find one – the system is not intelligent enough to distinguish between wanted and unwanted sounds. AGC systems are commonest on portable recorders, even DAT machines.

Having decided never to use AGC, you are now faced with using the manual gain control for recording. Imagine that a tape recorder looks at the amplitude of sound through a window. That window has a top and a bottom. The top and bottom are represented on the meter used in conjunction with the gain control. The window top is marked 0 dB and corresponds to the point where the meter scale usually changes to red (this applies both to meters like dials with pointers and to luminous displays). The bottom of the window is the far left of the meter (or bottom if it is mounted vertically). If you have the gain control too low the signal will be at the bottom of the window and insufficiently ‘seen’ by the recorder. On the other hand, if you have the gain control too high then the signal will overshoot the window resulting in considerable unwanted distortion, and giving an unusable recording. The control must be manipulated to get the signal within the window.

How do you do this? Consider what speech sound is again for a moment. It has a certain amplitude range, and we already know that most tape recorders can cope with that range. Some sounds have more amplitude than others, so tape recorders need to be set so that the loudest sounds just kick the display to the 0 dB mark and leave the rest to get on with it. So how do we know what the loudest sound is going to be before it has happened? Research shows that the speech sound usually with the highest intrinsic amplitude is the [ɑ] sound in a word like *cart*. If possible, get the subject to say this sound, or a word containing this sound, several times into the microphone before you begin the recording session proper. Adjust the gain control carefully so that the meter just registers 0 dB, and no more. Make sure the subject is talking in what you expect to be a normal voice. Let the subject practise using a microphone beforehand to make the voice as normal as possible.

Once you have set the gain control before the actual session begins do not touch it again, unless you can see during the session that you had obviously set it wrongly. The point here is that the gain control improves or worsens the recorder’s sensitivity to signals. If you change the setting during the recording, you will not be able to compare the amplitudes of anything recorded before the change with those of anything recorded after the change and you may want to do this. If it is absolutely necessary to make a change and the session cannot be restarted then do so deliberately and quickly, making a note of what you did and when you did it. But preferably start the session over again.

Listening to a recording

No tape recorder with built-in loudspeakers is good enough for listening purposes, except for the crudest monitoring. To listen seriously to any recording you need the highest fidelity playback system available or affordable. Only the best systems will accurately preserve the amplitude and frequency relationships that make up the speech to which you are trying to listen. Failing a good loudspeaker system, use headphones. The fidelity of headphones can often be deceptive; they sound better than they really are objectively. But many people prefer them for auditory analysis purposes because, by putting you closer to the signal being replayed, some find that concentration on listening is much better and that it is easier to be more objective in making judgements of what is being listened to. It is really up to you which you prefer, but try to ensure the best fidelity possible. Once again, it is a question of looking for the flattest adequate frequency response curves.

Tape editing

The need for editing arises when portions of a recording need to be removed, or sections from several different recordings need to be put together on to. a single tape. There are two ways of doing this: one is by physically cutting the tape and splicing it together again in the required sequence, and the other is to accomplish the same thing by electronic means. Cutting and splicing tape is a very time-consuming business and can really only be done successfully if the original recording is made in the open reel format at as high a tape speed as possible (to give the most room for locating the edit point on the tape). If you do go in for physically editing the tape in this way, make sure you get plenty of practice beforehand and never edit your original recording (you may make a mistake and destroy it). Always work on a copy of the tape. That way if you mess things up, you just make another copy and begin again.

Remember, though, that copying tape results in degradation of the signa1, so you must have an exceptionally clean recording to begin with. Furthermore, if your final spliced tape contains many joins or is to be kept for more than a few weeks, you cannot guarantee that your splices will hold and a copy of the edited tape must be made. You will then work from this final copy. This final tape is a copy of a copy, with attendant multiplied degradation. Having made these warnings, though, it is not likely that you will be using an open-reel recorder. It is more likely that you will have to deal with a cassette machine or a DAT recorder.

Electronic editing is better than physically splicing tape and is all that you can do with cassette and DAT. Electronic editing is done by connecting two tape recorders together, taking the signal out from the machine holding the original tape and putting it into the second machine. The sections of the original recording to be edited must be found by careful listening, and then copied on to the new tape on the second recorder with its controls set to record. It is worth noting that, unlike the situation with analogue recording, copying digital recordings should not result in any degradation in quality. Provided the binary representation can be read satisfactorily, a perfect copy can be made with no introduction of noise or distortion.

Displaying speech data signals

The output of instruments used for analysis in the laboratory is often presented in some visual form. This requires connecting the instruments or tape recorders to a computer via the sockets on the installed sound card, using analogue inputs for analogue material and digital inputs for digital material. It is important to make sure in advance that the signals being presented to the card are the type that it is expecting. In particular, it is important to distinguish between analogue and digital signals, and in the case of analogue signals to make certain that the voltage levels are correct.

The computer will need to be running a software package controlling the sound card drivers and checking that the signal is brought into the computer without introducing distortion. Often at this point there will be the option of making a copy of the original recording directly on to the hard disk for displaying and analysing later, or displaying now and saving to the hard disk later. Choosing which will depend on the purpose of the experiment. You may want to keep long unedited portions of the material on the hard disk; in which case you may want to make the disk copy now. Check to see there is enough file space. But you may want to inspect and edit the data before committing it to disk; in this case you will want to display now, inspect and then decide whether to commit to disk.

Temporary display

Software packages for inputting data, editing it and analysing it all have a means of displaying the data on the computer screen. The software provides for displaying both raw and processed data in separate windows. Often, data in various stages of processing or analysis can be displayed alongside one another; there are many possibilities. For complex and detailed data it pays to have large, high quality display monitors. This is the only part of the computer you actually look at, and it is worth spending a large proportion of the budget on it. It is not uncommon, especially with PCs, to spend more money on the display than on the computer itself in the laboratory environment. There is a great deal of difference between the requirements of office usage of PCs and their use in speech laboratories.

Software for displaying and analysing speech varies considerably, not just in terms of its functionality, but in terms of its ergonomics, or ease of use. You will sometimes get the impression that the software designer was unfamiliar with its actual usage in a speech laboratory. But having said that, there are many excellent general purpose packages available. These are too numerous to discuss here, but they range from simple programs that simply control the sound card for inputting and outputting sound waves, through waveform display and editing, to complex signal processing and displays. You will have to see what is available at the time of your experiment and choose accordingly.

In general, it is probably better to adopt programs that use standard computer configurations and also standard sound cards to ensure compatibility with other laboratories. Make sure that the software saves files in standard formats also; once again, this is to ensure compatibility with files produced by other researchers. You might want, for example, to exchange recordings in binary representation on floppy disks rather than as tape recordings.

For specialist work that goes beyond the simple editing of audio signals and perhaps spectrographic displays, you should look to software prepared by researchers in the field, and often marketed by themselves or made available free. There are many speech software packages available on the World Wide Web for free downloading and a systematic and regular search of the Internet is well worthwhile. You will soon get the hang of where to look and which WWW servers to visit regularly for the latest in software. But check on compatibility and how the data has been pre-processed, for example compression as mentioned above.

Permanent display

The computer screen is useful as a temporary display for viewing your sound files and is essential for editing sessions, but you will want to save many of the displays you generate as permanent records on paper. Although there are still a few around, the days of strip chart recorders have gone. These machines were very useful indeed; the paper format was ideal, matching the fact that speech unfolds in time, and the only limit to the length of time displayed was the length of paper on the roll. Today, however, page printers are almost universal.

A page printer is one that, as its name suggests, prints single pages, usually in A4 format with either portrait or landscape orientation. The commonest of these are the laser printer and the ink-jet printer. Laser printers compose an entire page prior to printing, whereas ink-jet printers generally print the page while data is being received (meaning you can actually see the page being continuously printed). Strictly speaking, the term ‘page printer’ should not be applied to an ink-jet printer for this reason. The print quality obtainable from a laser printer is superior to that from an ink-jet printer in both black-and-white and colour.

Most signal processing software will enable you to dump what is on the screen to a printer with automatic adjustments to format and aspect ratio to make sure that what is on the screen looks right when printed on paper. In the case of colour laser or ink-jet printers it is often possible to produce a printed image with different colours from those appearing on the screen. It may well be the case, for example, that colours appropriate for a screen are not suitable for a display on paper. The software will make adjustments automatically for printing the data at the appropriate printer resolution (the number of dots per inch). In general, the quality of print obtained by printing the file directly from disk is better than simply dumping a screen print to paper.

One additional piece of software that is very useful is a screen capture program. These enable you to ‘grab’ part of all or the screen display and send it to a printer or save it to a file in a standard graphical format. There are two reasons why you may want to do this:

1. To enable a quick printout of what is displayed in the screen or in a window on the screen – sometimes more useful than having the analysis software print the file from disk.
2. To produce an illustration in, say, a text article being produced in a word processor or desktop publishing program. The text processor simply loads and sometimes resizes and rescales the picture for placing at the desired location in the text.

Similarly, material on paper may be transformed into a graphics file for illustration purposes by using a scanner. These devices are able to accept paper illustrations in either colour or black-and-white for creating a standard format graphics file on your hard disk. Scanners operate at resolutions varying from 150 to 1200 dots per inch, but, of course, cannot improve on the resolution of the original picture.

Do not expect that a screen image will be more detailed on paper than it is on the screen. Screens have an aspect ratio of 4:3 and resolutions of 640 x 480, 800 x 600, 1024 x 786, 1280 x 1024, or 1600 x 1200 pixels, whereas printers have resolutions of 300, 400 or 600 dots per inch: the screen has comparatively low resolution, therefore, compared with what is possible on paper. In addition, the window in which your data is displayed on the screen will have fewer pixels available than the entire screen. This means that some of the details in the signal will not show on the screen.

Conclusion

These are the main points we have been making in this chapter covering the means of recording and displaying speech signals:

1. Be sure you are fully aware of the characteristics of the speech signals you wish to record and display.
2. Choose the right machine for the recording job in hand, making sure in particular that you understand the limitations of cassette recorders.
3. If possible now, and certainly in the future, use a digital recorder. You will never have any worries about quality if you do, although editing may be difficult. Failing a digital recorder, open-reel analogue is better than standard cassette analogue (although the former is now comparatively rare).
4. Check the frequency range of your data signal. If there are components lower than about 35 Hz you will need a frequency modulation tape recorder or digital machine designed for the purpose. Such signals can often be recorded in a direct-to-disk session using the sound card of a computer.
5. Make sure you have the highest quality display you can afford on your computer. You cannot expect to make accurate observations of the data if the display is unable to show the signals without distortion.
6. For the highest-quality permanent records choose a laser printer rather than an ink-jet recorder.

Reference

Code C, Ball M.J. (eds), *Experimental Clinical Phonetics*. London: Croom Helm, 1984.