Lecture 1-6: Audio Recording Systems

Overview

1. **Why do we need to record speech?** We need audio recordings of speech for a number of reasons: for **off-line analysis**, so that we can listen to and transcribe speech after the event; for **instrumental analysis**, so that we can make quantitative measurements of speech signal properties; to obtain a **permanent record** of a how a person sounded so that it can be analysed later, say after therapy or after treatment; to allow **distribution** of a recording to more than one listener.
2. **The recording chain.** You can consider audio recording and playback as a involving a chain of different components and a sequence of transformations of the signal: the **microphone** converts sound to electricity; the **pre-amplifier** increases the size of the microphone signal; the **recorder** converts the signal to a physical form which can be stored permanently, often magnetisation of ferromagnetic materials; the **storage medium** holds the recording but may itself deteriorate with time; the **player** reconverts the stored form back to a signal; the **amplifier** makes the signal large enough to drive a **loudspeaker** which converts signals back into pressure waves (see figure 1-6.1). Each component of the chain is a **system** that may change the signal passing through it. The overall quality of the chain is the combination of the quality of the components.
3. **Measuring Quality.** A number of standard measures are used to describe the quality of the component systems in the recording chain: the **frequency response** of a system shows us how it changes the spectrum of the signal passing through, typically this is described using the range of frequencies which have a response with 3dB of the peak response, a good audio system will have a 3dB range of more than 10,000Hz (see figure 1-6.3); the **signal-to-noise ratio (SNR)** is the ratio of the average signal level to the average noise level, where the noise is any unwanted signal added to the input by the system, a good audio system will have an SNR of better than 50dB; the **distortion** of a system is a measure of how many new harmonics are introduced by changes to the shape of the input signal, a good audio system will have distortion components smaller than -50dB ; the **crosstalk** of a system is how much signal leaks between the left and right channels (this is only important for simultaneous speech and laryngograph recordings), a good system will have crosstalk less than -50dB; the **wow and flutter** measure applies only to analogue tape recorders, and refers to the size of slow (wow) and rapid (flutter) changes in tape transport speed.
4. **Audio Components.** Microphones come in three main types: **crystal** or **piezo** microphones which are cheap and of poor quality; **dynamic** microphones which operate with a moving coil and can be found for moderate price and good quality; **condenser** or **electret** microphones which exploit capacitance changes and which tend to be more expensive but of higher quality than dynamic microphones. Most condenser microphones require a separate power supply which makes them less convenient unless they can be powered from the recorder. Microphones also vary in their directional sensitivity, (see Figure 1-6.4) Recorders also come in three main types: **Compact Cassette** is the most common type and uses analogue audio cassettes; cheap recorders should be avoided, but cassette recorders of good quality can be bought for £100 or so. **MiniDisc** recorders are becoming more popular and use a compressed digital format on small optical disks; minidisc recorders can be purchased for about £150 although they are not suited to

recordings needing instrumental analysis. **Digital Audio Tape (DAT)** recorders are professional broadcast-standard recorders which record onto 4mm tape cartridges; DAT recorders are of very high quality, but tend to be very expensive, usually more than £400. The advantage of digital recorders is that the reconstructed waveform is always of the same quality as when originally recorded (see figure 1-6.2).

5. **Noise reduction.** A significant problem with analogue cassette recordings is their poor signal to noise ratio which leads to recordings that have significant background "hiss". Noise reduction methods like Dolby exploit the fact that it is the high-frequency components of the hiss that are most troublesome. A Dolby noise reduction system boosts the high frequency components of the input signal going on to the tape (which are typically of low amplitude anyway) and then reduces the size of the high frequency components coming off the tape after recording which also reduces the size of the high-frequency noise generated by the tape (see figure 1-6.5).

Reading

Choose at least one from:

- ◻ Baken, Clinical Measurement of Speech and Voice (1st edition), Chapter 3: General Purpose Tools, pp 43-78. *Accessible description of audio equipment.*
- ◻ Rosen & Howell, Signals and Systems for Speech and Hearing (1st edition), Chapter 14: An introduction to digital signals and systems. *Non-mathematical introduction to digital processing of signals.*

Learning Activities

You can help yourself understand and remember this week's teaching by doing the following activities before next week:

1. Write definitions in your own words of frequency range, signal-to-noise ratio and harmonic distortion
2. Write a description of the characteristics of an "ideal" audio recorder. Be sure to think about usability as well as performance. List ways in which "real" recorders depart from this ideal and the consequences for recordings.
3. Find out about the differences between digital and analogue recording systems and summarise their advantages and disadvantages.
4. Find out how a noise reduction system like Dolby operates and draw a diagram describing the processing involved and its effect on the signal.

If you are unsure about any of these, make sure you ask questions in tutorial.

Reflections

You can improve your learning by reflecting on your understanding. Here are some suggestions for questions related to this week's teaching.

1. What is timbre? Can a tape recorder change the timbre of a sound?
2. What is pitch? Can a tape recorder change the pitch of a sound?
3. Does a tape recorder with a poor frequency response distort the signal? (think about the difference between reversible and irreversible changes)
4. What can you do to avoid distortion on your recordings?
5. What is the difference between 'system' noise and 'background' noise?
6. What does "Dolby" do? How does it work?
7. How might we measure the frequency response of a microphone?
8. Think of things you might do or check in the clinic to improve the quality of your recordings.

Figure 1-6.1 Analogue Tape Recorder

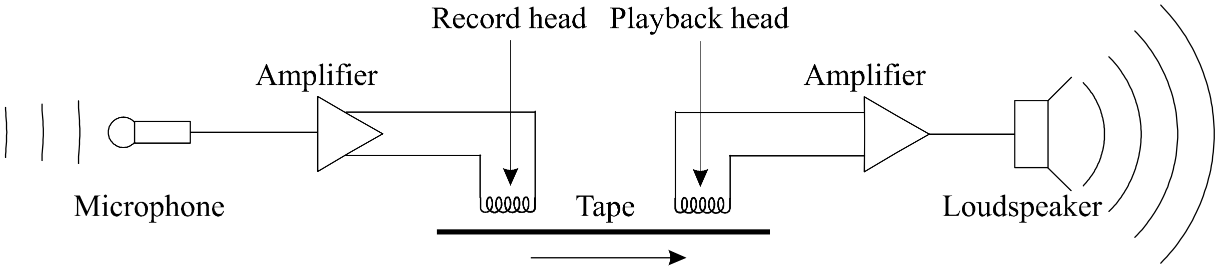


Figure 1-6.2 Digital Tape recorder

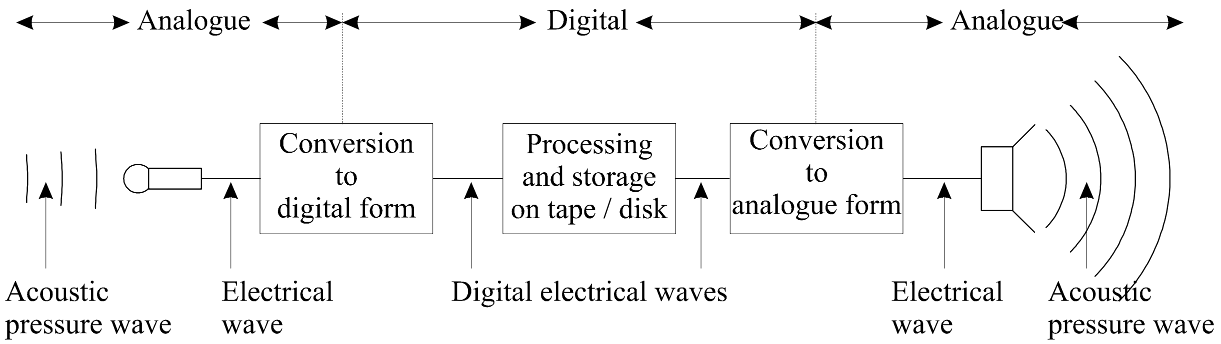
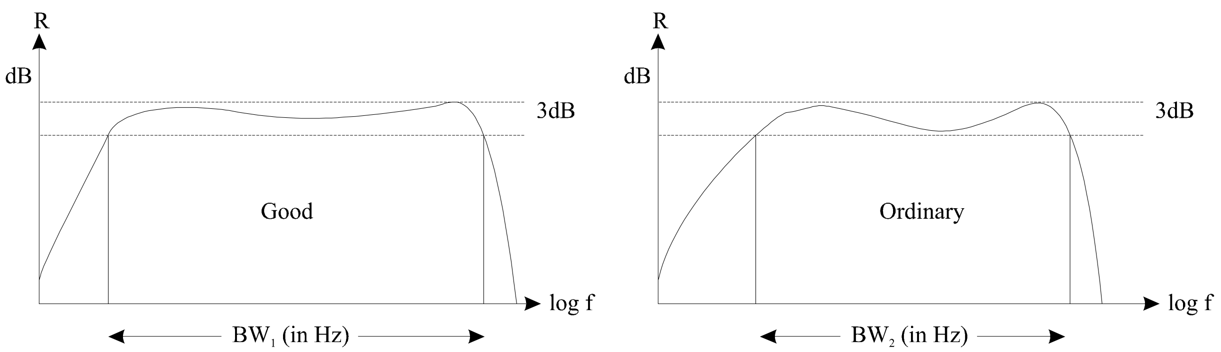
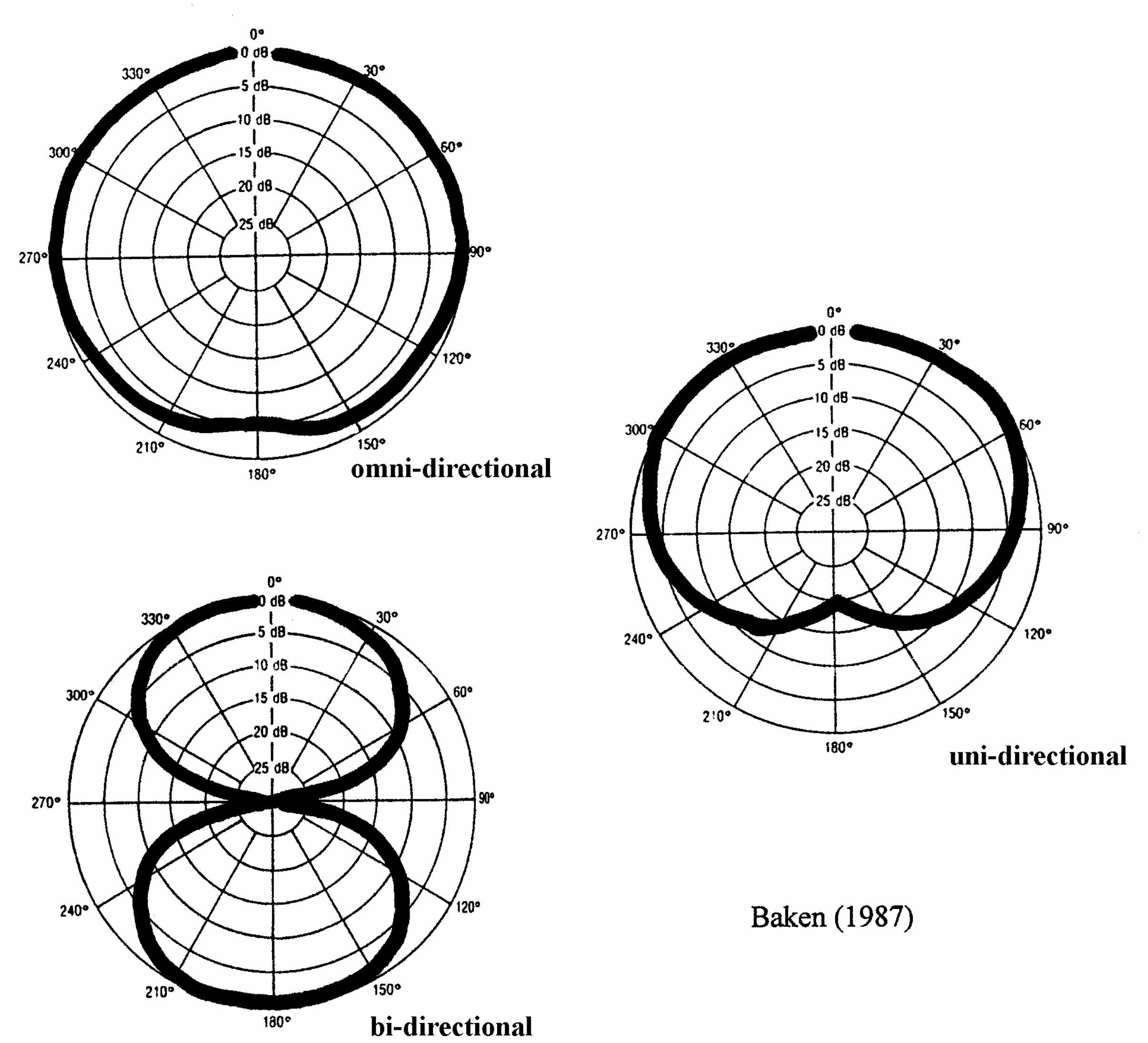


Figure 1-6.3 Comparison of Tape Recorder Frequency Response



Better systems have flatter response curves which have wider bandwidth. For speech, frequencies between 100 and 10,000Hz need to be faithfully reproduced.

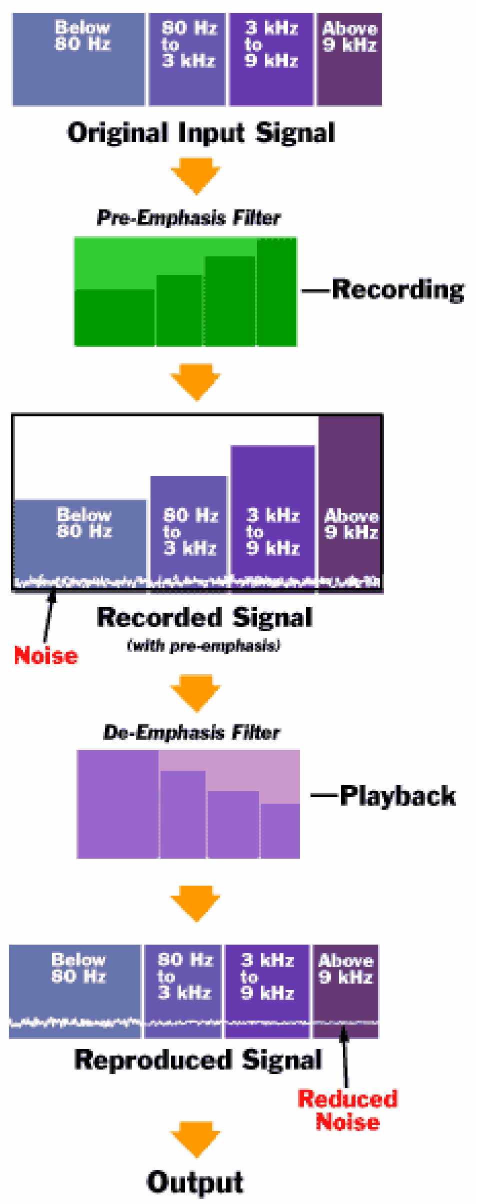
Figure 1-6.4 Directional Response of Different Microphone Types



Omni-directional microphones have similar sensitivity in all directions. They are good for the middle of a conference table. Uni-directional microphones have greatest sensitivity in one direction. They are good for making recordings in noisy places, where you can direct the microphone at the source. Bi-directional microphones have greatest sensitivity in two opposite directions. They are good for recording a conversation between two talkers.

Stereo microphone pairs (not pictured) are two microphones with directional sensitivities that cover the left and the right sound fields.

Figure 1-6.5 Dolby Noise Reduction



In Dolby noise reduction the high frequency components of the signal are amplified on recording and attenuated on playback. Since the most troublesome aspects of additive tape noise are at high frequency, the attenuation stage also reduces the perceptual impact of tape noise. (diagram adapted from www.howstuffworks.com)

Checklist for Speech Recording

1. Choice of subjects

- adult/child, male/female, normal/pathological,
- trained/untrained, accent, social background, educational background

2. Choice of material

- read/spontaneous,
- monologue/dialogue,
- controlled/uncontrolled,

3. Selection of environment

- quiet surroundings, soft furnishings

4. Selection of equipment

- external microphone,
- good quality tape & recorder

5. Operation

- distance to microphone from speaker
- set recording levels
- keep microphone out of airstream
- don't speak over subject's speech

6. Management

- keep details of recording, speak onto tape, record footage,
- label tape cassette and box,
- interpret unintelligible speech

7. Maintenance

- clean tape heads and rollers
- store tapes away from heat and magnetic fields

Lab 1-6: Tape Recorder Frequency Response

Introduction

A *Frequency Response* graph shows how the amplitudes of sinusoidal signals are modified when passed through a system. For each frequency, the graph shows the amplitude ratio of the output signal to the input signal at that frequency. Conventionally, this amplitude ratio is expressed in decibels, so a system that made no changes to a signal would have a response of 0dB at all frequencies. Such a response would be the ideal shape for a tape recorder where we would like the reproduced signal to be identical to the original signal. A frequency response that is flat to within 3dB from 100 to 10000Hz is adequate for speech.

The *Signal-to-noise ratio* of a recorder measures the amount of noise added to a recording by the audio recording system. Audio cassette recorders often have a poor signal-to-noise ratio. An SNR of 50dB or greater is adequate for instrumental analysis of speech.

The level of *Harmonic Distortion* in a recorder measures how much the shape of the recorded waveform changes during recording. It does this by measuring the size of any additional sinewave components that are added to a sinewave signal by the recorder. Harmonic distortion of -50dB or less is adequate for instrumental analysis of speech.

Scientific Objectives

- To measure the frequency response, signal-to-noise ratio and harmonic distortion levels for a given audio recording system.

Learning Objectives

- To appreciate how audio equipment can be assessed for quality.
- To learn how to interpret frequency response graphs of audio equipment.
- To see the effects of overload distortion.
- To gain familiarity with tape recording technology
- To gain familiarity with making tape recordings.

Apparatus

You will be given an audio recording system of a particular kind to assess. This may be an audio-cassette system, an audio cassette system with Dolby™, a Digital Audio Tape recorder, or a MiniDisc recorder. An oscillator is connected to the input of the recorder. The output of the recorder is connected to a milli-voltmeter to measure the output level and to an oscilloscope to monitor the waveform shape. The output can also be fed into a spectrum analyser to look at its harmonic content.

Measuring Frequency Response

1. Set the recording level so that a signal at 1000Hz is about 20dB less than the maximum allowed by the recorder.
2. Record about 10 seconds of each of the following pure tones on the recorder. Be sure to keep a record of the position counter at the start and end of each tone, and be wary about the "leader" tape on cassette recordings.

Start Position	End Position	Frequency (Hz)
		10
		20
		50

		100
		200
		500
		1000
		2000
		5000
		10000
		11000
		13000
		16000
		20000

3. Play back the recordings, noting the voltmeter reading for each frequency.
4. Set out your measurements in a table with headings:
   a) Counter reading
   b) Frequency (Hz)
   c) Output amplitude (mV)
   d) Output amplitude ratio to 1mV (the value in c) divided by 1mV)
   e) Output amplitude ratio to 1mv (dB) (20 x $\log_{10}$(the value in d)))
5. Plot the frequency response graph using the logarithmic graph paper supplied. Note the type, make and model number of the recorder on the graph.

Measuring Signal to Noise Ratio

1. Set the record level to just below the maximum allowed for each recorder (for example 0dB VU for the cassette recorder, -5dB for the digital recorders).
2. Record about 30 seconds of a 1000Hz tone.
3. Switch off the oscillator and record about 30 seconds of silence.
4. Play back the recording and note the average voltmeter reading for the maximum level recording and for the silence.
5. Calculate the signal to noise ratio from 20.$\log_{10}$(signal level/silence level).

Measuring Harmonic Distortion

1. Connect the output of your recorder to a PC running the spectral analysis program RTSpect. (http://www.phon.ucl.ac.uk/resource/sfs/rtspect/)
2. Replay the recording of the 1000Hz tone you recorded previously. Print out the spectral analysis for your file.
3. Note the level of the sinewave at 1000Hz.
4. Note the level of the largest harmonic of 1000Hz present.
5. Subtract the two levels to find the amount of harmonic distortion in dB.
6. If you have time, repeat this measurement with a 1000Hz tone recorded at 20dB *above* the maximum allowed by the recorder (do this by setting the input level to 0dB and then increasing the oscillator output voltage by 10). This should show some significant distortion components. Print out the spectral analysis for your file.

Concluding Remarks

Collect the data from the whole class, noting the type of recorder used by each group.

Write your report as if you were evaluating your own recorder for clinical use. Compare your results with the group results pretending that they are previously “published” data. Be sure to explain the limitations of your study. Discuss whether your recorder would be suitable for recording speech for clinical purposes.