Lecture 1-8: 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; or 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 electrical signal; the **recorder** converts the signal to a physical form which can be stored permanently, often by the 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 1-8.1). 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 within 3dB of the peak response. A good audio system for speech will have a 3dB range extending from 100Hz to at least 10,000Hz (see figure 1-8.3); this will allow it capture all the important frequency components in speech. The **signal-to-noise ratio (SNR)** of a system is the ratio of the average signal level to the average noise level at the output, 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; this will ensure that the noise does not mask any important information in speech. The **distortion** of a system is a measure of any changes to the shape of the input signal made by the system, usually estimated from the relative size of any new harmonics added to a periodic signal. A good audio system will have any added components –50dB below any true components; this will ensure they will not interfere with speech perception. Older analogue tape recorders were also assessed using **crosstalk** (how much signal leaks between the left and right channels) and **wow and flutter** (measures of the stability of tape transport speed). These measures are not relevant for modern digital audio recorders.
4. **Microphones.** 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)
5. **Recorders**. Recorders also come in three main types: **Compact Cassette** recorders use analogue audio cassettes and are only of moderate quality. Their use now is deprecated,

having been superseded by digital recorders. **MiniDisc** recorders are 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. **Solid-state digital recorders** are professional broadcast-standard recorders which record onto flash memory cards. Solid state recorders can be of very high quality, but tend to be expensive, usually more than £150. The advantage of digital recorders is that the reconstructed waveform is always of the same quality as when originally recorded (see figure 1-8.2). Solid state recorders often apply 'compression' to signals to increase the capacity of the recorder (common compression systems include MP3 and AAC). However compression always reduces quality, so for instrumental analysis it is best to choose an 'uncompressed' setting (sometimes this is called 'PCM' mode).

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 the lab or 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 1-8.1 Analogue Tape Recorder

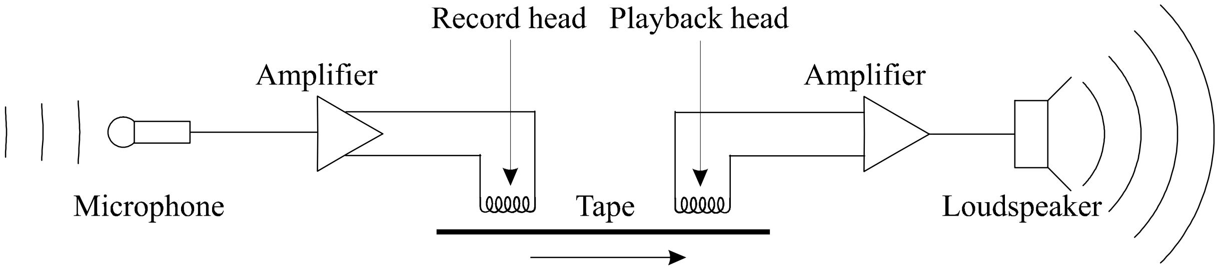


Figure 1-8.2 Digital audio recorder

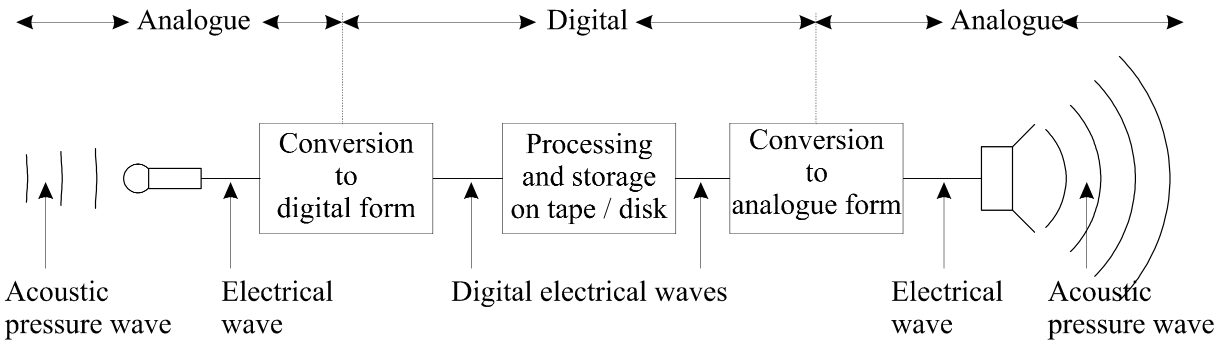
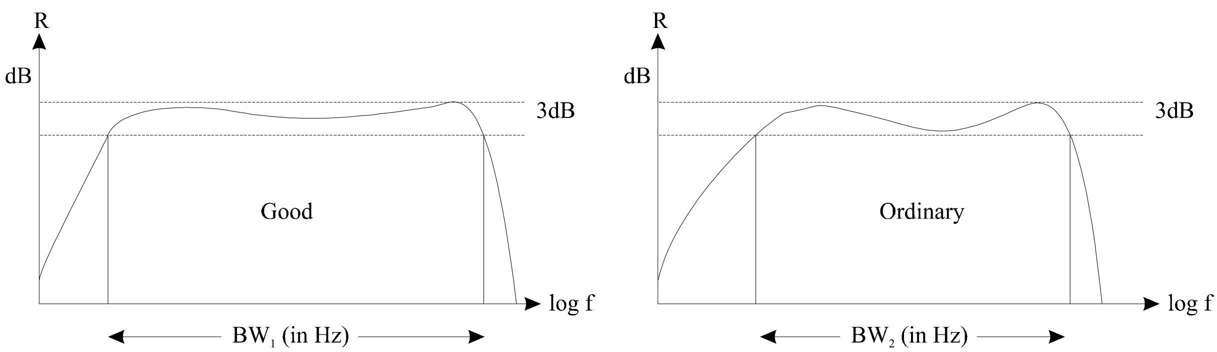
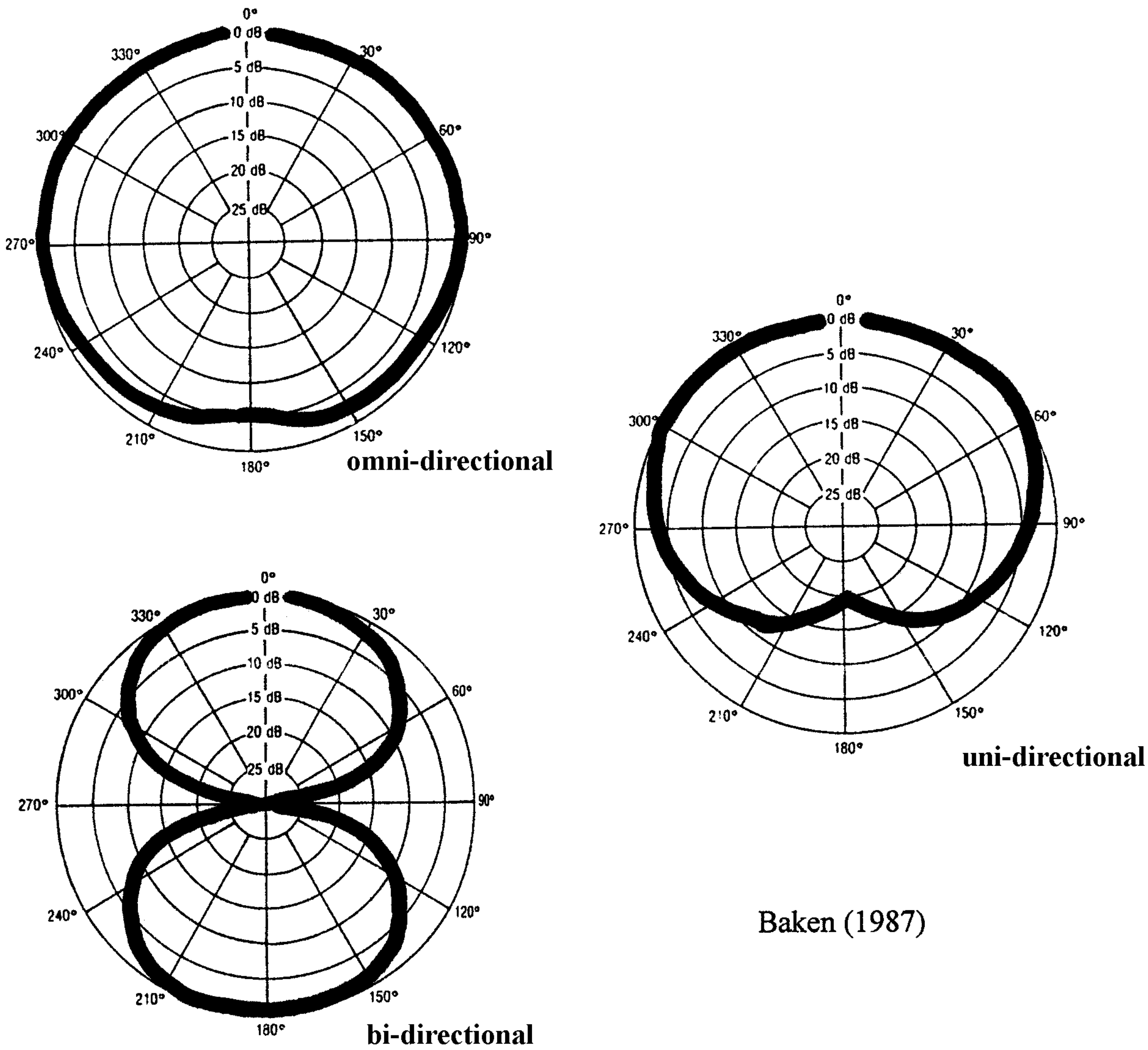


Figure 1-8.3 Comparison of Tape Recorder Frequency Response



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

Figure 1-8.4 Directional Response of Different Microphone Types



Baken (1987)

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.

Checklist for Speech Recording

1. Obtain details about the speakers

- physical character: e.g. age, sex, pathology, smoker, etc
- language background: e.g. L1/L2, dialect, accent, level of education, profession

2. Choose the right material to record

- read speech/spontaneous speech
- monologue/dialogue
- nonsense words, word lists, sentences, narratives

3. Select the right environment

- quiet surroundings
- non-reverberant surroundings, soft furnishings

4. Select the right equipment

- external microphone
- good quality recorder

5. Take care in operation

- keep distance to microphone < 0.5m
- keep microphone out of airstream
- set recording levels to get largest input that does not overload recorder
- don't speak over subject's speech

6. Manage your recordings

- keep details of recording, speak onto tape, record rough times
- label tape/disk/memory
- interpret unintelligible speech

Lab 1-8: Recording Speech

Introduction

Making a good recording of speech involves getting a large number of factors right. These include selecting the right equipment, choosing the speaker and material, making the recording in the right environment and in the right way.

In this laboratory session you will have an opportunity to make a number of recordings under a number of conditions and to compare the results.

Scientific Objectives

- To determine the most important factors for making a good quality recording of speech.

Learning Objectives

- to apply your knowledge of tape recording in the field
- to learn to operate analogue and digital recorders
- to understand the problems of speaker and material selection
- to appreciate the effect of different recording conditions
- to gain familiarity with the process of making a recording of a quality high enough for experimental phonetics research purposes

Method

You will work in a groups of 3. Each group will be given a portable tape recorder and a microphone. You should rotate the jobs of 'director', 'recording engineer' and 'speaker' over the different recording activities.

Once you have made your recordings, you should copy 20 second extracts of your recordings into a folder on the lab network so that the class can compare the efforts of all of the groups.

Observations

You should make three types of recording:

(a) Read sentences
(b) Spontaneous speech
(c) Dialogue

For each of these choose one of the following recording environments:

(a) Acoustically treated room (e.g. Research Laboratory Booths)
(b) Reverberant Room (e.g. Corridor)
(c) Noisy place (e.g. Kings X station or Euston Road)

You should end up with three different recordings. Each of a different type of speech and each in a different environment.

When the recordings are played back to the class, make a note of the conditions under which the best and worst recordings were made. What problems did people have?

Concluding Remarks

1. Having listened to the class recordings, write a list of 5 things you consider most important in making a good recording.
2. Write a list of the 5 most significant problems that occurred in the recordings.
3. Which problems with recorded signal quality do you think are most significant as far as experimental phonetics analysis is concerned?