LING 120 Introduction to Speech Analysis
Fall 2009

Week 1

Overview of the course
Recording and sampling

Sep. 9, 2009

Syllabus

http://www.ling.upenn.edu/courses/ling120/

The speech chain

- The speech communication consists of a chain of events linking the speaker's brain with the listener's brain.

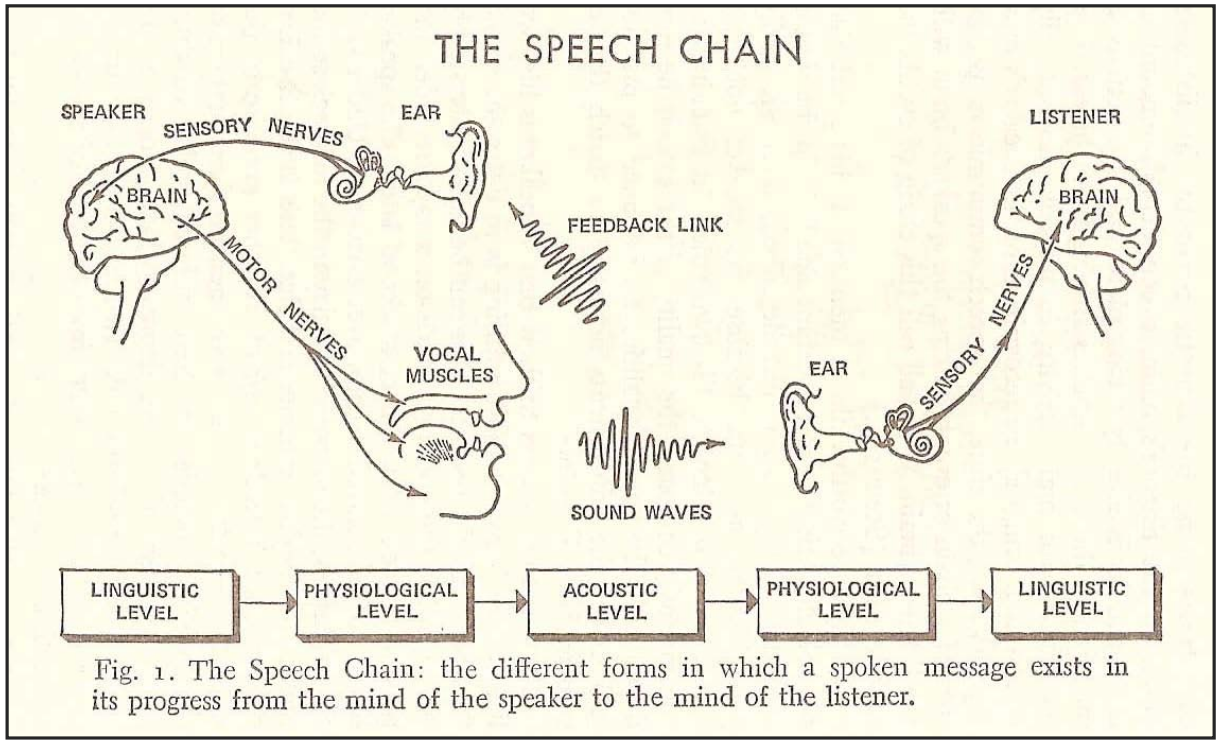


Fig. 1. The Speech Chain: the different forms in which a spoken message exists in its progress from the mind of the speaker to the mind of the listener.

(from Denes & Pinson, 1993)

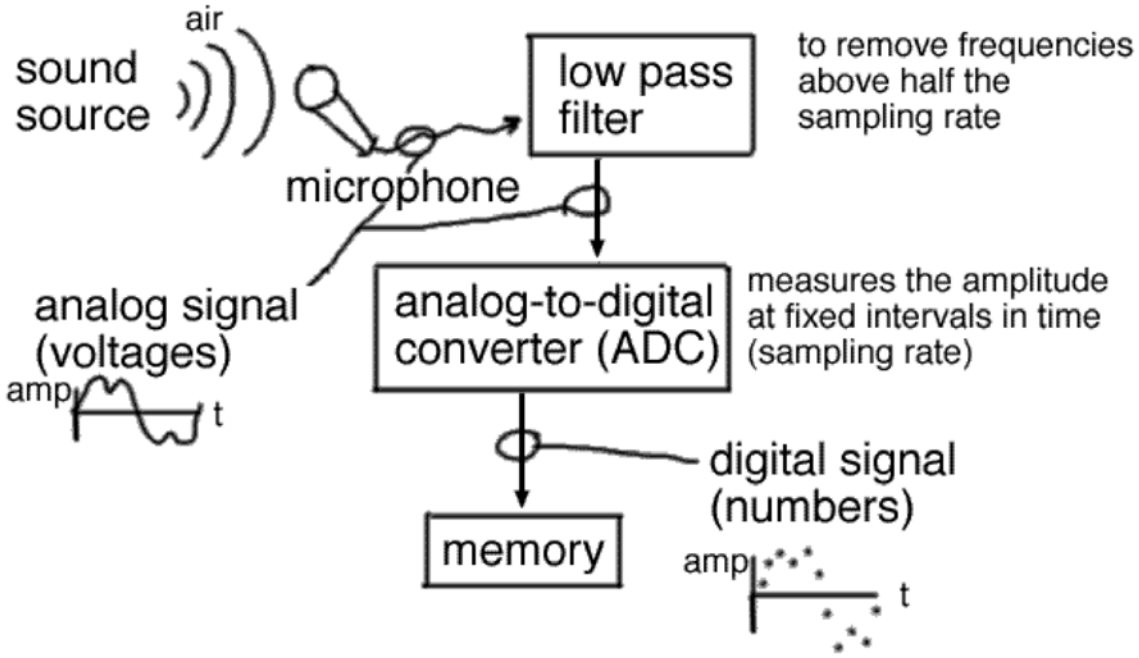
Linguistic Organization

Linguistics: The scientific study of human language

- Phonetics: physical nature of speech
- Phonology: sound structure of language
- Morphology: the structure of words
- Syntax: the structure of sentences
- Semantics: the meaning of words & sentences
- Pragmatics: how speakers and writers use language to do things

Recording

- **Digital recording:** The process of converting speech waves into computer-readable format is called *digitization*, or A/D conversion.



Recording

- **Signal to Noise Ratio (SNR):** Signal strength relative to background noise. The bigger the number, the better.

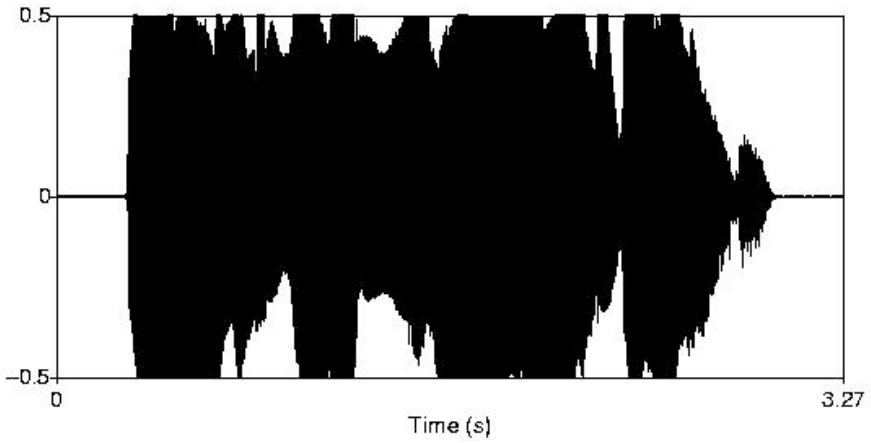
$$\mathrm{SNR(dB)} = 10\log_{10}\left(\frac{P_{\mathrm{signal}}}{P_{\mathrm{noise}}}\right) = 20\log_{10}\left(\frac{A_{\mathrm{signal}}}{A_{\mathrm{noise}}}\right)$$

- Classroom recording (SNR 29 dB)
- Laptop recording (SNR 44 dB)
- Professional recording (SNR 90 dB)

[From: Chilin Shih]

Recording

- **Clipping:** The sound is too loud (overloaded) for one or more components in the recording setup.



[From: Chilin Shih]

Sampling

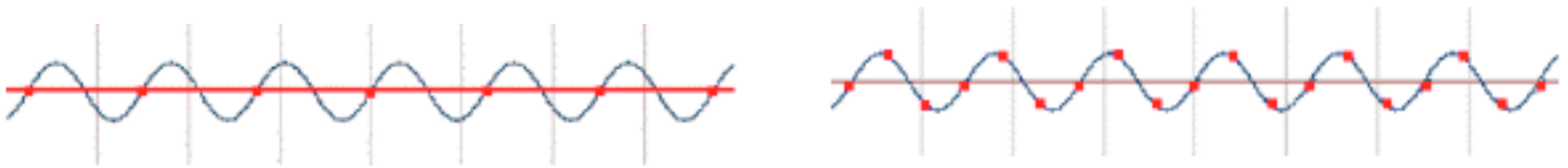
- In order to transform sound into a digital format, you must sample the sound. The computer takes a snapshot of the sound level at small time intervals while you are recording.
- The number of samples taken each second is called the **sampling rate**. The more samples that are taken, the better sound quality. But we also need more storage space for higher quality sound.
- For speech recordings, in most cases a sampling rate of 10k Hz is enough.

44100 Hz 
22050 Hz 
11025 Hz 
8000 Hz 
5000 Hz 

Sampling

- **Nyquist-Shannon theorem**:

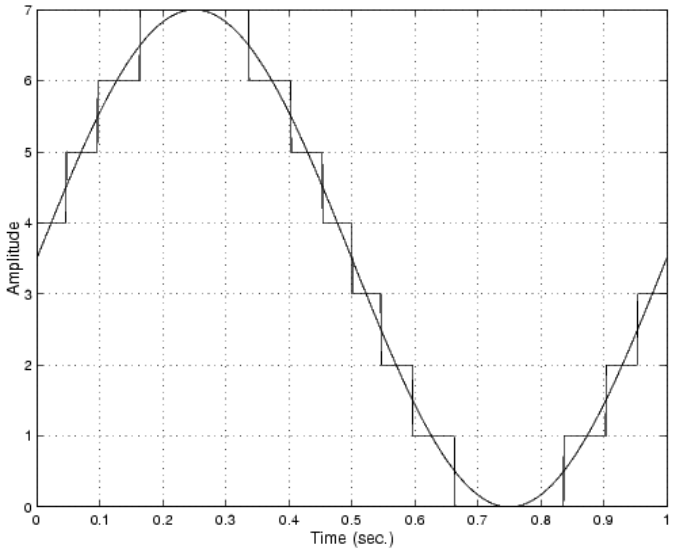
  When sampling a signal (e.g., converting from an analog signal to digital), the sampling frequency must be greater than twice the highest frequency in the input signal in order to be able to reconstruct the original perfectly from the sampled version.



- **Aliasing**: If the sampling frequency is less than twice the highest frequency component, then frequencies in the original signal that are above half the sampling rate will be "aliased" and will appear in the resulting signal as lower frequencies.
- **Anti-Aliasing filter**: typically a low-pass filter that is applied before sampling to ensure that no components with frequencies greater than half the sample frequency remain.

Quantization

- **Quantization:** Assigning a physical measurement to a binary number.
- **PCM (**Pulse-code modulation): the signal is sampled at **uniform** intervals, then quantized to a series of binary numbers.
- If eight bits are allowed for the PCM sample, this gives a total of 256 (2^8) possible values.



Audio file formats

- There are a number of different *types* of Audio files. The most common are Wave files (wav) and MPEG Layer-3 files (mp3).
- The way the audio is compressed and stored is called the *codec* which determines file size.

  ".wav" files are commonly used for storing uncompressed (PCM) sound files, which means that they can be large in size - around 10MB per minute of music.

  ".mp3" files use the "MPEG Layer-3" codec. "mp3" files are compressed to roughly one-tenth the size of an equivalent PCM file while maintaining good audio quality.

  ".aiff" is the standard audio file format used by Apple. It is like a wav file for the Mac.

  ".au" is the standard audio file format used by Sun, Unix and Java. The audio in au files can be PCM or compressed with the ulaw, alaw or G729 codecs.

Sound file formats

Type	Extensions	Codec
AIFF (Mac)	.aif, .aiff	*PCM
AU (Sun/Next)	.au	*u-law
CD audio (CDDA)	N/A	PCM
Midi	.mid	NA
MP3	.mp3	MPEG Audio Layer-II
Windows Media Audio	.wma	Proprietary (Microsoft)
QuickTime	.qt, .mov	Proprietary (Apple Computer)
RealAudio	.ra, ram	Proprietary (Real Networks)
WAV	.wav	*PCM

* Can be used with other codecs.