COMP47700
Speech and Audio

Dr Alessandro Ragano

Office: 3rd Floor, Science East, Insight Centre
School of Computer Science
University College Dublin

alessandro.ragano@ucd.ie

3.1.2
Speech Production: Characteristics of Speech

Characteristics of speech

General patterns and characteristics

- volume
- frequency distribution
- pitch rate
- syllabic rate

Differences in individuals and languages – accents (parts of Ireland), tempo (English vs. Italian), individuals (e.g. shy, brash), emotions, recognisable speakers

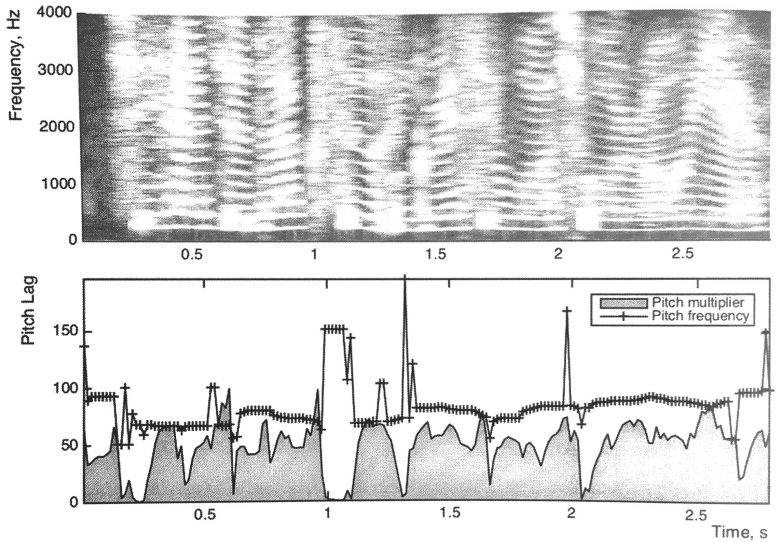
Speech as an encoding

vocal apparatus	medium	context
emotion	reverb	lexicon
urgency	noise	hearing ability
context		

Speech classification

Pitch contour

- It is effectively the fundamental vocal frequency variation with time
- fundamental frequency – denoted f0 (lower case f) – describes the tone of the voice
- The fundamental frequency is strongly related to the perceived frequency, even if it is not there in the spectrogram (more on this when we look at perception)



Pitch frequency varies over time (pitch contour) and amplitude for the spectrogram shown.

Pitch vs. Fundamental Frequency (F0)

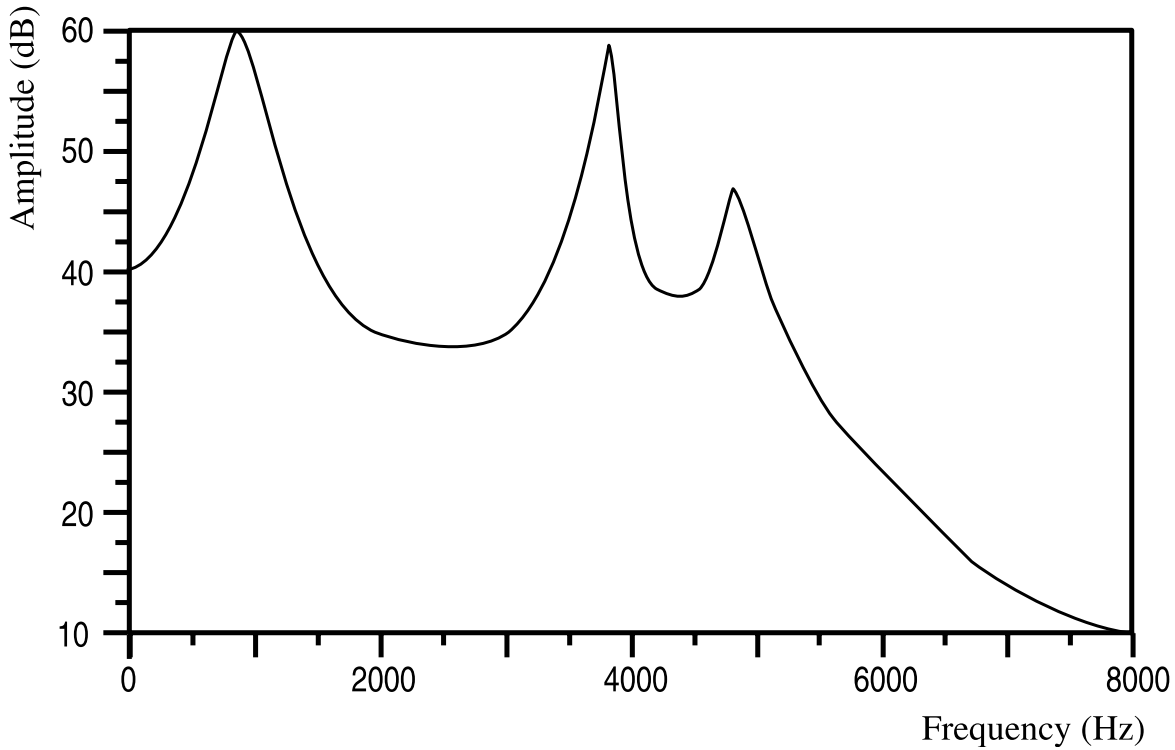
- **Fundamental Frequency (F0):**
  - *Physical*: Vocal fold vibration rate.
  - *Objective*: Measured in Hz.
  - *Acoustic*: Lowest periodic frequency.
- **Pitch:**
  - *Perceptual*: Highness/lowness.
  - *Subjective*: Listener-dependent.
  - *Psychoacoustic*: Not directly measurable.
- **Relationship:**
  - F0 **primary** determinant of pitch.
  - Strong, but not 1:1, correlation.
- **Other Factors:**
  - Intensity.
  - Timbre.
  - Duration.
  - Hearing.
  - Context
- **Key:** F0 = *Objective/Physical*; Pitch = *Subjective/Perceptual*

Speech classification

Formant frequencies

- Formants are resonant frequencies of the vocal tract which appear in the speech spectrum as clear peaks
- Formants are counted from the lowest frequency upwards, and usually only the first three (F1, F2 and F3)
- Contribute significantly to the intelligibility of speech
- F1 contains most of the speech energy while F2 and F3 between them contribute more to speech intelligibility

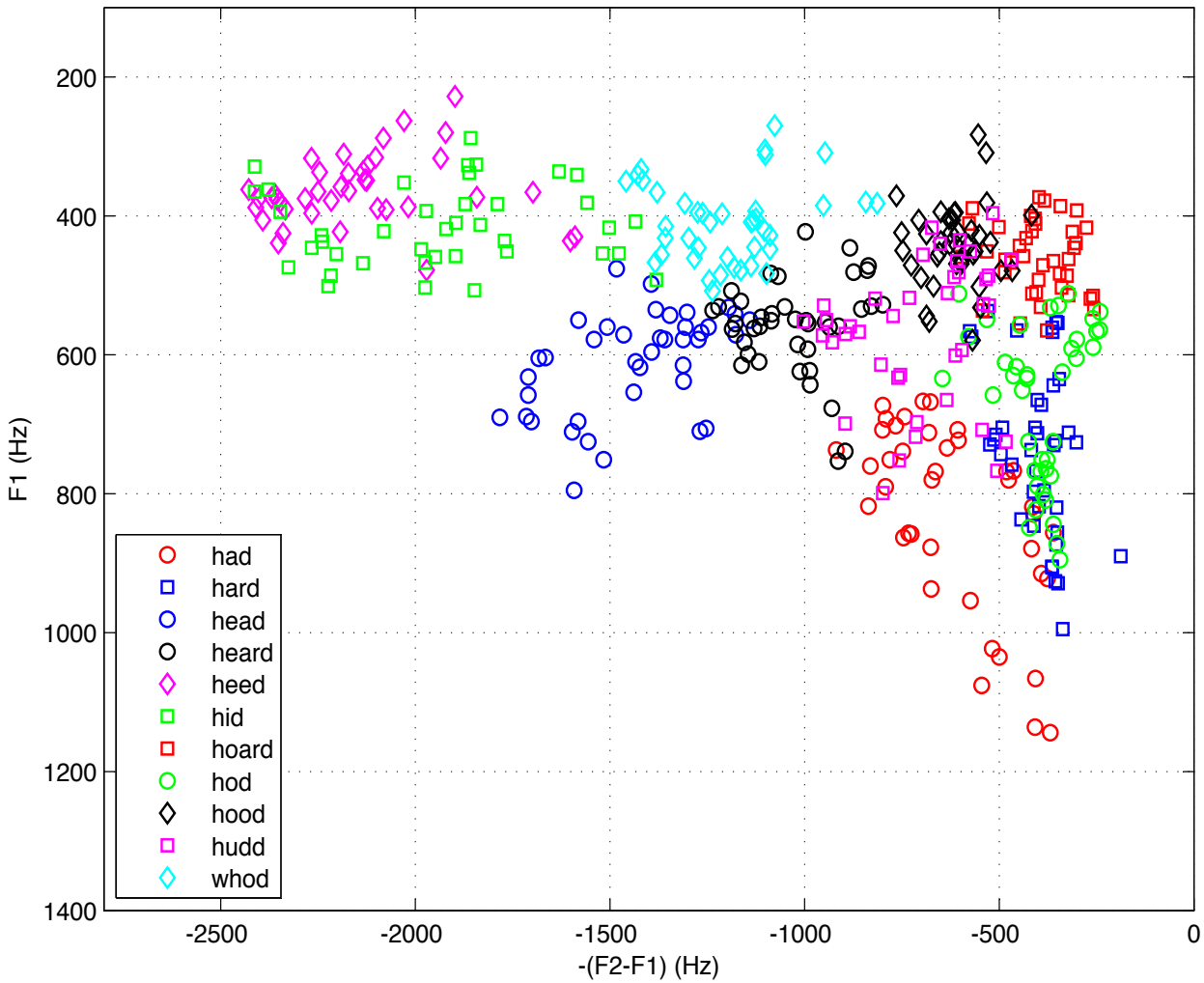
Speech Classification: Formant Frequencies



Spectrum plot of a 20 ms recording of voiced speech, showing three distinct formant peaks.

Vowel F1/F2 Scatter Plot

Plot of the first and second formant frequencies for various vowels.



Source: Sharifzadeh et al., 2010

Amplitude distribution of speech

The amplitude of speech is influenced by more than just the linguistic message it is encoding.

- Speaker personality
- Speaker mood
- Environmental noise
- Infection

Amplitude of speech in several environments

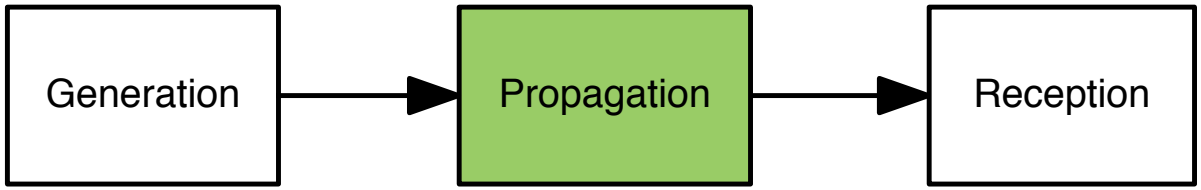


Table 3.1. Amplitude of speech in several environments, from [9].[a]

Location	Noise level (dB_{SPL})	Speech level (dB_{SPL})
school	50	71
home (outside, urban)	61	65
home (outside, suburban)	48	55
home (inside, urban)	48	57
home (inside, suburban)	41	55
department store	54	58
on a train	74	66
in an aircraft	79	68

Data Source: The Handbook of Hearing and the Effects of Noise, K. Kryter, Chapter 1, Copyright Elsevier (Academic Press) 1994.

Temporal Distribution

Physical Constraints

We are limited by the mechanics of the vocal system – breathing
Articulation speed is relatively independent of rate of speech – we just reduce the gaps
(This turns out to be very useful for speech analysis and synthesis!)
Native English speakers usually speak at 120 – 150 words per minute

Types of Speech

Whispering

Suitable for quiet situations but poor SNR in noise
Whispering is where all phonemes are unvoiced or unphonated.
Lack of glottal source as for pitch – substituted with a broadband excitation of air

Shouting

The **Lombard effect** or Lombard reflex is the involuntary tendency of speakers to increase their vocal effort when speaking in loud noise to enhance the audibility of their voice. This change includes not only loudness but also other acoustic features such as pitch, rate, and duration of syllables.
Recruitment: group speech adaptation to the environment