Carnegie Mellon University
Language Technologies Institute

11-324/11-624/11-724 Human Language for AI

Acoustic Phonetics I

David R. Mortensen

September 6, 2022

Language Technologies Institute
Carnegie Mellon University

Introduction

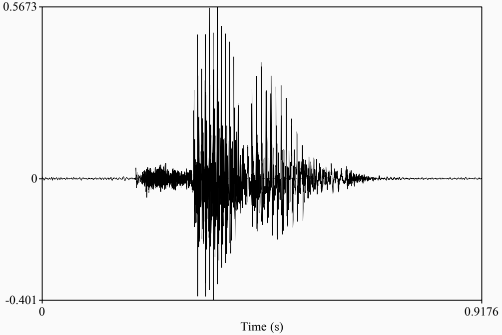
Introduction

At the end of this lecture, students will have learned:

- The basic elements of speech acoustics and digital processing of speech:
  - Complex waves and waveforms
  - Fourier analysis and spectra
  - Spectrograms (a visualization of spectra over time)
- The tube resonator model of speech acoustics
- The relationship between harmonics, resonances, and formants
- How to predict the frequencies of resonances from the length of the vocal tract
- How changes in the shape of the vocal tract affect the frequencies of resonances (perturbation theory)

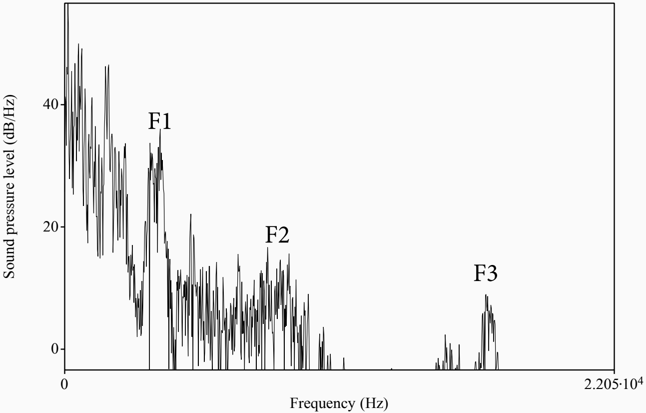
Representations of Speech Acoustics

Waveforms show sound pressure over time



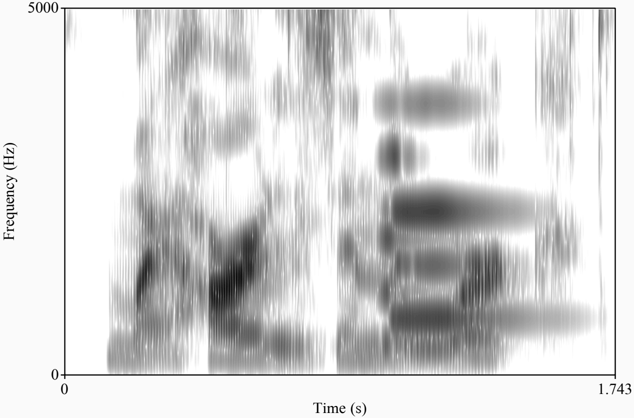
Sound pressure (measured in dB or decibels) over time (measured in seconds). Positive values indicate air “pushing” towards the hearer or microphone and negative values indicate air “pulling” away.

Spectra show sound pressure over frequency



Sound pressure varies over frequency. A spectrum shows this in a given slice of time.

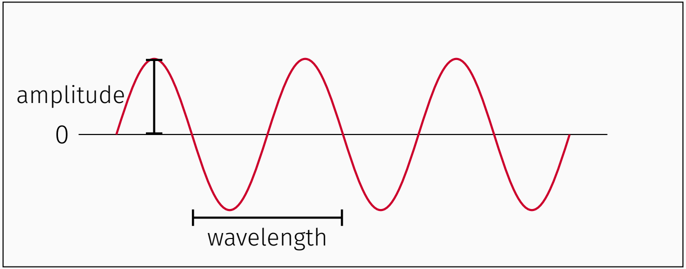
Spectrograms show sound pressure over time and frequency



Spectra are dynamic. A Spectrogram shows frequency (*x*) over time (*y*) over sound pressure (*z*)

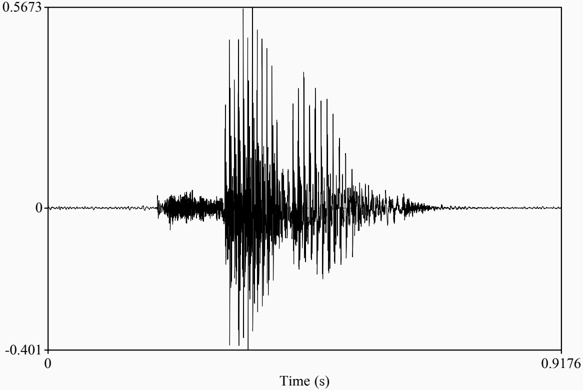
Complex Waves, Spectra, and Fourier Analysis

Sine waves are the simplest waves



These waves can be described by the sine function.

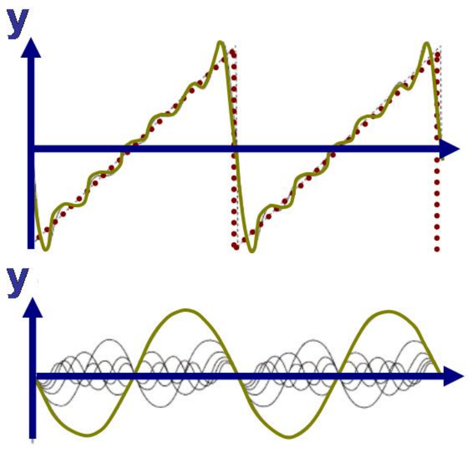
Complex waves are waves consisting over more than one component



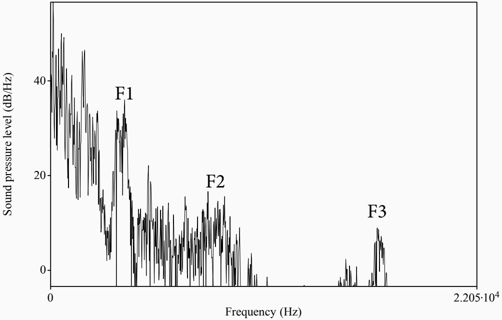
All waves but sine waves are complex waves.

Fourier analysis decomposes complex waves into a series of sine waves

Fourier analysis of a sawtooth wave

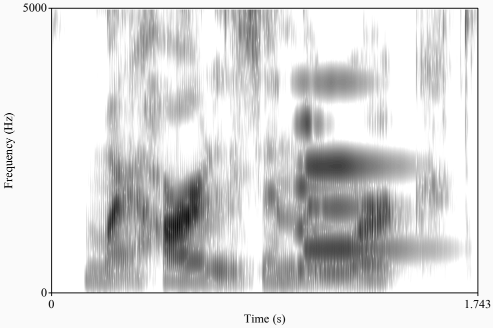


Spectra are visualizations of analyzed waves at single slices of time



Take a time-slice of a complex wave analyzed into sinusoidal components via Fourier analysis and plot the “average” amplitude (we’re approximating here) of each of the component waves according to their frequencies

Spectra over time are called spectrograms



- Time (seconds) is on the *x*-axis
- Frequency (Hz) is on the *y*-axis
- Amplitude (sound pressure) in on the *z*-axis, represented by the “darkness” of the pixels. “Darkness represents power!”

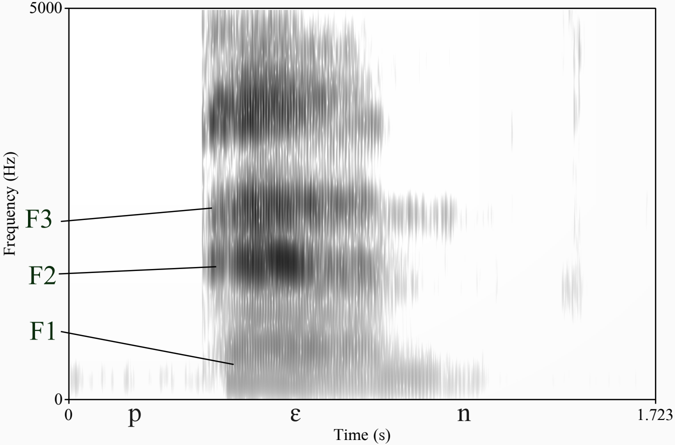
Harmonics, Resonances, and Formants

Some definitions of terms from acoustics

- **Harmonic:** One of the sinusoidal components of a complex wave; in speech, these are at whole-number ratios to the FUNDAMENTAL FREQUENCY
- **Fundamental frequency or F_0 or H_1:** Lowest component (harmonic) in a complex wave; in voiced speech, this is the rate at which the vocal folds are opening and closing
- **Resonance or resonant frequency:** A frequency at which, if energy is added, there will be a relatively large response; harmonics near resonances are relatively higher in amplitude for this reason
- **Formant:** a band of relatively high-amplitude harmonics (due to their proximity to a resonant frequency)

Formants are high-energy regions in a spectrum

Spectrogram of *pin* with the first three formants (F1, F2, and F3) labeled



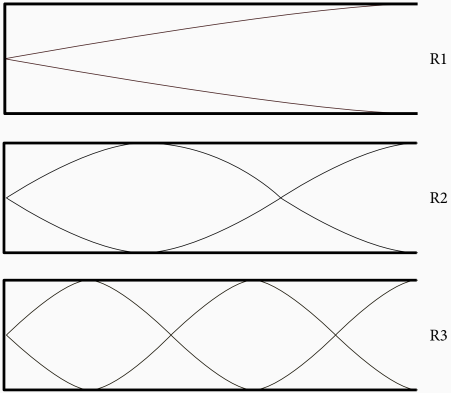
Frequency, wavelength, and velocity are mathematically related

$$f = \frac{c}{\lambda}$$

Where

- f = frequency in Hertz (Hz) (=cycles per second)
- c = velocity = 343 meters per second (the speed of sound)
- λ = wavelength (in meters)

The first three resonances of a closed tube



Why is this important? **Because, when the mouth is open, the vocal tract is essentially a tube closed on one end.**

Resonant frequencies can be calculated give the length of a tube

A tube closed at one end will have approximately the following series of resonances:

$$f = \frac{nc}{4L}$$

where

- f = frequency in Hertz (Hz) (=cycles per second)
- n = an odd integer (1, 3, 5, ...)
- c = velocity = 343 meters per second
- L = the length of the tube (in meters)

A more accurate formula is:

$$f = \frac{nc}{4(L + 0.4d)}$$

where d is the diameter of the tube.

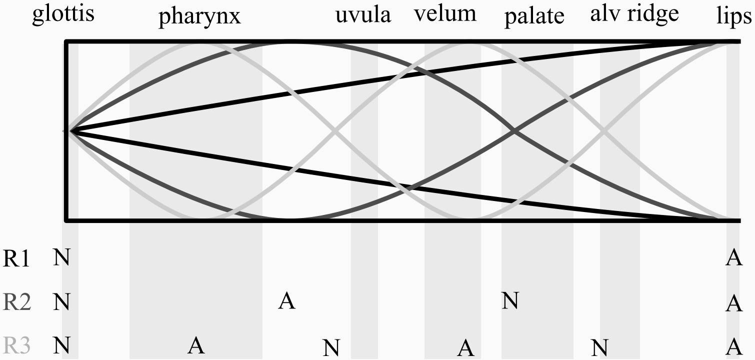
Perturbation Theory and Vowel Acoustics

Some more definitions

- **Standing wave:** is a wave which oscillates in time but whose peak amplitude profile does not move in space
- **Velocity node:** the points in a standing wave where the velocity of particles (etc.) is least (for example, at the closed end of a tube)
- **Velocity antinode:** the points in a standing wave where the velocity of particles (etc.) is the greatest (for example, at the open end of a tube)

The mapping between nodes/antinodes and anatomical landmarks allow for predictions regarding formants

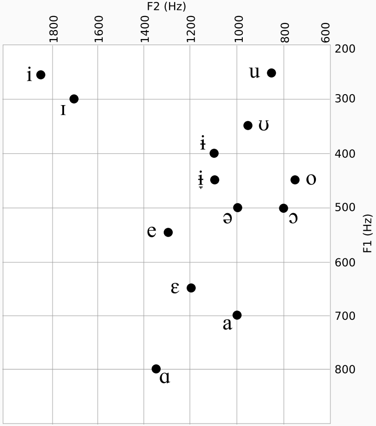
The relationship between velocity nodes and antinodes of the first three resonances and articulatory landmarks in the vocal tract



There are simple relationships between formants and vowel height

- The first formant (F1) is *inversely* correlated with vowel height
- The second formant (F2) is *inversely* correlated with vowel backness
- It is conventional to plot vowels so that high vowels are at the top, low vowels are at the bottom, front vowels are at the left, and back vowels are at the right

F2 × F1 vowel plots reveal the backness and height of vowels



Vowel plot for Welsh, a language of the United Kingdom

Steps for creating your vowel plots

Each of you will produce a vowel plot for your native language, like the previous plot for Welsh

1. Determine what and how many vowels your language has
2. Find carrier words for each of these vowels
3. Record at least 10 tokens of each of these carrier words
4. Annotate these recordings using Praat
5. Using the provided Praat script, extract formant measurements from these annotated recordings
6. Using a visualization tool of your choice, plot the vowels (following the standard conventions)

For the next lecture, we will have an exercise about using Praat.