

# 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](mailto:alessandro.ragano@ucd.ie)

### 3.1.3

Speech Production:

The Structure of Speech

# The Structure of Speech

## Phones

“The smallest discrete segment of sound in a stream of speech”  
(Crystal, 2009)

## Phonemes

Language specific building blocks of speech: concatenated to create words.

“The smallest phonetic unit in a language that is capable of conveying a distinction of meaning”

e.g. /t/ /b/ /d/ and /t/ in pad, pat, bad, bat

# Phonemes

A phoneme is generally regarded as an abstraction of a set (or equivalence class) of speech sounds (phones) which are perceived as equivalent to each other in a given language.

Phoneme /k/: **cat**, **kit**, **scat**, **skit**

An example is the English phoneme /k/, which occurs in words such as cat, kit, scat, skit. Although most native speakers do not notice this, in most English dialects the “c/k” sounds in these words are not identical: in **kit** the sound is aspirated, while in **skill** it is unaspirated

# Types of Speech

## Phoneme Amplitudes by Class

Phoneme class	Example	Amplitude (range), dB
vowel	<i>card</i>	26.0 (4.9)
glide	<i>luck</i>	21.6 (3.2)
nasal	<i>night</i>	17.1 (3.0)
affricative	<i>jack</i>	14.9 (2.6)
voiced fricative	<i>azure</i>	11.5 (2.2)
voiceless fricative	<i>ship</i>	10.0 (10.0)
voiced plosive	<i>bap</i>	9.6 (3.3)
voiceless plosive	<i>kick</i>	9.5 (3.3)

Average amplitude of phonemes by class, also showing amplitude range within each class, measured with respect to the quietest phoneme in English, the voiceless fricative /th/ in 'thought'.

# Phonemes

- A phoneme is the smallest structural unit of speech
- Written between slashes to distinguish them, e.g. /t/ at end of 'cat'.
- Phonemes often comprise distinctly recognisable phones which may vary widely to account for different spoken pronunciations.
- Two alternative pronunciations of a phoneme are usually the result of a choice between two phones that could be used within that phoneme.
- Alternative phone pairs termed allophones
- Phones which are identical except in their spoken tone, can be called allotones, (common in Mandarin Chinese)
- The International Phonetic Alphabet (IPA) is a way of encoding phonemes in writing

# Syllables

Single or clustered phonemes form units of sound organisation called syllables which generally allow a natural rhythm in speaking. Syllables usually contain some form of initial sound, followed by a nucleus and then a final. Both the initial and the final are optional, and if present are typically consonants, while the syllable nucleus is usually a vowel.

- A vowel is a sound spoken with an open vocal tract
- A consonant is one spoken with a constricted, or partially constricted vocal tract
- Definitions are unambiguous on paper but blurred substantially in practice

# Speech Articulation

## How does the vocal tract influence the sound produced?

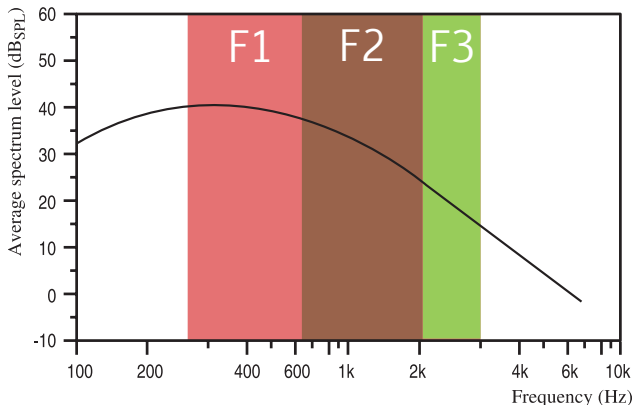
Speech sounds, both consonant and vowel, are defined by their place, or method of articulation within the vocal tract.

- **affricative** – a turbulent airflow fricative following an initial stop. E.g. /ch/ in 'chip'.
- **diphthong** – a two-part sound consisting of a vowel followed by a glide. E.g. /i//n/ in 'fine'.
- **fricative** – a very turbulent airflow due to a near closure of the vocal tract. E.g. /sh/ in 'ship'.
- **glide** – a vowel-like consonant spoken with almost unstricted vocal tract. E.g. /y/ in 'yacht'.
- **nasal** – a consonant spoken with velum lowered, so sound comes through the nasal cavity. E.g. /m/ in 'man'.
- **stop or plosive** – an explosive release of air upon rapid removal of a vocal tract closure. E.g. /p/ in 'pop'.



# Formant Bands

We saw formants as characteristics of speech with frequency and amplitude variations different phonemes.

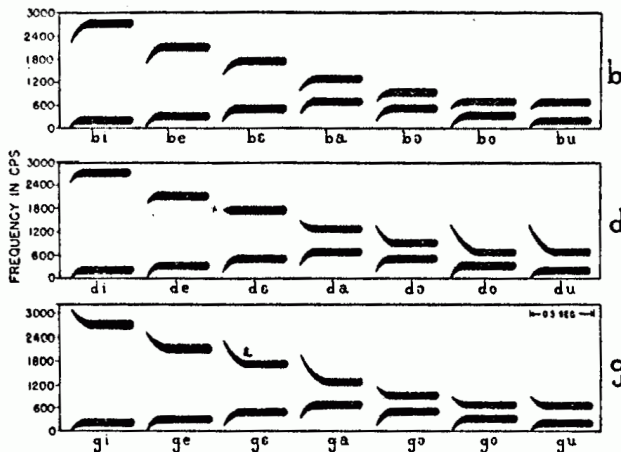


Long-time averaged speech power distribution plotted against frequency, with the approximate regions of the first three formants identified through vertical grey bands

# Formants and articulation

There is no one-to-one mapping between a given formant frequency and the place of articulation

Synthetic spectrograms showing second formant transitions that produce the voiced stops before various vowels



# Summary 3.1 Speech Production

## Speech Production

- Generation of speech: vocal apparatus
- Physical fundamentals of the speech signal (e.g. amplitude, frequency distribution, tempo)
- Units of speech: phonemes, syllables, words and sentences
- How units of speech convey information