The Speech Chain

The Human Vocal Organs

The vocal apparatus has evolved for …
– breathing
– eating
– vocalising
– speaking


Mammalian Vocalisation

Main source of energy
Airflow (determined by lung capacity & exhalation rate)

Main source of sound
Harmonic-rich ‘buzz’ (fundamental frequency determined by body size)

Main source of articulation
Vocal tract resonant frequencies (determined by body size & position of articulators)
The Human Vocal Tract


The Human Larynx

Vocal Fold

Epiglottis

Photo courtesy of Laryngograph Ltd.
Voice 'Source'

- Air pressure from the lungs builds up behind closed 'vocal folds' (often called 'vocal cords')
- The vocal folds are repeatedly forced apart and pulled together again, producing a series of small pulses of air
- This modulation of the airstream is known as 'phonation'
- The tension in the muscles attached to the vocal folds determines their rate of vibration and hence the 'fundamental frequency' ($F_X$ or $F_0$) of the speech waveform
- The fundamental frequency contributes to the perceived 'pitch' of the voice
- Because the vibration is not a pure sine wave, there is energy at frequencies that are multiples of the fundamental frequency – these are the 'pitch harmonics'

Two Cycles of Vocal Fold Vibration

[Image of vocal folds in vibration]
Air-Flow Through Glottis

The acoustic output from the vocal tract is initiated by the closure of the vocal folds.

More rounded corner at glottal opening

Closed phase

Sharp corner at glottal closure

Open phase

10 ms

0


Electroglottograph (EGG) / Laryngograph

COM3502/4502/6502 Speech Processing: Lecture 3, slide 9
modal voice - adult male

Sp  speech pressure microphone waveform

Lx  laryngograph/EGG vocal fold contact waveform

maximum contact
maximum separation

Analysing the Human Voice

Microphone placed on the larynx

Fundamental frequency
Pitch harmonics
Generating Pitch Pulses

Harmonic spectrum

Artificial Larynx / Electrolarynx

SolaTone™
The Human Vocal Tract

- The vocal tract forms a ‘**resonator**’ with a complex shape
- Resonances are known as ‘**formants**’
- Speech is produced by using the ‘**articulators**’ to change the shape of the vocal tract, hence modifying its resonant characteristics
- Different configurations of the vocal tract enhance some of the harmonics of the pitch, and suppress (**damp**) others
- The principal articulator is the **tongue**, but the jaw, lips, soft palate and teeth are also involved

The Excitation Spectrum

The Speech Spectrum

Pitch harmonics

'FORMANTS' (resonances)

F1
F2
F3
F4

Smoothed spectrum 'envelope'

Pitch harmonics 'illuminate' spectral shape of vocal tract

Fundamental frequency (pitch)
Vocal Tract Shape and Spectra

A Static Vocal Tract

Rhys Prosser

electrolarynx + 3D-printed tubes
The Moving Vocal Tract

Magnetic Resonance Imaging (MRI)

http://sail.usc.edu/span/index.php

The Moving Vocal Tract

https://youtu.be/yGV8az8np2U
The Moving Vocal Tract

[Diagram of the vocal tract]

https://dood.al/pinktrombone/

Analysing the Human Voice

Vocal tract resonances (‘formants’)

Microphone at the lips
Generating Formants

Formant 1
Formant 2
Sound Source

• Question …
  – is the larynx the only source of sound in the human vocal tract?

• Answer, no …
  – sound can be generated anywhere where there is a partial constriction (e.g. “sh”)
  – or by exciting a resonance (e.g. a whistle)
  – or by vibrating an articulator (e.g. the tongue)
  – or by releasing a blockage (e.g. the lips)

Types of Speech Sound

• A ‘voiced’ sound is one in which the vocal cords are vibrating

• An ‘unvoiced’ sound is one in which the vocal cords are not vibrating

• A ‘fricative’ sound results from a turbulent air flow at a constriction

• A ‘plosive’ sound occurs after a blockage is released
Sound Exercise

Which of the following English words end in a ‘voiced’ sound?

- bus  ❌
- breathe  ✔
- has  ✔
- off  ❌
- buzz  ✔
- breath  ❌
- rule  ✔
- of  ✔

Sound Exercise

Which of the following English words end in a ‘fricative’ sound?

- bus  ✔
- breathe  ✔
- hat  ❌
- off  ✔
- bun  ❌
- teeth  ✔
- rule  ❌
- of  ✔
Sound Exercise

Which of the following English words start with a ‘plosive’ sound?

- bus ✓
- give ✓
- has ✗
- sat ✗
- pull ✓
- teeth ✓
- rule ✗
- cough ✓

Part of a Speech Utterance

“briCKS FELL OVER”

Analysing the Human Voice

A 'voiced' sound

A 'fricative' sound

Point of glottal closure

Impulse response of the vocal tract

Frication (noise)

Generating All Speech Sounds

- The 'Holmes Parallel Formant Synthesiser'
- Capable of producing speech that is indistinguishable from the real thing (if properly controlled)
- 12 parameters (updated every 10 msecs):
  - FN: low-frequency formant (fixed at 250 Hz)
  - ALF: amplitude of low-frequency region
  - F1: frequency of 1st formant
  - A1: amplitude of 1st formant
  - F2: frequency of 2nd formant
  - A2: amplitude of 2nd formant
  - F3: frequency of 3rd formant
  - A3: amplitude of 3rd formant
  - AHF: amplitude of 4th formant (fixed at 3500 Hz)
  - V: degree of voicing
  - F0: fundamental frequency
  - MS: glottal pulse mark/space ratio (fixed)

John Holmes

The ‘Holmes’ Speech Synthesiser

This lecture has covered …

- The human vocal tract
- The larynx
- Generating pitch pulses
- Voice ‘source’ and voice ‘filter’
- The speech spectrum
- Resonances/formants
- Generating speech sounds
- Types of speech sound
- The ‘Holmes’ parallel formant synthesiser

Any Questions?
Next time …

Hearing