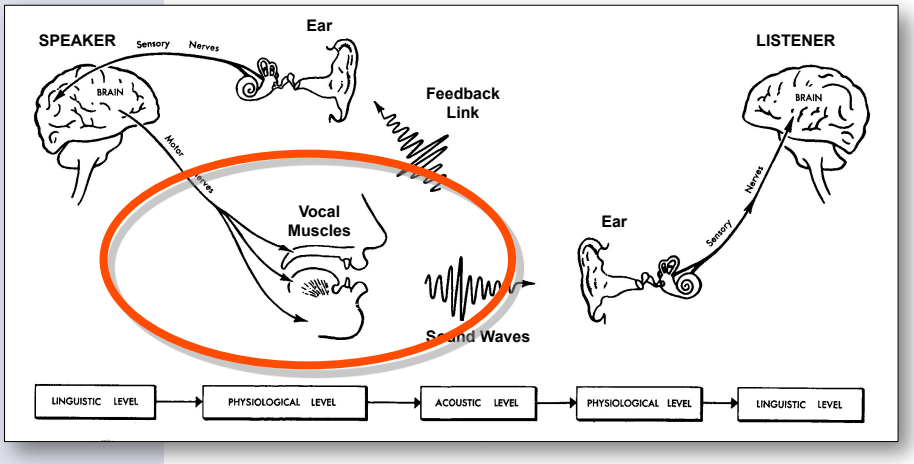
COM3502-4502-6502
SPEECH PROCESSING

Lecture 3
Speaking

The University Of Sheffield.

The Speech Chain



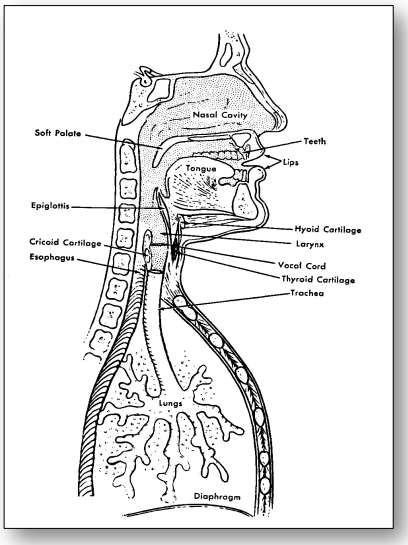
Taken from: Denes, P. B., & Pinson, E. N. (1973). ***The Speech Chain: The Physics and Biology of Spoken Language***: New York: Anchor Press.

The University Of Sheffield.

The Human Vocal Organs

The vocal apparatus has evolved for …

- breathing
- eating
- vocalising
- speaking



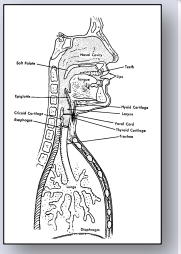
Taken from: Denes, P. B., & Pinson, E. N. (1973). *The Speech Chain: The Physics and Biology of Spoken Language*: New York: Anchor Press.

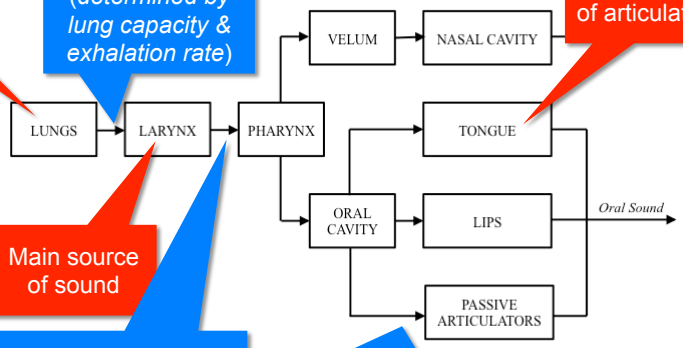
Mammalian Vocalisation

Main source of energy

Airflow (*determined by lung capacity & exhalation rate*)

Main source of articulation



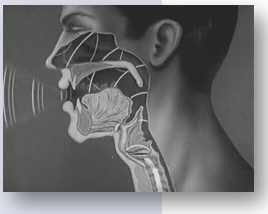


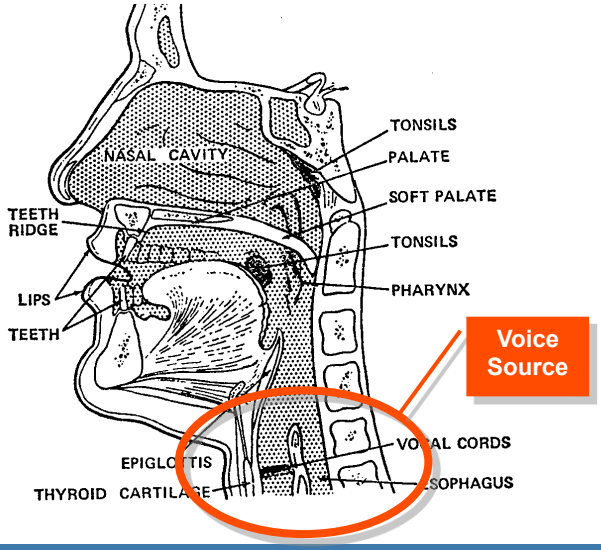
Main source of sound

Harmonic-rich 'buzz' (*fundamental frequency determined by body size*)

Vocal tract resonant frequencies (*determined by body size & position of articulators*)

The Human Vocal Tract





Taken from: Denes, P. B., & Pinson, E. N. (1973). *The Speech Chain: The Physics and Biology of Spoken Language*: New York: Anchor Press.

The University Of Sheffield.

The Human Larynx

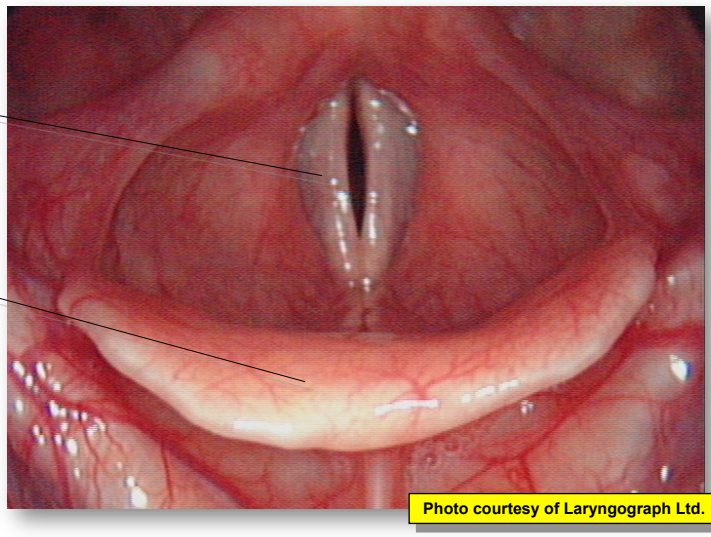
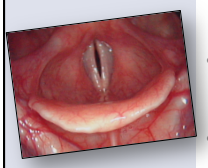


Photo courtesy of Laryngograph Ltd.

The University Of Sheffield.

Voice 'Source'

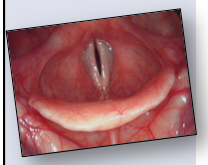


- Air pressure from the lungs builds up behind closed 'vocal folds' (often called 'vocal cords')
- The vocal folds are repeatedly forced apart and pulled together again, producing a series of small pulses of air
- This modulation of the airstream is known as 'phonation'
- The tension in the muscles attached 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 (*known as the 'pitch harmonics'*)

The University Of Sheffield.

Voice 'Source'

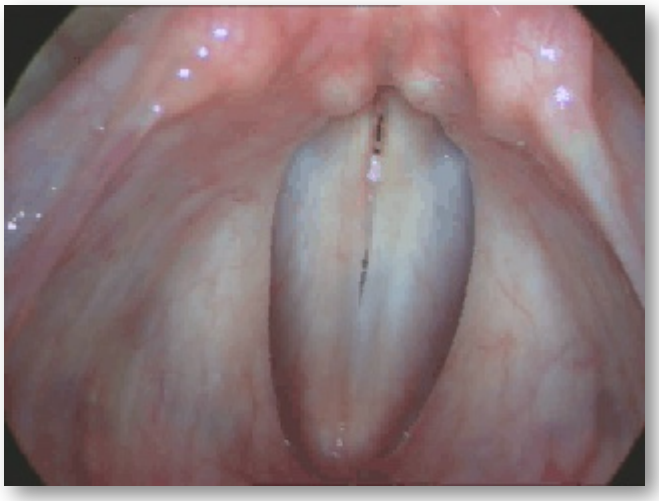


https://youtu.be/-XGds2GAvGQ

The University Of Sheffield.

Two Cycles of Vocal Fold Vibration

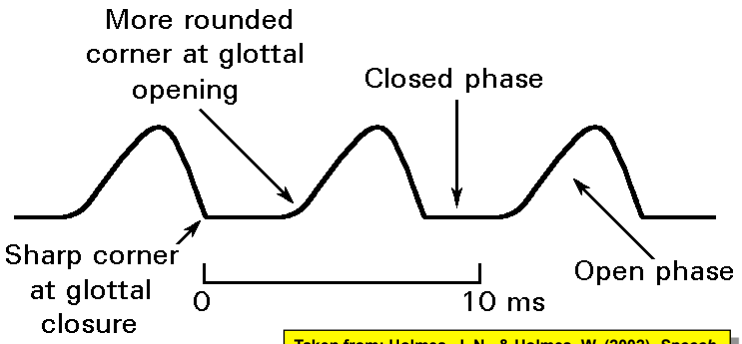
Laryngograph



The University Of Sheffield.

Air-Flow Through Glottis

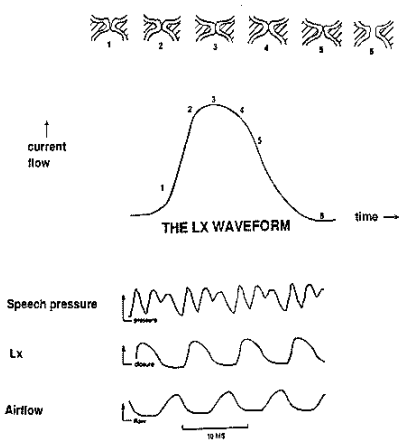
The acoustic output from the vocal tract is initiated by the closure of the vocal folds



Taken from: Holmes, J. N., & Holmes, W. (2002). *Speech Synthesis and Recognition*: Taylor & Francis.

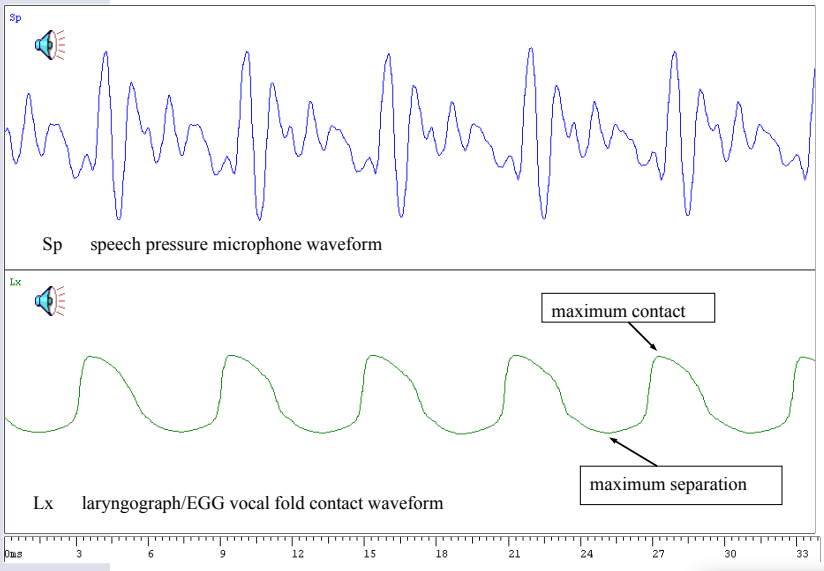
The University Of Sheffield.

Electroglottograph (EGG) / Laryngograph



Laryngograph

The University Of Sheffield.

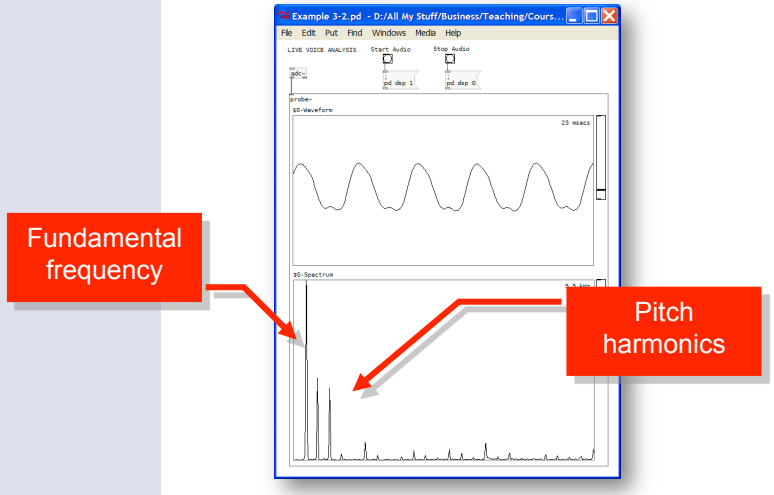


modal voice - adult male

Laryngograph

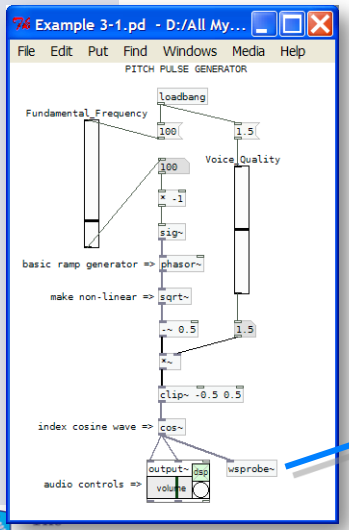
The University Of Sheffield.

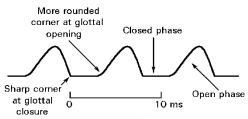
Analysing the Voice Source

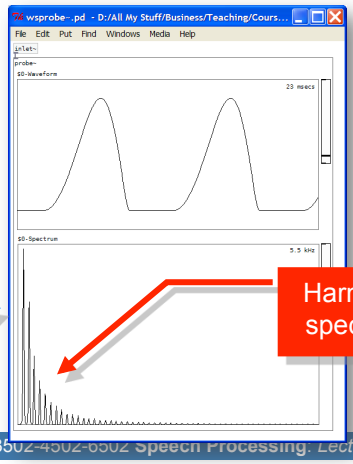


Microphone placed on the larynx

Synthesising the Voice Source



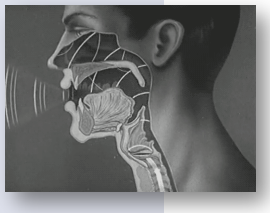




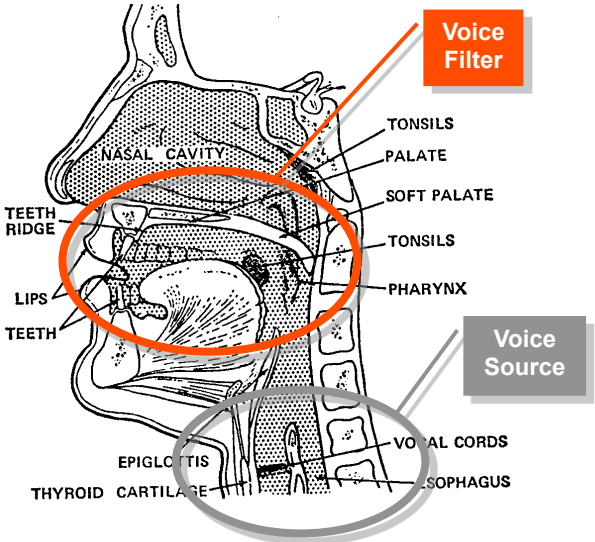
Artificial Larynx / Electrolarynx

SolaTone™

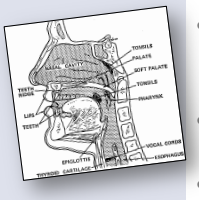
The Human Vocal Tract



Taken from: Denes, P. B., & Pinson, E. N. (1973). *The Speech Chain: The Physics and Biology of Spoken Language*: New York: Anchor Press.

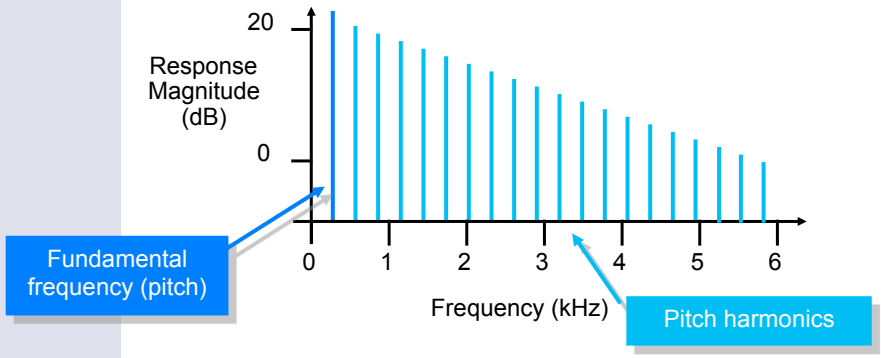


Voice 'Filter'

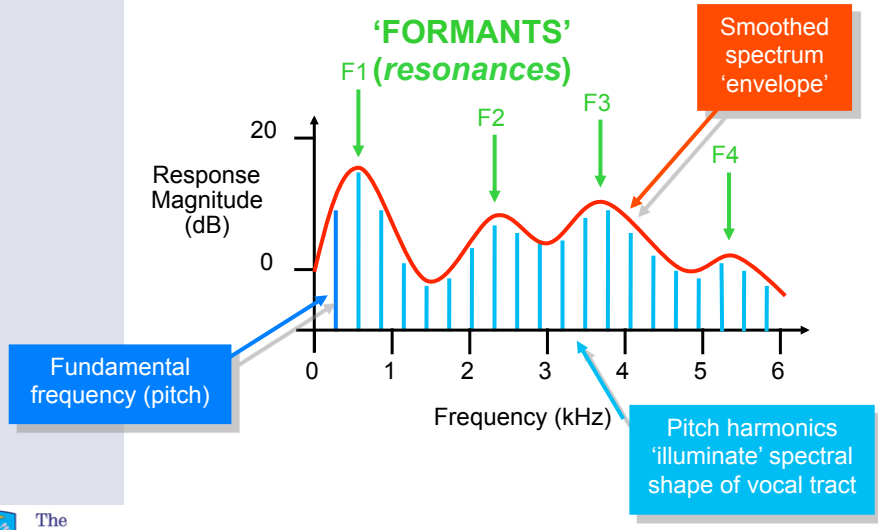


- The vocal tract forms a 'resonator' with a complex shape
- Resonances are known as 'formants'
- Speech is produced by using the 'articulators' to change the shape of the vocal tract, hence modifying its resonant characteristics
- Different configurations of the vocal tract enhance some of the harmonics 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

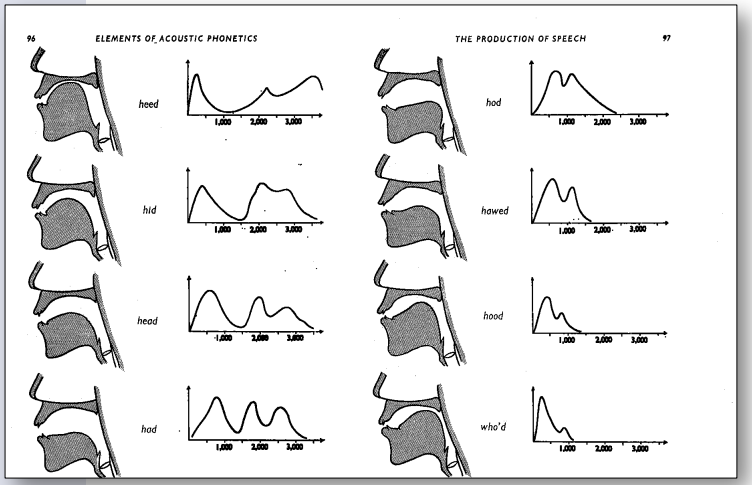


The Speech Spectrum



The University Of Sheffield.

Vocal Tract Shape and Spectra



Taken from: Ladefoged, P. (1962). *Elements of Acoustic Phonetics*: London: University of Chicago Press.

The University Of Sheffield.

A Static Vocal Tract

Rhys Prosser

electrolarynx

+

3D-printed tubes

The University Of Sheffield.

The Moving Vocal Tract

Magnetic Resonance Imaging (MRI)

University of Southern California

USC

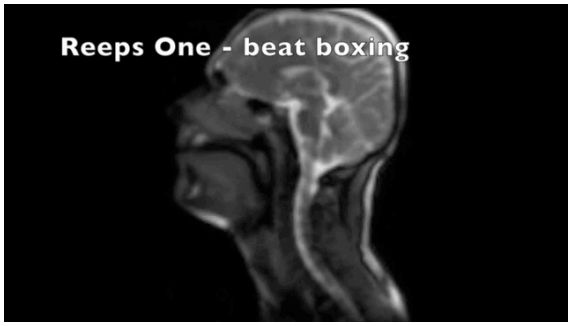
Speech Analysis and Interpretation Laboratory

USC Viterbi School of Engineering

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

The University Of Sheffield.

The Moving Vocal Tract

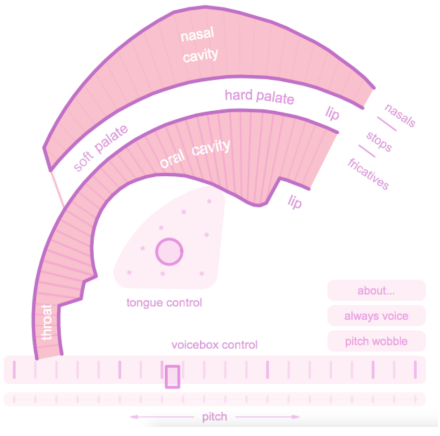


UCL

https://youtu.be/yGV8az8npZU

The University Of Sheffield.

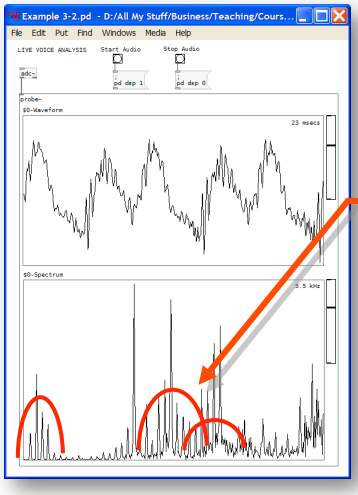
The Moving Vocal Tract



https://dood.al/pinktrombone/

The University Of Sheffield.

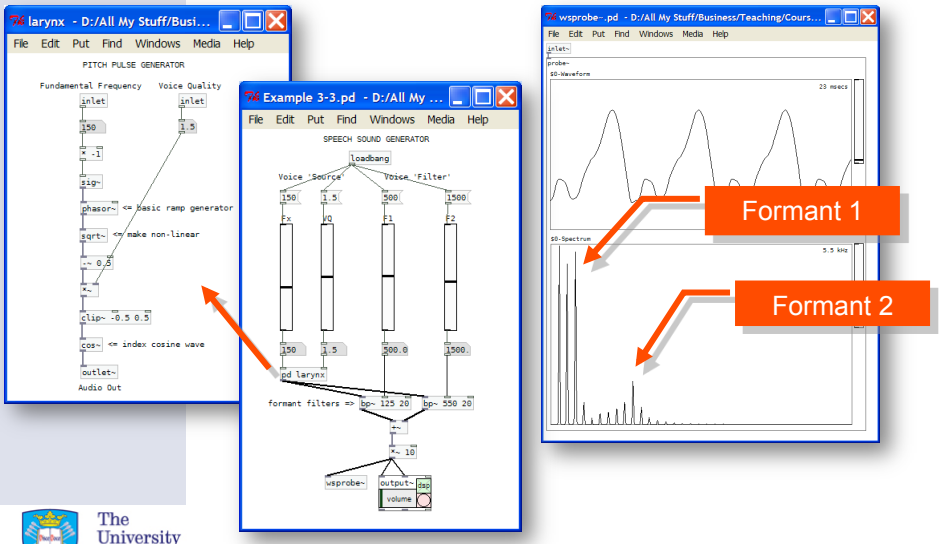
Analysing the Voice Filter



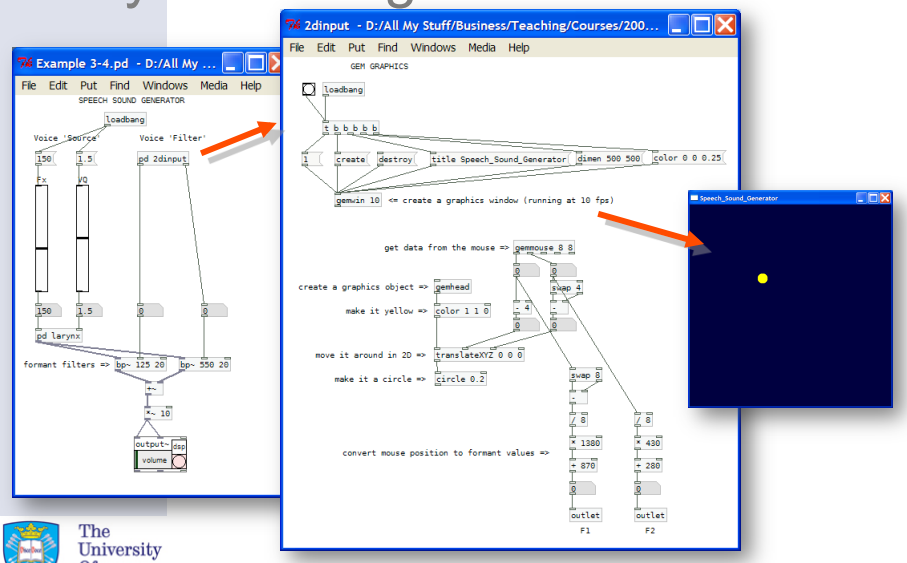
Vocal tract resonances ('formants')

Microphone at the lips

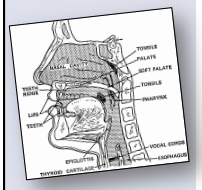
Synthesising the Voice Filter



Synthesising the Voice Filter

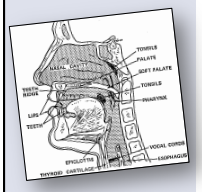


Sound Source



- 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)*

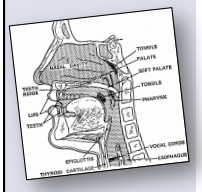
Types of Speech Sound



- 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

The University Of Sheffield.

Sound Exercise

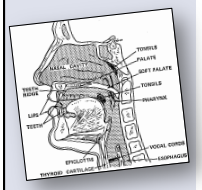


Which of the following English words end in a '**voiced**' sound ?

- bus ✗
- breathe ✓
- has ✓
- off ✗
- buzz ✓
- breath ✗
- rule ✓
- of ✓

The University Of Sheffield.

Sound Exercise

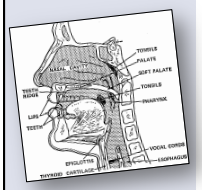


Which of the following English words end in a 'fricative' sound ?

- bus ✓
- breathe ✓
- hat ✗
- off ✓
- bun ✗
- teeth ✓
- rule ✗
- of ✓

The University Of Sheffield.

Sound Exercise

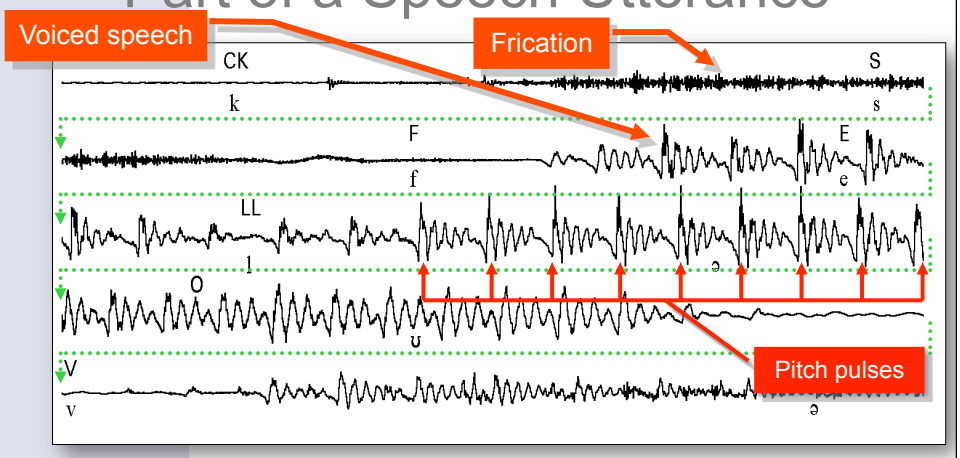


Which of the following English words start with a 'plosive' sound ?

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

The University Of Sheffield.

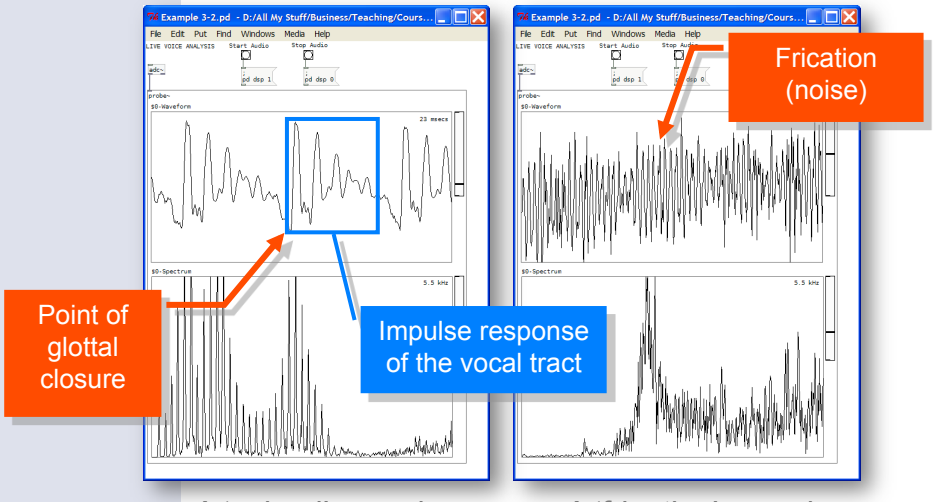
Part of a Speech Utterance



"briCKS FELL OVER"

Taken from: Holmes, J. N., & Holmes, W. (2002). *Speech Synthesis and Recognition*: Taylor & Francis.

Analysing the Human Voice

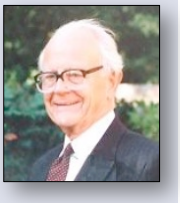


A 'voiced' sound

A 'fricative' sound

Synthesising the Human Voice

John Holmes

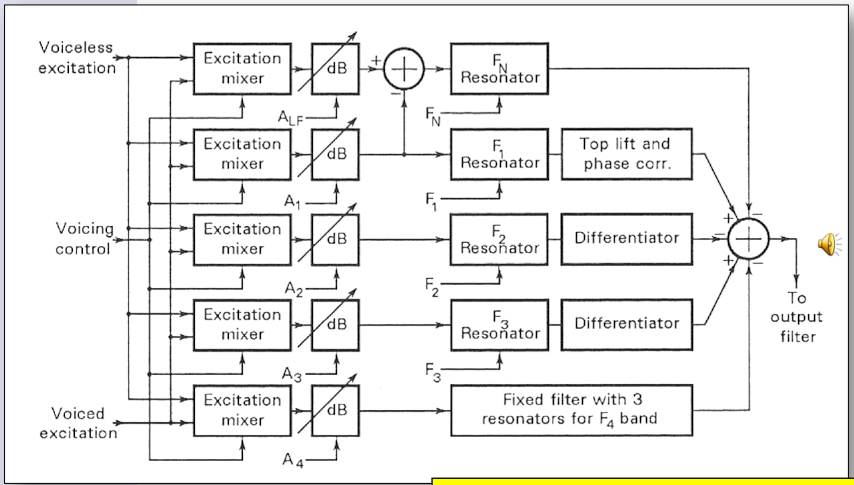


- The 'Holmes Parallel Formant Synthesiser'
- Capable of producing speech that is indistinguishable from the real thing (*if properly controlled*)
- 12 parameters (*updated every 10 msecs*):
  - FN low-frequency formant (*fixed at 250 Hz*)
  - ALF amplitude of low-frequency region
  - F1 frequency of 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)

Holmes, J. N. (1983). Formant synthesizers: cascade or parallel? *Speech Communication*, 2, 251-273.

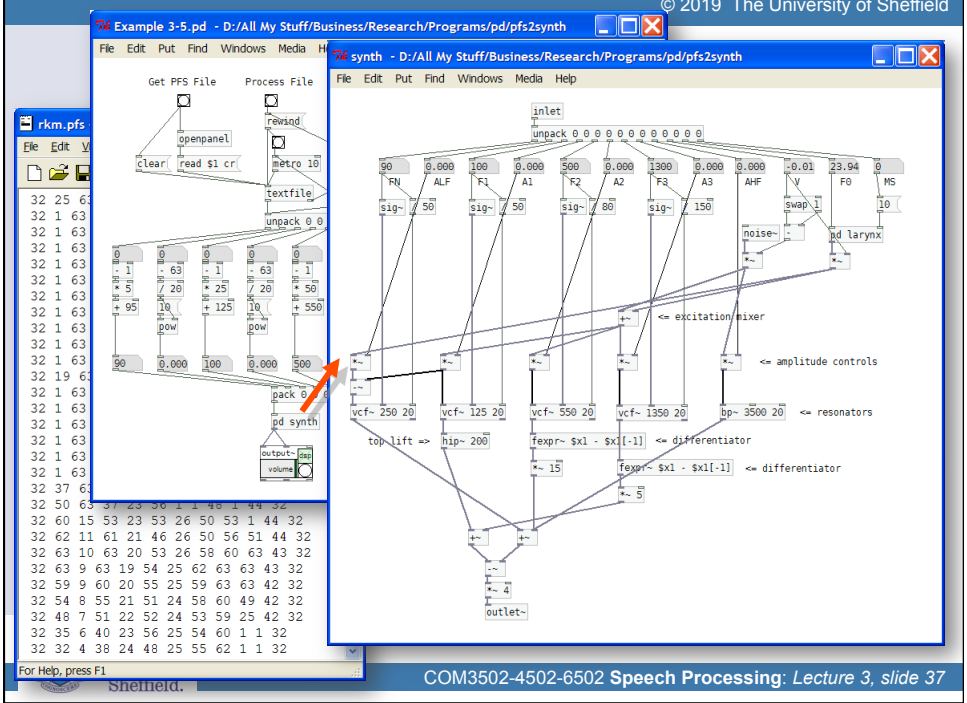
The University Of Sheffield.

The 'Holmes' Speech Synthesiser



Taken from: Holmes, J. N., & Holmes, W. (2002). *Speech Synthesis and Recognition*: Taylor & Francis.

The University Of Sheffield.



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