The Speech Chain Revisited

Acoustics of Speech and Hearing

Lecture 3-1

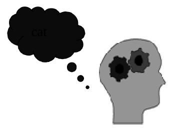
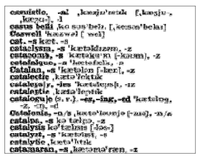
Aims of Lecture

- Remind you of topics we have covered on the course
- Show you how the topics fit together within the story of the speech chain
- Prompt you to write down questions to ask in the tutorial sessions later

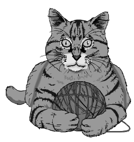
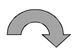
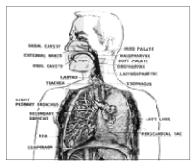
Overview

1. Phonology → Articulation
2. Articulation → Sound
3. Sound Transmission
4. Sound Analysis
5. Sound → Hearing
6. Hearing → Perception

1. Phonology → Articulation

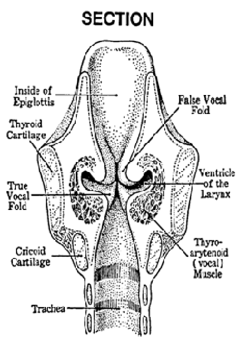
⇨ /kæt/

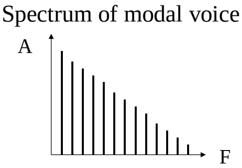
Synchronised control of articulators

- Phonological units
- Phonetic gestures
- Sound production
- Multiple articulators moving in continuous co-ordinated movement

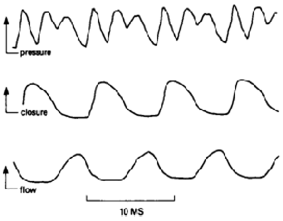
Larynx excitation



- Larynx structure
- Basic VF cycle
- Bernoulli effect
- Modal, Breathy, Creaky, Falsetto voice qualities
- Spectrum of modal voice

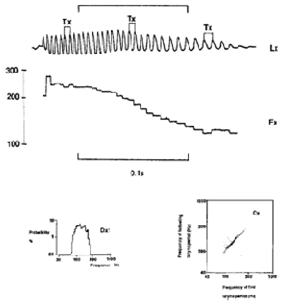


Larynx excitation



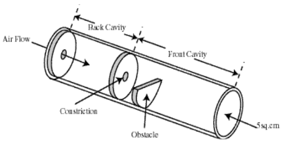
- Laryngograph operation
- Lx waveform shape
- Lx related to VF vibration and air-flow
- Measure jitter, shimmer, closed quotient and harmonics-to-noise ratio

Larynx excitation

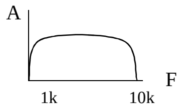


- Conversion of Lx to Tx period data
- Collect Tx data into Dx histograms, Cx scatterplot
- Measure mode, range, regularity

Fricative excitation



- Fricative noise generated by turbulence
- In constriction, when jet hits obstacle, mixing with stationary air
- Typically flat continuous spectrum



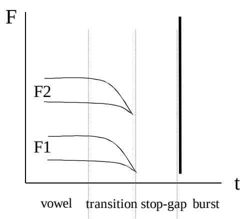
Continuants

- Steady vowels and fricatives
- “Target” shape for articulatory gesture
- Study output signal spectrum with source-filter model

Approximants and diphthongs

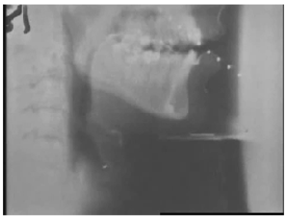
- Dynamic sounds requiring moving articulators
- Transitions to and from target shapes
- Need for spectrography to see spectral changes

Stops



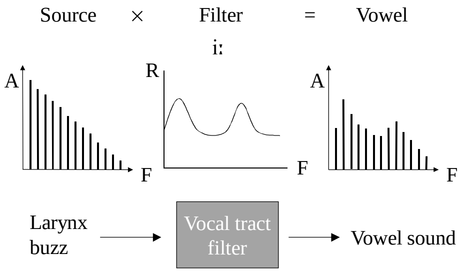
- Plosives and nasals
- Obstruction in vocal tract
- Closing and opening gestures
- Closing from a vowel, opening into a vowel
- Gesture affected by vowel environment

Dynamics

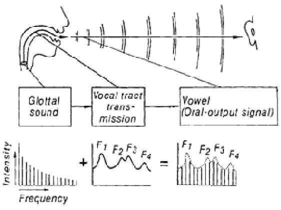


- Articulators constantly on the move
- Coarticulation: gesture for one phonological unit affected by gestures required for adjacent units
- Rapidity often causes undershoot
- Utterances have global rhythm and intonation

2. Articulation → Sound

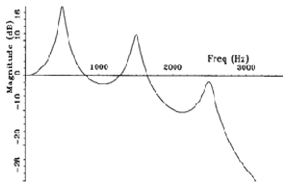


Source-filter model



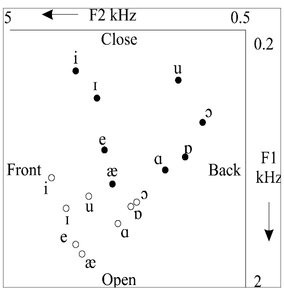
- Source = signal → source of sound
- Filter = system → changes sound
- Output = signal → output sound = Source × Filter
- Source-Filter model: assume source and filter are independent

Formant description of vocal tract



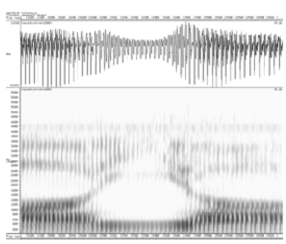
- Unobstructed vocal tracts have a frequency response made up from a small number of simple resonances called formants
- We can characterise the overall response by just measuring the frequencies and bandwidths of these formants

Formant description of vowels



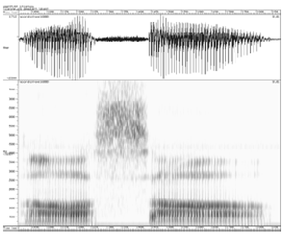
- Most variability in vowels found in F1 and F2
- F1-F2 plot looks a lot like vowel quadrilateral
- F1 associated with openness
- F2 associated with frontness
- Absolute frequencies depend on vocal tract length

Formant description of diphthongs and approximants



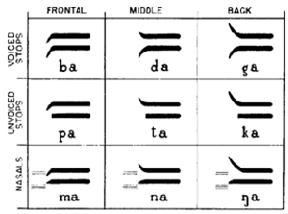
- Smooth movement of articulators is translated into smooth movement of formant frequencies

Source-filter model of fricatives



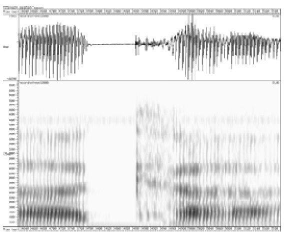
- Flat spectrum source
- Sibilant fricatives have air jet causing turbulence at teeth
- Most shaping done by front cavity
- More front: smaller cavity but increased damping

Formant transitions into/out-of obstruction



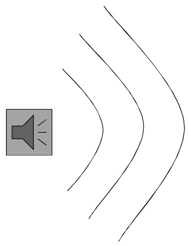
- Formant transitions caused by movement of articulators into/out-of obstruction
- F1 signals manner
- F2 & F3 signal place
- Voice onset time is an important cue to voicing in plosives

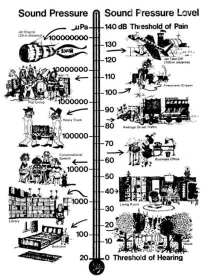
Redundant coding



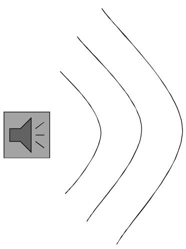
- Each articulatory gesture causes multiple changes to signal
- So each phonetic contrast cued by multiple features
- Place: F2, F3, burst frequency
- Voice: VOT, aspiration, F1 cut-back
- Multiple cues provide robustness through redundancy

3. Sound transmission



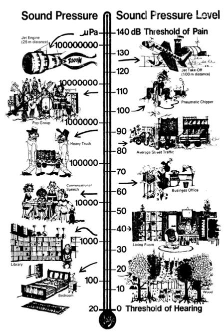


Sound as pressure waves in air



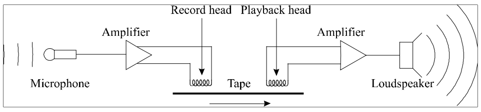
- Sound sources cause localised changes in air-pressure
- Changes propagate out as longitudinal pressure waves
- Ear detects small changes in air pressure
- Between 20μPa and 200Pa

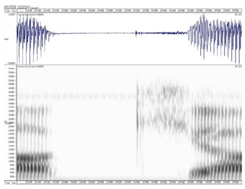
Measurement of intensity



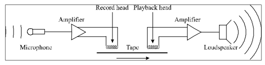
- Very wide range of pressure changes perceived as sound
- decibel scale is logarithmic ratio scale
- dBSPL scale based at threshold of audibility

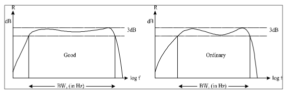
4. Sound Analysis





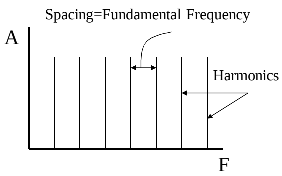
Recording





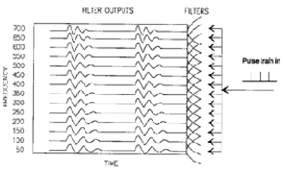
- Types of audio recorder
- Measures of recording quality: frequency response, SNR, harmonic distortion
- Check-list for good practice in making recordings

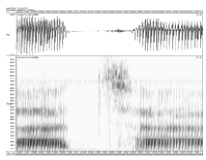
Spectral analysis



- Fourier analysis decomposes complex periodic signal into sum of sinewaves
- Sinewaves at whole-number multiple of repetition frequency = harmonics
- Phase unimportant

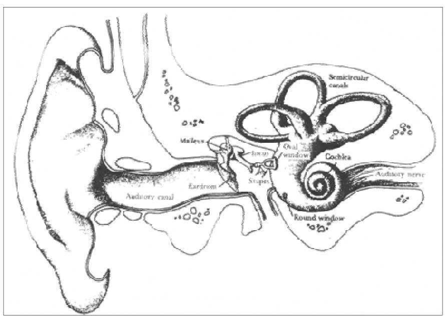
Spectrography



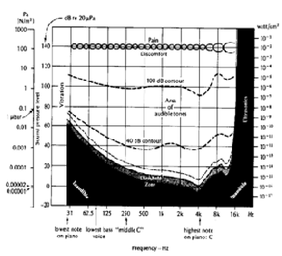


- Three-dimensional analysis along amplitude, frequency and time using wide or narrow bandpass filters
- Amplitude compressed to grey scale
- Shows how spectral content of speech signal changes with time

5. Sound $\rightarrow$ Hearing

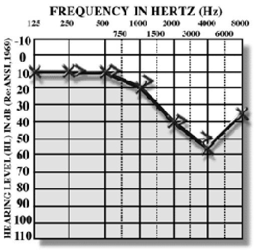


Intensity



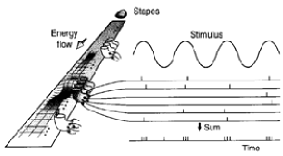
- Threshold of audibility
- Threshold of pain
- Outer ear and middle ear boost mid-frequency sounds, so more sensitive at 1-4kHz
- Loudness changes non-linearly with pressure, more like decibels

Pure-tone Audiometry



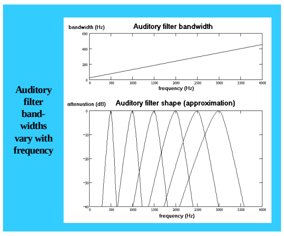
- Measure intensity of "just audible" pure tones at frequencies 125-8000Hz
- Compare thresholds in dBSPL to average normal thresholds
- Hearing level (dBHL) is difference to normal
- Plotted on audiogram

Frequency



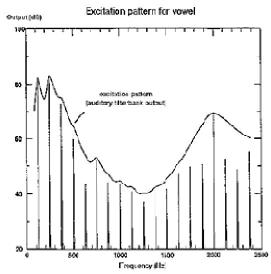
- Basilar membrane sorts sound components according to frequency = place coding
- Nerve firing synchronised to phase of stimulus cycle (up to 5kHz) = temporal coding

Auditory Filters



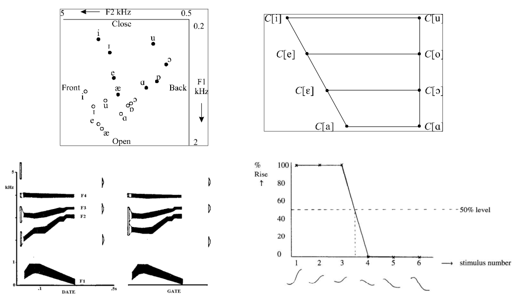
- Sound components that fall close together on the BM cannot be separately identified
- Better to view BM as small number of band-pass filters providing spectral envelope information
- Low-frequency: narrow bandwidth
- High-frequency: wide bandwidth

Complex sounds

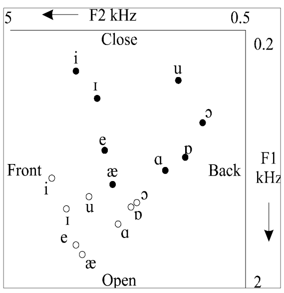


- Low frequency harmonics are resolved – important for pitch
- At high frequency only spectral envelope available – important for formant frequency estimation

6. Hearing → Perception

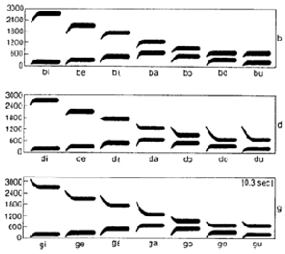


Vowels



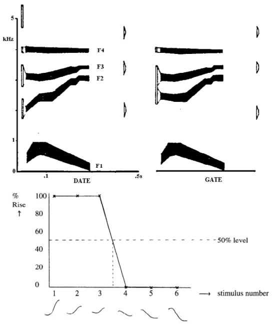
- Normalisation is process by which variable physical form mapped onto more constant perceptual labels
- E.g. vowels across vocal tract sizes; or use of Fx to signal pitch accents

Acoustic cues



- Single phonological units can be cued by variety of acoustic signals.
- Look for consistent patterning in the spectral features rather than absolute frequency
- E.g. formant transitions cueing plosive place have common locus

Perception tests



- Determine which spectral features are perceptually important using listening tests
- Phoneme confusion test
  - Look at confusion matrix patterns
- Phoneme contrast test
  - Labelling a stimulus continuum
  - Measure phoneme boundary and steepness

That's all folks!

The End

Acoustics Revision Sessions

- 09.00-10.00 Review Lecture (118)
- 10.15-11.00 Session 1 (116,101,Lab)
- 11.00-11.45 Session 2 (116,101,Lab)
- 11.45-12.30 Session 3 (116,101,Lab)
- 12.30-13.00 General/Exam technique (118)

Revision Sessions

- Topics (3 x 45min)
  - Signals & Systems
  - Speech Acoustics
  - Hearing & Spectrography
- Roughly follow Summary sheet
- Rooms
  - 116, 101, Lab
- No more than 13 in each room, please