Acoustic Phonetics – Part 2
In connected speech, many of the sounds are more difficult to distinguish.

Transcribe the segments in the following phrase “She came back and started again.” (American English)
I should have thought spectrograms were unreadable. (British English)

We first find obvious things first, i.e. [s, ʃ] which stand out.

Start at the beginning, and find the vowel [aɪ] in the first word.

The vowel in thought before [s]. And then the [t] in thought.

It seems as if the whole of the phrase “should have” was pronounced without any voicing: [aɪʃtʃ θɔt]

air ʃ t f θ o t s
I should have thought spectrograms were unreadable.

Try to transcribe spectrograms were unreadable, remembering that some of the sounds you might have expected to be voiced might be voiceless.

No aspiration after [p].

[ɛ] is very short but you can see the coming together of F2 and F3 for the [k].

[t] is highly aspirated, so the following [r] becomes voiceless. Same with [ə].
I should have thought spectrograms were unreadable.

The velar stop [g] is released into an [ɪ] located by the lowering of F3 and F4.

The fricative after the [m] appears to be voiceless and of less intensity than a [s].

The [w] is distinguishable by the low F2 of the following vowel.

The lowering of F3 marks the [r] in were.
I should have thought spectrograms were unreadable.
The lowering of F3 marks the [r] in unreadable.
[d] and [ə] are very short.
The final syllabic /l/ looks like a back vowel.
English sentence spoken by a British English speaker.
Try to identify segments in the sentence.
What do you observe at 14-15;
What must be there when the third formant is below 2000 Hz?
Can you discern a distinctive pattern of F2 and F3 at (26) and (24-25)?
**INTERPRETING SPECTROGRAMS (III)**

- (1): Small fricative noise near 3000 Hz.
- (2): A vowel that may be [i] or [ɪ].
- (3-4): A sharp break in the pattern and faint formants at about 250, 1300, and 2400 Hz → nasal or lateral
- (5): This vowel looks like [æ] or [ɛ].
- (6): Fricative of low energy, [f] or [θ]
(7): voiceless stop: [p], [t] or [k]
(8): Aspiration strong at high frequency, most likely a [t].
(9): The vowel has a low F1 and a high F2, so it’s either [i] or [ɪ].
(10): F2 falls slightly → diphthongization
INTERPRETING SPECTROGRAMS (III)

(1) (2) (3) (4) (5) (6) (7) (8) (9) (10)
?

æ θ

ε

he laugh/left

here
(13-14): Fricative [f] ħ [θ].

(15): Low F3, indicating [i].

(16-17): Vowel with low F1 and high F2 → [i]

(17-18): Voicing near the baseline and intense, high frequency burst → [d]

(20-21): Long, high and front vowel (diphthong) → [ɛi].

(23): Fricative like [s], but due to lack of intensity [z] with faint voicing.


(27-29): Long vowel (diphthong) ending in back low vowel → [u]
“Is Pat sad or mad?”

Types of Spectrograms

- wide-band spectrograms
- narrow-band spectrograms
TYPES OF SPECTROGRAMS

Wide-band spectrograms
✓ Very accurate in the time dimension
  + They show each vibration of the vocal folds as a separate vertical line.
  + They indicate the precise moment of a stop burst with a vertical spike.
✓ Less accurate in the frequency dimension
  + There are usually several component frequencies present in a single formant, all of them lumped together in one wide band on the spectrogram.

Narrow-band spectrograms
✓ More accurate in the frequency dimension (at the expense of accuracy in the time dimension).
  + The spikes of stop releases are smeared in the time dimension in the narrow-band spectrogram.
  + The frequencies that compose each formant are visible.
Women’s voices usually have a higher pitch.
The higher the F0 the more difficult it is to locate formants, because the harmonics interfere with the display of formants.

Greek phrase uttered by a male and a female Greek adult.
Λέγε «παππού» πάλι. (Say “grandfather” again)
It is important to know what sort of differences exist between different speakers.

1. When trying to measure features that are linguistically significant, one must know how to discount purely individual features.
2. When trying to find out whether a speaker has speech problems.
3. For valid speaker identification in forensic situations.

Individual variation is readily apparent when studying spectrograms \(\Rightarrow\) relative quality
7. INDIVIDUAL DIFFERENCES

- Same phonetic quality
  - Similar relative positions
  - Different absolute values

Vowels pronounced by 2 speakers of Californian English.
7. INDIVIDUAL DIFFERENCES

- No simple technique to average out individual characteristics so that a formant plot shows only the phonetic qualities of vowels.
- **F4** indicator of individual’s head size
  - Express values of other formants as percentages of the mean F4.
  - F4 values are not usually reported.
- Phoneticians do not really know how to compare acoustic data on the sounds of one individual with those of another.
- We cannot write a computer program that will accept any individual’s vowels as input and then output a narrow phonetic transcription.
A large part of applied phonetics work is concerned with computer speech technology directed towards improving speech synthesis systems.

The greatest challenges in the field of speech synthesis concern intonation and rhythm.
- Stereotyped intonation → unnatural speech
- To get correct pitch changes/rhythm
  - Speaker’s attitude towards world & specific topic
  - Emphasis
  - Syntax of the utterance
  - Higher level pragmatic considerations
  - Segmental influences
9. SPEECH RECOGNITION

- Systems can recognize
  - Single words
  - Limited sets of words in task specific situations with structured dialogue → limited set of possible answers

- Yet to achieve...
  - Accurate written transcript of ordinary speech as spoken by people with a wide range of accents and different personal characteristics
Speaker identification in legal proceedings.

**Voice-prints:** spectrograms of a person’s voice
- Said to be as individual as fingerprints
- Greatly exaggerated claim
- Some individual characteristics are recorded on spectrograms.

**Individual characteristics on spectrograms**
- Position of F4 and higher formants → speaker’s voice quality
- Locations of higher formants in nasals → individual physiological characteristics
- Speaker’s speech habits
  - Length and type of aspiration after initial voiceless stops
  - Rate of formant transition after voiced stops
- Mean pitch
- Range of F0
Nobody knows how many individuals share similar characteristics.

An expert’s opinion on the probability of two voices being the same has evidential value.

No two cases (recordings) are ever the same

- Recording quality
- Recording duration
- Word content
- Speech style (natural, emotional etc.)

Elaborate prior testing is needed.

Likelihood ratio: \[
\frac{\text{Likelihood voices are the same}}{\text{Likelihood voices are different}}
\]

Visit: Forensic Speech Science University of York
https://sites.google.com/site/yorkfss/home
Visit the websites:

- [https://corpus.linguistics.berkeley.edu/acip/](https://corpus.linguistics.berkeley.edu/acip/)
  Material for chapter 8 from UC Berkley Linguistics, “A course in phonetics” including online exercises

- [http://home.cc.umanitoba.ca/~robh/howto.html](http://home.cc.umanitoba.ca/~robh/howto.html)
  Monthly Mystery Spectrogram Webzone - Rob Hagiwara's professional web-space

- [http://www.youtube.com/watch?v=Gg4IHIbITd0](http://www.youtube.com/watch?v=Gg4IHIbITd0)
  Introduction to spectrogram analysis (FloridaLinguistics.com)

- [http://www.linguistics.ucla.edu/people/hayes/103/SpectrogramReading/Index.htm](http://www.linguistics.ucla.edu/people/hayes/103/SpectrogramReading/Index.htm)
  Spectrogram reading practice (by Bruce Hayes, UCLA)

  Text-to-Speech synthesis Avatars
Put a transcription of the segments in the phrase *Please pass me my book* above the waveform. Draw lines showing the boundaries between the segments.
EXERCISE B

- The spectrogram shows the phrase *Show me a spotted hyena*.
- Put a transcription above it, and show the segment boundaries. In places there are no clear boundaries (as in the first part of *hyena*), draw dashed lines.