Acoustic Phonetics – Part 1

Lecturer: Dr Anna Sfakianaki

HY578 Digital Speech Signal Processing
Spring Term 2016-17

CSD, University of Crete
QUESTIONS

- Which are the acoustic properties of speech?
- How do we “read” spectrograms?
1. Formants

- Sounds differ from each other in three ways
  - pitch
  - loudness/intensity
  - quality
- A vowel sound contains a number of different pitches simultaneously
  - pitch at which it was spoken
  - various overtone pitches that give it its distinctive quality

- Vowel Quality ↔ Overtone Structure
- **Overtones** = **Formants**
- The lowest 3 formants distinguish vowels from each other
  - F1
  - F2
  - F3
1.1 How do formants arise?

- The air in the vocal tract acts like the air in a bottle.
  - Tap on a bottle.
  - Open your mouth, make a glottal stop and flick a finger against your neck just to the side and below the jaw.
What do you observe?

- Articulate [i, e, a, o, u] without producing sound.
What do you observe?

Pitch of F1 going up for [i, e] and down for [a, o, u]
Formants that characterize different vowels are the result of the different shapes of the vocal tract.

Any body of air will vibrate in a way that depends on its size and shape.

- Blow across the top of
  - an empty bottle
  - partially filled bottle

What do you observe?

Great volume of air → low-pitched note
Small volume of air → high-pitched note

Adapted from Fant (1960)
1) Vocal folds open and close sending out pulses of acoustic energy at different pitches and amplitudes.

2) These pulses act like sharp taps on the air in the vocal tract.

3) The resonating cavities are set into vibration, enhancing or damping different frequencies.

The vocal tract has a complex shape \(\rightarrow\) contains several bodies of air with different volumes \(\rightarrow\) different overtones.
One pulse in the vocal tract produces **three** different waveforms.

In vowels, we actually hear the sum of these waveforms added together.

The air in the back of the vocal tract produces a low-frequency waveform.

The air in front of the tongue, a smaller cavity, produces a higher-frequency waveform.
Sum of waveforms
1.2 Fundamental Frequency (F0)

- **Fundamental frequency**: number of vocal fold vibrations per second.
- Vocal folds must be vibrating in order to have F0.
- It corresponds to variations in pitch (speech melody or intonation).
- Vocal folds may vibrate faster or slower giving higher or lower pitch to the sound, BUT the formants of the sound remain the same as long as vocal tract shape remains unchanged.

- Male voice: 120 Hz
- Female voice: 220 Hz
- Child voice: 260-280 Hz

- All voiced sounds are distinguishable due to their formants.
1.3 Speech synthesis (demo)

This speech was synthesized in 1971 by Peter Ladefoged on a synthesizer at UCLA.

<table>
<thead>
<tr>
<th>AUDITORY</th>
<th>PHYSIOLOGICAL</th>
<th>ACOUSTIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Pitch of voice</td>
<td>Rate of vibration of the vocal folds</td>
<td>Fundamental frequency</td>
</tr>
<tr>
<td>2. Overtone pitch (1)</td>
<td>First resonance of the vocal tract</td>
<td>Formant 1 frequency</td>
</tr>
<tr>
<td>3. Loudness of overtone (1)</td>
<td></td>
<td>Formant 1 amplitude</td>
</tr>
<tr>
<td>4. Overtone pitch (2)</td>
<td>Second resonance of the vocal tract</td>
<td>Formant 2 frequency</td>
</tr>
<tr>
<td>5. Loudness of overtone (2)</td>
<td></td>
<td>Formant 2 amplitude</td>
</tr>
<tr>
<td>6. Overtone pitch (3)</td>
<td>Third resonance of the vocal tract</td>
<td>Formant 3 frequency</td>
</tr>
<tr>
<td>7. Loudness of overtone (3)</td>
<td></td>
<td>Formant 3 amplitude</td>
</tr>
<tr>
<td>8. Pitch of noise</td>
<td>Fricative and stop bursts</td>
<td>Center of noise frequency</td>
</tr>
<tr>
<td>9. Loudness of noise</td>
<td></td>
<td>Amplitude of noise</td>
</tr>
</tbody>
</table>

The 8 links below demonstrate how speech can be built up using these parameters and additional (fixed) higher formants.

- Formant 1 alone (parameters 2 and 3 above)
- Formant 2 alone (parameters 4 and 5 above)
- Formant 3 alone (parameters 6 and 7 above)
- Formants 1, 2 and 3 (parameters 2 - 7 above)
- Formants 1, 2 and 3 plus additional overtones (parameters 2 - 7 above plus fixed overtones)
- Fricative and burst noises alone (parameters 8 and 9)
- Everything except the fundamental frequency (parameters 2-9)
- Everything including the fundamental frequency (parameters 1-9)
2. Acoustic Analysis

- It is possible to analyze sounds so that we can measure the actual frequencies of the formants and represent them graphically.
- Average of F1, F2 and F3 frequencies in eight American English vowels.

* heed, hid, head, had, hod, hawed, hood, who’d
2.1 Spectrogram

- Computer programs can analyze sounds and show their components. The display produced is called a spectrogram.

- In spectrograms
  - horizontal axis: time
  - vertical axis: frequency
  - degree of darkness or colour: formants

Spectrograms
Dark bands for concentrations of energy at particular frequencies showing the source and filter characteristics of speech
2.2 Computer Programs for acoustic analysis (free access)

- **Praat**
  http://www.fon.hum.uva.nl/praat/
  University of Amsterdam

- **Wavesurfer**
  http://www.speech.kth.se/wavesurfer/
  KTH (Royal Institute of Technology, Stockholm)
TRADITIONAL VOWEL CHART

Chart based on X-ray data

Ladefoged, *Vowels & Consonants*, p.115
2.3 Spectrograms of words (American English)

- The vertical scale goes up to 4000 Hz which is sufficient to show the component frequencies of vowels.
- The exact position of the higher formants varies a great deal from speaker to speaker. They are indicative of a person’s voice quality.
- Observe the effect of the consonant at the end of the vowel.

\[ \text{heed, hid, head, had, hod, hawed, hood, who’d} \]
2.4 Formants in relation to traditional articulatory descriptions

- **F1**
  - Increases from [i] to [u] – as vowel height decreases.
  - Decreases from [a] to [u] – as vowel height increases.
- Hence F1 is inversely related to vowel height.
F2
- higher for front vowels
- lower for back vowels
- affected by lip rounding → decrease of F2 & F3
2.5 F1 by F2 plot

- Zero frequency is placed at the top right corner because formants are inversely related to traditional articulatory parameters.
- F2 scale not as expanded as F1, due to less prominent energy (F1: 80% of vowel energy).
“Traditional vowel diagrams express *acoustic facts in terms of physiological fantasies.*” Oscar Russell (1930s)

- **Vowel height** ⇔ **F1**, not actually tongue height
- **Front – back dimension**
- Degree of backness ⇔ F1-F2 difference
- The closer together F1 and F2, the more “back” a vowel sounds.
Exercise: Make your own F1 by F2 plot
3. Acoustics of Consonants

- The acoustic structure of consonants is usually more complicated than that of vowels.
- In many cases, there is no distinguishable feature during the consonant articulation itself, e.g. silence part of [p, t, k].
- We have to look for the identity of the consonant at the beginning or the ending of the vowel beside it.
3.1 Stops

- Each of the stop sounds conveys its quality by its effect on the adjacent vowel.
- The formants of [p] correspond to the particular shape of the vocal tract.
- During the production of [b] the formants correspond to the particular shape that occurs the moment the lips come apart.
- Closure of the lips causes a lowering of all formants.
- The syllable [b b] will begin with formants in a lower position, then they will rapidly rise to the positions of [b], and finally descend again as the lip closure is formed.
Anticipatory Coarticulation

- For the production of e.g. [bib] or [bab], the tongue will be in position for the vowel even when the lips are closed at the beginning of the word.
- This happens because the part of the tongue not involved in the formation of the consonant closure is already in position for the following vowel.
- The formants at the moment of consonantal release will vary according to vowel.
- The apparent point of origin of the formant for each place of articulation is called the **locus** of that place of articulation.
- The locus depends on adjacent vowels.
3.2 Formant transitions

- Faint voicing striations near the baseline for each of the stops [b, d, g] (voice bar).
- In all three words, F1 rises from a low position due to consonant closure, hence it does not distinguish one place of articulation from another.
- What distinguishes the three stops are the onsets and offsets of F2 and F3.
3.2 Formant transitions

- [bEd]
  - F2 & F3 start at a lower frequency than in [dEd].
  - F2 & F3 are noticeably rising from a low locus.

- [dEd]
  - F2 is fairly steady at the beginning.
  - F3 drops a little.

- [gEg]
  - Characteristic coming together of F2 & F3 → velar pinch
3.3 Voiceless stops

- The release of aspirated stops is marked by a sudden sharp spike → lean vertical line.
- Period of aspiration noise → absence of energy in F1 & no vertical striations
- Frequency & intensity
  - Whisper [t, t, t, k, k, k, p, p, p]. What do you observe?
  - [t] > [k] > [p]
- Intensity of [p] burst is sometimes so low that there is no evidence of it on a spectrogram.

![Spectrogram showing voiceless stops](image)
3.3 Voiceless stops

- Formant transitions also present in aspiration noise.
- [pʰɛm]: F2 & F3 rising into the vowel.
- [tʰɛn]: F2 steady, F3 dropping and then rising.
- [kʰɛn]: characteristic velar pinch
3.4 Nasals

- A clear mark of a nasal (and a lateral) is an **abrupt change** in the spectrogram at the time of the formation of the articulatory closure.
- A nasal has a formant structure similar to that of a vowel. **Differences:**
  - Bands are fainter.
  - Bands located in particular frequency locations depending on characteristic resonances of the nasal cavities.
- **F1:** around 250 Hz
- Large region above F1 with no energy.
- **F2 etc:** varying according to speaker (here around 2000 Hz).
- Place cues sometimes not very clear.
3.5 Voiceless fricatives

- Highest frequencies in speech occur over fricatives.
- Frequency scale increased to 8000 Hz.
- Diphthong [aɨ] : F1 & F2 start close together for low central [a] and move apart for high front [ɨ].
- Fricatives: Random energy distributed over a wide range of frequencies.
Voiceless fricatives [f, θ]

- Same pattern in [f] and [θ].
- Difference: Movement of F2 into following vowel.
  - Very little movement in [f].
  - In [θ], F2 starts around 1200 Hz and moves down.
- Often confused in noisy settings.
- Fallen together in some accents of English, such as London Cockney
  - *fin* and *thin* both pronounced with a [f].
Voiceless fricatives [s, ɻ]

- The noise in [s] is centered at a high frequency, 5000 – 6000 Hz.
- In [ɻ] it is lower, extending down to about 2500 Hz.
- Both [s, ɻ] have **larger acoustic energy** and produce **darker patterns** than [f, θ]
- Both [s, ɻ] are marked with distinctive formant transitions.
- The locus of F2 transition increases throughout the words
  + [f] < [θ] < [s] < [ɻ] (see arrows in fig.)
- Before [ɻ] F2 of [a] is in a position comparable to its location in [i].
3.6 Voiced fricatives [v, ʰ]

- Voiced fricatives [v, ʰ, z, ɾ] have patterns similar to their voiceless counterparts [f, ɾ, s, ɾ].
- Voiced fricatives also have vertical striations indicative of voicing.
- Vertical striations due to voicing are apparent throughout [v] and [ʰ].
- The fricative component of [v] is very faint.
- F2 higher around [ʰ] than [v].
Voiced fricatives [z, ʐ]

- Fricative energy in higher frequencies very apparent in [z, ʐ].
- Voice bar
  - faint in [z]
  - hard to see in [ʐ] – vertical striations due to voicing in 6-8 kHz.
- F2 transition into [ɕ] is
  - level from [z]
  - descending from [ʐ]
3.7 Lateral and central approximants

- Voiced approximants have formants not unlike those of vowels.
- The initial [l] has formants with center frequencies of approx. 250, 1100 & 2400 Hz, which change abruptly in intensity at the beginning of the vowel.
- A marked change in formant pattern is characteristic of voiced nasals and laterals.
- A final lateral may have little of no central contact, making it not really a lateral but a back unrounded vowel.
- A formant around 1100 or 1200 Hz is typical of most initial laterals for most speakers.
3.7 Lateral and central approximants

- The most obvious feature of approximant [ɹ] is the low frequency of F2 and F3.
- F3 begins at 1600 Hz!
- There is great similarity between red and wed. Young children have difficulty trying to distinguish them.
- The approximant [w] also starts with a low position for all three formants.
- F2 of [w] has the sharpest rise, as if it were a very short [u].
- The movements of formants for [j] are like those of a very short [i].
- This is why [w] and [j] are appropriately called semivowels, that is, semi versions of vowels [u] and [i] respectively.
<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voiced</td>
<td>Vertical striations corresponding to the vibrations of the vocal folds.</td>
</tr>
<tr>
<td>Bilabial</td>
<td>Locus of both second and third formants comparatively low.</td>
</tr>
<tr>
<td>Alveolar</td>
<td>Locus of second formant about 1700–1800 Hz.</td>
</tr>
<tr>
<td>Velar</td>
<td>Usually high locus of the second formant. Common origin of second and third formant transitions.</td>
</tr>
<tr>
<td>Retroflex</td>
<td>General lowering of the third and fourth formants.</td>
</tr>
<tr>
<td>Stop</td>
<td>Gap in pattern, followed by burst of noise for voiceless stops or sharp beginning of formant structure for voiced stops.</td>
</tr>
<tr>
<td>Fricative</td>
<td>Random noise pattern, especially in higher frequency regions, but dependent on the place of articulation.</td>
</tr>
<tr>
<td>Nasal</td>
<td>Formant structure similar to that of vowels but with nasal formants at about 250, 2500, and 3250 Hz.</td>
</tr>
<tr>
<td>Lateral</td>
<td>Formant structure similar to that of vowels but with formants in the neighborhood of 250, 1200, and 2400 Hz. The higher formants are considerably reduced in intensity.</td>
</tr>
<tr>
<td>Approximant</td>
<td>Formant structure similar to that in vowels, usually changing.</td>
</tr>
</tbody>
</table>
Read & visit...

- Ladefoged & Johnson “A course in phonetics”, chapter 8
- Ladefoged “Vowels & Consonants”, chapter 7
- Lieberman & Blumstein “Speech physiology, speech perception & acoustic phonetics”, chapter 5, pp. 51-73
- Clark & Yallop “An Introduction to Phonetics & Phonology”, chapter 7

- Visit the websites:
  - [https://corpus.linguistics.berkeley.edu/aop/](https://corpus.linguistics.berkeley.edu/aop/) Material for chapter 8 from UC Berkeley Linguistics, “A course in phonetics”
  - [https://soundphysics.ius.edu/?page_id=812](https://soundphysics.ius.edu/?page_id=812) An Interactive eBook on the physics of sound (Indiana University Southeast)
  - [http://zonalandeducation.com/mstm/physics/waves/waveAdder/WaveAdder1.html](http://zonalandeducation.com/mstm/physics/waves/waveAdder/WaveAdder1.html) Wave Adder
  - [http://www.linguistics.ucla.edu/people/hayes/103/SpectrogramReading/ShortComparisons/](http://www.linguistics.ucla.edu/people/hayes/103/SpectrogramReading/ShortComparisons/) Spectrogram Reading Practice – Short Comparisons (by Bruce Hayes, UCLA)