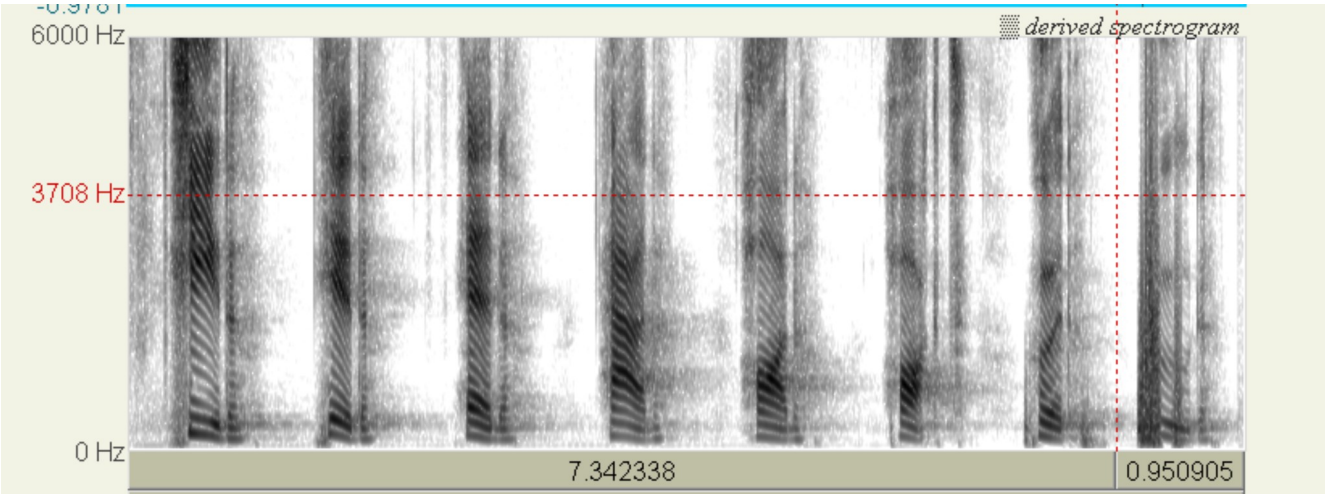
Ash Blum
2024-11-26

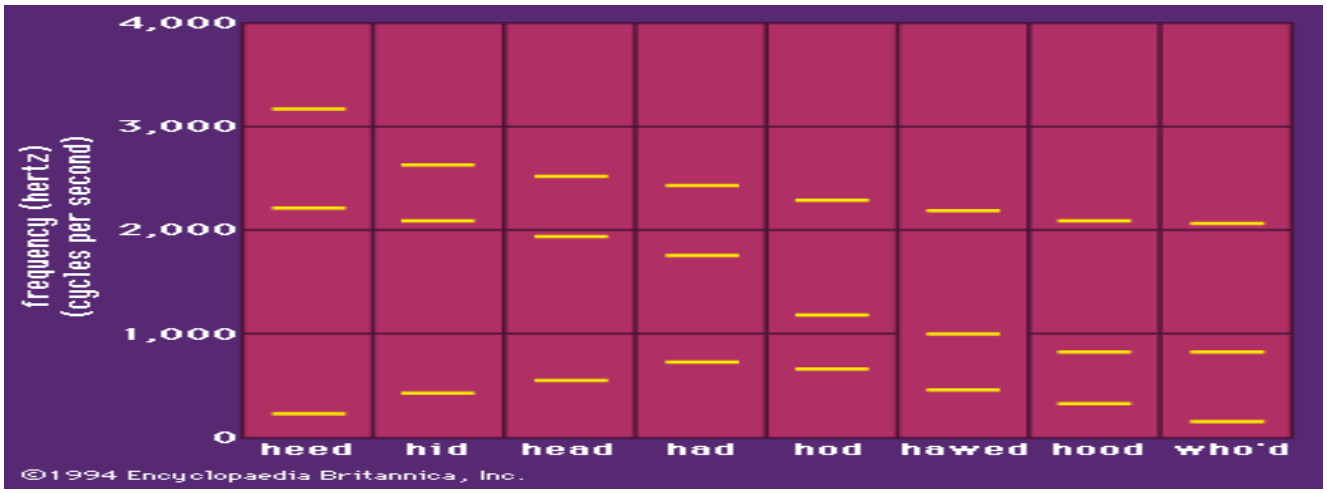
Day 2 - Speech Analysis and Perception. Formants with Praat.

HEED HID HEAD HAD HOD HAWED HOOD WHOD



My spectrogram (above) compared with the chart (below):

The spatial spread of my spectrogram is overall similar to the chart below -- the higher-order formants are more high and closely spaced where the vowel-consonant interactions are more tightly oriented, and as we open our mouths more by lowering the jaw, the first and second-order formants converge more than the higher order ones. So, the second formant is moving from high to low (left to right in the images above and below) as the x-y-z of the mouth moves from wide, flat and stretched to narrow, open, and long.



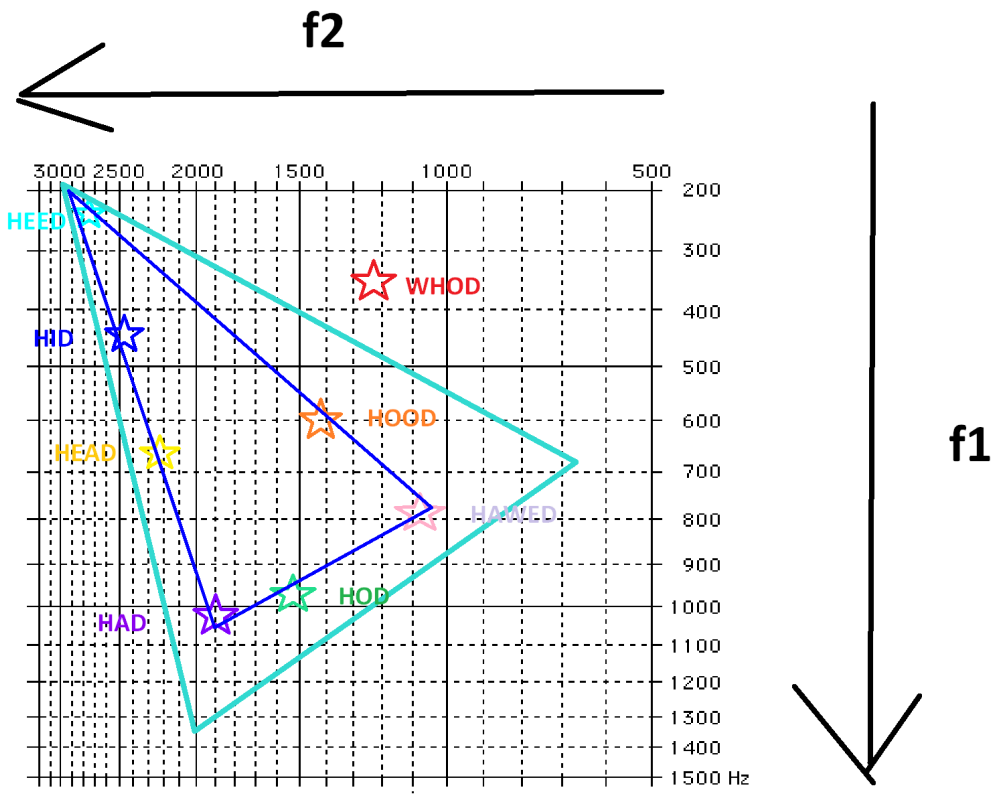
The values on my formant table correspond as well; for example, my F3 values start around 3600 (slightly higher than the chart above) and gradually decrease to 2500 (also higher than the chart above, but by a similar amount). My had/hod values for F2 don't experience quite as sharp a transition above (delta for mine are about 400, and the above

are maybe around 600Hz?). My F1 values also correspond, starting arround 230 and peaking at had (1036Hz - above values peak slightly lower around maybe 800Hz). The slight discrepancy in overall values makes me wonder if perhaps the above chart is for a male.

From Formant Table

	F1	F2	F3	F4
Heed	236	2675	3681	4466
Hid	450	2467	3138	4401
Head	647	2282	3062	4505
Had	1036	1921	2831	4427
Hod	994	1014	3289	4970
Hawed	795	1079	2693	3952
Hood	608	1431	2631	4243
Who'd	375	1252	2503	4153

Vowel Chart:



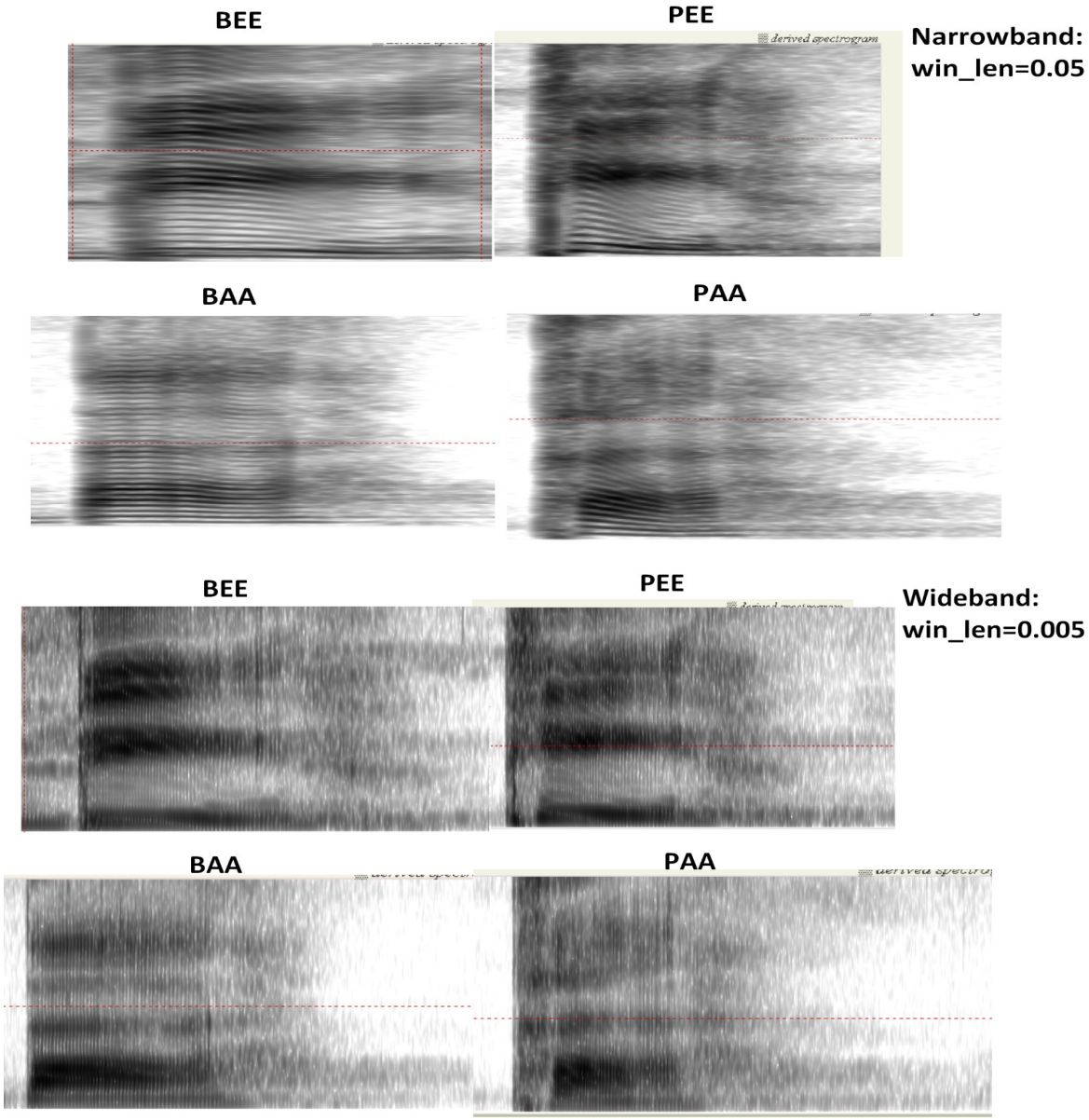
Comparisons between students:

Most other students had triangles of varying shapes and sizes -- some were bigger, and some were smaller, with slight differences in angular orientations. Most AFABs had an orientation similar to that above, and most AMABs had a triangle that was rotated clockwise about 20 degrees relative to that pictured above. The vowel chart illustrates the relationship between resonance of formants and differences in articulation with respect to vowel production.

The relative spread of f1 compared to f2 indicates a physical limitation to the perceived size and/or representation of articulation characteristics (f2) with respect to the

fundamental characteristics intrinsic to an individual speaker (f1). Translation: formant ratio indicates mouth size during phonation. As mentioned above, the second formant moves from high to low as the x-y-z of the mouth moves from wide, flat and stretched to narrow, open, and long. (Looking at the triangular shape, it even almost resembles a mouth!) The f1-f2 triangle gives a boundary to the front/back and high/low vowel production by the tongue in the oral cavity. From a linguistic standpoint, this reflects the tendency of languages to be divided into 3 main vowel categories: [a] [i] and [u]. From a binary sex evaluative standpoint, this also reflects the difference in vocal production organ size (vtl) and subsequent overall pitch variability between male- and female-assignees-at-birth.

BEE PEE BAA PAA



Narrowband (win_len=0.05) formant transitions:

- The narrowband analyses have finer resolution of frequency, which can let us see the pitched formant transitions more clearly.

- BEE and PEE first formant f1 rises from low to high, and f2 and f3 rise into the vowel, with the slope of PEE slightly sharper than that of BEE. BAA and PAA have a flatter formant transition into their subsequent vowels.

Wideband (win_len=0.005) formant transitions:

- The wideband analyses have a finer resolution of time, which can allow us to see the temporal transitions of the formants more clearly.

VOT differences between voiced vs. voiceless consonants:

- Voiced consonants onset transitions are shorter and more compact. There is almost an impulse-type signal at the beginning of the word, which occurs shortly after the voicing. F1 is present for the whole sound and formants are more continuous.
- Voiceless consonant onset transitions are softer and smoother in their onset transition. Voiceless transitions are longer, and there's more space between the onset and the release of the consonant and the rest of the word. F1 occurs later in the transition, and higher formants discretely emerge later in the signal, when the vocal folds begin to engage with sound production.

Cutting the Sound:

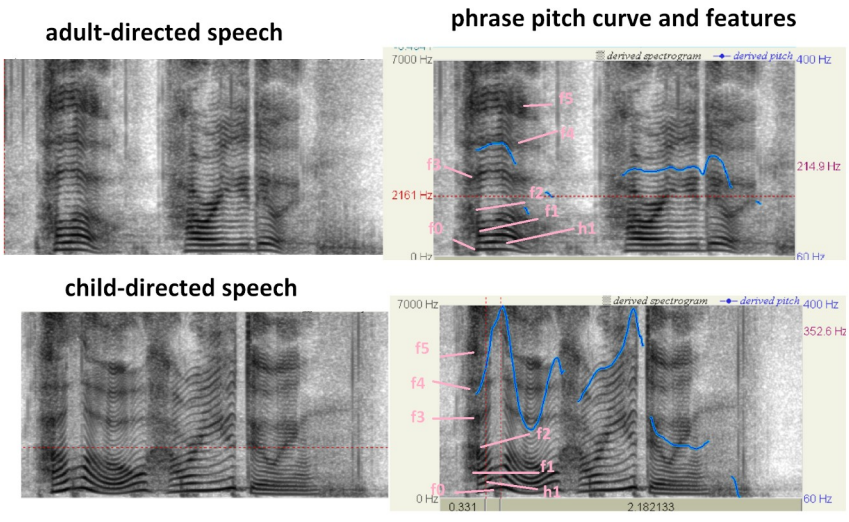
BEE turns into PEE
PEE turns into HEE
BAA turns into AAA
PAA turns into MBAA or MPAA (like in Greek, as Stelios was mentioning in class)

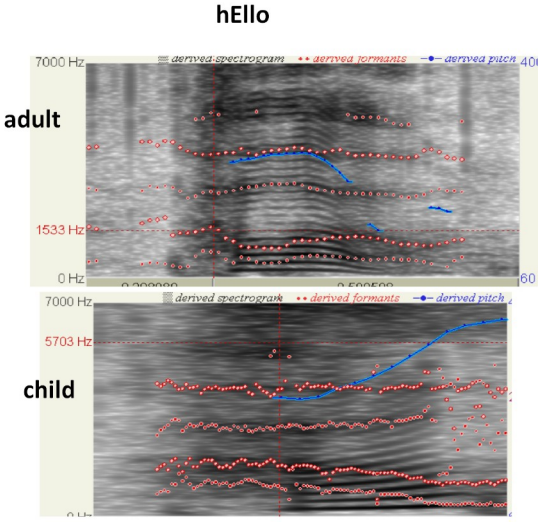
Perceptual consequences:

Cutting the silence from the sound makes the word sound overall cut off and more compact, but it's still recognizable. The prominence of the vowel is more distinct in the cutoff sound. It also has the effect of softening the first consonant somewhat, since our brains and ears "want" to hear something else.

HELLO HOW ARE YOU DOING

Adult vs. Child-directed Speech Pitch Contour Spectrogram:





Acoustic Components: representation, production, perception

fo

- The fundamental frequency, fo, is indicative of the source of a [human] vocal signal.
- It is produced by the vibrations of the vocal cords.
- Perceptual characteristics related to fundamental frequency include perceived gender and size. Generally, adult AMABs have a lower fundamental frequency than adult AFABs. Other physical characteristics that are related to fundamental frequency include vocal tract length, vocal cord length, and body height and size.

f1

- The first order harmonic, h1, is the first harmonic component of the source sound above the fundamental frequency.
- It is a fundamental property of sound, and is value is double the amplitude of the fundamental frequency.
- Perceptually, it is related to vowel production in human phonation, and is indicative of the timbre (pronounced tayum-bor) of the voice. The subsequent ratio of first harmonic h1 to the fundamental fo is also a measure of vocal quality: higher h1/fo values are more breathy or nonlinear, whereas lower h1/fo values are more modal and harmonic.

f2

- The first formant, f1, is indicative of the lowest-frequency spectral peak in the filter of a [human] vocal signal.
- It is a physical characteristic of an individual's supralaryngeal cavity and vocal tract length, and in speech is related to phonation in that it is determined, at least in part, by the linguo-palatal interactions within the filter structure.
- Perceptually, f1 is related to vowel production and vocal quality or timbre. As the interior of the mouth compresses or narrows, f1 lowers, and as the interior of the mouth enlargens or lowers, f1 rises. f1 is also related to vowel production, and is used in the construction of vowel plots (along with f2).

f3

- The second formant, f2, is indicative of the second-lowest frequency spectral peak (above f1).
- It is a physical property of the relative position of the tongue's flatness within the oral cavity as well as vocal tract length; f2 and vocal tract length are inversely related.
- Perceptually, f2 is related to vowel production and vocal quality or timbre. It is indicative of the front/back production of vowels, in that fronted or flattened vowels have higher f2, whereas backed or rounded vowels have a lower f2. f2 is also related to articulation, with sharp f2 slopes indicative of rapid articulation, and more gradual slopes indicative of slower articulation. Along with f1, f2 is used in the construction of vowel plots, as we did in this exercise.

From Voice Reports:

(in Hz)	Adult-dir	Child-dir	Difference
Source:	**Entire**	**Utterance**	
Min F0	135	147	12
Mean F0	217	255	38
Max F0	257	399	142
Std dev F0	26	78	52
Filter:	**first E of**	**hEllo**	
F1 H[e]llo	667	970	303
F2 H[e]llo	847	1647	800
F3 H[e]llo	2949	2952	3
F4 H[e]llo	4106	4157	51

Interpretation of observed differences:

SOURCE

Fundamental frequency is generally higher when talking to children compared to adults, and SD is greater when talking to children. This means that I raised the pitch of my voice to talk to the child, and that I varied my voice more widely. The greatest difference was seen in the max fo, indicating that the frequency peak might be a better indicator in determining from speech who the intended audience is.

FILTER

Frequency variance across formants varied as "small-large-small-large" -- between f1 and f2, then comparing that pair to the f3-f4 pair, the distance between adult-child formant variance seems to zigzag across formant pairs. This could indicate that more information is encoded in those first two "even" formants with respect to intended audience. The greatest variability was in F2 -- this is consistent with what we learned about articulation towards children, in that we exaggerate our articulation when speaking to children.

fin.