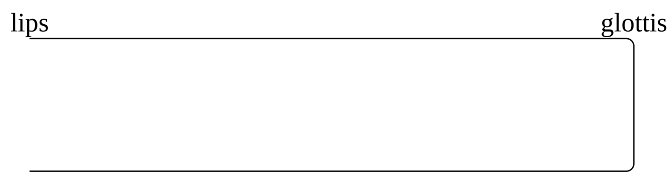
LING 210 **Handout: Tube Models of Vowels** **Spring 2009**

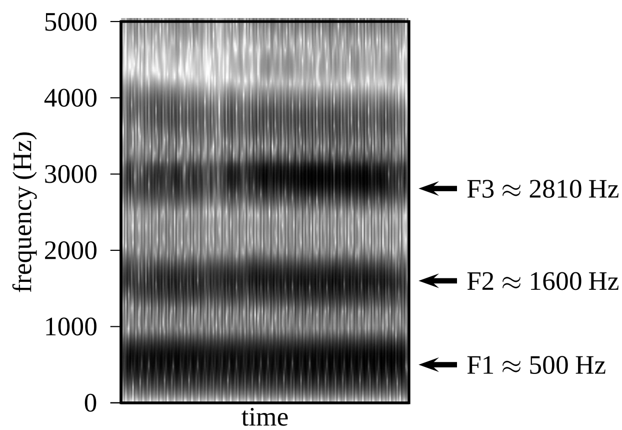
Monday, March 2nd

Since the vocal tract is roughly tube-shaped, we can predict the acoustic properties of speech sounds by creating a **tube model** of the mouth, making assumptions and approximations of the shape of the vocal tract based on articulation. For a mid-central vowel like [ə], the tube is basically uniform from the lips to the glottis. Since any sufficiently small opening can be approximated as a closure, this means that the articulation of [ə] can be approximated as a tube that is open at one end (the lips) and closed at the other (the glottis):

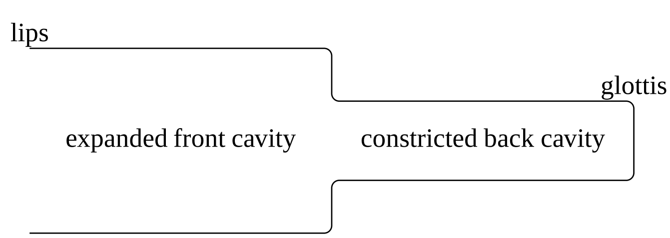


As we already know, the various resonant frequencies of such a tube are given by the formula $(2n-1)s/4L$. A value of 17.5 cm for L is typical of adult males, which means the first three resonant frequencies of this model of the mouth are 500 Hz, 1500 Hz, and 2500 Hz. For vowels, resonant frequencies are called **formants**, and are usually abbreviated as F1, F2, F3, etc.

Formants can be visualized in a special type of three-dimensional graph of the sound wave. If we plot time on the *x*-axis versus frequency on the *y*-axis, then showing intensity as darkness results in a **spectrogram**. Vertical striations in a spectrogram of speech correspond to glottal vibrations; the closer these striations are, the higher the pitch. Formants show up in spectrograms as dark bands, as in the following sample spectrogram for one pronunciation of [ə]:

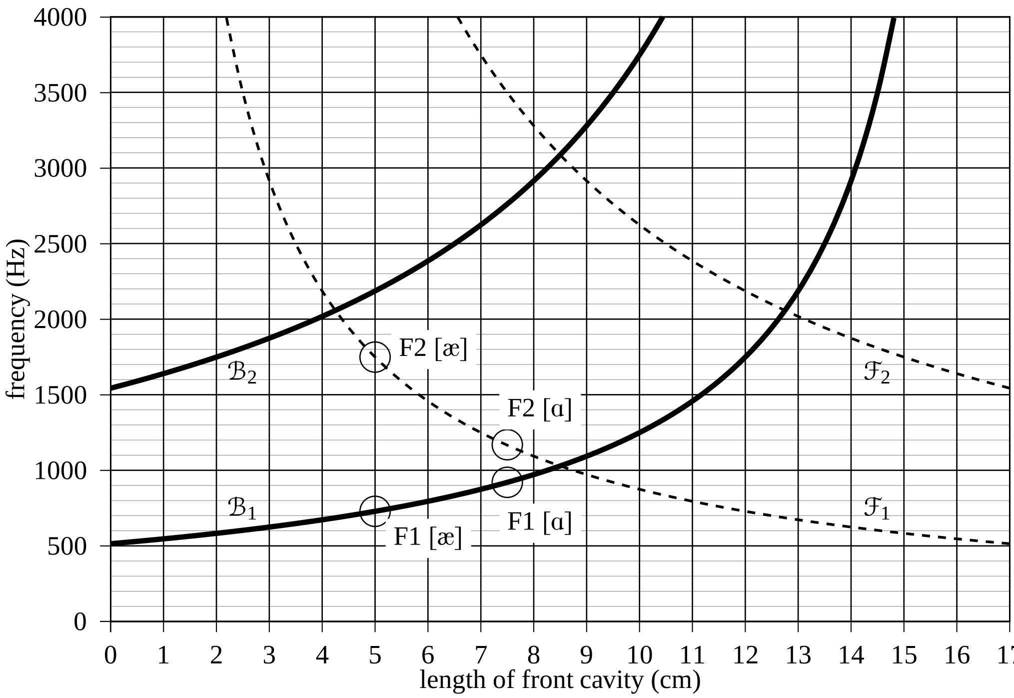


Vowels other than [ə] involve constrictions in the oral tract, so our simple tube model must be made more complex to account for these constrictions. When a low vowel like [æ] or [ɑ] is made, the tongue lowers in the front of the mouth, creating an expanded front cavity, in contrast to the back cavity, which is narrower, partly because of its comparative size with the front cavity, but also because as the tongue lowers, it bunches up in the back, which constricts the throat:



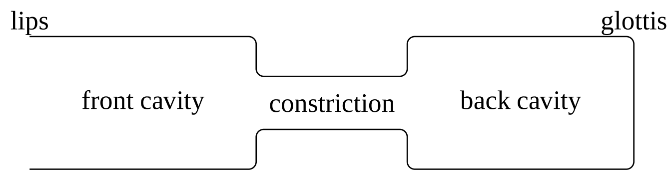
The difference between [æ] and [ɑ] lies primarily in the lengths of the cavities. For a front vowel like [æ], the length of the front cavity is short and the back cavity is long. In comparison, [ɑ] is a back vowel, so the front cavity is longer and the back cavity is shorter. Each of these two tubes affect the sound wave coming from the glottis. At a rough level of approximation, they each separately filter the sound wave with their own resonant frequencies. The final sound wave will have formants that are a simple mixture of those it would get from each tube individually.

F1 for the low vowel will be the lowest of all of the resonant frequencies for both tubes, F2 will be the second lowest of all of the resonant frequencies, and so on. We can see this visually in a graph called a **nomogram**, which shows the acoustic effects of a particular articulation. In the following nomogram for a 17 cm vocal tract, the length of the front cavity is measured on the *x*-axis (this marks the transition between the front and back cavities), while the resulting resonant frequencies are on the *y*-axis. Each line on the graph represents one of the resonant frequencies of one of the two tubes. The resonant frequencies of the front cavity are drawn in thin, dotted lines and labeled $\mathcal{F}_1$ and $\mathcal{F}_2$, while the resonant frequencies of the back cavity are drawn in thick, solid lines and labeled $\mathcal{B}_1$ and $\mathcal{B}_2$. Note that both tubes are open at one end, and essentially closed at the other, so their resonant frequencies are given by the formula $(2n-1)s/4L$:



The constriction for [æ] occurs about 5 cm into the mouth. As we can see on the graph above and plugging the appropriate values into the correct formulas, this results in F1 ≈ 729 Hz and F2 ≈ 1750 Hz. For [ɑ], the constriction is about 7.5 cm into the mouth, resulting in F1 ≈ 921 Hz and F2 ≈ 1170 Hz. Note that these values are only approximations; the mouth is much more complex, and the interactions between the two tubes are not this simple (in particular, where resonant frequencies of the tubes intersect, acoustic coupling occurs, and the true frequencies are somewhat different). However, for yielding a rough estimate, this model works well and can give us general information about the formants for low vowels.

Modeling a high vowel results in a very different configuration of tubes. When the tongue raises to create the vowel constriction, the front cavity remains relatively wide, but so does the back cavity. This results in a three-tube model, with a small tube corresponding to the constriction itself:

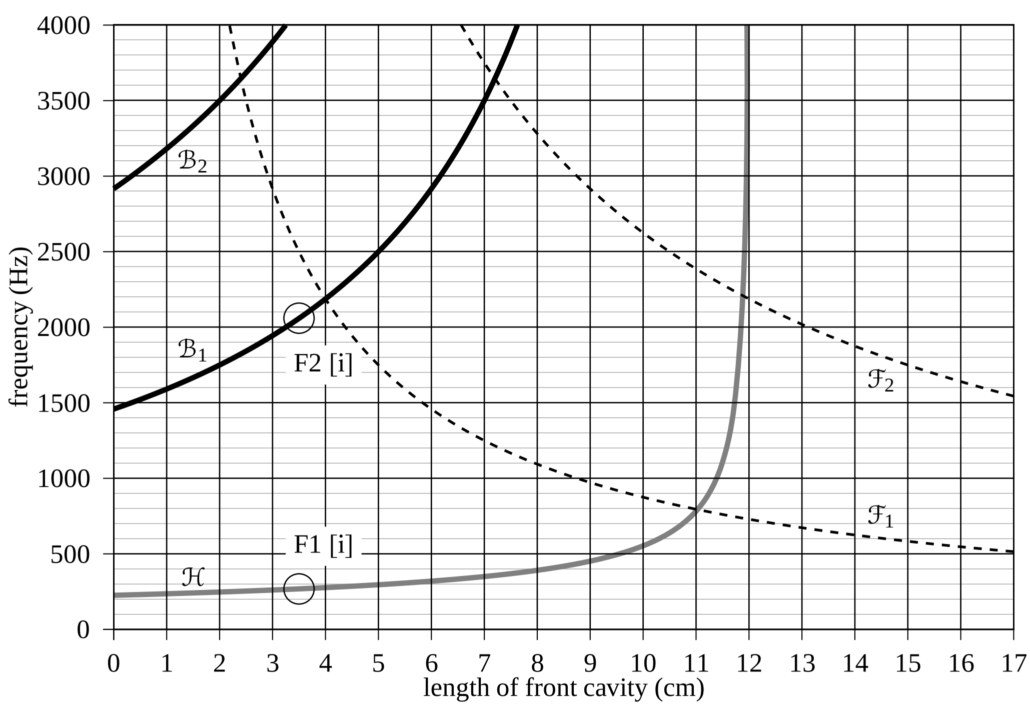


When a thin tube is attached to one end of a large tube that is otherwise closed at both ends, the resulting structure is called a **Helmholtz resonator**. The back cavity combined with the constriction can be approximated as a Helmoholtz resonator. A Helmholtz resonator has its own resonant frequency independent of its component tubes. This frequency is given by the following equation:

$$f_H = \frac{s}{2\pi}\sqrt{\frac{A_1}{A_2\ell_1\ell_2}}$$

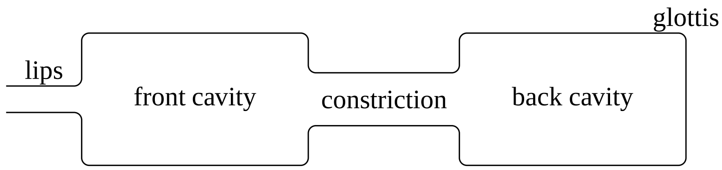
In the Helmholtz formula, A_1 is the cross-sectional area of the thin tube, A_2 is the cross-sectional area of the large tube, and ℓ_1 and ℓ_2 are their lengths. Like the other tubes in the model, the Helmholtz resonator also contributes to the filtering properties of the mouth, and thus, to the creation of formants. For our models of the mouth, we can simplify this formula to $f_H = \alpha s/\sqrt{\ell_1\ell_2}$, where ℓ_1 and ℓ_2 are as before, and α is a parameter that, for our purposes, ranges approximately between 0 and 0.16. When the constriction is more open, α is larger, so f_H is higher. Conversely, for a narrower constriction, α is smaller, so f_H is lower.

In the following nomogram, the resonant frequencies of the front cavity are drawn in thin, dotted lines and labeled $\mathcal{F}_1$ and $\mathcal{F}_2$, while the resonant frequencies of the back cavity are drawn in thick, solid lines and labeled $\mathcal{B}_1$ and $\mathcal{B}_2$. The Helmholtz frequency is drawn in a thick grey line and labeled $\mathcal{H}$. Note that front cavity is open at one end, and essentially closed at the other, so its resonant frequencies are given by the formula $(2n-1)s/4L$, but unlike for the low vowels, the back cavity for high vowels is essentially closed at both ends, so its resonant frequencies are given by the formula $ns/2L$. In this nomograph, the length of the constriction is set to 5 cm, and α is set to 0.05:

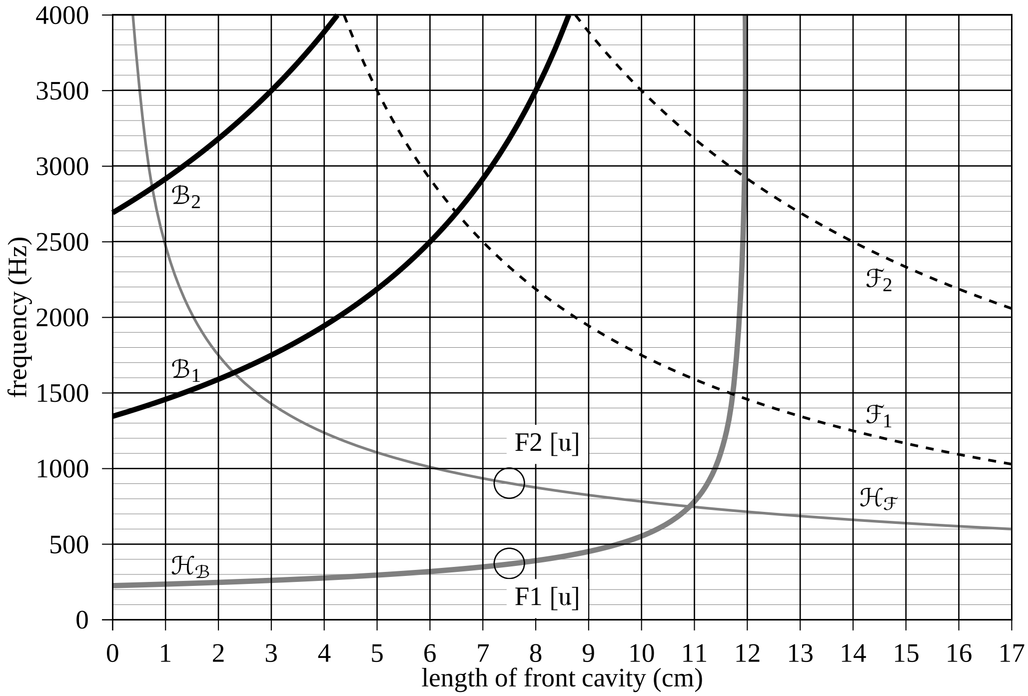


The constriction for [i] begins about 3.5 cm into the mouth. Using the nomogram and the correct formulas, we obtain F1 ≈ 268 Hz and F2 ≈ 2060 Hz.

For [u], the picture becomes more complicated. The lips are rounded, which has two effects: it turns the front cavity into a fully closed tube, and it creates a second Helmholtz resonator in the front of the mouth:



The following nomogram shows graphs for the resonant frequencies of the front cavity (thin, dotted lines, $\mathcal{F}_1$ and $\mathcal{F}_2$), the resonant frequencies of the back cavity (thick, black lines, $\mathcal{B}_1$ and $\mathcal{B}_2$), the Helmholtz frequency from the coupling of the back cavity with the tongue constriction (thick grey line, $\mathcal{H}_\mathcal{B}$), and the Helmholtz frequency from the coupling of the front cavity with the lip constriction (thin grey line, $\mathcal{H}_\mathcal{F}$), with the tongue constriction fixed at 5 cm, the lip constriction fixed at 0.5 cm, and α for both Helmholtz resonators fixed at 0.05:



With [u]'s constriction starting around 7.5 cm, we get F1 $\approx$ 373 Hz and F2 $\approx$ 904 Hz.

From all of these nomograms, we can see a few apparent patterns and make some predictions. Vowel backness is very closely linked to F2: front vowels have a higher F2, while back vowels have a lower F2. Vowel height is very closely linked to F1: low vowels have a higher F1, while high vowels have a lower F1. These predictions are borne out when we measure actual vowels, as in the following typical sample values for F1 and F2 for some vowels of English:

	[i]	[ɪ]	[ɛ]	[æ]	[ɑ]	[ɔ]	[ʊ]	[u]
F1	300	400	550	700	700	600	450	300
F2	2200	1900	1750	1700	1100	900	1000	850

F3 doesn't vary much in the English vowel system, so isn't very useful for distinguishing vowels in English. However, rhoticized vowels have a characteristically low F3. This is not easily accounted for by our simple tube models.