Lecture 8: Analysis of Nasal Consonants

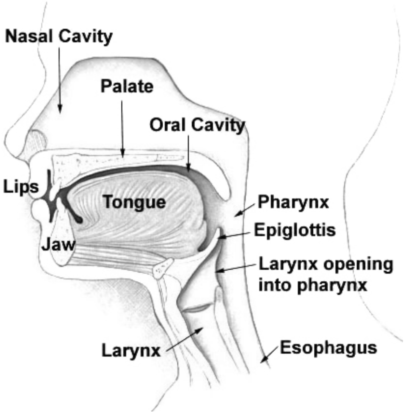
Mark Hasegawa-Johnson

ECE 537: Speech Processing Fundamentals

Outline

Basics of Voiced Speech Production

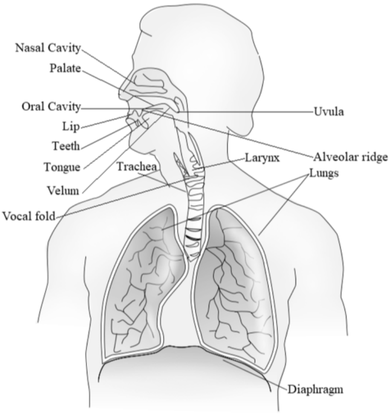
The purpose of the larynx, in most mammals, is to protect the lungs when you eat.
Voiced speech happens when you blow air through the closed larynx, vibrating the vocal folds.



Public Domain, PublicDomain,https://commons.wikimedia.org/wiki/File:Illu01_head_neck.jpg

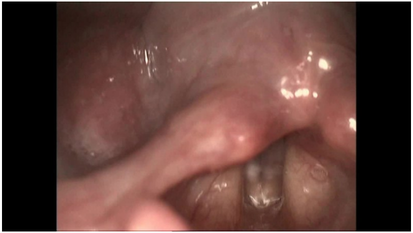
Lungs

The lungs compress in order to raise the lung pressure to about 800 Pascals higher than room pressure.
Most of that 800-Pascal pressure drop occurs across the larynx.



Glottis

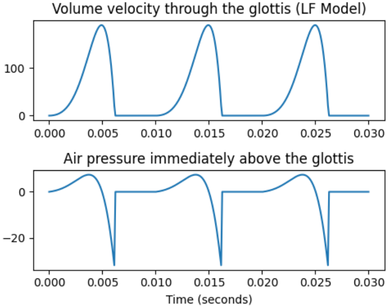
The vocal folds are closed so that they touch. Air is blown past them, so that they vibrate. Vibration frequency is about 100Hz for men, 200Hz for women, 300Hz for children.



→See Video←

Glottal Excitation

When the glottis closes, there is a discontinuity in the derivative of volume velocity.
Pressure is proportional to the derivative of volume velocity, thus there is a pressure discontinuity at closure, rather like a negative sawtooth wave.

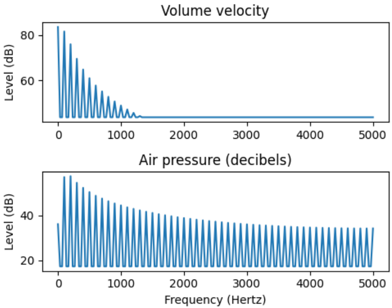


Glottal Excitation Spectrum

The Fourier transform of any periodic signal is an impulse train.

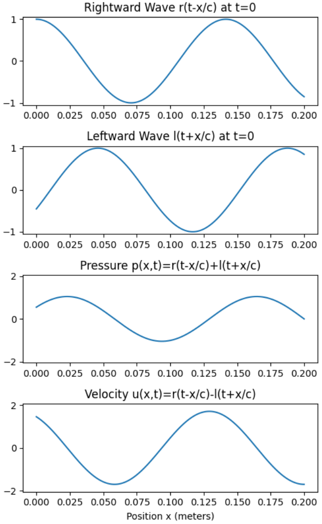
$u(t)$ has a slope-discontinuity, so its spectrum is $|U(f)| \propto 1/f^2$ ($|U(f)|^2 \propto 1/f^4$; $20\log_{10}|U(f)|$ drops at 12dB/octave).

$p(t)$ is discontinuous like a sawtooth wave, so its spectrum is Brownian, $|P(f)| \propto 1/f$ ($|P(f)|^2 \propto 1/f^2$; $20\log_{10}|U(f)|$ drops at 6dB/octave).



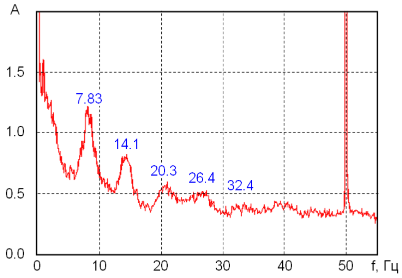
Reflections from the Lips

Volume velocity coming from the glottis activates standing wave patterns in the vocal tract.



Resonances of the Vocal Tract

Like any other system with standing waves, there are some frequencies that resonate (example shown at right has nothing to do with speech. This example shows Schumann resonance frequencies of the atmosphere between Earth and ionosphere, activated by a lightning storm).

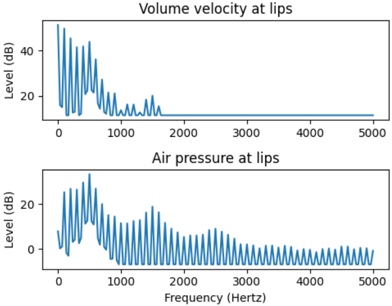


https://commons.wikimedia.org/wiki/File:Schumann_resonance_spectrum.gif

Volume Velocity & Pressure at the Lips

In the speech spectrum, the resonances of the vocal tract show up as a filter $H(s)$, with broad spectral peaks at around 500, 1500, and 2500Hz, multiplied times the glottal excitation spectrum:

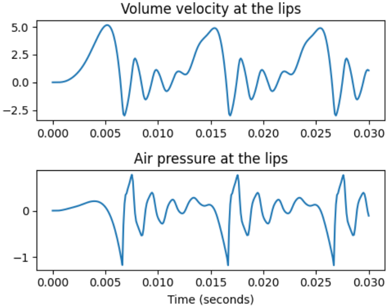
$$U_l(s) = H(s)U_g(s)$$



Volume Velocity & Pressure at the Lips

In the time domain speech signal, you see the resonances of the vocal tract as damped-sinusoid "ringing" that occurs after each glottal closure event.

$$u_l(t) = h(t) * u_g(t)$$



Outline

Rightward Waves, Leftward Waves, and Standing Waves

Remember that the pressure and volume velocity at any position, x, and frequency, $s = j2\pi f$, are

$$P(x,s) = R(s)e^{-sx/c} + L(s)e^{sx/c}$$
$$U(x,s) = \frac{A(x)}{\rho c}\left(R(s)e^{-sx/c} - L(s)e^{sx/c}\right)$$

Boundary Condition at the Glottis

The air flow through the glottis is restricted to a tiny amount. On the other hand, the air pressure at the glottis can be arbitrarily large.

For example, suppose we have $L(s) = 0.999R(s)e^{-2sd_g/c}$. Then the volume velocity through the glottis might be only

$$U(d_g, s) = \frac{A(d_g)}{\rho c}\left(R(s)e^{-sd_g/c} - L(s)e^{sd_g/c}\right)$$
$$= 0.001\frac{A(d_g)}{\rho c}R(s)e^{-sd_g/c}$$

...but the air pressure above the glottis could be...

$$P(d_g, s) = R(s)e^{-sd_g/c} + L(s)e^{sd_g/c} = 1.999R(s)e^{-sd_g/c}$$

Resonance

Resonance happens at frequencies, s_k, where the susceptance at any point in the vocal tract, $B(x,s)=\frac{U(x,s)}{P(x,s)}$, is zero.

$$B(x,s_k)=\frac{U(x,s_k)}{P(x,s_k)}=0.$$

Consider what happens if even a small nonzero volume of air, $U(x,s)=\epsilon$, is injected into the vocal tract at that location:

$$\begin{aligned}R(s_k)e^{-s_kx/c}+L(s_k)e^{s_kx/c} &= P(x,s_k)\\ &= \epsilon\times\left(\frac{P(x,s_k)}{U(x,s_k)}\right)\\ &= \frac{\epsilon}{B(x,s_k)}\rightarrow\infty\end{aligned}$$

So if we find the zeros of $B(x,s_k)$, those are the frequencies at which a small amount of volume velocity will cause an infinitely large standing wave in the vocal tract.

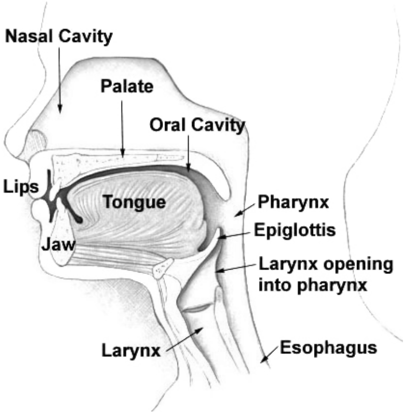
Resonance Frequencies are a Property of the System, Independent of the Measurement Location

Theorem: The frequencies s_k at which $B(x, s_k) = 0$ are independent of x.

These frequencies (the resonant frequencies) depend only on the overall shape of the system as a whole.

Proof for the case of a vowel

Suppose we set the zero point, $x = 0$, at the velum (the flap of tissue that separates the mouth from the nose.
We have two subcavities: the pharynx (looking backward toward the glottis), and the mouth (looking forward toward the lips).



Public Domain, PublicDomain, https://commons.wikimedia.org/wiki/File:Illu01_head_neck.jpg

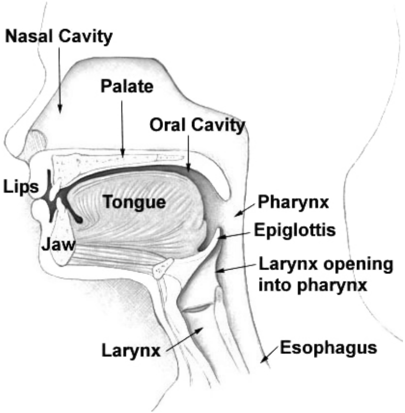
Proof for the case of a vowel

Air pressure at top of the pharynx equals air pressure at bottom of the mouth:

$$P_p(0,s) = P_m(0,s)$$

Resonance is defined by the condition that the volume velocity coming out of the pharynx can be perfectly absorbed by volume velocity going into the mouth (any mismatch between these would cause energy loss):

$$U_p(0,s) = -U_m(0,s)$$

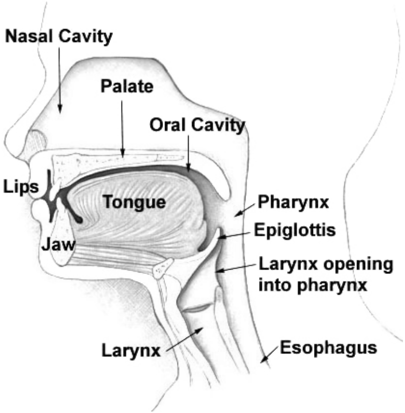


Public Domain, PublicDomain, https://commons.wikimedia.org/wiki/File:Illu01_head_neck.jpg

Proof for the case of a vowel

Putting those together, we find that, at resonance,

$$B(0, s) = B_m(0, s) + B_p(0, s) = 0$$



Public Domain, PublicDomain, https://commons.wikimedia.org/wiki/File:Illu01_head_neck.jpg

Proof for the case of a vowel

For example, remember that if the distance to the glottis is d_g, and the cross-section is uniform, then

$$B_p(0,s) = \frac{A(0)}{\rho c}\left(\frac{e^{sd_g/c} - e^{-sd_g/c}}{e^{sd_g/c} + e^{-sd_g/c}}\right)$$

Similarly, if the distance to the lips is d_l, and the cross-section is uniform, then

$$B_m(0,s) = \frac{A(0)}{\rho c}\left(\frac{e^{sd_l/c} + e^{-sd_l/c}}{e^{sd_l/c} - e^{-sd_l/c}}\right)$$

Proof for the case of a vowel

If we add those together, we get that

$$
\begin{aligned}
B(0,s) &= B_p(0,s) + B_m(0,s) \\
&= \frac{A(0)}{\rho c}\left(\frac{e^{sd_g/c} - e^{-sd_g/c}}{e^{sd_g/c} + e^{-sd_g/c}} + \frac{e^{sd_l/c} + e^{-sd_l/c}}{e^{sd_l/c} - e^{-sd_l/c}}\right) \\
&= \frac{A(0)}{\rho c}\left(\frac{2e^{s(d_l+d_g)/c} + 2e^{-s(d_l+d_g)/c}}{e^{s(d_l+d_g)/c} - e^{s(d_g-d_l)/c} - e^{-s(d_g-d_l)/c} + e^{-s(d_l+d_g)/c}}\right)
\end{aligned}
$$

The denominator depends on $d_g - d_l$, which changes depending on how far you are from either end of the tube. The numerator, however, depends only on $d_l + d_g$, which is the total length of the vocal tract.

Outline

Resonant Frequencies of a Vowel

The resonant frequencies of a vocal tract with uniform cross-sectional area, of length $L = d_l + d_g$, are given by

$$e^{sL/c} + e^{-sL/c} = 0$$

Plugging in $s = j2\pi f$, we get that

$$\cos\left(\frac{2\pi fL}{c}\right) = 0$$

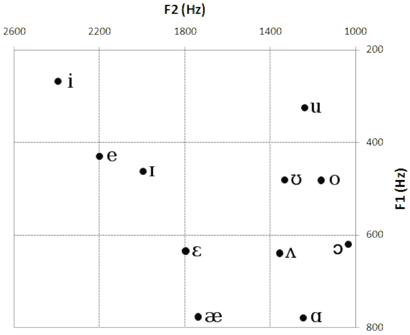
Thus the k^{th} formant frequency is

$$F_k = \frac{c}{4L} + \frac{c}{2L}(k-1)$$

For typical vocal tract lengths, these frequencies are roughly 500Hz, 1500Hz, 2500Hz, . . .

Resonant Frequencies of Vowels

The uniform tube configuration is characteristic of the vowel /ə/ (schwa), the unstressed vowel in "about."
Other vowels are distinguished by formant frequencies that are higher or lower, depending on the positions of the tongue and lips.



Public Domain,

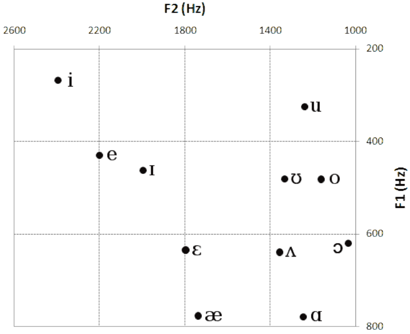
https://commons.wikimedia.org/wiki/File:

English_Monophthong_Formants_Bradlow1995.png

International Phonetic Alphabet

The distinct vowels of American English, using the symbols of the International Phonetic Alphabet, are:

IPA	Example	IPA	Example
/i/	beat	/u/	boot
/e/	bait	/o/	boat
/ı/	bit	/ʊ/	book
		/ɔ/	bought
/ɛ/	bet	/ʌ/	but
/æ/	bat	/ɑ/	baht

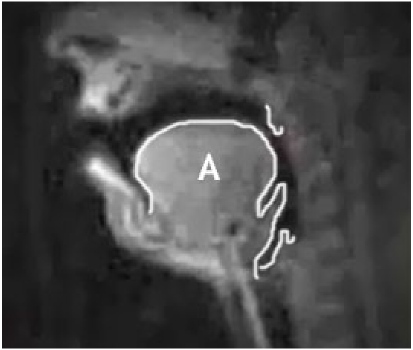


Public Domain, https://commons.wikimedia.org/wiki/File:English_Monophthong_Formants_Bradlow1995.png

Outline

Vowel: Velum Closed

During production of a vowel, the velum is closed, so sound only travels to the pharynx and mouth.



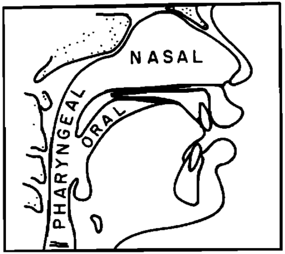
Nasal Consonant: Velum Open

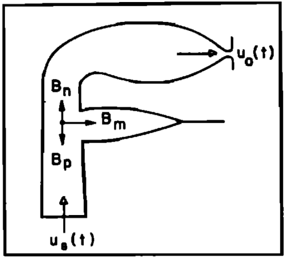
During production of a nasal consonant (/m/, /n/, or /ŋ/), the mouth is closed, and sound travels through the nose, instead.

Resonances of a Nasal Consonant

Fujimura proposed computing the resonances of a nasal consonant by finding the zeros of the total susceptance,

$$B(s) = B_p(s) + B_n(s) + B_m(s)$$

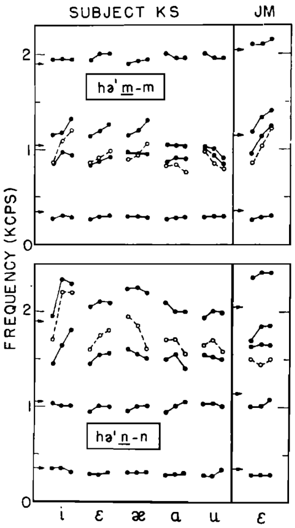




Resonances of a Nasal Consonant

His experimental test was a comparison of three speech sounds:

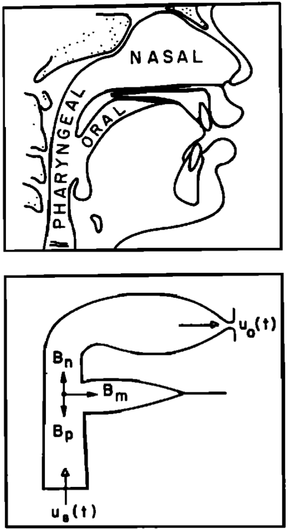
- /m/ (həmɑm)
- /n/ (hənɑn)
- /ŋ/ (hərɑŋ)



Resonances of a Nasal Consonant

For the consonant /ŋ/, Fujimura assumed that the mouth cavity has zero volume, thus $B_m(s) = 0$. The resonant frequencies of the /ŋ/ consonant therefore determine the zero frequencies of the "internal" susceptance,

$$B_i(s) = B_p(s) + B_n(s)$$

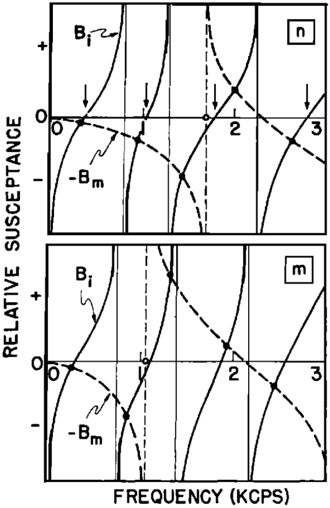


Resonances of a Nasal Consonant

The resonances of /m/ and /n/ are then modeled by the equation

$$B_i(s) = -B_m(s)$$

Fujimura fit a tangent-like curve, as shown at right, and then checked to see if the mouth length predicted by the acoustics was anatomically reasonable.



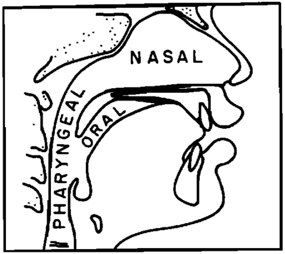
Outline

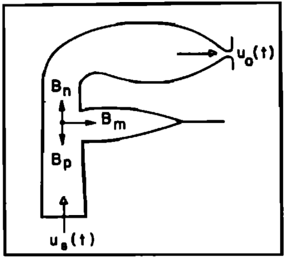
Anti-resonance

Any sound that enters the mouth is totally lost (it never radiates out).
Consider a frequency such that $B_m(s) = \infty$. At this frequency, the resonances of the mouth cavity are imposing the constraint:

$$P_m(s) = \frac{U_m(s)}{B_m(s)} = 0$$

Since $P_p(s) = P_n(s) = P_m(s)$, the mouth cavity essentially kills off all standing waves at this frequency, setting them to zero.

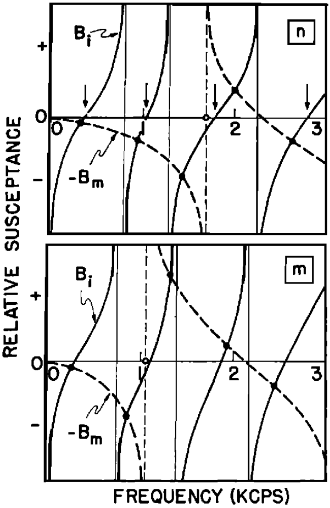




Anti-resonance

The anti-resonance (the zeros of the transfer function) are therefore the frequencies at which

$$B_m(s) = \infty$$



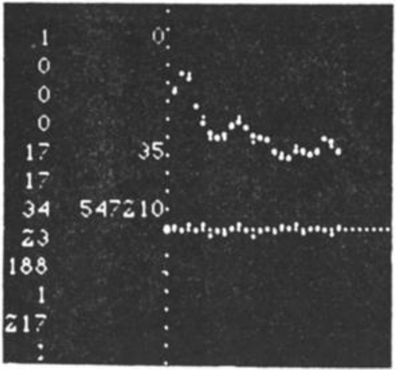
Outline

Transfer Function

The total transfer function is therefore

$$T(s) = \frac{\prod_{j=1}^{n}\left(1-\frac{s}{s_j}\right)\left(1-\frac{s}{s_j^*}\right)}{\prod_{i=1}^{m}\left(1-\frac{s}{s_i}\right)\left(1-\frac{s}{s_i^*}\right)}H(s)$$

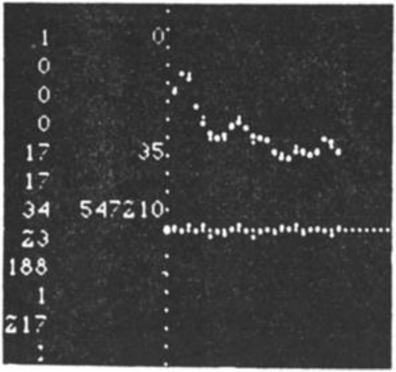
- $s_i = j2\pi F_i$ are the resonances,
- s_j are the antiresonances,
- $H(s)$ is an adjustment for higher-order poles and zeros (higher in frequency than s_m).



Transfer Function

$$T(s) = \frac{\prod_{j=1}^{n}\left(1-\frac{s}{s_j}\right)\left(1-\frac{s}{s_j^*}\right)}{\prod_{i=1}^{m}\left(1-\frac{s}{s_i}\right)\left(1-\frac{s}{s_i^*}\right)} H(s)$$

Fujimura's method was to adjust the zero and pole frequencies until the modeled $T(s)$ and measured $T(s)$ overlapped (as shown here), and the error went to zero (shown as the horizontal line: the number "23" is the mean-squared error, in dB).

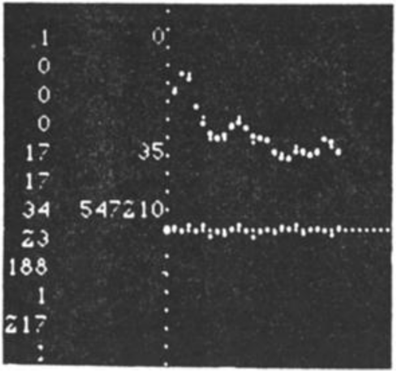


Copyright Journal of the Acoustical Society of America, 1962

Transfer Function

In this example (Fig. 3 from the article, an /m/ consonant), the model was constructed using:

- Pole frequencies at 300, 1050, 1450, 2000, 2650, 3300, and 3600Hz,
- One zero frequency, at 1600Hz.



Outline

Conclusions

- Resonant frequencies are frequencies where the volume velocity can be zero without causing the leftward and rightward waves to go to zero, thus, they are the zeros of $B(s)$.
- Antiresonances can be caused by a shunt cavity, $B_m(s) = \infty$.
- Resonant frequencies are the same, no matter where in the system you measure them.
- Fujimura estimated $B_i(s)$ using /ŋ/, then estimated $B_m(s)$ separately for /m/ and /n/, and showed that the resulting tangent-like functions were matched to reasonable mouth-cavity lengths d_m.