LING 576 Acoustic Phonetics
Spring 2009
Rod Casali
Topic number 6: Source-Filter Theory of Speech Production
1-27-09

Reading:
Fry, D.B. 1979. The physics of speech. Cambridge: Cambridge University Press. Read Chapters 6-8.
Lieberman, Philip & Sheila E. Blumstein. 1988. Speech physiology, speech perception, and acoustic phonetics. Cambridge: Cambridge University Press. Read Chapter 4.

1. Overview

Sound waves produced by the vocal tract are described by a theory known as the *source-filter* theory.

Central assumption: the properties of speech sound waves are determined by the interaction of two factors.

- A sound *source* that produces periodic or aperiodic noise at some point in the vocal tract.
- The vocal tract beyond the sound source, which acts as a *filter* to modify the original sound wave produced by the source.

To understand how sound waves are generated by the vocal tract, we need to understand the properties of both sound sources and the vocal tract filter and the ways in which they interact.

2. Vocal tract sound sources

There are three types of sound sources in the vocal tract:

- Vibrating vocal folds give rise to a periodic sound source.
- Narrow constrictions give rise to *turbulence*.
- A release of built up pressure behind a constriction generates a transient noise source.

What is the sound source for each of the following sound types?

Sound type	Source
Vowels	Vibrating vocal folds
Voiceless fricatives	
Voiceless stops	
[h]	
Voiced fricatives	

Turbulence and transients are both aperiodic noise sources involving random vibrations of air molecules, with no repeating pattern.

Their spectra include noise over a continuous range of frequencies.

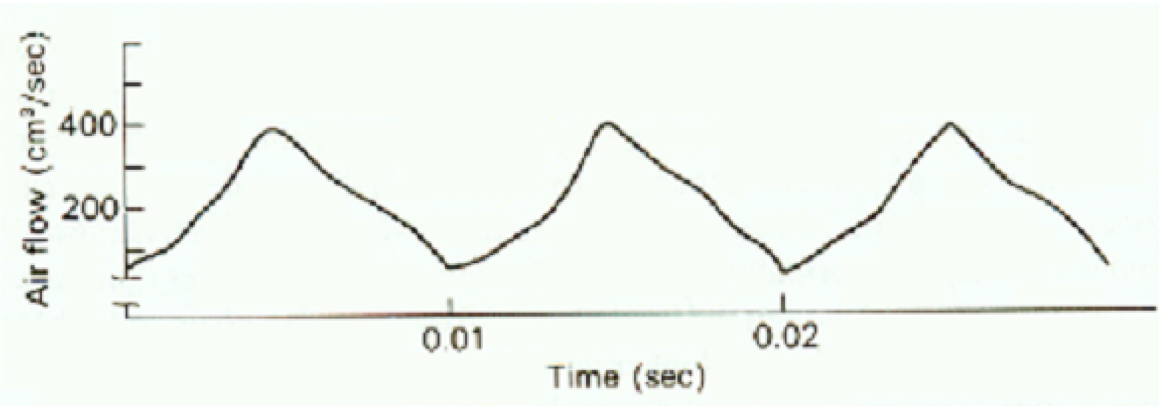
The spectrum produced by regular vibration of the vocal folds is more complex and interesting.

3. Vocal fold vibration as a sound source
The glottal waveform

The rapid opening and closing of the vocal folds during voiced sounds is known as *phonation.*

During phonation, the rate of air flow through the glottis increases as the vocal folds are forced apart, then decreases to a minimum as the vocal folds come together again.

Sample glottal waveform (from Lieberman & Blumstein 1988:35):



What is the fundamental frequency of this waveform?

At its minimum, is the airflow in this example equal to zero? What does this indicate about the degree of closure of the vocal folds?

Vowels and other speech sounds can be produced with different types of phonation, involving different degrees of contact of the vocal folds throughout a glottal cycle:

- *Creaky voice* involves a greater average degree of vocal fold contact than *modal* (normal) voicing.
- *Breathy voice* involves a lesser degree of vocal fold contact than modal voice.

Different phonation types will give rise to different degrees of air flow during the glottal cycle.

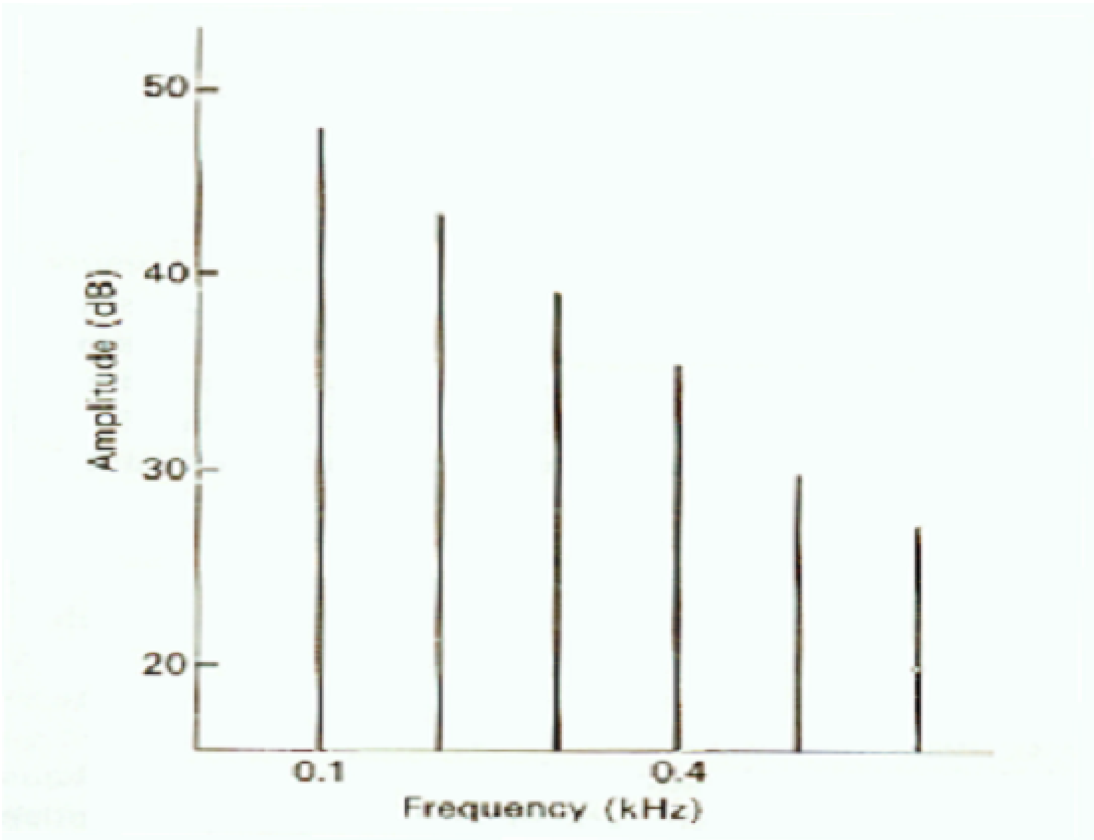
How might the glottal wave shown above change for a breathier vowel?

The glottal spectrum

The (unmodified) spectrum produced by vocal fold phonation is characterized by:

- Sound energy at the fundamental frequency and its harmonics.
- Generally, a loss of amplitude from one harmonic to the next.

Example (adapted from Lieberman & Blumstein (1988:35))



The rate at which the amplitudes of harmonics falls off is approximately 6 dB per octave.

An *octave* is an interval over which the frequency is doubled, e.g. the first octave above 100 Hz is 200 Hz, the second is 400 Hz, the third 800 Hz, etc.

4. The filtering effects of the vocal tract
Resonance, resonators and resonant frequencies

Many objects in nature have natural frequencies at which they will tend to vibrate if exposed to a source of vibration.

Such a frequency is called a *resonant frequency*.

Examples:

One of the more common cases, and the one that is most relevant to speech production, involves resonance of a body of air enclosed within a hollow space.

In such cases, the resonant frequency of the enclosed air column will depend on the shape and volume of the container that bounds it.

The fact that an enclosed air column has a resonant frequency is not sufficient to guarantee that sound at that frequency will be generated.

The enclosed air column is not itself a sound source! It can only act as a *filter* on some other sound source.

If a sound wave at the resonant frequency (with or without additional frequencies present) is produced by some sound source in the vicinity, then the enclosed air column will vibrate at its resonant frequency, serving as an amplifier for any components of the original sound source at that frequency.

Other frequencies that may have been present in the original sound source will not be amplified by the resonator (enclosed air column), and hence will in effect be filtered out.

Bandwidth

An ideal perfect resonator would only amplify sound sources that correspond exactly to its resonant frequency.

Real world resonators have an appreciable *bandwidth*, meaning that they will amplify sounds that are reasonably close to (within the bandwidth of) the resonant frequency, even though they may not be exactly equal to it.

Resonances of the vocal tract: The simplest (schwa vowel) case

The average length of the vocal tract, from the glottis to the opening of the mouth, in adult male speakers is about 17 cm.

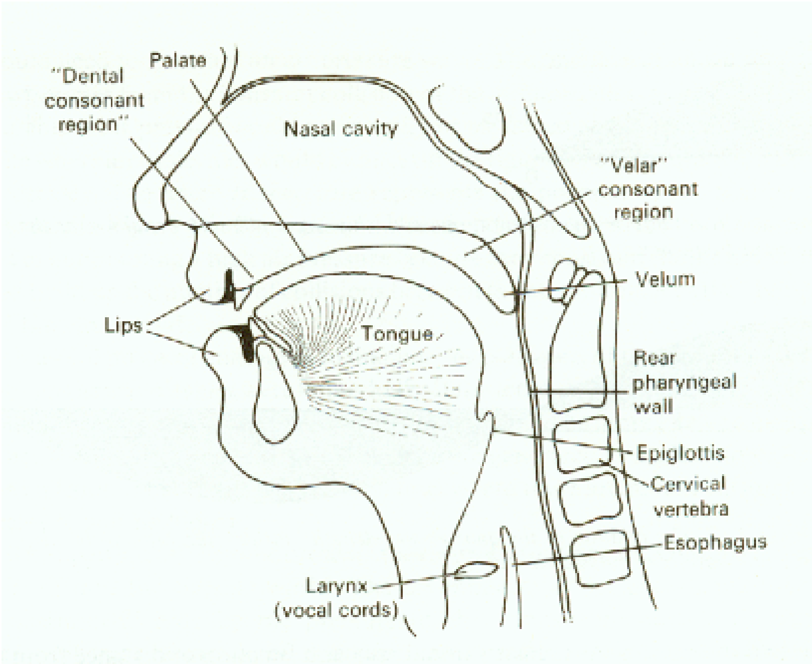
The vocal tract is curved. Also, its cross-sectional area varies from place to place for most speech sounds. (Typically, it will be wide at some places, where there is no constriction, and narrow at others.)

However, in the special case of a vowel of approximately schwa quality, we can ignore both of these factors, since:

- The curvature of the vocal tract turns out to be largely irrelevant acoustically.

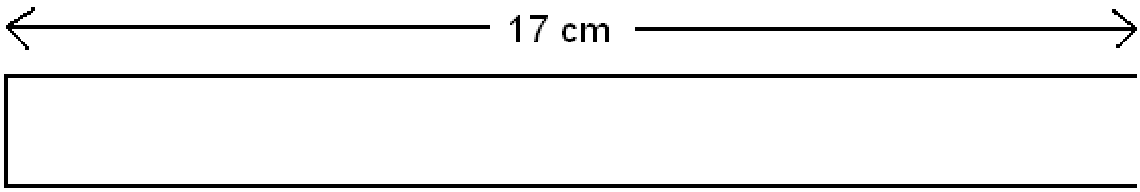
As far as acoustic output is concerned, the results are no different than if the vocal tract were a straight tube. (This is true in the case of other vowels also.)

- For a schwa vowel, the vocal tract is roughly of uniform cross-sectional length.



-- From Lieberman & Blumstein (1988:43)

This means that the acoustic properties of a schwa vowel will very closely resemble those of a straight tube 17 cm long, that is open at one end (the "mouth") and closed at the other (the "glottis").



Such a tube will tend to amplify sound sources at various resonant frequencies.

The lowest such resonant frequency is approximately 500 Hz.

How is this value arrived at?

Such a tube will favor vibration patterns that meet the following boundary conditions:

- Air pressure in the tube will be maximum at the closed end of the tube ("glottis").
- Air pressure at the open end of the tube will be equal to atmospheric pressure ("zero").

There are many different wave patterns that satisfy these "boundary" conditions.

However, the simplest such pattern is one in which the pressure simply falls steadily from a maximum at the closed end to zero at the open end.

This kind of pressure variation pattern is, in effect, a wave, albeit a special type of wave called a *standing wave*.

What is the wavelength of this wave?

The portion of the wave contained within the tube is equivalent to 1/4 of a full wave cycle. (Why?)

Thus, $1/4\ \lambda = 17$ cm, so that $\lambda = 4 \times 17$ cm $= 68$ cm.

The frequency corresponding to this wavelength can be found from the formula $V = \lambda F$, where V is the velocity of sound in air, which is about 33,500 cm / sec.

Thus:

33,500 cm/sec = 68 cm x F

F = 33,500 cm/sec ÷ 68 cm = 493 cycles/sec (Hz), which is close to 500.

This is the *lowest* resonant frequency of the tube.

What is the next simplest standing wave that satisfies the boundary conditions of the tube (pressure maximum at closed end, zero pressure at open end)?

What is the resonant frequency corresponding to this next simplest pattern?

In general, the resonant frequencies of a tube of this type are given by the following formula:

$(2n + 1)\ V \div 4L$

where n is an integer, V is the speed of sound in air (33,500 cm/sec) and L is the length of the tube (17 cm).

This yields 500 Hz, 1500 Hz, and 2500 Hz as the three lowest resonant frequencies. These correspond to the values obtained from the above equation when $n = 0$, 1, and 2 respectively.

Since the 17 cm tube corresponds closely in its acoustic properties to a schwa vowel pronounced by an adult male speaker with average vocal tract size, we predict that the vocal tract during a schwa vowel for such a speaker will have (approximately) 500 Hz, 1500 Hz, and 2500 Hz as its three lowest resonant frequencies.

Vocal tract resonant frequencies are known as *formants*.

Formants are most relevant for vowels, but are occasionally discussed in connection with other speech sounds as well.

Notation: Vowel formant frequencies are often symbolized F_1, F_2, F_3 etc. (with F_1 being the lowest formant, F_2 the next lowest, etc.).

Although vowels have formants above the first three, it is only the first three (and usually only the first two) that seem to be directly relevant to the perception of vowel quality. Formants above F_3 are usually ignored in phonetic studies. (Sometimes F_3 is ignored also.)

The effects of vowel formants on the output spectrum

Formant frequencies of a vocal tract configuration are those frequencies that will undergo maximum amplification if present in the glottal waveform.

In the case of the schwa vowel we are considering, this means that harmonics close to 500 Hz, 1500 Hz or 2500 Hz will be amplified.

It is important to note that formant frequencies can only act to *amplify* frequency components (harmonics) that are already present in the input. It cannot *create* frequency components that are not already present.

This means that if the glottal sound source spectrum lacks significant sound energy at or near a formant frequency, the formant frequency can have no effect whatsoever on the output spectrum.

This possibility usually does not arise when a sound is produced with relatively low F_0. For example, a glottal waveform with F_0 at or below 100 Hz is virtually certain to have some harmonic reasonably close to 500 Hz. (Why?)

However, this kind of effect (lack of harmonics near a formant frequency) can become important in cases where a speaker's F_0 is high.

5. Source-filter theory: putting it all together

The central assumption of the source-filter theory is that the resulting acoustic spectrum for a vowel (or other speech sound) is determined by the combined effects of two interacting factors:

- The spectrum of the vocal fold sound *source*.

This will largely depend on the fundamental frequency of vocal fold vibration (phonation), which will in turn determine the harmonic frequency components that are present.

- The *filtering* properties (formant frequencies) of the vocal tract in the configuration (size & shape) it has for the sound in question.

This is often referred to as the *vocal tract transfer function*.

Neither factor alone is fully determinant; the output is, in effect, a "compromise" between the two factors.

By taking both factors into account, it is possible to approximately predict the output spectrum for a given combination of F_0 and vocal tract formants in simple cases.

As a detailed example, consider the expected output for a schwa vowel produced at a fundamental frequency of 100 Hz.

What output frequency component will have the greatest amplitude?

Questions to consider:

- Which of the source harmonics has the greatest amplitude?
- Which harmonics would the vocal tract filter most strongly amplify?
- What is the expected net effect of these two factors?

How would the output in this example change if F_0 were 350 Hz instead of 100 Hz?

6. More complicated vowel cases
Vowels other than schwa

For other vowels, the vocal tract does *not* have a shape that is approximately that of a single uniform tupe.

Generally, the highest portion of the tongue body is raised toward the roof of the mouth, forming a constriction or narrowing at some point.

This serves to divide the vocal tract into two regions, one in front of the constriction (the "front cavity") and the other behind it (the "back cavity"). The relative sizes and also the shapes of these regions differ for different vowels.

For a low back vowel, as in the word *father* (British pronunciation), the vocal tract can be modeled acoustically as two tubes of roughly equal length, as in the diagram below (from Ladefoged 1996:123):

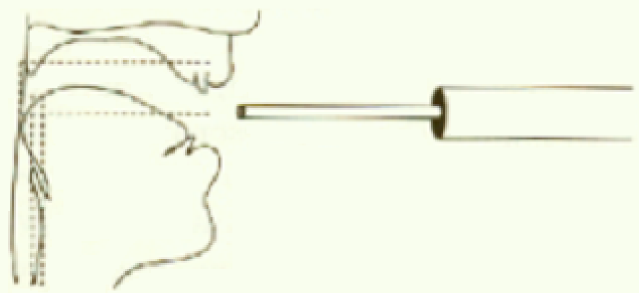
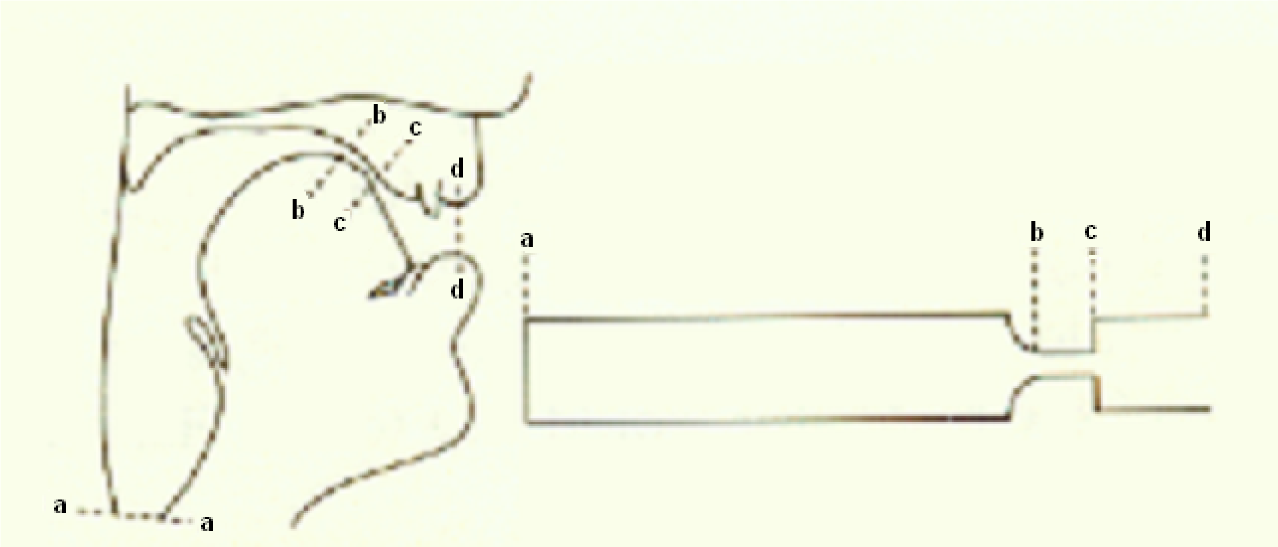


Fig. 8.7. The shape of the vocal tract in the vowel [ɑ] as in *father* schematized as two tubes.

The situation for this vowel is complicated however by the fact that predicted formant frequencies must be adjusted to take into account acoustic interaction between the two regions ("tubes").

A high front vowel /i/ is better modeled as two tubes (one of which is almost completely closed) connected by a short channel, as in the diagram below (adapted from Ladefoged 1996:127):



A complicating factor in this case is that the back (nearly closed) tube does not behave acoustically like the simple case of a tube open at one end. Rather, it's acoustic properties more closely approximate those of a configuration known as a *Helmholtz resonator*. (See the Ladefoged reading.)

The resonant frequency of this region is estimated to be about 280 Hz for typical vocal tract size and shape for [i] (Ladefoged 1996:127-128)

Though the calculation of predicted formant frequencies must take into account a number of complicating factors for non-schwa vowels, it is true in general that F_1 is determined by the size and shape of the back oral cavity and F_2 by the size and shape of the front cavity.

Average formant values of some American English vowels (from Ladefoged 1993):

Vowel	First formant (F1)	Second Formant (F2)
i	280	2250
ɪ	400	1920
ɛ	650	1770
æ	690	1660
a	710	1100
ɔ	590	880
ʊ	450	1030
u	310	870

Effect of lengthening the vocal tract

The effective length of the vocal tract can be increased in either of two ways

- Protruding the lips.

- Lowering the larynx.

In either case, the effect of the increased length is to lower the vocal tract formant frequencies.

7. Application of source-filter theory to consonants

The acoustic behavior of consonants presents more variety than that of vowels, in that:

- The sound source is not always vocal fold phonation.
- The articulatory configuration of the vocal tract varies considerably for different consonants.

Nevertheless, the same basic principle of source-filter theory applies; the output spectrum for a consonant is the result of the filtering effect of the vocal tract on an input source of sound energy at a range of frequencies.

An example: Voiceless fricatives

What is the sound source?

What is the filter?

How should the output spectrum for [s] compare with that of [ʃ]?

References

Ladefoged, Peter. 1993. A course in phonetics. 3rd edition. Fort Worth: Harcourt Brace College Publishers.

Ladefoged, Peter. 1996. Elements of acoustic phonetics. Second edition. Chicago: University of Chicago Press.

Lieberman, Philip & Sheila E. Blumstein. 1988. Speech physiology, speech perception, and acoustic phonetics. Cambridge: Cambridge University Press.