ACOUSTIC THEORY OF SPEECH PRODUCTION
SUPPLEMENT TO AND EXTENSION OF PAPER 3 LECTURES

1. Background and aim
This handout brings together information covered so far on acoustics and the acoustic theory of speech production. Much is repeated, to let us review. Some material is new e.g. consonant acoustics.

It is brought together here to provide a single resource that gives enough understanding of acoustics for you to understand general principles of sound generation in speech, keeping in focus that our purpose is to be able to analyse and interpret speech sounds.

Some people find this material fascinating; others can be offput. If you fall into the latter group, you need only take notice of Sections 1-5, though look also at Sections 7.2, 8 and 9.

2. The acoustics underlying our earlier observations about speech
Individual high-amplitude frequency components determine the detailed shape of waveforms and the formant frequencies visible in spectrograms and spectra; these are important acoustic correlates of what we hear as phonetic quality of speech sounds. What are these separate frequency components like, and why can a particular frequency have high amplitude in one speech sound and very low amplitude in another?

2.1 What is sound?
Noise is caused by oscillation of air particles: alternating regions of compression & rarefaction moving outwards. Figure 1 illustrates two types of particle movement:

- In free field, these waves radiate outwards from the sound source like water ripples when a pebble is dropped in a pond. Concentric circles of compression and rarefaction form (Figure 1a)
- When the source is not close to where the sound is measured, or when the sound is travelling along a tube (and obeys certain conditions that are true for speech) the path of the oscillating air particles can be considered a straight line. This is a travelling wave or plane wave (Fig. 1b.)

Figure 1a
An illustration of the concentric spheres of compression and rarefaction surrounding a sound source in a free sound field.

Figure 1b
Schematic diagram of a travelling wave. Greater particle density (compression) is indicated by closer vertical lines.
2.2 Sound at a single frequency: sine waves and simple harmonic motion

Figure 2. At their simplest, these particle oscillations are symmetrical around the centre point, like the route described by the end of a pendulum. The route taken by the end of the pendulum can be described in terms of progress at a constant rate around a circle, rather than a back and forth movement: this is called simple harmonic motion (SHM): the linear projection of circular movement.

Another way to think about SHM: as the shadow, projected against a wall, of a tall funnel on a toy train as it moves once round a circular track. See demo.

SHM is the simplest form of movement that gives rise to sound. Such sounds are called sine waves: their shape is described by the sines of the angles formed as a point progresses round the 360° cycle at a constant rate. A sine wave has only one frequency. (All other sounds are complex waves.)

Each sine wave can be completely described in terms of its
- **frequency** number of complete cycles per second; measured in Hertz (Hz); psych. experience ≅ pitch.
- **amplitude** extent of vibration or displacement; various measures e.g. volts; psych. experience ≅ loudness.
- **phase** point in the 360° cycle where a sine wave starts, relative to another sine wave; used to describe the relative timing of two sine waves.

Like any periodic complex wave, each sine wave also has a
- **period** duration of one complete cycle: \(1/f\) (\(f=\text{frequency}\)): 100 Hz cycle = 10 ms period; 125 Hz = 8 ms
- **wavelength** distance between equivalent points in two adjacent periods; i.e. distance travelled in one cycle: \(c/f\) \(c=\text{speed of sound in a medium}\); lower-frequency sounds have longer wavelengths.

Figure 3.
The spectrum of a sine wave thus has a single frequency component (of a particular amplitude).

Assuming the axes are on the same scale, the right hand spectrum shows a sine wave with about twice the frequency and two-thirds the amplitude of that in the left hand spectrum.
Figure 4. Sinusoidal oscillation of air particles can be described in different ways, as:
- displacement
- velocity
- acceleration
- instantaneous sound pressure

- Acceleration leads velocity by 90°; velocity leads displacement by 90°.
- When the particle is close to the source, displacement is in phase (0°) with the source.
- For a plane wave which has no reflections, instantaneous sound pressure is in phase with particle velocity.
- When the sound is in a resonator such as a tube, there are reflections from the ends of the tube, and then velocity and pressure are out of phase. This is important in creating formant, or resonant frequencies, as we see later.

2.3 Combining sine waves by adding them together

When sine waves are combined, the resultant wave can be calculated simply by adding the amplitudes of the two waves at enough points in time to allow the new waveform to be plotted.

When sine waves of the same frequency are combined, the result is another sine wave.

Figure 5. Combination of sinusoids of the same frequency. In (a), (b) and (c) the top two sinusoids (1 & 2) are added, thus forming the lower one (R). The amplitudes of (1) and (2) are identical, but have a phase difference of 0°, 90° and 180° in (a), (b) and (c) respectively.

This figure shows that when all component sine waves have the same frequency, another sine wave results, usually with a different amplitude and phase. The frequency remains the same.

Or else silence results...just when the two waves have identical amplitudes and frequencies but are 180° out of phase.
When sine waves of different frequencies are combined, the result is a complex wave – so called because the waveshape is complex compared with that of a sine wave.

**Figure 6.** When the sine waves to be added together have different frequencies, the result is a complex wave: that is, each frequency is preserved in the resultant wave. The waveshape of this complex wave depends on the relative frequency, amplitude and phase of each component sine wave. Top left panel (from Fry) shows two sine waves. Adding the amplitudes together by simply summing the two amplitudes at suitable points in time (remembering the positive or negative sign of each) produces the complex waves at the bottom left. This is like an [u] sound. The right panel (Fig. 1.5 from Johnson) is a complex periodic waves composed of a 100 Hz sine wave and 1,000 Hz sine waves. One cycle of the fundamental frequency (f0) is labelled. The waveshape of this is rather like an [i] sound, although the frequencies are too low to have come from a human vocal tract: what would they be for a human [j]?

### 2.4 Making periodic and aperiodic complex waves

When the component frequencies are related such that higher frequencies are integer (whole-number) multiples of the lowest one, then the complex waveform is periodic, and we hear a pitch.

Each component frequency is called a **harmonic**. The lowest is H1; next is H2 = 2 (H1). H3 = 3 (H1).

The highest common factor of the harmonics (normally, the lowest frequency) is the **fundamental frequency** (f0, = H1). It is f0 that determines the period of the complex waves.

Under most conditions, the perceived pitch is directly related to the fundamental frequency.

The f0 does not have to be physically present in a periodic complex wave in order for us to hear its pitch. This has interesting implications for hearing, and how the ear/brain processes sound to hear pitch. But for speech, the way the vocal fold vibrate means that f0 is always the lowest frequency component of the periodic glottal waveform.
Table 1. Harmonic components in three different (periodic) complex waves, two with f0 = 100 Hz, one with f0 = 200 Hz.

<table>
<thead>
<tr>
<th>Component</th>
<th>Component frequencies (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>H7</td>
<td>700 1400 700</td>
</tr>
<tr>
<td>H6</td>
<td>600 1200</td>
</tr>
<tr>
<td>H5</td>
<td>500 1000 500</td>
</tr>
<tr>
<td>H4</td>
<td>400 800 400</td>
</tr>
<tr>
<td>H3</td>
<td>300 600 300</td>
</tr>
<tr>
<td>H2</td>
<td>200 400 200</td>
</tr>
<tr>
<td>H1 (f0)</td>
<td>100 200</td>
</tr>
</tbody>
</table>

Amplitude and phase changes in the components of a complex wave affect its shape but not its period i.e. not its fundamental frequency.

Most complex waves in the natural world contain many frequencies that are not mathematically related in any way. Such random combinations of frequencies produce aperiodic waveforms, with no regularly repeating pattern in the waveshape (thus no period) and no true pitch.

The dft spectra of Figures 2 and 3 in handout lab3_M08_SpeechandSpectralAnalysis show the f0 and the harmonic frequencies of the vowels they are taken from.

If we were to plot the spectrum of an aperiodic speech waveform, there would be no harmonic frequencies, and we would normally plot only the spectral envelope. This is done by using the Long-Term Average (LTA) spectrum (which is really the average of a succession of spectra with short time windows taken over the duration of the aperiodicity, as if stepping through it. Usually, successive spectra used in this way overlap one another. Thus an LTA spectrum smooths out any temporary random fluctuation in the spectrum.)

3. The basics of resonance

Why can a particular frequency have high amplitude in one speech sound and very low amplitude in another?

It is because the complex wave travels through a tube (the vocal tract), which acts as a resonator.

Resonance can be defined as vibration of an object at its natural frequency (or frequencies) in response to the same or similar frequencies applied by a driving force (either transient (impulse) or continuous). A driving force is an external source of sound. Frequencies in the external sound that are close to the resonance frequencies of the object will be amplified, because it takes least energy to move the air particles at those frequencies i.e. there’s a large sound output for a small input of energy.

You can check this by singing a glissando into a bottle or a tube. Certain pitches will sound louder than others. When they sound loud, you will feel the bottle vibrate: it is resonating in sympathy with the excitation from your voice. Tubes/bottles of different lengths and shapes resonate at different frequencies. A resonator’s natural frequencies can be calculated using well-understood principles.

Likewise, as the vocal tract changes shape, or the sound source changes from the glottis to somewhere in the oral cavity, then the resonance or formant frequencies change.

3.1 Bandwidth and damping of resonance frequencies

Objects vary in what frequencies they respond to, and how strongly they respond.

**Bandwidth**: the frequency range over which an object will respond

**Damping**: how quickly the object’s response dies away after excitation by a single impulse.

When the input is an impulse:

- narrow-bandwidth response — lightly damped (e.g. tuning fork); high amplitude; response rings on
- wide-bandwidth response — heavily damped (e.g. table); low amplitude; short response
Figure 7. Left panel: Waveforms and spectral envelopes from narrow-bandwidth (N) and wide-bandwidth (W) systems excited by a single impulse. The single impulse produces a sound that dies away, hence the waveforms are of the decaying sinusoidal waveform that would be produced by each resonator. The narrowband one is lightly damped; the wideband one is heavily damped.

Right panel: The effect on the resultant waveform of passing a second impulse, 7.5 ms after the first one, through the same resonators as in the upper panel. The narrowband filter produces continuous sound; the wideband filter produces discrete sounds, heard as clicks.

Figure 8. Spectra showing resonance curves for lightly damped (narrow bandwidth) and highly damped (wide bandwidth) systems, responding to the same sound input.

This is what is found in speech: for the same centre frequency and amplitude of the input sound, a wide-bandwidth spectral prominence is normally much lower in amplitude than a narrow bandwidth prominence.

Wide bandwidths are associated with nasal and lateral sounds i.e. whenever two cavities are “in parallel”.

In less technical language:
- a narrowband system “rings on”: it “sees” frequencies very well, but is vague about time
- a wideband system decays quickly: it discriminates poorly between similar frequencies, but is precise about the time at which a sound happens

That is, frequency and temporal resolution are inversely related:
- narrowband: good frequency resolution, poor temporal resolution
- wideband: poor frequency resolution, good temporal resolution

Resonances are regions of extra intensity in the frequency spectrum, so they appear as:
- prominent regions in a spectrum (showing amplitude x frequency);
- black bars in a wide-band spectrogram (showing frequency x time, with amplitude as blackness).

Resonances in speech are called formants. They are labelled F1, F2, F3, etc, starting with the lowest frequency.
Each resonance, or formant, is described in terms of its centre frequency, amplitude, and bandwidth. How it actually appears in a spectrum or spectrogram depends also on the properties of the filter system used to make the spectrum/spectrogram.

Consider only voiced speech sounds, which have complex periodic waveforms (from vocal fold vibration), and hence, by definition, a fundamental frequency f0 and harmonics. The lowest f0 is likely to be greater than 80 Hz:
- Narrow-bandwidth sgm has a bandwidth of c. 45 Hz. It shows f0 and harmonic components (more-or-less horizontal stripes), but not precise temporal events.
- Wide-bandwidth sgm has a bandwidth of c. 300 Hz. It shows only broad regions of resonance, but precise temporal events (each short-term vertical striation reflects a single glottal pulse (vocal-fold vibration).

**INTERIM SUMMARY: THE BASICS**

Sound is either periodic the same pattern repeats more or less regularly: phonated sounds like [i ɑ u].
or aperiodic there is no regularly repeating pattern: voiceless sounds lie [s].

A **sine wave** is periodic but has only one frequency.
A **complex wave** comprises two or more frequencies. All speech sounds are complex, and all have many frequencies.
If a complex wave is periodic, then its components are harmonics.
(If there are a few nonharmonic frequencies (enharmonic partials) amongs the harmonic partials, then the signal is almost periodic, or quasi-periodic.)

**Table 2**

<table>
<thead>
<tr>
<th>Types of Complex Wave</th>
<th>Periodic</th>
<th>Aperiodic</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Definition</strong></td>
<td>Regularly repeating waveform i.e. repeated cycles of the same shape.</td>
<td>Irregular waveform; i.e. no cyclic repetition.</td>
</tr>
<tr>
<td><strong>Properties</strong></td>
<td>The waveform comprises a fundamental frequency (f0) and harmonic frequencies (nf0). The harmonics are integer multiples of the f0. Thus f0 is the slowest component, and also the highest common denominator of the harmonics. We hear a pitch.</td>
<td>Has components at non-integer multiples of the lowest frequency (i.e. there is no f0, and no true pitch).</td>
</tr>
<tr>
<td><strong>In speech</strong></td>
<td>Voiced sounds e.g. vowels, nasals. Produced at the glottis (phonation)</td>
<td>Voiceless sounds e.g. /f, s, p, t/. Produced at the glottis (aspiration noise) or above the glottis (friction noise)</td>
</tr>
</tbody>
</table>

Each periodic waveform in a speech signal contains many frequencies because the laryngeal sound source contains many frequencies. The frequencies with the greatest amplitudes—usually the formants, or resonance frequencies—can normally be seen and measured in the waveform, spectrogram, and spectrum. Because each speech sound has distinctive formant frequencies, it also has a distinctive characteristic waveshape (pattern).

**Resonance** is fundamental to speech acoustics because most differences in phonetic quality stem from differences in resonance patterns of the vocal tract as it changes shape. Thus we study the resonance properties of tubes, representing a simplified model of the vocal tract. The length and shape of the vocal tract are the main determinants of the formant frequencies of speech sounds.

Examples of different types of source and vocal tract shape. (Mid-sagittal sections)

Output depends on:
1. **Source spectrum:**
   - periodic (narrow/complete constriction at glottis → regular vocal fold vibration)
   - aperiodic (incomplete laryngeal or supralaryngeal constriction → turbulence noise)
   - mixed (periodic (phonated/voiced) + turbulence noise)

2. **Transfer function:** determined by length and shape of vocal tract
   a. vocal tract length – longer vocal tracts have lower natural frequencies (i.e. resonances)
   b. vocal tract shape – strictly, the cross-sectional area at each point along the VT, simply modelled as the cross-sectional area at the point or points of maximal constriction

3. **Radiation function:** attenuates low frequencies as they emerge at the lips (c. +6 dB per octave)

A resonator acts as a filter on the original source of sound. Think of it as rearranging the input energy so that frequencies that are at or near the resonance frequencies are amplified, at the expense of those frequencies that are not near the resonance frequencies.

Figure 9 illustrates these principles. It shows spectra of
- the glottal source (periodic, dropping off at -12 dB per octave for modal voice)
- the vocal tract transfer function, when the vocal tract is unconstricted, as for schwa.
- the radiation function (as the sound emerges at the lips: always the same: +6 dB per octave)
- the final output spectrum

At its simplest, just think about the process as one of melding a series of independent patterns so that the final pattern shows the combined influence of all of the initial and intermediate patterns.
Figure 10. Upper panels, left to right. A glottal source spectrum with f₀ = 100 Hz. An idealized transfer function for an unconstricted tube of about 17 cm: this has formant frequencies at about 500 Hz, 1500 Hz, and 2500 Hz, and sounds like schwa. The output spectrum. Lower panels, left to right. The same, except that the f₀ of the glottal source is 200 Hz. Formant definition is not so clear. Think about what this could mean for intelligibility at high fundamentals.

5. Resonance in more detail

Resonance is fundamental to speech acoustics because most differences in phonetic quality stem from differences in resonance patterns of the vocal tract (VT) as it changes shape. We study the resonance properties of tubes, representing a simplified model of the vocal tract. Two types of tube are relevant: bottle-shaped tubes, and straight-sided tubes. Most speech sounds are best modelled using a series of straight-sided tubes. Bottle-shaped tubes model only some special cases in speech, but as they are more familiar, we start with them.

5.1 Helmholtz resonators: cavities with narrow necks e.g. bottles.

They give a single resonance whose frequency depends on the relationship between the acoustic mass (the plug of air in the neck) and compliance (the relatively springy particles in the body of the bottle) in the system. Helmholtz resonances in speech are always low frequency and only occur in special cases, e.g. lowest resonance (F1) of high vowels ([i u]). (Narrow mouth opening with one or two large cavities behind it.)

Figure 11. Some Helmholtz resonators:

Formula: only bother with it if you want to!

\[ \text{Resonance frequency} = \frac{c}{2\pi} \sqrt{\frac{A}{Vl}} \]

where
- \( c \) = speed of sound in [air]
- \( A \) = cross-sectional area of neck
- \( V \) = volume of bottle area (back cavity)
- \( l \) = length of neck

Model for [i]:
a single Helmholtz resonator

Draw the model for [u] here as two Helmholtz resonators:
5.2 Simple tubes of uniform cross-sectional area

The tube shape with the most general application to speech is a straight tube of uniform cross-sectional area—i.e. no constrictions. The vocal-tract shape for schwa ([ə]) can be modelled as a single such unconstricted tube:

\[ \text{Figure 12} \]

Vocal tract (VT) shapes for all other sounds must be modelled by more than one tube, because they involve at least one constriction, but the principles are essentially the same.

6. What causes resonance in a straight-sided tube?

For ideal tubes of uniform cross-sectional area, resonance arises when standing waves of pressure (and velocity) occur.

A small volume of air in a tube can have a velocity (hence kinetic energy); it can also be compressed and expanded so that there are variations in sound pressure (hence in potential energy). A wave of sound pressure or velocity travelling down a tube is called a travelling, or plane wave. In an ordinary travelling wave of any given frequency, velocity and pressure (kinetic and potential energy) fluctuate together, either 0° or 180° out of phase.

When a sound wave travels down a tube, it is reflected back upon reaching the end of the tube. Cf. slinky spring.

Reflections occur because the ends of tubes form acoustic boundaries. (There are acoustic boundaries even when a tube’s ends are open enough to allow air to flow out, as the lips are in the production of a vowel. Compare the way it can be hard to hear someone who’s talking inside a car with the window open, when you are outside the car.)

All travelling waves in a tube are reflected from the ends. The amplitude of the incident (→) and reflected (←) waves sum to give the overall pattern of pressure and velocity fluctuation.

Reflections of waves at most frequencies tend to create complex but variable patterns that do not excite resonances. But at certain frequencies, the pattern of incident and reflected waves sums to a simple pattern whose peak amplitude at any given place in the tube is maintained (“stands still”) over time. Hence the term standing waves.

Standing waves thus arise when incident (→) and reflected (←) waves sum in such a way that:

- their combined peak amplitude is constant over time at any one point in the tube, but
- there are differences in peak amplitude at different places in the tube.

There are only certain frequencies at which this can happen in any given tube (of uniform cross-sectional area). In order for it to happen, the the wavelength of the sound must be the right length to fulfil certain boundary conditions.
Wavelength = distance (cm) travelled in one cycle
Thus, wavelength depends on speed of sound (referred to as \( c \))
In air, \( c = 34000 \text{ cm/s} \)
Thus, higher frequencies (e.g. A) have shorter wavelengths than lower frequencies (e.g. B).

### 6.1 Boundary conditions

When the wavelength of a sound is the right length to fulfill these boundary conditions, the sound will keep being reflected by each end of the tube (it will keep bouncing back and forth along it) and it will be amplified because it takes less energy to move the air particles. A frequency at which this amplification arises is a resonance frequency.

You can think of it as happening when the boundary conditions are such that the particular frequency keeps bouncing back and forth along the tube, reinforcing itself as it goes, rather than causing random fluctuation or cancelling itself out as it goes back and forth. More technically, reinforcement rather than random fluctuation or cancellation takes place when the boundary conditions cause pressure and velocity waves of the same frequency to become 90° rather than 0° out of phase.

The boundary conditions necessary for standing waves to occur depend on the state of the tube's ends.

![Diagram showing boundary conditions](M4_AcThSpProd)
The three lowest-frequency pressure waves that meet the boundary conditions for resonating in an unconstricted vocal tract as in schwa. They are labeled F1, F2 and F3 for vowel formants 1, 2 and 3.

The formula for the lowest resonance of such a tube is
\[ F_1 = \frac{c}{4L} \]
where \( c \) = speed of sound (c. 34,000 cm/s in air); \( L \) = length of tube (16-17 cm for a man; c. 14 cm for a woman.)
in speech, \( L \) represents distance from glottis;
glottis is modelled as closed; lips are modelled as open

Successively higher resonances occur at odd-number integer multiples:
\[ F_2 = \frac{3c}{4L}, \quad F_3 = \frac{5c}{4L} \]
& so on.

Work out that this must be so if the above statements are right. Check your reasoning vs. Figure 15.

Figure 14. Amplitude envelopes of standing waves of velocity for a tube of uniform cross-sectional area, closed at one end, open at the other. Draw in the standing waves of pressure for yourself. (They are 90° out of phase with velocity (see last point in Figure 4), so should have maxima at the closed end, minima at the open end.)

A tube that is closed at one end and open at the other is called a “quarter-wave resonator”.

An average man’s vocal tract is 17 cm long, so for [ə], which is modelled as a single unconstricted tube, his formant frequencies are:
\[ F_1 : \frac{34,000}{4 \times 17} = 500 \text{ Hz} \]
\[ F_2 : 3(500) = 1500 \text{ Hz} \]
\[ F_3 : 5(500) = 2500 \text{ Hz} \]
6.2 For a tube that is closed at both ends, or a tube that is open at both ends:

The same boundary conditions apply: there must be a pressure maximum at a closed end, and a minimum at an open end. The lowest frequencies that meet these boundary conditions are shown in Figure 16.

Figure 16. The lowest three standing waves of pressure for a tube that is closed at both ends (left) and one that is open at both ends (right)

We call these tubes “half-wave resonators”.

Half-wave resonator models are used for obstruent consonants, and for [i].

7. Two ways to apply this information to speech

1. tube models
2. perturbation theory

7.1 Tube models

You can estimate the formant frequencies of a given vocal-tract shape if you know:

(a) the overall length of the vocal tract (glottis-to-lips);
(b) how many tubes of uniform cross-sectional area reasonably approximate that shape (Usually 2 for vowels and 2 or 3 for consonants. For 2 tubes, the back cavity extends from the glottis to the major constriction, and the front cavity from the major constriction to the lips. See below.)
(c) what the acoustic boundaries of each tube are (closed or open).
A constriction effectively divides the vocal tract into two tubes, each with its own resonance frequencies.

- the location of the major constriction in the vocal tract determines the length of the front and back cavities;
- the cross-sectional area of one tube relative to the adjacent tube determines whether its ends are modelled as closed or open. This principle is illustrated in Figure 17. It’s not crucial that you understand it, but if you do, then you can work out formant frequencies for yourself using a tube model of the VT.

**Figure 17:** To determine boundary conditions: (c = closed; o = open)

The type of boundary condition is what you would mainly see if you were inside one tube, looking out from one end (or, towards its end): mainly “closed” wall, or mainly “open” air.

### 7.1.1 Examples for vowels

Two tubes, each of uniform cross-sectional area.

**Figure 18:** Model for [a]

Lb = length of back cavity (glottis to point of maximum constriction)
Lf = length of front cavity (point of maximum constriction to lips)

Boundary conditions: both back and front cavities are modelled as closed-open.

So the quarter-wavelength model (as for schwa) is used for each tube separately. Roughly:

\[
\begin{align*}
\frac{c}{4L} & = \frac{34000}{28} = 1215 \text{ Hz} \\
\frac{3c}{4L} & = \frac{3 \times 1215}{3} = 3645 \text{ Hz}
\end{align*}
\]

First 3 formant frequencies measured at lips for this vowel should be:
- F1 = 850 Hz (lowest front-cavity resonance)
- F2 = 1215 Hz (lowest back-cavity resonance)
- F3 = 2550 Hz (second front-cavity resonance)

These values are slightly high for an adult male, because the model is too simple to allow precise prediction of formant values. But the general pattern of formants is right: for [a], F1 is high and close to F2.
**Figure 19: Model for [i]**

F1 is a Helmholtz resonance. (“Bottle-shaped” tube, with a single low resonance.)


\[
\begin{align*}
L_b &= 9 \text{ cm} \\
L_f &= 8 \text{ cm} \\
c/2L_1 &= \frac{34000}{18} = 1890 \text{ Hz} \\
c/2L_2 &= \frac{34000}{16} = 2125 \text{ Hz}
\end{align*}
\]

These formants are raised further in frequency because the back cavity is tapered towards the constriction, and a tapered end raises resonance frequencies.

N.B. If you start calculating formant frequencies to see if you understand, don't give up if your answers don't match real speech well. Rather, talk with your supervisor. Your calculations may be right, but your estimates of tube lengths may be wrong, and additionally, a number of adjustments have to made to match real speech (like fat in the cheek walls, warm moist air, overall VT shape in some special cases......).

7.1.2 Obstruents

**Figure 20. Model for [s]**

Three cavities (tubes):
- back cavity: closed-closed
- constriction: open-open
- front cavity: closed-open.

\[L_c = \text{length of constriction (when it’s long enough, it acts as another tube with its own resonances)}\]

The back cavity is excited relatively little by a supralaryngeal noise source, and the amplitudes of its resonances are further reduced by antiformants (zeros) arising from interactions between front and back cavity resonances. Therefore, the back cavity contributes little to the output.

Greater accuracy is achieved when other factors are taken into account (eg length of constriction, shape of back cavity near constriction, wall compliance, whether the airstream hits a sharp edge like teeth).
Thus, for a voiceless obstruent like [s], source and transfer functions combine as in Figure 21.

Noise source (supralaryngeal except for [h])

Transfer function
(for the [s] vocal tract shape shown above)

Radiation function  (+6 dB per octave)

Output spectrum
7.2 Perturbation theory

You can estimate the general shape and range of the vowel quadrilateral using perturbation theory. This predicts the effect on formant frequencies of squishing a tube (recall the electrolarynx and tube demo). The comparison is with the formant frequencies of the unconstricted tube. There are two basic principles:

- if you constrict a tube at a place along its length where there is a minimum in a standing wave of pressure, then the frequency of the corresponding resonance will fall, relative to its frequency in the unconstricted tube.
- conversely, if you constrict a tube at a place along its length where there is a maximum in a standing wave of pressure, then the frequency of the corresponding resonance will rise, relative to its frequency in the unconstricted tube.

The main principles can be summarised as follows:

Since most vowels can be adequately modelled using only the lowest two formants (relative to a fairly constant F3), then you can constrict the tube in different places such that you create all 4 possible patterns of change relative to their values in the unconstricted tube (schwa):

- $F_1$ falls and $F_2$ falls in frequency
- $F_1$ falls and $F_2$ rises
- $F_1$ rises and $F_2$ falls
- $F_1$ rises and $F_2$ rises

In a 17 cm vocal tract, $F_1$ for schwa is about 500 Hz ($c/4L$), and $F_2$ is about $3(500) = 1500$ Hz. So, constricting the tube will produce changes that shift $F_1$ away from 500 Hz (up or down) and $F_2$ away from 1500 Hz (up or down).

Figure 22.

\[
\Delta \text{ is the standard symbol meaning “change in” the variable being measured. In Figure 22 and hereafter, } \Delta F \text{ means the change in frequency of a formant above (+) or below (-) a reference frequency, which in this case is the formant frequency that would be the case in an unconstricted tube.}
\]
Figure 23. Effects on formant frequencies of changes in vocal-tract shape away from schwa

Making a **local constriction** in a tube of uniform cross-sectional area has the following results, relative to the natural frequencies of schwa (the unconstricted tube):

- Constriction at a pressure minimum (points x above) → decrease in natural frequency relative to that of schwa (the unconstricted condition)
- Constriction at a pressure maximum (points y above) → increase in natural frequency
- Widening the tube at these places causes the opposite change in natural frequency (an increase if widened at a pressure minimum; a decrease if widened at a pressure maximum)
- Changes in cross-sectional area midway between a pressure maximum and a pressure minimum (points z above) cause very little change in natural frequency.

Changes in cross-sectional area at velocity maxima and minima cause the opposite effects from at pressure max and min: e.g. constraining at a velocity maximum cause the natural frequency to decrease.

So, for example, a constriction at the front of the tube will decrease all formant frequencies relative to those of schwa (the unconstricted condition).

Higher formant frequencies change little because their pressure maxima and minima are closely spaced in the vocal tract, and, since most vocal-tract constrictions extend 1-2 cm, both maxima and minima tend to be affected by changes in cross-sectional area, leading to a net change of zero.
### 7.2.1 LIMITS ON THE VOWEL QUADRILATERAL USING PERTURBATION THEORY

F1 and F2 frequencies change most (they have widely-spaced maxima and minima). Four ways that F1 and F2 can change: both can go up; both down; F1 up and F2 down; F1 down and F2 up. Each of these changes leads to a different quality of sound. By making the largest possible changes, we can approximate the four extreme vowels of the vowel quadrilateral.

---

#### Figure 24

<table>
<thead>
<tr>
<th>Change from vowel space in</th>
<th>Spectrum</th>
<th>Vocal tract mid-sagittal section</th>
<th>Tongue body</th>
<th>Vowel</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1 F2</td>
<td>dB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>↑ ↓</td>
<td><img src="#" alt="Spectrum" /></td>
<td><img src="#" alt="Vocal tract" /></td>
<td>open back</td>
<td>[a]</td>
</tr>
<tr>
<td>↓ ↑</td>
<td><img src="#" alt="Spectrum" /></td>
<td><img src="#" alt="Vocal tract" /></td>
<td>close front</td>
<td>[i]</td>
</tr>
<tr>
<td>↓ ↓</td>
<td><img src="#" alt="Spectrum" /></td>
<td><img src="#" alt="Vocal tract" /></td>
<td>close back</td>
<td>[u]</td>
</tr>
<tr>
<td>↑ ↑</td>
<td><img src="#" alt="Spectrum" /></td>
<td><img src="#" alt="Vocal tract" /></td>
<td>open front</td>
<td>[a]</td>
</tr>
</tbody>
</table>

---

In the diagram at the bottom, F2 cannot be above the specified limits, and F1 must always be less than F2.
Standing waves of pressure

Change in natural frequencies as a function of the place at which the tube is constricted. Changes are shown relative to natural frequency of unconstricted tube, represented as zero.

<table>
<thead>
<tr>
<th>Change from</th>
<th>Spectrum</th>
<th>Vocal tract</th>
<th>Tongue body</th>
<th>Vowel</th>
</tr>
</thead>
<tbody>
<tr>
<td>school ( F_1 ) ( F_2 )</td>
<td>( \Delta F_1 )</td>
<td></td>
<td>( \Delta F_2 )</td>
<td>( \Delta F_3 )</td>
</tr>
<tr>
<td>( \uparrow ) ( \downarrow )</td>
<td>( \uparrow )</td>
<td>( \Delta F )</td>
<td>( \Delta F )</td>
<td>( \Delta F )</td>
</tr>
<tr>
<td>( \downarrow ) ( \uparrow )</td>
<td>( \downarrow )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \Delta F )</td>
<td>( \Delta F )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \Delta F )</td>
<td>( \Delta F )</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Distance of the place of constriction from the glottis

M4_0809_AcousticTheorySpeechProduction.doc
The resultant frequencies of such constrictions can be calculated fairly precisely, but for our purposes it is enough to understand that, by making a constriction in the appropriate location along the tube, you can shift the formant frequencies of schwa so that you hear another vowel quality, as we did with the electrolarynx and tube, and that the limits of such shifts for a given vocal tract length define the limits of the vowel quadrilateral.

Other vowels can be added using various principles, for example:

- “filling in” spaces between the extremes
- nasalization (normally broadens and flattens spectral prominen ces in F1 region, tending to make most nasalized vowels sound slightly centralized)
- changing the lip rounding
- changing duration, voice quality, etc

As you learn more about the acoustics of speech sounds, this will begin to make a lot of sense. (E.g. you know that lip rounding will lower formants. Think where a lip-rounded vowel will be in the quadrilateral relative to one that is identical but has spread lips.)
8. **Quantal theory: using acoustic principles to predict why some sounds are more common than others in languages**

**Figure 26**
Part of schematic nomogram showing the change in frequency of lowest two resonance frequencies of a two-tube model for vowels, as a function of length of back cavity.

1) At certain frequencies, F1 is associated with the back cavity and F2 with the front cavity, and vice versa. The same principle applies to higher pairs of formants (e.g., F2 and F3).

2) Formants cannot intersect; when close in frequency, they “push each other apart” (at region x).

**Quantal Theory** (Stevens 1989):
Regions where formants “push each other apart” are among the regions from which languages favour choosing sounds. In these regions, fairly large variations in the length of the front and back cavities result in only small changes in the frequency of a formant, or pair of formants. Thus, a certain amount of articulatory “slop” can be tolerated, without compromising the acoustic output.

Proposed quantal sounds include:
- Vowels in which two formants are close in frequency (e.g., high front vowels, high back rounded vowels, low vowels)
- Consonants where two formants are close in frequency during the transitions into and out of the consonant (for velars, F2 and F3 – the “velar pinch”. For pharyngeals, F1 and F2. For retroflexes, normally F3 and F4.)
- Consonants produced with turbulence at a constriction (e.g., fricatives). The amplitude of the noise source remains fairly constant over a wide range of constriction degrees (0.03 – 0.2 cm²). Frication can be maintained stably as the articulators move into and out of a constriction.
- Voice qualities: modal, breathy and pressed

Quantal Theory is based on solid acoustic theory and modelling (though for a limited range of contexts). There is support for some of its predictions from articulatory investigations. However, it is controversial, and is not complete as an account of sound systems:
- Some “quantal” sounds are rare across languages (e.g., retroflexes), and there are common sounds that the theory doesn’t explain (e.g., alveolars).
- Quantal Theory is narrow, whereas speech is a multifaceted system that probably involves competing constraints. Lindblom (1989) argues that Quantal Theory wrongly focuses on those sounds that contrast maximally with others; instead, the emphasis should be on “sufficient contrast”, combined with “economy of effort”.

Nonetheless, Quantal Theory may explain certain basic contrasts in speech sounds well, and is worth developing.
9. Two special sounds

9.1 [h]
- A glottal fricative.
- Source: aperiodic (turbulence noise)
- Transfer function: excites the whole vocal tract, therefore full formant structure
- F1 may have a wider bandwidth, and be slightly higher than with voiced equivalent, because the trachea can be coupled in to the system
- Can be thought of as a voiceless vowel.
- Also functions as aspiration noise at the release of voiceless stops

9.2 [ɹ] and other retroflex sounds
- Made with the tongue tip raised towards the palate.
- When this small structure makes a short constriction just at a pressure minimum in a standing wave, it should lower the higher formants (F3, F4 and higher). Higher formants are normally unaffected by longer constrictions, because the standing wave maxima and minima are close together, so both are equally affected by longer constrictions (see perturbation theory).
- Lip rounding will further lower all formant frequencies.
- (But, though this works well in outline, in reality it is much more complicated, at least for [ɹ].)
Reading

Standard – the place to start
The following are relatively nontechnical accounts of acoustics for phoneticians. They cover the same material with different approaches and emphases. Read at least one—whichever suits you best. Others are on your supervision reading list.

Rosen, S. and Howell, P. (1991). *Signals and Systems for Speech and Hearing*. Academic Press. Clearly written and relatively nontechnical, but much more detail than the above books. Try it if you enjoy more scientific approaches. Relevant sections are spread over several chapters—find them in Table of Contents & Index.

CD ROM. *Speech Production and Perception 1*. Sensimetrics. Available from MML Library front desk, for use in CALL lab. Good for self-paced, hands-on work in a number of areas—try it out.

Advanced If you are very interested, or if your maths or physics is A level or better, then try some of these:
Ask me for advanced notes on mathematical basis for resonance.

Quantal theory
If you get interested in this area, ask me for pointers towards new, advanced material.

Reading waveforms
All textbooks discuss spectrogram reading to some extent. Some useful short papers on reading waveforms:

Electronic resources
A standing wave demo: [http://www.walter-fendt.de/ph11e/stlwaves.htm](http://www.walter-fendt.de/ph11e/stlwaves.htm)
Another one, using styrofoam pellets in a tube

See also excellent demos from music re e.g. violin making. Try the YouTube demo on [http://chambermusictoday.blogspot.com/2007/05/violin-physics-chladni-patterns.html](http://chambermusictoday.blogspot.com/2007/05/violin-physics-chladni-patterns.html) and playing the saw (which exploits standing waves): [www.sawlady.com/whatis.htm](http://www.sawlady.com/whatis.htm)

(N.B. I can demonstrate the reality of pressure maxima and minima associated with standing waves by placing a microphone in a tube that is open at one end, and closed at the other end by a loudspeaker. With the loudspeaker emitting a sine wave at one of the tube’s resonance frequencies, you can move the microphone down the tube, and watch/listen to the amplitude of that frequency changing dramatically. If interested, ask me to do it in a lab.)