

## Problem Set 3 Solutions

### Problem 1

(a)

Decibels (dB) are defined by the following relation:

$$\text{dB} = 10 \cdot \log(\text{power}), \text{ equivalently, } \text{power} = 10^{\frac{\text{dB}}{10}}$$

In order to estimate the difference in overall sound pressure levels of the two vowels, we first need to calculate the total power in all of the harmonics of each spectrum in Figure 1.2, and then convert the total power back to dB units. We can ignore harmonics whose amplitudes are more than 20dB below the highest peak of the spectrum.

	Catsoft		Catloud2	
Harmonic	Amplitude (dB)	Power	Amplitude (dB)	Power
1	40	10000	55	316227
2	47	50119	55	316227
3	46	39811	55	316227
4	43	19953	68	6309573
5	56	398107	53	199526
6	54	251189	56	398107
7	36	3981	56	398107
8	40	10000	61	1258925
9	38	6310	45	31623
10	42	15849	45	31623
11	48	63096	49	79433
12	50	100000	54	199526
13	40	10000	49	79433
14	36	3981	-	-
15	35	3162	-	-
16	38	6310	-	-
17	39	7943	-	-
<b>Total</b>		<b>999811</b>		<b>9934557</b>

Catsoft: 999811 => 60 dB

Catloud2: 9934557 => 70 dB.

The difference in dB of overall sound pressure of the two utterance is 10 dB
--

(b)

There are several differences between how the two utterances are produced:

**Fundamental frequency:** The harmonics of “catloud2” are wider apart than those of “catsoft”, indicating the louder utterance having a higher fundamental frequency. This can also be seen from the period of the glottal waveform. When we speak loudly, we tend to tense up the vocal folds.

**Amount of high frequency components:** The spectrum of “catloud2” has more high frequency components than “catsoft”. The glottis closes more abruptly for louder utterances, which is indicated by the steeper slope in the glottal waveform. This translates to more energy in the high frequency range of the louder utterance.

**Intensity:** The intensity of loud speech is higher because a higher subglottal pressure is used.

**Frequency of F1:** F1 is higher for the loud version, suggesting that the mouth opening is wider, possibly with a lower jaw.

**Bandwidth of F1:** The first formant of the loud utterance is much more prominent than that of “catsoft”, which means less acoustic loss and smaller bandwidth for “catloud2”. This makes sense because with louder speech, the open quotient is smaller, resulting in less acoustic loss.

## Problem 2

(a)

From the problem, we know that:

$$A = 4\text{cm}^2$$

$$l = 15\text{cm}$$

$$l_c = 0.8\text{ cm}$$

$$\rho = 0.00114\text{ gm/cm}^3$$

$$c = 35400\text{ cm/s}$$

(1) when the lips are closed, we have a closed-off tube of length,  $l = 15\text{cm}$

assuming hard walls, resonant frequencies are:  $f = 0, \frac{c}{2l}, \frac{c}{l}$   
 $= 0\text{ Hz}, 1180\text{ Hz}, 2360\text{ Hz}$

Correcting for the effect of walls on F1 by using

$$F1 = \sqrt{(F1')^2 + 180^2} = \sqrt{(0)^2 + 180^2} = 180\text{Hz.}$$

We get  $F1 = 180\text{ Hz}, F2 = 1180\text{ Hz}, F2 = 2360\text{ Hz}$

- (2) After the lip opening has reached an area of  $4 \text{ cm}^2$ , we have a tube with one open end and one closed end, of length  $l = 15.8 \text{ cm}$

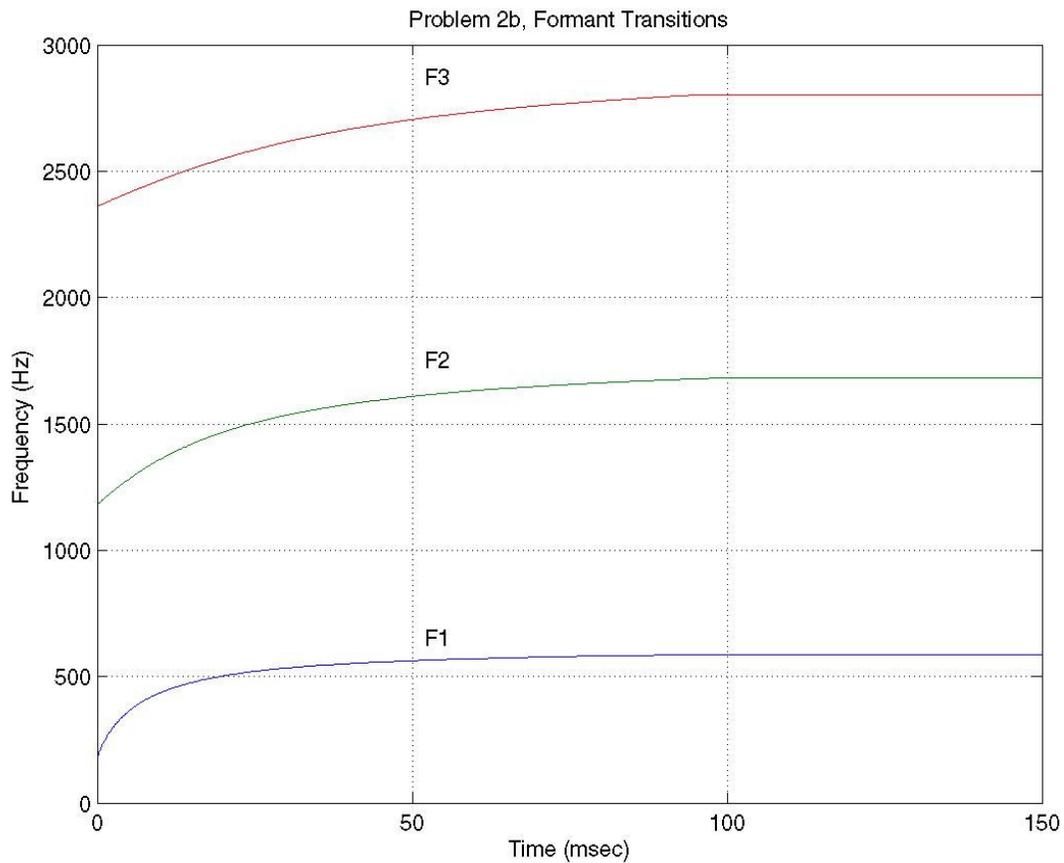
assuming hard walls, resonant frequencies are:  $f = \frac{c}{4l}, \frac{3c}{4l}, \frac{5c}{4l}$   
 $= 560.13 \text{ Hz}, 1680.38 \text{ Hz}, 2800.63 \text{ Hz}$

Correcting for the effect of walls on F1 by using

$$F1 = \sqrt{(F1')^2 + 180^2} = \sqrt{(560.13')^2 + 180^2} = 588.34 \text{ Hz.}$$

We get  $F1 = 588.34 \text{ Hz}, F2 = 1680.38 \text{ Hz}, F3 = 2800.63 \text{ Hz}$

- (b) Matlab script is on the next page. As one can see from the graph, the formants do not transition at the same rate.



```

%Problem Set 3, problem 2B

%parameters

f = 1:1:3000;
rho = 0.00114;
c = 35400;
A = 4;
l=15;
lc = 0.8;

index1 = zeros(1, 1500);
index2 = zeros(1, 1500);
index3 = zeros(1, 1500);

%loop for every 1/10 of a second
for t = 1:1:1500          % time in tenths of msec
if (t<1000), Ac = 0.004*t;    % setting Ac values
else Ac = 4.0; end;

dif = abs((2*pi.*f*rho*lc)/Ac - ((rho*c)/A)*cot((2*pi.*f*l)/c));

    [min1, index1(t)] = min(dif(1:589));      %picking out min. value for F1
    [min2, index2(t)] = min(dif(1180:1681)); %picking out min. value for F2
    [min3, index3(t)] = min(dif(2360:2801)); %picking out min. value for f3

end;

%correct F1 for soft walls

F1_corrected = sqrt(f(index1).^2 + 180^2);

%plot formants vs. time

time = [1:1:1500]/10;
plot(time, F1_corrected, '- ', time, f(index2+1179), '- ', time, f(index3+2359), '- ')

grid on

title('Problem 2b, Formant Transitions');
xlabel('Time (msec)');
ylabel('Frequency (Hz)');
text(52, 650, 'F1');
text(52, 1750, 'F2');
text(52, 2870, 'F3');

```

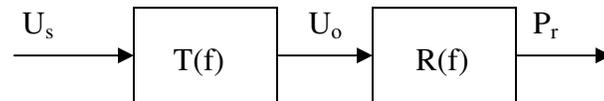
**Problem 3**

We know from the problem that our transfer function of the vocal tract has:

F1 = 500 Hz	F2 = 1500 Hz	F3 = 2500 Hz
B1 = 100 Hz	B2 = 150 Hz	B3 = 200 Hz

(a)

From the source filter theory we know that:



We want to find  $|U_o(f=1000\text{Hz})| = |U_s(f=1000\text{Hz})| |T(f=1000)|$

From Figure 3.1 in the problem set, we know that

$$|U_s(f=1000\text{Hz})| = 10^{\frac{14}{20}} = 5.012 \text{ cm}^3 / \text{s}$$

To calculate  $T(f=1000)$ , recall the transfer function (3.60) from textbook

$$T(f) = \frac{U_o}{U_s} = \frac{1}{\cos kl} = \frac{1}{\cos\left(\frac{2\pi fl}{c}\right)} \quad (1)$$

Since we know  $T(f)$  goes to infinity at the natural frequencies of 500 Hz, 1500 Hz, and 2500 Hz, etc, then  $\frac{2\pi fl}{c} = \frac{\pi}{2}, \frac{3\pi}{2}, \frac{5\pi}{2}, \text{etc.}$  at those frequencies, which gives us,  $l = 17.7$  cm. Plugging  $l$  back into (1), we get

$$|T(f=1000)| = \left| \frac{1}{\cos\left(\frac{2\pi fl}{c}\right)} \right| = \left| \frac{1}{\cos\left(\frac{2\pi(1000)17.7}{35400}\right)} \right| = 1, \text{ since } 1000\text{Hz is a valley.}$$

$$\text{So, } |U_o(f=1000\text{Hz})| = |U_s(f=1000\text{Hz})| = 5.012 \text{ cm}^3 / \text{s}$$

(b)

Using eq. 3.4 from the textbook

$$|R(f)| = \frac{f\rho}{2r} = \frac{1000 \cdot 0.00114}{2 \cdot 20} = 0.0285 \frac{\text{gm}}{\text{sec} \cdot \text{cm}^4}$$

$$p_r(f) = 0.0285 \cdot 5.012 = 0.143 \frac{\text{dyne}}{\text{cm}^2}$$

(c)

We want to find  $|p_r(f=1500\text{Hz})| = |U_s(f=1500\text{Hz})| |T(f=1500)| |R(f=1500)|$

From Figure 3.1 in the problem set, we know that

$$|U_s(f=1500\text{Hz})| = 10^{\frac{4}{20}} = 1.585 \frac{\text{cm}^3}{\text{s}}$$

to calculate  $|T(f=1500\text{Hz})|$ , we use the peak-to-valley relation, where

$$\frac{\text{peak}}{\text{valley}} = \frac{2S}{\pi B} = \frac{2(1000\text{Hz})}{\pi(150\text{Hz})} = 4.244$$

since we know the valley has a magnitude of 1, then  $|T(f=1500)| = 4.244$

similar to part (b),  $|R(f)| = \frac{f\rho}{2r} = \frac{1500 \cdot 0.00114}{2 \cdot 20} = 0.04275 \frac{\text{gm}}{\text{sec} \cdot \text{cm}^4}$

finally, we have  $p_r(f=1500 \text{ Hz}, r=20\text{cm}) = 0.288 \frac{\text{dynes}}{\text{cm}^3}$

(d)

Now that F1 has shifted down to 350 Hz, while the rest of the formants remain the same, recalling that the spectrum at higher frequencies rides on the “skirt” of the component of the transfer function contributed by F1, we can approximate how much the magnitude decreases by for the transfer function at 1000 Hz.

Refer to Figure 3.3 in the textbook, which approximately represents the vocal transfer function given in this problem. If F1 is shifted down from 500 Hz to 350 Hz, the amplitude shown at 1150 Hz will be the amplitude at 1000 Hz when F1 is shifted. We see that F1 at 1150 Hz is about 4dB lower than that at 1000 Hz. We can then re-calculate  $U_o$ .

$$\begin{aligned} |U_o(f=1000\text{Hz})| &= |U_s(f=1000\text{Hz})| |T(f=1000)| \\ &= 10^{\frac{14-4}{20}} = 3.162 \text{cm}^3 / \text{s} \end{aligned}$$

The radiation characteristic remains the same, so we have

$$\begin{aligned} p_r(f=1000\text{Hz}, r=20) &= |U_s(f=1000\text{Hz})| * |R(f=1000\text{Hz})| \\ &= 3.162 \frac{\text{cm}^3}{\text{s}} \cdot 0.0285 \frac{\text{gm}}{\text{s} * \text{cm}^4} \end{aligned}$$

$$p_r(f=1000\text{Hz}, r=20) = 0.0901 \frac{\text{dyne}}{\text{cm}^2}$$

(e)

$$|p_r(f=1500\text{Hz})| = |U_s(f=1500\text{Hz})| |T(f=1500)| |R(f=1500)|$$

$U_s(f=1500\text{Hz})$  and  $R(f=1500\text{Hz})$  are the same as in part(c)

$$|U_s(f=1500\text{Hz})| = 10^{\frac{4}{20}} = 1.585 \frac{\text{cm}^3}{\text{s}}$$

$$|R(f)| = \frac{f\rho}{2r} = \frac{1500 \cdot 0.00114}{2 \cdot 20} = 0.04275 \frac{\text{gm}}{\text{sec} \cdot \text{cm}^4}$$

Refer to Figure 3.3 in the textbook, which approximately represents the vocal transfer function given in this problem. If F1 is shifted down from 500 Hz to 350 Hz, the amplitude shown at 1650 Hz will be the amplitude at 1500 Hz when F1 is shifted. We see that F1 at 1650 Hz is about 2.5dB lower than that at 1500 Hz. Therefore, our vocal tract transfer function's magnitude decreases by 2.5 dB at 1500 Hz.

From part c,  $|T(f=1500)| = 4.244$  equivalent to 12.56 dB before shifting F1. When F1 is shifted, the new  $|T(f=1500)|$  is approximately  $12.56 - 2.5 = 10.06$  dB.

$$|T(f=1500)| = 10^{\frac{10.06}{20}} = 3.184$$

$$p_r(f=1500) = 1.585 \frac{\text{cm}^3}{\text{s}} * 3.184 * 0.04275 \frac{\text{gm}}{\text{sec} \cdot \text{cm}^4}$$

$$p_r(f=1500) = 0.216 \frac{\text{dynes}}{\text{cm}^2}$$