

Topics of this homework: 2-tube transmission line simulation of a vowel sound. Questions and corrections to: [jontalle @ uiuc.edu](mailto:jontalle@uiuc.edu).

Basic Description: In this assignment you are to synthesize four different vowel sounds: /i/ as in eve, /æ/ as in at, /a/ as in father, and IPA symbol /ə/ (“IPA upside-down-e”) as in b/i/rd. You will be using what we have learned about wave models and transmission lines to generate these speech samples. The model will be a 2-tube model of the vocal tract. You also need two boundary conditions. First is the glottal boundary condition consisting of a velocity source in parallel with a flow resistance. Second is a lip boundary condition, consisting of a radiation impedance load (as in HW03).

I would like a copy of the Matlab program that *you* write, the speech files that generate, along with your final report. Your program should be “runable” by me. See below for the details.

Time required: Since this course is on speech processing, the synthesis of speech samples is a rather important topic. In fact, everything we have done up to this point culminates in this homework. Treat this assignment with corresponding respect, by giving it the required time. I estimate that this assignment should take you five to eight hours to complete the simulations, and three to four hours to write the report. The “good news” is that some of the program has already been written, from the previous two homeworks, so if you did a good job on those, this will cut the time significantly. The Matlab program should be your own.

Detailed Description: We first need to lay out the problem by picking a time sampling rate. This choice will be determined by the degree of quantization along the vocal track that we feel we might tolerate. For example, if we pick too low a sampling frequency, the vocal tract would only be one sample long. This clearly would not be enough. If we pick the sampling rate too high, we could suffer from long simulations and numerical error.

Vowel	Section	Length [cm]	Area [cm ²]
/e/ve	1	9	8
	2	8	1
/æ/t	1	4	1
	2	13	8
f/a/ther	1	9	1
	2	8	7
b/ə/rd	1	17	6
	2	0	6

Table 1: Table of lengths from Flanagan for the various vowel sounds.

The average male vocal tract is $L = 17$ [cm] long (see Table 1), so we shall begin with this number. Assuming the speed of sound in air, at body temperature (37 degrees C), is 367

m/sec. The time taken is T seconds to traverse the length of the 17 [cm] length of the vocal tract. Take the sampling period $dT = 1/F_s$ to be $T/10$. Round F_s up. Remember that you will need to listen to your speech, at a reasonable sampling rate, once it has been generated, thus a reasonable thing to do is to take 10 times 44.1 kHz and then round the 17 cm length to the nearest sample (make the tube an integer number of samples, slightly different from (less than one sample period) 17 cm..

Once the sampling rate has been established, determine the length (i.e., 1:K) of the wave arrays. Since we are driving the vocal tract with velocity pulses coming from the simulated glottis, it is reasonable to call the forward wave `up` and the retro wave `um`. Let `up(1)` correspond to $u_p(x = 0, t)$, and `up(K)`, correspond to $u_p(x = L, t)$, and the same for `um(1)`.

Glottal boundary condition: In this initial simulation we shall keep the glottis very simple, and treat it as a velocity source in parallel with an acoustic admittance. (In the real vocal tract the area of the glottis is time-varying. This means that the boundary condition is time varying and the reflection coefficient depends on the state of the glottis opening. I suspect this is an important effect, that would dramatically change the quality or timbre of the sound. We will not attempt a simulation of this effect.)

Treat the boundary condition at the glottis as a reflection coefficient $r_g=0.95$. At each time sample, add the glottis volume velocity $u_g(t)$ to the forward going velocity wave `up(1:K)`. Thus at each time sample is:

$$u_p(0, t) = r_g * u_m(0, t) + u_g(t).$$

Glottal velocity $u_g(t)$: To generate the $u_g(t)$ lets assume the fundamental frequency $f_0(t)$ (denoted “the pitch” by speech researchers) is slowly rising from 150 to 200 Hz, and has a slight vibrato to it. (If you don’t know the word “vibrato,” look it up in the dictionary.) Compute an $f_0(t)$ “pitch” signal over the duration of the speech sample. Lets assume that the speech will be D seconds long. How long will your array need to be that defines $f_0(t)$?

Compute this with a vibrato of 8 Hz, having a 0.2 Hz deviation, and with some variation in it, as you might use when you speak, namely

$$f_0(t) = 150 + 100 * t/\text{Duration} + 25 * \sin(2 * \pi * t)/\text{Duration} + 0.2 * \sin(2 * \pi * 20 * t);$$

Your speech sample should be at least $D = 1.5$ [sec].

Plot $f_0(t)$ on linear-linear coordinates, and label your plot. Add this figure to your report, with a figure caption describing what it is. Matlab’s `print` command is used to generate a figure. Since I use L^AT_EX (<http://en.wikipedia.org/wiki/LaTeX>), the command I must use to generate a postscript file is `print -depsc2 FigureName`. Free advice: When writing reports with figures and equations, L^AT_EX works *much* better than MS-WORD. If you don’t believe me, ask a L^AT_EX user. It is free software (which is the best software, since it is written by users).

Now that you have the $f_0(t)$ description, generate the actual glottal waveform using the formula in the code example I put on the website, at

<http://auditorymodels.org/537/Assignments/mfiles/>,
file `VowelStart.m`, and plot a few periods of this waveform.

Verify that the period starts out with a duration of 6.6 ms (150 Hz), given your sample rate (do this by plotting $u_g(t)$, and look to see that the period is correct (use the Matlab command `plot(t,ug)` after defining t and $u_g(t)$). Also plot the spectrum of one glottal pulse, on a log frequency, log amplitude scale (Matlab command `loglog()`). To get the spectrum, use the `fft()` (Matlab command `fft(ug(1:N))`). I have placed a better (faster) fft program (`fast.m`)

and fsst.m) on the website. These ffts assume the time signal is real and therefore only return positive frequencies. This is very a very convenient assumption!

Label your figures. *Do not plot the negative frequencies. I'll take 1 point off every time you do this!* Use [kHz], and log-log plots (log frequency, log amplitude) (the Matlab `loglog()` function).

Lip boundary condition: Next we must define the boundary conditions at the lips. As in HW3, the radiation load at the tragus was assumed to be the parallel of two impedances sm and r , where s is the complex (Laplace) frequency, m is the mass and r is the resistance. Use a lip area taken from the above table, section 2 Don't forget to convert to MKS units. Find the corresponding radius $a = \sqrt{A/\pi}$, and then use this to find the radiation mass m and resistance r (as in HW2). Let $r = 0.459\rho_0c/a^2$ [mks acoustic ohms] and $m = 0.27\rho_0/a$ [kg/m⁴] [Beranek *Acoustics* (1954), page 121], where a is the radius at the lips, $\rho_0 = 1.18$ [kg/m³] and $c = 367$ [m/s] at body temperature.

Find the bilinear transform of this parallel combination of impedances, as in HW2, to design a filter that implements the boundary condition, given the radiation load impedance.

Do this all before you start writing your code, and fully document your choice of parameters, giving the formula you used to get these values. Do this before you start writing your code! Why am I repeating this? Start writing your report before you start writing the code.

You will find the discussion of the bilinear transform given in Oppenheim and Schaffer's book *Digital signal processing* useful, or you might look in your lecture notes for 410. Not everyone may know how to use the bilinear transform, which corresponds to the "trapezoidal integration rule" used in numerical analysis (Oppenheim and Schaffer, page 207).

The full procedure is to start from the function $R(s)$ of complex frequency s , and to do the following substitution

$$s \rightarrow 2F_s \left(\frac{1 - z^{-1}}{1 + z^{-1}} \right).$$

Fully write out the coefficients of the lip boundary condition filter, namely find $[b_0, b_1]$ and $[a_0, a_1]$ for the filter having the form

$$R(z) = R(s) \Big|_{2F_s \left(\frac{1-z^{-1}}{1+z^{-1}} \right)} \equiv \frac{b_0 + b_1 z^{-1}}{a_0 + a_1 z^{-1}}.$$

The denominator is typically normalized such that $a_0 = 1$.

Given these definitions, what will the recursion relationship look like in the time domain? Put your detailed analysis of the lip boundary condition in your report (5 pts here). Discuss in a short paragraph what would be necessary to do if one were to consider the time varying aspect of the lip area. I am not looking for a full analysis of this problem, just your level of awareness, of how the problem might change, if one were to fully account for such a time varying boundary condition.

Tube model: The model of the vocal tract is a two tube approximation. Therefore we need boundary conditions between the two tubes, which depends on the wave variable, velocity or pressure. In my *lecture notes on wave models* I give these conditions. I also derived these conditions in class. If you do not understand this part, TALK TO ME. Since we are using velocity waves $u_p(x, t)$ and $u_m(x, t)$, corresponding to forward and backward traveling waves, we must use velocity reflection coefficients which slightly differ from pressure reflection coefficients.

Feel free to reproduce the figure from my notes in your report, if you wish. Split the 17 cm tube into two sections. Quantize on a sample point. Simulate 4 different vowels [i/ as in eve, /æ/ as in at, /a/ as in father, and /ə/ “upside-down-e” as in bird], as shown in Table 1.

What to compute: Start by computing the pressure impulse response at the lips ($x=L$) for an impulse of volume velocity at the glottis. Describe how you calculate the pressure from $u_p(x=L, t)$ and $u_m(x=L, t)$.

Take the FFT of the pressure impulse response and plot it on log-log coordinates. Only plot the positive frequencies. I recommend you use `fast.m`, I have provided. Label all the coordinates. The x axis (abscissa) should be in kHz, and the y axis (ordinate) should be in [mks acoustic ohms]. In the figure caption note the formant frequencies for each of the cases. Compare these to the known frequencies, from the famous graph of F1 and F2 from Peterson and Barney JASA (1952) Table II, page 183. Assume male speech.

Next drive the model with your glottal velocity source $u_g(t)$, and save the lip pressure. Convert this to a wave file, and listen to it. Make a plot of 100 ms of the waveform, from the middle of the sound, and make a spectrogram of 0.5 secs of each sound.

Summarize and conclude, about the work you did and what you learned.

Your report: Start this exercise by writing out your assumptions in your report. In other words, *begin your report before you start to code*. Describe what it is that you plan to do, and determine all the parameters and numerical values in the introduction. Thus I will know what you have done. Better yet, you will know what you have done. This introduction should contain a detailed description of the problem, your assumptions (I have largely dictated these to you), and all the numerical values you plan to use in the simulations. Provide your code by sending it to me as an email attachment (zip or tar). I will grade this assignment 20% for its completeness and form and 80% for its content, so both count.

The wave model exercises from last week were a warmup exercise for this one.

The text book by Rabiner and Schafer “Digital processing of speech” (Prentice–Hall) has relevant material which you might find helpful.