# ECES-631 Project
# Discrete-Time Models for the Speech Signal
# Due: December 4th, 2014

## 1 Introduction

One of the most fruitful areas of application of digital signal processing is in the processing of speech signals. The basis for most digital speech processing algorithms is a discrete-time system model for the the production of the speech waveform. There are many useful models that have been used as the basis for speech synthesis, speech coding and speech recognition algorithms. One such model is depicted in Figure 1. The purpose of this project is to show how such a model can be related to a specific speech waveform.

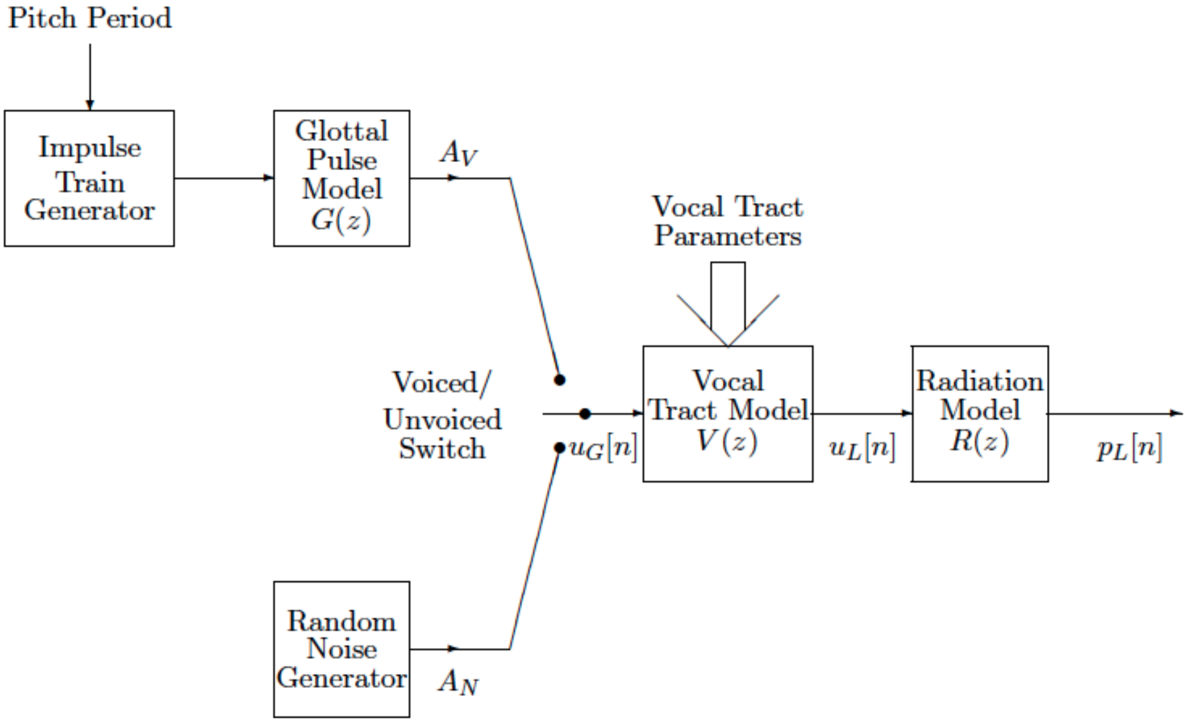


Figure 1: Discrete-time system model for speech production.

### Background Reading

The following references provide appropriate background for this project.

(a) G. Fant, *Acoustic Theory of Speech Production*, Mouton, The Hague, 1970.

(b) T. F. Quatieri, *Discrete-Time Speech Signal Processing*, Prentice-Hall, Inc., 2002.

(c) L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals*, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1978.

(d) A. E. Rosenberg, "Effect of glottal pulse shape on the quality of natural vowels", *Journal of Acoustical Society of America*, Vol. 49, No. 2, pp. 583-590, February, 1971.

## 2 Getting Started

You will need to download the file project2_stuff.zip from Bblearn. This file contains the M-files mentioned in the project description.

## 3 Glottal Pulse Models

### Project Description

The model of Figure 1 often underlies our thinking about the speech waveform, and in some cases, such a system is explicitly used as an speech synthesizer. In this part, we will study the part labeled Glottal Pulse Model $G(z)$ in Figure 1.

### Hints

In speech production, the excitation for voiced speech is a result of the quasi-periodic opening and closing of the opening between the vocal cords (the glottis). This is modeled in Figure 1 by the combination of the impulse train generator and the glottal pulse model filter. The shape of the pulse affects the magnitude and phase of the spectrum of the synthetic speech output of the model.

### Exercise 3.1: The Exponential Model

A simple model that we will call the *exponential model* is represented by

$$G(z) = \frac{az^{-1}}{(1-az^{-1})^2} \tag{0.1}$$

Write an M-file to generate Npts samples of the corresponding glottal pulse waveform $g[n]$ and also compute the frequency response of the glottal pulse model. The calling sequence for this function should be

```
[gE,GE,W]=glottalE(a,Npts,Nfreq)
```

where gE is the exponential glottal waveform vector of length Npts, GE is the frequency response of the exponential glottal model at the Nfreq frequencies W between 0 and $\pi$ radians. You will use this function later.

## Exercise 3.2: The Rosenberg Model

Rosenberg [4] used inverse filtering to extract the glottal waveform from speech. Based on his experimental results, he devised a model for use in speech synthesis, which is given by the equation

$$g_R[n] = \begin{cases} \frac{1}{2}[1 - \cos(\pi n/N_1)] & 0 \le n \le N_1 \\ \cos[\pi(n - N_1)/(2N_2)] & N_1 \le n \le N_1 + N_2 \\ 0 & \text{otherwise} \end{cases} \tag{0.2}$$

This model incorporates most of the important features of the time waveform of glottal waves estimated by inverse filtering and by high-speed motion pictures.

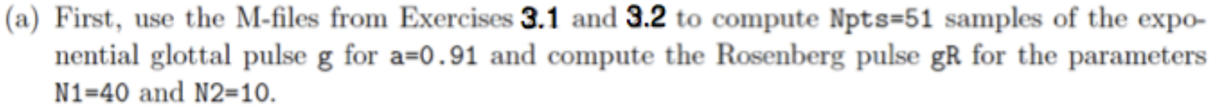
Write an M-file to generate all $N_1 + N_2 + 1$ samples of a Rosenberg glottal pulse with parameters $N_1$ and $N_2$, and compute the frequency response of the Rosenberg glottal pulse model. The calling sequence for this function should be

```
[gR,GR,W]=glottalR(N1,N2,Nfreq)
```

where gR is the Rosenberg glottal waveform vector of length N1+N2+1, GR is the frequency response of the glottal model at the Nfreq frequencies W between 0 and $\pi$ radians.

## Exercise 3.3: Comparison of Glottal Pulse Models

In this exercise you will compare three glottal pulse models.



(a) First, use the M-files from Exercises **3.1** and **3.2** to compute Npts=51 samples of the exponential glottal pulse g for a=0.91 and compute the Rosenberg pulse gR for the parameters N1=40 and N2=10.

(b) Also compute a new pulse gRflip by time reversing gR using the MATLAB function fliplr( ) for row vectors or flipud( ) for column vectors. This has the effect of creating a new causal pulse of the form

$$g_{Rflip}[n] = g_R[-(n - N_1 - N_2)] \tag{0.3}$$

Determine the relationship between $G_{Rflip}(e^{j\omega})$, the Fourier transform of $g_{Rflip}[n]$, and $G_R(e^{j\omega})$, the Fourier transform of $g_R[n]$.

(c) Now plot all three of these 51-point vectors on the same graph using plot( ). Normalize the exponential glottal pulse by dividing by its maximum value before plotting. Also plot the frequency response magnitude in dB for all three pulses on the same graph.

Experiment with the parameters of the models to see how the time-domain wave shapes affect the frequency response.

(d) The exponential model has a zero at $z = 0$ and a double pole at $z = a$. For the parameters N1=40 and N2=10, use the MATLAB function roots( ) to find the zeros of the $z$-transform of the Rosenberg model and also the zeros of the flipped Rosenberg model. Plot them using the MATLAB function zplane( ) or the function zpl( ) that is supplied in project2_stuff.zip. Note that the Rosenberg model has all its zeros outside the unit circle (except one at $z = 0$). Such a system is called a *maximum-phase* system. The flipped Rosenberg model, however, should be found to have all its zeros inside the unit circle, and thus, it is a *minimum-phase* system. Show that in general, if a signal is maximum-phase, then flipping it as in Eq. (0.3) produces a minimum-phase signal and vice-versa.

# 4 Lossless Tube Vocal Tract Models

## Project Description

One approach to modeling sound transmission in the vocal tract is through the use of concatenated lossless acoustic tubes as depicted in Figure 2.

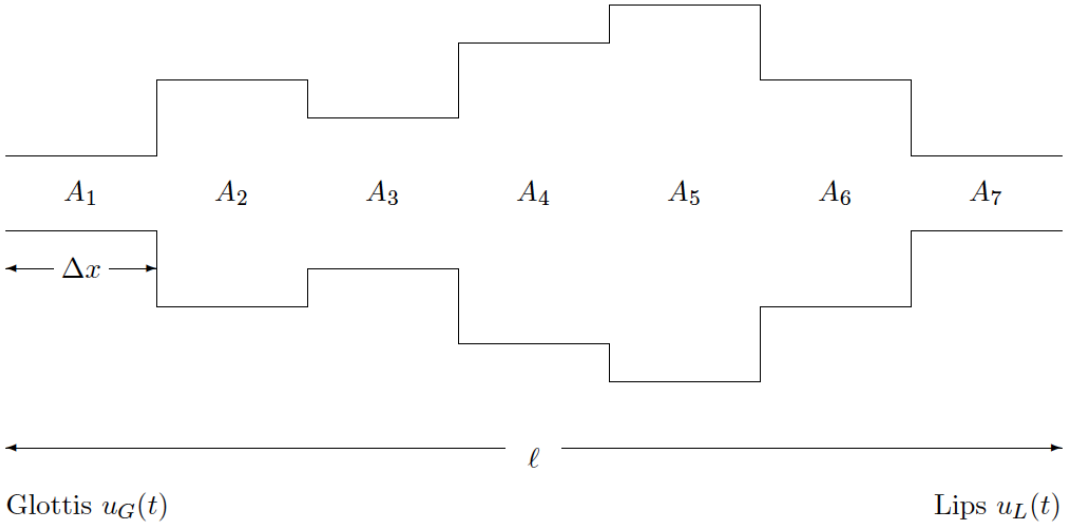


Figure 2: Concatenation of (N=7) lossless acoustic tubes of equal length as a model of sound transmission in the vocal tract.

Using the acoustic theory of speech production [1-3], it can be shown that the lossless assumption and the regular structure leads to wave simple equations and simple boundary conditions at the tube junctions so that a solution for the transmission properties of the model is relatively straightforward, and can be interpreted as in Figure 3(a) where $\tau = \Delta x/c$ is the one-way propagation delay of the sections. For sampled signals with sampling period $T = 2\tau$, the structure of Figure 3(a) (or equivalently Fig. 2) implies a corresponding discrete-time lattice filter as shown in Figure 3(b) or 3(c).[2,3]

## Hints

Lossless tube models are useful for gaining insight into the acoustic theory of speech production, and they are also useful for implementing speech synthesis systems. We have shown that if $r_G = 1$, the discrete-time vocal tract model consisting of a concatenation of $N$ lossless tubes of equal length has system function

$$V(z) = \frac{\prod_{k=1}^{N}(1+r_k)z^{-N/2}}{D(z)} \tag{0.1}$$

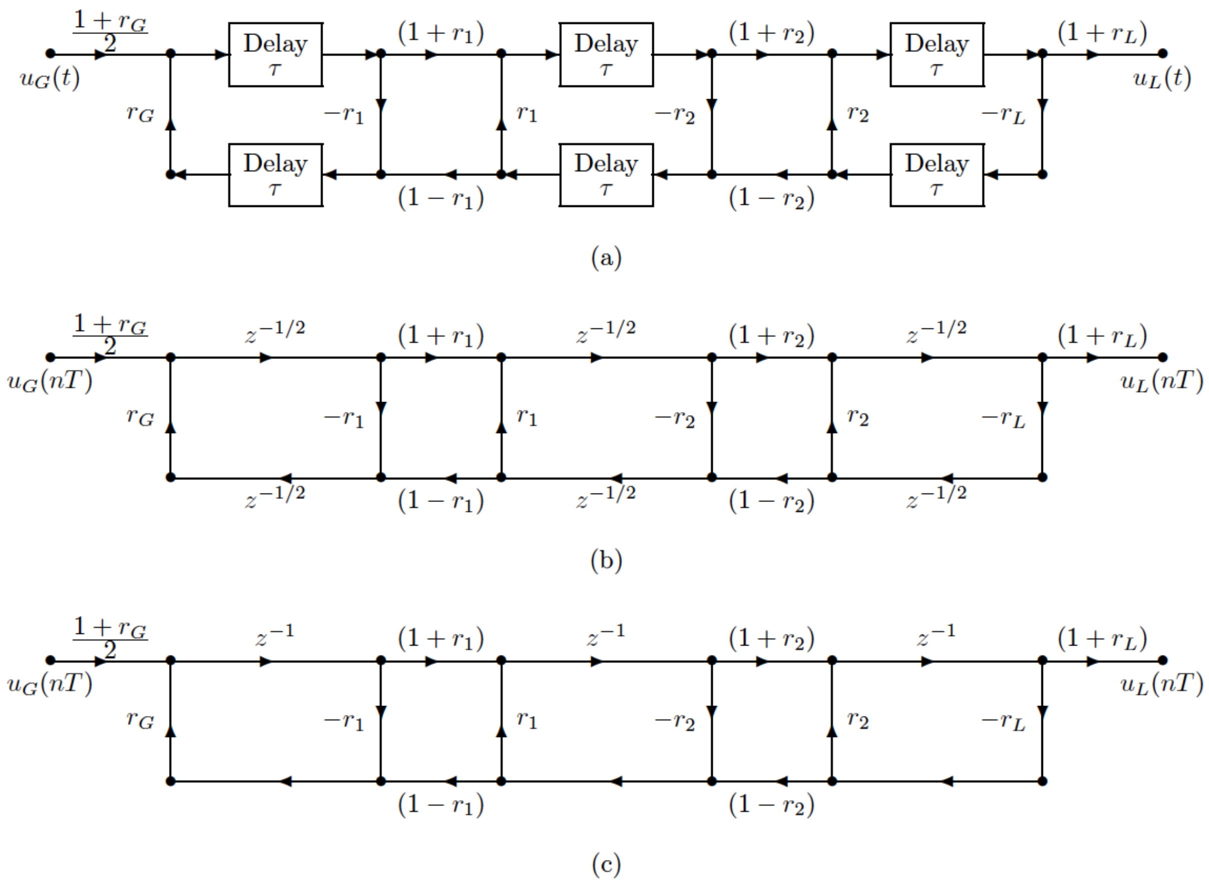


Figure 3: (a) Signal flow graph for lossless tube model ($N = 3$) of the vocal tract; (b) equivalent discrete-time system; (c) equivalent discrete-time system using only whole delays in ladder part.

The denominator polynomial $D(z)$ in Eq. (0.1) satisfies the polynomial recursion [2,3]

$$\begin{aligned} D_0(z) &= 1 \\ D_k(z) &= D_{k-1}(z) + r_k z^{-k} D_{k-1}(z^{-1}) \qquad k = 1, 2, \ldots, N \\ D(z) &= D_N(z) \end{aligned} \tag{0.2}$$

where the $r_k$'s in Eq. (0.2) are the reflection coefficients at the tube junctions,

$$r_k = \frac{A_{k+1} - A_k}{A_{k+1} + A_k} \tag{0.3}$$

In deriving the recursion in Eq. (0.2), it was assumed that there were no losses at the glottal end ($r_G = 1$) and that all the losses are introduced at the lip end through the reflection coefficient

$$r_N = r_L = \frac{A_{N+1} - A_N}{A_{N+1} + A_N} \tag{0.4}$$

where $A_{N+1}$ is the area of an impedance-matched tube that can be chosen to introduce a loss in the system.

Suppose that we have a set of areas for a lossless tube model, and we wish to obtain the system function for the system so that we can use the MATLAB `filter( )` function to implement the model; i.e., we want to obtain the system function of Eq. (0.1) in the form

$$V(z) = \frac{G}{D(z)} = \frac{G}{1 - \displaystyle\sum_{k=1}^{N} \alpha_k z^{-k}} \tag{0.5}$$

(Note that we have dropped the delay of $N/2$ samples, which is inconsequential for use in synthesis and would be impossible to implement when $N$ is odd.) The following MATLAB M-file implements Eqs. (0.2) and (0.3); i.e., it takes an array of tube areas and a reflection coefficient at the lip end and finds the parameters of Eq. (0.5) along with the reflection coefficients.

```
function        [r,D,G]=AtoV(A,rN)
%       function to find reflection coefficients
%       and system function denominator for
%       lossless tube models.
%               [r,D,G]=AtoV(A,rN)
%               rN = reflection coefficient at lips (abs value < 1)
%               A = array of areas
%               D = array of denominator coefficients
%               G = numerator of transfer function
%               r = corresponding reflection coefficients
%       assumes no losses at the glottis end (rG=1).
[M,N]=size(A);
if(M~=1) A=A'; end      %make row vector
N=length(A);
r=[];
for m=1:N-1
        r=[r (A(m+1)-A(m))/(A(m+1)+A(m))];
end
r=[r rN];
D=[1];
G=1;
for m=1:N
        G=G*(1+r(m));
        D=[D 0] + r(m).*[0 fliplr(D)];
end
```

As test data for this project, the following area functions were estimated from data obtained by Fant.[1]

| Section | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| vowel AA | 1.6 | 2.6 | .65 | 1.6 | 2.6 | 4 | 6.5 | 8 | 7 | 5 |
| vowel IY | 2.6 | 8 | 10.5 | 10.5 | 8 | 4 | .65 | .65 | 1.3 | 3.2 |

## Exercise 4.1: Frequency Response and Pole/Zero Plot

(a) Use the M-file **AtoV( )** to obtain the denominator *D(z)* of the vocal tract system function, and make plots of the frequency response for both area functions for `rN=0.71` and also for the totally lossless case `rN=1.`

(b) Factor the polynomials *D(z)* and plot the poles in the z-plane using **zplane( )** or **zpl( )**. Convert the angles of the roots to analog frequencies corresponding to a sampling rate of *1/T=1000 samples/sec.*, and compare to the format frequencies expected for these vowels (see supplementary figures 1-3 at the end of this document). For this sampling rate, what is the effective length of the vocal tract in cm?

## Exercise 4.2: Problem on making a model from a system function

Suppose that we know the system function (transfer function) of a discrete-time model of the vocal tract; i.e.,

$$V(z) = \frac{G}{D(z)} = \frac{G}{1 - \sum_{k=1}^{N} \alpha_k z^{-k}}$$

We may wish to obtain the areas and reflection coefficients for a lossless tube model given $D(z)$. Figure 4.18 in Quatieri shows the flow graph of such a model for $N = 2$. For the case $r_G = 1$ and $r_N = r_L$, the denominator of the system function ($D(z)$) satisfies the following equations (given near the middle of p. 147 in Quatieri):

$$\begin{aligned} D_0(z) &= 1 && (5) \\ D_k(z) &= D_{k-1}(z) + r_k z^{-k} D_{k-1}(z^{-1}), \qquad k = 1, 2, \ldots, N && (6) \\ D(z) &= D_N(z) && (7) \end{aligned}$$

If we are given the reflection coefficients (equivalently the tube areas) for the model, these equations can be interated to obtain $D(z)$. In this problem we will use these equations to develop an algorithm for finding the reflection coefficients and the areas of a lossless tube model having a given $D(z)$.

**(a)** Show that $r_N$ is equal to the coefficient of $z^{-N}$ in the denominator of $V(z)$; i.e. $r_N = -\alpha_N$.

**(b)** Use the above equations to show that

$$D_{k-1}(z) = \frac{D_k(z) - r_k z^{-k} D_k(z^{-1})}{1 - r_k^2} \qquad k = N, N-1, \ldots, 2$$

**(c)** How would you find $r_{k-1}$ from $D_{k-1}(z)$?

**(d)** Using the results of parts (b) and (c), state an algorithm for finding all of the reflection coefficients $r_k$, $k = 1, 2, \ldots, N$ and all of the tube areas $A_k$, $k = 1, 2, \ldots, N$. Are the $A_k$'s unique?

## Exercise 4.3: Finding the Model from the System Function

The inverse problem arises when we want to obtain the areas and reflection coefficients for a lossless tube model given the system function in the form of Eq. (0.5). We know that the denominator of the system function, *D(z)*, satisfies Eq. (0.2). In this part, we will use Eqs. (0.2) to develop an algorithm for finding the reflection coefficients and the areas of a lossless tube mode having a given system function. Parts (a)-(c) are covered in Exercise 4.2.

(a) Show that $r_N$ is equal to the coefficient of $z^{-N}$ in the denominator of $V(z)$; i.e. $r_N = -\alpha_N$.

(b) Use Eqs. (0.2) to show that

$$D_{k-1}(z) = \frac{D_k(z) - r_k z^{-k} D_k(z^{-1})}{1 - r_k^2} \qquad k = N, N-1, \ldots, 2$$

(c) How would you find $r_{k-1}$ from $D_{k-1}(z)$?

(d) Using the results of parts (a), (b), and (c), state an algorithm for finding all of the reflection coefficients $r_k$, $k = 1, 2, \ldots, N$ and all of the tube areas $A_k$, $k = 1, 2, \ldots, N$. Are the $A_k$'s unique? Write a MATLAB function to implement your algorithm for converting from $D(z)$ to reflection coefficients and areas. This M-file should as defined by the following:

```
function        [r,A]=VtoA(D,A1)
%       function to find reflection coefficients
%       and tube areas for lossless tube models.
%        [r,A]=VtoA(D,A1)
%               A1 = arbitrary area of first section
%               D = array of denominator coefficients
%               A = array of areas for lossless tube model
%               r = corresponding reflection coefficients
%       assumes no losses at the glottis end (rG=1).
```

For the vowel AA, the denominator of the 10th-order model should be (to 4 digit accuracy)

$$\begin{aligned} D(z) &= 1 - 0.0460z^{-1} - 0.6232z^{-2} + 0.3814z^{-3} + 0.2443z^{-4} + 0.1973z^{-5} \\ &\quad +0.2873z^{-6} + 0.3655z^{-7} - 0.4806z^{-8} - 0.1153z^{-9} + 0.7100z^{-10} \end{aligned}$$

Use your MATLAB program to find the corresponding reflection coefficients and tube areas and compare to the data for the vowel AA in the table above.

# 5 Vowel Synthesis

## Project Description

For voiced speech, the speech model of Fig. 1 can be simplified to the system of Fig. 4. The exitation signal $e[n]$ is a quasi-periodic impulse train and the glottal pulse model could be either the exponential or the Rosenberg pulse. The vocal tract model could be a lattice filter of the form of Fig. 3(c) or it could be an equivalent direct form difference equation as implemented by MATLAB's `filter( ) or \verb`conv( )+ function.

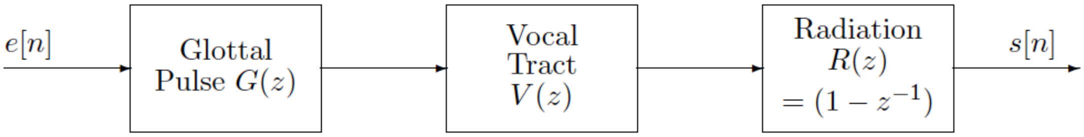


Figure 4: Simplified model for synthesizing voiced speech.

## Hints

In this project we will use the MATLAB `filter( )` and `conv( )` functions to implement the system of Fig. 4 and thereby synthesize periodic vowel sounds.

### Exercise 5.1: Periodic Vowel Synthesis

Assume a sampling rate of 10000 samples/sec. Create a periodic impulse train vector `e` of length 1000 samples with period corresponding to a fundamental frequency of 100 Hz. Then use either `filter( )` or `conv( )` to implement the system of Fig. 4.

Use the excitation `e` and radiation system $R(z) = (1 - z^{-1})$ to synthesize speech for both area functions given above, and for all three glottal pulses studied in Project 2. Use `subplot( )` and `plot( )` to make a plot comparing 1000 samples of the synthetic speech outputs for the exponential glottal pulse and the Rosenberg minimum-phase pulse. Make another plot comparing the outputs for the two Rosenberg pulses.

### Exercise 5.2: Frequency Response of Vowel Synthesizer

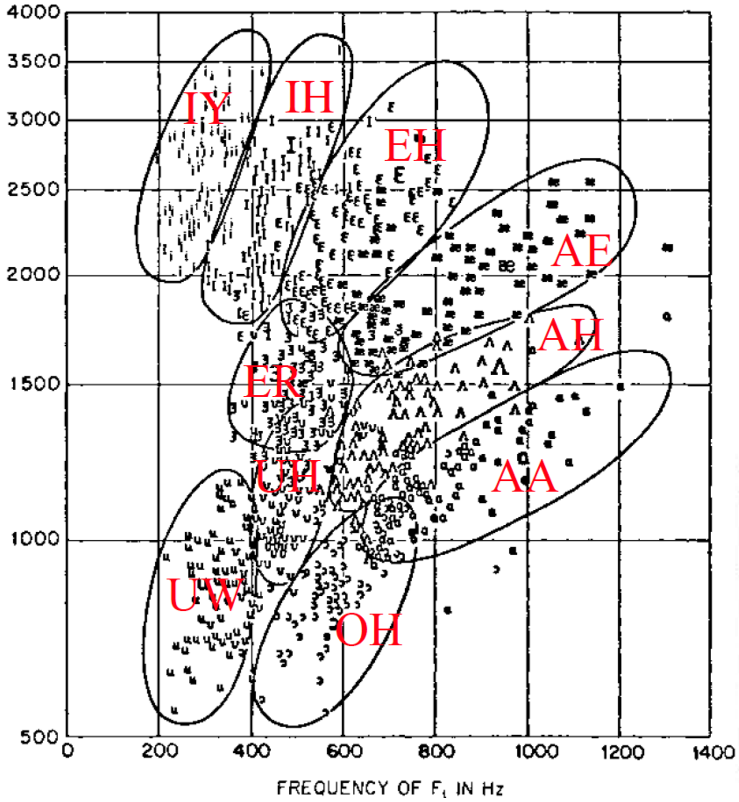
Plot the frequency response of the overall system with system function $H(z) = G(z)V(z)R(z)$ for the case of the Rosenberg glottal pulse, $R(z) = (1 - z^{-1})$, and vocal tract response for the vowel IY.

### Exercise 5.3: Listening to the Output

Create a file of length corresponding to 0.5 sec. duration and play it out through the D/A system using MATLAB's `soundsc( )` function. Does the synthetic speech sound like the desired vowels?

## 6 Report

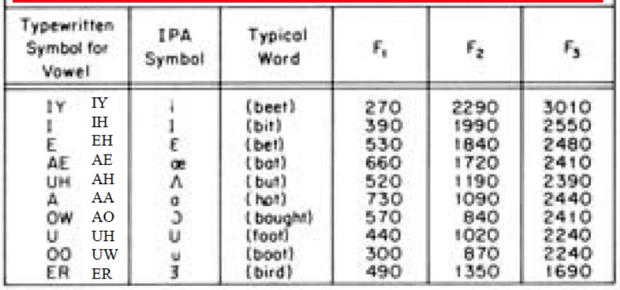
Submit a typewritten report including appropriate plots and images to illustrate your work. Learn to include graphics in your report either with LaTeX or MS Word, or whatever you use for this sort of thing. You should structure your report along the lines of the sections of this project assignment. Be sure to answer all the specific questions asked above and provide graphs and MATLAB coded wherever appropriate.



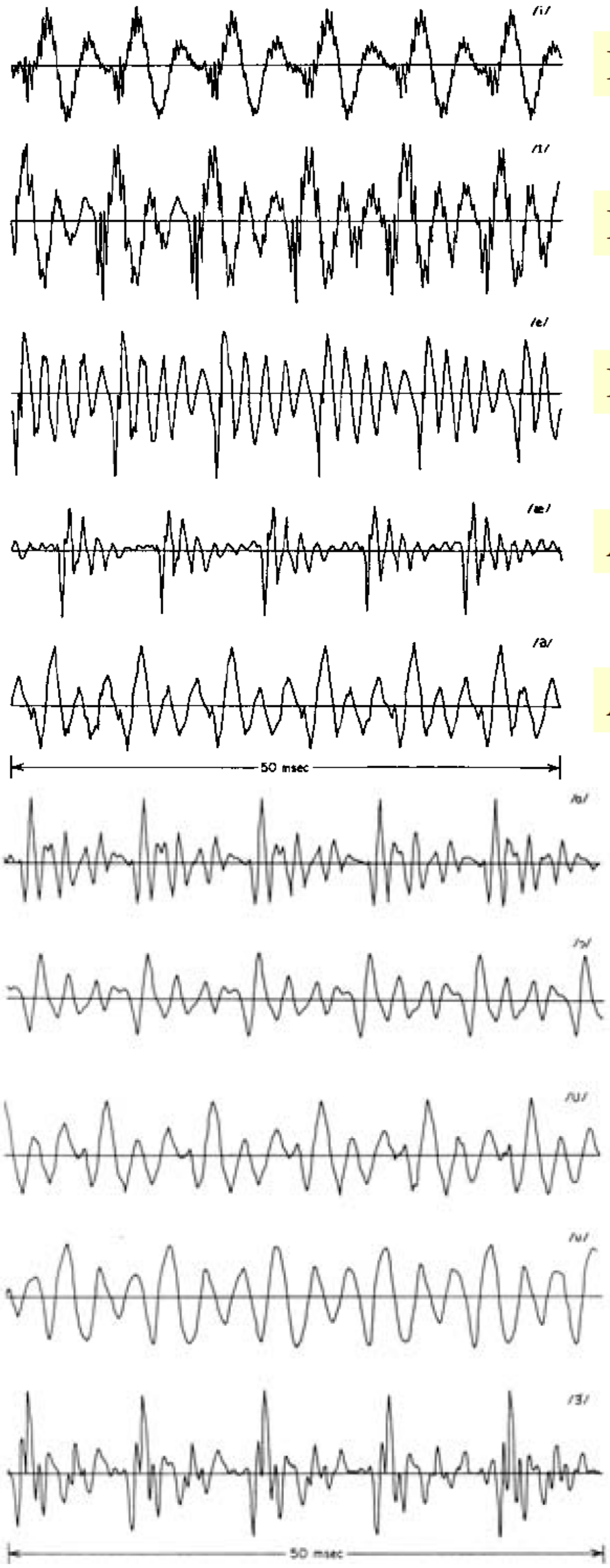
Supplementary Fig. 1:  Formant Frequences of Vowels

Peterson and Barney measured these on spectrograms.

Average Format Frequencies

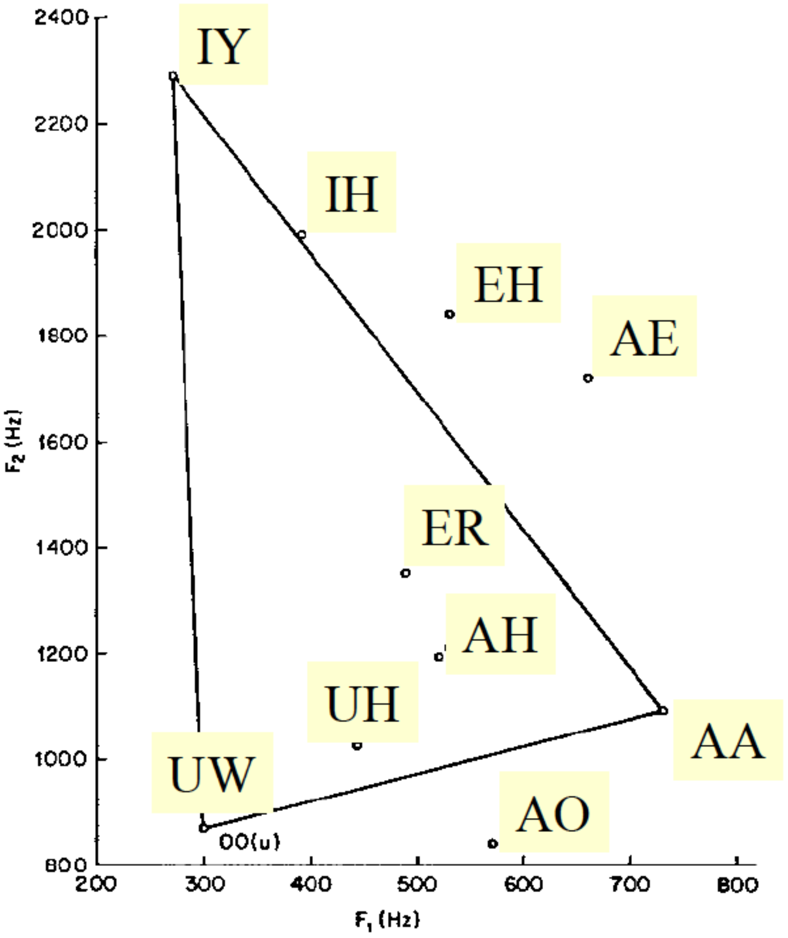


| Typewritten Symbol for Vowel | IPA Symbol | Typical Word | $F_1$ | $F_2$ | $F_3$ |
|---|---|---|---|---|---|
| IY IY | i | (beet) | 270 | 2290 | 3010 |
| I IH | ɪ | (bit) | 390 | 1990 | 2550 |
| E EH | ɛ | (bet) | 530 | 1840 | 2480 |
| AE AE | æ | (bat) | 660 | 1720 | 2410 |
| UH AH | ʌ | (but) | 520 | 1190 | 2390 |
| A AA | ɑ | (hot) | 730 | 1090 | 2440 |
| OW AO | ɔ | (bought) | 570 | 840 | 2410 |
| U UH | ʊ | (foot) | 440 | 1020 | 2240 |
| OO UW | u | (boot) | 300 | 870 | 2240 |
| ER ER | ɝ | (bird) | 490 | 1350 | 1690 |



| | |
|---|---|
| IY | $F_1$=270, $F_2$=2290, $F_3$=3010 |
| IH | $F_1$=390, $F_2$=1990, $F_3$=2550 |
| EH | $F_1$=530, $F_2$=1840, $F_3$=2480 |
| AE | $F_1$=660, $F_2$=1720, $F_3$=2410 |
| AH | $F_1$=520, $F_2$=1190, $F_3$=2390 |
| AA | $F_1$=730, $F_2$=1090, $F_3$=2440 |
| AO | $F_1$=570, $F_2$=840, $F_3$=2410 |
| UH | $F_1$=440, $F_2$=1020, $F_3$=2240 |
| UW | $F_1$=300, $F_2$=870, $F_3$=2240 |
| ER | $F_1$=490, $F_2$=1350, $F_3$=1690 |

Supplementary Figure 2: Vowel Waveforms and first 3 formants of each vowel.



Supplementary Figure 3: The Vowel Triangle (F1 and F2 are formant frequencies)