

**SPEECH COMMUNICATION**  
**6.541J-24.968J-HST710J**  
**Spring 2004**

Lab 2

Assigned: 03/30/04

Due: 04/06/04

The purpose of this laboratory is to provide some experience in the synthesis of speech with a terminal-analog speech synthesizer. This synthesizer, which is implemented on the computers of the speech group running Linux, consists of an array of sources and filters that simulate the acoustic sources and transfer functions that describe human speech generation.

Your task in this lab is to design a set of synthesis control parameters to synthesize a consonant-vowel syllable. You should first record a phrase like “**Say that CV please**”. The CV can be a syllable like /pa/, /tu/, /sa/, /di/, /ne/, etc. You should analyze the CV part to determine fundamental frequency vs. time, formant tracks, noise spectra, etc. Using this analysis as a guide, design control parameters of the synthesizer to generate this CV. Then embed your synthesized CV in the original phrase in place of the original CV. If you do your job carefully, you should hear no difference between the original phrase and the one with the synthesized CV.

You will need to use several programs to do this lab. The main synthesizer program is accessed from XKL by clicking on “**synthesize**”. This is described in a handout, which contains excerpts of a chapter written by Dennis Klatt. If you are synthesizing a female voice, you should set the sampling rate (SR) to 11,500 Hz. For a male voice, SR should be at the default value of 10,000 Hz. The original utterance *file.wav* can be edited using XKL. This will allow you to divide the utterance into three separate files *saythat.wav*, *CV.wav*, and *please.wav*. When you have finished your synthesis of an utterance, named *xcv.wav*, then you can put it all together with

*concat saythat xcv today xfile*

This will give you a new waveform *xfile.wav* which is a concatenation of the other three.

As a start, you should analyze the file *CV.wav* to determine the timing, the F0 contour, the formant frequencies, and the times when frication noise occurs. This analysis can be assisted by making a spectrogram (*lspecto CV*, followed by *lpr CV.ps*), by making a F0 track (*lspecto CV -f0*, followed by *lpr CV.ps\_f0*) and formant tracks (*lspecto CV -pf*, followed by *lpr CV.ps\_pf*). You can also get nnumbers for F0 and formants by typing *lspecto CV -syn*, followed by *enscript CV.doc*. The numbers that are printed cannot always be trusted, and may need smoothing. (In particular, you should pay no attention to the AV values on this printout.)

(If the schedule can be arranged, you should come to the lab for about 30 minutes a day or two before beginning your synthesis, in order to make your recording and to make hard copies of your analysis of the CV utterance.)

Before trying to synthesize the entire CV, you should attempt to match the overall spectrum shape of the synthetic vowel to the spectrum shape in the middle of the spoken vowel in the file *CV.wav*. This match is achieved by first setting the fundamental frequency, the formant frequencies and the amplitude AV, and then adjusting the bandwidths and the glottal parameters OQ and TL. OQ is usually between 40 and 60 percent, and TL is usually in the range 0-20 dB. The rest of the vowel is then synthesized by entering the formant and F0 parameters as a function of time.

If your CV has a fricative or a stop consonant, this portion of the synthesis is accomplished by entering appropriate time variation for AF (amplitude of frication noise). The spectrum of the noise is shaped by adjusting the amplitude parameters A2F, A3F, A4F, A5F, and AB.

Your lab report should describe your procedures for determining the control parameters, particularly the parameters that gave you the most trouble. The report should give plots or tables of the control parameters you used, and spectrograms of the original and synthesized utterances. You should aim for a length of no more than 4 pages plus figures.

## Commands on Speech Synthesis on XKL

### The Basics

1. To begin synthesis, type **Y** or select **Synthesize** from the pull down menus.
2. For your very first synthesis, choose the **defaults** option. For all subsequent iterations of your synthesis, choose the **.doc file** option and pick the correct file to modify.
3. After you are finished modifying the appropriate parameters, synthesize the file and name it.
4. You can open the new file in XKL and see how well you did? Typing **p** or clicking the middle mouse button will play your utterance.
5. Use some of the spectral analysis techniques learned in lab 1 to help you make changes in your next iteration of synthesis.
6. Synthesis usually takes quite a few iterations before there is little or no audible distinction between the original utterance and your synthesized version.

### Synthesis Commands

All of the synthesis commands are listed and described in the newly opened dialog box. The most useful commands for this lab are as follows:

- c** Changes the default value for a parameter. This value is constant throughout the entire duration of the utterance.
- d** Allows you to draw time-varying values for a parameter.
- s** Synthesizes the changes you have made to the original .doc file and creates a new .doc file and .wav file.

## Parameters

For a list of all of the parameters and their descriptions, refer to the Klatt paper or open any .doc file (using emacs). The more useful parameters are as follows:

<b>DU</b>	Duration of the utterance, in milliseconds
<b>SR</b>	Sampling rate, in samples per second
<b>AV</b>	Amplitude of voicing, in dB
<b>OQ</b>	Open quotient, in percentage
<b>TL</b>	Spectral tilt of voicing, in DB down @ 3 kHz
<b>AH</b>	Amplitude of aspiration, in dB
<b>AF</b>	Amplitude of frication, in dB
<b>NF</b>	Number of formants
<b>FO</b>	Fundamental frequency, in Hz, times 10
<b>F1</b>	Frequency of first formant, in Hz
<b>F2</b>	Frequency of second formant, in Hz
<b>F3</b>	Frequency of third formant, in Hz
<b>F4</b>	Frequency of fourth formant, in Hz
<b>F5</b>	Frequency of fifth formant, in Hz
<b>B1</b>	Bandwidth of first formant, in Hz
<b>B2</b>	Bandwidth of second formant, in Hz
<b>B3</b>	Bandwidth of third formant, in Hz
<b>B4</b>	Bandwidth of fourth formant, in Hz
<b>B5</b>	Bandwidth of fifth formant, in Hz
<b>A2F</b>	Amplitude of frication-excited parallel to the 2 <sup>nd</sup> formant, in dB
<b>A3F</b>	Amplitude of frication-excited parallel to the 3 <sup>rd</sup> formant, in dB
<b>A4F</b>	Amplitude of frication-excited parallel to the 4 <sup>th</sup> formant, in dB
<b>A5F</b>	Amplitude of frication-excited parallel to the 5 <sup>th</sup> formant, in dB