EEEM030 Assignment 1

Assessment: 15% of module

Linear Predictive Speech Synthesizer

Reports should be submitted electronically to SurreyLearn (https://surreylearn.surrey.ac.uk/) by the deadline as specified in SurreyLearn. See the submission method section below for more details.

1 Basics

You are asked to synthesize some vowels (monophthongs) using the source-filter model of speech production. The formant structure of each vowel will be estimated directly from a real speech sample, using linear predictive coding (LPC). Synthesized vowels can then be generated by passing a periodic impulse train through the all-pole filter obtained.

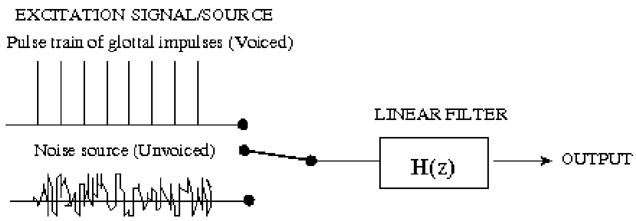

Figure 1: Source-filter model of speech production

It is recommended to use Matlab for this assignment, since it contains many in-built functions for LPC, filter design, audio I/O, etc.. You will need a set of headphones to listen to speech samples.

2 Task

Download a set of male and female vowel phoneme samples in .wav format from:

`http://personal.ee.surrey.ac.uk/Personal/W.Wang/assignment.html`

Choose one male vowel and one female vowel from the sample set, and then choose a quasi-stationary segment of around 100ms in length from each vowel for the LPC estimation.

For each of the vowel segments:

Model Estimation

Use one of the Matlab functions for autoregressive (AR) modelling (e.g. `lpc`, `aryule`, `arcov`; the help files should contain details of each method), to estimate the LPC coefficients that provide a least-squares error fit to the speech waveform. Plot the frequency response of the LPC filter, and include in your report a figure showing the amplitude spectrum of the speech segment (in dB) and the corresponding LPC filter response, as a function of frequency in Hz. The model order should be chosen such that the LPC filter provides a reasonable fit to the formant structure. Estimate the first three formant frequencies of the vowel, and the mean fundamental frequency.

Synthesis

Generate a periodic impulse train (each impulse should be of unit-amplitude and one sample in width) with the same fundamental frequency as the original vowel segment, and of roughly a second in length. This will be the simulated periodic source for voiced sounds in the source-filter model of speech production. Filter the pulse train using the LPC filter determined above, and listen to the synthesized speech. Experiment with different AR model orders and segment lengths. Make an informal subjective assessment of the final synthesized speech quality.

To complete the task, you may find the following Matlab functions useful:

`fft`	Fast Fourier Transform.
`lpc`	Uses autocorrelation method to find LPC coefficients.
`arcov`	Uses covariance method to find LPC coefficients.
`filter`	One dimensional digital filter.
`freqz`	Plots the frequency response of a digital filter.
`plot`	Plot graph.
`abs`	Absolute value of complex number.
`log10`	Logarithm to base 10.
`audioread`	Read in wav file.
`audiowrite`	Write to wav file.
`wavplay`	Play sound using Windows audio output device.
`sound`	Play vector as sound.
`play`	If neither `wavplay` nor `sound` works, try e.g.: `ap = audioplayer(y,fs,16); play(ap);`.

3 Submission method

Submit a written final report (between 1000-2000 words, excluding appendices, in .pdf or .doc format) to SurreyLearn (https://surreylearn.surrey.ac.uk/) or by email to w.wang@surrey.ac.uk. This should contain a brief description of the task, details of the implementation (including any calculations or values of parameters obtained), and a summary of the results (containing graphs of any frequency responses, and an informal assessment of the synthesized speech quality). The appendices should contain any Matlab code implemented by yourself. Along with your report, please also submit synthesized audio samples (e.g. in .wav files).

To log into SurreyLearn, you need to use the same account number and password as you are using to access your Surrey email box.

4 Assessment method

Your assignment mark will be given based on the quality of the report, code, and re-synthesized audio samples. Specifically, we will give marks based on the quality and/or correctness of the following items, each of which will contribute **ten percent** to the final mark of your assignment.

```
M1:  LPC coefficient estimation;
M2:  Plots of filter response and speech amplitude spectrum;
M3:  Formant frequency estimation;
M4:  Mean fundamental frequency estimation;
M5:  Periodic impulse train generation and speech synthesis;
M6:  Evaluations on different AR orders and segment lengths;
M7:  Informal subjective assessment;
M8:  Quality of presentation and references;
M9:  Quality of your programming code;
M10:  Quality of synthesized speech samples.
```