Speech Analysis

V 2.3.1, March 19, 2018
Martin Hagmüller, hagmueller@tugraz.at
Signal Processing and Speech Communication Laboratory
http://www.spsc.tugraz.at/
Graz University of Technology
Inffeldgasse 16c, 8010 Graz, Austria

Abstract

This is the first part of the Speech Communication Laboratory. Its goal is to provide some insight into the methodology of speech analysis. It covers time-domain as well as frequency-domain methods, and Linear Predictive Coding (LPC). Most of the tasks are performed in Praat - a powerful speech processing tool.

Equipment:

- PC with a sound card installed

- Praat software and a copy of source files (these you can download from our website https://www.spsc.tugraz.at/sites/default/files/file/scl/analysis.zip)

- Headsets and necessary cables.

Before you start...

1. You are expected to write a detailed report about your work. The report should be prepared on the fly, as you proceed with the tasks and handed over at the end of the laboratory.

2. Do not forget to answer the questions asked in experiments and those asked by the lab assistant. Your answers should be provided in full, with all the necessary explanations and, if required, plots and tables. Try to explain what you observe! If you can't, do not hesitate to ask your lab assistant!

Experiment 1 – Time-domain analysis

The power of time-domain methods is low computational complexity and simplicity of interpretation. However, they suffer from the increased noise sensitivity and often require a certain amount of synchronization.

(1.1)Short-Time Average Energy (Intensity)

1. Start Praat. In the main menu select $\boxed{\textbf{Read}} \rightarrow \boxed{\textbf{Read From File}}$ and load one of the `*.wav` files from the `speech_samples/sentences` directory. You should see the created sound object in the Praat object browser.

2. Select this sound object and press $\boxed{\textbf{To Intensity...}}$ button. In the dialogue window set ***Minimum pitch*** to 100Hz and press $\boxed{\textbf{Ok}}$. The result will be another object that contains the desired intensity profile.

3. Now select the created intensity object and plot the intensity curve by pressing $\boxed{\textbf{Draw...}}$.

4. In the Praat picture window select the graphical area you used for plotting the intensity profile. Then, select the sound object and press $\boxed{\textbf{Draw...}}$. This should plot the signal waveform overlaid with the corresponding intensity profile.

5. Repeat the previous steps trying ***Minimum pitch*** values from 50Hz to 500Hz. How does it influence the resulting plot? Explain why it happens.

6. What could be the possible application of these computations? Explain your choice.

Experiment 2 – Frequency-domain analysis

The frequency domain deliver the most useful parameters for the speech processing. Human hearing appears to pay much more attention to spectral properties of the speech rather than to phase/timing aspects.

(2.1)Wide- and narrow-band spectrograms

1. From the Praat main menu select $\boxed{\textbf{Read}} \rightarrow \boxed{\textbf{Read from file}}$ and load the `a_8000.wav` sound file from the `speech_samples/sounds` directory.

2. In the object browser select the corresponding sound object and press $\boxed{\textbf{Edit}}$ button. This will create a new analysis window.

3. In the created window select $\boxed{\textbf{Spectrum}} \rightarrow \boxed{\textbf{Spectrogram settings}}$. In the dialogue window set ***Window length*** to 30msecPress $\boxed{\textbf{Ok}}$ button and carefully study the spectrogram. What is the source of the spectrogram ripples?

4. Change ***Window length*** to 5msec and recompute the spectrogram. What has changed, compared to the previous case, and why?

5. Perform the above steps for the `s_8000.wav` file. What has changed in this case? **Hint**: recall that the sound *"a"* is voiced, unlike the clearly unvoiced sound *"s"*.

(2.2) f_0 **and formants**

1. Recall, what is the physical meaning of formants and a fundamental frequency f_0.

2. From the Praat object browser select `a_8000` sound and press $\boxed{\textbf{Edit}}$ button.

3. Select $\boxed{\textbf{Formant}} \rightarrow \boxed{\textbf{Formant settings}}$ and set ***Number of formants*** to 4, ***Maximum formant*** to 4000Hz, and ***Window length*** to 30msec. If necessary, set a check mark on $\boxed{\textbf{Formant}} \rightarrow \boxed{\textbf{Show formants}}$. You should see red dots that mark the formant traces. What are the values of F_1, F_2, F_3 and F_4 for the given signal?

4. Now, select $\boxed{\textbf{Pitch}} \rightarrow \boxed{\textbf{Pitch settings}}$. Set the ***Pitch range*** between 75Hz and 400Hz, and press $\boxed{\textbf{Ok}}$. If necessary, set a check mark on $\boxed{\textbf{Pitch}} \rightarrow \boxed{\textbf{Show pitch}}$. You should see a blue line that marks the f_0 trace (the corresponding values are depicted on the right side of the plot). How high is the values of f_0?

5. Repeat the same analysis for the `u_8000.wav` and `e_11025.wav` files from the `speech_samples/sounds` directory. Fill in the following table:

Sound	f_0	F_1	F_2	F_3	F_4
/a:/					
/u:/					
/e:/					

(2.3) **Formant and** f_0 **behaviour for different sounds**

1. Using Praat perform similar analysis as in the previous task for the following sounds:

   - Diphthongs: `ai_8000.wav`, `ou_8000.wav`
   - Liquids: `l_8000.wav`
   - Nasals: `n_8000.wav`
   - Fricative unvoiced: `s_8000.wav`
   - Fricative voiced: `z_8000.wav`

- Stops voiced: `aba_8000.wav`
- Stops unvoiced: `ata_8000.wav`

2. Describe the behavior of formants and f_0 in these cases?

Experiment 3 – Linear Predictive Coding (LPC)

Linear prediction analysis takes a crucial role in speech processing algorithms. The LPC coefficients make up a model of the vocal tract that produced the original speech signal, thus allowing us to analyse speech more effectively.

(3.1) LPC models

1. If necessary, load the `a_8000.wav` sound file from the `speech_samples/sounds` directory and select the corresponding sound object from the object browser.

2. Press $\boxed{\textbf{Analyse spectrum \& LPC}} \rightarrow \boxed{\textbf{To LPC (autocorrelation)}} \rightarrow \boxed{\textbf{Standards}}$. This will create a new LPC-object, constructed from the selected signal.

3. Select the created LPC object and press $\boxed{\textbf{To Spectrum (slice)}}$. In the dialogue window set ***Time (seconds)*** to the middle of the signal duration (0.3sec might be a good guess). The resulting object contains a single slice of the spectrum computed from the LPC coefficients at the selected time instance ***Time (seconds)***. You can plot it by selecting the corresponding object and pressing the $\boxed{\textbf{Draw...}}$ button.

4. Where do the peaks in the spectrum come from? **Hint**: recall that the LPC coefficients model the vocal tract of a speaker.

(3.2) LPC residual signal

1. Now, in the object browser select simultaneously the created LPC object and the corresponding sound signal. Press $\boxed{\textbf{Filter (inverse)}}$. The newly created object represents the reconstructed glottal excitation sequence. Press $\boxed{\textbf{Edit}}$ button to plot it. Each outbreak in the sequence corresponds to the glottal excitation pulse.

2. Measure the distance between two consecutive outbreaks, i.e., the pitch period. Is it consistent with the f_0 estimate for the original sound computed by Praat?