Exercise #2:
Basic signal features, linear prediction, analysis-synthesis

İlknur Baş
151226814
01.04.2023

Task 1: Spectrogram, energy, and zero-crossing rate

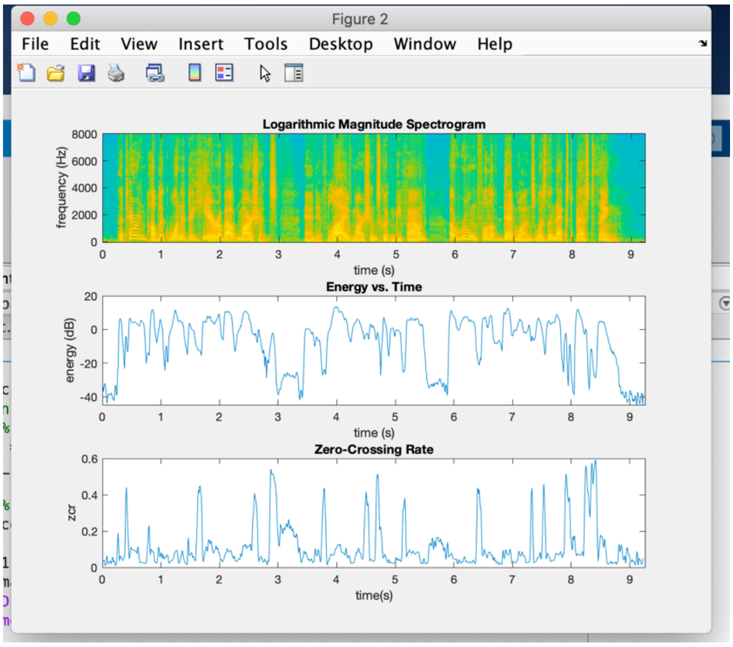

Figure 1: The plots of a spectrogram (top), energy envelope (middle) and zero-crossing rate (bottom) for the given speech sample.

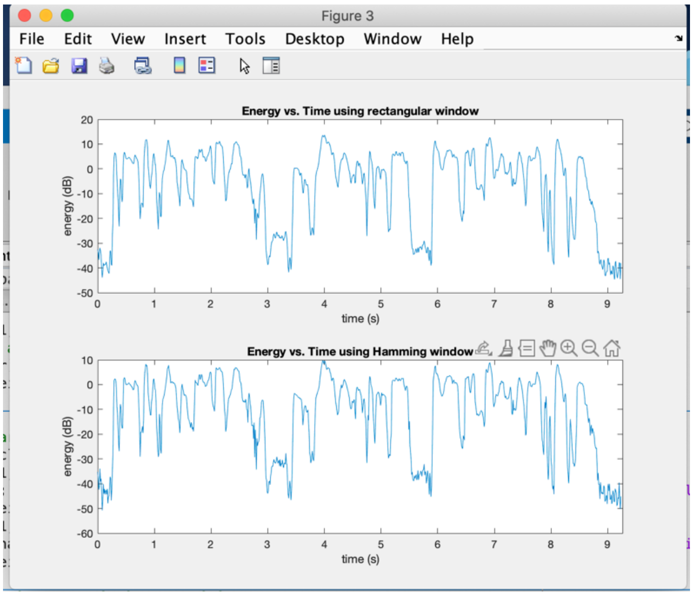

Figure 2: The plots of energy envelope when rectangular window (top) and Hamming window (bottom) is applied

Q1.1:

Firstly, when we split the given signal into short segments, the borders of each segment are seen as discontinuities, which is an unwanted behavior, compared to real-world signal [1]. For that reason, the windowing functions is applied to those segments. Windowing functions (can be referred as smoothing functions) go to zero at the border; meaning that in the case of multiplication of it with a particular segment, we won't be observing discontinuities at the border since the borders in the resulting multiplication will go to zero as well.

The assignment requires to use rectangular window for the calculation of energy and zero-crossing rate. If we were to think what happens if Hamming window etc. is used, as I explained above, I would expect the border of each segment to go zero. This will happen in a case where the first couple sample amplitudes are positive. We won't be observing any zero-crossing since the sign won't be changed. Assume, in a specific segment, the first couple sample amplitudes are negative. When multiplication is performed, similar behavior is expected, sign remains the same, so again no sign changes are detected. Lastly, assume zero-crossing is detected at the middle of the frame and Hamming window is applied, the zero crossing remains the same. We could try to think different cases like these. Yet, I have also tried this in MATLAB code and zero-crossing rate stays the same for each frame, which make sense. Also, when we visualize both windowing functions, hamming and rectangular, we see that in both there is zero value at the edge, but hamming reaches that zero value smoothly. This is also good indicator of why zero-crossing rate remains same.

The use of non-rectangular window in calculation of energy affects the resulting energy. I have also added the plot of comparison in MATLAB code. It is observed that the energy is decreased when the Hamming window is applied. It makes sense as we smoothen out the borders of a signal (a frame of a signal), and as a result, the signal at the edges do not contribute much anymore. That is why energy decreases.

All my assumptions above are done according to Hamming window, the use of different window type might affect the results differently.

Task 2: Linear prediction

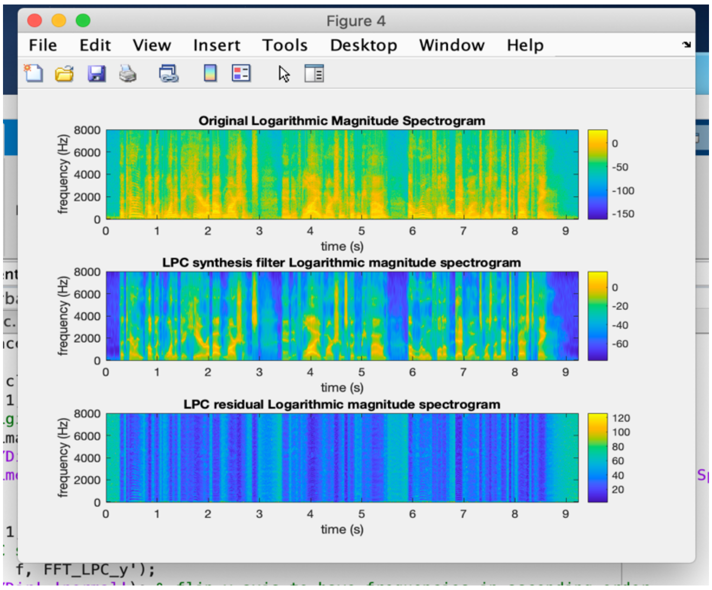
Figure 4
File Edit View Insert Tools Desktop Window Help
Original Logarithmic Magnitude Spectrogram
frequency (Hz)
8000
6000
4000
2000
0
0 1 2 3 4 5 6 7 8 9
time (s)
0
-50
-100
-150
LPC synthesis filter Logarithmic magnitude spectrogram
0
-20
-40
-60
LPC residual Logarithmic magnitude spectrogram
120
100
80
60
40
20

```
f, FFT_LPC_y');
```

Figure 3: The plots of original magnitude spectrogram of the utterance (top), logarithmic magnitude spectrogram of LPC filters (middle), and logarithmic magnitude spectrogram of the LPC residual (bottom) when LPC order is 20

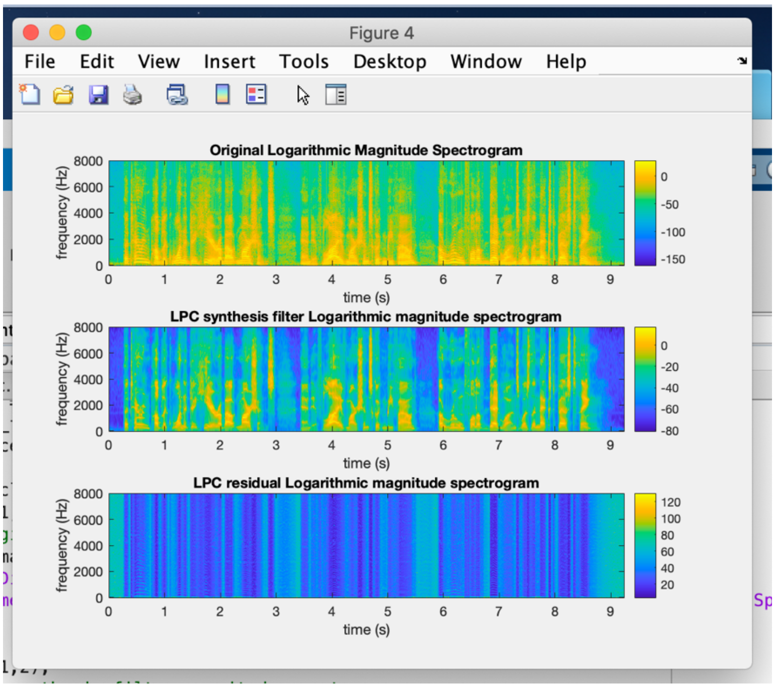
Figure 4
File Edit View Insert Tools Desktop Window Help
Original Logarithmic Magnitude Spectrogram
frequency (Hz)
8000
6000
4000
2000
0
0 1 2 3 4 5 6 7 8 9
time (s)
0
-50
-100
-150
LPC synthesis filter Logarithmic magnitude spectrogram
0
-20
-40
-60
-80
LPC residual Logarithmic magnitude spectrogram
120
100
80
60
40
20

Figure 4: The plots of original magnitude spectrogram of the utterance (top), logarithmic magnitude spectrogram of LPC filters (middle), and logarithmic magnitude spectrogram of the LPC residual (bottom) when LPC order is 40

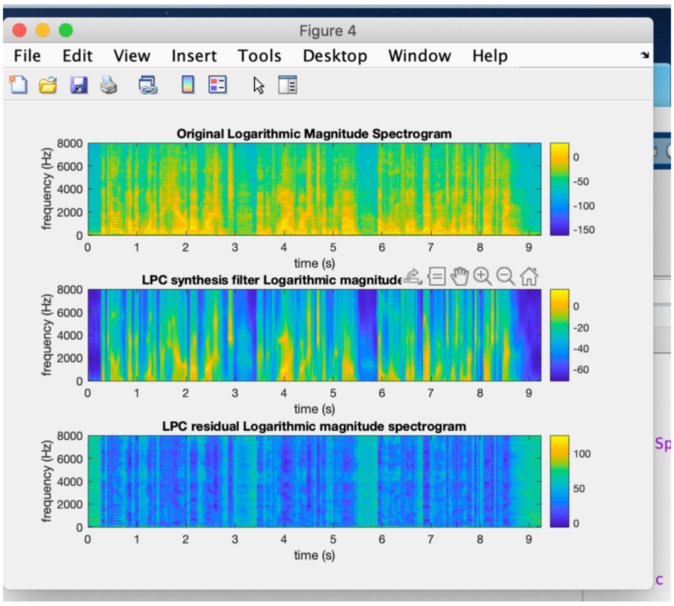
Figure 4
File Edit View Insert Tools Desktop Window Help
Original Logarithmic Magnitude Spectrogram
frequency (Hz)
8000
6000
4000
2000
0
0 1 2 3 4 5 6 7 8 9
time (s)
0
-50
-100
-150
LPC synthesis filter Logarithmic magnitude
frequency (Hz)
0
-20
-40
-60
time (s)
LPC residual Logarithmic magnitude spectrogram
frequency (Hz)
100
50
0
time (s)

Figure 5: The plots of original magnitude spectrogram of the utterance (top), logarithmic magnitude spectrogram of LPC filters (middle), and logarithmic magnitude spectrogram of the LPC residual (bottom) when LPC order is 5

Q2.1:

Firstly, in the lectures, we have learned that LPC analysis breaks down the signal into spectral envelope and the residual components. LPC spectrogram asked in the assignment shows the spectral envelope of the original signal, and from the spectral peaks, we could find the formant structure. And the residual part gives information that cannot captured by the spectral envelope, in other words the part where cannot be modeled by LPC. For example, we could do F0 estimation using residual. Hence, we could say that they provide complementary information which results in original signal as it can see from the Figure 3.

In Figure 3, when we look at the top plot, we would see lower energy (very little yellow coloring) in certain time steps, since they indicate silences in the speech. We see the same pattern in middle plot. Also in middle plot, there are high energy in lower frequencies. I think it is because lower frequencies carry much information about the speech since they are closer to the fundamental frequency. On the other hand, in bottom plot, we observe the lowest energy (compared to top and middle plots) throughout the speech. It makes sense because residual shows the part that is not captured by the LPC coefficients. In the lectures, we even said that most of the time there is noise in the residual.

As the LPC order increases, we expect to see more accurate result when it comes to shape of spectral envelope, meaning that we would get more detailed information about the formants, peaks etc., Hence, the spectrogram plots will be shown as sharper. The reason for all these is that as the LPC order increase we are considering more previous samples when predicting. (Refer to Figure 3 and 4, where LPC order is 20 and 30 respectively) In fact, we are not seeing much difference, but it is a little bit visible that the spectrogram got sharper. I have also attached Figure 5, LPC order is 5, where the difference between plots is much more observable. I think when we increase LPC order a lot, we will encounter with overfitting problem. That is why we don't see many changes LPC order increase after 30-35.

I have seen similar patterns in residual when the LPC order increases. There is not much visible difference in Figure 3 and 4, where LPC order is 20 and 30 respectively, but we could say that the spectrogram got sharper, smoother, and have less fluctuations. This can be easily observed when the plot is compared Figure 5, LPC order is 5 (even though it is not asked in the assignment). I think, again, the reason we don't see much observable difference is that, as the LPC order increase a lot, we encounter with overfitting problem. I also think the reason we see sharper, smoother and not much fluctuated spectrogram for residual is that spectral envelope gets very accurate and capture most of the information (from both low and high frequencies) as the LPC order increases. That means there is not much left for residual to capture.

Task 3: Signal re-synthesis with LPC

Q3.1:

In my case, speech quality in the residual-based resynthesized signal was better than the impulse-based resynthesized signal. It makes sense as the residual represents the information that cannot captured by the spectral envelope, in other words the part where cannot be modeled by LPC. To synthesize the waveform for each frame, we make use of LPC coefficients and excitation signal, which is residual. The predicted LPC coefficients will keep the spectral envelope of the original signal and in the case where residual is used, excitation signal will include the remaining information which is not captured by LPC model/vocal track. As a result, we get a better-quality resynthesized speech. On the other hand, the impulse-based resynthesized signal has less natural sound, more like a robotic. It also makes sense because we define the excitation as a periodic signal that consists of T time internals.

This creates limitation and as a result we get worse quality compared to the residual-based resynthesized signal.

Q3.2:

First of all, as I explained above sections, LPC coefficients/vocal track gives information about the spectral envelope which also gives information about formants (applying LPC analysis). We could make use of this formant information in other speech tasks. Also, during the exercise session I have learned that vocal track shape is different for each individual and these differences are captured by predicted LPC coefficients. So, I think by analyzing those, we could also get information about the characteristics of the speaker much easier. Similarly, we could analyze the excitation signal/voiced sounds as well. Also, as we implemented in current section, we could make use of this separation in the application of speech synthesis by using different excitation signals. Lastly, we can also make use of this separation in speech coding domain in an efficient way. In general separation makes each task in speech processing domain much easier and efficient.

REFERENCES

[1] "Windowing" *Aalto University Wiki*, wiki.aalto.fi/display/ITSP/Windowing.