dipartimento di ingegneria unipa

Università degli Studi di Palermo

Drill 5

Problem 1

Develop a Python program to carry out Analysis-Synthesis using Linear Prediction. Specifically, work with the waveform file *samplespeech.wav*, which is provided in the class materials. Your program should load the .wav file and process it frame-by-frame using an LPC (Linear Predictive Coding) order of 24. Once the analysis is completed, your program has to reconstruct the waveform (synthesis). After completing the synthesis, save the reconstructed speech signal. Finally, play both the original and the processed audio using the appropriate Python commands, and briefly note your observations on how they sound.

Note: Which would be the most appropriate window to use?

Problem 2

You are required to carry out a frame-by-frame analysis. In particular consider an unvoiced and a voiced frame, namely frames 0 and 166, and compare the magnitude of the frequency response of the estimated linear prediction filter with the magnitude of the Fourier
Transform of the speech frame. Use the following values for the LPC orders: 6, 12, 24, 48, and 96.

You should observe that when the LPC order is very high—for example, 96—the magnitude spectrum of the LP filter closely matches almost all the peaks of the Fourier transform of the speech frame. In voiced frames, all formants are enveloped, and in unvoiced frames, the spectral peaks are fully covered. This shows that high LPC orders can effectively capture the entire magnitude spectrum. However, this level of detail is not ideal; we prefer an LPC envelope that captures the general shape of the spectrum without fitting every small peak. An LPC order of 24 demonstrates this balance well and serves as a good example of an optimal case. Conversely, with very low orders such as 6, the LP filter fails to adequately model the spectral shape for both voiced and unvoiced frames. In summary, low LPC orders underfit the spectral envelope, while very high orders tend to overfit.

Problem 3

Make modifications in the excitation signal, specifically create a whisper voice - a type of speech produced **without vibration of the vocal cords**. Instead of using vocal fold vibration to generate sound (as in normal voiced speech), whispering relies solely on **turbulent airflow** through a **partially open glottis** to create audible speech.

Problem 4

Make modifications in the vocal tract. Select the most significant (according to their magnitude) three poles which we may assume they correspond to the first three formants. Modify accordingly these formants so that you can get an elderly and a younger voice than the input speaker.

Estimate the formant frequencies of the signal by selecting the three most prominent complex-conjugate pole pairs (i.e., the six poles with the highest magnitudes that are not real). These poles represent the formants -- real poles are not formant. To modify the voice, multiply these selected poles by a scaling factor — typically between 0.8 and 1.2 — which corresponds to shifting the formant frequencies down or up by up to 20%. This adjustment influences the perceived age of the speaker: lowering the formants (scaling closer to 0.8) makes the voice sound older, while raising them (scaling toward 1.2) gives the voice a more child-like quality.

Importantly, the LPC (Linear Predictive Coding) order remains constant throughout. With higher LPC orders (around 25–30), the effect of formant manipulation on perceived age is more pronounced due to the greater number of coefficients available to model the vocal tract. However, this can also introduce artifacts or noise, making the voice sound more "glitchy." On the other hand, with lower LPC orders (around 7–8), formant changes have a subtler effect on age perception but tend to preserve a cleaner speech signal.