Assignment 2 Report

Alexandra Gianni
STUDENT ID: 03382

Exercise 1: LP Analysis-Synthesis

In this exercise i need to work with an audio signal *"sample.wav"* and perform LP analysis and synthesis.
In general, in the analysis section, we extract the reflection coefficients, as known as LP coefficients, from the signal and use it to compute the residual signal. In the synthesis section, we reconstruct the signal using the residual signal and the reflection coefficients.

But, what are reflection coefficients and the residual signal?

- Reflection coefficients (LP coef.): They represent how the signal can be modeled as a linear combination of its previous samples. These coefficients capture the formant structure of the speech and i can get them by performing **Levinson Recursion**
- Residual signal: The prediction error or the part of the speech signal that can't be modeled using the LP coefficients. This signal contains unvoiced sounds, noise, excitation energy. Here, the residual is obtained by inverse filtering the original signal with the LPC filter:
  $ex = filter(a, 1, sigLPC)$

Subsequently, we will briefly mention the steps that we followed in order to complete the LPC exercise.

Analysis Steps

1. Frame the signal: Divide the input speech signal into short overlapping frames ($30ms$ in our case) for analysis and apply a Hanning window to smooth each frame. After frame setup, begin analysis frame-by-frame.
2. Compute Autocorrelation: Calculate the autocorrelation of each frame using MATLAB's *xcorr* function to capture the signal's self-similarity. Each time keep the non-negative lags.
3. Derive LPC coefficients: Use the *Levinson-Durbin* recursion to estimate the Linear Predictive Coding (LPC) coefficients from the autocorrelation values. These coefficients model the vocal tract.
4. Compute the gain: Calculate the LPC filter gain to scale the residual signal based on the LPC model's energy. The gain is equal to the square root of the LPC coefficients and autocorrelation values of the signal.

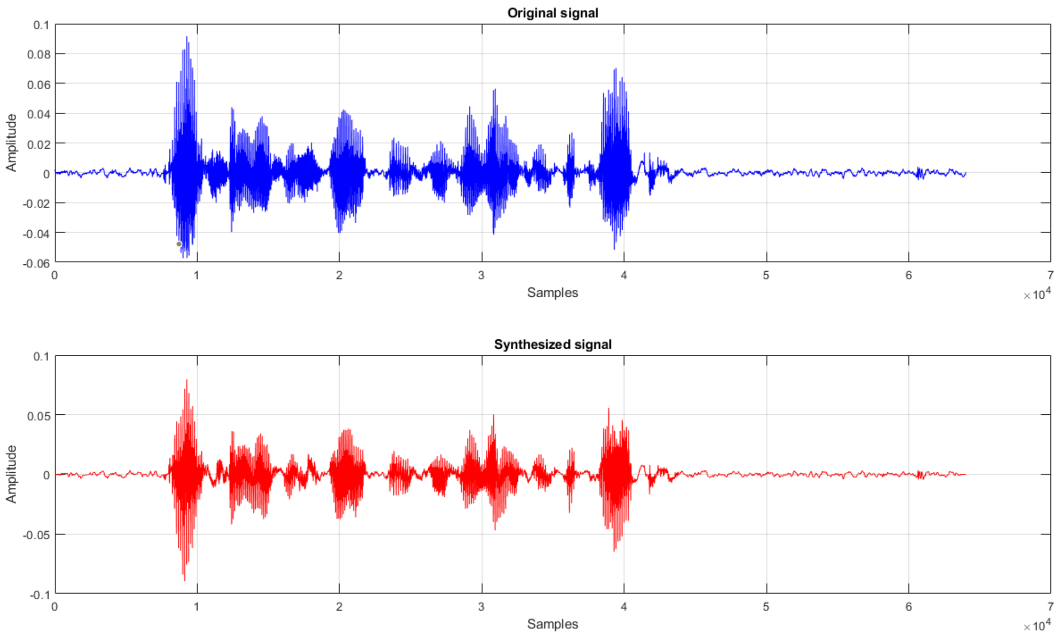
5. Calculate Residual signal: Apply inverse filtering to remove the predictable part of the speech signal, leaving the residual signal. This step isolates the excitation source (glottal pulse or noise).

Synthesis Steps

1. Scale the Residual: Multiply the residual signal by the gain computed during analysis to restore the signal's energy.
2. Reconstruct the Signal: Pass the scaled residual signal through the LPC synthesis filter (defined by the LPC coefficients) to reconstruct the speech signal.
3. Energy Normalization: Normalize the energy of the synthesized signal to match the energy of the original frame.
4. Overlap and Add: Combine overlapping frames to ensure smooth transitions between frames and reconstruct the final output signal.

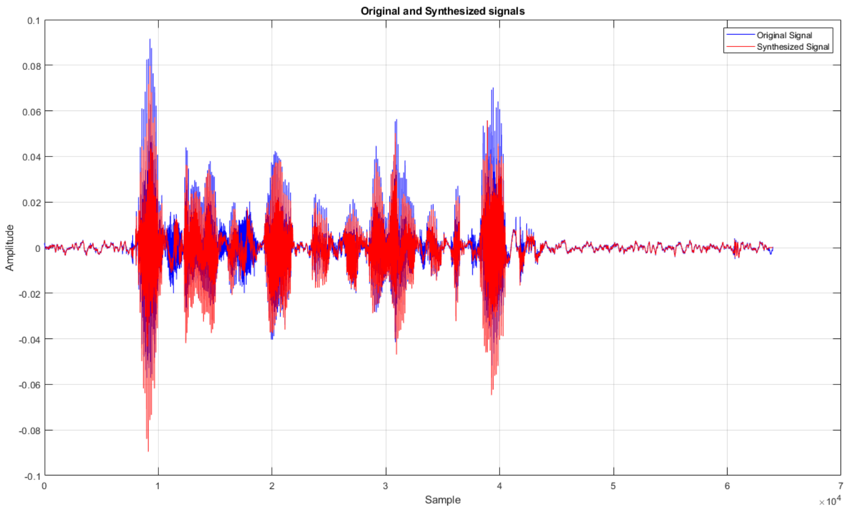
After completing the above steps, we save the processed signal using the *"audiowrite"* command and then we play both the original and synthesized signals to compare their perceptual quality using *soundsc* command.

These are the plots i generate by my code:



Σχήμα 1: Subplot of Original and Synthesized signals

We can observe some similarities between our signals, but of course they are not identical. More analytically:



Σχήμα 2: Original and Synthesized signals on one plot

Observations

1. Similarity Between Signals:
   The overall structure and waveform shape of the synthesized signal closely follow the original signal. The periodic and non-periodic components -representing voiced and unvoiced sounds- are preserved in the synthesized signal.

2. Amplitude Differences:
   There are slight differences in amplitude between the original and synthesized signals, with the original signal visually having a larger amplitude than the synthesized one. These differences may result from energy normalization or numerical approximations introduced during the LPC analysis and synthesis process.

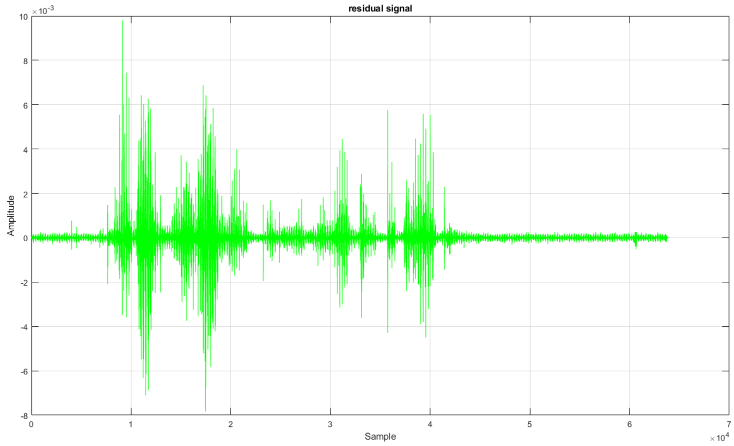
   Additionally, the difference in amplitudes is noticeable when listening to the audio, as the synthesized audio has a lower volume compared to the original signal.

3. Noise Reduction and Signal Density:
   The LPC process smoothens the signal, potentially reducing noise or irregularities. The LPC process models the speech signal using a limited number of coefficients- in our case 24-. This reduces the amount of information retained, particularly in higher-frequency components, which can result in smoother synthesized signal. Furthermore, LPC is particularly efficient at modeling voiced sounds but can struggle with unvoiced sounds . These unvoiced sounds have more high-frequency energy and irregular patterns, which may not be fully captured in the synthesized signal.

   About density, if i zoom on the waveforms i can observe that the waveforms of the original signal are more dense, possibly containing more information. The synthesized signal is less ”dense” as sharp, fine-grained variations in the residual signal are reduced and higher-frequency components are mostly lost.

We can visually see the differences at Figure 2 and by the following plot. By analyzing residuals, one can identify patterns that indicate model inadequates.

When the estimate is not perfect, the difference between the *original x[n]* and the *reconstruction x[n]* is non-zero. This difference signal $r[n] = x[n] - \hat{x}[n]$ is termed the residual.



Σχήμα 3: Plot of Residual Signal

Exercise 2: Create Robot Voice

Thereinafter, we will modify the excitation signal in **line 46** to create a synthesized *"robot voice."*. LPC orders of 23, 24, and 25 were tested to observe how different parameter values affect the synthesized signal.
The procedure we followed for this experiment was the same as on the previous exercise with the only addition of a *pitch_period = 100hz* and the creation of a *constant pitch excitation signal*.

Observations

After completing the experiments, we observed some differences between the synthesized signals. While the differences might seem small, they are still significant and impactful:

1. <u>Pitch Period:</u>
   The pitch period directly affects the perceived pitch of the robotic voice.
   A shorter pitch period (higher pitch frequency) produces a higher-pitched robotic voice. For example, using a pitch period corresponding to 200 Hz gives a sharper and more artificial sound.

   A longer pitch period (lower pitch frequency) results in a deeper robotic voice, which can sound mechanical but less sharp.

2. <u>LPC order value:</u>

   - **LPC order = 23:**
     May lead to a more distorted signal, as fewer coefficients are used to model the signal, reducing the precision and simplifying the model with a risk of losing important signal details.

Lower LPC orders failed to capture the finer spectral details of the original signal, leading to a simplified harmonic structure. With lower LPC orders, the robot-voice effect tends to sound more artificial and less natural because the model oversimplifies the signal.

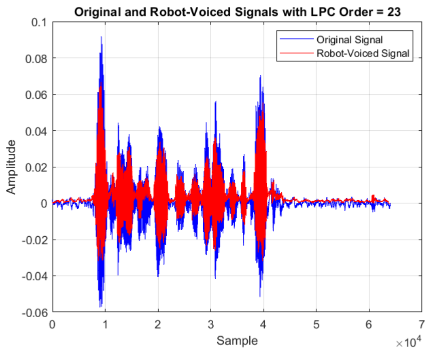
The residual signal retained higher energy, indicating a less precise prediction of the signal and showed more significant amplitude variations and required energy normalization to maintain a consistent output.

- **LPC order = 25:**
  On the other hand, the signal was the smoothest and most accurate compared to the original. The robotic effect was less pronounced but still identifiable. It created a smoother signal, which slightly diminished the robotic feel by introducing more natural characteristics of the original signal.

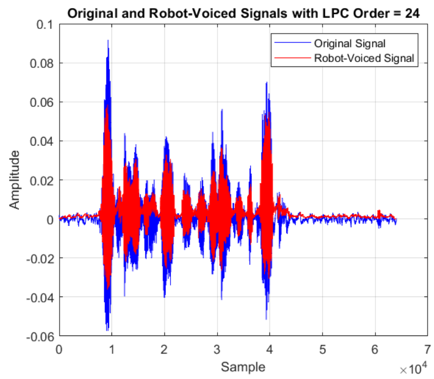
  As the LPC order increased to 24 and 25, the residual energy decreased, reflecting better predictive modeling and smoother synthesis with a more stable amplitude that closely matched the original signal.

  It retained more harmonic details, resulting in a fuller spectrum. In general, higher LPC orders lead to better excitation matching, making the signal slightly more natural.

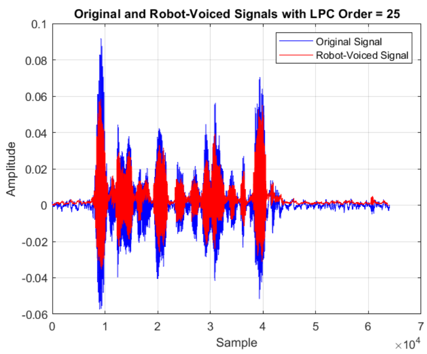
The following figures illustrate the comparison of the original and robot-voiced signals for LPC orders of 23, 24, and 25.



(i) LPC Order = 23



(ii) LPC Order = 24



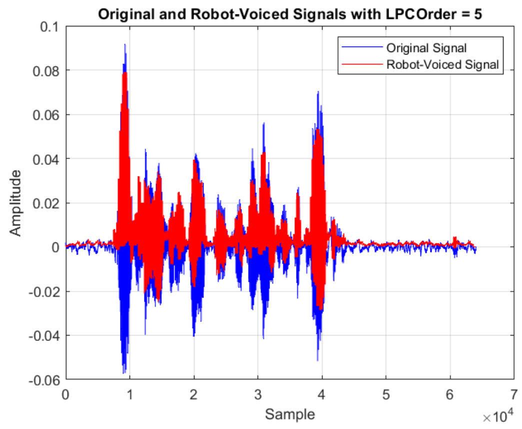
(iii) LPC Order = 25

Σχήμα 4: Comparison of Original and Robot-Voiced Signals for Different LPC Orders.

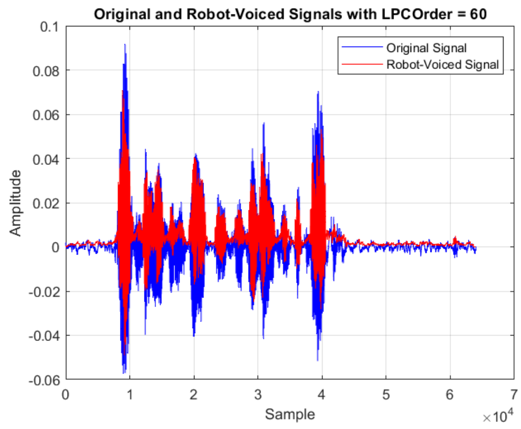
In conclusion, an overall perception is that:

- The LPC order changes the balance between the clarity of the output and the strength of the robotic effect.
  With **Lower LPC orders** the synthesized signal is less accurate, with more pronounced robotic distortions.
  With **Higher LPC orders** the signal is smoother and more natural, with better preservation of original details and the robotic effect is slightly diminished due to improved reconstruction.
- The pitch period adjusts the tone and pitch of the robotic voice, influencing how "sharp" or "deep" the robotic effect sounds.

As extra values to showcase how LPC orders affect the produces signal, here i tried with a very low value(*LPCorder = 5*) and with a very high one (*LPCorder = 60*).



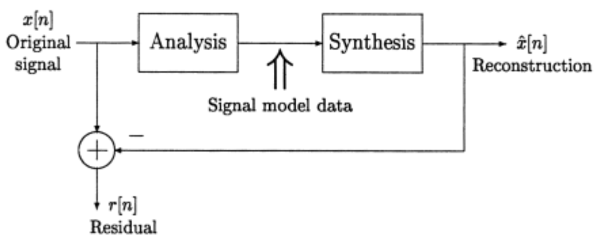
(i) LPC Order = 5



(ii) LPC Order = 60

Overall

Overall, this experiment demonstrates the fundamental principles of speech audio analysis and synthesis. The assignment involved breaking down the speech signal into smaller frames, analyzing its characteristics using Linear Predictive Coding (LPC), and synthesizing a robotic version of the speech using a predefined excitation signal. Through experimentation with different LPC orders and pitch periods, we gained insight into how these parameters influence the quality and characteristics of the synthesized speech. This process reflects the analytical and creative approach we applied throughout the assignment



Σχήμα 6: LPC Analysis and Synthesis of Speech