CS578 Speech Processing
Laboratory 2
Sinusoidal Modeling

George Manos
csd4333@csd.uoc.gr

Alexandros Angelakis
csd4334@csd.uoc.gr

5 February 2023

1 Implementation Details

We implemented the required parts, and applied the full sinusoidal modeling, as suggested by McAulay and Quatieri. There aren't any specific experiments implemented in this project, so the report will be sort. Note that we used the default suggested hyperparameters, as suggested by the provided implementation. However, we did attempt to experiment with them and we got errors in code sections that were not implemented by us, and therefore avoided to make any further attempts.

2 Results

2.1 Differences

First of all, we modeled the given speech signal. The results are presented in figure 1. Note that, to properly compute the MSE scores, we zero-padded the end of the resulting signal as they had different lengths.

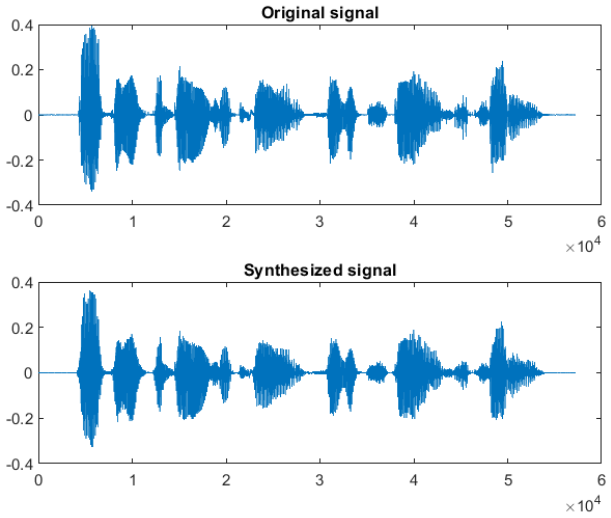

Figure 1: Sinusoidal Modeling for the arctic_bdl_snd_norm signal. MSE: 0.000088

Aurally, there aren't any major differences. However, with good speakers one may notice a slight difference on the overall audio quality. This is a result of the phase and amplitude interpolations we performed during the peak matching analysis part. The MSE score is fairly low, yet higher than the other ones.
Next, we have the synthesized George's voice after SM in figure 2. Its the result that has the lowest overall MSE. Aurally, the signal had no noticable difference to the original one.

Finally, we have the synthesized Alex' voice after SM in figure 3. The results once again are pretty similar, although we can hear something like a reverb on the beginning of the synthesized signal added to it.

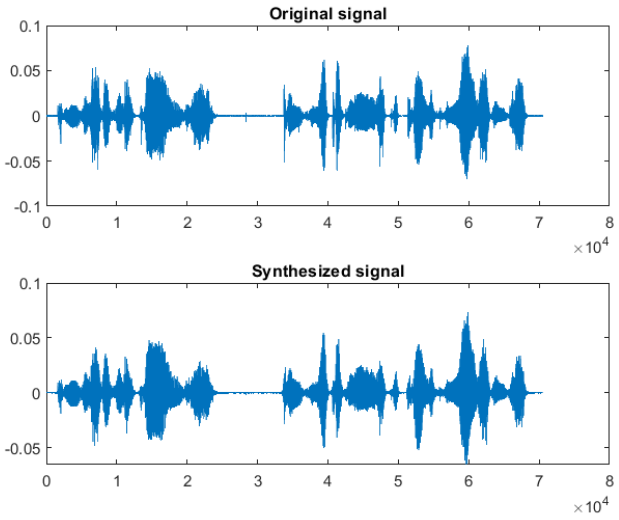

Figure 2: Sinusoidal Modeling for George's voice. MSE: 0.000007

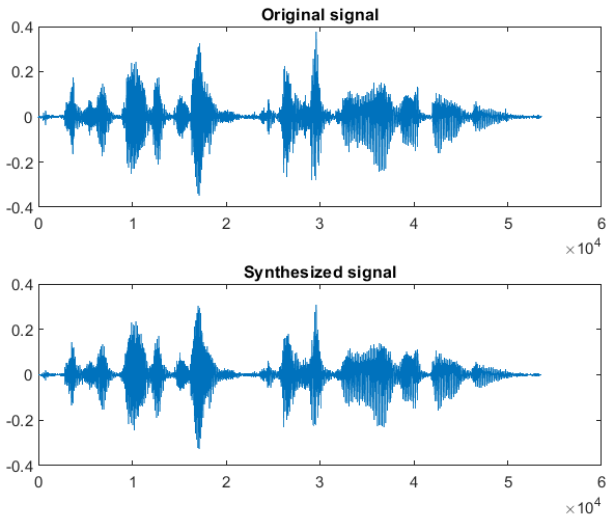

Figure 3: Sinusoidal Modeling for the Alex' voice. MSE: 0.000079

2.2 Signal-Noise-Ratio (SNR)

We also plotted the SNR over all frames as resulted from the SM. The results are presented below, for each speaker. The results look similar, except for Alex' signal as it lacks the 2 big SNR drops. Alex' voice apparently has a higher SNR over all frames, essentially meaning that his speech sample is cleaner overall than the other 2.

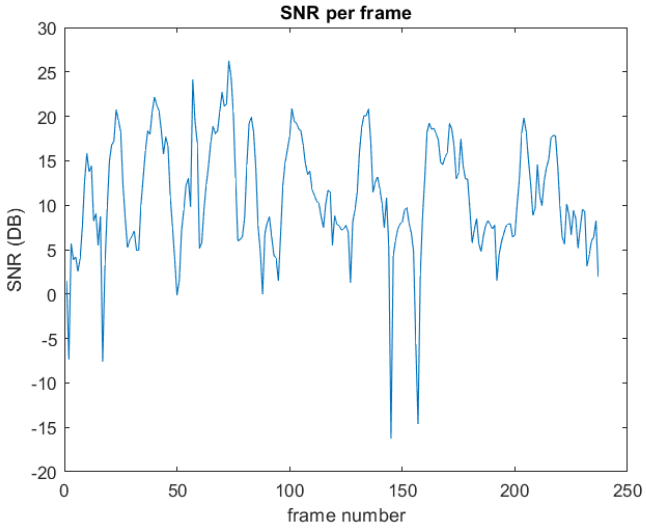

Figure 4: SNR plot over all frames for arctic_bdl_snd_norm signal.

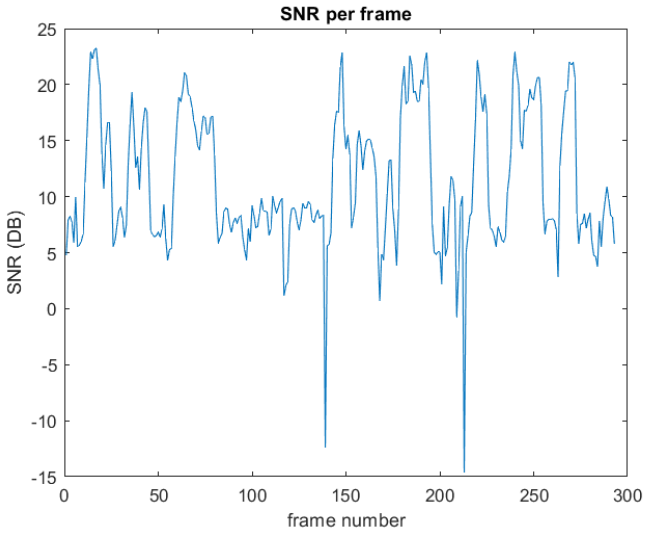

Figure 5: SNR plot over all frames for George's voice.

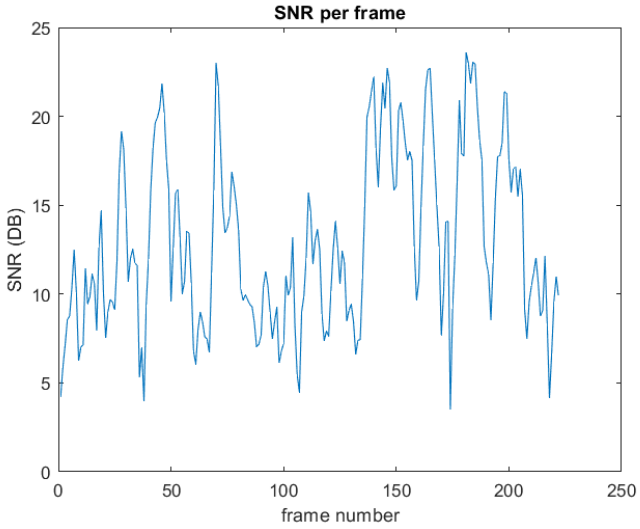

Figure 6: SNR plot over all frames for Alex' voice.

3 Appendix

(George) This is probably my last report for a course that Dr. George Kafentzis is related to. Its actually fun to realize that I was enrolled at 1 every academic year. All the memes included on my reports were only added for him and his amazing TAs, hopefully cheering them up through this boring process of grading our assignments. Thank you for making these courses great, and I really hope to see more as such down the line, inspiring such creativity and thirst for knowledge and experimentation. I hope to see our tradition of meme appendix passed on the future generations.

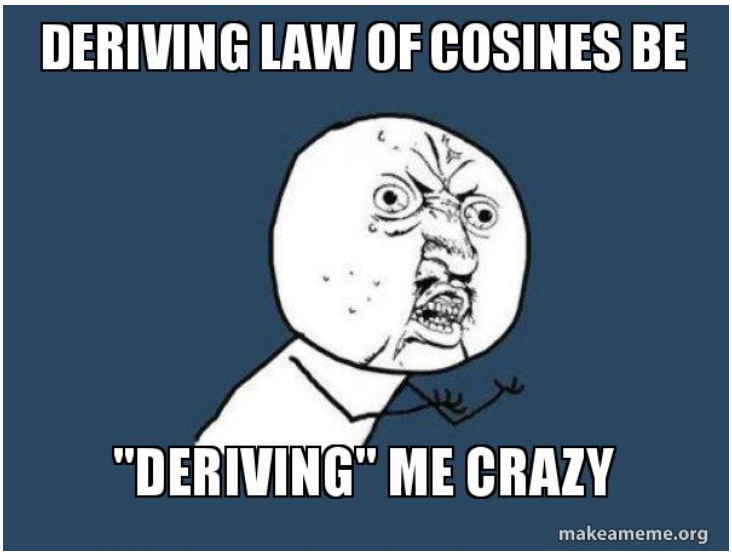

Figure 7: This will be my last meme.