Audio processing: lab sessions

Session 3: Noise Reduction in the STFT domain

Alexander Bertrand, Jesper Brunnström and Julian Schott.[1]

November 2021

Introduction

In Session 2, both the OLA and WOLA method were implemented in order to do frequency domain processing, to successfully obtain the resynthesised binaural signals. In this session, a noise source is added in the environment which affects the intelligibility and quality of the resynthesised binaural signals. The WOLA method is used in order to implement a Multi-channel Wiener Filter (MWF) in the Short-Time Fourier Transform (STFT) domain. Throughout the session, complete the missing sections from the Matlab code provided, *Noise_Reduction_WOLA_skeleton.m*.

Exercise 3.1: Generating the noisy microphone signals and WOLA analysis

From Sessions 1 and 2, the resynthesised binaural speech signals, *speech_L* and *speech_R*, are already available. First, create a set of noisy microphone signals by introducing a noise source into the scenario considered in Session 1 and 2. The following tasks are to be completed in the first two sections of *Noise_Reduction_WOLA_skeleton.m*.

1. Open the scenario from Session 1 in the GUI. Add a noise source to the existing scenario and generate the new set of RIRs.

2. Generate correlated noise in each of the microphones, using the RIR from the noise source. Use `Babble_noise1.wav` as the noise source. Obtain a noise signal for each of the microphones such that the SNR of *speech_L* to the noise signal in the left microphone = 5 dB.

3. Generate uncorrelated noise in each of the microphones such that the SNR of *speech_L* to the uncorrelated noise in the left microphone = 30 dB.

4. Combine the speech signal with the two noises, to create a noisy speech signal for the left ear and a corresponding noisy speech signal for the right ear. Listen to the binaural noisy signal. The location of the speech signal and the correlated noise signal should be identifiable.

5. Apply the WOLA analysis to the noisy speech signals, speech-only signals, and noise-only signals to convert the time-domain signals to the STFT domain and examine the corresponding spectrograms.

[1]For questions and remarks: `jesper.brunnstroem@kuleuven.be` and `jschott@esat.kuleuven.be`.

6. A function (*spp_calc.m*) is provided to compute the speech presence probability (SPP)[2] , i.e. the probability that speech is present in a particular time-frequency segment. The SPP is an important aspect to the noise reduction process as it can be used to distinguish between periods of speech-plus-noise and periods of noise-only , which is used to update the speech-plus-noise and noise-only correlation matrix for computing the MWF. Compute the SPP for the speech-only signal in the left microphone and observe the SPP output.

7. Instead of the speech-only signal, compute the SPP for the noisy speech signal in the left ear. How does this compare to the SPP for the speech-only signal? How does this differ between using a white noise signal or the babble noise as the noise source?

Exercise 3.2: Implementation of dual-microphone noise reduction with the MWF.

With the signals in the STFT domain, the MWF can be computed and applied to the noisy-speech signals in each time-frequency segment. The initial focus is on producing an enhanced speech signal for the left ear. The SPP, that was computed for the speech-only signal in the left-ear microphone, will also be used. In the Matlab code of *Noise_Reduction_WOLA_skeleton.m*, a nested for-loop runs through each time frame and each frequency bin. Within this nested loop do as follows:

1. Update the speech-plus-noise and noise-only correlation matrix. Set a threshold for the SPP in order to distinguish between speech-plus-noise periods and noise-only periods.

2. With the correlation matrices obtained, compute the MWF filter for the left-ear microphone using the generalised eigenvalue decomposition (GEVD) procedure.

3. Filter the noisy speech signals, the speech-only signals, and the noise-only signals. In practice the speech-only and noise-only signals are not available. For learning purposes, they are made available here, as they are useful for computing objective metrics to evaluate the performance of various algorithms.

4. Observe the spectrogram of the enhanced speech signal. How does it compare to the noisy speech signal input?

5. Apply the synthesis stage of the WOLA method to convert the STFT representations back into the time-domain. Plot the signals and listen to them.

6. Compute the input signal-to-noise ratio using the input speech-only signal and the noise-only signal for the left ear. Using the filtered speech-only and noise-only signals, compute the output SNR. What is the SNR improvement?

[2] More details on how the SPP is derived is given in: Gerkmann T, Hendriks RC. Noise power estimation based on the probability of speech presence. *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics* 2011; 145–8.

Exercise 3.3: Modifications: Binaural, single-channel, parameter settings

With the basic MWF properly working, proceed to make several modifications and investigate the corresponding performance.

1. The first modification is to obtain a binaural enhanced speech signal. Previously, noise reduction was only performed for the left-ear signal. Append the code so that the corresponding MWF is computed for the right-ear signal (*Hint: This only involves a very minor modification*). Filter the left-ear noisy speech signal with the MWF for the left ear and filter the right-ear noisy speech signal with the MWF for the right ear. The noise should be reduced and the binaural cues of the speech maintained.

2. So far two microphone signals have been used to obtain the enhanced speech signals. In some cases, there may only be access to one of these microphone signals at a time[3], and hence single-channel processing needs to be used instead. Modify the code once again to perform single-channel noise reduction. In order to use the same MWF code, the number of channels need to be adjusted accordingly . Obtain a single-channel MWF for the left-ear microphone signal, and a corresponding single-channel MWF for the right-ear microphone signal. How does this single-channel processing compare to the multi-channel case?

3. As seen, the MWF depends on a number of parameters. Investigate the effect of these parameters on the noise reduction performance (using the SNR metric and also by listening) by doing the following:

   ◦ Compute the SPP on the noisy speech signal of the left ear and apply the MWF. Experiment with various thresholds for the SPP.

   ◦ Apply the MWF for various values of the forgetting factor in updating the correlation matrices.

   ◦ Apply the MWF for various input SNRs.

   ◦ Apply the MWF for different FFT sizes.

[3] This can happen for instance if both signals have not been transmitted to one central processing unit.