P-D-ISSP Phase 1 Audio Processing

Group 4

Bing-Ruei Tsai r0772916

Kyrloglou Alexandros r0785071

I. Introduction

Hearing aids are really important for the socialisation of people with any hearing disability, but just amplifying all the sounds around them is not a realistic solution. What is needed is the ability for them to be able to focus on someone speaking while reducing the overall noise around them. A starting point for this is, using multiple microphones, to be able to know where each signal is coming from and afterwards to amplify that direction while reducing all others. For this two different setups are used in this report. First a, simpler, linear array and then, a simulation of a real situation using hearing aids and a dummy head. The situation set, is a setup where two people are talking from different sides, one on the left and one on the right, while wanting to focus on one of the two. For this the left side will be the target signal while the right speaker will be regarded as noise. The final result should be adaptable so even if the speaker is moving the hearing aids should be adapting to the location the sound is coming from.

II. Experimental Setup

For this project two different 4 microphone setups are used, the first one is a linear array and the second is a dummy head with two microphones on each ear.

A. Linear Microphone Array (LMA)

A schematic of the LMA can be seen in figure 1. In this the distance between the microphones is constant and equal to 5cm.

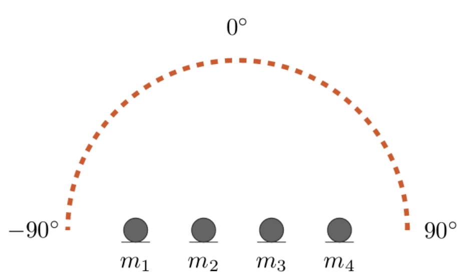

Fig. 1. Setup of linear Microphone array

B. Head-Mounted Microphone Array (HM)

A schematic of the HM can be seen in figure 2. The distance between the microphones in each ear is 1.3cm while the distance between the two ears is 21.5cm.

Fig. 2. Setup of Head-Mounted Microphone array

C. Measurements

In order to get a simulation of a situation where two people are talking, with the objective being to focus on one of the two. Loud-speakers are placed in the angles of -90° or -60° for one speaker and 90° or 60° for the other. These signals are recorded individually and in the same environmental conditions.

III. Theoretical Background & Implementation

The following methods are used in this report, here a quick explanation will be given and they way they are used. For an explanation on the mathematical background behind these algorithms check the citations on each part.

A. DOA-estimation: MUSIC algorithm

Here a wideband version of the MUSIC algorithm is used, in order to get the angle from which the signal is coming the total pseudospectrum is calculated from an weighted average of the pseudospectrums per frequency bin. The weights for this sum is the power of that frequency bin. Other important values in this part are the fact that a Hann window was used with a length of 1024 and 50% overlap in order to calculate the Fourier transform of each bin. Finally the DOA's selected are the top peaks in the total pseudospectrum, as meany peaks as sources.[1]

B. Beam Forming (BF)

Beam Forming is used to focus on one of the two sources. The BF is done in two steps discussed below. Results are only shown from the final version, that being the combined result, with both parts. A figure of the method can be seen in figure 3.

1) Delay-and-Sum (DAS): A preliminary result is gotten from the DAS, this basic algorithm removes part of the interfering noise by adding the delayed signals and averaging afterwards. This algorithm takes as input the DOA found from the MUSIC algorithm [III-A] and from that calculates the delay of each microphone signal. The output of this can be seen in figure 3 labeled as "Target audio signal reference", the input is the microphone inputs also seen in the same figure.[2]

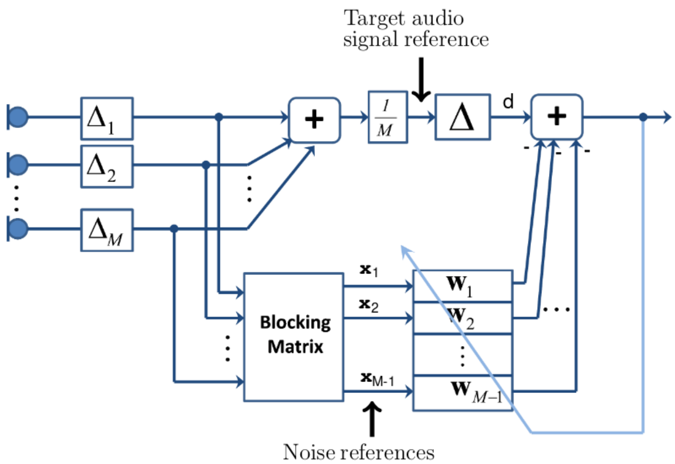

Fig. 3. Block diagram Implementation of the GSC algorithm

2) Generalised Sidelobe Canceler(GSC): After DAS in order to remove more of the sidelobes GSC is also implemented. For this, as seen in the lower part of [3], a Griffiths-Jim blocking matrix is implemented. That matrix calculates the difference between the different delayed microphone signals, which theoretically should be the sounds coming from all other angles other than the target one. After the blocking matrix an adaptive filter is used to optimally filter these noise references, so to remove as much noise as possible. The error of the adaptive filter is selected from the output at times where no speech can be seen, that is calculated by applying a voice activity detection filter (VAD).[3]

IV. System tuning

A. MUSIC

In the wideband MUSIC algorithm, the way the total pseudospectrum is calculated was improved. At start a geometrically averaged method was applied throughout all bins. However, some of the frequency bins are mostly noisy and thus mislead the DOA result. Referring to the narrowband version of MUSIC algorithm, the power of each frequency bin is calculated and used as the weights in a weighted average to find the final pseudospectrum. From this the effect from the noise is reduced.

B. GSC

The VAD and the NLMS filter determine the result of the GSC. The VAD estimation is challenged by the very noisy signal. A loose threshold of standard-deviation is used to label all non-speech parts. And the mean filter eliminates the quickly switching of NLMS which often cancels the speech signal.

V. Results

First, the ideal RIRs (Room Impulse Responses) are generated to estimate the DOAs in the different environments. There are 5-microphones array with 5cm inter distance which are put in center. The distance between microphones and source is 2m, and the room is a 5m squared room. In table I the results can be seen, as the reverberation time increases the resulting angle gets more inaccurate, additionally bigger angles have more substantial errors.

TABLE I
Simulation results of actual angle vs resulting angle from wideband MUSIC for different reverberation times

T60	0s	0.1s	0.2s	0.3s	0.4s	0.5s	1s
90°	90	90	62	48	-	-	-
60°	58	58	50	7.5	24	-	-
30°	29.5	29.5	30.5	28.5	39.5	36.5	19
0°	0	0	3	3.5	1	0.5	0
-30°	-30.5	-30.5	-27.5	-24.5	-18	-19.5	-11.5
-60°	-61.5	-61.5	-48.5	-41.5	-36	-	-
-90°	-90	-90	-59	-46	-	-	-

After estimating the performance of MUSIC algorithm, the DAS and the GSC is applied to do Beam Forming. There are two methods to label speech signal for calculating the SNR. One is generating the VAD from the original speech signal, and the other is from the processed-signal. The VAD from the original is the most accurate, but it wouldn't be available in a real-time application. In simulations the VAD is calculated from the original signal. For the experimental part, where actual recordings are used the processed signal is used to find the VAD. The SNR of each stage should be increasing meaning that the GSC is less noisy compared to the DAS and the similarly the DAS with the original microphone recording. However, when the angle between the source and microphone is larger, the result of DAS gets worse. A problem can be observed when the angle of the source to the microphone array gets bigger. The resulting SNR of the DAS is worse than that of the original microphone, as seen in table II. By listening to the output signals the perceived noise in the DAS output is still less than the original so there probably is some mistake in the calculation of the SNR.

TABLE II
SNR of simulation signals for different angles and reverberation times.

T60	0s			0.2s			0.5s		
Sig.	Mic	DAS	GSC	Mic	DAS	GSC	Mic	DAS	GSC
0°	7.3	7.5	8.0	7.4	8.9	17.4	5.9	4.9	35.5
-30°	7.6	5.1	15.1	7.6	5.8	14.7	6.0	4.6	20.6
-60°	7.0	2.8	13.7	7.1	4.0	13.0	5.8	4.6	12.4
-90°	6.9	3.3	12.8	7.0	3.3	12.7	6.1	4.3	14.4

Finally, the algorithms are applied on the recorded real cases. Due to the critical reverberation, the MUSIC algorithm fails to get the correct DOA, as seen in table III. Therefore, the speech and white noise signals have big errors. To be able to analyse the effect of the DAS and the GSC the correct angle is given to the DAS in order remove the errors made in the MUSIC algorithm. The final resulting SNR from the real recordings using the linear microphone array and the head mounted array can be seen in table IV.

TABLE III
ESTIMATED ANGLE WITH MUSIC ALGORITHM FROM THE REAL RECORDINGS.

Mic. config.	Linear Mic		Head Mounted	
signal	Speech	Noise	Speech	Noise
60°	14°	-	61.5°	-
-60°	-18.5°	-41°	-56.8°	72°
90°	18.5°	16°	69.2°	59.2°
-90°	-15.5°	-19°	-92°	-73.7°

TABLE IV
SNR OF REAL SIGNALS FOR DIFFERENT ANGLES AND SIGNAL TYPE.

type	Linear Mic			Head Mounted		
Sig.	Mic	DAS	GSC	Mic	DAS	GSC
60°	16.8	17.1	18.8	18.2	18.2	18.1
-60°	16.1	15.5	18.9	20.7	20.5	21.2
90°	19.3	18.9	22.1	19.1	18.5	20.1
-90°	20.7	20.7	23.8	20.1	19.9	20.3

VI. POSSIBLE IMPROVEMENTS

As it can be seen from the above stated results, there is still quite some room for improvement for any of the three parts of the setup. There has been extensive research on each of the parts and while implementing a better algorithm for any of them is possible, it would require a lot more time, that we did not have.

Firstly, for he the MUSIC algorithm, improving the resistance to reverberations would yield much better results. That can be done in various ways with the most important one being having a better analysis and processing of the resulting individual pseudospectrums.

Secondly, in the DAS section it would be possible to improve the accuracy of the time delays making the target audio signal reference much cleaner and at the same time the noise references have less leakage.

Finally, the GSC, this part has the biggest room for improvement. In this part having a very accurate VAD yields in a better adapting filter. Thus removing more interference's from the result, while not distorting the signal itself. This can be done by changing the VAD variables over time. Another important part is what blocking matrix is used, while we stayed with the basic Griffiths-Jim blocking matrix there are many other options that one can experiment with. In the end also the type of adaptive filter can be altered to get better results here NLMS was used but there are more complicated and possibly better options.

VII. CONCLUSION

In conclusion, the results seen in this report are a good indication that such a method to make a hearing aid would be possible, but it would require a lot more time, research and be more computationally expensive. Also since any change in the environment makes an increase in complexity to adapt to it using such a theoretical and mathematical method would make the final product not very adaptable. As such other methods to calculate the DOA or in general to apply beam forming can be looked into, such methods could be convolutions neural networks, or any other machine learning algorithms.

APPENDIX

The MATLAB scripts, input files and final results can all be found in the GitHub repository https://github.com/AKyrl/P-D-ISSP. This repository will be updated with the future parts of the project going deeper and making a better version for hearing aids.

REFERENCES

[1] DOA estimation based on MUSIC algorithm, https://pdfs.semanticscholar.org/5ff7/806b44e60d41c21429e1ad2755d72bba41d7.pdf, 16.05.2014.

[2] Delay Sum Beamforming, http://www.labbookpages.co.uk/audio/beamforming/delaySum.html, 12.12.2012, The Lab Book Pages

[3] Beamspace Adaptive Beamforming and the GSC, https://www.uio.no/studier/emner/matnat/ifi/INF5410/v12/undervisningsmateriale/foils/INF5410_beamspace-handouts-CIN.pdf, 27.04.2011, Carl-Inge Colombo Nilsen (carlingn), UiO.