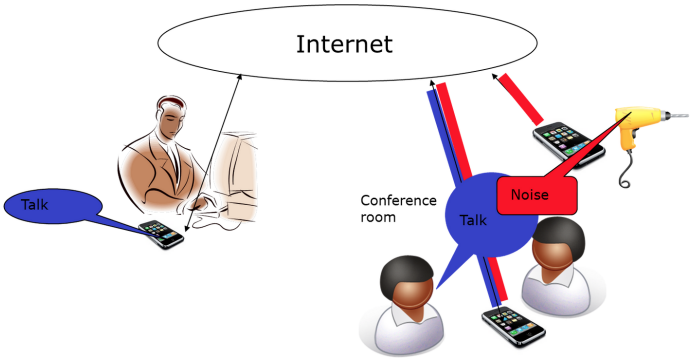
Kungliga Tekniska Högskolan

Project Report

Noise and echo cancellation in a teleconference



Authors:
Animesh DAS
Jonas SEDIN
Mohammad ABDULLA
Thomas GAUDY
Xavier BUSH

Advisor:
Per ZETTERBERG

Spring 2015

Contents

Chapter 1

Background

1.1 Introduction of noisy environments

It is a fact that the scenarios with phone calls involved are increasing every day. This situation implies an increase of the probability of being in a noisy scenario, specially in big cities. As a result of the discomfort that the users suffer in these noisy environments, engineering and science have worked with different approaches to solve this problem.

The diversity of noise nature and its sources lead the engineering to a big challenge: develop high performance soltions in these diverse environments. When facing noise cancellation is very important to take into account the variability that the noise may experience, as previously said. Duration of the noise sequences (from *ms* to long sequences), color of the noise and stationarity are possible classifications of the noise and each classification implies different ways of treating it. Therefore, a lot of systems are using combined techniques to reach the best possible performance, which has been naturally the case of this project.

1.2 Historical Overview

Before presenting the proposed solutions and approaches of the project, it is needed a historical overview to understand how have the group been influenced and which have been the patterns of research.

1.3 Description of the project

The problem proposed by the course *EQ2440* has been a "Noise and echo cancellation of a teleconference". The general scenario is that the first of the two speakers of the teleconference is in a noisy environment and the clear goal is to cancel as much noise as possible in order that the second speaker could receive a cleaner speech and make the conversation more comfortable. As said in **??**, there are different approaches to solve this problem, where several of them require the availability of pure noise recordings, in our case recorded with a third phone placed close to the noise source. To have a clearer overview of the scenario the Figure 1.1 shows an approximate scheme easy to understand.

When talking about denoising a teleconference there are two factors to take into account, techniques to cancel the noise and the possibility of their implementation in a real time application. The real time application has been, as expected, a big challenge because

it implies good performance in terms of cancellation with the minimum reachable delay to conserve the naturalness of the conversation.

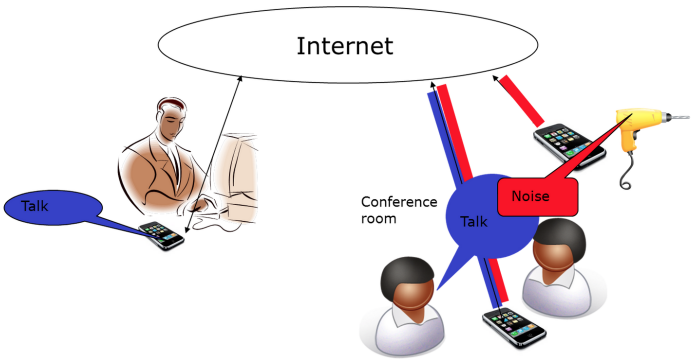


Figure 1.1: Scenario to solve

1.4 Goal

As commented in 1.3, the goal is to cancel the noise contribution in the conversation between the two speakers of the teleconferences. With the purpose to simplify the scenario, it will be assumed that only one of the speakers is surrounded by noise and the main noise source is known as well.

As in every engineering project, the group had to find a compromise between performance in noise cancellation and viability of implementation in real life. As it will be explained in 2, the computational cost is a big constrain and the best performance of certain approaches (3) introduce too much delay because of this reason. As a consequence, not always the best solution will be possible to implement in the real time version of the project.

As a contrast, the personal goals of the project members are to learn form the teamwork environment, learn a research methodology, research criteria and certain skills of management that might be used in the performance of a Master Thesis (as an inmidate future) and in a research or business environment.

The new knowledge acquisition is obviously another personal goal of all the team members.

1.5 Organizationn and Human Resources

The organization of the project consists in electrical engineering students at different stages of the studies and within different specializations. In order to make the team as efficient as possible, the project has been divided in four different groups: *Theory Group*, *Android Group*, *Multimedia Group* and *Management Group*, all of them explained in detail in 2.

The distribution of the team members has been as follows.

- Animesh Das
  - Role: Management Group
  - e-mail:animeshu1989@gmail.com (animeshd@kth.se)

  - Telephone: +46 737155575

- Jonas Sedin
  - Role: Theory Group & Android Group
  - e-mail: sedinjo@gmail.com (jonassed@kth.se)
  - Telephone: +46 704252951

- Mohammad Abdulla
  - Role: Android Group & Multimedia Group
  - e-mail: hamodiilatch@gmail.com (mabdulla@kth.se)
  - Telephone: +46 737393276

- Thomas Gaudy
  - Role: Android Group
  - e-mail: gaudy.thomas@gmail.com (gaudy@kth.se)
  - Telephone: +46 760936034

- Xavier Bush
  - Role: Theory Group & Management Group (Project Leader)
  - e-mail: xavier.bush@gmail.com (xbush@kth.se)
  - Telephone: +46 764141834

The sponsor members as Project Examiner/Supervisor and Project Support are:

- Per Zetterberg
  - Role: Project Examiner
  - e-mail: perz@ee.kth.se
  - Telephone: +46 8 790 77 85

- Hadi Ghauch
  - Role: Group Assistant
  - e-mail: ghauch@kth.se

- Martin Ohlsson
  - Role: Android Guru
  - e-mail: martinoh@kth.se
  - Telephone: +46 87907818

Chapter 2

Methodology

This chapter shows the methodology that the group has followed since the projecte started. On the first hand, it goes without saying that the project group has followed the *Scientific Method* in the implementation of the project. On the second hand, as commented in 1.5, the group has been divided in three groups explained in the following subsections.

2.1 Theory Group

The*Theory Group* had as its main goal finding solutions to cancel the present noise in the teleconference. Nevertheless, a constraint of the group has been the computational cost that the implementation have, where all the details may be found in 3.

The fact that three members of the project had recent and good background in Adaptive Signal Processing, which has been one of the chosen approaches to face the noise cancellation, made easier the making of the groups. Moreover, the *Theory Group* avoided the first stage in theory research, which is the most difficult part when starting a new project.

In terms of methodology, the *Theory Group* has followed next steps:

- Make research in suitable algorithms.
- Record with the given mobile phones both signals: 'voice+noise' and 'noise'.
- Test the performance in MATLAB.
- Check the computational cost in MATLAB.
- Check the possibility to transfer the solutions from MATLAB to Android.

Because of the presence of Jonas Sedin in the *Theory Group* and the *Android Group*, it has been possible to design theory solutions think in the availability to transfer them to Android coding.

2.2 Android Group

This group has had two different duties:

- Android tasks: these tasks aimed to be an introdution to the Android programming.

- Noise cancellation in Android: has been the transfer from the proposed solution in the *Theory Group* to Android.
- Real time application: optimize the code to decrease the computational cost and make possible a real time application.

In this case, even if Thomas Gaudy has not been a part of the *Theory Group*, his background in Adaptive Signal Processing has helped in the implementation of the proposed solutions.

2.3 Multimedia Group

This group is only formed by Mohammad Abdulla who has taken care of all the technical preparation of the presentations of the project:

- Video of the project: explanation of the project with real examples.
- Power Point Presentation: power point to use in the Grand Final Presentation.
- Bloopers Video: video containing bloopers and funny moments during the performance of the project.

2.4 Management Group

The *Management Group* has taken car of the drawing up of those tasks that were oriented to plan, follow up and present the work of the project.

The main documents are listed below:

- Project Plan: document that cointained a brief description of the project, time resources and planning and tasks planning.
- Progress Report: document containing the follow up of the projects in terms of achievements and resources. It is a updated version of the Project Plan. The last Progress Report can be found in 6.
- Project Report: this precise document. It contains the description of all the project in all the areas.

Besides the main documents, the *Management Group* has taken care of next tasks:

- Meetings: there has been no need to schedule meetings because of the small size of the group and the cross-group duties.
- Scheduling: the planning of all the tasks was designed by this group and it has been properly followed.

2.5 Cross-Groups Duties

In terms of making easier the transfer of information between groups, the makeup of the groups has followed a cross-specialization criteria. Therefore, there are team members in more than one group, and it has been possible due to the wide scope of skills that the team members have (1.5).

With this makeup, there have

Chapter 3

Theory

The *Theory Group* has been working within different scenarios but mainly has followed what is called *Adaptive Signal Processing*, as it has been explained in **??**. Some of them have ended in good performance algorithms where only two of them have been transferred to Android and some other have not reached the wanted outcome.

In terms of goals, it has been very important for the group to reach a good noise cancellation and preserving as much as possible the voice quality.

3.1 Successful Approaches

3.1.1 LMS Algorithm

Introduction

The LMS (Least Mean Square) Algorithm [1] is an adaptive filtering technique used to estimate the unwanted component in the signal that wants to be filtered and, a posteriori, subtract it to "clean" the wanted signal.

To use this technique it is a needed to have both signals, the one that contains the useful information ($y(n)$) and the noisy one ($x(n)$).

The process used in this technique is a recursive calculation samples by sample were N previous samples of the wanted signal are used to calculate the current filter sample, used later on for the subtraction.

The original problem to solve in LMS is shown in the Equation 3.1, where is clearly visible that the LMS minimizes the MSE using the instantaneous error.

$$\hat{\theta}(n) = \hat{\theta}(n-1) - \frac{\mu}{2})\frac{\partial}{\partial\theta}MSE(n,\theta)|_{\theta=\theta(\hat{n-1})} \tag{3.1}$$

Where after the appropriate calculations, the final version ready to implement in MATLAB is next:

$$\hat{\theta}(n) = \hat{\theta}(n-1) + \mu Y(n)(x(n) - Y^T(n)\hat{\theta}(n-1)) \tag{3.2}$$

In the Equation 3.2:

- $\theta(\hat{n})$: estimation of the current sample of the filter
- $\theta(\hat{n})$: estimation of the previous sample of the filter
- $Y(n)$: vector that contains the N samples used in the implementation

- $x(n)$: sample of the wanted signal
- μ: step size

A very important parameter to consider is the *Step Size* μ. The value of μ has a direct impact on the stability and convergence of the algorithm. The general rule of choosing the right *step size* is shown in **??**

$$0 < \mu < \frac{2}{\lambda_1} \tag{3.3}$$

where λ_1 is the maximum eigen value of $I - \mu\Sigma_{YY}$.

In terms of cost operation, LMS is very efficient thus only needs O(N) operations in each iteration. A priori, this is feasible for a real time implementation and this one of the reason that made us start with this approach.

Nevertheless, the implementation of LMS for noise cancellation is a little bit different. As showed in Figure 3.1, the purpose is to estimate $x_1(n)$ from $x(n)$, i.e. calculate $\hat{x_1}(n)$ to then subtract if from $z(n) = s(n) + x(n)$ from the received signal at the receiver phone. In conclusion, it is necessary that de *Adaptive Filer* emulate the *Channel*.

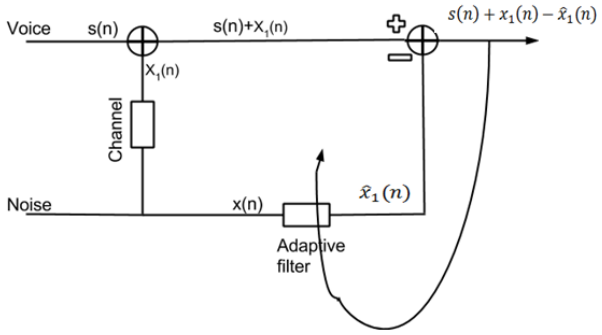


Figure 3.1: Simplified boxes scheme of LMS (Noise Cancellation)

Implementation & Results

The implementation used in MATLAB is based on the one that Jonas, Thomas and Xavier have done during the course of *Adaptive Signal Processing (EQ2400)*.

Once all the recordings where done, the first tests of the LMS gave different performances depending on the number of previous samples used to calculate the filter.

- $N = 100$: good performance
- $N = 1000$: better performance with a big computational cost

Even though there is an upper limit for the value of μ (**??**), there is not a lower limit (besides 0). Therefore, the group followed a test-and-set empirical criteria where different μ values where tested and the value who gave the best auditive performance was the chosen. In the last case, the best performance was given by $\mu = 5 \cdot 10^{-5}$.

These results are reasonable: it exists a linear dependency of the performance with the number of previous samples used to calculate the current estimation.

To show some graphic results of the performance of this algorithm, on the one hand Figure 3.2 and 3.3 show the time and frequency response of the filtered sequence with

$N = 100$ samples in comparison with the original recording. There is some audible noise reduction that in the figure can be seen in the noisy frames where there is more presence of noise in the original plot than in the filtered one. When it comes to the frequency domain, we can appreciate that there are several small peaks at low frequencies and the contribution of the peak at $f = 0.6$ (where $fs = 44.1$ KHz) has been slightly decreased.

On the other hand, the Figures 3.4 and 3.5 show the same comparison as before but with a filter of $N = 1000$ samples. In this case the Figure 3.4 shows that the noise has less presence. The drawback comes with a certain distortion in the audio file that can be seen in the Figure 3.5 as a modification in the frequency domain with new peaks in a coarser spectral response at lower frequencies.

In terms of computational time, in MATLAB the LMS with $N = 100$ samples takes 2.22 seconds, while the LMS with $N = 1000$ takes 3.78 seconds (increasing of a 70%). This does not make a huge difference in MATLAB but in a real time application it might, therefore it will be treated by the *Android Group*.

As a clarification, the reason why the group has tested different number of samples in the LMS is because the lack of synchronization between $z(n)$ and $x(n)$: with a bigger N the delay between both signals is solved in terms of filtering.

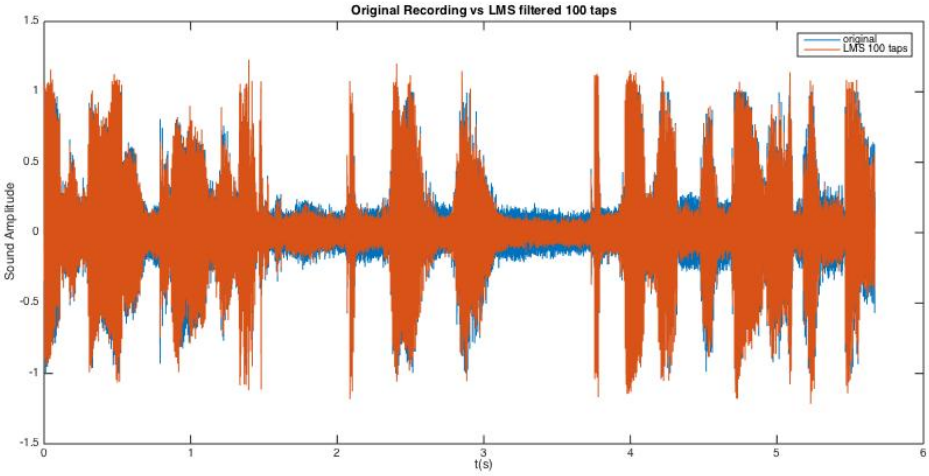


Figure 3.2: Time Domain - LMS result with N=100

3.1.2 Speech Enhancement Using a-Minimum Mean- Square Error Short-Time Spectral Amplitude Estimator

3.1.3 Wiener Non-Causal Filtering

Introduction

Wiener Filtering is another approach used in Signal Processing with many purposes and noise cancellation is one of them. Inside Wiener Filtering there are many different algorithms to use from which we have chosen Non-Causal infinite dimensional Wiener filtering [1].

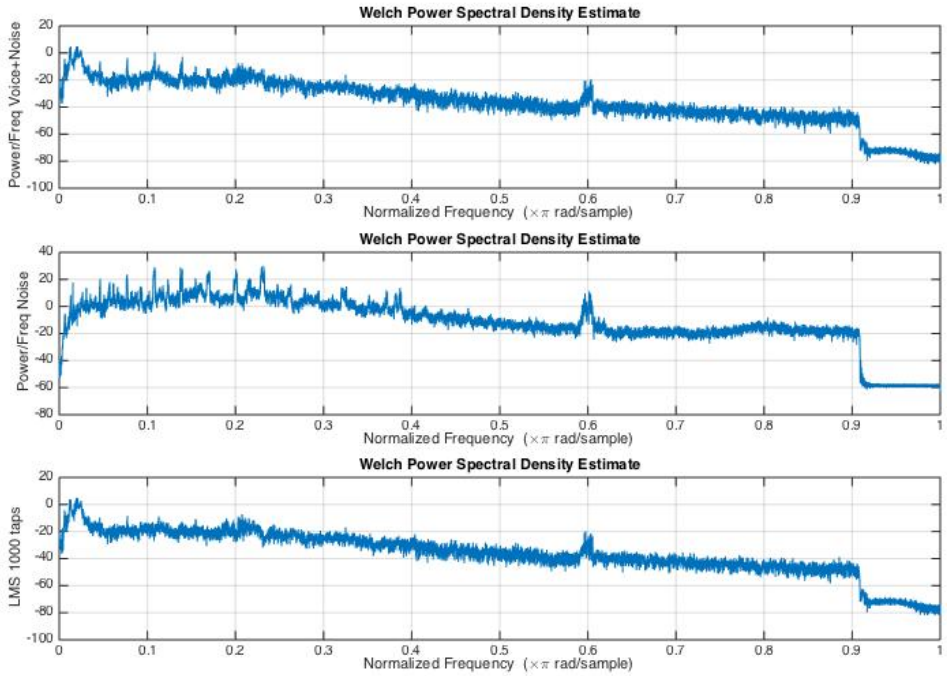


Figure 3.3: Frequency Domain - LMS result with N=100

A general non-causal linear estimator is given by

$$\hat{x(n)} = \sum_{k=-\infty}^{\infty} \theta(n,k)y(n-k) \tag{3.4}$$

To solve this problem, given the estimator, the group has followed a spectral factorization technique that defines the optinar linear filter as

$$H(z) = \frac{\Phi_{xy}(z)}{\Phi_y(z)} \tag{3.5}$$

Therefore, the group's goal is to find $\Phi_{xy}(z)$ and $\Phi_y(z)$.

Implementation & Results

There are many differences between LMS and Non-Causal Filtering. These are the most important:

- Number of phones=2. In this case, to extract the needed information about the noise it is not needed the third phone that was used to record pure noise.
- The processing is done by frames.
- A voice unvoice detector is needed.

The developed voice-unvoice detector used is energy-based. As the filtering is done frame by frame, the detector will calculate the variance of the frame and, on behalf of a

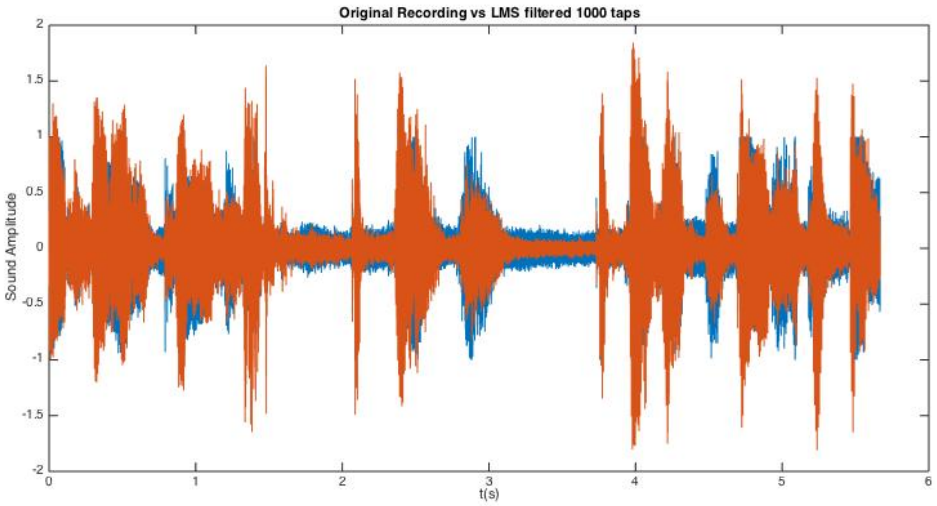


Figure 3.4: Time Domain - LMS result with N=1000

σ_{ref} previously set. A compromise is needed because a high σ_{ref} has a good accuracy in the speech detection (less voice presence) but there are some noise frames with part of consonants at the end or the beginning of a noise block. Contrary, a small σ_{ref} guarantees a pure noise sample but it adds noise presence in the voiced samples. As a result, we used a high σ_{ref}, thus what we want to estimate the better is the noise, so it is what we really want to update and filter afterwards. An example of how the detector works is shown in Figure 3.6.

Once the frame is decided to be voiced or unvoiced the proper calculation of $\Phi_{xy}(z)$ and $\Phi_y(z)$ is next step. To do so, the group has used MATLAB functions used previously in the *Adaptive Signal Processing (EQ2400)* course. These functions require information about the noise and voice frames such as their variances and means. Therefore, to calculate them the MATLAB code follows next steps:

- It is assumed that the teleconference starts with a noise frame
- Voice-Unvoice detection of the frame
  - If the frame is unvoiced: $\mu_{unvoiced}$ and $\sigma_{unvoiced}$ are updated
  - If the frame is detected as voiced: μ_{voiced} and σ_{voiced}
- Filter the frame
  - If the frame is unvoiced should be totally removed
  - If the frame is voiced the noise should be reduced
- No overlap between frames

After applying these steps the results have been partly satisfactory. As can be appreciated in the Figure 3.7, the noise is highly reduced in the unvoiced frames. On the other hand, the drawback comes in shape of unstable regions, as can be seen at first sight between seconds 4 and 5. Figure 3.8 shows a zoom in in an unstable region, and a possible

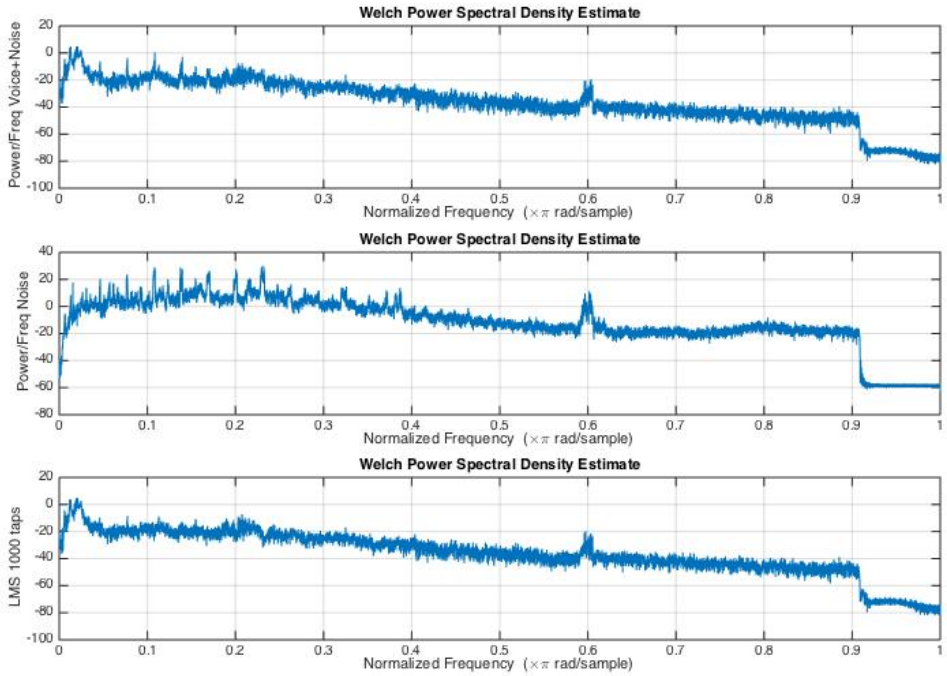


Figure 3.5: Frequency Domain - LMS result with N=1000

explanation of it is that the functions used in the non-causal filtering is the inserting of a zero 60 samples-vector at the end of each filtered signal, in each frame in our case.

To try to solve this problem, a solution to filter overlapped frames has been implemented. The main purpose of if was to avoid unstable regions and the zero-sample vectors. Meanwhile the second problem has been successfully solved, a bigger unstable appeared, as shown in Figure 3.9. A curious results is the perfect shape of a decreasing exponential that it has.

However, the members of the theory group have been working in parallel, while this problem was trying to be solved the performance of 3.1.2 gave extraordinary results and the research to solve this problem was stopped.

3.2 Unsuccessful Approaches

In this section are presented some of the unsuccessful approaches of the theory group. The reasons of the unreached success come from different natures, as will be explained above in each algorithm or technique.

3.2.1 Kalman Filtering

Kalman Filtering is an Adaptive Filtering technique highly used in tracking and predicting, for instance. Therefore it was considered as a suitable option to face the noise cancellation problem of this project. But, unfortunately this technique showed two main drawbacks:

- Computational cost too high. Therefore not possible the transfer to the Android coding.

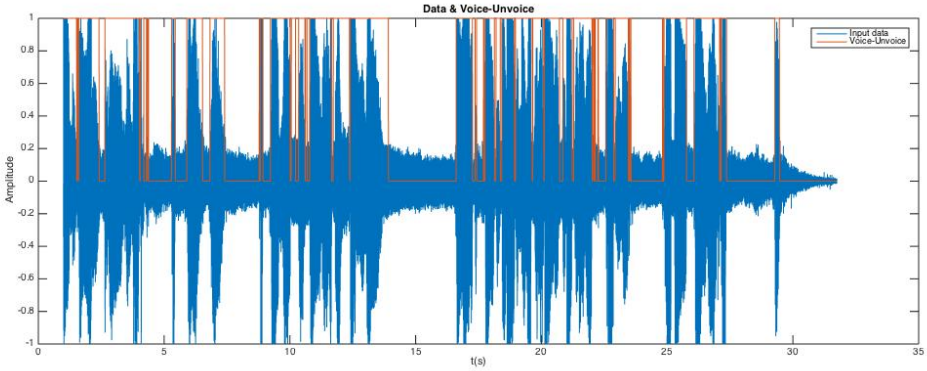


Figure 3.6: Voice-Unvoice detector

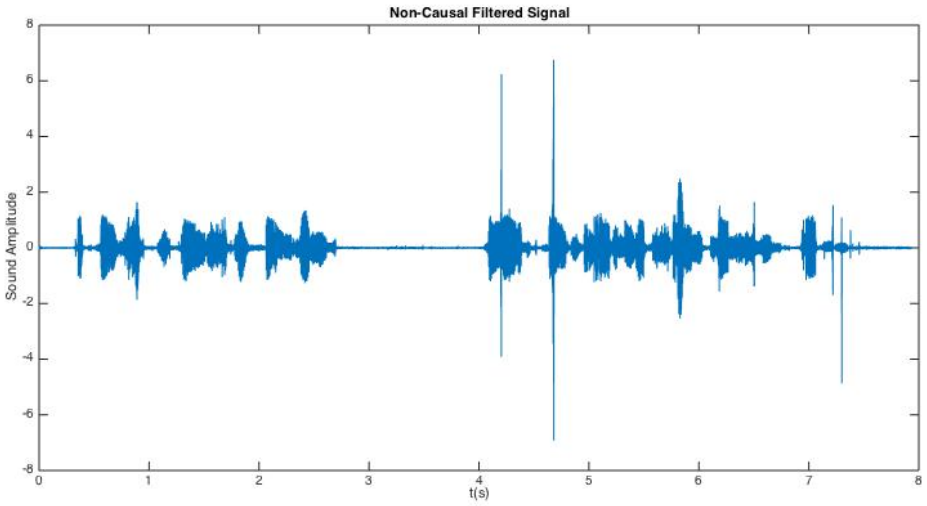


Figure 3.7: Unstable region in non-causal filtering

- Difficulties to find a right State-Space Model showed in Equation 3.6, i.e. problems to find *F, G, H, v(n) and w(n)*

$$x(n+1) = Fx(n) + Gw(n) \tag{3.6a}$$
$$y(n) = Hx(n) + v(n) \tag{3.6b}$$

For the reasons above, this way was dropped in the first weeks of the project.

3.2.2 RLS

The REcursive Least Squeares (RLS) Algorithm [1] is supposed to have a better convergence and result than the LMS Algorithm. The problem that RLS tries to solve is the FIR Wiener one, where the aim is to estimate the proxess $x(n)$ using a fixed number of

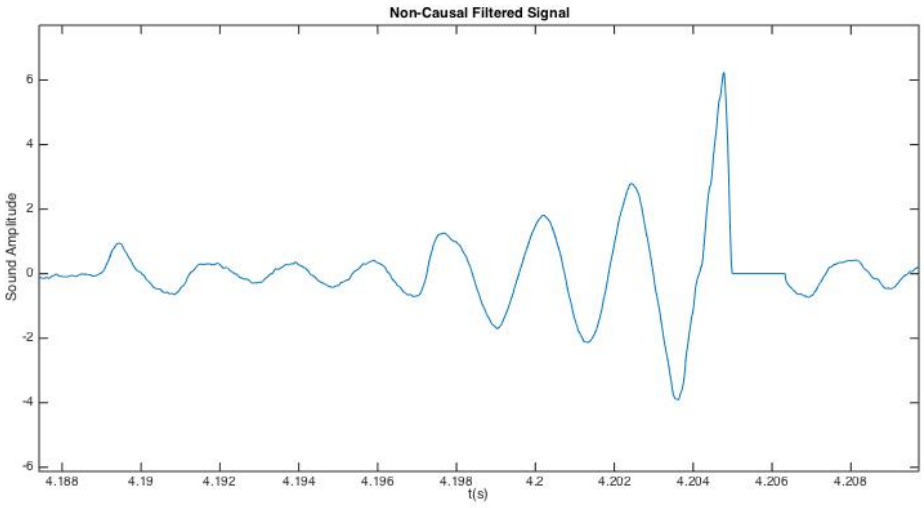


Figure 3.8: Unstable region in non-causal filtering

observations from another process $y(n)$ using a liner estimator

$$\hat{x}(n) = \sum_{k=-0}^{N-1} \theta(k) y(n-k) = Y^T(n)\theta \tag{3.7}$$

where $Y(n) = [y(n), y(n-1), \ldots, y(N-N+1)]^T$ are the saples that have to be filtered and the crierion is $MSE(\theta) = E(x(n) - \hat{x(n)})^2$.

Nevertheless, the option was tested during the first days of the project and dropped after checking the Computational Cost: $\mathrm{O}(N^2)$ vs the $\mathrm{O}(N)$ of the LMS.

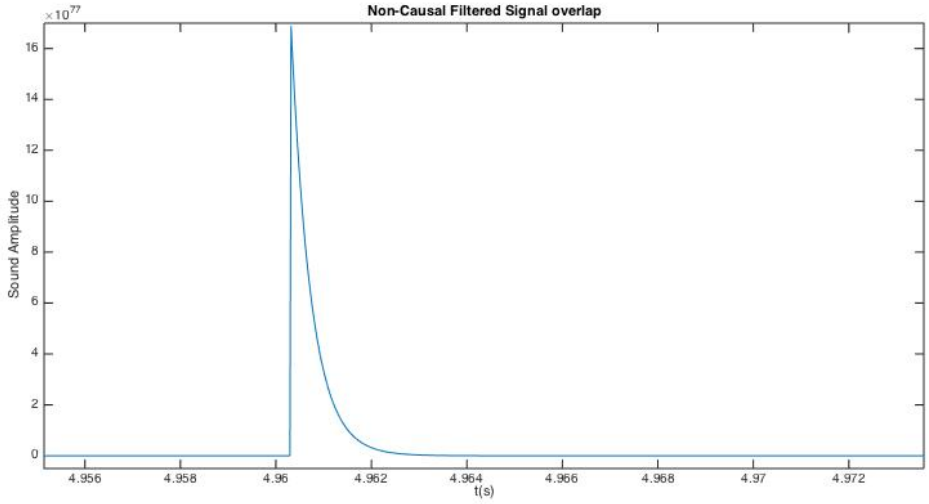


Figure 3.9: Unstable region in non-causal filtering

Chapter 4

Android

The Android part has been the most challenging in the whole project: implement a real time application with signal treatment implies code optimisation skills. Even though Android environment coding was a new experience for all the members of the group, the results have been outstanding.

The coding environment of Android is Eclipse and the language used is JAVA, objecte oriented and assumed to be known by the reader. Moreover, the used *Android* platform is *Android 5.1*

I this chapter it is going to be presented the followed steps and reached results during the performance of the project.

4.1 Code Training

For the *Android Group*, the project began with a set of tasks that needed to be finished before the *Midterm Evaluation*. It goes without saying that the group did do it in less time than expected and could start working on the transfer from MATLAB to Android coding two weeks before than expected.

4.2 Coding for Noise Cancellation

Once the group was familiarized with the environment, the procedure of treating the information needed to be set. As the communication path for the teleconference is the internet, the information is going to be sent with TCP frames, therefore the signal processing is going to be frames/blocks oriented, i.e. a buffering is needed for such an application.

In terms of time, the buffer is a non-variant parameter and it lasts around *92 ms*. Contrary, the buffer length in terms of samples is frequency f_s dependent and proportional. Typical buffer lengths are 1024 and 2048 samples which correspond to f_s=*11025 Hz* and f_s=*22050 Hz* respectively.

Other technical requirements are:

-
-
-

4.2.1 State Diagram

To have a full understanding about how does the coding work, it is necessary to explain the different states that take place in the coding (see Figure 4.1). The list of states and their explanation are above:

- **Waiting**: starting state. This state is used to wait for both signal and noise phone to be online and to send buffers. Then, buffers are been filled until T1 happens.
- **Correlation**: state where the cross-correlation between the signal and noise buffers is calculated. The maximum of the cross-correlation gives the total delay between the two samples' flows. Then, by a simple algorithm, this total delay is divided into two variables. These two variables are important thus the coding works with dynamic lists of buffers. The two variables are:
  - *Number of buffers between the two corresponding buffers*: this is the number of buffers needed to throw, either for the signal or the noise, to synchronize the buffers' lists.
  - *Delay inside the synchronized buffers*: this is the number of samples we need to throw, either for the signal or the noise, to synchronize the samples.
- **Empty**: useless buffers (used for the cross-correlation) are removed. This state also helps fixing the length of the dynamic lists.
- **Play**: the most complex state. It can be divided as follows:
  - *Initialization phase*: variables that will be used are initialized, particularly the noise, signal and *to_be_played* arrays. The*to_be_played* arrays is basically the same as the signal array, but of a smaller length since more samples are needed on the signal array to apply the NLMS (**??**) algorithm, otherwise we some samples could be missed.
  - *Noise cancellation phase*: the algorithm goes through this phase if we ask it to do it; otherwise it skips to the next phase. This phase applies the NLMS algorithm between the signal and the noise arrays. Then the result is subtracted to the "to be played" array.
  - *Voice detection phase*: this algorithm uses the *to_be_played* array (after NLMS or not) to detect if there is voice or not. If there is no voice, some parts of the array are set to zero.
  - *Playing phase*: the processed array *to_be_played* is played and the noise cancellation is analysed.

Once the states are explained, the transitions are the last concepts to have a full understanding of this section:

- T1: happens when enough buffers are received from the signal and the noise phones to calculate the cross correlation between the buffers.
- T2: happens when the cross-correlation has been calculated and the delay computed.
- T3: happens when the useless buffers are removed (the one we used to calculate the cross correlation)

- T4: happens when someone tap the button "reset" on the phone's screen.

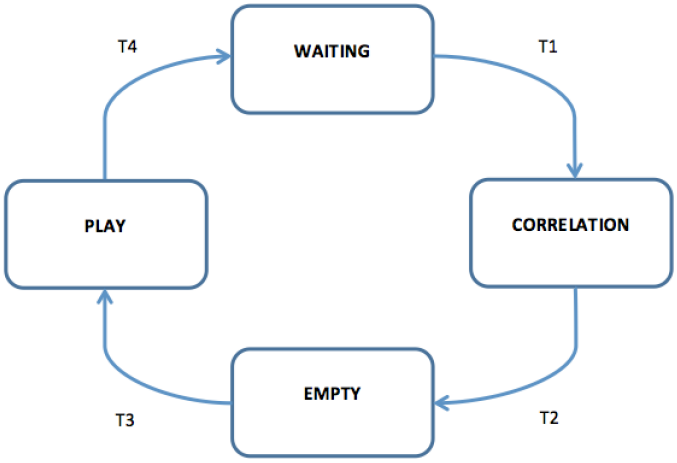


Figure 4.1: State Diagram in the Android Coding

4.2.2 NLMS

The *Android Group* had to made some changes from the material sent by the *Theory Group*. Due to some saturation, the *NLMS Algorithm* was successfully tested.

A few problems that may appear on the LMS are related to stability due to the signal power. Hence the step-size may have to be chosen unnecessarily small. However, the LMS can be made insensitive to the signal power by using a normalized step-size [1]:

$$\mu(n) = \frac{\bar{\mu}}{c + \|Y(n)\|^2} \tag{4.1}$$

The NLMS version in Android has been challenging because the buffers used on the framework last approximately *92 ms*, as commented in 4.2 and the processing part has to last less than this time. Because the NLMS version was imported from MATLAB, it has been first translated to JAVA and later on the code has been improved to reduce the execution time. These have been two implemented solutions to do so:

- **Intelligent 'for-loops'**: The for-loop executes a test on the index variable for every iteration. This test consists of calculating the subtraction between the index variable and the tested value, and then compare it to 0. To really reduce the execution time, the compared value has to be zero and thus any subtraction will be calculated, which helps if there are a lot of iterations. It is done as follows:
  - Replacing
  - With
- **Moving every "static" calculation out the loops:** operations inside loops take time. Therefore, all the constant values respect to the loop index are taken out of the loop and define them as local variables to avoid useless operations. It costs less to point at a local variable than to calculate.

Chapter 5

Conclusions

Chapter 6

Appendices

Chapter 7

Bibliography

[1] Håkan Hjalmarsson and Björn Ottersten. *Lecture Notes in Adaptive Signal Processing (EQ2400)*. KTH Edition, 2010.