Master's Thesis

Drone Detection using Audio Analysis

Louise Hauzenberger
Emma Holmberg Ohlsson

Department of Electrical and Information Technology,
Faculty of Engineering, LTH, Lund University, June 2015.

Drone Detection using Audio Analysis

Louise Hauzenberger
ada10lha@student.lu.se

Emma Holmberg Ohlsson
ada10eho@student.lu.se

Department of Electrical and Information Technology
Lund University

Advisor:
Dr. Mikael Swartling
mikael.swartling@eit.lth.se

June 15, 2015

Printed in Sweden
E-huset, Lund, 2015

Abstract

Drones used for illegal purposes is a growing problem and a way to detect these is needed. This thesis has evaluated the possibility of using sound analysis as the detection mechanism. A solution using linear predictive coding, the slope of the frequency spectrum and the zero crossing rate was evaluated. The results showed that a solution using linear predictive coding and the slope of the frequency spectrum give a good result for the distance it is calibrated for. The zero crossing rate on the other hand does not improve the result and was not part of the final solution. The amount of false positives increases when calibrating for longer distances, and a compromise between detecting drones at long distances and the number of false positives need to be made in the implemented solution. It was concluded that drone detection using audio analysis is possible, and that the implemented solution, with linear predictive coding and slope of the frequency spectrum, could with further improvements become a useable product.

Acknowledgments

We would like to express our gratitude to Dr. Mikael Swartling, our supervisor at LTH, and our examiner Doc. Nedelko Grbic for their help and guidance during the thesis work. We would also like to thank Axel Landgren, for the thesis proposal, his help and his enthusiasm during the whole project. Without all of them this thesis would not have been possible.

Louise Hauzenberger and Emma Holmberg Ohlsson
Lund, May 2015

Table of Contents

List of Figures

List of Tables

List of Abbreviations

DSP	Digital Signal Processing/Digital Signal Processor
FFT	Fast Fourier Transform
FLOPs	Floating-point operations per second
LPC	Linear Predictive Coding
MFCC	Mel-Frequency Cepstrum Coefficients
PLP	Perceptual Linear Predictive
PSD	Power Spectral Density
ZCR	Zero Crossing Rate

Chapter 1
Introduction

In the last few years the use of small drones have increased dramatically. Illegal activity with the assistance of these drones has increased as well, or at least become more obvious than before. There have lately been reports of drones being used to transport drugs across borders, transport contraband into prisons and take aerial footage of secure facilities. [1–3] To help protect from these activities, a product for detecting drones could warn about a security breach in time to take action. This thesis concerns the creation of a drone detector which uses audio analysis in the detection process.

1.1 Problem Definition

The goal of the thesis is to create a drone detector based on audio analysis. Questions to answer are

- Is it possible to detect drones using audio analysis?
- Which characteristics of sounds are interesting to compare with?
- Will other sounds be detected if all drones are detected?
- How much better does the detector get with added characteristics in the detection?
- Can multiple microphones give any benefits?

A database of drone sounds has to be created to compare the observed sound against. The solution should be implemented on a DSP and tested with drone sounds and other sounds to evaluate the quality of the detector.

To decide on suitable parameters two drones were used, one WLtoys V262 Cyclone Quadcopter (a large quadcopter of 52×52 cm) and one Hubsan X4 (a small quadcopter of 11×11 cm) as of now called Cyclone and X4 respectively, seen in figure 1.1. They were chosen to see if the characteristics were similar despite the size difference. 50% of the drones' throttle level was chosen as the focus since that is approximately when the drones start to hover.

Figure 1.1: The drones used in the project, WLtoys V262 Cyclone Quadcopter on the left and Hubsan X4 on the right.

1.2 Related Work

No research on detection of drones using audio as the only source of detection has been found. However a number of articles trying to detect other sounds using audio analysis have been encountered, the most related concerning speech detection since they use similar techniques. Detection of vehicles have also been a popular subject among searched articles [4][5]. Articles about drone detection have been found when other attributes than sound have been used in the detection.

Although no articles on the subject have been found, there are products on the market that market themselves as drone detectors and claim to use sound as the only source of detection, such as DroneShield [6] and DRONE-DETECTOR [7]. The product Drone Detector [8] use multiple factors, including sound, to detect drones.

1.3 Disposition

Chapter 2 introduces the background needed for the remainder of the report.

Chapter 3 discusses how the authors chose which parameters to use for the processing of the signal before performing linear predictive coding.

Chapter 4 introduces the considered additions to the detector.

Chapter 5 describes how the database is created and how the comparison is done.

Chapter 6 presents how the final solution was implemented on the DSP.

Chapter 7 describes the test procedure and presents the results.

Chapter 8 discusses the results and possible further improvements.

Chapter 9 summarizes the thesis.

Appendix A describes how the recordings used in the thesis were carried out.

Appendix B contains the limits used in the database.

Appendix C presents the hardware used in the implementation.

Chapter 2
Scientific Background

The goal of the thesis is to create a drone detector which can detect drones using audio analysis. Detection is performed by comparing the similarities on selected parameters between the incoming sound and values in a database. Multiple different techniques exist in finding parameters of sound to compare. The focus of the thesis became coefficients gained from speech recognition methods. When detecting human speech the three most used techniques are linear predictive coding (LPC), mel-frequency cepstrum coefficients (MFCC) and perceptual linear predictive (PLP) analysis. MFCC and PLP analysis are both based on how humans perceive sound. The methods could therefore be unsuitable for detecting drones as the detector should not interpret the sound it receives [9]. LPC requires the analyzed sounds to have spikes in the spectral envelop as human speech does. Studied drone sounds showed similarities to human speech, which made LPC a possibility, consequently LPC was chosen. The thesis will also consider the slope of the frequency spectrum and the zero crossing rate (ZCR) as additions to the detection.

2.1 Linear Predictive Coding

Linear predictive coding is mostly used when analyzing speech, the name comes from that at a given time n the speech signal $s(n)$ can be approximated as a linear combination of the past p speech samples and an excitation term, $Gu(n)$. The equation is given by

$$s(n) = \sum_{i=1}^{p} a_i s(n-i) + Gu(n) \tag{2.1}$$

where a_i are the LPC coefficients which are assumed constant over the frame being analyzed. From (2.1) the transfer function for the speech signal can be derived:

$$H(z) = \frac{S(z)}{GU(z)} = \frac{1}{1 - \sum_{i=1}^{p} a_i z^{-i}}. \tag{2.2}$$

Linear predictive coding gives the possibility of estimating the sound using past samples. The estimation of the signal is given by

$$\tilde{s}(n) = \sum_{i=1}^{p} a_i s(n-i). \tag{2.3}$$

From the exact relation and the estimation of the signal, the prediction error, $e(n)$, can be defined as [10, pp.100-102] [11, pp.115,117]

$$e(n) = s(n) - \tilde{s}(n) = s(n) - \sum_{i=1}^{p} a_i s(n-i). \tag{2.4}$$

The problem in linear prediction analysis is to find predictor coefficients, a_i, that minimize the mean square error signal,

$$\sum_{n} (s(n) - \tilde{s}(n))^2. \tag{2.5}$$

These coefficients need to be estimated from sufficiently short segments since the characteristics only are stationary during short periods of time. When minimizing the mean square error with respect to a_i, a system of p equations to solve for the coefficients is obtained. This system can be solved using different methods. The LPC coefficients give an accurate representation of the signal when the method is used on stationary signals. Since the input signal is seldom stationary, it is divided into smaller frames that are quasi-stationary to improve the result.

The covariance method and the autocorrelation method are two common approaches to calculate a_i, their difference lies in the estimation of the autocorrelation $r_n(i)$. The covariance method does not always give a solution. The autocorrelation method on the other hand always does, and for that reason the autocorrelation method was chosen. [10, pp.102] [11, pp.117]

2.1.1 Autocorrelation Method

The most common technique to calculate the LPC coefficients is the autocorrelation method. This method needs to window the signal to obtain a good result. The signal is divided into frames by multiplying it by a windowing signal that is zero outside the interval of interest:

$$s_n(m) = \begin{cases} s(m+n) * w(m), & 0 \leq m \leq N-1 \\ 0, & \text{otherwise} \end{cases} \tag{2.6}$$

where N is the desired window length in samples. The windowing is also supposed to taper the signal at the ends of the desired interval to avoid smearing in the frequency domain and minimize the error at section boundaries. A hamming window is often used as the windowing function:

$$w(n) = \begin{cases} 0.54 - 0.46\cos(\frac{2\pi n}{N}), & 0 \leq n \leq N-1 \\ 0, & \text{otherwise} \end{cases} \tag{2.7}$$

where N is the desired window length in samples. There are other windowing functions that satisfy these requirements. The hamming window was selected because of its frequent use in LPC. [10, pp.103,105][11, pp.115-116]

The autocorrelation of a signal for the autocorrelation method is calculated by

$$r_n(i) = r_n(-i) = \sum_{k=0}^{N-i-1} s(k)s(k+i) \tag{2.8}$$

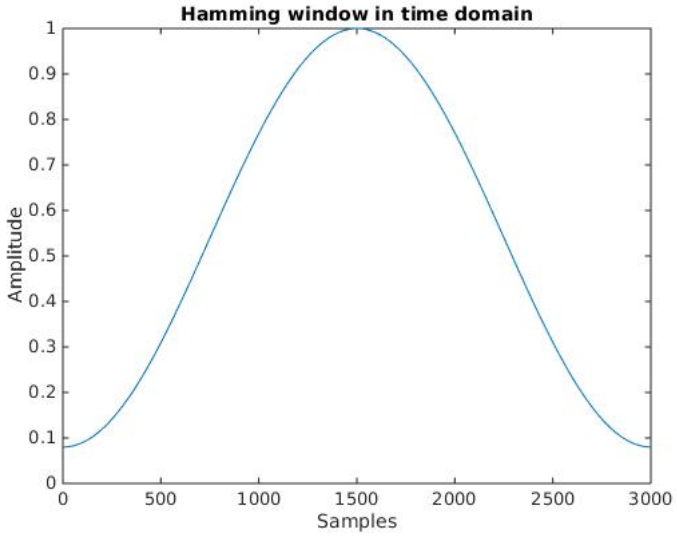


Figure 2.1: Hamming window of length 3000 in the time domain.

where i is a value greater than or equal to one. The autocorrelation equation for linear predictive coding is derived from the attempt at minimization of the mean square error signal in (2.5). Since the autocorrelation is symmetric the resulting equations to be solved for a_i can be written as

$$r_n(i) = \sum_{k=1}^{p} a_k r_n(\mid i-k \mid) \quad 1 \leq i \leq p \tag{2.9}$$

which in matrix form becomes

$$\begin{bmatrix} r_n(0) & r_n(1) & r_n(2) & \cdots & r_n(p-1) \\ r_n(1) & r_n(0) & r_n(1) & \cdots & r_n(p-2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_n(p-1) & r_n(p-2) & r_n(p-3) & \cdots & r_n(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} r_n(1) \\ r_n(2) \\ \vdots \\ r_n(p) \end{bmatrix} \tag{2.10}$$

or

$$R \cdot \boldsymbol{a} = \boldsymbol{r}. \tag{2.11}$$

The R matrix is a Toeplitz matrix, a matrix with identical elements in its diagonals. This type of matrix always has an inverse, therefore the coefficients can always be calculated by

$$\boldsymbol{a} = R^{-1} \cdot \boldsymbol{r}. \tag{2.12}$$

The Toeplitz matrix also makes it possible to use fast methods for the computation, for example the Durbin recursion which can solve (2.12) in $\mathcal{O}(p^2)$ instead of $\mathcal{O}(p^3)$ which is the complexity of the Gaussian method. [10, pp.106][11, pp.117-119]

2.1.2 Durbin Algorithm

The filter coefficients can be computed recursively by the following equations:

$$E^{(0)} = r(0) \tag{2.13}$$

$$k_i = \left\{ r(i) - \sum_{j=1}^{i-1} \alpha_j^{(i-1)} r(i-j) \right\} / E^{(i-1)}, \qquad 1 \leq i \leq p \tag{2.14}$$

$$\alpha_i^{(i)} = k_i \tag{2.15}$$

$$\alpha_j^{(i)} = \alpha_j^{(i-1)} - k_i \alpha_{i-j}^{(i-1)}, \qquad 1 \leq j \leq i-1 \tag{2.16}$$

$$E^{(i)} = (1 - k_i^2) E^{(i-1)}, \tag{2.17}$$

where $r(i)$ is the ith autocorrelation for frame l (frame index left out of the equations) and $E^{(i)}$ is the prediction error for a predictor of order i. $E^{(i)}$ can be used to decide on the number of coefficients to use. The above equations results in LPC coefficients, $a_m = \alpha_m^{(p)}$, and reflection coefficients, k_m, for $1 \leq m \leq p$. [10, pp.115,412][11, pp.121]

2.1.3 Different Coefficients

From LPC different coefficients can be derived. All of them are equivalent to each other, but they have some different characteristics which makes them more suitable to different areas of usage. Two of the most common are LPC coefficients and reflection coefficients.

LPC Coefficients

The LPC coefficients are the a_i from the Durbin recursion. All a_i depend on each other, which means that all calculations must be redone if a different number of coefficients are required.

Reflection Coefficients

The reflection coefficients are the k_i from the Durbin recursion. They are also called PARCOR (PARtial CORrelation) coefficients and Lattice coefficients. The reflection coefficients are ordered from most important to least important. The first coefficient holds the most information and the second then holds some of the information that is left in the signal and so forth. When the coefficients are close to zero, there is nothing more to predict and increasing the number of coefficients will not improve the result.

All reflection coefficients will always have the same values for the same recording independent of the number of coefficients used, therefore one coefficient at a time can be calculated without affecting the result. The reflection coefficients are a variant of more robust LPC coefficients.

2.2 Power Spectral Density

This thesis has used power spectral density (PSD) for analyzing signals. The contents of a signal can be represented as the frequencies it is composed of by transforming the time-domain signal into the frequency domain. A higher spike for a frequency means that it is more prominent in the signal than a lower spike. The power spectral density can be used for finding periodicity or observing which frequencies are present in the signal.

2.3 Simple Linear Regression

Simple linear regression is used to find the line with the least amount of distance to all data points under consideration. The result of the regression is a trend line for the points. The line is calculated by considering,

$$y_i = \alpha + \beta x_i + \epsilon_i \tag{2.18}$$

where ϵ_i is the least square error, that minimizes

$$\sum_{i=1}^{n} \hat{\epsilon}^2 = \sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2. \tag{2.19}$$

From (2.19) β can be derived as

$$\hat{\beta} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} \tag{2.20}$$

where

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{2.21}$$

and

$$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i. \tag{2.22}$$

After a signal has been transformed with FFT from the time domain to the frequency domain, simple linear regression can be used to calculate the trend of the frequency spectrum. The trend varies depending on which frequencies are more significant in the signal.

2.4 Zero Crossing Rate

Zero crossing rate (ZCR) is the rate a sampled sound, $s(n)$, switches sign as given by [12, pp.127]

$$\mathrm{ZCR} = \frac{1}{2N} \sum_{n=1}^{N} \mid \mathrm{sign}[s(n)] - \mathrm{sign}[s(n-1)] \mid \tag{2.23}$$

where

$$\mathrm{sign}[s(n)] = \begin{cases} = 1, & s(n) \geq 0 \\ = -1, & s(n) < 0. \end{cases} \tag{2.24}$$

Chapter 3
Base of the Detection

The first step in creating a drone detector was to study audio from different drones to find what differentiate them from other types of sound. Since the power spectral density had similarities to that of speech, the LPC method which is mostly used for speech recognition was considered. LPC is only dependent on spikes in the frequency spectrum which audio from drones contain, for that reason coefficients gained from LPC were chosen as the base of the detection. The reflection coefficients were chosen instead of the LPC coefficients because of their robustness.

3.1 Recordings Used

Sounds of Cyclone and X4 recorded in an anechoic chamber at 30%, 40%, 50%, 60%, 70%, 80% and 90% throttle level, were used to decide on which parameters to use for the LPC analysis. See appendix A.1 for more information about the recordings.

3.2 Deciding Parameters

The sampling frequency, filter, frame size and the number of reflection coefficients to use were decided in parallel. Therefore the figures shown use the decided values for the other parameters for clarity. In addition to these parameters the hamming window needs the frames to overlap as it taper the signal at the ends, otherwise information in the signal is lost. The overlap was chosen to be one third of a frame size to counter the effect of the necessary tapering.

3.2.1 Sampling Frequency

The PSD of the drones were analyzed to decide which frequencies were necessary to keep in order to not loose needed information from the signal. Another concern was that the sampling frequency should be suitable for both of the drones. The similarity of the drones' PSD can be seen in figure 3.1.

There are spikes at frequencies near 0 Hz which is considered to be noise. From the PSD it can be seen that there are interesting frequencies from 400 Hz and after 8 kHz there are less information, except for the X4 which have spikes around 12 kHz as well.

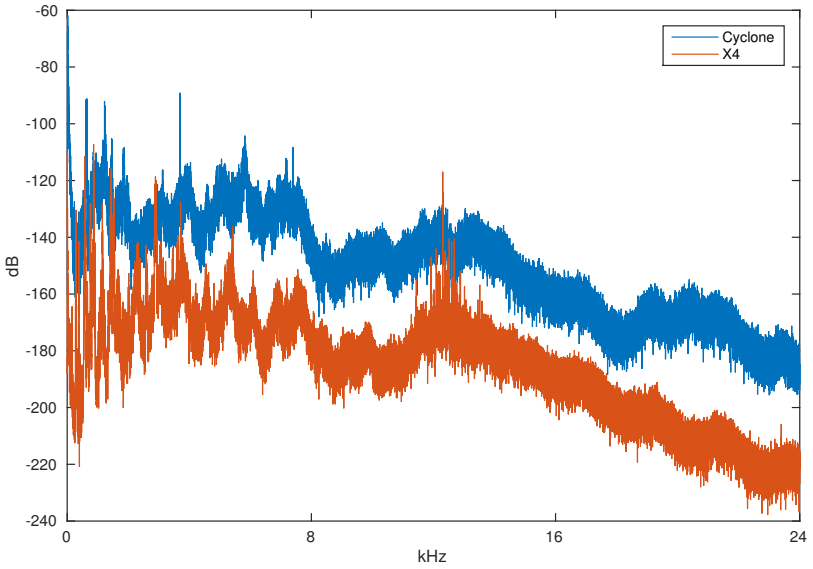


Figure 3.1: Power spectral density of Cyclone and X4 created from the recordings in the anechoic chamber at 50% throttle level. Created with Matlab's pwelch method.

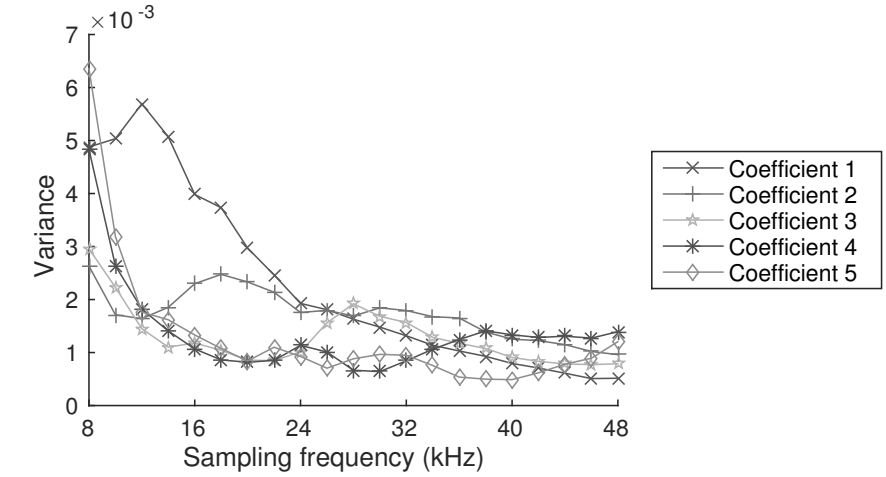
The sampling frequency was decided by checking the prediction error and how the coefficients varies for seemingly constant drone sound. From the PSD in figure 3.1 it could be seen that the lowest sampling frequency should be at least 16 kHz. A high sampling frequency would require the detection algorithm to be faster, therefore the highest limit was set to 48 kHz.

The variance was studied per coefficient and the variance of the lower coefficients was considered more important than that of the higher, because the reflection coefficients have more information in the lower coefficients. It can be seen in figure 3.2 that the variance for all the coefficients gets lower as the sampling frequency gets higher. After 32 kHz the overall variance gets stable but the variance of coefficient one and two decreases all the way to 48 kHz. In table 3.1 the mean prediction error for five coefficients is shown. Similar to the variance the mean prediction error decreases as the sampling frequency increase. From these figures and corresponding figures for other throttle levels, a sampling frequency of 48 kHz was selected. These results were independent of the frame size.

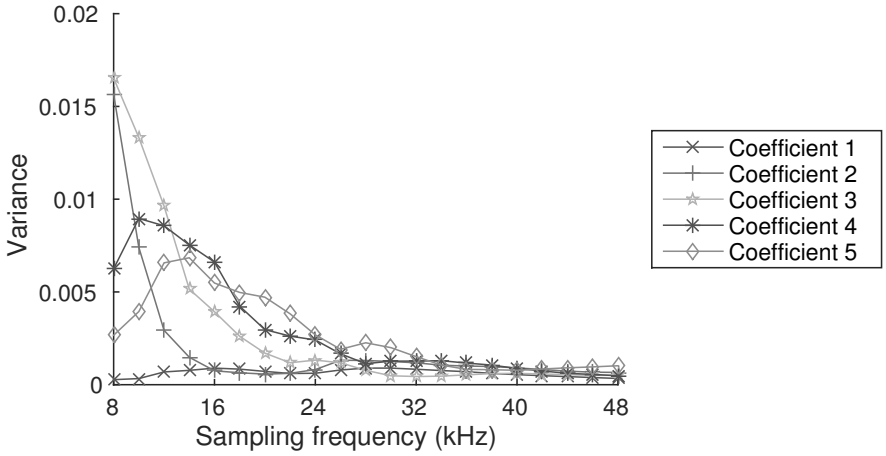
3.2.2 Filter

The incoming sound to the detector could need filtering to get rid of unwanted frequencies. On the other hand the filter should not affect the reflection coefficients and therefore needs to be tested.

At first a Butterworth filter was considered in order to get more similar reflec-



(a) Cyclone



(b) X4

Figure 3.2: How the variance of the coefficients changes based on the sampling frequency for the drones in the anechoic chamber at 50% throttle level. A frame size of 3000 samples and an overlap of 1000 samples were used.

	Sampling Frequency (kHz)								
	16	20	24	28	32	36	40	44	48
Cyclone	8.86	7.72	7.21	7.97	7.42	5.49	4.03	3.34	3.09
X4	0.18	0.15	0.15	0.17	0.15	0.11	0.09	0.07	0.05

Table 3.1: The mean prediction error for Cyclone and X4 in the anechoic chamber at 50% throttle level for different sampling frequencies in kHz. Five reflection coefficients, a frame size of 3000 samples and an overlap 1000 samples were used.

		Coefficient				
		1	2	3	4	5
Cyclone	Unfiltered	0.0026	0.0022	0.0018	0.0022	0.0017
	Filtered	0.0008	0.0014	0.0011	0.0016	0.0015
X4	Unfiltered	0.0009	0.0015	0.0019	0.0014	0.0015
	Filtered	0.0009	0.0010	0.0015	0.0011	0.0011

Table 3.2: The variance of the coefficient values with and without the notch filter (3.1) of Cyclone and X4 in the anechoic chamber at 50% throttle level.

tion coefficients for different frames by removing unwanted frequencies. It turned out that the coefficients were affected by the filter and became almost identical to the filter's transfer function. This gave coefficients from all types of sound a similar look which means that no significant filtering could be applied. Some filtering still seemed appropriate and therefore a check against white noise was introduced to control the effect of using a filter. White noise receives values close to zero for all reflection coefficients and should after being filtered also have coefficients close to zero. Otherwise the filter have affected the coefficients and any coefficients gained using the filter are not reliable.

Since the lower frequencies in the figure 3.1 seems to be noise, a filter to remove them seemed appropriate. A notch filter with limiting frequency at zero to only remove the lowest frequencies was tested. The check against white noise showed that a notch filter which cut off a small frequency band did not affect the coefficients negatively. After some tests a notch filter with transfer function

$$H(z) = \frac{1 - z^{-1}}{1 - 0.99z^{-1}} \tag{3.1}$$

was selected. This filter only filters out the very lowest frequencies, its response can be seen in figure 3.3. It lowers the variance of the coefficients as can be seen in table 3.2 but the prediction error is only lowered for the Cyclone, as seen in table 3.3.

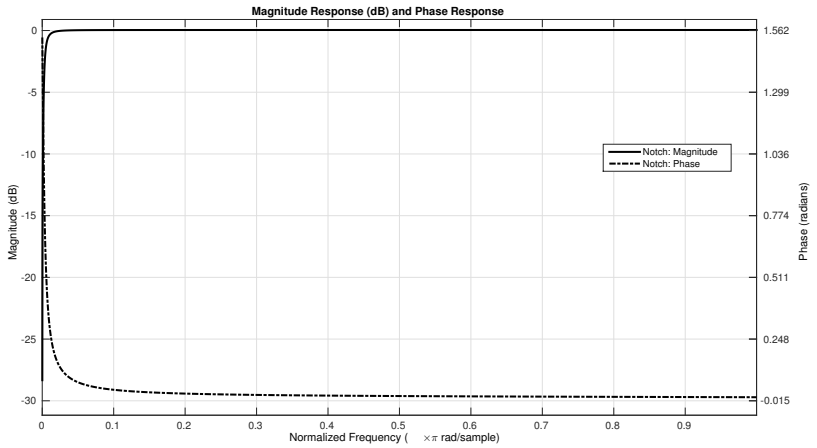


Figure 3.3: The magnitude and phase response for the notch filter given in (3.1).

		Number of Coefficients				
		1	2	3	4	5
Cyclone	Unfiltered	2.4	2.0	1.6	1.6	1.4
	Filtered	2.3	1.7	1.5	1.4	1.3
X4	Unfiltered	0.06	0.05	0.04	0.03	0.03
	Filtered	0.06	0.05	0.04	0.03	0.03

Table 3.3: The mean prediction error for different number of coefficients with and without the notch filter (3.1) of Cyclone and X4 in the anechoic chamber at 50% throttle level.

	Frame size (samples)										
	1000	1400	1800	2200	2600	3000	3400	3800	4200	4600	5000
Cyclone	1.01	1.43	1.84	2.25	2.66	3.08	3.49	3.90	4.31	4.75	5.15
X4	0.0166	0.0234	0.0302	0.0369	0.0435	0.0504	0.0570	0.0638	0.0707	0.0775	0.0841

Table 3.4: The mean prediction error for Cyclone and X4 in the anechoic chamber at 50% throttle level for different frame sizes. Five reflection coefficients, a sampling frequency of 48 kHz and an overlap of one third of the frame size were used.

3.2.3 Frame Size

A frame size that minimize both the prediction error and the variance of the reflection coefficients would be ideal. A small frame size decrease the prediction error while a large frame size does the opposite as can be seen in table 3.4. The variance of the coefficients behaves in the opposite way as can be seen in figure 3.4. It is therefore a compromise between reliable coefficients and a good prediction, and a difficult choice. The decision was based on the figures presented and corresponding figures for other throttle levels which showed equivalent results. A frame size of around 3000 samples was chosen because the variance and the prediction error were low enough to give a good result.

3.2.4 Number of Reflection Coefficients

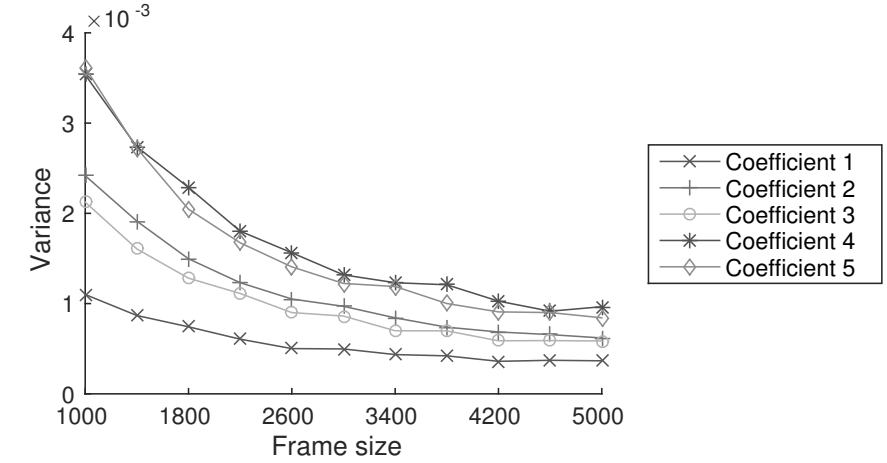
To decide the number of reflection coefficients of interest in the linear predictive coding, the size of the prediction error was studied to see where the difference between using n or $n+1$ coefficients made no significant difference on the prediction error. There is no use in having coefficients that do not lower the error significantly since it takes more computing power. The value of the coefficients should also be considered. When the coefficients approaches zero there is nothing more to predict in the signal that the previous coefficients have not already taken into account. This comes from that the first reflection coefficient contains the most information and then the next contains some of what is left in the signal and so forth.

In figure 3.5 it can be seen that after coefficient five for X4, and coefficient six for Cyclone, there is less information. This can also be seen in figure 3.6. This indicates that five or six coefficients are suitable. The evaluation was done for both drones at 50% and 60% and the decision was to use the first five coefficients.

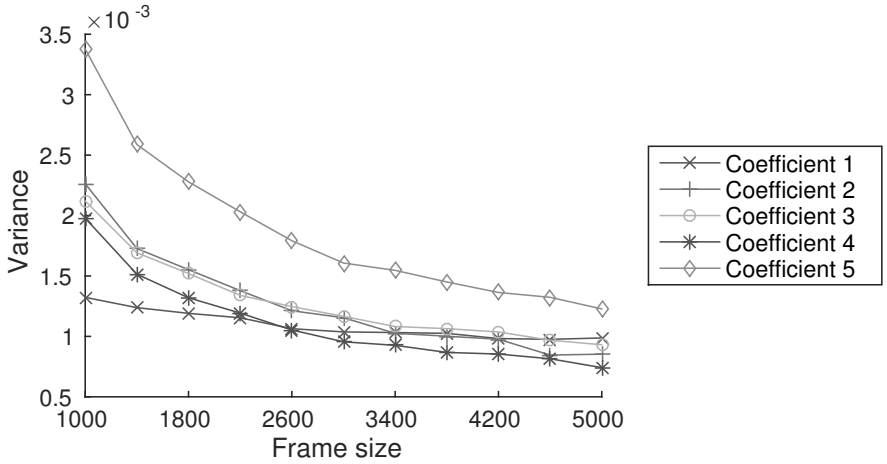
To add additional coefficients in order to rule out non-drone sounds was considered. The coefficient values after coefficient five are affected by surrounding noise which made them unpredictable, and adding more coefficients to the detection risk eliminating drones.

3.2.5 Computing Reflection Coefficients

One set of reflection coefficients are computed for each frame. The selected frame size and sampling frequency make one frame 61 ms long. The mean value is taken of the coefficients of eight subsequent frames. These frames have an overlap of



(a) Cyclone



(b) X4

Figure 3.4: How the variance of the coefficients changes based on the frame size for the drones in the anechoic chamber at 50% throttle level. A sampling frequency of 48 kHz and an overlap of one third of the frame size were used.

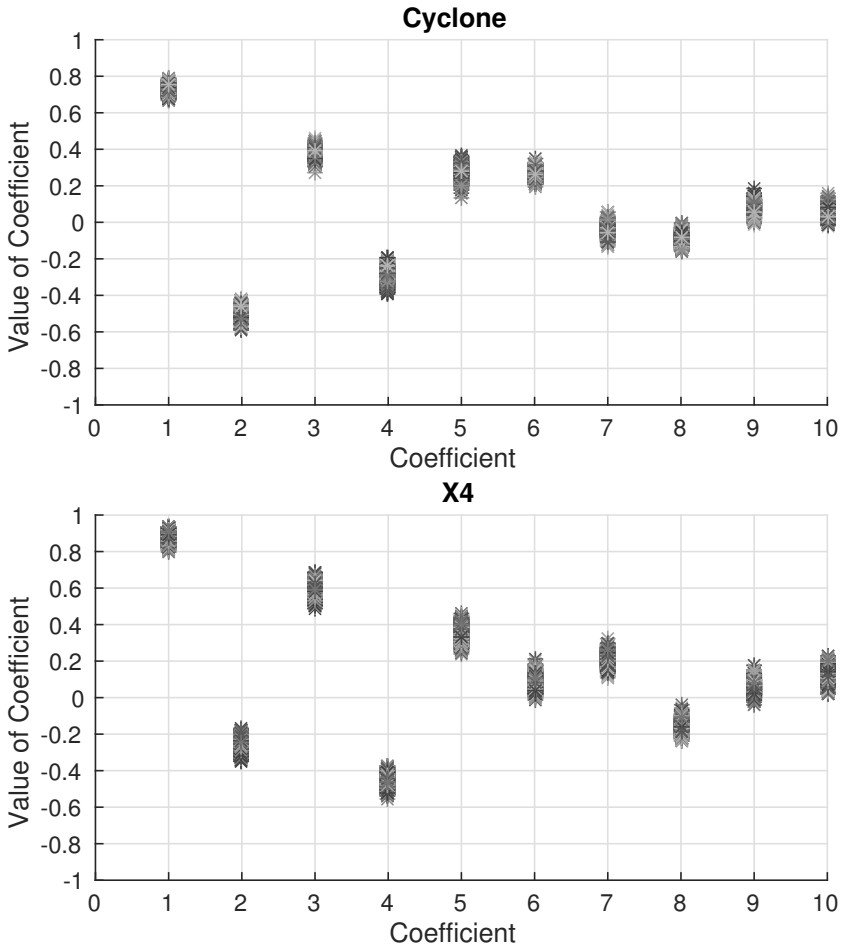


Figure 3.5: Values of the 10 first coefficients for a whole audio file containing X4 and Cyclone at 50% throttle level in the anechoic chamber. A frame size of 3000 samples, an overlap of 1000 samples and a sample frequency of 48 kHz were used.

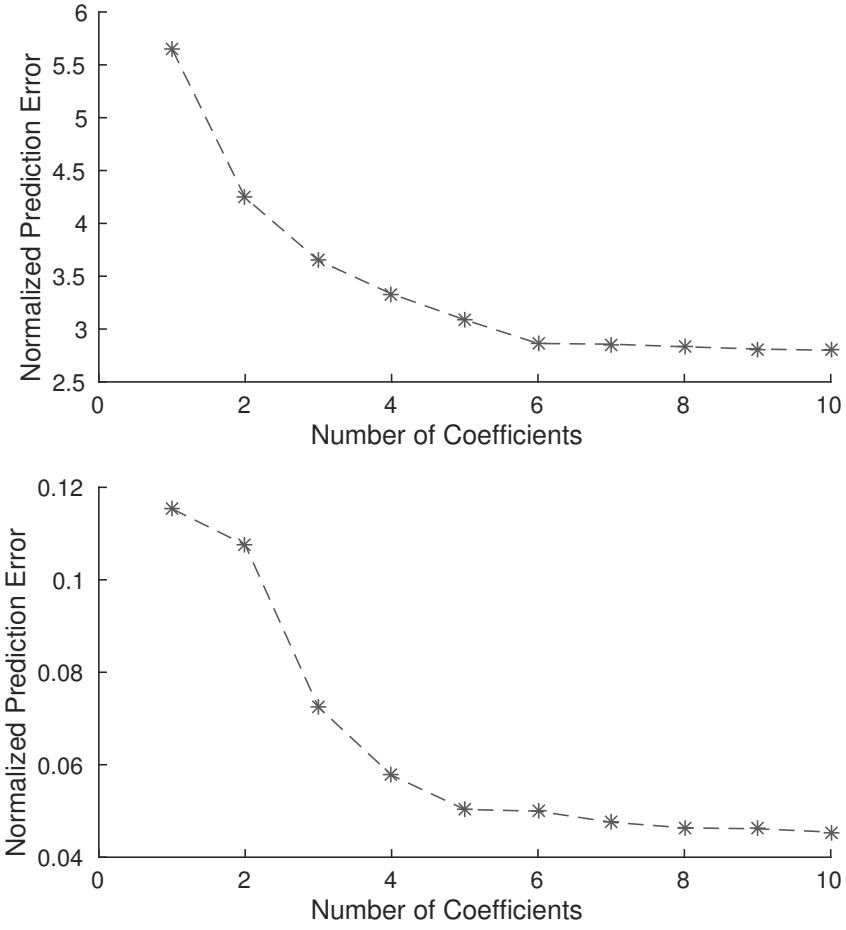


Figure 3.6: The mean prediction error for a whole audio file containing X4 and Cyclone at 50% throttle level in the anechoic chamber. A frame size of 3000 samples, an overlap of 1000 samples and a sample frequency of 48 kHz where used.

1024 samples which results in that we get the coefficients for 2^{14} samples (341 ms) at a time during the comparison.

This method applied to incoming sound is also the method used to create the values in the database because the comparison should be made against something equivalent.

Chapter 4

Possible Additions to the Detection

By using the reflection coefficients as the single characteristic to compare the incoming sounds against, the detector need to have quite large allowed coefficient intervals to take into account that sounds outside of the anechoic chamber is affected by their surroundings. This could mean that the amount of false positives becomes relatively high. A solution to this problem would be to add characteristics specific to drones to check against in order to minimize the false positives. The proposed additions in this chapter deals with this possible problem.

4.1 Slope of the Frequency Spectrum

The shape of a power spectrum depends on the frequencies present in the considered sound. For this reason different sounds have different power spectra. The difference in power spectra could be advantageous in distinguishing drones from other similar sounding sounds as the overall trend could differ. Figure 4.1 shows how four sounds that a human ear might consider similar to a drone has differing trends.

At a frequency range of 0 to 12 kHz the frequencies for Cyclone in figure 4.1 are close to being equally distributed. Applying a simple linear regression to find the slope would give a result close to zero. The other sounds, except for the white noise which is there for comparison, have a steeper slope. This result applies to X4 and other throttle levels as well. Based on these observations the slope of the frequency spectrum up to 12 kHz was considered as an extra characteristic to distinguish drones from other sounds.

4.2 Zero Crossing Rate

When reading about different aspects that may differ in sounds, the ZCR was found and considered as a possible addition. After a brief analysis of one recording of each drone, sound from a lawn mower, white noise and one file recorded outside the results shown in table 4.1 were obtained.

These results indicate the possibility for removing false positives using the zero crossing rate as an improvement to the detector. For that reason zero crossing rate was added to the detector and tested to evaluate its effect.

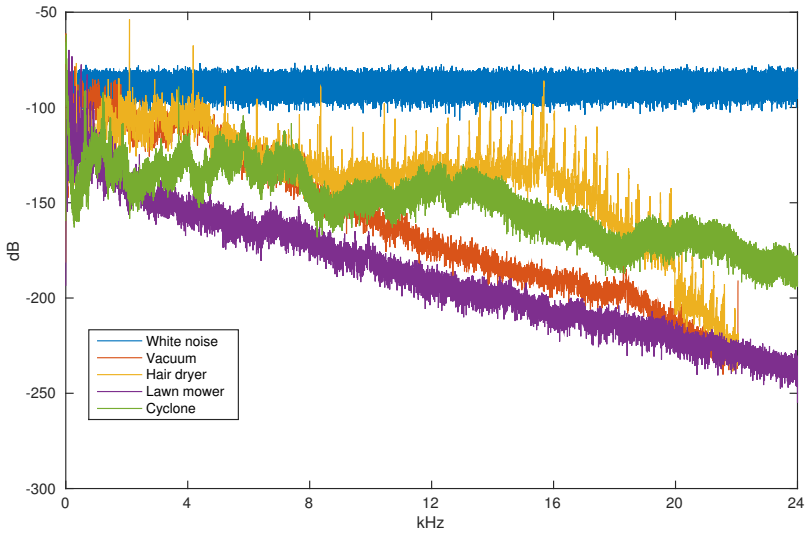


Figure 4.1: Power spectral density of sounds similar to drones created with Matlab's pwelch method. The recording of Cyclone was created in the anechoic chamber at 50% throttle level.

Sound	**min ZCR**	**max ZCR**
Cyclone	0.203	0.272
X4	0.090	0.201
Lawn mower	0.023	0.055
White noise (randomly generated)	0.470	0.537
Outside (no particular sound)	0.009	0.074

Table 4.1: Minimum and maximum ZCR for a frame for different recordings. A sampling frequency of 48 kHz, a frame size of 2048 samples and recordings of the drones at 50% throttle level in the anechoic chamber were used.

4.3 Source Localization

If the detector could calculate the angle the sound originates from, all sounds below a certain angle could be eliminated. If this is possible it means that no previous false positives that originated from the ground would be considered a drone, which would improve the detector.

The only interesting angle to consider would be the angle between the horizontal plane and the origin of the sound. To be able to obtain this angle, a minimum of two microphones would have to be used. They should be placed straight above each other in the vertical plane to be able to perform this calculation.

Chapter 5
Database

The detector should work by containing a database with a set of parameter values of known drones to compare incoming sound against. When the detector is up and running it should create parameter values for a sound of a suitable length and then these parameter values will be compared against the saved values. If they are close enough, a match will be signaled, otherwise nothing will happen. Parameter values are saved for both Cyclone and X4 for different throttle levels to be able to detect them at all times.

5.1 Comparison with Reflection Coefficients

The database of reflection coefficients will be created from the recordings from the anechoic chamber for each drone and throttle level, see appendix A.1. From each recording, the mean values for the reflection coefficients for eight subsequent frames are calculated, these values are then used to create the database. The middle point of the mean values for one recording and the distance to the outermost value are saved for each coefficient, see figure 5.1. This equals two values per coefficient and recording. The expectation is that most values during the detection should be within the interval created from the recordings from the anechoic chamber. In reality there will be frames with coefficient values outside the interval and for that reason the coefficients of the incoming sound are allowed to be a specified distance from the interval, this distance is called the limit, see figure 5.1. This limit is specified per coefficient, independent of drone and throttle level. To decide this limit, recordings of the drones outside at different distances from the microphones were used, see appendix A.2. The coefficient values of the frames of these recordings were evaluated and the limits set in order for close to 100% of the frames being detected. The limits were set large because the goal of using the reflection coefficients is to detect the drones at all times. This could possibly mean that detection is signaled for some non-drone sounds as well, but in that case it will hopefully be removed by the suggested additions.

Cyclone was given more weight when deciding the limits because the size of Cyclone is more likely to be the size of drones of interest. This conclusion was drawn because X4 is sensitive to wind and have quite low signal to noise ratio, which makes it hard to detect outside.

When the detector is running, a mean value of the reflection coefficients of

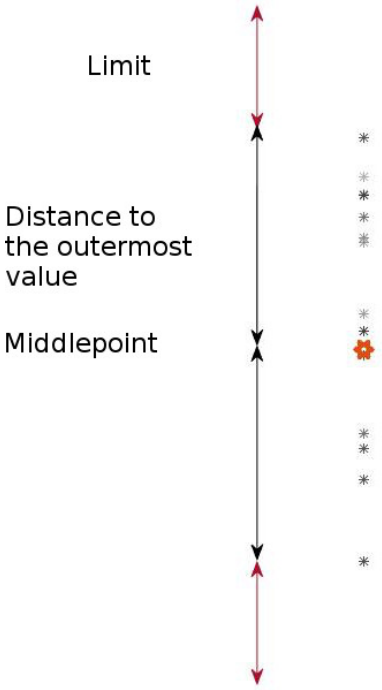


Figure 5.1: An illustration of how the intervals in the database were created from the values of the reflection coefficients.

eight subsequent frames will be taken and compared against the saved values in the database. If the values of all the coefficients fall within the allowed intervals it is considered to be a drone.

5.2 Comparison with Slope of the Frequency Spectrum

There is only one value stored in the database related to the slope of the frequency spectrum. The value is the lowest possible slope that the authors want to consider to belong to a drone. Only frames with a slope between that value and zero will be considered drones, this means that both the reflection coefficients and the slope of the frequency spectrum must fall within the allowed intervals.

The value was chosen by studying the slope in table 5.1 of the recordings of the drones outside at different distances from the microphones, see appendix A.2 for a description of the recordings.

5.3 Comparison with Zero Crossing Rate

Only one value is stored in the database for the ZCR addition. This value is used to make sure that the ZCR is below the value to be considered a drone. The value is set quite high in order to not reject any drone but hopefully it can still reject some non-drones.

The value was decided by studying the ZCR in table 5.1 for the recordings of the drones outside at different distances from the microphones, see appendix A.2 for a description of the recordings. A lower limit except for zero was impossible to set as the ZCR varied too much. Because of the lack of a lower limit other than zero, the table 4.1 showed less promise. The ZCR was still included in the testing.

	Distance	Slope (min)	Slope (max)	ZCR (min)	ZCR (max)
Cyclone	2m	-0.0013	-0.0008	0.1839	0.2210
	5m	-0.0024	-0.0014	0.0037	0.0940
	10m	-0.0025	-0.0014	0.0083	0.1653
	15m	-0.0026	-0.0012	0.0020	0.1321
	20m	-0.0026	-0.0014	0.0044	0.1851
	25m	-0.0026	-0.0017	0.0022	0.1175
	30m	-0.0031	-0.0019	0.0005	0.1043
	35m	-0.0032	-0.0018	0.0015	0.1067
	30m	-0.0031	-0.0023	0.0024	0.0930
X4	2m	-0.0023	-0.0012	0.0005	0.1197
	5m	-0.0023	-0.0018	0.0012	0.1079
	10m	-0.0028	-0.0021	0.0002	0.0777
	15m	-0.0033	-0.0017	0.0002	0.0381
	20m	-0.0033	-0.0023	0.0042	0.0444
	25m	-0.0039	-0.0021	0.0000	0.0359
	30m	-0.0033	-0.0014	0.0012	0.0735
	35m	-0.0031	-0.0011	0.0005	0.2232
	40m	-0.0038	-0.0011	0.0017	0.1087

Table 5.1: The minimum and maximum slope of the frequency spectrum and ZCR for the drones at different distances from the microphone.

Chapter 6
Implementation

The drone detector was implemented on a DSP from Analog Devices (ADSP-21262) together with a audio codec from Texas Instruments (TLV320AIC32). The specification for these can be seen in tables C.1 and C.2.

6.1 Software

The software is divided into four parts: the analog to digital converter, the buffer, the calculation of values, and the actual decision of detection.

The audio codec samples the incoming sound, and when the chosen block size of 960 samples is reached it signals the buffer interrupt that there are samples to attend to. The buffer interrupt then filters the samples and copies them to the buffer. When there are 1920 new samples, the number of samples needed for the next frame, it signals the calculation interrupt to start calculation of the values. The calculation interrupt calculates reflection coefficients from 2944 samples. Since the overlap was chosen to be 1024 samples this means that reflection coefficients are calculated at each calculation interrupt. The slope of the frequency spectrum and the ZCR are calculated of 2048 samples, taken without an overlap. This means that they are calculated every time a calculation interrupt occurs as well. When eight frames have been calculated of reflection coefficients, slope, and ZCR, a mean value of these calculated values are taken and the calculation interrupt signals the detection interrupt to check the database for a match. If the detector finds a match a simple sinus wave is sent to the DSP output. If no match is found the detector is quiet.

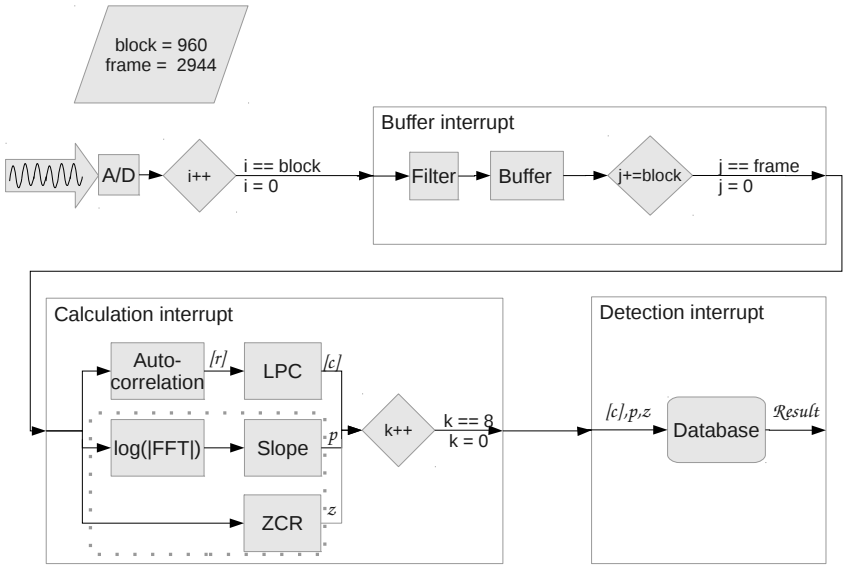


Figure 6.1: An overview of the program flow on the DSP. The parts within the dotted rectangle are the parts pertaining to the possible additions.

Chapter 7
Evaluation

The goal of the thesis was to have an implemented drone detector that runs on a DSP. To evaluate how well the chosen method works the detector will be simulated in Matlab on prerecorded sounds. The simulation during the tests is equivalent to the code implemented on the DSP, and is done in Matlab to be able to analyze the results and use the same recordings when the tests are repeated for the selected additions.

The authors are aware that the testing is not extensive enough to clearly show how well the detector will perform in a real setting, but believe that this testing will show how much potential a solution based on these methods have for the future. By using the same recordings for both the base with only the reflection coefficients and later on, with the additions made, at least a measurement of how much each addition improves the detector can be evaluated.

7.1 Recordings for Testing

The recordings used for the testing were of both Cyclone and X4 recorded as they fly in different directions. More information about the recordings can be found in appendix A.3.

7.2 Test Procedure

The program is run on each recording in Matlab. The values are calculated as described in chapter 5. They are then compared against the values in the database, and if they are within the acceptable interval, detection, called a hit, is signaled. Otherwise no hit is signaled.

For each decision made there are four possible events that can occur:

A: A hit is signaled and there was something to detect

B: A hit is signaled but there was nothing to detect (false positive)

C: No hit is signaled and there was nothing to detect

D: No hit is signaled but there was something to detect (false negative)

Where event A and C are positive outcomes since the detector made the correct decision.

The results will be measured in the quantity of correct calls divided by the number of calls made,

$$\frac{A+B}{A+B+C+D}. \tag{7.1}$$

This quotient will hereinafter be called correctness. A result close to 1 will mean that the detector works well. For the recordings containing drone sound the correctness is equal to $\frac{A}{A+B+C+D}$ since they only contain drones, and for the other recordings the correctness is equal to $\frac{C}{A+B+C+D}$ since they do not contain any drones.

7.3 Test Results

The thought behind the detector was that when only using the reflection coefficients for the detection, close to all frames containing drones should be detected. There is a large probability that quite a few of the frames not containing drones will also be detected if the sounds are similar. The plan was to decrease the number of false positives by introducing new characteristics to compare against in addition to the LPC analysis. Note that these additions will never make the detector find more drones than it did before, they will only remove false positives and possibly also correct positives if the allowed intervals are too small.

The limits for the different characteristics were optimized for two different distances. One version was made when drones should be detected up to 40 meters from the detector at all times, and one version was made for up to 20 meters. These different limits were created while prioritizing Cyclone because its size is more likely for approaching drones, the limits can be seen in appendix B.1. The result of the optimization for long distances can be seen in figure 7.1 and 7.3, and for short distances in figure 7.2 and 7.4. A description of the recordings used can be found in appendix A.3.

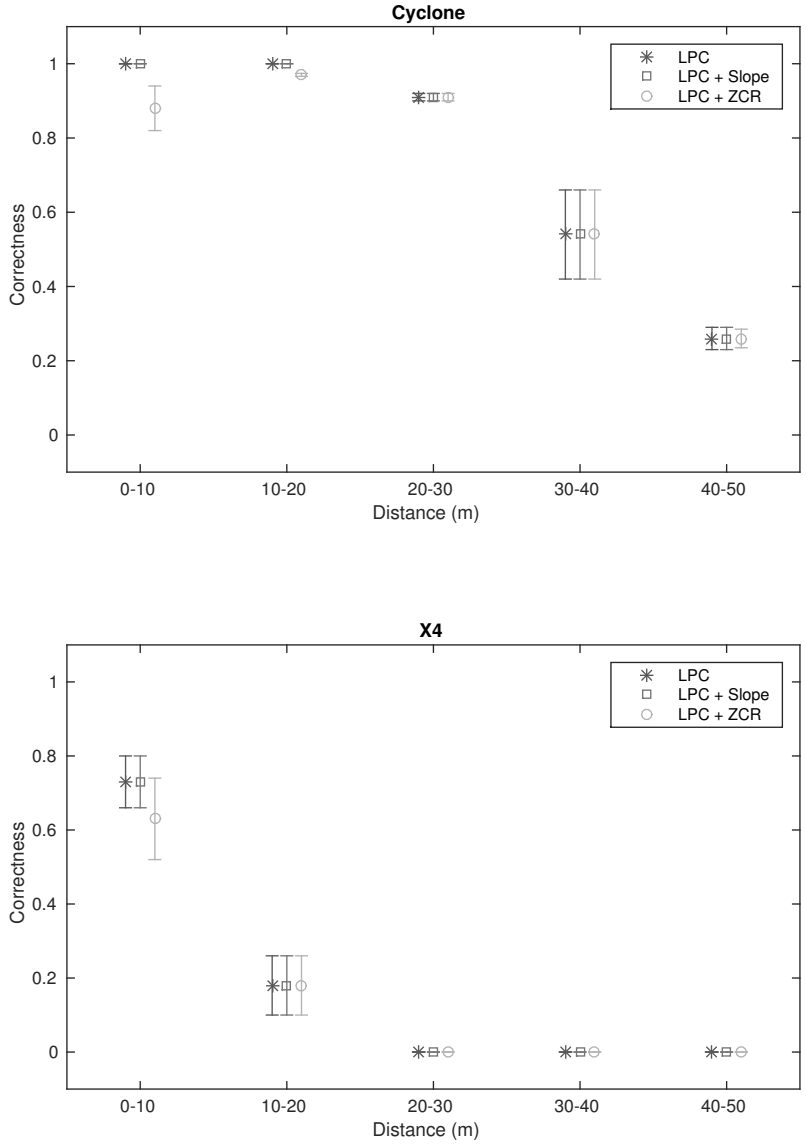


Figure 7.1: The results of the detector for Cyclone and X4 when optimized for long distances. The mean value and the variance of the correctness are shown for the detector when using only LPC, LPC and slope of the frequency spectrum, and LPC and ZCR for the comparison.

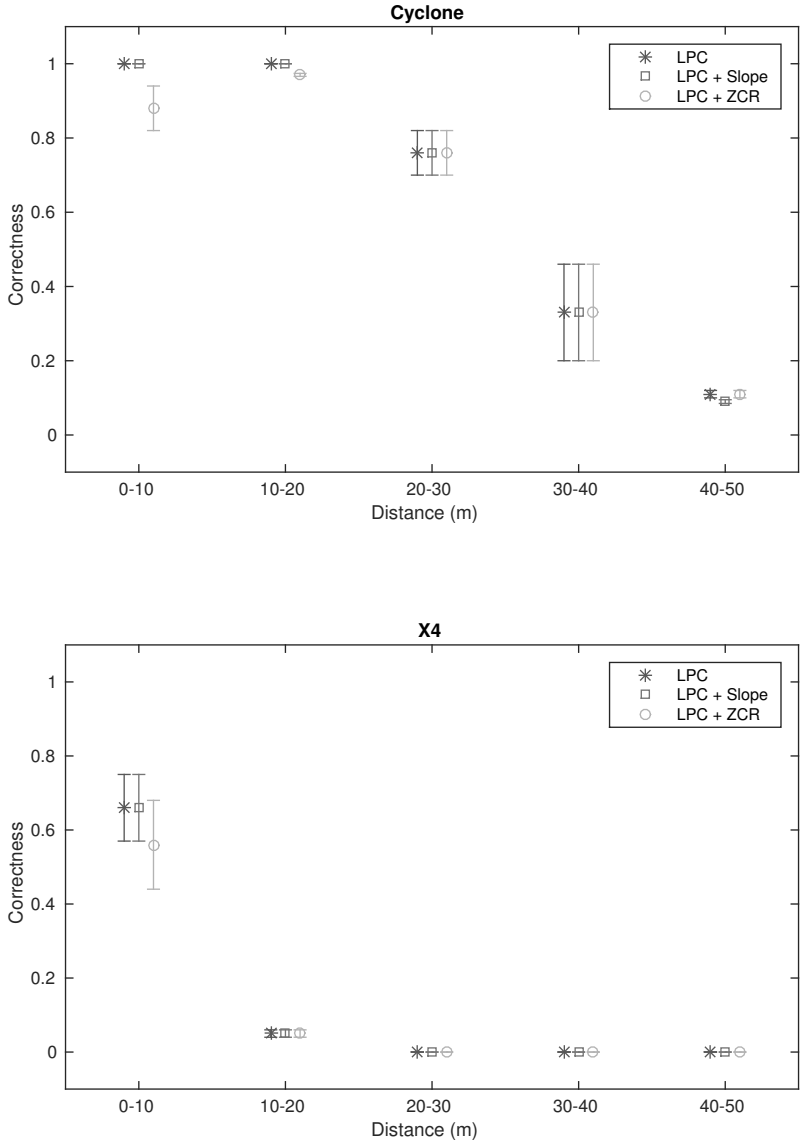


Figure 7.2: The results of the detector for Cyclone and X4 when optimized for short distances. The mean value and the variance of the correctness are shown for the detector when using only LPC, LPC and slope of the frequency spectrum, and LPC and ZCR for the comparison.

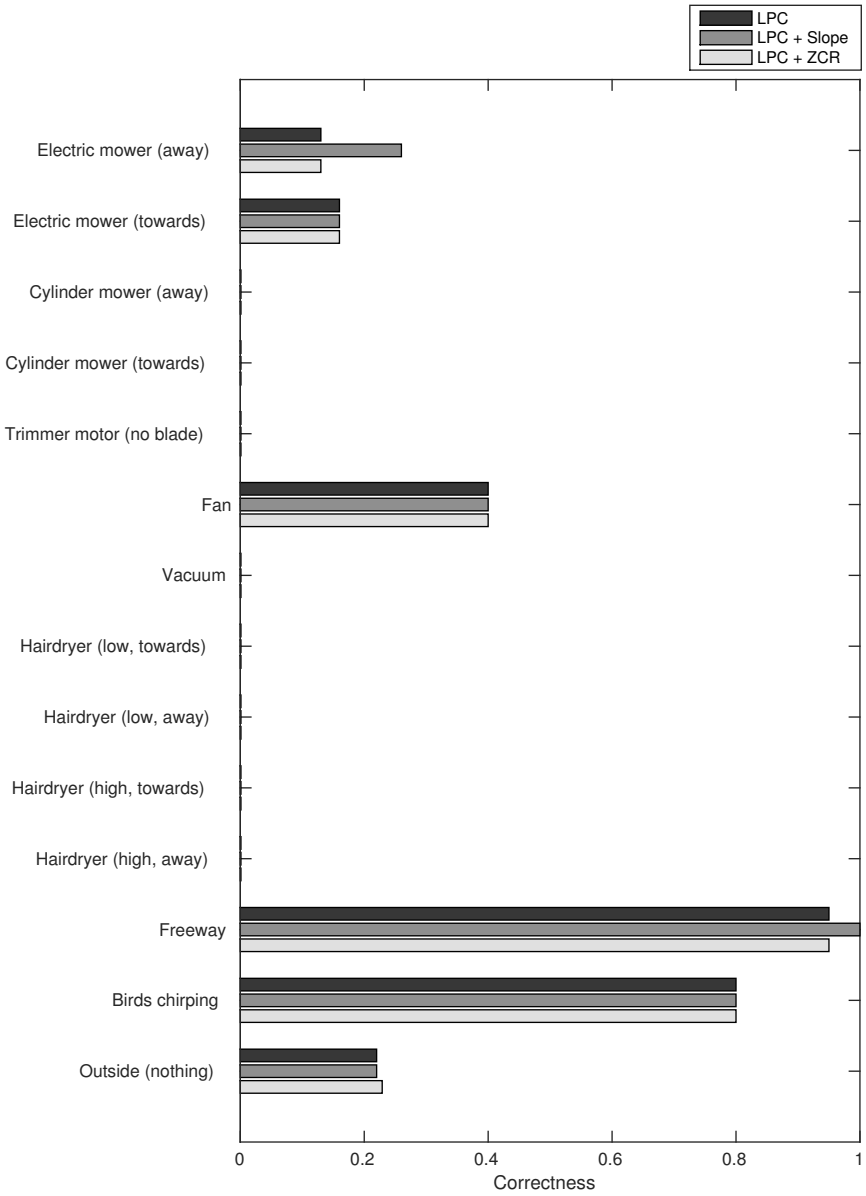


Figure 7.3: The results of the detector on recordings not containing drone sound when optimized for long distances. The correctness is shown for the detector when using only LPC, LPC and slope of the frequency spectrum, and LPC and ZCR for the comparison.

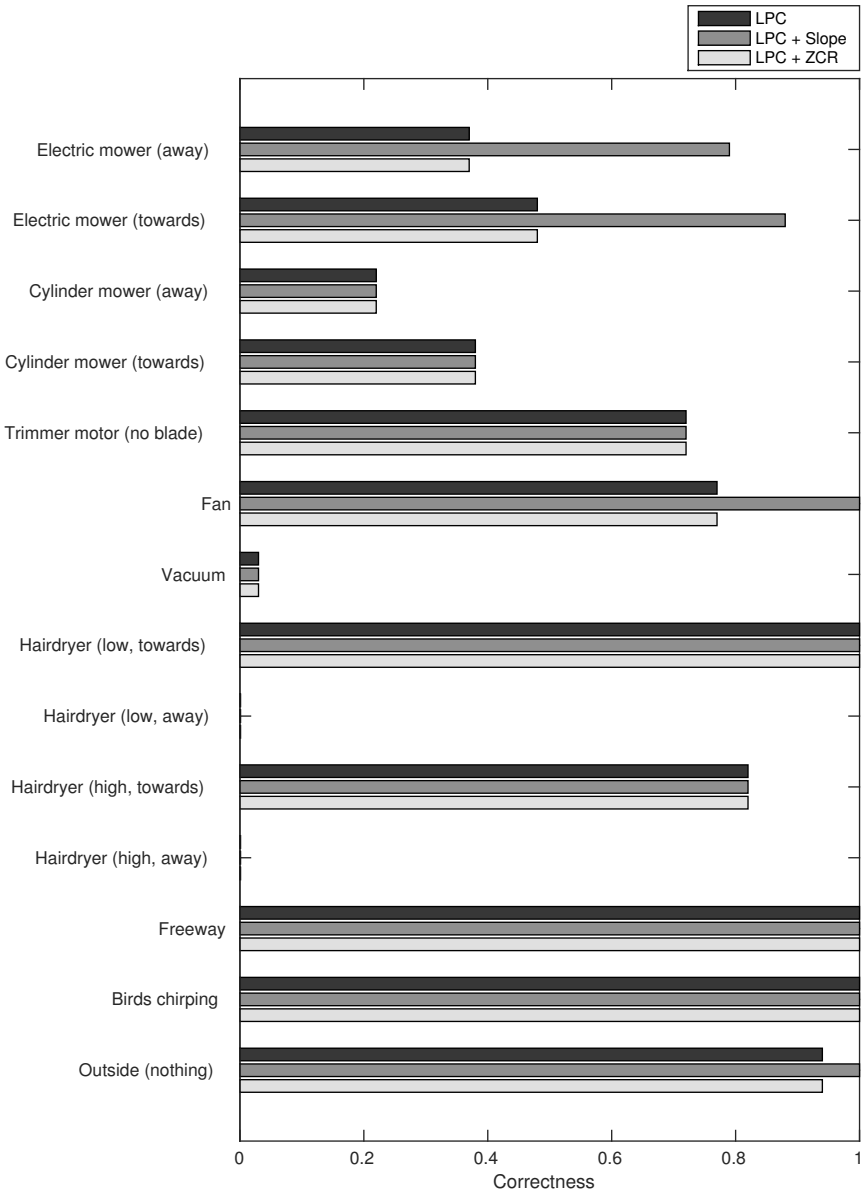


Figure 7.4: The results of the detector on recordings not containing drone sound when optimized for short distances. The correctness is shown for the detector when using only LPC, LPC and slope of the frequency spectrum, and LPC and ZCR for the comparison.

Chapter 8
Discussion

The reflection coefficient intervals let through Cyclone satisfactorily for the distance they are optimized for, X4 on the other hand get a quite low correctness. The authors believed that X4 should not be considered as important as Cyclone because of its small size and signal to noise ratio, and therefore the limits set were more adapted for Cyclone, which makes these results explainable. Sadly the intervals let through a substantial number of false positives. The difference between choosing limits with respect to long distances compared to for short distances show that fewer other sounds are detected with limits for short distances, as can be seen in figures 7.3 and 7.4. This comes as no surprise since the allowed intervals are somewhat smaller as can be seen in appendix B.

Sound that consist of a small electric motor and a fan will be hard to distinguish from a drone as they share similar hardware, this could explain some of the results. To allow less other sounds through, a different approach to the intervals for the reflection coefficients could perhaps be used. At the moment the intervals are symmetrical around the middle point which makes them quite large. It is possible that all the drones' reflection coefficients are shifted in some fixed direction from the middle point since the database middle point has been calculated from sounds from a different environment than the test environment.

The addition of the slope of the frequency spectrum to the detection did not affect the drones results negatively within the desired distance from the microphone, as can be seen in figures 7.1 and 7.2. The slope limit adapted for long range did not remove a large amount of false positives which is not surprising as the slope limit is large, but lowering it would possibly remove correct hits within the desired distance. The slope of the frequency spectrum did however perform well in removing more of the false positives when optimized for short distances. In both optimizations slope improved the result for some unwanted sounds and more importantly it did not worsen the result for the drones, except on one instance on a distance it was not optimized for. Since only positive effects can be seen with adding the slope of the frequency spectrum to the detection, the slope check was implemented in the final detector.

Adding ZCR to the detection did affect some of the drone results negatively, this should have stemmed from setting the limit for ZCR too low which means that some of the recordings close to the microphone were affected. The ZCR addition did not improve the results for the non-drone recordings either, and is therefore not included in the final version of the implementation. Had it been possible to set a

lower limit other than zero on the ZCR, the non-drone recordings should have been affected. This was not done since the recordings used to set the database limits showed that it could eliminate a substantial amount of correct hits for drones.

In the final version of the implementation the authors have chosen the limits optimized for the shorter distance, since they consider 20 meters to be an acceptable distance to discover the drones on for the first time. The coefficient limits for the shorter distance also give a better result for other sounds, and let through a substantial amount of drone sound even on the longer distances. The slope limit used when optimizing for shorter distances remove more other sounds and close to no correct hits, therefore it was chosen in the final solution.

8.1 Further improvements

To improve the drone detector in the future there are a number of different additions that could be made. The authors suggest that source localization should be implemented first as it should improve the results significantly. Another improvement that could possibly give a large effect is changing the structure of the reflection coefficient intervals. At the moment the limits are the same on both sides of the middle point, but limits that are different depending on the side of the middle point could be introduced as a substitute. This could potentially exclude some non-drone sounds that are close to the middle points.

One addition that could improve the detector as well would be to add another type of identification as a complement, for example a camera that could check if the there is something to detect on a signaled hit. If this was to be added, multiple microphones for identifying the position in space would be beneficial for positioning the camera. Since the microphones used gives the best result when directed directly at the incoming sound, having a function that can direct the microphones as well towards the incoming sound could make an improvement to the correctness.

And lastly the authors feel that more extensive tests need to be performed on more drones of different sizes to evaluate its performance. During this testing, to add a measure of how certain the detector is of its decision and what it is matching against in the database could help the evaluation of how well the detector performs.

Chapter 9
Conclusion

The authors consider it possible to detect drones using audio analysis. The reflection coefficients as a base give close to no false negatives for drones in the database for optimized distances if the limits are chosen with enough care. If the database is expanded with additional drone types, the authors believe that the detector could find the drones contained in the database, with a high enough signal to noise ratio, close to a 100 percent of the time.

The slope of the frequency spectrum improves the detector's correctness and is considered suitable for the detection. The variation in values for the slope of the frequency spectrum for the drones was smaller than for the coefficients and a limit easier to choose. The ZCR on the other hand vary to much to be of use as a parameter in the detection. Had the drones had a more stable ZCR when further away from the microphone, a lower limit could have been possible. This should have improved the result for the non-drone sounds, but due to the reasons mentioned it is not possible to use.

To be able to have a detector with nearly no false positives at all, another improvement needs to be implemented. The authors would have implemented source localization to distinguish the sound coming over a selected angle from unwanted sounds. This should improve the detector's correctness substantially as most of the sounds coming from the ground could be filtered out. With some additional work a detector that does not give false positives and detect drones with a high enough signal to noise ratio, within the interval optimized for, could be created.

The result of the thesis is a detector that detect drones of a larger size within the optimized interval, and exclude similar sounding objects a portion of the time. Among the sounds it gives false positives for are objects not often found just outside secure facilities. This, together with the fact that mostly sounds similar to drones have been tested, should mean that the percentage of time it would give an incorrect result on-site is lower than what figures 7.3 and 7.4 implies. With additional work the authors believe the drone detector could become a useful product in the future.

Bibliography

[1] "Drug delivery drone crashes in mexico," *BBC*, 2015-01-22. [Online]. Available: http://www.bbc.com/news/technology-30932395

[2] A. Phillip, "Delivery drone carrying marijuana, cellphones and tobacco crashed outside a s.c. prison," *Washington Post*, 2014-07-31. [Online]. Available: http://wapo.st/1AEqrgR

[3] A. Neslen, "Three arrests fail to staunch mystery of drones flying over french nuclear plants," *The Guardian*, 2014-09-06. [Online]. Available: http://gu.com/p/434yx/sbl

[4] A. Y. Nooralahiyan, M. Dougherty, D. McKeown, and H. R. Kirby, "A field trial of acoustic signature analysis for vehicle classification," *Transportation Research Part C: Emerging Technologies*, vol. 5, no. 3, pp. 165–177, 1997.

[5] N. Bhave and P. Rao, "Vehicle engine sound analysis applied to traffic congestion estimation," in *Proc. of International Symposium on CMMR and FRSM*, 2011.

[6] "Droneshield, drone detection and response," http://www.droneshield.org/products/, accessed April 27,2015.

[7] "Drone-detector," http://www.drone-detector.com/, accessed April 27,2015.

[8] "Drone detector," http://www.dronedetector.com/compare-drone-detector/, accessed May 4,2015.

[9] N. Dave, "Feature extraction methods lpc, plp and mfcc in speech recognition," *International Journal for Advance Research in Engineering and Technology*, vol. 1, 2013.

[10] R. Lawrence and J. Biing-Hwang, *Fundamentals of speech recognition*. PTR Prentice Hall Englewood Cliffs, 1993.

[11] P. E. Papamichalis, *Practical approaches to speech coding*. Prentice-Hall, Inc., 1987.

[12] L. R. Rabiner and R. W. Schafer, *Digital processing of speech signals*. Prentice-hall Englewood Cliffs, 1978.

[13] "Adsp-21262," http://www.analog.com/en/products/processors-dsp/sharc/adsp-21262.html, accessed May 20,2015.

[14] "Audio codec specification, tlv320aic32," http://www.ti.com/product/tlv320aic32, accessed May 20,2015.

[15] "Microphone specification," https://pro.sony.com/bbsc/ssr/product-ECMVG1/, accessed May 12,2015.

Appendix A
Recordings

All sound files were recorded with a TASCAM DR-680 audio recorder and SONY ECM-VG1 condenser microphones.

A.1 Recordings for Database Coefficients

The recording took place in an anechoic chamber and all equipment on the floor were padded with soft material underneath to affect the recordings as little as possible. Cyclone and X4 were mounted one at a time to a metal pole between two tripods, as can be seen in figure A.1 and figure A.2. They were mounted to be able to operate them from outside the chamber without visual contact and to have constant distance to the microphones. Three microphones were used, all three directed towards the drone. Both the authors and the recorder were outside of the chamber during the recordings. The recorder was connected to the microphones by cables through a hole in the wall. The drones were operated with their regular transmitters.

Seven different throttle levels were recorded for each drone, 30%, 40%, 50%, 60%, 70%, 80% and 90% as indicated on the transmitters. The recordings were made with a sampling frequency of 48 kHz. They were made in the anechoic chamber in order to avoid noise and reverberations from the surroundings to be able to gain reliable insights about the characteristics of the drone sounds.

These recordings were used to create the allowed values of the reflection coefficients in the database.

A.2 Recordings for Database Limits

During the recording of files to use to set the limits the weather varied some: small amounts of rain and wind in different directions, and there were some traffic noise in the background. On account of the wind the drones were held by one of the authors in their surrounding shell to stabilize them during the recordings, this should not affect the results.

One microphone, directed straight ahead onto the drones, was used. It was attached to a large tripod approximately 1,5 meters above the ground. A total of 9 and 7 different recordings were made for Cyclone and X4, respectively. The recordings contained:

Figure A.1: Cyclone in the anechoic chamber.

Figure A.2: X4 in the anechoic chamber.

- Drone at a distance of 2 meters from the microphone.
- Drone at a distance of 5 meters from the microphone.
- Drone at a distance of 10 meters from the microphone.
- Drone at a distance of 15 meters from the microphone.
- Drone at a distance of 20 meters from the microphone.
- Drone at a distance of 25 meters from the microphone.
- Drone at a distance of 30 meters from the microphone.
- Drone at a distance of 35 meters from the microphone (only for Cyclone).
- Drone at a distance of 40 meters from the microphone (only for Cyclone).

These recordings were used to calculate the limits for the reflection coefficients, the slope of the frequency spectrum and the zero crossing rate.

A.3 Recordings for Final Testing

Recordings of drone sounds, other sounds that have similarities with drones and some sounds that could be present in a future environment for a drone detector were made. One microphone, directed straight ahead onto the object, was used. It was attached to a large tripod approximately 1,5 meters above the ground. The recordings contained:

- The drones flying from 50 meters away to the microphone and flying from the microphone to 50 meters away.
- A hairdryer on low and high speed directed towards and away from the microphone at a distance of 2 meters.
- Two different electric lawn mowers, one with blades underneath and one with a cylinder for blades, moving towards and away from the microphone.
- An electric trimmer motor without a blade 2 meters from the microphone.
- A vacuum 3 meters from the microphone.
- Birds chirping.
- Sound recorded beside a freeway.
- Sound recorded outside without anything specific happening nearby.

Appendix B
Database Values

The parameter limits used for the optimizations can be seen in table B.1.

	C1	**C2**	**C3**	**C4**	**C5**	**Slope**	**ZCR**
Optimized for long distances	0.20	0.35	0.35	0.30	0.30	-0.0033	0.25
Optimized for short distances	0.20	0.25	0.35	0.20	0.15	-0.0027	0.25

Table B.1: The limits used for the two optimizations.

Appendix C
Hardware

C.1 Analog Devices ADSP-21262 DSP

The specification of the DSP used for the implementation can be seen in table C.1. [13]

C.2 TI TLV320AIC32 Audio codec

An audio codec from Texas Instruments was used, its specification can be seen in table C.2. [14]

C.3 SONY ECM-VG1 Condenser Microphone

The characteristics of the microphones used for all recordings can be seen in table C.3 and figure C.1. [15]

Feature	**Analog Devices ADSP-21262 DSP**
RAM	2Mbit
ROM	4Mbit
Clock speed	200MHz
FLOPs	up to 1.2G FLOPs

Table C.1: Specification of the DSP.

Feature	TI TLV320AIC32 Audio codec
Sample rate	From 8kHz to 96kHz
Chosen sample rate	48kHz
Block size	960

Table C.2: Specification of the audio codec.

Feature	SONY ECM-VG1 Condenser Microphone
Frequency response	40 Hz to 20 kHz
Sensitivity (at 1 kHz)	-33dB ± 3dB

Table C.3: Specification of the microphones.

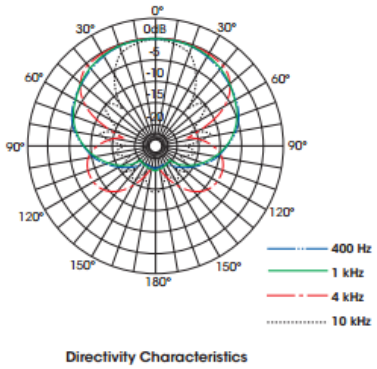


Figure C.1: Directivity characteristics of the SONY ECM-VG1 condenser microphone

LUND
UNIVERSITY

Series of Master's theses
Department of Electrical and Information Technology
LU/LTH-EIT 2015-448

http://www.eit.lth.se