2013

Analysis of Human Echolocation Waveform for Radar Target Recognition

Kandarp Patel
Wright State University

Follow this and additional works at: https://corescholar.libraries.wright.edu/etd_all

Part of the Electrical and Computer Engineering Commons

Repository Citation
https://corescholar.libraries.wright.edu/etd_all/716

This Thesis is brought to you for free and open access by the Theses and Dissertations at CORE Scholar. It has been accepted for inclusion in Browse all Theses and Dissertations by an authorized administrator of CORE Scholar. For more information, please contact corescholar@www.libraries.wright.edu, library-corescholar@wright.edu.
ANALYSIS OF HUMAN ECHOLOCATION WAVEFORM FOR RADAR TARGET RECOGNITION

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Engineering

By

KANDARP PATEL
B.Eng., University of Ontario Institute of Technology, Canada, 2011

2013
Wright State University

___________________________________________
Arnab K. Shaw, Ph.D.
Thesis Director

___________________________________________
Kefu Xue, Ph.D.
Chair, Department of Electrical Engineering

Committee on Final Examination

___________________________________________
Arnab K. Shaw, Ph.D.

___________________________________________
Chris J. Baker, Ph.D.

___________________________________________
Fred D. Garber, Ph.D.

___________________________________________
R. William Ayres, Ph.D.
Interim Dean, Graduate School
ABSTRACT


Some blind humans have developed the remarkable capability of echolocation, similar to the type used by mammals such as the bat, dolphin and whale. This population of human has shown the ability to classify targets based on their location, size, shape and material in diverse environmental conditions simply by listening to the reflected echoes of tongue clicks generated by their mouth. To date, much of the research into human echolocation has been confined exclusively to behavioral science and the analysis is inconsistent with the approaches used in engineering. The waveforms used in current radar systems appear different to those typical of mammal echolocation. It is speculated that the lack of robust success in radar target recognition may therefore be attributed to application of an inappropriate waveform. This research focuses on the analysis of human echolocation waveforms and their reflected echoes from different objects to investigate what properties of the waveform may carry target information.

Results based on the analyses of echo data collected for various targets and their extracted features suggests that normalized target signatures cannot provide target classification in efficient manner. The normalized frequency spectrum has some potential for target classification, but it does not lead to confident classification results. The absolute difference between normalized frequency spectrum of transmit signal and normalized frequency spectrum of echoes performs much better than the two features discussed previously. It should be noted that the tongue click waveform performs much better at classifying objects made of hard materials from objects made
of soft materials. However, they cannot be classified based on their shape or size by utilizing this feature. The chirp waveform provides superior classification performance for this feature, however, it is unclear which broad categories the targets can be put in for classification. The chirp, certainly, cannot classify all the targets as distinct from each other using this feature. From this analysis, it can be concluded that the frequency spectrum of echoes contain the most useful information for the task of target classification based on its material.

Moreover, a moment invariant (MI) based automatic target recognition system has been developed that can potentially be implemented using efficient table-look up hardware. The phenomenon that echolocation experts utilize multiple perspectives of the objects to classify a target has inspired the development of this ATR algorithm. The main advantages of MIs are the translation and scale invariant properties and the possibility of implementation using table-look up hardware. The results of applying the theory of moment invariants on High Range Resolution profile data for the purpose of target classification are presented in work. For the developed ATR system, the performance is as high as 98% when the azimuth angle search range is 2 degrees while it reduces to approximately 58% for azimuth angle search range of 172 degrees.
# TABLE OF CONTENTS

1 INTRODUCTION ..................................................................................................................... 1

1.1 Objective ............................................................................................................................. 1

1.2 Literature Review ................................................................................................................. 3

1.2.1 Human Echolocation ......................................................................................................... 3

1.2.2 Automatic Target Recognition (ATR) in Radar ................................................................. 6

1.3 Thesis layout .......................................................................................................................... 8

2 FUNDAMENTALS OF RADAR SYSTEMS ........................................................................... 9

2.1 Range resolution ................................................................................................................... 9

2.2 Maximum unambiguous range ............................................................................................. 9

2.3 Radar waveforms .................................................................................................................. 10

2.4 Pulse compression ............................................................................................................... 10

2.5 Pulse integration .................................................................................................................. 12

2.6 Wideband ambiguity function ............................................................................................. 13

3 HUMAN TONGUE CLICK ANALYSIS ............................................................................. 17

3.1 Analysis of human tongue click .......................................................................................... 17

3.2 Producing a synthetic human tongue click ........................................................................ 21

3.2.1 Deconstruction of human tongue click ............................................................................ 21
3.2.2 Construction of synthetic tongue click signal ........................................... 22

3.3 Comparison of biological tongue click with synthetic tongue click signal............. 27

4 ACOUSTIC RADAR SYSTEM................................................................................. 31

4.1 Acoustic radar system ....................................................................................... 31

4.2 Laboratory setup ................................................................................................. 33

4.3 Acoustic radar calibrations .................................................................................. 34

4.3.1 Background noise ........................................................................................... 34

4.3.2 Pulse integration .............................................................................................. 35

4.3.3 Clutter mitigation ............................................................................................. 36

4.3.4 Baffles to eliminate strong returns ................................................................. 38

4.3.5 Evaluation of multipath environment ............................................................. 38

5 CLASSIFICATION EXPERIMENTS ....................................................................... 41

5.1 Experimental setup ............................................................................................. 41

5.2 Types of targets used .......................................................................................... 42

5.3 Classification ........................................................................................................ 43

5.3.1 Radar range profile ......................................................................................... 43

5.3.2 Target signatures .............................................................................................. 44

5.3.3 Absolute value of target echoes ....................................................................... 48
7.2 Conclusion ......................................................................................................................... 79

7.3 Future work .................................................................................................................... 80

APPENDIX A .......................................................................................................................... 82

EXPERIMENTAL AND LABORATORY SETUP ................................................................... 82

BIBLIOGRAPHY ..................................................................................................................... 87
# TABLE OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Spectrogram of LFM Chirp</td>
<td>11</td>
</tr>
<tr>
<td>2</td>
<td>Output of pulse compression (matched) filter</td>
<td>12</td>
</tr>
<tr>
<td>3</td>
<td>Real part of human tongue click</td>
<td>18</td>
</tr>
<tr>
<td>4</td>
<td>Spectrogram of Tx tongue click</td>
<td>19</td>
</tr>
<tr>
<td>5</td>
<td>Spectrogram of Rx tongue click</td>
<td>19</td>
</tr>
<tr>
<td>6</td>
<td>WAF of tongue click</td>
<td>19</td>
</tr>
<tr>
<td>7</td>
<td>Zero-velocity and zero-range cuts of WAF</td>
<td>19</td>
</tr>
<tr>
<td>8</td>
<td>WCAF of Tx and Rx tongue click</td>
<td>19</td>
</tr>
<tr>
<td>9</td>
<td>Zero-velocity and zero-range cuts of WCAF</td>
<td>19</td>
</tr>
<tr>
<td>10</td>
<td>Component at 1.54 kHz</td>
<td>22</td>
</tr>
<tr>
<td>11</td>
<td>Component at 3.33 kHz</td>
<td>22</td>
</tr>
<tr>
<td>12</td>
<td>Component at 4.62 kHz</td>
<td>22</td>
</tr>
<tr>
<td>13</td>
<td>Component at 11.03 kHz</td>
<td>22</td>
</tr>
<tr>
<td>14</td>
<td>Synthetic component at 1.54 kHz</td>
<td>25</td>
</tr>
<tr>
<td>15</td>
<td>Synthetic component at 3.33 kHz</td>
<td>25</td>
</tr>
<tr>
<td>16</td>
<td>Synthetic component at 4.62 kHz</td>
<td>25</td>
</tr>
<tr>
<td>17</td>
<td>Synthetic component at 11.03 kHz</td>
<td>25</td>
</tr>
<tr>
<td>18</td>
<td>Frequency response for component at 1.54 kHz</td>
<td>26</td>
</tr>
<tr>
<td>19</td>
<td>Frequency response for component at 3.33 kHz</td>
<td>26</td>
</tr>
</tbody>
</table>
Figure 20: Frequency response for component at 4.62 kHz ........................................ 26
Figure 21: Frequency response for component at 11.03 kHz ...................................... 26
Figure 22: Tongue click comparison in time domain ...................................................... 29
Figure 23: Tongue click comparison in frequency domain ............................................ 29
Figure 24: Zero-velocity and zero-range cuts of WAF for human tongue click .............. 29
Figure 25: Zero-velocity and zero-range cuts of WAF for synthetic tongue click .......... 29
Figure 26: Frequency spectrum of whole synthetic tongue click .................................. 29
Figure 27: Zero-velocity and zero-delay cuts of WAF for whole synthetic tongue click .... 29
Figure 28: Acoustic radar system .............................................................................. 32
Figure 29: Laboratory Setup ...................................................................................... 34
Figure 30: Pulse integration gain .............................................................................. 36
Figure 31: Background subtraction ............................................................................ 37
Figure 32: Multipath Effect ...................................................................................... 39
Figure 33: Range profile for metal plate .................................................................... 40
Figure 34: Experimental Setup .................................................................................. 42
Figure 35: Radar range profile for metal plate using LFM chirp as Tx signal ............ 43
Figure 36: Radar range profile for metal plate using synthetic tongue click as Tx signal 43
Figure 37: Target signature of metal dihedral (chirp) .................................................. 46
Figure 38: Target signature of metal dihedral (tongue click) ........................................ 46
Figure 39: Target signature of metal plate (chirp) ....................................................... 46
Figure 40: Target signature of metal plate (tongue click) ............................................ 46
Figure 41: Target signature of metal sphere (chirp) ......................................................... 46
Figure 42: Target signature of metal sphere (tongue click) .................................................. 46
Figure 43: Target signature of human (chirp) ....................................................................... 47
Figure 44: Target signature of human (tongue click) ............................................................. 47
Figure 45: Target signature of plastic box (chirp) ................................................................. 47
Figure 46: Target signature of plastic box (tongue click) ....................................................... 47
Figure 47: Target signature of chair front (chirp) ................................................................. 47
Figure 48: Target signature of chair front (tongue click) ....................................................... 47
Figure 49: Target signature of chair back (chirp) ................................................................. 48
Figure 50: Target signature of chair back (tongue click) ....................................................... 48
Figure 51: Absolute value of target echo for dihedral (chirp) ............................................... 50
Figure 52: Absolute value of target echo for dihedral (tongue click) ..................................... 50
Figure 53: Absolute value of target echo for plate (chirp) .................................................... 50
Figure 54: Absolute value of target echo for plate (tongue click) ........................................ 50
Figure 55: Absolute value of target echo for sphere (chirp) ................................................. 50
Figure 56: Absolute value of target echo for sphere (tongue click) ...................................... 50
Figure 57: Absolute value of target echo for human (chirp) .................................................. 51
Figure 58: Absolute value of target echo for human (tongue click) ...................................... 51
Figure 59: Absolute value of target echo for plastic box (chirp) .......................................... 51
Figure 60: Absolute value of target echo for plastic box (tongue click) ............................... 51
Figure 61: Absolute value of target echo for chair front (chirp) .......................................... 51
Figure 62: Absolute value of target echo for chair front (tongue click) ............................................. 51
Figure 63: Absolute value of target echo for chair back (chirp) ................................................................. 52
Figure 64: Absolute value of target echo for chair back (tongue click) ......................................................... 52
Figure 65: Frequency spectrum - dihedral (Chirp) .................................................................................. 55
Figure 66: Frequency spectrum - dihedral (tongue click) ........................................................................ 55
Figure 67: Frequency spectrum - plate (chirp) ....................................................................................... 55
Figure 68: Frequency spectrum - plate (tongue click) ............................................................................ 55
Figure 69: Frequency spectrum - sphere (chirp) ................................................................................. 55
Figure 70: Frequency spectrum - sphere (tongue click) ......................................................................... 55
Figure 71: Frequency spectrum - human (chirp) .................................................................................. 56
Figure 72: Frequency spectrum - human (tongue click) ......................................................................... 56
Figure 73: Frequency spectrum - plastic box (chirp) ........................................................................ 56
Figure 74: Frequency spectrum - plastic box (tongue click) ................................................................. 56
Figure 75: Frequency spectrum - chair front (chirp) ........................................................................... 56
Figure 76: Frequency spectrum - chair front (tongue click) ................................................................. 56
Figure 77: Frequency spectrum - chair back (chirp) ............................................................................ 57
Figure 78: Frequency spectrum - chair back (tongue click) ................................................................. 57
Figure 79: Example HRR Trajectory Plot ......................................................................................... 69
Figure 80: Moment Invariant MI(4,1) ............................................................................................... 71
Figure 81: Moment Invariant MI(4,2) ............................................................................................... 71
Figure 82: Moment Invariant MI(6,1) ............................................................................................... 71
Figure 83: Moment Invariant MI(6,2)........................................................................................................71

Figure 84: Sum of MI(p,q) for p = 8, 10, 12, 14, 16, 18, 20 and q = 1, 2 exhibit target separability
........................................................................................................................................................72

Figure 85: Probability of correct classification versus number of test profiles for moment
invariant based ATR .....................................................................................................................................73

Figure 86: Probability of correct classification versus azimuth angle search range for moment
invariant based ATR .....................................................................................................................................73

Figure 87: Probability of correct classification versus number of test profiles for template based
ATR .............................................................................................................................................................76

Figure 88: Probability of correct classification versus azimuth angle search range for template
based ATR ......................................................................................................................................................76

Figure 89: Speaker (transmitter) and microphone (receiver) acoustic radar system ..................82

Figure 90: Metal sphere on a baffle stand ...............................................................................................83

Figure 91: Metal sphere facing acoustic radar system ...........................................................................83

Figure 92: Plastic box on a baffle stand ....................................................................................................84

Figure 93: Baffles on the floor between acoustic radar and target .........................................................84

Figure 94: Baffles around pillars .............................................................................................................85

Figure 95: Baffles around pillars .............................................................................................................86
LIST OF TABLES

Table 1: Parameters used to estimate envelop of component at 1.54 kHz........................................23
Table 2: Parameters used to estimate envelop of component at 3.33 kHz.................................23
Table 3: Parameters used to estimate envelop of component at 4.62 kHz.................................23
Table 4: Parameters used to estimate envelop of component at 11.03 kHz.................................24
Table 5: RMSE between actual envelop and estimated envelop ....................................................27
Table 6: Number of pulses integrated to equalized SNR...............................................................53
Table 7: Absolute Correlation coefficient between normalized target signatures (chirp)...........58
Table 8: Absolute Correlation coefficient between normalized target signatures (tongue click)..59
Table 9: Absolute correlation coefficient between normalized frequency spectrums of echoes (chirp)..................................................................................................................60
Table 10: Absolute correlation coefficient between normalized frequency spectrums of echoes (tongue click)...........................................................................................................61
Table 11: Absolute correlation coefficient between absolute difference between normalized frequency spectrum of transmit signal and normalized frequency spectrum of echo (chirp)......62
Table 12: Absolute correlation coefficient between absolute difference between normalized frequency spectrum of transmit signals and normalized frequency spectrum of echo (tongue click)..........................................................................................................................63
Table 13: MSTAR targets used for training and testing ..................................................................70
Table 14: Confusion matrix for classification using average MIs from 8 adjacent test profiles and + 5 degree of azimuth angle search. .............................. 73
ACKNOWLEDGEMENTS

It is my pleasure to acknowledge and thank people who helped me accomplish my goal to pursue graduate studies. First I would like to thank my parents for their constant emotional & financial support and encouragement. I will always be grateful to them and without them this would have not been possible.

I would like to thank Professor Arnab K. Shaw, Wright State University, and Professor Chris J. Baker, The Ohio State University, for their guidance throughout my thesis. I would, especially, like to thank Dr. Graeme E. Smith, The Ohio State University, for his daily guidance and encouragement throughout my thesis. I would also like to thank Professor Fred D. Garber for agreeing to be on my thesis committee.

Lastly, I would wish to thank all the faculty members of the Electrical Engineering Department at Wright State University for their generous help and tremendous support through the course of my M.S.E. program.
1 INTRODUCTION

1.1 Objective

Radar is an electromagnetic system for the detection and location of reflecting objects such as aircraft, ships, spacecraft, vehicles, people, and the natural environment. It can operate in darkness, haze, fog, rain, and snow. Its ability to measure distance with high accuracy and in all weather is one of its most important attributes. However, early radar only detected the presence of a target and gave its location in range and angle, but could not provide much else about the type of target being detected. The desires for target classification lead to transponder systems on aircrafts and ships that allowed them to identify themselves on a regular basis or when asked for its identification by an interrogator [1].

Over time, radar resolution in range and cross-range improved and it became possible to resolve the individual scattering centers of a target and make inference about its nature. But even today, automatic target recognition of objects is a difficult and important problem in ATR (Automatic Target Recognition) field. However, many animals (i.e. bats, dolphins, etc.) use echolocation as their primary method of navigation. They can both detect and classify the world based on their processing of the backscattered signals. In fact, some species of bat obtain their entire perception of the world from echolocation [1] [2].

Research has shown that humans are also able to sense the environment around them using reflected sound waves. Basset et al. [3] found that when sound containing many frequencies is reflected from a flat surface, an observer in the field of both the incident and reflected sounds
hears a broad tone with an associated pitch that varies inversely with distance from the surface. Rice et al. [4] show that various participants were successful in identifying the presence or absence of metal flat plate using their choice of signal in a series of trials. Rice [5] revealed that six participants, blind from birth, were significantly superior to either adventitiously blind and either sighted participants in their ability to detect small targets using self-generated sonar signals.

Rice [6] observed in his experiment that when the blind participants were allowed to generate their preferred sound for echolocation, they made either a “hissing” sound or a “tongue click” sound. In both cases, Rice noticed that the performance of echolocation was similar regardless of the choice of sound. However, it is fascinating to know that both human tongue click sound and clicking sound of certain species of bat have similarities regardless of the fact that their echolocation abilities were evolved independently and in different time frames. Both sounds have three distinct frequency components located at their resonant frequencies [2] [7]. However, the frequency components are hyperbolically modulated in frequency in bat click sound, whereas the components in human tongue click are not modulated in frequency. These modulations are in contrast to manmade waveforms commonly used in radar such as chirp [1]. Therefore, it might be assumed that this particular structure of clicking sound (i.e. consisting of three different frequency components) makes it better suited for target classification activities than the conventional man-made radar waveforms (i.e. chirp).

A recent study by Thaler et al. [8] revealed remarkable abilities of blind human echolocation experts who locate different objects by listening to the returning echoes of their mouth clicks. These echolocation experts perform activities such as mountain biking, exploring cities while traveling, playing basketball, or hiking using their echolocation skills. Based on their study of brain activity of echolocation experts, they suggested that echoes of these clicks are being processed in region of brain which is responsible for processing visual information, and not the audio information.
Current radar systems do not attempt to exploit any features of typical mammal echolocation. To date, much of the research into natural human echolocation has been confined exclusively to behavioral science. As a result, the analysis of the transmitted signals and their echoes are inconsistent with the methods used in radar and automatic target recognition. However, there appears to be an enormous potential for the radar community to improve the task of automatic target recognition. Better understanding and proper incorporation of some diversity schemes into advanced radar systems may lead to autonomous navigation and automatic target classification that have remained elusive thus far. In order to exploit echolocation concepts in radar, it is important to understand how humans make decisions to classify different objects.

Hence, this research focuses on the analysis of human echolocation waveform and their reflected echoes from different objects to investigate what properties of the waveform may carry target information. Moreover, a moment invariant based automatic target recognition system has been developed after inspiring from the echolocation expert’s ability to classify objects utilizing multiple perspectives of the object.

1.2 Literature Review

In this section, the covered literature on human echolocation and automatic target recognition (ATR) in radar provides the discussion on most relevant topics. The literature review on human echolocation discusses the research performed on human echolocation from the point of view of behavioral sciences. The literature review on ATR in radar discusses only the most relevant research on ATR.

1.2.1 Human Echolocation

Stroffregen and Pitternger [9] argued that echolocation may be a basic perception-action ability of humans. Their study suggested that, regardless of whether a human is blind or sighted, he/she is able to perceive various properties of objects, such as shape, distance, size, relative motion and substance, from a distance. Their study suggested that pulse-to-echo delay contains information for distance to object by
observation of how long it takes the echo to arrive to ears. They suggested that perception of the intensity of received echo will change depending on the constructive/destructive interference of the pulse and echo at the ear. As the observer moves, the frequency of change in intensity of echo will change and it is directly related to the frequency of pulse. Therefore, this difference has been defined as pulse-to-echo frequency. They also suggested that the three dimensional shape of an object might be detected by knowing how far the object is located. For an observer observing a three dimensional objects, the parts of objects which are nearer and further from the observer will provide different pulse-to-echo frequency and pulse-to-echo delay information. A particular frequency of the tone provides the information about distance to the stationary observer, whereas the rate of change of frequency of this tone provides the information about distance to the moving observer. Since different elements reflect and absorb different bands of frequencies, the frequency spectra of pulse and echo are also useful in specifying elements.

Rosenblum et al. [10] performed experiments to test whether a moving listener has any advantage while echolocating compared to the stationary echolocator. In their experiment, the ecolocators were asked to locate an object while either remaining stationary or moving. Then, after removing the object, ecolocators were asked to walk to the location of the object. Based on their results, they suggested that the moving ecolocators may have been able to utilize the information about arrival time of echoes, and hence, were able to perform better than stationary ecolocators.

Schwitzgebel and Gordon [11] argued that sighted human beings do echolocate as part of their normal intercourse with the world. Although they do echolocate, most people’s knowledge of that experience is limited. Therefore, it can be argued that it most likely depends on how conscious one is about his/her experience of echolocation. Even though echolocation is not as vivid as visual experience, it is important, pervasive, and distinctive feature of our sensory phenomenology.

An experimental study was conducted by Schenkman and Nilsson [12] to analyze blind and sighted persons’ ability to detect sounds recorded in the presence of a reflecting object. In their study, the sounds
of different durations were produced by a loud speaker and recorded using a microphone in the presence and absence of a flat, circular aluminum plate. The plate was placed at the ranges between 0.5 m and 5 m. When these recorded sounds were presented to 10 sighted and 10 visually handicapped participants, they were able to identify the presence or absence of the plate for recordings which included plate at less than 2 meter distance. They also found that the participants performed better for sounds which were of longer duration than those of shorter duration. The same experiments performed on two blind participants revealed that they were able to perform better than all of the previous 20 participants, even in the cases which included plate at greater than 2 meter distance. However, just like previous 20 participants, the blind participants also performed better for recordings with longer duration sounds compared to recordings with shorter duration sounds.

Thaler et al. [8] employed two echolocation experts to study their functional brain activity. In their study, the echolocation recordings were played in the ears of echolocation experts while their functional brain activity was measured. Thaler et al. [8] mentioned that both of the participants EB (Early Blind) and LB (Late Blind) had lost their vision at the age of 13 months and 14 years, respectively, and were 43 and 27 years old, respectively, at the time of experiments. Both were able to locate the auditory source and had no hearing defects. They were also able to perform daily tasks using their echolocation ability. During their studies, they found activities in calcarine cortex of both individuals while the sounds containing echoes and clicks were played, as opposed to the sounds without echoes. In human brain, calcarine cortex is responsible for processing visual information. Surprisingly, the researcher did not find any change in functional activity of auditory cortex – the part of human brain responsible for processing of auditory information. Moreover, the early blind participant’s brain showed more activity in calcarine cortex, than in case of late blind participant’s brain, when the sounds containing echoes from surrounding space were played. Based on these findings, Thaler et al. [8] suggested that, for the EB and LB participants, the region of brain devoted to processing of visual information processes the echoes of these clicks instead of the region of the brain devoted to processing of audio information.
Rojas et al. [13] suggests that among various ways to produce clicks, clicks produced by moving the tongue downwards and backwards from the palatal region of mouth is most suitable for human echolocation. In the experiments by Thaler et al. [8], the EB and LB participants were producing tongue clicks in the same manner and showed remarkably accurate detection capabilities. Therefore, it can be argued that this particular signal may have some distinct properties which provide EB and LB participants with remarkably accurate results.

Teng and Whitney [14] suggested that skilled echolocators could retrieve various properties of the object, such as texture, size, shape, distance and position, from the returning echoes. The EB participant, an echolocation expert, employed by Thaler et al. [8] is successfully able to provide all of the above mentioned information for an object using echolocation. Therefore, it can be argued that the tongue click used by EB participant must have distinct properties which allow him to provide information regarding the position, distance, size, shape, and texture of objects.

1.2.2 Automatic Target Recognition (ATR) in Radar

Automatic target recognition of objects regardless of their color, size, orientation and position in environment is a difficult and important problem in ATR fields. A straightforward solution to this problem could involve the use of a library which contains target signatures from all the possible viewing angles, positions in the environment, spectral bands and contrast conditions leading to extremely large set of combinations to consider for a target [15]. Hence, there is a need to develop ATR algorithms that are computationally efficient, fast and possibly invariant to rotation, translation or scaling. However, blind echolocation experts are very efficient at classification of different targets regardless of their position in the field of view while navigating in their surrounding environment. Therefore, this thesis utilized the theory of moment invariants (MI) on High Range Resolution (HRR) profile data for the purpose of target classification.
A research by Bhanu and Jones [16] discusses a method to optimize the recognition of vehicles in SAR imagery by using multiple SAR recognizers at different look angles using MSTAR public data. The recognition system uses SAR images of actual military vehicles from MSTAR public data to identify a specific type of vehicle. The fundamental azimuth variance of SAR scatter location was successfully used as a feature to design an effective ATR algorithm. The experiment in this research demonstrates that the system can recognize the vehicle independent of the azimuth look angle.

Bhanu and Jones [17] used the location of peaks and nulls corresponding to the azimuth angles of the objects as invariant features and developed SAR ATR which could classify articulated objects using templates created by non-articulated objects. The system could also classify efficiently in the cases of high noise and occlusion. The system performance was better for less occluded objects compared to the performance for more occluded objects. Their limited experiments show that scaling to model more objects provides similar results, although performance will degrade depending on the number of coincidental similarities found in the radar signatures of the objects.

Mitchell and Westerkamp [18] presented a statistical feature based (StaF) classifier using HRR profiles of aircrafts. The StaF classification algorithm was designed to successfully utilize features such as location and amplitudes of peaks in HRR profiles, peak amplitude probability function and the peak location probability function.

Shaw and Bhatnagar [19] presented an ATR algorithm which formed templates via Singular Value Decomposition (SVD) of HRR profiles. They demonstrated that SVD provides an optimal feature set in the mean-squared sense. The algorithm was implemented using MSTAR and XPATCH data and classification was performed using matched filter and linear least-square based classifiers.

Sadjadi [15] considers a different approach to ATR by using the theory of invariant algebra which used the seven moments proposed by Hu [20]. In his approach, he utilized the properties of algebraic invariants to identify the features which remain unchanged regardless of the linear geometrical and spectral
transformation of objects. While navigating in the surrounding environment, the echolocation experts experience the similar transformation by observing the target from different aspect angles. Hu moments are invariant to rotation also, making them highly appropriate for two-dimensional images. However, HRR radar profiles being used in this work are obtained by mapping three-dimensional target information into one-dimensional signals that represent reflected radar intensity along target extend in range at a particular angle. Therefore, the property of rotation invariant is not applicable to HRR data, as well as echolocation data. Nevertheless, in case of HRR profiles as well as for echo returns, translation and scale invariance are critical.

1.3 Thesis layout

Chapter 2 presents the fundamental theory of radar system which the reader needs to know in order to understand the work presented in the thesis. Chapter 3 presents the analysis of human tongue click waveform and discusses how a synthetic version of human tongue click waveform has been created. It also provides the comparison between two waveforms. Chapter 4 discusses the acoustic radar system used for this research and the calibrations performed to ensure reliability of collected data. Chapter 5 discusses the classification experiments performed using different targets. It also discusses how the features extracted from these experimental data can be utilized for target classification. Chapter 6 presents the moment invariance based automatic target recognition system, which is developed by utilizing the theory of invariant algebra. Finally, Chapter 7 discusses the accomplishments, conclusions and future work that can be done based on this research.
2 FUNDAMENTALS OF RADAR SYSTEMS

This chapter discusses a number of radar processing concepts that are relevant to the research reported here. For a detailed description of radar, the reader is referred to, [1] [21] [22] [23].

2.1 Range resolution

The most common radar signal, or waveform, is a series of short-duration, somewhat rectangular-shaped pulses modulating a sinewave carrier. This is sometimes called a pulse train. The range to a target is determined by the time $T_R$ it takes the radar signal to travel to the target and back. Electromagnetic energy in free space travels with the speed of light, which is $c = 3 \times 10^8 \text{ m/s}$. Thus the time for the signal to travel to a target located at a range $R$ and return back to the radar is $\frac{2R}{c}$. The range to a target is then

$$R = \frac{cT_R}{2}. \quad (1)$$

2.2 Maximum unambiguous range

Once the signal is radiated into space by a radar, sufficient time must elapse to allow all echo signals to return to the radar before the next pulse is transmitted. The rate at which pulses may be transmitted, therefore, is determined by the longest range at which targets are expected. If the time between pulses $T_p$ is too short, an echo signal from a long-range target might arrive after the transmission of the next pulse and be mistakenly associated with that later pulse rather than the actual pulse transmitted earlier. This can result in an incorrect or ambiguous measurement of the range. Echoes that arrive after the transmission of the next pulse are called second-time-around echoes (or multiple-time-around echoes if from even earlier pulses). Such an echo would appear
to be at a closer range than actual and its range measurement could be misleading if it is not known to be a second-time-around echo. The range beyond which targets appear as second-time-around echoes is the maximum unambiguous range $R_{un}$, and is given by

$$R_{un} = \frac{cT_p}{2} = \frac{c}{2f_p} \quad (2)$$

where $T_p = pulse\ repetition\ period(interval)(PRI) = 1/f_p$, and $f_p = pulse\ repetition\ frequency\ (PRF)$, usually given in hertz or pulse per second (pps).

2.3 Radar waveforms

The typical radar utilizes a pulse waveform. This pulse is composed of a sine wave interrupted by turning a switch on and off. The pulse waveform has a duration of $\tau$ and it extends in space over a distance $ct$. Two equal targets can be recognized as being resolved in range when they are separated by a distance half this value, or $\frac{ct}{2}$. The factor of one-half results from the two-way travel of the radar wave.

A very long pulse is needed for some long-range radars to achieve sufficient energy to detect small targets at long range. A long pulse, however, has poor resolution in the range dimension. Frequency or phase modulation can be used to increase the spectral width of a long pulse to obtain the resolution of a short pulse. This is called pulse compression, and is described in the next section.

2.4 Pulse compression

The pulse compression technique to be explained here is called LFM chirp, with LFM for linear frequency modulation and chirp for the fact that the ramp waveform that changes in frequency is used; if it were in the audible range, it would make a chirping sound. Figure 1 represents a spectrogram of a LFM chirp which can be used as a transmit signal. Spectrogram is a graph of energy content of a signal expressed as function of frequency and time, where the vertical axis is frequency, the horizontal axis is time, and amplitude is shown on a color-scale. In Figure 1, the frequency increases linearly from $f_1 = 1.54 \ kHz$ to $f_2 = 4.62 \ kHz$ over the pulse duration $T = 3.5 \ ms$. 
The frequency increases linearly from $f_1$ to $f_2$ over the duration of the pulse. This is sometimes known as an *up-chirp*. On reception, the frequency-modulated signal is passed through the pulse-compression filter, which is a delay line whose velocity of propagation is proportional to frequency. It speeds up higher frequencies at the trailing edge of the pulse relative to the lower frequencies at the leading edge so as to compress the signal to a width $1/B$, where signal bandwidth $B = f_2 - f_1$. The pulse compression filter is a matched filter; hence, its output envelope (neglecting noise) is the autocorrelation function of the input. Figure 2 presents the output of pulse compression (or matched) filter for an LFM chirp signal. In this case, the output is proportional to $(\sin \pi B t)/\pi B t$. The peak power of the pulse is increased by the pulse compression ratio $BT \approx T/\tau$ after passage through the filter. The matched filtering process will result in the maximum attainable signal-to-noise ratio (SNR) at the output of the filter when the signal to which it was matched, plus white noise, are passed through it. Another technique to improve SNR is called pulse integration and it is discussed in the next section.

**Figure 1: Spectrogram of LFM Chirp**
2.5 Pulse integration

Integration is the process of combining multiple samples of a signal, each contaminated by noise or other interference, to “average down the noise” and obtain a single combined signal-plus-noise sample that has a higher SNR than the individual samples. The integration can be either coherent, meaning that the signal phase information is used, or non-coherent, meaning that only the magnitude of the signal is processed. In this work, the method of coherent integration is used since it retains the phase information.

Assume a measured signal $x$ consists of a signal component, $Ae^{j\phi}$, and white, Gaussian noise component, $w$. The sample $x$ is a single range sample from a single pulse representing echo return. If the measurement that gave $x$ is repeated $N$ times, a sequence of measurements $x[n] = Ae^{j\phi}$ can be formed. The noise $w[n]$ in each sample is assumed independent and identically distributed (i.i.d.) with variance $\sigma_w^2$ and zero mean, but the signal component is the same in each sample. The SNR of each individual sample is therefore $SNR_1 = \frac{A^2}{\sigma_w^2}$. Now consider the integrated signal

$$x_N = \sum_{n=0}^{N-1} (x[n] + w[n]) = \sum_{n=0}^{N-1} (Ae^{j\phi} + w[n])$$

$$= NAe^{j\phi} + \sum_{n=0}^{N-1} w[n]$$
For a short period of time, the phase of a sample stays the same for the same environment and the samples are in phase with one another. Because all the signal samples add in phase with one another, the amplitude of the coherently integrated signal component is now $NA$ and the signal power will be $(NA)^2$. The power in the noise component is

$$E\left[\left|\sum_{n=0}^{N-1} w[n]\right|\right] = E\left((\sum_{n=0}^{N-1} w[n])(\sum_{l=0}^{N-1} w^*[l])\right)$$

$$= \sum_{n=0}^{N-1} \sum_{l=0}^{N-1} E\{w[n]w^*[l]\}$$

$$= N\sigma_w^2$$

where in the last step the common assumptions that the noise process $w$ is zero mean, white, and stationary have been used for simplicity. The SNR of the coherently integrated data is therefore

$$SNR_N = \frac{N^2A^2}{N\sigma_w^2} = N \frac{A^2}{\sigma_w^2} = N(SNR_1)$$

where $SNR_1$ is the SNR of a single pulse. Thus, coherent integration of $N$ data samples increases the SNR by a factor of $N$. This increase is called the integration gain [24].

### 2.6 Wideband ambiguity function

It was mentioned in Sec. 2.4 that the output of the matched filter is the cross correlation between (1) the received signal plus noise and (2) a replica of the transmitted signal. When the signal-to-noise ratio is large (as it must be for detection), the output of the matched filter can usually be approximated by the autocorrelation function of the transmitted signal; that is, if the noise is ignored. This assumes there is no Doppler shift so that the received echo signal has the same frequency as the transmitted signal. In many radar applications, however, the target is moving so that its echo signal has a Doppler frequency shift. The output of the matched filter, therefore, will not be the autocorrelation function of the transmitted signal. Instead, it must be considered as the cross correlation between the Doppler-shifted received signal and the transmitted signal, with noise being ignored since the signal-to-noise ratio is assumed to be large.
When the received echo signal is large compared to noise, the output of the matched filter may be written as the following cross-correlation function:

\[
\text{output of the matched filter} = \int_{-\infty}^{\infty} s_r(t)s^*(t - T_R)\,df
\]

where \(s_r(t)\) is the received echo signal, \(s(t)\) is the transmitted signal, \(s^*(t)\) is its complex conjugate, and \(T_R^*\) is the estimate of the time delay (considered a variable). Complex notation is assumed, so that the transmitted signal can be written as \(u(t)\exp[j2\pi f_0 t]\), where \(u(t)\) is the complex modulation function whose magnitude \(|u(t)|\) is the envelope of the real signal, and \(f_0\) is the carrier frequency. The received echo signal \(s_r(t)\) is assumed to be the same as the transmitted signal, except for a Doppler frequency shift \(f_d\) and a delay equal to the true time delay \(T_0\). Therefore,

\[
s_r(t) = u(t - T_0)\exp[j2\pi(f_0 + f_d)(t - T_0)]
\]

The change in amplitude due to system and environmental factors is ignored here. With the above definitions, the output of the matched filter is

\[
\text{output} = \int_{-\infty}^{\infty} u(t - T_0)u^*(t - T_R^*)e^{j2\pi(f_0 + f_d)(t - T_0)}e^{-j2\pi f_0(t - T_R)}\,dt
\]

For simplicity in understanding this equation, we take the origin to be the true time delay and the transmitted frequency to zero; hence, \(T_0 = 0\) and \(f_0 = 0\). Then \(T_0 - T_R^* = -T_R^* = T_R\). The output of the matched filter is then

\[
\chi(T_R, f_d) = \int_{-\infty}^{\infty} u(t)u^*(t + T_R)e^{j2\pi f_d t}\,dt
\]

A positive time delay \(T_R\) indicates a target beyond the true target time delay \(T_0\), and a positive Doppler frequency \(f_0\) indicates an approaching target. The squared magnitude of Eq. (12), \(|\chi(T_R, f_d)|^2\), is called the ambiguity function.

\[
|\chi(T_R, f_d)|^2 = \left|\int_{-\infty}^{\infty} u(t)u^*(t + T_R)e^{j2\pi f_d t}\,dt\right|^2
\]
Its three-dimensional plot as a function of time delay $T_R$ and Doppler frequency $f_d$ is the ambiguity diagram [1] [21] [23].

When the speed of propagation is very high (such as for RF waves) compared to the target speed, the pulse compression or expansion due to target movement is very minor relative to the actual length of the pulse. However, in the case of sonar, the speed of propagation in air (340.3 m/s) is not very high compared to the target speed. Therefore, even a little movement by target induces the pulse compression or expansion which cannot be ignored. The (narrowband) ambiguity function does not take into account this effect because it assumes that the speed of propagation is very high compared to the speed of target which leads to minor pulse compression or expansion. Hence, it ignores the pulse compression or expansion. Therefore, the ambiguity function should be modified, as shown in equation (14), to consider this effect.

When the transmitted signal is a wideband waveform, the echo from a static target is still a delayed and attenuated copy of the transmitted waveform but when the target moves the Doppler effect induces a time compression of the signal. The ambiguity function in this case is defined as

$$|\chi(T_R, \eta)|^2 = \frac{1}{|\eta|^2} \left| \int_{-\infty}^{\infty} u(t) u^* (\eta(t + T_R)) dt \right|^2$$  \hspace{1cm} (14)$$

in order to take this effect into account. In the equation, $\eta$ is the parameter that represents the Doppler compression and is equal to $\eta = \frac{c + v}{c - v}$, where $c$ is the speed of propagation and $v$ is the target velocity. It is self-evident that the ambiguity function is directly related to the range and Doppler resolution. Indeed, the range resolution corresponds to the width of the main lobe of the ambiguity function as a function of $T_R$ computed in $f_D = 0$, that is $|\chi(T_R, 0)|^2$. The Doppler resolution is the width of the main lobe of the ambiguity function as a function of $f_D$ computed in $T_R = 0$, that is $|\chi(0, f_D)|^2$ [25].

Similarly, a wideband-cross ambiguity function (WCAF) between two signals $u(t)$ and $s(t)$ can be defined as
\[ |\chi(T_R, \eta)|^2 = \frac{1}{|\eta|^2} \left| \int_{-\infty}^{\infty} s(t) u^*(\eta(t + T_R)) dt \right|^2 \]
3 HUMAN TONGUE CLICK ANALYSIS

As discussed in section 1.2.1 above, there are significant differences between the waveforms used in mammal, and particularly human, echolocation and manmade waveforms. Since human echolocation experts are able to undertake classification and recognition with high success rate it would seem appropriate to evaluate the performance of human waveforms. Some preliminary analysis has been undertaken [8] [12]. However, these studies have been done exclusively in the behavioral science discipline. As a result, the analysis of the transmitted signals and their echoes were undertaken using techniques inconsistent with radar engineering. There was a preliminary analysis done using techniques consistent with radar engineering [26], but this chapter presents a more detailed analysis.

3.1 Analysis of human tongue click

To perform an analysis on human echolocation waveform, a tongue click recording of a human echolocation expert (early blind – EB) participant has been used. It is a “Sound S1” recording by Thaler et al. [8] and is available to download as part of the PLoS One Journal’s open access policy. The waveform was digitized with a sampling frequency of 44.1 kHz using an in ear microphone on the left side of the head. The signal was Hilbert transformed and downsampled by a factor of 2 in order to obtain I and Q samples [27], and the real part is shown in Figure 3.
Figure 3: Real part of human tongue click

Figure 4 presents a spectrogram of the tongue click. It contains three high power non-frequency modulated components at 1.54 kHz, 3.33 kHz and 4.62 kHz. It also has a low power component at 11.03 kHz. It is also visible that different signal components have slightly different durations due to different frequency components fading in power at different times. The component of 1.54 kHz is the longest with the duration of 3 ms, whereas the component at 4.62 kHz is the shortest with the duration of 2 ms. These properties make the tongue click waveform very different from the typical radar waveforms with defined duration and frequency modulation (i.e., LFM chirp).

Figure 5 presents the spectrogram of received echo from a stationary target straight ahead of EB. The same signal components are present, but the component at 11.03 kHz seems to have higher power relative to the other frequency components than its relative power in transmitted signal. This visible change in power is merely an artifact of rescaling of the MATLAB color scale such that the maximum power of signal is represented by 0 dB, and the fact that overall power level of echo signal has been decreased due to absorption of power by target. However, it should be noted that power of component at 11.03 kHz has been increased relative to the power of component at 3.33 kHz.
Figure 4: Spectrogram of Tx tongue click
Figure 5: Spectrogram of Rx tongue click
Figure 6: WAF of tongue click
Figure 7: Zero-velocity and zero-range cuts of WAF
Figure 8: WCAF of Tx and Rx tongue click
Figure 9: Zero-velocity and zero-range cuts of WCAF
Figure 6 presents the wideband ambiguity function (WAF) of the EB’s tongue click. It is calculated using Equation (14) of section 2.6,

\[ |\chi(T_R, \eta)|^2 = \frac{1}{|\eta|^2} \left| \int_{-\infty}^{\infty} u(t) u^*(t + T_R) dt \right|^2 \]  

where \( u(t) \) is complex sampled tongue-click,

\[ v = c(\eta - 1)/(\eta + 1) \]  

is the target speed, which is positive during approach. \( \eta \) is the Doppler compression factor and \( c \) is the speed of sound in air \((340.3 \text{ms}^{-1})\). In Figure 6, the y-axis represents velocity and has been calculated using Equation (17). Figure 7 presents zero-velocity and zero-range cuts of WAF. The 3 dB bandwidths in zero-velocity and zero-range cuts of WAF provide the range resolution and velocity resolution capacity of tongue click, respectively. Zero-velocity cut in Figure 7 has 3 dB range resolution of 6 cm, whereas the primary bandwidth of EB’s tongue click \((4.62 \text{kHz} – 1.54 \text{kHz}) = 3.08 \text{kHz}\) corresponds to a range resolution of 5.5 cm. This suggests that the higher frequency component located at 11.03 kHz plays little part in range resolution observed in zero-velocity cut of Figure 7. Zero-range cut in Figure 7 has 3 dB velocity resolution of 16.5 ms\(^{-1}\).

Figure 8 presents the wideband cross-ambiguity function (WCAF) of the transmitted tongue click with received echo. The zero-velocity and zero-range cuts from the WCAF are shown in Figure 9. The zero-velocity cut in Figure 9 has 3 dB range resolution of 6.8 cm, and zero-range cut in Figure 9 has 3 dB velocity resolution of 17.2 m/s. From the comparison of Figure 7 with Figure 9, it is evident that the zero-velocity and zero-range cuts have altered in form in Figure 9 due to the alteration in received echo signal compared to transmitted click.

The significance of the 11.03 kHz frequency component visible in Figure 4 and Figure 5 is uncertain. It appears to play a little role in the range resolution capacity of the tongue click signal. The 3 dB range-resolution of tongue click is 6 cm which is very close to theoretical range resolution capacity of 5.5 cm.
for 3.08 kHz bandwidth, as mentioned above. However, the Rx signal in Figure 5 shows that this component contains higher power relative to component at 3.33 kHz than it is in Tx signal shown in Figure 4.

3.2 Producing a synthetic human tongue click

In order to understand the role of 11.03 kHz component, the nature of other components in human tongue click and obtain a noise free version of human tongue click, a synthetic human tongue click is designed and compared with the actual tongue click. Next few sections explain the process of synthetic tongue click generation and its comparison with actual human tongue click signal.

3.2.1 Deconstruction of human tongue click

Figure 4 presents that there are four main frequency components in human tongue click signal located at 1.54 kHz, 3.33 kHz, 4.62 kHz and 11.03 kHz. 100-th order FIR bandpass filters, with passbands of (900 Hz, 2000 Hz), (2400 Hz, 3700 Hz), (4000 Hz, 5200 Hz) and (10400 Hz, 11700 Hz), respectively, were used to separate the components so that they could be analyzed independently.

The time domain signals for the individual components are shown in Figure 10 to 13 as the solid lines and the signal envelop is shown as the dotted lines. It should be noted in Figure 10 to Figure 13 that the envelopes have a basic shape of the Gaussian distribution curve. The envelop of lowest frequency component shows a quick rise, followed by a nearly linear decay. The high frequency component appears to have a near Gaussian envelop while the in between components appear to show a gradual transition between the two extremes. The next section explains how this information is utilized to construct a synthetic version of the signal components and a synthetic tongue click.
3.2.2 Construction of synthetic tongue click signal

As noted in the previous section, all four components of human tongue click have a near Gaussian envelop. Therefore, a possible approach to estimate the signal could begin with the estimation of their envelops using a Gaussian curve. However, it was mentioned in previous section that the envelops do not have a shape of perfect Gaussian curve. Therefore, a novel technique of estimation has been used which involves the estimation of envelop by addition of multiple time-shifted Gaussian curves. The estimation of envelopes involves the estimation of the parameters $a_n$, $b_n$, and $c_n$ in Equation (18) via curve-fitting.

\[ e(t) = \sum_{n=1}^{N} a_n e\left(\frac{t-b_n}{c_n}\right)^2 \]  

(18)
Using curve-fitting, the estimated envelops of components at 1.54 kHz, 3.33 kHz, 4.62 kHz and 11.03 kHz, namely \( e_1(t) \), \( e_2(t) \), \( e_3(t) \) and \( e_4(t) \), were created by using parameters listed in Table 1 to 4, respectively.

Table 1: Parameters used to estimate envelop of component at 1.54 kHz

<table>
<thead>
<tr>
<th>n</th>
<th>( a_n )</th>
<th>( b_n )</th>
<th>( c_n )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-2.107</td>
<td>63.02</td>
<td>25.31</td>
</tr>
<tr>
<td>2</td>
<td>8.417</td>
<td>37.47</td>
<td>13.57</td>
</tr>
<tr>
<td>3</td>
<td>-8.16</td>
<td>37.48</td>
<td>13.4</td>
</tr>
<tr>
<td>4</td>
<td>2.025</td>
<td>62.07</td>
<td>23.67</td>
</tr>
<tr>
<td>5</td>
<td>0.3108</td>
<td>79</td>
<td>24.6</td>
</tr>
</tbody>
</table>

Table 2: Parameters used to estimate envelop of component at 3.33 kHz

<table>
<thead>
<tr>
<th>n</th>
<th>( a_n )</th>
<th>( b_n )</th>
<th>( c_n )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-65.41</td>
<td>38.91</td>
<td>11.21</td>
</tr>
<tr>
<td>2</td>
<td>0.2706</td>
<td>54.59</td>
<td>19.94</td>
</tr>
<tr>
<td>3</td>
<td>65.94</td>
<td>38.9</td>
<td>11</td>
</tr>
</tbody>
</table>

Table 3: Parameters used to estimate envelop of component at 4.62 kHz

<table>
<thead>
<tr>
<th>n</th>
<th>( a_n )</th>
<th>( b_n )</th>
<th>( c_n )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td>38.95</td>
<td>11.88</td>
</tr>
<tr>
<td>2</td>
<td>0.1685</td>
<td>51.36</td>
<td>16.97</td>
</tr>
<tr>
<td>3</td>
<td>-9.531</td>
<td>39.05</td>
<td>11.77</td>
</tr>
</tbody>
</table>
Table 4: Parameters used to estimate envelop of component at 11.03 kHz

<table>
<thead>
<tr>
<th>n</th>
<th>(a_n)</th>
<th>(b_n)</th>
<th>(c_n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-2.918</td>
<td>37.8</td>
<td>9.417</td>
</tr>
<tr>
<td>2</td>
<td>0.01775</td>
<td>50.48</td>
<td>16.54</td>
</tr>
<tr>
<td>3</td>
<td>3.046</td>
<td>37.76</td>
<td>9.512</td>
</tr>
</tbody>
</table>

The curves created using parameters of Table 1 -4 provide very good estimates of the envelop as evident in Figure 10, Figure 11, Figure 12 and Figure 13, respectively.

Then, four sinusoidal signals with constant frequencies of 1.54 kHz, 3.33 kHz, 4.62 kHz and 11.03 kHz were amplitude modulated using the estimated envelopes and are shown in Figure 14, Figure 15, Figure 16, and Figure 17, respectively, as synthetic components (blue).
It is evident from the comparison of Figure 10 to Figure 13 with Figure 14 to Figure 17, respectively, that the synthetic components are reasonably good estimates of the actual components of the signal at 1.54 kHz, 3.33 kHz, 4.62 kHz and 11.03 kHz.
Moreover, Figure 18, Figure 19, Figure 20 and Figure 21 show the comparison of the frequency spectra of the four frequency components and their synthetic components. It is evident from these figures that the synthetic components provide reasonably good estimates of the actual components while eliminating artifacts present at other frequencies.

In order to quantify the results, the root mean square error (RMSE) was found between the actual and estimated envelopes and are listed in Table 5 below. The small RMSE values indicate that selected parameter values of $a_n$, $b_n$ and $c_n$ provide reasonably good estimates of the actual envelopes.
Table 5: RMSE between actual envelop and estimated envelop

<table>
<thead>
<tr>
<th>Frequency component</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.54 kHz</td>
<td>0.0022</td>
</tr>
<tr>
<td>3.33 kHz</td>
<td>0.0073</td>
</tr>
<tr>
<td>4.62 kHz</td>
<td>0.0025</td>
</tr>
<tr>
<td>11.03 kHz</td>
<td>0.0017</td>
</tr>
</tbody>
</table>

The whole synthetic tongue click was created by using the envelopes found to amplitude modulate pure tones at frequencies of the four components and then summing these signals. This provides a clean version of tongue click, which does not contain spurious artifacts. Equation (19) presents a whole synthetic tongue click.

\[
\text{Whole synthetic tongue click} = e_1(t) \cdot \cos(2\pi \cdot 1540t) + e_2(t) \cdot \cos(2\pi \cdot 3330t) + e_3(t) \cdot \cos(2\pi \cdot 4620t) + e_4(t) \cdot \cos(2\pi \cdot 11030t)
\]  

Equation (19)

It should be noted from Figure 4 and Figure 13 that the component at 11.03 kHz has very low power relative to other components and it does not have any major contribution in signal’s range resolution. Therefore, a reduced complexity synthetic tongue click was created by removing this component and is shown in Equation (20).

\[
\text{Synthetic tongue click} = e_1(t) \cdot \cos(2\pi \cdot 1540t) + e_2(t) \cdot \cos(2\pi \cdot 3330t) + e_3(t) \cdot \cos(2\pi \cdot 4620t)
\]  

Equation (20)

3.3 Comparison of biological tongue click with synthetic tongue click signal

The original recording of the human tongue click was compared with the two synthetic version created above.
Figure 22 presents the time domain comparison of human tongue click and synthetic tongue click. As it is evident from the figure that synthetic tongue click is reasonably good estimate of the human tongue click signal. Figure 23 presents the frequency-domain comparison of human tongue click and synthetic tongue click. It should be noted that the three frequency components at 1.54 kHz, 3.33 kHz and 4.62 kHz are present in both the human tongue click and synthetic tongue click. However, the synthetic tongue click does not contain the component at 11.03 kHz, which is in agreement with Equation (20). Figure 26 presents a frequency spectrum of whole tongue click. From comparison of Figure 23 and Figure 26, it is evident that the relative power in frequency components of both versions of synthetic tongue click matches with power levels in human tongue click.
Figure 22: Tongue click comparison in time domain
Figure 23: Tongue click comparison in frequency domain
Figure 24: Zero-velocity and zero-range cuts of WAF for human tongue click
Figure 25: Zero-velocity and zero-range cuts of WAF for synthetic tongue click
Figure 26: Frequency spectrum of whole synthetic tongue click
Figure 27: Zero-velocity and zero-delay cuts of WAF for whole synthetic tongue click
Figure 24 and Figure 25 presents zero-velocity and zero-range cuts of the WAF for human tongue click and synthetic tongue click, respectively. Comparing the zero-velocity cuts in both figures, it should be noted that the first side lobe in synthetic tongue click is higher from -5.5 dB to -3.8 dB. Figure 27 presents zero-velocity and zero-range cuts of the WAF for whole synthetic tongue click. The zero-velocity cut in Figure 27 also has the first side lobe level at -3.8 dB. However, the second sidelobe in zero-velocity cut has remained the same at -5.5 dB level in synthetic tongue click and whole synthetic tongue click. The 3-dB range resolution in both the whole synthetic tongue click and synthetic tongue click has remained the same as human tongue click, which is 5.5 cm.

It should be noted that the 3-dB velocity-resolution has been decreased to 26 m/s in both versions of synthetic tongue click, compared to the 3-dB velocity resolution of 16 m/s in human tongue click. This decrease in velocity resolution is due to the decrease in duration of both versions of synthetic tongue clicks, compared to the duration of human tongue click. Since all of the experiments in this research are confined to stationary transmitter/receiver and stationary targets, the poorer velocity-resolution of synthetic tongue click is not a concern.

In conclusion, both whole synthetic tongue click and synthetic tongue click provide very good estimates of the human tongue click. The elimination of 11.03 kHz component does not affect the range resolution of synthetic tongue click from whole synthetic tongue click. Hence, the component at 11.03 kHz does not play role in range resolution capacity of the signal. Moreover, the synthetic tongue click provides an estimate of the human tongue click with reduced complexity. Therefore, the acoustic radar system explained in the next section involves the use of synthetic tongue click signal as a transmit signal for the purpose of target classification.
4 ACOUSTIC RADAR SYSTEM

As discussed in section 1.2.1, blind echolocation experts use tongue clicks generated using their mouth to perform echolocation. They are able to retrieve information from the echoes of their tongue click reflected from different objects in the environment. The fact that blind echolocation is acoustic; the lower speed of propagation of sound in air gives fine resolution compared to radio-frequency waves. The finer resolution is essential for indoor navigation, since the indoor environment is more cluttered.

The acoustic radar system can be efficiently used to transmit tongue click signals and record their echoes from the environment. The fact that it works at acoustic frequency enables it to provide fine resolution at reduced data rates lowering the data processing needs. Therefore, an acoustic radar system has been designed to mimic the human echolocation behavior. This chapter discusses the designed acoustic radar system and its calibration.

4.1 Acoustic radar system

The acoustic radar system is built using National Instrument (NI) hardware and LabVIEW software. The NI hardware consists of a PXIe-1082 chassis which interface over the PCIe bus with a Windows 7 workstation computer. To interface with the workstation, an NI PCIe card has been installed into a full 16x PCIe slot within the workstation to ensure maximum throughput between the PXIe-1082 chassis and the LabVIEW software installed on workstation. To drive and monitor audio hardware, two multifunction DAQ expansion cards (NI PXIe-6368 X-series) have been installed into the PXIe-1082 chassis. Each of these expansion cards supports up to 4 analog....
outputs with sampling rates up to 3.33 MS/s/channel as well as 16 analog inputs at sampling rates up to 2 MS/s/channel. For this research, only one output channel and one input channel is used. Both the input and output channels are samples with 16-bit resolution. Each DAQ expansion card is then connected with two BNC-2110 interface boards which are used to interface the audio hardware with the DAQ expansion cards.

The acoustic radar system has been equipped with one speaker as a transmitter (Tx) and one microphone as a receiver (Rx). This setup allows for imitation of echolocating human where speaker represents the mouth of human and the microphone represents the ear of a human. Figure 28 presents a diagram of the acoustic radar system.

*Refer to Appendix A for pictures of the acoustic radar system.*

![Diagram of Acoustic Radar System](image)

The radar control software allows transmission of any desired signal after importing the signal from an ASCII file format and outputs the transmitted signal and signals captured by the receiver in ASCII file formats.

For this research, two signals are imported into LabVIEW, namely, LFM chirp and synthetic tongue click. Both signals are 3.5 ms long in duration and have the bandwidth of 3.08 kHz spanning from 1.54 kHz to
4.62 kHz. Since these signals propagate in air at a speed of 340.3 m/s, the physical length of these signals in air is \((3.5 \times 10^{-3} \text{ s}) \left( 340.3 \frac{m}{s} \right) = 1.19 \text{ m} \).

The system operates at a pulse repetition interval (PRI) of 0.096 seconds and a pulse repetition frequency (PRF) of 10.42 Hz. This leads to an unambiguous range \(R_{unambiguous} = \frac{c \cdot PRI}{2} = \frac{(340.3 \frac{m}{s})(0.096 \text{ s})}{2} = 16.3 \text{ m} \). This unambiguous range is sufficient to perform experiments in the laboratory where the maximum possible range is 12.5 meters, as discussed in the next section.

### 4.2 Laboratory setup

Figure 29 presents the diagram of laboratory setup. The laboratory room is 16.38 m wide and 12.5 m long and the ceiling is 3 m high. The acoustic radar is placed into this room such that the maximum feasible range is 12.5 m. There are three pillars located left of the acoustic radar system along the 12.5 m dimension. These pillars are approximately 4.1 m apart from each other. The floor is made of concrete cement.

The laboratory room is also equipped with other steel storage racks as shown in diagram. There are also some equipment placed near backwall, as indicated in Figure 29. The thick lines on both sides of the acoustic radar represent divider screens. There are also some equipment and tables located on the right side of the screen. However, those are not shown for simplicity. It should be noted that a steel storage rack is shown on the right side of the screen, because it is taller than the screen and could reflect signals back to the radar.

It should be noted that there are no objects present between the acoustic radar and the target. The target is placed on a stand created using baffles such that the height of the target aligns with the height of the acoustic radar transmitter and receiver.
4.3 Acoustic radar calibrations

In order to ensure the reliability of classification experiments’ results, it is necessary to ensure that the data being collected by the acoustic radar are correct. Therefore, a series of acoustic radar calibration experiments were performed and are discussed below. Emphasis is placed on evaluating the background noise in the laboratory, to ensure that the equipment operates in-line with basic theory, mitigating the clutter in the laboratory and evaluating the multipath environment.

4.3.1 Background noise

The HVAC system installed in the laboratory produces a prominent sound that can be heard easily by a person in the laboratory. It was, therefore, potential source of interference for the acoustic radar system and its characteristics were analyzed. The HVAC sound was recorded by operating the radar as normal, but disconnecting the transmit loudspeaker. In this manner the background sounds were digitized as they would be during an experiment, but in absence of any transmissions or target echoes.
A discrete Fourier transform was performed on the received data to evaluate its power spectrum. The HVAC interference was observed to be present in the [0, 500] Hz frequency interval. Above 500 Hz the spectrum was noise only.

The HVAC noise was completely removed by application of a digital high pass filter design to pass frequencies above 700 Hz. Since the transmit signals to be used covered frequency interval [1.54, 4.62] kHz, the filter was not expected to affect them. As such, it was concluded that the HVAC interference did not constitute a problem for the experimental trials.

### 4.3.2 Pulse integration

To overcome the low transmit power of speakers in acoustic radar system, pulse integration was employed. A metal plate target was placed at 2 meter range, and data were collected using LFM chirp and synthetic tongue click waveforms. Using this data, the radar range profile was created by matched filtering the transmitted signal with the received echoes. Then, a target peak to mean noise ratio (SNR) was calculated from the range profile. This task was repeated with various number of pulse integration and SNRs were found for each case.

Figure 30 presents the pulse integration gain achieved by system using the linear frequency modulated (LFM) chirp and synthetic tongue click signals. In Figure 30 x-axis shows number of pulses integrated and y-axis shows SNR in dB. The blue line represents theoretical gain in SNR, the green line represents gain in SNR when LFM chirp was transmitted, and the red line represents gain in SNR when synthetic tongue click was transmitted.
Figure 30: Pulse integration gain

For pulse integration of 1024 pulses, the theoretical gain is \(10 \times \log_{10} 1024 = 30\) dB. In this system, 1024 pulse integration gains for LFM chirp and synthetic tongue click are 28 dB and 25 dB, respectively. The experimental gain of 25 dB is more than sufficient to acquire high target detectability; therefore, all the data acquisition has been done to allow for integration of 1024 pulses.

4.3.3 Clutter mitigation

In order to detect a target, data was acquired for a metal plate placed at 2 meter range. Figure 31 presents the power distribution of data across the range.

The blue line represents the data acquired with the plate as target. Green line represents the data acquired after removing the target from scene. Therefore, it is referred as a background (of target) data. Red line represents the result of subtracting background data from target data. A strong peak in Figure 31 (green) at 1 meter range corresponds to the “cross-talk” between microphone and speaker. The cross-talk refers to sound travelling directly from speaker to microphone which can be misinterpreted as a strong power return from a target between 0 and 2 meter range.
Observing that the cross-talk is constant in both the target data and background data, the method of background subtraction was employed to eliminate the cross-talk in the range of 0 to 2 meter [1] [28]. In background subtraction, the data without the target is subtracted from the data taken with a target in the same environment so that all the returns coming from the other objects present in both data acquisitions will be eliminated. This will lead to higher SNR by reduction of peaks from objects present in both data acquisitions. The red line presents result of background subtraction showing the successful elimination of cross-talk in the range of 0 to 2 meter. The peak in the range of 2 to 3 meter is the target, and it was unaffected by background subtraction. Therefore, the background subtraction can be successfully used to eliminate “cross-talk” interference in acoustic radar system without affecting target peak. It should also be noted in Figure 31 that the target peak to mean noise floor ratio is approximately 30 dB, which is in agreement with the theoretical result for 1024 pulse integration as shown in Figure 30.
4.3.4 Baffles to eliminate strong returns

To further eliminate the cross-talk, baffles created using soft sponge were placed between the speaker and the microphone which resulted in the SNR gain of couple of dB [29]. While observing the data, strong peaks in range profile corresponding to the pillars in laboratory and the floor between radar and target were noticed. Therefore, the pillars and the floor between radar and target were also covered using baffles which resulted in the decrease of peak levels representing pillars and the floor by couple of dB. However, it should be evident from Figure 31 that background subtraction provides very good results in terms of cross-talk and environment elimination, compared to just a couple of dB reduction in power levels achieved by the use of baffles.

4.3.5 Evaluation of multipath environment

The transmitted signal may propagate to the target by routes other than the direct path, an effect known as multipath [30]. In classical multipath the transmitted signal is reflected from the ground, or a large clutter object, as it propagates to and from the target. When operating in an indoor environment, multipath may arise due to reflections from the floor, ceiling, walls and other objects in the room. This multipath returns may be identified as false target peaks, in case, if they are not known to be due to multipath returns. It is nearly impossible to avoid illuminating the ceiling with the transmitter in the laboratory, therefore it is essential that we understand how the ceiling reflection influences the detection performance in range profiles.

Figure 32 presents the ground (or ceiling)-plane range geometry showing the direct path $D$ and the indirect path $S_1 + S_2$ (or $S_3 + S_4$) due to the ground (or ceiling) reflection. As shown in Figure 32, energy reaches the target via a direct path $D$ and an indirect reflection path whose length is $S_1 + S_2$ (or $S_3 + S_4$). Energy scattered by the target travels back to the acoustic radar along the same path but in the opposite direction. Therefore, the ground (or ceiling)-plane effect involves no fewer than four distinct roundtrip propagation paths, two of which involve a monostatic geometry and two a bistatic geometry.
Note that the direct path $D$ and the second leg $S_2$ of the indirect path converge on the target from slightly different directions in the vertical plane. However, the bistatic angle in that plane is so small (less than $1^\circ$) that the target echo we measure on the ground-plane range is indistinguishable from the monostatic echo. Of the four possible roundtrip propagation paths, there are only three distinct path lengths:

1. Roundtrip distance along the direct path: $2D$
2. Roundtrip distance along the indirect path: $2I$
3. Roundtrip distance along the direct and indirect paths: $D + I$

where $I = S_1 + S_2$ (or $S_3 + S_4$) is the indirect path length.

These parameters can be calculated using Equations (21) and (22) [30].

\[ I = \left( (h_t + h_a)^2 + R^2 \right)^{1/2} \approx R + \frac{(h_t+h_a)^2}{2R} \]  \hspace{1cm} (21)

\[ D = \left( (h_t - h_a)^2 + R^2 \right)^{1/2} \approx R + \frac{(h_t-h_a)^2}{2R} \]  \hspace{1cm} (22)

Figure 33 presents a range profile of metal plate target located at 2 meter range. The range profile is a result of matched filtering the transmitted signal with the received echoes. It should be noted that the
pulse integration of 1024 pulses and background subtraction have also been performed. In Figure 33, x-axis represents the range in meter and y-axis represents the power in dB.

![Range Profile for Metal Plate](image)

**Figure 33: Range profile for metal plate**

In this case, the floor between acoustic radar and the target was covered with acoustic baffles to eliminate indirect path returns \((S_1 + S_2)\). The peak at 2 meter range represents the metal plate target, whereas the peak at 12.5 meter range represents back-wall of laboratory. The small peak at approximately 6 meter range corresponds to a multipath reflection from the ceiling because its location reasonably agrees with the expected location of return from path \((S_3 + S_4)\) obtained using Equation (21). The peaks at 8.5 meters and 9 meters are assumed to be the reflections from other objects in the laboratory.

Finally, the calibrations of acoustic radar system discussed above ensure that the data obtained by the system are reliable. The next step is to perform classification experiments using different types of targets which are discussed in Chapter 5.
5 CLASSIFICATION EXPERIMENTS

It has been evident from the literature review given in section 1.2.1 that human echolocation experts can successfully classify objects using their tongue clicks. The classification is reported as being based on the target size, shape, and material. To investigate the capability of the tongue click waveform for the purpose of classification, the acoustic radar system was programmed to transmit the synthetic replica of the human tongue click described in section 3.2.2. The radar illuminated a target set, which exhibited diversity in the characteristics described as allowing humans to classify different targets, and the received echoes were recorded for analysis. Moreover, these measurements were also performed with LFM chirp signal to allow comparison with the most common manmade waveform.

5.1 Experimental setup

Figure 34 presents the diagram of experimental setup for the classification experiments. The target was placed at 2 meter range from the acoustic radar between the transmitter (Tx) and receiver (Rx). The target was placed on a stand created using acoustic baffles such that its height is aligned vertically with the height of acoustic radar’s transmitter and receiver. The floor between the acoustic radar and target is covered with acoustic baffles to eliminate multipath returns from the floor. The thick horizontal lines on both sides of acoustic radar represent divider screens, as mentioned in section 4.2. The transmitter and receiver are separated by placing baffles between them to reduce the cross-talk, as discussed in section 4.3.4, and are represented using square blocks between Tx and Rx in Figure 34.
5.2 Types of targets used

The literature [8] [9] [14] discussed in section 1.2.1 suggested that the human echolocation experts have ability to distinguish objects based on their size, shape, and material. Therefore, various types of targets were selected based on their size, shape and material for the purpose of target classification. Three targets made of metal were selected which had the shapes of sphere, flat plate, and dihedral. A hard plastic box with some surface texture was selected to introduce a change in material type and texture. The plastic box represents a cubic shape. A chair was selected which was made of soft cushion on the front and hard plastic on the back. This served as two distinct targets made of two different materials, namely, soft cushion and hard plastic, when looked at from front and back. The last target was a standing human, which provided a different shape as well as soft material base due to human’s soft tissue of skin and clothing. The sphere is smaller in size than a dihedral, but they are made of the same material. The metal plate and plastic box are similar in size, but they are made of different materials. The chair front and back are same in size and shape, but the material on front side is softer than the material on back side. The human is made of soft tissue and is different in size and shape than any other targets. Thus, this set of targets provides a diversified class of targets which are used in classification experiments discussed in next section.
5.3 Classification

In order to classify different targets, various features of returned echoes were analyzed when the targets were illuminated with LFM chirp and synthetic tongue click signals. These features are discussed in the next few sections.

5.3.1 Radar range profile

Figure 35 and Figure 36 presents the radar range profiles created by matched filtering the received echoes with LFM chirp and synthetic tongue click, respectively, as transmit signals and using the signal processing chain described in section 4.3. In both figures, x-axis represents range in meters and y-axis represents power in dB.

![Range Profile for Metal Plate 1](image1)
![Range Profile for Metal Plate 1](image2)

Figure 35: Radar range profile for metal plate using LFM chirp as Tx signal

Figure 36: Radar range profile for metal plate using synthetic tongue click as Tx signal

The peaks at 2 meter and 12.5 meter represent metal plate target and back-wall, respectively. As discussed in Section 4.3.5, the peak at 6 m represents the multipath return, and the peaks at 8.5 m and 9 m represent returns from objects in the laboratory. It should be noted that the peak corresponding to target at 2 meter range is narrower in Figure 35 than it is in Figure 36. For LFM chirp signal, the zero-velocity cut of WAF has main lobe to first sidelobe ratio of 14.02 dB, which leads to narrow peak appearance of target peak (only main lobe) in range profile. For synthetic tongue click signal, the zero-velocity cut of WAF
has main lobe to first sidelobe ratio of 4.02 dB, which leads to the wider appearance of target peak (main lobe and first sidelobes together) in range profile. The range profile created using synthetic tongue click signal does not seem to provide any additional information than the range profile created using LFM chirp signal.

5.3.2 Target signatures

Figure 37–50 show target signatures created using LFM chirp and synthetic tongue click signals for the targets used in classification experiments. The x-axis represents the range and y-axis represents the absolute power. The target signature is created by extracting the peak representing a target in the range profile. In the process of target signature extraction, enough samples were extracted such that the target signature represents the range of 1.2 meter which is larger than the length of any of the targets used in classification experiments. The target signature is a zoomed in view of the peak representing the target in range profile.

It should be noted from the target signatures that the peak power changes for different types of targets. Regardless of waveform, the principal separation between the targets is their back scatter power. The back scatter power is directly proportional to the physical size of the target for the same material type. The sphere, dihedral and plate are made of the same metal and the peak power levels in their signatures created using synthetic tongue click are approximately 0.3, 0.5 and 3.4, respectively. This also holds true for the signatures created using chirp, where the peak power levels for sphere, dihedral and plate are 0.35, 0.7 and 4.2, respectively.

The plastic box is similar in size to metal plate, but both of them are made of hard material. The target signatures created using synthetic tongue click show that plastic box and metal plate have peak powers of 3 and 3.4, respectively, which are very close. This also holds true for their target signatures created using chirp, since the plastic box and metal plate have the peak power levels of 4 and 4.2, respectively.
The human, chair front and chair back have peak power levels of 0.2, 0.5, and 0.9, respectively, for the signatures created using chirp and peak power levels of 0.23, 0.25 and 0.8, respectively, for the signatures created using synthetic tongue click. It should be noted that all three objects are made of different materials. The human is made of a soft tissue and is relatively large in size, whereas the chair front is made of a soft cushion. Both of these targets are made of softer material than the hard plastic material of chair back. Therefore, the peak power levels of human and chair front are lower than the peak power levels of chair back.

It can be concluded that, regardless of the waveform, different targets can be distinguished based on their backscatter power, which is proportional to their size, if they are made of the same material. Furthermore, targets made of different materials can also be distinguished, since the back scattered power for objects made of softer materials is less than the back scattered power of objects made of hard materials.
Figure 37: Target signature of metal dihedral (chirp)

Figure 38: Target signature of metal dihedral (tongue click)

Figure 39: Target signature of metal plate (chirp)

Figure 40: Target signature of metal plate (tongue click)

Figure 41: Target signature of metal sphere (chirp)

Figure 42: Target signature of metal sphere (tongue click)
Figure 43: Target signature of human (chirp)

Figure 44: Target signature of human (tongue click)

Figure 45: Target signature of plastic box (chirp)

Figure 46: Target signature of plastic box (tongue click)

Figure 47: Target signature of chair front (chirp)

Figure 48: Target signature of chair front (tongue click)
5.3.3 Absolute value of target echoes

It is discussed above that the acoustic radar system records the raw ADC signal allowing the extraction of the raw target echo before matched filtering. It is also known that the exact delay time for the received echo to arrive is represented by the peak of the target in radar range profile. By knowing the duration after which the peak appears in radar range profile, it can be said that the echo from the target arrives after the same duration in the raw received signal. In the case of a stationary, point-like target, extracting a part of the received signal which starts at the duration of the peak and ends at after some duration such that the length of the extracted part equals the length of the transmit signal would provide the target echo.

Figure 51 – 64 provides the absolute value of the extracted target echoes for different targets. The x-axis provides length of the signal in metres and y-axis provides the absolute value of power of the signal. It can be observed that the absolute values of the echoes for chirp signal have two major peaks at approximately 0.2 m and 0.4 m range for all the targets except the human and chair front, both of which are made of soft material. However, the absolute values of the echoes for tongue click signal have multiple peaks across the range axis and the plots representing human and chair front are not so distinct from others, when compared to the case for chirp signal.
The difference between peak power levels of the absolute value of target echoes for different targets does not appear to provide distinguishing features as it did for the matched filtered signatures. Therefore, it was concluded that absolute value of target echoes was not a useful target classification feature. However, the echo signatures can also be analyzed in the frequency domain, rather than the time domain approach used here. The result of such an investigation is discussed in the next section.
Figure 51: Absolute value of target echo for dihedral (chirp)

Figure 52: Absolute value of target echo for dihedral (tongue click)

Figure 53: Absolute value of target echo for plate (chirp)

Figure 54: Absolute value of target echo for plate (tongue click)

Figure 55: Absolute value of target echo for sphere (chirp)

Figure 56: Absolute value of target echo for sphere (tongue click)
Figure 57: Absolute value of target echo for human (chirp)

Figure 58: Absolute value of target echo for human (tongue click)

Figure 59: Absolute value of target echo for plastic box (chirp)

Figure 60: Absolute value of target echo for plastic box (tongue click)

Figure 61: Absolute value of target echo for chair front (chirp)

Figure 62: Absolute value of target echo for chair front (tongue click)
5.3.4 Frequency spectrum of echoes

Figure 65–78 present the normalized frequency spectrum of echoes and transmit signal. The frequency spectrum is normalized such that the maximum power is equal to 0 dB. The x-axis shows frequency in Hz, and y-axis shows the power in dB.

As it is discussed in section 2.5, the pulse integration can be used to increase SNR. While performing the pulse integration of 1024 pulses, it was observed that integration of the same number of pulses for different targets leads to different SNRs due to the difference in reflected power for different objects, as discussed above. This change in SNR could lead to visible differences in normalized frequency spectra of different targets, which is merely an artifact of the difference in SNRs. Therefore, the SNR of the received echo signals were equalized by integration of different number of pulses. It was noticed that the chair front target provided the lowest SNR of 15 dB with 1024 pulse integration, but it was more than sufficient for the classification experiments. Hence, different numbers of pulses were integrated for different targets to achieve an SNR of approximately 15 dB. The number of pulses integrated for each target to obtain the approximate SNR of 15 dB is shown in Table 6.
Table 6: Number of pulses integrated to equalized SNR

<table>
<thead>
<tr>
<th></th>
<th>Dihedral</th>
<th>Plate</th>
<th>Sphere</th>
<th>Human</th>
<th>Plastic box</th>
<th>Chair front</th>
<th>Chair back</th>
</tr>
</thead>
<tbody>
<tr>
<td>LFM Chirp</td>
<td>16 pulses</td>
<td>1 pulse</td>
<td>64 pulses</td>
<td>32 pulses</td>
<td>1 pulse</td>
<td>32 pulses</td>
<td>8 pulses</td>
</tr>
<tr>
<td>Synthetic tongue click</td>
<td>64 pulses</td>
<td>1 pulse</td>
<td>64 pulses</td>
<td>64 pulses</td>
<td>1 pulse</td>
<td>1024 pulses</td>
<td>8 pulses</td>
</tr>
</tbody>
</table>

Comparing the frequency spectrum of received signals for the chirp, it can be seen that the frequency spectrum of echoes from human, chair front and chair back have distinctive shape than any other targets. For all other targets, the frequency spectrum of echo is very similar in shape to the frequency spectrum of transmitted chirp signal.

Comparing the frequency spectrum of echoes for the synthetic tongue click, it can be seen that there are three distinct peaks at 1.54 kHz, 3.33 kHz and 4.62 kHz, representing the main frequency components available in synthetic tongue click signal. It should be noted that the central peak representing 3.33 kHz component has the same power in the frequency spectra of transmit signal and echoes, which is the result of normalization by a maximum power. However, the power of the 1.54 kHz and 4.62 kHz components in echo spectra relative to the spectrum of transmit signal have changed for different targets. This suggests that the ratios of powers in 1.54 kHz and 4.62 kHz component relative to the power in 3.33 kHz component has changed in spectra of echoes compared to the spectra of transmit signal. This change in ratios suggests that there is information about the target in frequency spectra of echoes. It is difficult to analyze the distinctive behavior of any particular target based on these ratios just by looking at the plots of frequency spectrum for different targets.
In order to better understand the differences between the frequency spectra of different targets’ echoes, the correlation-coefficients have been found for different features and are discussed in next section.
Figure 65: Frequency spectrum - dihedral (Chirp)

Figure 66: Frequency spectrum - dihedral (tongue click)

Figure 67: Frequency spectrum - plate (chirp)

Figure 68: Frequency spectrum - plate (tongue click)

Figure 69: Frequency spectrum - sphere (chirp)

Figure 70: Frequency spectrum - sphere (tongue click)
Figure 71: Frequency spectrum - human (chirp)

Figure 72: Frequency spectrum - human (tongue click)

Figure 73: Frequency spectrum - plastic box (chirp)

Figure 74: Frequency spectrum - plastic box (tongue click)

Figure 75: Frequency spectrum - chair front (chirp)

Figure 76: Frequency spectrum - chair front (tongue click)
5.3.5 Correlation-coefficient for different features

In this section the correlation coefficient is used to quantify the differences observed between the different target features discussed above.

The correlation coefficient between two signals $X$ and $Y$ can be defined as

$$\rho_{XY} = \frac{\text{cov}(X,Y)}{\sqrt{\text{var}(X)\text{var}(Y)}}$$  \hspace{1cm} (22)

where $\text{var}(X) = E[(X - E[X])^2]$ and $\text{cov}(X,Y) = E[(X - E[X])(Y - E[Y])]$.

The $E[X]$ measures the mean of $X$, and $E[(X - E[X])^2]$ measures the average squared deviation from the mean. Covariance is a measure of how much two variables change together. When $X$ and $Y$ have $\rho_{XY} \neq 0$, then $X$ and $Y$ are said to be correlated. If, however, the covariance is zero and hence $\rho_{XY} = 0$, then $X$ and $Y$ are said to be uncorrelated. The correlation coefficient is always less than 1 in magnitude or $|\rho_{XY}| \leq 1$ [31].

For the purpose of target classification, the higher correlation coefficient indicates that the two signals are highly correlated and it is difficult to classify them as distinct signals, whereas the lower correlation
coefficient indicates that the two signals are less correlated and it is easier to classify them as distinct signals.

Table 7 and 8 show the absolute correlation coefficient between normalized target signatures created using chirp and tongue click signals. In these tables, the cells with correlation coefficient higher than 0.7 are represented using red color, where the red indicates that the two targets cannot be classified as different using this feature. The cells with correlation coefficient lower than 0.4 are represented using the green color, where green indicates that the two targets can be classified as different using this feature. The cells with correlation coefficients between 0.4 and 0.7 are represented with yellow color, where yellow indicates that the two targets may or may not be classified as different using this feature. This convention is also applied to the correlation coefficients between other features of the other tables as well.

Table 7: Absolute Correlation coefficient between normalized target signatures (chirp)

<table>
<thead>
<tr>
<th></th>
<th>Dihedral</th>
<th>Plate</th>
<th>Sphere</th>
<th>Human</th>
<th>Plastic Box</th>
<th>Chair Front</th>
<th>Chair Back</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dihedral</td>
<td>1</td>
<td>0.9599</td>
<td>0.9435</td>
<td>0.6079</td>
<td>0.9492</td>
<td>0.8028</td>
<td>0.6921</td>
</tr>
<tr>
<td>Plate</td>
<td>0.9599</td>
<td>1</td>
<td>0.9342</td>
<td>0.5450</td>
<td>0.9748</td>
<td>0.8111</td>
<td>0.6734</td>
</tr>
<tr>
<td>Sphere</td>
<td>0.9435</td>
<td>0.9342</td>
<td>1</td>
<td>0.6060</td>
<td>0.9300</td>
<td>0.7595</td>
<td>0.6665</td>
</tr>
<tr>
<td>Human</td>
<td>0.6079</td>
<td>0.5450</td>
<td>0.6060</td>
<td>1</td>
<td>0.5694</td>
<td>0.4858</td>
<td>0.3748</td>
</tr>
<tr>
<td>Plastic Box</td>
<td>0.9492</td>
<td>0.9748</td>
<td>0.9300</td>
<td>0.5694</td>
<td>1</td>
<td>0.8150</td>
<td>0.6796</td>
</tr>
<tr>
<td>Chair Front</td>
<td>0.8028</td>
<td>0.8111</td>
<td>0.7595</td>
<td>0.4858</td>
<td>0.8150</td>
<td>1</td>
<td>0.6037</td>
</tr>
<tr>
<td>Chair Back</td>
<td>0.6921</td>
<td>0.6734</td>
<td>0.6665</td>
<td>0.3748</td>
<td>0.6796</td>
<td>0.6037</td>
<td>1</td>
</tr>
</tbody>
</table>
### Table 8: Absolute Correlation coefficient between normalized target signatures (tongue click)

<table>
<thead>
<tr>
<th></th>
<th>Dihedral</th>
<th>Plate</th>
<th>Sphere</th>
<th>Human</th>
<th>Plastic Box</th>
<th>Chair Front</th>
<th>Chair Back</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dihedral</td>
<td>1</td>
<td>0.9650</td>
<td>0.9549</td>
<td>0.7480</td>
<td>0.9507</td>
<td>0.7727</td>
<td>0.9168</td>
</tr>
<tr>
<td>Plate</td>
<td>0.9650</td>
<td>1</td>
<td>0.9752</td>
<td>0.7721</td>
<td>0.9594</td>
<td>0.7669</td>
<td>0.9168</td>
</tr>
<tr>
<td>Sphere</td>
<td>0.9549</td>
<td>0.9752</td>
<td>1</td>
<td>0.8328</td>
<td>0.9669</td>
<td>0.7752</td>
<td>0.9309</td>
</tr>
<tr>
<td>Human</td>
<td>0.7480</td>
<td>0.7721</td>
<td>0.8328</td>
<td>1</td>
<td>0.8722</td>
<td>0.8218</td>
<td>0.7635</td>
</tr>
<tr>
<td>Plastic Box</td>
<td>0.9507</td>
<td>0.9594</td>
<td>0.9669</td>
<td>0.8722</td>
<td>1</td>
<td>0.8525</td>
<td>0.9299</td>
</tr>
<tr>
<td>Chair Front</td>
<td>0.7727</td>
<td>0.7699</td>
<td>0.7752</td>
<td>0.8518</td>
<td>0.8525</td>
<td>1</td>
<td>0.8196</td>
</tr>
<tr>
<td>Chair Back</td>
<td>0.9168</td>
<td>0.9168</td>
<td>0.9309</td>
<td>0.7635</td>
<td>0.9299</td>
<td>0.8196</td>
<td>1</td>
</tr>
</tbody>
</table>

As it is evident from Table 7, majority of the targets cannot be classified easily using normalized target signatures. The human and chair back can be easily classified as different from each other based using normalized target signature, as indicated by green color. Moreover, human and chair back may be classified as different from all other targets, as indicated by yellow color. However, the normalized target signatures of all other targets are highly correlated and therefore, they cannot be classified as distinct targets.
In Table 8, all of targets normalized signatures are highly correlated, as indicated by red color. Therefore, they cannot be classified as different targets using normalized target signatures created by transmission of synthetic tongue click signal.

It can be concluded, regardless of the transmit waveform, that the performance of normalized target signature is very poor for the task of target recognition.

**Table 9: Absolute correlation coefficient between normalized frequency spectrums of echoes (chirp)**

<table>
<thead>
<tr>
<th>Absolute Correlation Coefficient - Normalized Frequency Spectrum of Echoes (Chirp)</th>
<th>Dihedral</th>
<th>Plate</th>
<th>Sphere</th>
<th>Human</th>
<th>Plastic Box</th>
<th>Chair Front</th>
<th>Chair Back</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dihedral</td>
<td>1</td>
<td>0.9183</td>
<td>0.6933</td>
<td>0.7057</td>
<td>0.7177</td>
<td>0.7182</td>
<td>0.5497</td>
</tr>
<tr>
<td>Plate</td>
<td>0.9183</td>
<td>1</td>
<td>0.7325</td>
<td>0.7260</td>
<td>0.8411</td>
<td>0.7082</td>
<td>0.6339</td>
</tr>
<tr>
<td>Sphere</td>
<td>0.6933</td>
<td>0.7325</td>
<td>1</td>
<td>0.5587</td>
<td>0.7265</td>
<td>0.5348</td>
<td>0.5771</td>
</tr>
<tr>
<td>Human</td>
<td>0.7057</td>
<td>0.7260</td>
<td>0.5587</td>
<td>1</td>
<td>0.5269</td>
<td>0.5397</td>
<td>0.2287</td>
</tr>
<tr>
<td>Plastic Box</td>
<td>0.7177</td>
<td>0.8411</td>
<td>0.7265</td>
<td>0.5269</td>
<td>1</td>
<td>0.7318</td>
<td>0.6598</td>
</tr>
<tr>
<td>Chair Front</td>
<td>0.7182</td>
<td>0.7082</td>
<td>0.5343</td>
<td>0.5397</td>
<td>0.7318</td>
<td>1</td>
<td>0.2961</td>
</tr>
<tr>
<td>Chair Back</td>
<td>0.5497</td>
<td>0.6339</td>
<td>0.5771</td>
<td>0.2287</td>
<td>0.6598</td>
<td>0.2961</td>
<td>1</td>
</tr>
</tbody>
</table>
Table 10: Absolute correlation coefficient between normalized frequency spectrums of echoes (tongue click)

<table>
<thead>
<tr>
<th></th>
<th>Dihedral</th>
<th>Plate</th>
<th>Sphere</th>
<th>Human</th>
<th>Plastic Box</th>
<th>Chair Front</th>
<th>Chair Back</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dihedral</td>
<td>1</td>
<td>0.8523</td>
<td>0.6094</td>
<td>0.3520</td>
<td>0.5812</td>
<td>0.5585</td>
<td>0.4490</td>
</tr>
<tr>
<td>Plate</td>
<td>0.8523</td>
<td>1</td>
<td>0.5268</td>
<td>0.3471</td>
<td>0.6645</td>
<td>0.4139</td>
<td>0.4109</td>
</tr>
<tr>
<td>Sphere</td>
<td>0.6094</td>
<td>0.5268</td>
<td>1</td>
<td>0.5046</td>
<td>0.5912</td>
<td>0.6300</td>
<td>0.5973</td>
</tr>
<tr>
<td>Human</td>
<td>0.3520</td>
<td>0.3471</td>
<td>0.5046</td>
<td>1</td>
<td>0.6126</td>
<td>0.6827</td>
<td>0.5418</td>
</tr>
<tr>
<td>Plastic Box</td>
<td>0.5812</td>
<td>0.6645</td>
<td>0.5912</td>
<td>0.6126</td>
<td>1</td>
<td>0.5648</td>
<td>0.5552</td>
</tr>
<tr>
<td>Chair Front</td>
<td>0.5585</td>
<td>0.4139</td>
<td>0.6300</td>
<td>0.6827</td>
<td>0.5648</td>
<td>1</td>
<td>0.4747</td>
</tr>
<tr>
<td>Chair Back</td>
<td>0.4499</td>
<td>0.4109</td>
<td>0.5973</td>
<td>0.5418</td>
<td>0.5552</td>
<td>0.4747</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 9 -10 shows the absolute value of correlation coefficients between the normalized frequency spectrums of echoes for different targets, created using chirp and tongue click, respectively. As it can be seen from Table 9 that human and chair front can be easily classified as different targets using normalized frequency spectrum of echoes created using chirp signal. The chair front and chair back targets can also be classified as different targets, as indicated by green color in Table 9. It should be noted, in Table 9, that chair back may be classified as different target from all the other targets, as indicated by yellow color.

Table 10 indicates that human can be easily classified as a different target from dihedral and plate and is shown by green color. The other targets may be classified as different targets from each other using normalized frequency spectrum of echoes created by transmission of tongue click signal. It should be
noted that the performance of tongue click seems to be better than the performance of chirp signal using normalized frequency spectrum of echoes for target classification because Table 10 has more yellow (and less red) cells than Table 9.

Overall, the target classification performance seems to be better using normalized frequency spectrum of echoes than normalized target signatures.

Table 11: Absolute correlation coefficient between absolute difference between normalized frequency spectrum of transmit signal and normalized frequency spectrum of echo (chirp)

<table>
<thead>
<tr>
<th></th>
<th>Dihedral</th>
<th>Plate</th>
<th>Sphere</th>
<th>Human</th>
<th>Plastic Box</th>
<th>Chair Front</th>
<th>Chair Back</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dihedral</td>
<td>1</td>
<td>0.7877</td>
<td>0.3774</td>
<td>0.1110</td>
<td>0.1942</td>
<td>0.7882</td>
<td>0.3172</td>
</tr>
<tr>
<td>Plate</td>
<td>0.7877</td>
<td>1</td>
<td>0.5907</td>
<td>0.1355</td>
<td>0.5152</td>
<td>0.5917</td>
<td>0.5557</td>
</tr>
<tr>
<td>Sphere</td>
<td>0.3774</td>
<td>0.5907</td>
<td>1</td>
<td>0.0118</td>
<td>0.5609</td>
<td>0.3252</td>
<td>0.4435</td>
</tr>
<tr>
<td>Human</td>
<td>0.1110</td>
<td>0.1355</td>
<td>0.0118</td>
<td>1</td>
<td>0.1973</td>
<td>0.2696</td>
<td>0.1516</td>
</tr>
<tr>
<td>Plastic Box</td>
<td>0.1942</td>
<td>0.5152</td>
<td>0.5609</td>
<td>0.1973</td>
<td>1</td>
<td>0.3502</td>
<td>0.4220</td>
</tr>
<tr>
<td>Chair Front</td>
<td>0.7882</td>
<td>0.5917</td>
<td>0.3252</td>
<td>0.2696</td>
<td>0.3502</td>
<td>1</td>
<td>0.2434</td>
</tr>
<tr>
<td>Chair Back</td>
<td>0.3172</td>
<td>0.5557</td>
<td>0.4435</td>
<td>0.1516</td>
<td>0.4220</td>
<td>0.2434</td>
<td>1</td>
</tr>
</tbody>
</table>
While observing frequency spectrum of echoes, in Section 5.3.4, it was noted that the normalized frequency spectrum of echoes have altered relative power in three main frequency components of the signal (1.54 kHz, 3.33 kHz, and 4.62 kHz) compared to the relative powers in transmitted signal. This change seems to be significant for different types of targets. Therefore, in order to identify if this change in relative power of different frequency bands compared to the relative power in transmit signal can be used for the purpose of target classification, the absolute different between normalized frequency spectrum of transmit signal and normalized frequency spectrum of echoes were found for different targets and used it as a feature. Then, the absolute correlation coefficient was found between these features of
targets. The results are shown in Table 11 and 12 for the features created using chirp and tongue click waveforms, respectively.

From Table 11, it is evident that human can be easily classified as a different target from all other targets and is indicated by green color of cells. Dihedral and chair front can be easily classified as different targets from sphere, human, plastic box and chair back, however, they cannot be classified as different from each other. The sphere and chair back can be easily classified as different targets from dihedral, human and chair front, however, they may be classified as different from each other. Chair back can be easily classified as different from dihedral, human and chair front, however, it may be classified as different from plate, sphere and plastic box. Plate can only be classified as different from human, and it cannot be classified as different from dihedral. However, it may be classified as different from all other targets.

From Table 12, it is evident that chair front can be easily classified as different from all the other targets, expect human. Human can be easily classified as different target from all the targets expect chair front and chair back. Also, human may be classified as different target from chair front and chair back. Plastic box may be classified as different only from sphere. All the other targets cannot be classified as different from each other.

Overall, it should be noted that tongue click signal performs better at classifying objects made with hard material from the objects made with soft materials. This is in agreement with the findings of pervious researchers [8] [9], where they conclude that humans can successfully classify objects based on their material. However, the chirp waveform does not seem to display any particular pattern for the purpose of target classification. It should be noted that some of the targets can be classified as different using chirp waveform, but they cannot be classified as different using tongue click waveform. It can be concluded, from Table 11 and 12, that human tongue click waveform provides much better performance for classification between hard and soft materials when compared to chirp waveform, and hence, it could be
one of the reasons why many echolocators prefer tongue click waveform for the task of target classification.

5.4 Summary

The results of Section 5.3 indicates that, regardless of the waveform, the absolute difference between the normalized frequency spectrum of transmit signal and normalized frequency spectrum of echoes provide much better classification performance than the normalized target signatures or normalized frequency spectrum of echoes.

From Table 11 and 12, it could be concluded that absolute difference between normalized frequency spectrum of transmit signal and normalized frequency spectrum of echo can be used to classify objects made from soft materials and object made of hard materials using tongue click waveform, but it may not be possible using chirp waveform. Hence, tongue click waveform performs much better for target classification based on material than chirp waveform. This could be one of the main reasons why humans have adopted tongue click waveform for the purpose of echolocation. It is also evident that humans can successfully navigate in environment and classify objects utilizing multiple perspectives of the objects in the environment.
6 MOMENT INVARiANCE BASED AUTOMATIC TARGET RECOGNITION (ATR)

It is evident that humans can successfully navigate in their environment using echolocation [8]. While navigating in their environment, they are constantly able to classify different objects by utilizing multiple perspectives of the objects [32] [33]. Taking cue from this phenomenon, the effect of using multiple perspectives in distinguishing targets has been studied with moment invariant (MI) based HRR-ATR. The algebra of moment invariance has been utilized to develop a novel ATR system which can classify targets using multiple perspectives, and can potentially be implemented using efficient table-look up hardware. This chapter presents the results of applying the theory of moment invariants (MI) on High Range Resolution (HRR) profile data for the purpose of target classification and utilizes confusion matrix as a performance metric.

6.1 Invariant algebra and justification for use in HRR-ATR

The HRR profiles of each object change with aspect look-angle. This has driven an investigation for utilizing the invariant features for a particular type of object that differs for different target classes but do not vary significantly for a limited range of aspect angles. The other goal is to minimize the number of features used to represent a target for efficient ATR implementation.

It may be noted that in image processing literature, the seven invariant moments proposed by Hu are commonly used [20]. The Hu moments are invariant to rotation also, making them highly appropriate for Two-dimensional images. However, HRR radar profiles being used in this work are obtained by mapping three-dimensional target information into one-dimensional signals that
represent reflected radar intensity along target extent in range at a particular aspect angle. It may be noted that echo returns to be exploited in ATR are also one-dimensional signals from which targets need to be classified. The phenomenology of HRR and echo return data appears to be similar in the sense that both are range vs. range return data. Therefore, in case of HRR profiles as well as for echo returns, translation and scale invariance are critical. Target-rotation or directional information is incorporated in the particular aspect look-angle at which the HRR profile (or echo return) is collected.

For Table-Look-up based efficient ATR implementation, it will be beneficial if the look-angle dependent information can be represented with minimal number of features or parameters which do not vary or are slowly varying with aspect angle. With these goals in mind, in this research, a novel approach of HRR-ATR by exploiting the theory of invariant algebra to develop ATR algorithm using HRR profiles has been considered. Specifically, the use of Central Moments based on range-locations \(x\) and corresponding range returns \(y\) has been considered and is defined as,

\[
M_{p,q} = \iint (x - \mu_x)^p (y - \mu_y)^q f(x,y)dxdy
\]

(23)

where, \(x\) and \(y\) represent relative range locations and corresponding range returns, respectively. It is straight-forward to show that these moments are translation invariant because \(\mu_{x+a} = E[x+a] = \mu_x + a, \mu_{y+b} = E[y+b] = \mu_y + b\), and

\[
\iint (x + a - \mu_{x+a})^p (y + b - \mu_{y+b})^q f(x,y)dxdy = \iint (x - \mu_x)^p (y - \mu_y)^q f(x,y)dxdy = M_{p,q}
\]

(24)

where, \(a\) represents translation in range-location \(x\) and \(b\) represents DC-offset in the range returns \(y\). To achieve scale (or gain) invariance, the normalized two-dimensional \((p, q)\)-th order moment invariants are estimated using HRR profiles as,
\[ M_{pq} = \frac{1}{N} \frac{\sum_{i=1}^{N} (x_i - \bar{x})^p (y_i - \bar{y})^q}{\left( \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})^2 \right)^{\frac{p}{2}} \left( \frac{1}{N} \sum_{i=1}^{N} (y_i - \bar{y})^2 \right)^{\frac{q}{2}}} \]  

(25)

where,

\( N \) = number of HRR returns

\( x_i \) = Range-location (relative or absolute)

\( y_i \) = Magnitude of range returns sampled at range location \( x_i \)

\( \bar{x} \) and \( \bar{y} \) = sample means of \( x_i \) and \( y_i \), respectively.

In this research, the moments are extracted from HRR profiles using equation (25). It has been observed that there is separation of the 4-target case for even values of \( p \) (see Figure 80 and Figure 81 below). Hence, only the MIs for even \( p \) values have been used in the implementation of the MI-based HRR-ATR.

### 6.2 Data used to conduct HRR-ATR study

For this research, it would be ideal to work with actual human tongue click data and corresponding echo returns. However, the data of human tongue clicks and echoes were not available at the time this work was initiated. Therefore, the MSTAR HRR data is used for this research.

The human tongue clicks and their echo are one-dimensional. The High Range Resolution (HRR) profiles are also one-dimensional. The fact that both human tongue click echoes and HRR profiles are one-dimensional representation of three-dimensional objects justifies our use of HRR profiles instead of actual human tongue click data to conduct the HRR-ATR study.
6.2.1 MSTAR Data

The MSTAR data used in this study was collected using the Sandia National Laboratories Twin Otter SAR sensor payload operating at X-band and it was distributed under the DARPA moving and stationary target recognition (MSTAR) program. Two sets of data were distributed under the MSTAR program: one for template formation and another for testing, collected at $17^\circ$ and $15^\circ$ depression angles, respectively. Figure 79 shows the HRR trajectory plot which displays range vs. aspect of a target as the target moves through 360 degrees of azimuth. Note that the profile length is shortest at broadside (90 and 270 degrees) and longest at the head-or-tail on (0, 180, 360 degrees). Brighter pixels represent stronger peaks in the profile.

![HRR Trajectory for BMP2, SN#C21 @ depression = 15](image)

Figure 79: Example HRR Trajectory Plot

6.3 Classification using Moment Invariants (MI)

For classification, the moment invariants of training data at $17^\circ$ depression angle are pre-calculated using all the HRR profiles in each aspect angle and stored. For testing the MIs are calculated using one or more HRR profiles at $15^\circ$ observed (test) profiles and compared with the trained MIs. The ATR system finds the distance between the MI of test target with the available stored training MIs within a range of azimuth.
angle. For example, the available MI of test target \( bmp2\_c21 \) with known azimuth angle of 35° will be compared with the MIs of template targets \( bmp2\_c21, t72\_132, 2s1\_b01, \) and \( brdm2\_e71 \) located at 35°±D°, where the value of D is typically ±5°. The template target class for which the MI values are closest (minimum absolute difference) to the test profile MI is selected as the detected target. In subsequent ATR performance evaluation and testing, the value D was varied from ±2° to ±180° and the results are given in a later section.

6.4 Results and observations on MI-based HRR-ATR

The moment invariants are found for the targets listed in Table 13.

<table>
<thead>
<tr>
<th>Train Target</th>
<th>Test Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>d17_t72_132</td>
<td>d15_t72_132</td>
</tr>
<tr>
<td>d17_bmp2_c21</td>
<td>d15_bmp2_c21</td>
</tr>
<tr>
<td>d17_2s1_b01</td>
<td>d15_2s1_b01</td>
</tr>
<tr>
<td>d17_brdm2_e71</td>
<td>d15_bdm2_e71</td>
</tr>
</tbody>
</table>

While calculating the moment invariants MI(p, q) using equation (25), it was found that even values of p provided distinct and separable moment values for different targets. For odd values of p, however, the moments of distinct targets were overlapping at different azimuth values, leading to possible conflicts in classification of targets. Hence, only the moments with even values of p are used in developing the MI-based ATR algorithm. Note also the MIs for different targets were found to be better separated for \( p \geq 8 \). Therefore, only the moments with values of \( 8 \leq p \leq 20 \) were considered to achieve higher target separability.
6.4.1 Choice of moment invariants for storing training features on look-up table

Typical examples of MIs for even p-values over entire azimuth angle span of 0 to 360 degrees for all four targets are given in Figure 80-83. They provide superimposed plots the MIs of test targets (d17s) and the corresponding train targets (d15s) at same aspect angles over the entire range of the azimuth angles (0 to 360 degrees). It can be seen that the MIs for the same target class match very closely while the MIs of different targets differ at each aspect.

It should be noted that for same p values, the shape of the MIs tend to get inverted for subsequent q values. After analyzing the data, it was found that for the same value of p, adding the absolute value of
certain values of q (1 and 2) provided further separability for the MIs. Therefore, in order to maximize target separability, the moment invariants for, p = 8, 10, 12, 16, 18, 20 and q = 1, 2, were used to create the test and template data sets. Figure 84 shows the MI(p, q) summed over p = 8, 10, 12, 14, 16, 18, 20 and q = 1, 2.

![MI average all](image)

**Figure 84**: Sum of MI(p,q) for p = 8, 10, 12, 14, 16, 18, 20 and q = 1, 2 exhibit target separability

### 6.4.2 Moment invariance based HRR-ATR performance study

Confusion matrices have been used as the performance metric for classification of typical 4-class MSTAR targets. Table 14 shows a typical confusion matrix where average MIs from 8 adjacent HRR test profiles were used for classification decision with ±5 degree of azimuth angle search range with the Look-Up table.
Table 14: Confusion matrix for classification using average MIs from 8 adjacent test profiles and ± 5 degree of azimuth angle search

<table>
<thead>
<tr>
<th></th>
<th>D17.bmp2_c21</th>
<th>D17_t72_132</th>
<th>D17_2s1_b01</th>
<th>D17_brdm2_e71</th>
</tr>
</thead>
<tbody>
<tr>
<td>D15_bmp2_c21</td>
<td>0.9592</td>
<td>0.0051</td>
<td>0.0204</td>
<td>0.0153</td>
</tr>
<tr>
<td>D15_t72_132</td>
<td>0.0153</td>
<td>0.9745</td>
<td>0</td>
<td>0.0102</td>
</tr>
<tr>
<td>D15_2s1_b01</td>
<td>0</td>
<td>0.0547</td>
<td>0.9307</td>
<td>0.0146</td>
</tr>
<tr>
<td>D15_brdm2_e71</td>
<td>0.0182</td>
<td>0.0146</td>
<td>0.0584</td>
<td>0.9088</td>
</tr>
</tbody>
</table>

From Table 14 it can be seen that the targets are correctly recognized with > 90% probability in all cases when 8 HRR test profiles are used to make the classification decision. The overall $P_{cc}$ obtained by averaging the diagonals of the confusion matrix is 94.3% in this case. Next, in order to study the effect of utilizing multiple perspectives, the number of HRR test profiles were varied from single-profile testing to up to 101 range profiles with a fixed ±5 degree of azimuth angle search, and the overall 4-target $P_{cc}$ (average of diagonals of the corresponding confusion matrices) are plotted in Figure 85.

![Figure 85: Probability of correct classification versus number of test profiles for moment invariant based ATR](image1)

![Figure 86: Probability of correct classification versus azimuth angle search range for moment invariant based ATR](image2)
It can be seen that relatively large $P_{cc}$ is maintained in all cases. Figure 85 also shows that even when a single test profile is used, the MI-based ATR system provides correct classification approximately 94% of the times. Also, the system performance improves almost consistently as the MIs from multiple perspectives of adjacent test profiles are used to perform classification.

Figure 86 plots the overall $P_{cc}$ vs. azimuth angle search range where in each case 8 profiles were used for making a classification decision but the azimuth search space was increased from ±2° to ±180°. Figure 86 indicates that the system performance is as high as 98% when the azimuth angle search range is 2 degrees while it reduces to ~58% for azimuth angle search range of 165°.

Based on this preliminary analysis for 4-target ATR using moment invariants, it can be concluded that the ATR performance improves as the number of test profiles increases and/or the azimuth angle search range decreases.

### 6.5 Performance Comparison with Baseline Eigen-Template-Based Approach

In order to verify the effectiveness of the proposed new MI-based HRR-ATR algorithm, we have performed apples-to-apples comparison with an existing Template-based approach [19] [34], using identical test and training target data set under identical operating scenarios. Specifically, the comparison has been performed with the classical Eigentemplate-based Matched Filter (ETMF) approach originally developed by one of the co-authors of this summer effort [19] [34].

#### 6.5.1. Eigen-Template Formulation

Let $Y$ denote the contiguous set of training profiles at an aspect angle $\alpha$, defined as $Y = [y_1, y_2, ..., y_M]$, where $M$ is the total number of range profiles in aspect $\alpha$. The Singular Value Decomposition (SVD) of a matrix of detected range profiles ($Y$) over a sector of aspect angles produces three matrices $U$, $\Lambda$, and $V$:

$$Y \xrightarrow{SVD} U\Lambda V^T = \sum_{i=1}^{M} \lambda_i u_i v_i^T$$  \hspace{1cm} (26)
where, \( \lambda_1 \) denotes the \( i \)-th eigenvalue, while \( u_i \) and \( v_i \) denote the left and right eigenvectors, respectively. The eigenvector \( (u_1) \) corresponding to the largest eigenvalue \( \lambda_1 \) is used as the feature template for each degree sector.

6.5.2 Matched-Filter based Classification

The recognition step for ETMF is based on the Matched filter technique [19] [34]. The decision determines the target type for which the normalized correlation between its template \( (m) \) and the given observation \( (a, \text{ or test}) \) profile is maximized among all template choices.

\[
\text{Accept } H_i \text{ for which } \sum_{k=1}^{N} a_k m_k = \mathbf{a}^T \mathbf{m}_i, \quad i = 1, \ldots, K \text{ is maximized}
\]

with constraint \( \mathbf{m}_i^T \mathbf{m}_i = \mathbf{a}^T \mathbf{a} = 1, \forall_1 \)

where, the eigen-templates \( \mathbf{m}_i \) are the dominant eigenvectors \( u_1 \) of HRR training profile matrices formed with \( 1^0 \) aspect angles for each target. More details can be found at [19] [34].

6.5.3 HRR-ATR Performance Comparison Results

In Figure 87, the number of HRR test profiles were varied from single-profile testing to up to 101 range profiles and the overall 4-target \( P_{cc} \) for the ETMF are plotted for \( \pm 5 \) degree of azimuth angle search. Comparing the MI-based results in Figure 85 with the ETMF-based results Figure 87, it can be seen that the performance of the MI-based approach improves from 94% to 96% as the number of test profile increases, whereas the performance of eigen-template based method stays \( \sim 93\% \) for all the cases. Hence, the MI-based HRR-ATR algorithm exhibits enhanced performance when multiple contiguous perspectives are used, i.e., when the number of test profiles is increased when compared with the template-matching based approach.
Figure 87: Probability of correct classification versus number of test profiles for template based ATR

Figure 88: Probability of correct classification versus azimuth angle search range for template based ATR

Figure 88 plots the overall $P_{cc}$ vs. azimuth angle search range where in each case 8 test-profiles were used for making a matched-filter based classification decision but the azimuth search space was increased from $\pm2^\circ$ to $\pm180^\circ$. Comparing Figure 86 with Figure 88, it can be seen that the performance of our system has not degraded much despite using very small data set for templates. Figure 86 indicates that the system performance is as high as 93.42% (compared to 98% in our system) when the azimuth angle search range is 2 degrees while it reduces to 81% (compared to ~58% in our system) for azimuth angle search range of 172°.

Based on the preliminary work, the MI-based HRR-ATR approach appears to perform slightly better than ETMF in case of smaller search angle range, whereas it tends to perform worse for higher search angle range. This may be considered one of the trade-offs required to limit the trained feature information to be stored to only 360 composite MI values at 360 aspect angles, as opposed to the raw HRR data to be stored in case of template based ATR method. The reduced data storage is an essential feature required to make the system hardware more efficient. It should be noted here that in case of the MI based system, the MI’s for test (observation) profiles need to be calculated using equation (25), to make relatively quick ATR decisions using lookup-table method. Further computational requirement study needs to be undertaken to
explore the feasibility of implementing MI-based HRR-ATR in hardware using table-lookup to make the proposed MI-based system practically useful.

6.6 Conclusion

The results indicate that the MI based ATR method provides very good results with the probability of correct classification of up to 98% in certain cases. The performance improves when multiple perspectives, i.e., increased number of test profiles and smaller azimuth angle search range are used. The performance has been compared with a baseline template-based approach. The performance comparison indicates that the MI-based HRR-ATR approach appears to perform slightly better than ETMF in case of smaller search angle range, whereas it tends to perform worse for higher search angle range. This may be considered one of the trade-offs required to limit the trained feature information to be stored as opposed to the raw HRR data to be stored in case of template based ATR method.

The proposed MI-based approach uses a relatively small number of higher-order moment features, which leads to reduced data storage requirements. Thus the MI-based approach should have reduced storage requirements than the template-based approach that relies on the entire range-template profiles in the training database. It should be noted however that in case of the MI based system, the high-order MI's for test (observation) profiles need to be calculated to perform classification. Further computational requirement study needs to be undertaken to explore the feasibility of implementing MI-based HRR-ATR in hardware using table-lookup to make the proposed MI-based system practically useful.
7 ACCOMPLISHMENTS, CONCLUSION AND FUTURE WORK

7.1 Accomplishments

The following list document the principal accomplishments from the research reported on this manuscript.

1. A human tongue click waveform was analyzed in time and frequency domain. It was found from the spectrogram of the human tongue click waveform that there are four frequency components located at 1.54 kHz, 3.33 kHz, 4.62 kHz and 11.03 kHz. The ambiguity function of human tongue click was analyzed to understand it better. While observing the zero-delay and zero-velocity cuts of the ambiguity function, it was found that -3 dB bandwidth corresponds to 3.08 kHz bandwidth, which raised the questions about significance of 11.03 kHz component in human tongue click signal.

2. To evaluate the significance of 11.03 kHz component in human tongue click signal, a synthetic version of human tongue click was created from which individual components could be removed. Analysis of the wideband ambiguity function revealed the -3dB range resolution was unaffected by omitting the 11.03 kHz component. As such it was not regarded as significant in the case of stationary target detection.

3. An acoustic radar system was developed using National Instrument kit which had a maximum unambiguous range of 16.3 m. The system was capable of transmitting the synthetic tongue click and LFM chirp signals. As a part of the system calibration, the acoustic baffles were used to reduce strong returns from pillars in laboratory and the floor between acoustic radar and target. The baffles were successfully able to reduce the noise
by approximately 2-3 dB. They were also used to reduce the amount of cross-talk between the transmitter and receiver. In order to eliminate the constant cross-talk and environment, a method of background subtraction was utilized which provided very good results. The cross-talk and environment were completely eliminated using this technique. Moreover, 1024 pulses were integrated for the received signals, which lead to the SNR gain of approximately 30 dB.

4. The backscatter signals from diverse set of targets containing different shapes, size and material were measured by LFM chirp and synthetic tongue click signal and the data were recorded.

5. The collected dataset was analyzed by creating different features. These features are normalized target signatures, normalized frequency spectrum of echoes, absolute value of echoes, and the absolute different between normalized frequency spectrum of transmit signal and normalized frequency spectrum of echo. Then absolute value of correlation coefficients was found for the same feature of different targets to understand how correlated they are and whether they can be separated easily based on these features.

6. From all the features analyzed, the absolute difference between normalized frequency spectrum of transmit signal and normalized frequency spectrum of echo provided very good results.

7. Then, a MI-based ATR system was developed using MSTAR dataset and its performance was analyzed using confusion matrices as a performance metric for classification.

7.2 Conclusion

From the results presented here we may conclude that the human tongue click signal is composed of three main frequency components located at 1.54 kHz, 3.33 kHz and 4.62 kHz and has a bandwidth of 3.08 kHz and it provides a range resolution of 5.5 cm. The component at 11.03 kHz does not have any significant impact on the range resolution capacity of the human tongue click signal.

From the analyses of data collected for various targets and their extracted features, it can be concluded that normalized target signatures cannot provide target classification in efficient manner. The normalized frequency spectrum has some potential for target classification, but it does not lead to confident
classification results. The absolute difference between normalized frequency spectrum of transmit signal and normalized frequency spectrum of echoes performs much better than the two features discussed previously. It should be noted that the tongue click waveform performs much better at classifying objects made of hard materials from objects made of soft materials. However, they cannot be classified based on their shape or size by utilizing this feature. The chirp waveform provides superior classification performance for this feature, however, it is unclear which broad categories the targets can be put in for classification. The chirp, certainly, cannot classify all the targets as distinct from each other using this feature. From this analysis, it can be concluded that the frequency spectrum of echoes contain the most useful information for the task of target classification based on its material.

For the MI-based ATR, the performance of MI-based ATR system has not degraded much despite using very small data set for templates. Figure 86 indicates that the system performance is as high as 98% when the azimuth angle search range is 2 degrees while it reduces to ~58% for azimuth angle search range of 172°.

The MI-based ATR system performs better in the cases of smaller search angle range, whereas it performs worse for higher search angle range. This is one of the trade-offs required to make in order to reduce the template dataset to only 360 points, which are calculated MIs. The reduced dataset is an essential feature required to make the system hardware implementable. It is also possible to make quick decisions using binary lookup-table method for small dataset. The fact that it can be implementable in hardware using table-lookup method makes the MI-based system very important in battlefield.

7.3 Future work

In future, it would be beneficial to analyze the data and identify any features that can lead to target classification based on target’s size or reflectivity. Moreover, the human tongue click signal should be analyzed for moving targets since this work only focused on stationary targets. It would be beneficial to analyze the echoes received in two receivers while observing a moving target or approaching a stationary
target, as opposed to only one receiver in this research. As the literature studies revealed that the humans approaching target do hear the change in pitch of sound which can lead to better target classification.

It is also essential to combine the features used in this research with any new features to verify if a better classification result can be obtained. This is suggested because, knowingly or unknowingly, human successfully brain process all of these information together while they are performing echolocation.

In future, if capable hardware is available, it would be useful to perform the same tests using bat click waveforms, since they are very similar in structure to human tongue click waveforms. The comparison and contrasts of the features observed by transmitting both signals may lead to valuable information for the purpose of target classification.

In future, after creating a similar dataset to MSTAR, the MI-based ATR can also be implemented using human tongue click waveforms and bat click waveforms, and their performance can be compared to the system developed in this research.

Finally, due to the multidisciplinary nature of this work, it is advisable to advance this research in collaboration with the experts of cognitive science and speech processing. It would also be advisable to investigate how speech processing concepts can be applied to the dataset created using human tongue click waveforms for the purpose of target classification.
APPENDIX A
EXPERIMENTAL AND LABORATORY SETUP

Figure 89: Speaker (transmitter) and microphone (receiver) acoustic radar system
Figure 90: Metal sphere on a baffle stand

Figure 91: Metal sphere facing acoustic radar system
Figure 92: Plastic box on a baffle stand

Figure 93: Baffles on the floor between acoustic radar and target
Figure 94: Baffles around pillars
Figure 95: Baffles around pillars
BIBLIOGRAPHY


