An analysis of blind signal separation for real time application

Daniel Smith
University of Wollongong

UNIVERSITY OF WOLLONGONG
COPYRIGHT WARNING

You may print or download ONE copy of this document for the purpose of your own research or study. The University does not authorise you to copy, communicate or otherwise make available electronically to any other person any copyright material contained on this site. You are reminded of the following:

This work is copyright. Apart from any use permitted under the Copyright Act 1968, no part of this work may be reproduced by any process, nor may any other exclusive right be exercised, without the permission of the author.

Copyright owners are entitled to take legal action against persons who infringe their copyright. A reproduction of material that is protected by copyright may be a copyright infringement. A court may impose penalties and award damages in relation to offences and infringements relating to copyright material. Higher penalties may apply, and higher damages may be awarded, for offences and infringements involving the conversion of material into digital or electronic form.

Unless otherwise indicated, the views expressed in this thesis are those of the author and do not necessarily represent the views of the University of Wollongong.

Recommended Citation

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: research-pubs@uow.edu.au
An Analysis of Blind Signal Separation for Real Time Application

A thesis submitted in fulfilment of the requirements for the award of the degree

Doctor of Philosophy

from

THE UNIVERSITY OF WOLLONGONG

by

Daniel Smith
Bachelor of Engineering (Honours Class I)
University of Wollongong, 2001

SCHOOL OF ELECTRICAL, COMPUTER AND TELECOMMUNICATIONS ENGINEERING
2006
Abstract

The ‘cocktail party problem’ is the term commonly used to describe the perceptual problem experienced by a listener who attempts to focus upon a single speaker in a scene of interfering audio and noise sources. Blind Signal Separation (BSS) is a blind identification approach that can offer an adaptive, intelligent solution to the ‘cocktail party problem’. Audio signals can be blindly retrieved from the mixture, that is, without a priori knowledge of the audio signals or the location of the audio sources and sensors. Hence, BSS exhibits greater flexibility than other identification approaches, such as adaptive beamforming, which require precise knowledge of the sensors and/or signal locations.

Speech enhancement is a potential application of BSS. In particular, BSS is potentially useful for the enhancement of speech in interactive voice technologies. However, interactive voice technologies, such as mobile telephony or teleconferencing, require real time processing (on a frame-by-frame basis), as longer processing delays are considered intolerable for the participants of the two-way communication. Hence, BSS applications with interactive voice technologies require real-time operation of the algorithm.
BSS primarily employs Independent Component Analysis (ICA) as the criteria to separate speech signals. Separation is achieved with ICA when statistical independence between the signal estimates is established. However, investigations in this Thesis, that study the relationship between the ICA criteria and speech signals indicate that significant statistical dependencies can exist between short frames of speech. Hence, it was found that the ICA criteria could be unreliable for real-time speech separation.

This Thesis proposes a number of BSS algorithms that improve real-time separation performance in acoustic environments. In addition, these algorithms are shown to be better equipped to handle the dynamic nature of acoustic environments that contain moving speakers. The algorithms exhibit higher data efficiency, that is, these approaches accurately separate the acoustic scene with smaller amounts of data. The higher data efficiency is the result of BSS models that better represent the underlying characteristics of audio, and in particular speech in the mixture.

Sparse Component Analysis (SCA) algorithms are proposed to exploit the sparse representation of audio in the time-frequency (t-f) domain. Conventional SCA approaches generally place strong constraints upon signals, requiring them to be highly sparse across their entire t-f representation. This constraint is not always satisfied by broadband audio, particularly speech, and hence separation performance is reduced. The SCA algorithms developed in this Thesis relax this constraint, such that signals can be estimated from sparse sub-regions of the t-f representation rather than the complete t-f representation. A SCA algorithm that employs K-means clustering of
the t-f space is proposed in order to improve the accuracy of estimation. In addition, an exponential averaging function is used to reduce the influence of poor estimates when separation is performed on a frame by frame basis.

Sequential approaches to SCA are proposed in this Thesis where only a sparse sub-region of one signal in the mixture is required for estimation at one time. This relaxes the sparsity constraints that are placed upon broadband signals in the mixture.

A BSS algorithm that jointly models the production mechanisms of speech (pitch and spectral envelope) is also presented in this Thesis. This produces a more accurate model of speech than existing algorithms that individually model the pitch or spectral envelope. An investigation of this algorithm then determines the parameter set that optimally models the underlying speech signals in the mixture.

Finally, an algorithm is proposed to exploit both the sparse t-f representation of audio and the joint model of speech production. This unified approach compares the SCA and speech production mechanism criteria, switching to the criteria that provides the most accurate estimate. Results indicate that this unified algorithm offers a superior data efficiency to its constituent algorithms, and to three benchmark ICA algorithms.
Statement of Originality

This is to certify that the work described in this thesis is entirely my own, except where due reference is made in the text.

No work in this thesis has been submitted for a degree to any other university or institution.

Signed

Daniel Vaughan Smith

April, 2007
Acknowledgments

Firstly, I would like to thank my supervisors, Dr. Jason Lukasiak and Dr. Ian Burnett, for their guidance and support throughout the course of my research. I would also like to thank my fellow colleagues in the Whisper Laboratories for creating a relaxed, friendly atmosphere to work in. In particular, I would like to thank Ms Eva Cheng for proof reading my Thesis.

More personally, I would like to thank my family and friends for allowing me to maintain a balanced lifestyle and showing interest in my research, despite their claims about having no idea what I was talking about. Finally, I would like to thank my parents for their support and encouragement as I pursued this path of higher learning.
# Contents

1 Introduction 1
   1.1 Blind Signal Separation ................................. 1
   1.2 Motivation for BSS in an Acoustic Environment ........ 2
   1.3 Thesis Outline ........................................ 5
   1.4 Contributions ......................................... 7
   1.5 Publications .......................................... 10
      1.5.1 Journal Publications .............................. 10
      1.5.2 Book Chapter .................................. 10
      1.5.3 Conference Publications ......................... 10

2 Literature Review 12
   2.1 Introduction .......................................... 12
   2.2 General BSS Framework ................................. 13
      2.2.1 Structure of the BSS Algorithm ................. 15
      2.2.2 Ambiguities of BSS ................................. 17
   2.3 Extensions of the BSS Framework for Audio .......... 18
      2.3.1 Propagation Models in an Audio Environment . 18
      2.3.2 BSS in a Convolutive Mixing Environment .... 20
      2.3.3 The Dynamic Nature of an Audio Environment 26
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.4</td>
<td>The Separation Criterion of BSS</td>
<td>29</td>
</tr>
<tr>
<td>2.4.1</td>
<td>Whitening</td>
<td>30</td>
</tr>
<tr>
<td>2.5</td>
<td>Independent Component Analysis</td>
<td>31</td>
</tr>
<tr>
<td>2.5.1</td>
<td>Statistical Independence</td>
<td>32</td>
</tr>
<tr>
<td>2.5.2</td>
<td>Information Theory Connection to ICA</td>
<td>36</td>
</tr>
<tr>
<td>2.5.3</td>
<td>Maximum Likelihood</td>
<td>37</td>
</tr>
<tr>
<td>2.5.4</td>
<td>Information Maximisation</td>
<td>39</td>
</tr>
<tr>
<td>2.5.5</td>
<td>Mutual Information</td>
<td>40</td>
</tr>
<tr>
<td>2.5.6</td>
<td>Non-Gaussian Maximisation</td>
<td>41</td>
</tr>
<tr>
<td>2.5.7</td>
<td>Higher Order Approximations</td>
<td>44</td>
</tr>
<tr>
<td>2.5.8</td>
<td>Limitations of ICA Separation</td>
<td>46</td>
</tr>
<tr>
<td>2.6</td>
<td>Temporal BSS</td>
<td>47</td>
</tr>
<tr>
<td>2.6.1</td>
<td>Temporal Correlation</td>
<td>49</td>
</tr>
<tr>
<td>2.6.2</td>
<td>Sequential Separation with Linear Prediction</td>
<td>53</td>
</tr>
<tr>
<td>2.6.3</td>
<td>A Set of Non-Stationary Statistics</td>
<td>58</td>
</tr>
<tr>
<td>2.6.4</td>
<td>Unification of the Temporal Approaches</td>
<td>62</td>
</tr>
<tr>
<td>2.7</td>
<td>Sparse Component Analysis</td>
<td>64</td>
</tr>
<tr>
<td>2.7.1</td>
<td>Preprocessing in SCA</td>
<td>67</td>
</tr>
<tr>
<td>2.7.2</td>
<td>Estimation of the Mixing System</td>
<td>69</td>
</tr>
<tr>
<td>2.7.3</td>
<td>Retrieving Signals from the Mixture</td>
<td>79</td>
</tr>
<tr>
<td>2.7.4</td>
<td>Limitations of SCA Separation</td>
<td>82</td>
</tr>
<tr>
<td>2.8</td>
<td>Combining Different Separation Criteria</td>
<td>83</td>
</tr>
<tr>
<td>2.9</td>
<td>Performance Measures</td>
<td>85</td>
</tr>
<tr>
<td>2.9.1</td>
<td>Interference Measure</td>
<td>86</td>
</tr>
<tr>
<td>2.9.2</td>
<td>Signal to Noise Ratio</td>
<td>87</td>
</tr>
</tbody>
</table>
## Contents

2.10 Limitations of Current BSS Research in Audio Environment .............................................. 87

### 3 Limitations of Independent Component Analysis for Real Time Separation of Speech

3.1 Introduction ......................................................................................................................... 91
3.2 Mutual Information ............................................................................................................. 94
3.3 Analysis of the Relationship between Statistical Independence and Speech ................................. 95
  3.3.1 MI Analysis Data Set ..................................................................................................... 95
  3.3.2 MI - Frame Size Relationship for Signal Classes ........................................................... 97
  3.3.3 Deterministic and Harmonic Speech Signal Effects on MI .......................................... 98
  3.3.4 Influence of the Speech Production Model on MI ......................................................... 102
3.4 ICA Application with Speech in Relation to Frame Size ...................................................... 106
3.5 Conclusion ......................................................................................................................... 109

### 4 Block Adaptive Algorithms using Sparse Component Analysis

4.1 Introduction ......................................................................................................................... 111
4.2 TIFROM and TIFCORR Estimation .................................................................................... 114
  4.2.1 TIFROM Estimation ..................................................................................................... 114
  4.2.2 TIFCORR Estimation .................................................................................................. 116
4.3 Limitations of TIFROM and TIFCORR Estimation ........................................................... 119
  4.3.1 Bias Caused by the Variance Measure in TIFROM Estimation ..................................... 119
  4.3.2 Bias Caused by the Fluctuation of Signal Sparsity ...................................................... 121
4.4 Outline of the K-Means Modified Architecture for TIFROM and TIFCORR Estimation ........... 124
4.5 Experiments with the K-means Modified Algorithm ............................................................ 126
  4.5.1 Experimental Setup ..................................................................................................... 126
  4.5.2 Discussion of the Results for the K-means Modified Algorithm ................................. 129
4.6 Adaptive Block Based Architecture .......................... 136
4.7 Experiment with the Block Adaptive Algorithm ............... 138
  4.7.1 Experimental Setup for the Time-Varying Mixtures ....... 139
  4.7.2 Discussion of the Results for the Block Adaptive Algorithm 140
4.8 A Comparison of the Variance and Correlation Based Algorithms . 145
  4.8.1 Comparison with the Stationary Mixing Systems ........... 145
  4.8.2 Comparison with the Time-Varying Mixtures ............... 147
4.9 Conclusion ...................................................... 149

5 Blind Signal Separation using a Joint Model Of Speech Production 152
  5.1 Introduction .................................................. 152
  5.2 Blind Signal Extraction Problem ............................. 154
  5.3 Speech Production Mechanisms ............................... 154
  5.4 Separation of Speech Signals ................................ 157
  5.5 Derivation of the Learning Algorithms ....................... 160
    5.5.1 Preprocessing of the Mixture ......................... 161
    5.5.2 Calculation of the Fundamental Frequency ............ 162
  5.6 Outline of the AR-F0 Algorithm ............................ 163
  5.7 Results of the AR-F0 Algorithm .............................. 164
    5.7.1 Experimental Setup ................................. 164
    5.7.2 Experiments with Voiced Speech ....................... 166
    5.7.3 Experiments with Unvoiced Speech ..................... 169
    5.7.4 Experiments with Natural Speech ...................... 171
  5.8 Investigation of Temporal Modeling ......................... 173
    5.8.1 Analysis Data Set ................................. 174
    5.8.2 Investigation with Artificial Voiced-Unvoiced Speech . 176
5.8.3 Investigation with Natural Speech .................................. 179
5.9 Conclusion ........................................................................... 181

6 Sequential Approaches to Blind Signal Separation .......................... 183

6.1 Introduction ........................................................................... 183
6.2 Formulation of a Sequential BSS Problem .............................. 186
6.3 Sequential SCA Approach ....................................................... 187
   6.3.1 The Source Cancellation Approach .................................. 187
   6.3.2 The Deflation Technique ................................................. 188
6.4 Outline of the Sequential Algorithm ........................................ 190
   6.4.1 A Related Sequential SCA Approach ............................... 191
6.5 Results of the Sequential and Simultaneous Algorithm Analysis . . 193
   6.5.1 Experiments with the Stationary Mixing Systems .......... 193
   6.5.2 Experiments with the Time-Varying Mixing Systems .... 199
6.6 Comparison of the Variance and Correlation Based Sequential Approaches ................................................. 203
   6.7 A Switched Approach to Combine Separation Criteria ........ 206
      6.7.1 Switching between the SCA and Temporal Criteria .... 207
      6.7.2 Outline of the Switched Algorithm ............................... 209
      6.7.3 Results of the Switched Algorithm ............................... 210
      6.7.4 Experimental Setup ................................................... 211
      6.7.5 A Comparison with the SCA and Temporal Algorithms ... 212
      6.7.6 A Comparison with the Benchmark Algorithms .......... 214
6.8 Conclusion ........................................................................... 216

7 Conclusions and Suggestions for Future Work ................................. 218

7.1 Overview .............................................................................. 218
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.2 An Analysis of ICA for Real Time Operation with Speech</td>
<td>220</td>
</tr>
<tr>
<td>7.3 Modified SCA Approaches that Improve the Separation Performance</td>
<td>221</td>
</tr>
<tr>
<td>of the TIFROM and TIFCORR Algorithms</td>
<td></td>
</tr>
<tr>
<td>7.4 A Sequential Approach to SCA that Improves the Separation Performance of Simultaneous SCA Algorithms</td>
<td>223</td>
</tr>
<tr>
<td>7.5 Improved Modeling of the Temporal Structure of Speech</td>
<td>225</td>
</tr>
<tr>
<td>7.5.1 A Joint Model of the Production Mechanisms of Speech</td>
<td>225</td>
</tr>
<tr>
<td>7.5.2 An Analysis of AR Modeling for Temporal Algorithms Separating Speech Mixtures</td>
<td>226</td>
</tr>
<tr>
<td>7.6 A Combined Framework of Different Separation Criteria that improves the Data Efficiency of Single Criteria Algorithms</td>
<td>227</td>
</tr>
<tr>
<td>7.7 Future Work</td>
<td>228</td>
</tr>
<tr>
<td>7.7.1 Simulation with more Extensive Data Sets</td>
<td>228</td>
</tr>
<tr>
<td>7.7.2 Extensions to Accommodate Convolutive Mixtures</td>
<td>229</td>
</tr>
<tr>
<td>7.7.3 Constraints of the System</td>
<td>232</td>
</tr>
<tr>
<td>7.7.4 Under-determined Systems</td>
<td>233</td>
</tr>
<tr>
<td>Bibliography</td>
<td>236</td>
</tr>
<tr>
<td>A The Complete Set of Separation Results for the SCA Algorithms in</td>
<td>259</td>
</tr>
<tr>
<td>Chapter 4</td>
<td></td>
</tr>
</tbody>
</table>
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>General formulation of the BSS problem.</td>
<td>14</td>
</tr>
<tr>
<td>2.2</td>
<td>The BSS algorithm consists of three main components; the demixing system $W$, separation criterion and learning algorithm [6].</td>
<td>16</td>
</tr>
<tr>
<td>2.3</td>
<td>Two realistic models for mixing in an acoustic environment [29]. In an anechoic model (a), sources are observed at sensors with different intensities and arrival times. In an echoic model (b), sources are observed at sensors with different intensities, arrival times and multiple arrival paths.</td>
<td>21</td>
</tr>
<tr>
<td>2.4</td>
<td>The Frequency Domain approach to BSS [45]. In each of the $T$ frequency channels, an instantaneous BSS algorithm is independently employed. After separation, the permutation inconsistencies across the $T$ independent BSS problems can result in signals being incorrectly formed from the frequency components.</td>
<td>23</td>
</tr>
<tr>
<td>2.5</td>
<td>The joint pdf of a pair of statistically dependent signals. This signal pair comprises of a sine wave of 1Hz and a sine wave of 2Hz. When the value of one signal is given, the value of the other signal belongs to a limited set of 2-4 values.</td>
<td>33</td>
</tr>
<tr>
<td>2.6</td>
<td>The joint pdf of a pair of statistically independent signals. The pair of signals include a sine wave of 1Hz and a uniform distribution of noise with a range of -1 and 1. When the value of one signal is given, the other signal can be any value within its range of -1 and 1.</td>
<td>34</td>
</tr>
<tr>
<td>2.7</td>
<td>A comparison of super-Gaussian, sub-Gaussian and Gaussian pdfs. The super-Gaussian and sub-Gaussian pdf shapes are commonly used to identify separated signals in ICA approaches. A Gaussian shape generally indicates signals are still mixed in ICA.</td>
<td>37</td>
</tr>
</tbody>
</table>
2.8 Linear Prediction can be employed to separate temporally correlated signals from the mixture. The separation column $W_i$ can be obtained by minimising the M.S.E between the estimated signal and the predicted estimated signal. ........................................... 54

2.9 BSS algorithms that exploit the non-stationary structure of signals, must ensure that a unique set of second order statistics are obtained for each frame across time. These frames correspond to the light coloured segments of the mixed speech observations. A covariance matrix $R_{x_1 x_2}$ is then computed between the mixed channels for each of the frames. The separation matrix $W$ is estimated by the JAD of the set of covariance matrices. ........................................... 61

2.10 Two channels of the mixture are plotted against one another. When the pair of signals in the mixture are sparse, with only 20 non-zero values, the plot points have a clear orientation in the two straight lines shown. The gradient of each of these straight lines corresponds to the mixing column ratio of a source. ........................................... 66

2.11 The structure of the DPWT where each level of the tree represents a different time-resolution of the wavelet transform with scale $j$ and shift $k$ parameters, and additionally, a number of nodes representing the different frequency sub bands $n$ [123]. ........................................... 78

2.12 Binary t-f masks can be used to retrieve signals from a t-f representation of the mixture. When signals are non-overlapping in the t-f domain, the frequency components belonging to a specific signal can be passed, while all other frequency components can be blocked by the mask. The binary mask determines whether a frequency component should be passed or blocked by comparing its attenuation and delay parameters with the parameters of other frequency components. 81

3.1 Average Mutual Information estimated for speech and Gaussian classes for frame sizes ranging from 20ms to 0.5s. ......................... 97

3.2 Average Mutual Information estimated for harmonic artificial vowels, harmonic natural vowels and the entire class of natural vowels for frame sizes 20ms-0.5s. ........................................... 99

3.3 Joint pdf of two artificial vowels with a harmonic pitch relationship of 242.42Hz and 121.21Hz. ........................................... 100
3.4 Mutual Information estimated between all combinations of frames belonging to two 1s sections of speech signals, Speaker 1 and Speaker 2, for frame sizes of 200ms (Figure 3.4(a)), 80ms (Figure 3.4(b)) and 20ms (Figure 3.4(c)). In Figure 3.4(c), label i corresponds to the unvoiced frames of Speaker 1 and Speaker 2. Label ii refers to frames of voiced speech between Speaker 1 and Speaker 2, while label iii corresponds to voiced frames that have formed harmonic pitch relationships.

3.5 The 1s sections of Speaker 1 (a) and Speaker 2 (b) which were used in the MI analysis in Figure 3.4. The labels i, ii, iii are the regions of the speakers corresponding to the MI sections in Figure 3.4(c). Label i corresponds to the unvoiced portions of Speaker 1 and Speaker 2. Label ii refers to the voiced portions of Speaker 1 and Speaker 2, while label iii refers to the voiced sections that form harmonic pitch relationships.

3.6 The average IM obtained by applying JADE and FastICA to the set of speech signals and Laplacian data for frame sizes 20ms to 5s.

4.1 The procedure for estimating a mixing column $C_{ie}$ using the TIFROM algorithm.

4.2 TIFROM estimation space in terms of the variance and mean of series ($\Upsilon_{u}, k$). A mixing column is estimated from each cluster, where $C_{1e} = 0.5$ and $C_{2e} = 0.62$. The dotted lines correspond to the true mixing columns of 0.5 and 1.

4.3 TIFROM estimation space when K-means clustering is conducted across the mean of the series. When a mixing column is estimated from each cluster, $C_{1e} = 0.5$ and $C_{2e} = 1.11$. The dotted lines correspond to the true mixing columns of 0.5 and 1.

4.4 rectangular, Hanning and Hamming windows of 160 samples were used in the analysis.

4.5 The separation performance $IM$ was compared across the rectangular (1), Hanning (2) and Hamming (3) windows for the TIFmod and TIFCmod algorithms. The separation performance was averaged across all 144 trials, $seriesnum = \{1...180\}$ and $fps = 4, 6, 8$.

4.6 The separation performance $IM$ was compared across $fps = \{2, 4, 6, 8\}$ for the TIFmod and TIFCmod algorithms. The separation performance was averaged across all 144 trials, $seriesnum = \{1...180\}$ and three windows.
4.7 The separation performance $IM$ (averaged across all 144 trials and three windows) was compared across all $seriesnum$ for the variance and correlation based algorithms for $fps = 6$. The original algorithms (TIFROM and TIFCORR), modified K-means algorithms (TIFmod and TIFCmod) and the block adaptive algorithms (adTIFmod and adTIFCmod).

4.8 The physical path of the acoustic environment in which the mixing system $A1$ was generated. Both speakers moved in a circular path at constant velocities of $2ms^{-1}$ and $4ms^{-1}$, respectively. $x1$ and $x2$ correspond to the two sensors.

4.9 The separation performance ($IM$) of the variance and correlation based algorithms were compared between the original (TIFROM and TIFCORR) and block adaptive algorithms (adTIFmod and adTIFCmod). The experiments were averaged across 144 trials and the two window types when $fps = 6$.

4.10 The $A1$ mixing system tracked by the TIFROM (a) and adTIFmod (b) algorithms.

4.11 The $A1$ mixing system tracked by the TIFCORR (a) and adTIFCmod (b) algorithms.

5.1 A section of voiced speech is shown in the time domain in subplot (a). In subplot (b), the spectrum of the voiced speech segment is shown.

5.2 A section of unvoiced speech is shown in the time domain in subplot (a). In subplot (b), the spectrum of the unvoiced speech segment is shown.

5.3 The joint AR-F0 algorithm separates speech by learning $W_j$ that optimally predicts the short term and long term temporal structure of speech.

5.4 The MMSE and separation performance $IM$ (subplot (a) and (b) respectively) of the joint AR-F0, AR and F0 models, averaged over 8 pairs of sustained vowels and 3 mixing simulations (24 mixed pair trials). In each simulation, the sustained vowels where mixed by a different mixing system $A$.

5.5 The MMSE and separation performance $IM$ (subplot (a) and (b) respectively) of the joint AR-F0, AR and F0 models, averaged over 8 pairs of fricatives and 3 mixing simulations (24 mixed pair trials). In each simulation, the fricatives where mixed by a different mixing system $A$. 

135

139

141

143

144

155

156

159

167

169
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.6</td>
<td>The MMSE and separation performance $IM$ (subplot (a) and (b) respectively) of the joint AR-F0, AR and F0 models, averaged over 10 pairs of natural speech and 3 mixing simulations (30 mixed pair trials). In each simulation, the natural speech was mixed by a different mixing system $A$.</td>
<td>171</td>
</tr>
<tr>
<td>5.7</td>
<td>Average $IM$ across 15 mixed pairs of artificial unvoiced speech. Prediction order ranged from 1-50.</td>
<td>177</td>
</tr>
<tr>
<td>5.8</td>
<td>Average $IM$ across 15 mixed pairs of artificial voiced speech. Prediction order ranges from 1-133.</td>
<td>178</td>
</tr>
<tr>
<td>5.9</td>
<td>Average $IM$ across 15 mixed pairs of natural speech. Prediction order ranges from 1-133.</td>
<td>179</td>
</tr>
<tr>
<td>6.1</td>
<td>The structure of the sequential SeqTIF and SeqCOR algorithms. The mixing column of signals are estimated and the contribution of each signal is cancelled from the mixture, until only one signal remains. This retrieved signal is then deflated from the mixture. This process is repeated until all signals are retrieved.</td>
<td>192</td>
</tr>
<tr>
<td>6.2</td>
<td>The average SNR of the SeqTIF and TIFROM algorithms across 40 different trials (mixtures), where each mixture consists of three speech signals. The analysis is conducted across $fps = 6, 8$ and $seriesnum = {1...180}$.</td>
<td>197</td>
</tr>
<tr>
<td>6.3</td>
<td>The average SNR of the SeqCOR and TIFCORR algorithms across 40 different trials (mixtures), where each mixture consists of three speech signals. The analysis is conducted across $fps = 6$ and $fps = 8$, and $seriesnum = {1...180}$.</td>
<td>198</td>
</tr>
<tr>
<td>6.4</td>
<td>The physical path of the acoustic environment in which the $A2$ mixing system was generated. The first two speakers moved in a circular path at constant velocities of $0.85ms^{-1}$ and $1.5ms^{-1}$. The third speaker moved in a straight line at a constant velocity of $2ms^{-1}$. $x_1$, $x_2$ and $x_3$ correspond to the sensors.</td>
<td>200</td>
</tr>
<tr>
<td>6.5</td>
<td>The average SNR of the SeqTIF and TIFROM algorithms across ten time-varying mixtures of speech for $fps = 6$ and 8.</td>
<td>201</td>
</tr>
<tr>
<td>6.6</td>
<td>The average SNR of the SeqCOR and CORTIFF algorithms across ten time-varying mixtures of speech for $fps = 6$ and 8.</td>
<td>202</td>
</tr>
</tbody>
</table>
6.7 The structure of the sequential heuristic algorithm which switches between the SeqTIF and joint AR-F0 criteria. The switching is based upon a comparison of each criteria’s estimation quality, that is, comparing the variance of the SeqTIF estimates and MMSE of the AR-F0 estimates. ........................................... 205

6.8 A comparison of the separation performance ($SNR$) of the SCAtemp, SeqTIF and AR-F0 algorithms proposed in this Thesis, along with the benchmark FastICA, Extended Infomax and TIFROM algorithms for block sizes spanning from 70ms to 0.56s. The experimental set consisted of 10 mixtures each consisting of three different speech signals. The mixtures changed every 125ms, as shown by the dotted vertical line. ............................................................ 213

7.1 A sub band approach to AR-F0 separation, where mixtures are decomposed using an analysis filter bank and the AR-F0 algorithm is independently applied to each sub band. A synthesis filter bank is then used to recover the full band separated signals. ......................... 229

A.1 The average separation performance $IM$ of the TIFROM, TIFmod and adTIFmod algorithms across 144 trials with pairs of audio signals. 260

A.2 The average separation performance $IM$ of the TIFCORR, TIFCmod and adTIFCmod algorithms across 144 trails with pairs of audio signals. ............................................................ 261

A.3 The average separation performance $IM$ of the TIFROM and adTIFmod algorithms across a time varying mixture (updated every 90ms) and 6 pairs of audio signals. ................................................ 262

A.4 The average separation performance $IM$ of the TIFCORR and adTIFCmod algorithms across a time varying mixture (updated every 90ms) and 6 pairs of audio signals. ................................. 262
List of Tables

4.1 The parameters used for the experiment in Section 4.5 between TIFROM, TIFCORR and their modified TIFmod and TIFCmod algorithms. .............................................. 128

4.2 The parameters used for the experiment in Section 4.7 between TIFROM, TIFCORR and their modified adTIFmod and adTIFCmod algorithms. .............................................. 140

4.3 A comparison of the average $IM$ of the variance and correlation based algorithms for stationary mixtures across $fps = 4, 6, 8$, three windows and $seriesnum = \{1...180\}$. ......................... 146

4.4 A comparison of the average $IM$ of the variance and correlation based algorithms for time-varying mixtures across $fps = 6, 8$, two windows and $seriesnum = \{1...180\}$. ......................... 148

6.1 The parameters used for the experiment in Section 6.5.1 between TIFROM, TIFCORR and the modified sequential algorithms SeqTIF and SeqCOR. .............................. 194

6.2 A comparison of the average $SNR$ of the SeqTIF and SeqCOR algorithms for both the stationary and time-varying mixtures. The average $SNR$ was computed across the ten speech mixtures, all $seriesnum$ and $fps = 6, 8$. ......................... 203

6.3 The results of an empirical study conducted to determine the effect that the threshold value $c_{\text{comp}}$ has on separation performance. The SCAtemp algorithm is applied to a set of 20 stationary mixtures as $c_{\text{comp}}$ is varied between 0.004 and 0.4. The $SNR$ performance (in dB) is shown for a subset of $c_{\text{comp}}$ values for analysis blocks spanning from 70ms to 0.56s. ......................... 211
List of Abbreviations

ADF      Adaptive Decorrelation Filtering
AR       Autoregressive
ASR      Automatic Speech Recognition
BSS      Blind Signal Separation
cdf      cumulative density function
DOA      Direction of Arrival
DWPT     Discrete Wavelet Packet Transform
DWT      Discrete Wavelet Transform
EVD      EigenValue Decomposition
FIR      Finite Impulse Response
fps      frames per series
ICA      Independent Component Analysis
iid      independent identically distributed
IIR      Infinite Impulse Response
IM       Interference Measure
ISTFT    Inverse Short Time Fourier Transform
JAD      Joint Approximate Diagonalisation
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>JADE</td>
<td>Joint Approximate Diagonalisation of Eigenmatrices</td>
</tr>
<tr>
<td>LP</td>
<td>Linear Prediction</td>
</tr>
<tr>
<td>LS</td>
<td>Least Squares</td>
</tr>
<tr>
<td>MAP</td>
<td>Maximum A Posteriori</td>
</tr>
<tr>
<td>MI</td>
<td>Mutual Information</td>
</tr>
<tr>
<td>ML</td>
<td>Maximum Likelihood</td>
</tr>
<tr>
<td>MSE</td>
<td>Mean Squared Error</td>
</tr>
<tr>
<td>MMSE</td>
<td>Minimum Mean Squared Error</td>
</tr>
<tr>
<td>pdf</td>
<td>probability density function</td>
</tr>
<tr>
<td>SCA</td>
<td>Sparse Component Analysis</td>
</tr>
<tr>
<td>STFT</td>
<td>Short Time Fourier Transform</td>
</tr>
<tr>
<td>t-f</td>
<td>time-frequency</td>
</tr>
<tr>
<td>TIFCORR</td>
<td>Time Frequency of CORrelation</td>
</tr>
<tr>
<td>TIFROM</td>
<td>Time Frequency Ratio Of Mixtures</td>
</tr>
<tr>
<td>SNR</td>
<td>Signal to Noise Ratio</td>
</tr>
<tr>
<td>SVD</td>
<td>Singular Value Decomposition</td>
</tr>
</tbody>
</table>