Blind Source Separation via Independent and Sparse Component Analysis with Application to Temporomandibular Disorder

---

Thesis submitted to the University of Cardiff in candidature for the degree of Doctor of Philosophy.

Clive Cheong Took

Centre of Digital Signal Processing
Cardiff University
2007

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

DECLARATION

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed ..... (candidate) Date 09/10/07

STATEMENT 1

This thesis is being submitted in partial fulfillment of the requirements for the degree of PhD.

Signed ..... (candidate) Date 09/10/07

STATEMENT 2

This thesis is the result of my own investigation, except where otherwise stated. Other sources are acknowledged by giving explicit reference. A bibliography is appended.

Signed ..... (candidate) Date 09/10/07

STATEMENT 3

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organizations.

Signed ..... (candidate) Date 09/10/07

Abstract

Blind source separation (BSS) addresses the problem of separating multichannel signals observed by generally spatially separated sensors into their constituent underlying sources. The passage of these sources through an unknown mixing medium results in these observed multichannel signals. This study focuses on BSS, with special emphasis on its application to the temporomandibular joint disorder (TMD). TMD refers to all medical problems related to the temporomandibular joint (TMJ), which holds the lower jaw (mandible) and the temporal bone (skull). The overall objective of the work is to extract the two TMJ sound sources generated by the two TMJs, from the bilateral recordings obtained from the auditory canals, so as to aid the clinician in diagnosis and planning treatment policies.

Firstly, the concept of 'variable tap length' is adopted in convolutive blind source separation. This relatively new concept has attracted attention in the field of adaptive signal processing, notably the least mean square (LMS) algorithm, but has not yet been introduced in the context of blind signal separation. The flexibility of the tap length of the proposed approach allows for the optimum tap length to be found, thereby mitigating computational complexity or catering for fractional delays arising in source separation.

Secondly, a novel fixed point BSS algorithm based on Ferrante's affine transformation is proposed. Ferrante's affine transformation provides the freedom to select the eigenvalues of the Jacobian matrix of the fixed point function and thereby improves the convergence properties of the fixed point iteration. Simulation studies demonstrate the improved convergence of the proposed approach compared to the well-known fixed point FastICA algorithm.

Thirdly, the underdetermined blind source separation problem using a filtering approach is addressed. An extension of the FastICA algorithm is

devised which exploits the disparity in the kurtoses of the underlying sources to estimate the mixing matrix and thereafter achieves source recovery by employing the ℓ_1-norm algorithm. Additionally, it will be shown that FastICA can also be utilised to extract the sources. Furthermore, it is illustrated how this scenario is particularly suitable for the separation of TMJ sounds.

Finally, estimation of fractional delays between the mixtures of the TMJ sources is proposed as a means for TMJ separation. The estimation of fractional delays is shown to simplify the source separation to a case of instantaneous BSS. Then, the estimated delay allows for an alignment of the TMJ mixtures, thereby overcoming a spacing constraint imposed by a well-known BSS technique, notably the DUET algorithm. The delay found from the TMJ bilateral recordings corroborates with the range reported in the literature. Furthermore, TMJ source localisation is also addressed as an aid to the dental specialist.

Contents

Acknowledgements

I would like to thank both my supervisors Dr. Saeid Sanei and Prof. Jonathon Chambers, for their invaluable support and guidance throughout this PhD. The creativity and enthusiasm of Dr. Sanei for biomedical signal processing has inspired me to think 'outside of the box'. On the other hand, I have not only benefited from the technical expertise of Prof. Chambers, but he has also ensured that I could focus on my research, free from financial constraints.

I am indebted to Prof. Stephen Dunne from King's College London, who has contributed significantly to the provision of materials required for this project. Funding for my study was provided by the Overseas Research Students Award Scheme (ORSAS) of 'Universities UK', and a complimentary scholarship from Cardiff University.

I am grateful to all my friends for providing a stable environment within the Centre of Digital Signal Processing. Special thanks to Ms. Linda Wan and Mr. Ritchie Ramsamy for their continuous support and saint-like patience in putting up with me during these last three years. They always believe in me. Finally, I must express my gratitude to my family for their unconditional love & prayers for my success and to God who gives nothing to those who keep their arms crossed.

Publications

The publications listed below account partially for the originality of the work presented herein.

Conference publications:

- C. Cheong Took, S. Sanei, and S. Dunne, "Computationally efficient classification of temporomandibular joint sounds," *in Proc. 12th Int. Conf. on Biomedical Engineering (ICBME)*, Suntec, Singapore, Dec. 2005.

- C. Cheong Took, S. Sanei, and J. Chambers, "A filtering approach to underdetermined blind source separation with application to temporomandibular disorders," *in Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, Toulouse, France, May 2006.

- C. Cheong Took, S. Sanei, J. Chambers, and S. Dunne,"A geometrically constrained anechoic model for blind separation of temporomandibular joint sounds," *14th European Signal Processing Conf. (EUSIPCO)*, Florence, Italy, Sep. 2006.

- C. Cheong Took, K. Nazarpour, S. Sanei, and J. Chambers, "Blind Separation of TMJ Sounds by Incorporating Fractional Delay Estimation," *Institute of Mathematics & its applications (IMA) Conf. on Mathematics in Signal Processing*, Cirencester, UK, Dec. 2006.

- L. Spyrou, S. Sanei and C. Cheong Took, "Estimation and location tracking of the P300 subcomponents from single-trial EEG signals," *in Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, Hawaii, USA, Apr. 2007.

- C. Cheong Took and S. Sanei, "Variable Tap Length Convolutive Blind Source Separation," *IEEE Asilomar Conference on Signals, Systems, and Computers*, California, USA, Nov. 2007 [Invited Paper].

Journal publications:

- C. Cheong Took, S. Sanei, J. Chambers and S. Dunne, "Underdetermined blind source separation of temporomandibular joint sounds," *IEEE Trans. on Biomedical Engineering*, vol. 53, pp. 2123 - 2126, Oct. 2006.

- C. Cheong Took, and S. Sanei, "Exploiting sparsity, sparseness, and super-Gaussianity in underdetermind blind identification of temporomandibular joint sounds," *Journal of Computers*, vol. 2, issue 6 , August 2007 [Invited Paper].

- C. Cheong Took, S. Sanei, J. Chambers, S. Rickard, and S. Dunne, "Fractional delay estimation for blind source separation and localisation of temporomandibular joint sounds," Accepted for publication in *IEEE Trans. on Biomedical Engineering*, August 2007.

Under review:

- C. Cheong Took, S. Sanei, and J. Chambers, "Application of Ferrante's affine transformation to improve the convergence of fixed point iterations within ICA algorithms," submitted to *IEEE Trans. on Neural Networks*, Mar. 2007.

List of Acronyms

BSS	Blind Source Separation
cdf	Cumulative Density Function
CMT	Contraction Mapping Theorem
CONTRA	Opposite side
dB	Decibel
det	Determinant
DUET	Degenerate Unmixing Estimation Technique
FIR	Finite Impulse Response
FT	Fractional Tap
GGD	Generalised Gaussian Distribution
HTM	Hierarchical Temporal Memory
ICA	Independent Component Analysis
IC	Independent Component
IPSI	Same side
MI	Mutual Information
ML	Maximum Likelihood

MMLETDE	Mixed Modulated Lagrange Explicit Time Delay Estimator
MSE	Mean Square Error
PCA	Principal Component Analysis
pdf	Probability Density Function
PI	Performance Index
PM	Performance Measure
r.v.	Random Variable
s.t.	subject to
SCA	Sparse Component Analysis
SIR	Signal to Interference Ratio
SNR	Signal to Noise Ratio
TMD	Temporomandibular Disorder
TMJ	Temporomandibular Joint
UBSS	Underdetermined Blind Source Separation
w.r.t.	with respect to

List of Symbols

$\|.\|$	Absolute value
$\|\|.\|\|_2$	Euclidean norm
$\odot$	Schur-Hadamard (elementwise) product
$(.)^H$	Hermitian transpose operator
$(.)^T$	Transpose operator
$(.)^\dagger$	Pseudo-inverse
$\lfloor . \rfloor$	Floor operator
$\Gamma(.)$	Gamma operator
$\Psi(.)$	Digamma function
$\Psi(.)'$	Trigamma function
Λ	Diagonal matrix
$\mathbf{a}_i$	ith column vector of mixing matrix
$\mathbf{A}$	Mixing matrix
$diag(\mathbf{b})$	Diagonal matrix with vector $\boldsymbol{b}$ on its main diagonal
$C_{\mathbf{u}}$	Covariance matrix of signals $\mathbf{u}$
$E\{\cdot\}$	Expectation operator

$\mathbf{I}$	Identity matrix
$I(y_i, y_j)$	Mutual information between variables y_i and y_j
$kurt(.)$	Kurtosis
L	Order of filter
m	Number of mixtures
n	Number of sources
$Neg(.)$	Negentropy
ω	Angular frequency
R	Covariance matrix
$Re\{.\}$	Real part
$\mathbf{S}$	Source matrix
$sgn(.)$	Signum function
sup(.)	Supremum function
T	Number of time samples
$\mathbf{V}$	Whitening matrix
$Var(.)$	Variance
$\mathbf{w}_i$	Estimated unmixing vector corresponding to the ith IC
$\mathbf{W}$	Estimated unmixing matrix
$\mathbf{X}$	Input data matrix
$\mathbf{Y}$	Estimated output data matrix
$\mathbf{Z}$	Whitened input data matrix

List of Figures

List of Tables

Chapter 1

INTRODUCTION

1.1 Blind Source Separation

Blind source separation (BSS) is currently one of the most exciting areas of research in statistical signal processing and unsupervised machine learning due to its potential applications in various areas such as financial time series analysis, biomedical signal processing, and digital communications [7–10]. As the appellation 'source separation' suggests, it is concerned with the recovery of the underlying sources from a set of observations. These observations are generated when the sources are mixed through an unknown medium. However, the main appeal of BSS lies in the word 'blind', which points out that source separation has to be achieved without any training data. Instead, only weak assumptions regarding the sources and the unknown medium are permitted.

The most common example to introduce BSS is the *Cocktail Party Problem* [8, 11, 12]. The setting is in a cocktail party, where many people are talking simultaneously. Yet, a listener in the party can discern the voice of a particular speaker from a myriad of other voices. This ability to select one voice in such an uncontrolled acoustic environment is possible, as the human brain learns how to exploit several physical factors such as the probability of recurring words, the accent of the speaker, the movement of the lips of the speaker, the distinction between male and female voices and so forth. This scenario illustrates both the aim and the properties of a BSS algorithm. The

objective, as mentioned earlier, is to extract the unknown sources from their observed mixtures. To undertake such a task, the BSS algorithm must also be adaptive and operate in a blind fashion.

Unlike the human brain, most BSS algorithms rely on one modality (e.g. audio information), although a few bi-modal BSS algorithms have been proposed [13–15]. Mimicking mother nature in this particular task is challenging, especially when one is faced with the computational complexity of processing video data, synchronising the video data with that of audio, and selection of the right criterion to correlate the visual information with audio data. Therefore, the most practical and well-known algorithms have the same common denominator, i.e. uni-modality. But, this uni-modality and the 'blindness' of BSS implies that BSS techniques have to fully exploit the weak assumptions concerning the sources and the mixing environment.

One of these assumptions is the statistical independence of the sources, which lays the foundation for most BSS algorithms. The term for the operation of this family of algorithms is independent component analysis (ICA). ICA is a powerful statistical tool, that seeks to transform data into a set of signals that are mutually statistically independent. However, the BSS problem is even further complicated when there are fewer sensors than sources. In this case, a less realistic but practical assumption regarding the sources is generally made, i.e. sparse sources. The sparsity of the sources refers to the situation where only a given source is active for a particular time interval, thus enabling one to exploit the structure of the mixing process [11]. Subsequently, sparse component analysis (SCA) has appeared. SCA is generally a non-linear technique which converts data into a set of sparse signals. Both ICA and SCA are explained in further detail in the following chapter.

Blind source separation dates back to the work of Herault and Jutten in

the French conference GRETSI in 1985 [16, 17]. Since that time, blind source separation has had a wide range of applications ranging from geophysical exploration to genomics, making it a 'hot' topic within the signal processing and the machine learning communities [12, 18]. In the sequel, BSS has evolved into three main classes, notably instantaneous, anechoic, and echoic/convolutive BSS. Whenever a high signal propagation velocity allows the assumption that the mixtures impinge on the sensors without any relative delay and there exists only single paths form the sources to sensors, it is termed instantaneous blind source separation. This arises in a number of biomedical applications, such as in electrocardiograms, electroencephalograms, and magnetoencephalograms [9, 19, 20]. On the other hand, anechoic BSS can be seen as the intermediate situation between instantaneous and convolutive/echoic BSS. It refers to the situation whereby a delay is associated with each source in the mixtures $\mathbf{X}$ through only direct paths. Examples of such a scenario are: a group of people talking in an open area, the acoustics in an anechoic chamber, Doppler frequency-shifts differing between mobile sensors and sources [21], and spatial shifts from reflections through window glass [22]. In echoic or convolutive BSS, each element of the mixing matrix $\mathbf{A}$ is in fact a linear filter to simulate multipaths from sources to sensors. In this case, the past as well as the present samples of the source signals contribute to the current mixture sample. Multipaths occur in communication systems and echoic chambers [5, 17].

Nevertheless, several issues need to be addressed, such as in time-domain convolutive BSS where the optimum tap-length of the filters is unknown, underdetermined BSS where the number of sources is greater than the number of mixtures, and its potential application to the vast disciplines within biomedicine, e.g. its relevance in the context of temporomandibular sounds. All these three issues are addressed in this thesis, with the goal of improving existing BSS techniques and their applicability to monitoring TMD. Next,

the relevance of BSS in the context of temporomandibular joint sounds is explained.

1.2 Temporomandibular Joint Disorder

The temporomandibular disorder encompasses most medical problems related to the region of the mandible (lower jaw) and the temporal bone. TMD is the most common non-dental related chronic source of oral-facial pain [23–25]. 75% of the USA population will at some time have some of the signs and symptoms of TMD [26], with a similar figure in the UK [23]. There are two well-known sounds generated by the temporomandibular joint (TMJ) of a TMD patient, namely click and crepitus. These TMJ signals classified into four classes (hard click, soft click, hard crepitus, and soft crepitus) are illustrated in Fig. 1.2, and the TMJ is shown in Fig. 1.1. It is noted that these classified TMJ sounds analysed in this thesis were obtained from the dental institute, Kings College London in the person of Prof. S. Dunne. There exist three main types of acoustic sensors for TMJ acquisition: piezoelectric accelerometers, a special type of two channel stethoscopes and small condenser microphones. A comparison study of these sensors is given in [27]. In this work, a special two channel stethoscope connected to microphones as used in [27] was employed. However, it is noted that the TMJ sounds presented only for the last part of Chapter Six is recorded from a pair of microphones placed in the auditory canals of the patient. Throughout this thesis, the sampling frequency to record the TMJ sounds is understood to be 12 kHz. Generally, the click is related to the displacement of the disc which holds the mandible and the temporal bone, and hence conveys the dysfunction of the TMJ. Likewise, the crepitus suggests at the presence of a degenerative joint disease (e.g. osteoarthrosis). Furthermore, a 'hard' TMJ sound is generally associated with a mature stage of the TMD, while a 'soft' TMJ sound hints a mild TMD [23]. Therefore, poor detection of these

sounds can lead to misdiagnosis of TMDs. A dental specialist has to differentiate between the TMJ sounds such as click, crepitus and noise produced by a 'normal' joint. Besides, the inherently *subjective* classification of these TMJ sounds makes it hard for the clinicians to determine the correct pathology. This has led to controversy as pointed out in [23, 28–30]. It was also highlighted in [31] that patients, who did not exhibit symptoms of TMD, suffered from this disorder, hence illustrating that the diagnosis of TMD can be quite challenging. Hence, there is a need for an objective and automated detection of these TMJ sources. To aid the dental specialist in making a prognosis of TMD, Guo *et al.* proposed to utilise a BSS algorithm called Infomax to separate two mixtures of crepitus [1]. This was the only work on TMD in the context of BSS, prior to the work presented within this thesis.

1.3 Signal Processing Techniques for TMJ Sounds

There are three main approaches to analyse TMJ sounds, notably classification or characterisation of TMJ sounds, TMJ source separation and lo-

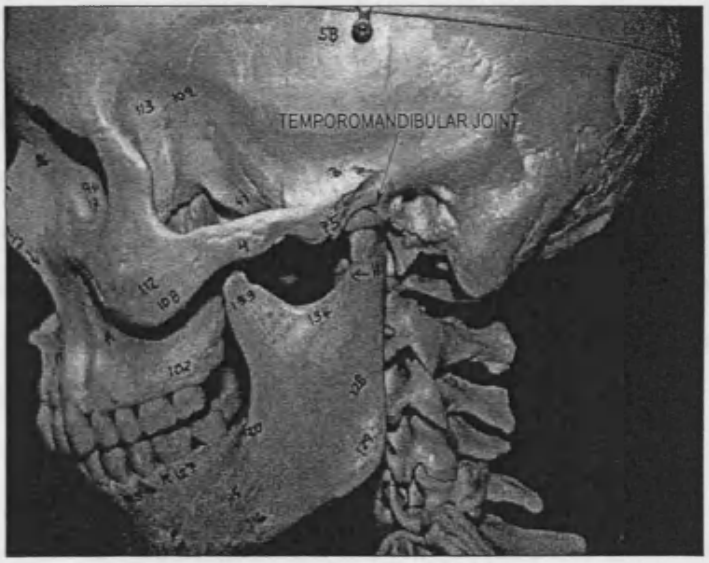

Figure 1.1. The temporomandibular joint.

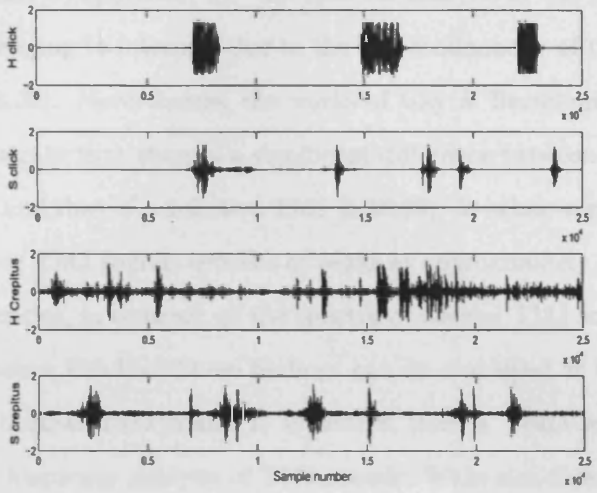

Figure 1.2. From top to bottom: hard click, soft click, hard crepitus, and soft crepitus.

calisation. Most of the work on TMJ has been undertaken in the context of TMJ classification. Prior to the study presented herein, only the work of Guo *et al.* addressed the problem of TMJ source separation [1]. On the other hand, a pair of studies by Widmalm *et al.* dealt with TMJ localisation [32, 33]. This has motivated the work presented herein. Thus, in this thesis, different scenarios of source separation of temporomandibular joint sounds are envisaged and simulated to demonstrate the potential of BSS techniques in this biomedical field. Furthermore, the last chapter of this thesis tackles the problem of TMJ source localisation as a post-processing step to TMJ separation from bilateral TMJ recordings. A brief review of these signal processing techniques is provided next.

1.3.1 Classification and characterisation of TMJ sounds

Due to the non-stationarity of the TMJ sounds, there has been tremendous work on the time-frequency analysis of TMJ sounds for the purpose of classification [30, 34–37]. These approaches are briefly reviewed in Chapter six.

A more 'primitive' approach, i.e. the spectral analysis of the TMJ sources is more challenging to interpret due to the non-stationarity of the TMJ signals [6, 27, 38, 39]. Nevertheless, the works of Gay & Bertolami, and Gallo *et al.* demonstrate that there is a significant difference between spectra of a normal TMJ and that of a defective TMJ [6,38,39]. In other words, the spectra of defective TMJ sounds consists of peaks at approximately 800 Hz or at higher frequencies, in contrast to the spectra of normal TMJ sounds, which are mainly below 800 Hz. These findings can be employed as benchmarks, whenever a normal TMJ sound is separated from a TMD sound source. Through the frequency analysis of TMJ sounds, Widmalm deduced that the best acquisition of the TMJ sounds can be performed by utilising microphones located at the auditory canals [27]. Furthermore, Leader *et al.* and Watt tackled the TMJ classification from a quantitative perspective [40,41]. Watt examined the waveforms of TMJ sounds [41], while Leader *et al.* fused several features such as the number of sound events, energy in each sound event, and time interval between the sound events in each TMJ signals to categorise the TMJ sounds. Next, the work of Guo *et al.* is examined.

1.3.2 Source separation of TMJ sounds

The two TMJ joints generate two sound sources, while background noise such as noise generated by dental equipments, breathing of the patient, movement of masticatory muscles, and the blood flow of the temporal artery can contribute to another acoustic source. Guo *et al.* addressed the source separation of crepitus [1]. They employed the convolutive Infomax algorithm proposed by Torkkola [17] to solve this particular BSS problem. This algorithm was derived, based on the assumption that the sources are statistically independent. The Infomax was implemented via a feedback neural network shown in Fig. 1.3. This algorithm will be reviewed on the algorithmic level in the following chapter. In their work, Guo *et al.* considered the mixing model within the brain to be convolutive [1]. More specifically, they consid-

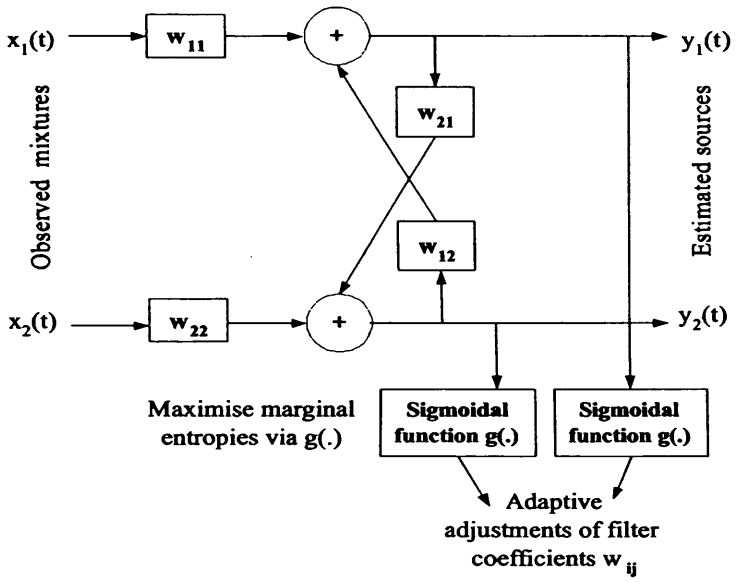

Figure 1.3. The feedback neural network employed to separate the TMJ sources [1].

ered the propagation of crepitus from the ipsi (originating) side to the contra (opposite) side to be characterised by multipaths. This *intuitive* assumption did not consider any physiological aspects of the human head and no existing literature review supports this convolutive model. The hypothesis of acoustic multipaths within the human head is plausible, although the acoustic attenuation within the brain reported in the literature suggests that these multipaths from one side of the head to the opposite side are negligible. The mixing model of the head is examined in greater detail in Chapter six of this thesis. Guo *et al.* then demonstrated that a particular synchronised peak present in both TMJ mixtures was attenuated in one of the extracted crepitus, while in the other crepitus, it was still present. Thereafter, they argued that the separation of crepitus sources has been achieved successfully. Firstly, this work illustrates that TMJ source separation can be achieved. Secondly, there is scope for more TMJ source separation scenarios, such as the separation of a normal TMJ sound from a click source. Thirdly, the fact that the convolutive mixing model was formulated rather intuitively leads

to the following question: is the convolutive mixing model appropriate for TMJ BSS? The convolutive model is generally over-parameterised (due to the consideration of hundreds of delay units), when the mixing model instead can be instantaneous or anechoic. Henceforth, instantaneous as well as anechoic mixing models for TMJ sounds are also addressed in this thesis.

1.3.3 Localisation of TMJ sounds

Widmalm *et al.* examined the particular case where a patient suffered from TMD at only one temporomandibular joint [32, 33]. Therefore, the objective of their studies was to locate the defective joint, based on the bilateral recordings obtained from the auditory canals. The delay between the ipsi TMJ sound and the contra TMJ sound was calculated. This was estimated via the gradient of the phase between the two bilateral recordings. According to their studies, the delay was found to be in the range of 0.2-1.2 ms. This range illustrates that the acoustic medium of the human brain depends on the individual. Widmalm *et al.*, however, did not consider the possibility that the recording on the contralateral side of the TMD is in fact a mixture of the 'click' and the sound produced by the normal joint. This possibility is supported by the fact that the normal TMJ also produces a sound, as investigated in [6, 38, 39]. Hence, one of the aims of this thesis is to consider such a scenario, i.e. the separation of click and normal TMJ sound from the bilateral recordings. Furthermore, localisation of the defective TMD joint is also addressed in this thesis.

1.3.4 Conclusions and objectives

In the light of the above survey on TMD from a signal processing perspective, it can be deduced that:

1. The non-stationarity property exhibited by the TMJ signals can be exploited in TMJ source separation. This non-stationarity also implies the super-Gaussianity (as shown by Parra and Spence [8]) of the TMJ

sources. Their proof is included in Appendix A.3. Thus, the super-Gaussianity of the TMJ sources can also be utilised as a statistical criterion to perform TMJ source separation.

2. Several possibilities for TMJ BSS can be envisaged and simulated, because Guo *et al.* considered only the separation of crepitus sources. More precisely, instantaneous and anechoic mixing of TMJ sources are simulated. This is undertaken by the synthetic mixing of the ipsilateral TMJ sounds pertaining to a particular TMJ class, i.e. *soft* clicks & crepitus or *hard* clicks & crepitus. However, the ethical protocol for the acquisition of TMJ sounds has limited most of the work presented herein to synthetic mixing.

3. The localisation of the TMJ infected source, as a post-processing step to the separation of the TMJ sources is also addressed in this thesis, as an aid to the dental specialist.

1.4 Organisation of the thesis

Chapter two lays the foundation for blind source separation. The objective of this chapter is to introduce the techniques pertaining to BSS. These techniques are illustrated by the commented outlines of some well-known algorithms.

Chapter three proposes a variable tap length convolutive BSS algorithm. This concept has been adopted from the LMS algorithm. In contrast to the LMS algorithm, the proposed *blind* technique does not have a priori the 'desired' signal to adaptively determine the optimum tap length. Therefore, the optimum tap length is defined as the minimum tap length that minimises the off-term elements of the covariance of the estimated sources. This adaptive property of the variable tap length approach opens a new field of research within the BSS area.

Chapter four applies Ferrante's affine transformation to a fixed point

algorithm to improve its convergence properties. Moreover, the transformation allows for the application of fixed point concepts such as attractive fixed point, or the contraction mapping theorem. In that respect, Chapter four analyses the convergence properties of the proposed fixed point BSS algorithm, which converges faster than the well-known fixed point FastICA algorithm.

Chapter five addresses a particular underdetermined TMJ BSS scenario. A filtering approach based on the FastICA algorithm is proposed to solve this particular TMJ separation scenario. The proposed approach is robust to noise modelled as a non-sparse source at a signal-to-noise ratio of 0 dB. It was found that the filtering of the mixtures attenuates one of the sources such that two sources can be estimated at a time, thereby enabling the full identification of the underdetermined BSS.

Chapter six presents two methodologies in incorporating fractional delays in the context of TMJ separation. The first part reduces the anechoic source separation to an instantaneous separation, thereafter a conventional instantaneous BSS algorithm can be employed. The second part deals with the extraction of the sound source produced by a normal TMJ from a click source generated by an infected TMJ. Both parts address the problem of TMJ source localisation to pinpoint the location of each of the estimated sources.

The last chapter summarises the work presented herein and draws general conclusions. It also suggests opportunities for future work.

Chapter 2

FUNDAMENTALS OF BLIND SOURCE SEPARATION

2.1 Problem Statement

The BSS problem is to recover the constituent sources $\mathbf{s}(t)$ from a given set of observed or mixture signals $\mathbf{x}(t)$, with minimum assumptions about the mixing medium and the underlying sources. In effect, the generative models are summarised in the Table 2.1 [11]:

Table 2.1. Generative mixing models for instantaneous, anechoic, and convolutive blind source separation.

Generative Mixing Model	Mathematical Model
Instantaneous	$x_i(t) = \sum_{j=1}^{n} a_{ij} s_j(t) + v_i(t)$
Anechoic	$x_i(t) = \sum_{j=1}^{n} a_{ij} s_j(t - \tau_{ij}) + v_i(t)$
Convolutive	$x_i(t) = \sum_{j=1}^{n} \sum_{p=1}^{L} a_{ijp} s_j(t - \tau_{ijp}) + v_i(t)$

In this table $i = 1, ..., m$, $x_i(t)$ denotes the ith element of the mixture column vector $\mathbf{x}(t) \in \Re^m$, $s_j(t)$ denotes the jth element of the source column vector $\mathbf{s}(t) \in \Re^n$, $v_i(t)$ denotes the ith element of the noise column vector $\mathbf{v}(t) \in \Re^m$, t denotes the discrete time index, and a_{ijp} is the attenuation element of the mixing matrix $\mathbf{A}$ corresponding to its ith row, jth column, and its corresponding delay τ_{ijp}. In the context of TMJ source separation, the range of τ_{ijp} corresponds to 2.4-14.4 samples at a sampling frequency of 12 kHz. The absence of subscript p in a_{ij} implies that there is at most

one delay. It is noteworthy to say that **A** is called the signal dictionary or basis matrix in the sparse component analysis (SCA) literature [18,42]. Note however, that $v_i(t)$ corresponding to the additive noise of the ith sensor is negligible in the context of the source separation of TMJs. In this work, it is reasonable to assume that **A** is stationary, i.e. it does not vary with time. In the dipole source model ([43] and the references therein), the stationarity of **A** leads to the fact that the sources have fixed locations and orientations. Similarly, in this work, it is assumed that the sources do not move. For example, in the context of TMD study, the sound sources are generated by the temporomandibular joints, which indeed are fixed in location.

The two main trends within the BSS community are to either investigate instantaneous or convolutive source separation. In theory, convolutive BSS algorithms should perform much better in an anechoic scenario, while the over-parameterisation of convolutive BSS with regard to the instantaneous BSS explains why instantaneous BSS is still an ongoing topic in the BSS area. Likewise, many instantaneous techniques have been extended to the anechoic case without too much effort [11, 21, 22, 44–46]. Hence, the only techniques pertaining to these two main topics are discussed in this chapter. In Chapter six, however, an important algorithm is overviewed for anechoic BSS, namely the DUET algorithm [5]. On the basis of the generative models in Table 2.1, the source separation problem for the instantaneous and convolutive cases can be solved as follows:

Table 2.2. Unmixing models for instantaneous and convolutive BSS.

Unmixing Model	Mathematical Model
Instantaneous	$y_j(t) = \sum_{i=1}^{m} w_{ji} x_i(t)$
Convolutive	$y_j(t) = \sum_{i=1}^{m} \sum_{p=1}^{L} w_{jip} x_i(t - \tau_{jip})$

where $j = 1, ..., n$, $y_j(t)$ denotes the jth element of the estimated source column vector $\mathbf{y}(t)$, and w_{jip} is the gain element of the so-called separating

or unmixing matrix $\mathbf{W}$ corresponding to its jth row, ith column, and its corresponding delay τ_{jip}. On the other hand, the anechoic BSS does not offer a unified elegant solution. For instance, Yeredor estimates the mixing matrix and its corresponding delays via his non-orthogonal joint diagonalisation approach [21, 22, 44], while Torkkola provides two solutions via a feedforward and a feedback network in [46], and in the DUET algorithm, Yilmaz and Rickard perform source separation via binary masking. Next, the ambiguities inherent to the source separation problem are highlighted.

2.2 Indeterminancies of the Problem

There exists two indeterminancies inherent to BSS, namely the permutation and the scaling ambiguities. In other words,

1. The order of the recovered sources cannot be determined, mainly due to the 'blindness' of the problem, i.e. both the mixing matrix and the sources are unknown [7, 12]. Thus, a change in the order of the recovered sources also implies a permutation of the corresponding columns of the mixing matrix. Alternatively, this can be viewed as a change in the order of the terms for the outer summations in Table 2.1 does not affect the result of the summations.

2. From the following equation:

$$\mathbf{x}(t) = (\frac{1}{\gamma_k}\mathbf{a}_k)(\gamma_k s_k(t)) + \sum_{j \neq k} \mathbf{a}_j s_j(t) \tag{2.2.1}$$

   It is clear that an arbitrary multiplying factor γ_k to the kth source can be cancelled out by dividing the kth column of the mixing matrix by the same factor γ_k. This demonstrates that the sources can be estimated only up to a scaling constant.

Some researchers exploit the scaling ambiguity to simplify their algorithms, by enforcing the variances of the estimated sources to be unity [7, 12, 47].

Moreover, it is noted that the scaling ambiguity also includes the sign ambiguity, i.e. the BSS model will not be altered, if any of the sources is multiplied by -1. These ambiguities show that the separating matrix $\mathbf{W}$ is not necessarily the exact inverse of the mixing matrix $\mathbf{A}$. Instead,

$$\mathbf{W} = \mathbf{P}\mathbf{\Lambda}\mathbf{A}^{\dagger} \tag{2.2.2}$$

where the superscript $(.)^{\dagger}$ denotes the pseudo-inverse to cater for over-determined BSS as well, $\mathbf{P}$ is a permutation matrix, and $\mathbf{\Lambda}$ is a diagonal matrix to convey the scaling ambiguity.

2.3 Techniques for BSS

This section overviews two techniques, notably independent component analysis (ICA) and sparse component analysis (SCA) for BSS. ICA estimates statistically independent sources, whilst SCA recovers sparse sources. These will be discussed in more detail in the following two sections. ICA traces back to the early work of Herrault and Jutten [16], when the latter introduced BSS to the signal processing community in 1985. This is why some researchers consider ICA and BSS as one entity, and use these two terms interchangeably. According to Hyvärinen *et al.* [7], it was Infomax proposed by Bell and Sejnowski [48] that sparked much enthusiasm for this problem. However, prior to the formulation of SCA, ICA was limited to exactly-determined (i.e. equal number of sources and sensors, $n = m$), and over-determined (i.e. more sensors than sources) cases [11].

Whenever the number of sensors is less than the number of sources, SCA is a more practical tool to separate the sources. In such cases, the number of active sources at each time instant should be generally at most equal to the number of sensors. This particular situation is termed as the sparsity of the sources. Therefore, the sparsity of the sources make the under-determined

BSS a pseudo-determined BSS at a particular time instant. The sparsity of the sources can be enforced by selecting an appropriate domain such as the frequency domain, and the wavelet domain. In 2000, the work of Bofill and Zibulevsky [49] illustrated that SCA can solve under-determined BSS (UBSS) without too much difficulty. Consequently, SCA attracted much attention and attained a much wider audience. The concept of SCA was already applied in mid-1990's, although it is not clear in the literature which work initiated SCA. Further details on the history of both ICA and SCA can be consulted from the following literature [7,8,11]. It should be stressed that most ICA and SCA algorithms do not cater for the noise $v_i(t)$ in the mixing models tabulated in Table 2.1. In other words, they perform source separation, without cancelling out the noise from the estimated sources as can be seen in Table 2.2. This is because the system is under-determined, if each $v_i(t)$ the additive noise of the ith sensor was considered as a source, making the BSS more complex and less tractable. Therefore, the common approach is to consider $v_i(t)$ negligible, which is the case of TMJ source separation. In the following section, ICA is defined, illustrated by an instructive example of ICA, and concluded by a survey on existing methods. Likewise for SCA in section 2.5.

2.4 Independent Component Analysis

2.4.1 Definition

Independent component analysis is a statistical approach designed to decompose multivariate data into components that are as statistically independent as possible. In the literature [7,8,17], ICA normally refers to a linear transform, i.e. the instantaneous BSS model. Nevertheless, within the same literature, some authors address convolutive BSS and implicitly convey the idea that these convolutive BSS algorithms form part of the ICA family. For simplicity in this thesis, ICA refers to the techniques which solve BSS based

on the statistical independence of the sources assumption. In effect, ICA implies that the joint probability density function $p(\mathbf{s}(t))$ of the sources can be factorised as:

$$p(\mathbf{s}(t)) = \prod_{j=1}^{n} p_j(s_j(t)) \tag{2.4.1}$$

where $p_j(s_j(t))$ is the marginal distribution of the jth source. Furthermore, the statistical independence of the sources implies the uncorrelatedness of the sources, but the reverse is not necessarily true. As a pre-processing step, most ICA algorithms decorrelate the mixtures via spatial whitening, before optimising their separating criteria known as contrast/cost functions. This spatial whitening is achieved by employing the well-known principal component analysis (PCA), which is explained next.

Principal Component Analysis

In the context of BSS, PCA seeks to remove the cross-correlation between the observed signals, and ensuring that they have unit variance. It operates by finding the projections of the mixture data in orthogonal directions of maximum variances [7]. A vector $\mathbf{z}$ is said to be spatially white iff

$$E\{\mathbf{z}(t)\mathbf{z}^T(t) - \mathbf{I}\} = 0 \tag{2.4.2}$$

where $E\{.\}$ denotes the expectation operator and $\mathbf{I}$ the identity matrix. The separating matrix, $\mathbf{W}$ can be decomposed into two components, i.e.

$$\mathbf{W} = \mathbf{UV} \tag{2.4.3}$$

where $\mathbf{V}$ is the whitening matrix and $\mathbf{U}$ is a rotation matrix [50]. Assuming $m = n$, there are n^2 unknown parameters in $\mathbf{W}$. PCA requires the n diagonal elements of the covariance $\mathbf{C_z}$ to be unity, and due to the symmetric property of $\mathbf{C_z}$, it suffices that only $(n^2 - n)/2$ of its off-terms to be zero. Therefore, spatial whiteness imposes $n(n + 1)/2$ constraints. This leaves $n(n - 1)/2$

unknown parameters. Hence as Cardoso describes it, *prewhitening only does half of the BSS job* [50]. The whitening matrix $\mathbf{V}$ can be computed as follows:

$$\mathbf{V} = \mathbf{Q}^{-\frac{1}{2}}\mathbf{E}^T \tag{2.4.4}$$

$\mathbf{E}$ is the eigenvector matrix of the covariance matrix of $\mathbf{x}$, $\mathbf{C_x}$. It projects the data into the n-dimensional source space. $\mathbf{Q}$ is a diagonal matrix storing the eigenvalues of $\mathbf{C_x}$. $\mathbf{Q}^{-\frac{1}{2}}$ makes the projections have unit variance. However, it is important to notice that the whitening matrix $\mathbf{V}$ is not unique because it can be pre-multiplied by an orthogonal matrix to obtain another version of $\mathbf{V}$.

2.4.2 An illustrative example

In the previous subsection, it was mentioned that PCA does half of the job of ICA. The other half is to effectively find the rotation matrix $\mathbf{U}$ in Eq. (2.4.3). In order to visualise how ICA performs, consider the following example: Two uniformly distributed sources s_1 and s_2 are mixed by the following matrix, at an angle $\theta = \pi/4$:

$$\mathbf{A} = \begin{bmatrix} \cos(\theta) & -2\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -2 \\ 1 & 1 \end{bmatrix} \tag{2.4.5}$$

This mixing matrix is in effect a rotation of 45 degrees, followed by a stretch in the same direction of the horizontal line joining the point (1,1) and (-1,1) by a factor of $2/\sqrt{2}$. The upper left scatter plot of Fig. 2.1 shows that the two sources are independent. For example, whenever s_1 is at its minimum value -1, s_2 has several possible values. This means that the knowledge of the value of one of the sources does not give any information on the value of the other sources, demonstrating the statistical independence [51]. On the other hand, in the scatter plot of the mixtures x_1 and x_2, whenever x_1

is at its minimum value, there is only one corresponding value of x_2, hence illustrating their dependence. The lower left plot of Fig. 2.1 demonstrates that PCA reverses the stretching effect of the mixing matrix, while the final step of the ICA rotates back the whitened data $\mathbf{z}$ to yield $\mathbf{y}$.

2.4.3 ICA approaches to BSS

ICA relies on fundamentally two factors: 1) A statistical criterion expressed in terms of a cost/contrast function $C(\mathbf{y}(t))$, which requires to be either minimised or to be maximised, 2) An optimisation technique to carry out the minimisation or maximisation of the cost function.

Many researchers have focused mainly on formulating new cost functions to propose novel BSS algorithms. In doing so, it is common in the BSS community to employ either the traditional steepest descent/ascent, or those more specific to the BSS field, such as the natural gradient algorithm (NGA) [52].

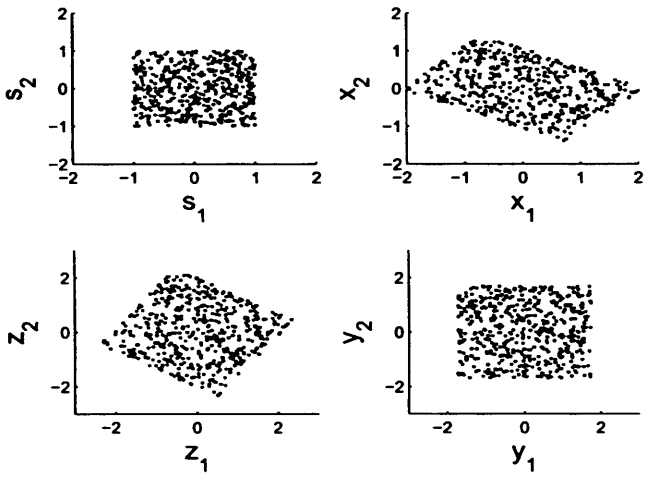

Figure 2.1. Scatter plots of original sources (upper left), mixtures (upper right), whitened mixtures (lower left), recovered sources with ICA (lower right).

The natural gradient can be expressed as:

$$\nabla_{NGA}\mathbf{W} = \frac{\partial C(\mathbf{y}(t))}{\partial \mathbf{W}}\mathbf{W}^T\mathbf{W} \tag{2.4.6}$$

where $C(\mathbf{y}(t))$ is the cost function to be either minimised or maximised, and $\nabla_{NGA}\mathbf{W}$ is the natural gradient w.r.t. to separating matrix $\mathbf{W}$. This gradient is derived based on the fact that the optimisation space is Riemannian or curved [52]. The concept of Riemannian is intrinsically related to differential geometry, which is the mathematics of curved space. The NGA has been shown to work more efficiently in terms of convergence than the normal gradient approach [52], and therefore it has been used extensively [47]. However, this thesis does not address the aspects regarding this algorithm or any other gradient-based approaches. Instead, fixed point iteration is the subject of Chapter four. This iterative optimisation technique can simply be summarised as:

$$u_{k+1} = f(u_k) \tag{2.4.7}$$

where $f(.)$ is a function of u_k. It is noted that at the solution u

$$f(u) = u \tag{2.4.8}$$

and therefore, unlike gradient-based approaches, its performance does not depend on any step-size parameter. An example of a fixed point iteration for a logistic function is demonstrated in Fig. 2.2. This example is taken from Prof. Moon's book [2]. Note that the fixed point iteration operates explicitly on the axis $y = f(x) = x$. The fixed point is where logistic function intersects the axis $y = x$, and is the point where the iteration terminates.
Fixed point iteration whose theory is well-established has so far resulted in one prominent fixed point BSS algorithm, i.e. FastICA [7], although there exists derivatives of FastICA [53–55]. In the existing literature, there does not seem to be any work which analyses such optimisation technique from

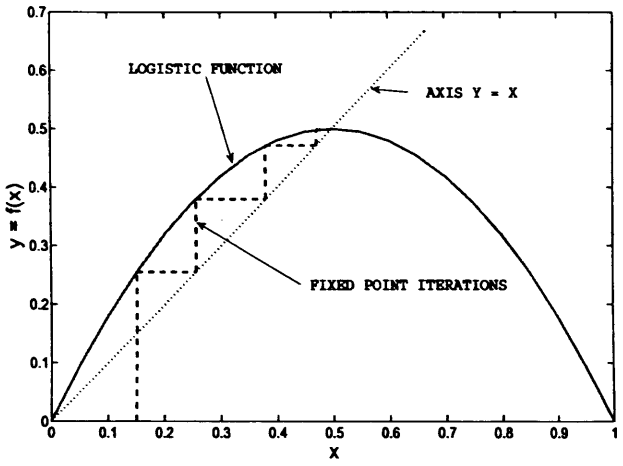

Figure 2.2. An example of fixed point iteration corresponding to the logistic function $f(x) = 2x(1-x)$ taken from [2].

a fixed point theory perspective. Hyvärinen *et al.* improve the convergence of their FastICA algorithm by adopting Lagrange method to estimate the separating matrix $\mathbf{W}$ [7], while Regalia and Kofidis examine the convexity of the contrast functions of FastICA [56]. Fixed point theory encompasses concepts such as attractive or repulsive fixed points, and theorems such as Contraction Mapping Theorem (CMT) [2, 57]. These are included in Appendix A.1 for clarity. Chapter four exploits these concepts to analyse a novel fixed point BSS algorithm.

Here, a survey on the statistical criteria employed by several ICA algorithms is provided.

- Many techniques such as second order blind identification (SOBI) [58], algorithm for multiple unknown signals extraction (AMUSE) [59] employ second order statistics to exploit the temporal structure of the sources, mainly the temporal correlation of the sources. While another class of second order techniques such as Parra's algorithm [4] exploit the statistical non-stationarity of the source signals. These techniques are particularly attractive, as they involve only second order statistics,

which are computationally less intensive than the methods based on higher order statistics.

- Another class of ICA algorithms utilise higher order statistics to maximise the statistical independence. For example, the Joint Approximate Diagonalisation of Eigenmatrices (JADE) algorithm jointly diagonalises a set of fourth-order cumulant matrices, such that the sum of squared cross-cumulants is minimised [60]. The reason why these algorithms employ higher order statistics lies in the fact the sources are statistically independent. In other words, uncorrelatedness at higher order statistics entails statistical independence, whilst uncorrelatedness at second order statistics does not imply independence, except if the sources are Gaussian [7].

- The last class of ICA algorithms are derived from an information-theoretic perspective. This family of ICA algorithms exploits concepts borrowed from information theory such as entropy, and mutual information. It is noted that two variables are said to be statistically independent, whenever their mutual information is zero [61]. Examples of this ICA category are the Infomax algorithm of Bell and Sejnowski [48], which attempts to maximise the entropy of the estimated sources, and FastICA of Hyvarinen *et al.* that utilises differential entropy, negentropy [7]. These two algorithms will be discussed in more detail later.

This concludes ICA, but well-established algorithms such as Infomax [48], FastICA [7], and that of Parra [4] will be examined in greater detail later. In the following section, SCA for BSS is explained. As mentioned in section 2.3, it is a relatively new field compared to ICA. Therefore, the literature pertaining to SCA is not as structured as that of ICA and no textbook on the subject is yet available.

2.5 Sparse Component Analysis

2.5.1 Definition

The fundamental assumption of sparse component analysis rests upon the sparsity of a multivariate data or that it can be sparsified by a given transformation such as Fourier-transform, wavelet transform, and so forth. Sparsity "implies" that most of the values of the sources are zero, with only a few sources taking significant values [62]. SCA estimates the basis vectors of the mixing matrix, by exploiting the geometric constraint entailed by sparsity. This geometric constraint can be viewed as follows: in a scatter plot of the mixtures $\mathbf{x}(t)$, the points lie in the direction of the basis vectors of the mixing matrix. The next section visually demonstrates such a scenario. O'Grady *et al.* regarded this geometric constraint as the structure of the mixing matrix 'appearing' in the mixture signals [11]. BSS involves blind channel identification and source recovery. Most ICA algorithms regard these two problems as one indivisible operation by finding one separating matrix. On the other hand, SCA treats them as two distinct problems, while relaxing the statistical independence of the sources. Next, a typical example of SCA is illustrated.

2.5.2 An illustrative example

In this section, an instructive example of SCA is given to; 1) understand what the underlying principle behind SCA is and, 2) demonstrate how it can solve the under-determined BSS (i.e. more sources than sensors). Three uniformly distributed sparse sources s_1, s_2 and s_3 are mixed by the following matrix:

$$\mathbf{A} = \begin{bmatrix} 0 & -1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \tag{2.5.1}$$

In Fig. 2.3, the left plot demonstrates the scatter plot of the sparse sources. It is worth noting that sparsity is conveyed by the fact that there is no

overlap between the sources, as compared with the upper left plot of Fig. 2.1. In other words, the samples of s_1 in the x-axis overlap with the samples of s_2 in the y-axis and s_3 in the z-axis only at origin. This demonstrates a degree of dependence between the sources permitted in SCA, but not allowed in ICA. The right plot of Fig. 2.3 illustrates the scatter plot of the resulting mixtures. It is clear that the samples of s_1 lie in the direction of the first basis vector $[0\ 1]^T$ of the mixing matrix **A**. Similarly, this can be observed for s_2 and s_3. This graph gives a good insight on how to estimate the mixing matrix, as well as separating the sources. The columns of the mixing matrix can be estimated by finding the directions of each of those three lines. Also, the source recovery can be achieved by clustering the samples pertaining to each of those lines. This example looks trivial. Nevertheless, this concept lays the foundation for most SCA algorithms, which are mostly designed to solve the under-determined BSS (UBSS).

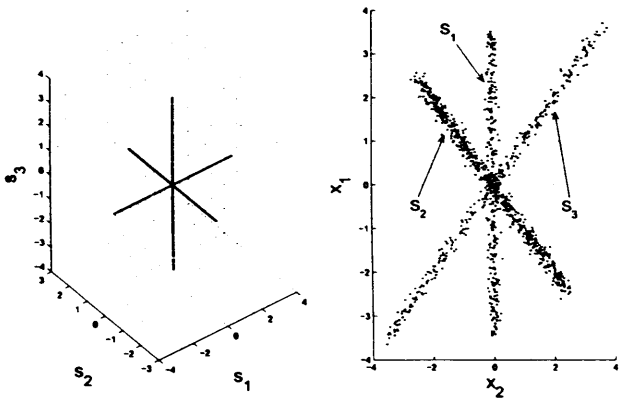

Figure 2.3. The left plot illustrates the three dimensional scatter plot of the three uniformly distributed sources. Note the non-overlap structure of this plot due to the sparsity of the sources. The right plot demonstrates how the mixtures align in the direction of the basis vectors of the mixing matrix in the scatter plot of the mixtures.

2.5.3 SCA approaches to BSS

As mentioned earlier, research within the SCA community does not enjoy the same maturity as ICA. Further to the previous section, most of SCA algorithms attempt to perform clustering on the mixture data to solve either partially (identification of mixing matrix only) or fully (estimation both mixing matrix and the sources) UBSS. For example, utilisation of traditional clustering algorithms such as C-Means (see Chapter seven [8]), and K-means in the DUET algorithm [5] has already been reported in the BSS literature. These two algorithms have also been modified for the purpose of SCA [8, 63]. Furthermore, more advanced clustering techniques namely Gap-statistics and self splitting competitive learning have been proposed to solve UBSS in [64]. These two advanced clustering techniques enable blind detection of the number of active sources over a given time-frequency interval. For source recovery only, it is common to use the ℓ_1-norm minimisation algorithm [42,65–67]. It estimates the sparse sources by minimising their ℓ_1-norm at a particular time instant. Nonetheless, it requires a priori the mixing matrix. This algorithm is attractive mainly because it finds the sparsest solution and hence can solve the sparse UBSS. A more comprehensive survey can be found in [11]. Now, some of the most significant BSS algorithms are examined.

2.6 Protagonists in BSS

2.6.1 Infomax

A good starting point is to discuss an algorithm which exploits explicitly the statistical independence of the sources such as Infomax [48]. Bell and Sejnowski endeavour to maximise the statistical independence by minimising the mutual information between the source estimates. Two independent variables y_1 and y_2 are said to be statistically independent, whenever their mutual information is zero [61]. The upper Venn diagram of Fig. 2.4 illus-

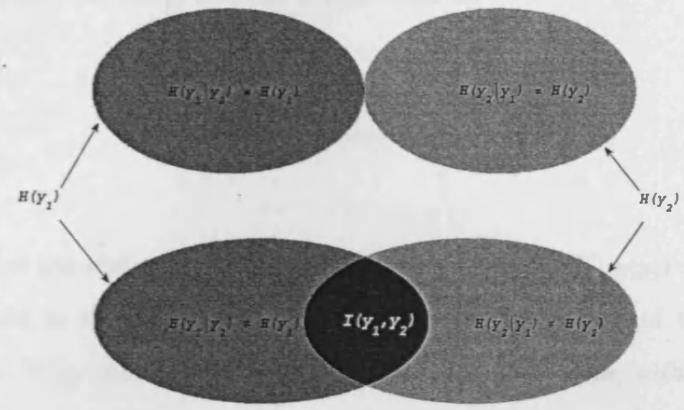

Figure 2.4. The upper Venn diagram shows the entropy relationships between two statistically independent variables y_1 and y_2. The lower Venn diagram corresponds to two dependent y_1 and y_2. $I(y_1, y_2)$ denotes mutual information between y_1 and y_2; $H(y_i|y_j)$ is conditional entropy of probability of y_i, given y_j, where $i \neq j$; $H(y_k)$ corresponds to entropy of y_k, for all k.

trates such a case. From this diagram, it is clear that

$$H(y_i|y_j) = H(y_i) \tag{2.6.1}$$

where $H(y_i|y_j)$ denotes the conditional entropy of y_i, given y_j, $i \neq j$; $H(y_k) = -E\{\log P(y_k)\}$ is the entropy of y_k $\forall$ k, with E{.} as the expectation operator. In contrast to the upper diagram, the lower Venn diagram demonstrates the statistical dependence of two correlated variables y_1 and y_2 through their intersection. The latter conveys their mutual information $I(y_1, y_2)$. Eq. (2.6.1) no longer holds, but instead,

$$\begin{aligned} H(y_i) &\neq H(y_i|y_j) \\ &= H(y_i|y_j) + I(y_i, y_j) \end{aligned} \tag{2.6.2}$$

Alternatively, the above formula is equivalent to:

$$\underbrace{H(y_i|y_j)}_{maximise} = \underbrace{H(y_i)}_{maximise} - \underbrace{I(y_i, y_j)}_{minimise} \tag{2.6.3}$$

To achieve the statistical independence, the area under $H(y_i|y_j)$ should be maximised in the lower Venn diagram of Fig. 2.4 to that of the upper diagram. Since mutual information is a non-negative value, with its minimum possible value being zero, maximising $H(y_i|y_j)$ leads to minimisation of $I(y_i, y_j)$. This implicitly corresponds to maximising the marginal entropy $H(y_i)$. Hence, the aim is to maximise the marginal entropies of the source estimates to attain statistical independence, thereby achieving separation. Similarly, the infomax of Tony and Bell endeavour to maximise the marginal entropies of the output $u_i = g(y_i)$ of a non-linear neural network [48]. One way to achieve this is to ensure that the outputs of the neural network have approximately uniform distributions by applying sigmoidal-like non-linear functions $g(y_i)$, such as $\tanh(y_i)$ and $1/(1 + e^{-y_i})$ on the estimated sources y_i. To tackle the convolutive BSS, Torkkola proposed to use Infomax based on a feed-back neural network as follows [17]:

$$\begin{aligned} y_1(t) &= \sum_{p=0}^{L_{11}} w_{11p} x_1(t-p) + \sum_{p=1}^{L_{12}} w_{12p} y_2(t-p) \\ y_2(t) &= \sum_{p=0}^{L_{22}} w_{22p} x_2(t-p) + \sum_{p=1}^{L_{21}} w_{21p} y_1(t-p) \end{aligned} \tag{2.6.4}$$

where w_{ijp} denotes the corresponding separating filter coefficient, $y_j(t)$ denotes the jth output at discrete time t, and $x_i(t)$ denotes the ith convolutive mixture. The resulting increments to learn the parameters of the network

can be summarised as

$$\begin{aligned}
\Delta w_{ii0} &\propto \Phi(y_i(t))x_i(t) + 1/w_{ii0} \\
\Delta w_{iip} &\propto \Phi(y_i(t))x_i(t-p) \\
\Delta w_{ijp} &\propto \Phi(y_i(t))y_j(t-p) \quad i \neq j, \quad \forall i,j,p
\end{aligned} \tag{2.6.5}$$

where $\Phi(y_i(t)) = \frac{\partial}{\partial u_i(t)} \frac{\partial g(y_i(t))}{\partial y_i(t)}$, noting that $u_i(t) = g(y_i(t))$. Its derivation has been included in Appendix A.2. This example portrays a 2×2 convolutive source separation, i.e. two sources and two mixtures. Nonetheless, it can be generalised to higher number of sources and mixtures, because the learning rule is local [17]. The only work on TMJ BSS employed convolutive Infomax [1], and it motivates why this 2×2 BSS technique has been utilised. Furthermore, in this technique, the filter tap lengths were assumed to be constant and no adaptive technique was proposed to mitigate the computational complexity entailed by long filter lengths. In fact, an adaptive variable tap length time-domain convolutive BSS allows to cater for fractional tap lengths, which leads to longer filter lengths [68]. This issue is clearly addressed in the following chapter. Another way to achieve statistical independence is to consider the non-Gaussianity of the sources, which is the basis of FastICA explained in the following subsection.

2.6.2 FastICA

Now, a well-known *instantaneous* ICA algorithm is described. Another approach to maximise the statistical independence of the estimated sources is maximising non-Gaussianity. FastICA is inspired by the Central Limit Theorem, in which the distribution of the sum of independent random variables tends towards a Gaussian distribution [7]. It is reasonable to assume that the distributions of the mixtures are closer to Gaussian distribution than that of any of the underlying sources. Thus, within the ICA community, the equiv-

alence between non-Gaussianity and statistical independence is understood. As a result, the main limitation of this criterion is that at most one source can possess a Gaussian distribution. In fact, most ICA algorithms rely on this assumption. The aim of ICA algorithms is to maximise statistical independence by ensuring uncorrelatedness at higher order statistics. However, if the sources are Gaussian, uncorrelatedness at second order statistics is equivalent to independence, which is not the aim of ICA [12]. Moreover, it is shown in Chapter five that non-Gaussian sources ensure the uniqueness of the BSS model.

The non-negative measure negentropy $Neg(y)$ quantifies how much a random variable (r.v.) y deviates from Gaussianity and can be formulated as [7]:

$$Neg(y) = H(\nu) - H(y) \tag{2.6.6}$$

where ν is a Gaussian r.v. of the same variance as y , and $H(.)$ denotes the differential entropy. This measure of non-Gaussianity underpins the basis of FastICA. Due to the computational complexity of negentropy, Hyvärinen *et al.* proposed instead to use an approximation [7]:

$$Neg(y) \propto [E\{G(y)\} - E\{G(\nu)\}]^2 \tag{2.6.7}$$

where $G(u)$ can be either $\cosh(u)$ or $-e^{u^2/2}$. Taking the derivative of (2.6.7) w.r.t. the separating vector $\mathbf{w}_i$, corresponding to the ith source yields:

$$\nabla \mathbf{w}_i = \alpha E\{\mathbf{z} g(\mathbf{w}_i^T \mathbf{z})\} \tag{2.6.8}$$

where $\alpha = E\{G(\mathbf{w}_i^T \mathbf{z})\} - E\{G(\nu)\}$, $\mathbf{w}_i$ corresponds to the separating vector, $\mathbf{z}$ is the whitened mixtures, and $g(.)$ the derivative of G(.). Then, Hyvärinen

et al. suggested intuitively the following fixed point iteration:

$$\mathbf{w}_{i,k+1} \leftarrow E\{\mathbf{z}g(\mathbf{w}_{i,k}^T\mathbf{z})\} \tag{2.6.9}$$

However, if the right-hand side term of equation (2.6.9) is examined carefully, it can be expressed as:

$$E\{\mathbf{z}g(\mathbf{w}_i^T\mathbf{z})\} = E\{\mathbf{VA}\mathbf{s}g(y_i)\} \tag{2.6.10}$$

where **VA** corresponds to the mixing matrix in the whitened space, s is the source vector, and $g(y_i)$ is the non-linear function on ith estimated source y_i. It turns out that **VA** also corresponds to the transpose of the separating matrix [7]. On the other hand, calculation of $E\{\mathbf{s}g(y_i)\}$ results in an *element column vector*. An element column vector can be defined as a vector that has a non-zero element at the ith location and zero elsewhere. This follows from statistical independence of the sources, i.e. the cross-correlation between y_i and s_j is zero , for $j \neq i$. Subsequently,

$$E\{\mathbf{z}g(\mathbf{w}_i^T\mathbf{z})\} = \gamma\mathbf{w}_i \tag{2.6.11}$$

where $\gamma = E\{s_i g(y_i)\}$, and thus it justifies the fixed point iteration in (2.6.9). However, the convergence of this fixed point iteration (2.6.9) is not satisfactory [7]. It is shown analytically why it does not converge to the fixed point in Chapter four, subject to some conditions. Hence, Hyvärinen *et al.* proposed to employ a Lagrangian approach to yield a convergent fixed point iteration as:

$$\mathbf{w}_i \leftarrow E\{\mathbf{z}g(\mathbf{w}_i^T\mathbf{z})\} - E\{g'(\mathbf{w}_i^T\mathbf{z})\}\mathbf{w}_i \tag{2.6.12}$$

where $g'(.)$ is the derivative of $g(.)$ w.r.t. to $\mathbf{w}_i$. FastICA has been extensively utilised in biomedical signal processing [43, 55, 69–71]. Its popularity within the biomedical BSS community stems from the fact that most natural

signals are non-Gaussian (mainly super-Gaussian). This suggests that it can be a potential candidate for the separation of TMJ sounds.

FastICA has often been strongly linked with projection pursuit [7]. Projection pursuit is a statistical tool to determine the linear combination of a multivariate data, such that the new projected data reveals the most 'interesting' direction. In the case of FastICA, the most interesting direction is in fact non-Gaussianity. As FastICA might sometimes fail to perform separation, owing to its bias towards maximal non-Gaussianity of the data, it is said to perform projection pursuit [7, 8]. Now that examples of both an instantaneous BSS algorithm and a time-domain convolutive BSS have been given, Parra's convolutive frequency-domain algorithm which exploits the non-stationarity of the sources is reviewed next.

2.6.3 Parra's frequency domain algorithm for convolutive BSS

Parra and Spence showed that a non-stationary signal possesses a heavy-tail distribution (i.e. super-Gaussianity), and therefore they exploited the non-stationarity of the sources to maximise the statistical independence [8]. For the purpose of clarity, their proof is included in Appendix A.3. Consider firstly the covariance matrix of the mixtures for an instantaneous BSS as follows:

$$\begin{aligned} \mathbf{C}_{\mathbf{x}}(t) &= E\{\mathbf{x}(t)\mathbf{x}(t)^T\} \\ &= \mathbf{A}E\{\mathbf{s}(t)\mathbf{s}(t)^T\}\mathbf{A}^T + \mathbf{C}_{\nu} \\ &= \mathbf{A}\mathbf{C}_{\mathbf{s}}\mathbf{A}^T + \mathbf{C}_{\nu} \end{aligned} \tag{2.6.13}$$

where $\mathbf{C}_{\mathbf{x}}(t)$ denotes the covariance matrix of the mixtures $\mathbf{x}(t)$, likewise for the source and noise covariance matrices $\mathbf{C}_{\mathbf{s}}(t)$ and $\mathbf{C}_{\nu}(t)$. The non-stationarity of $\mathbf{s}(t)$ implies that $\mathbf{C}_{\mathbf{s}}(t) \neq \mathbf{C}_{\mathbf{s}}(t+\tau)$. Due to the statistical independence assumption, $\mathbf{C}_{\mathbf{s}}(t)$ and $\mathbf{C}_{\nu}(t)$ are diagonal. Hence, source sep-

aration can be achieved through the diagonalisation of multiple covariances $\mathbf{C_x}(t+\tau)$, taken at different lags. Convolution in the time-domain translates to multiplication in the frequency domain. Therefore, Parra and Spence tackled convolutive BSS in the frequency domain so that they could apply the same concept to each frequency bin:

$$\mathbf{C_x}(z) = \mathbf{A}(z)\mathbf{C_s}(z)\mathbf{A}^H(z) + \mathbf{C}_\nu(z) \tag{2.6.14}$$

where z denotes the z-transform [4]. However, the process of applying ICA individually to each frequency bin poses a serious problem in terms of permutation. For example, the first estimated source in one frequency bin does not necessarily correspond to the first estimated source in another frequency bin. Parra and Spence argued that this permutation problem can be solved by imposing a smoothness constraint on the unmixing filters. In other words, the unmixing filter length $L \ll Q$, which is the discrete Fourier transform window length. According to Parra and Spence, this restricts the solutions to be continuous in the frequency domain [4], and therefore mitigates the permutation problem. Till now, techniques pertaining to only exactly-determined BSS or overcomplete BSS have been overviewed. The next step is to illustrate examples of algorithms utilised to solve underdetermined BSS, specifically, SCA algorithms.

2.6.4 Implication of K-means clustering algorithm

In under-determined source separation, $n > m$ implies that there are less number of equations than variables. Thus, UBSS is an ill-posed problem. This ill-posed condition implies that the inverse of the mixing matrix $\mathbf{A}$ is not unique or may not exist. Therefore, the best that can be achieved is to estimate the mixing matrix $\mathbf{A}$, and not the separating matrix $\mathbf{W}$. The example given in 2.5.2 demonstrates that as a result of the sparsity of the sources, clustering can be performed to estimate the mixing matrix.

In other words, the samples of the sources will lie in the direction of the columns of the mixing matrix. K-means clustering algorithm figures amongst the most conventional clustering methods, and thus has been utilised in SCA [5,11,18]. In this algorithm, it is assumed that the number of centroids, corresponding to the number of sources is known. Once these are initialised, K-means algorithm assigns samples of the mixture signals which are nearest to a particular cluster centroid, to that cluster [72]. It then updates the coordinates of the new centroid, which correspond to the mean distance between these samples and the centroid. This procedure is repeated until the coordinates of the centroid does not alter significantly. The caveat is that its performance depends on the initialisation. Thus, it is a common practice to employ the k-means algorithm a number of times on the same data and take the average of the centroids estimated from each application of the k-means algorithm. The resulting averaged centroids yield the mixing matrix. In the following sub-section, a more recent clustering technique that has been proposed to solve UBSS, is examined.

2.6.5 Li's clustering UBSS technique

Li *et al.* proposed recently a clustering technique, which operates well, provided there exists only one active source over any given time instant [42]. Consider the following 2×3 UBSS scenario, whereby the sources are sparse:

$$\mathbf{X} = \mathbf{AS}$$

$$\begin{bmatrix} Za & Kb & Lc & Ia \\ Zd & Ke & Lf & Id \end{bmatrix} = \begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix} \begin{bmatrix} Z & 0 & 0 & I \\ 0 & K & 0 & 0 \\ 0 & 0 & L & 0 \end{bmatrix} \quad (2.6.15)$$

where $\mathbf{X}$ is the mixture matrix, with each of its rows corresponding to a mixture signal, similarly for the sources $\mathbf{S}$, and $\mathbf{A}$ is the mixing matrix. If the first row of $\mathbf{X}$ is divided by its second row elementwise, then the following

is obtained:

$$[\, a/d \;\; b/e \;\; c/f \;\; a/d \,] \tag{2.6.16}$$

The first and last element of the above row vector are identical, due to the fact that they both correspond to the active source s_1, at $t = 1$ and $t = 4$. Li *et al.*'s algorithm detects those identical elements and cluster them into different classes. Each class yields an estimate of one of the columns of the mixing matrix. This technique is robust, as it only requires the sparsity of the sources to be satisfied for 10 % of the whole duration of the sources [42]. Furthermore, this technique does not need the active source to be 'on' uninterruptedly for an interval of time. Nonetheless, this technique needs the sparsity of the sources to be restricted to a single active source. Now that the two algorithms for the identification of the mixing matrix have been outlined, source recovery in UBSS is addressed in the following section.

2.6.6 ℓ_1-norm minimisation algorithm

Source recovery in the context of underdetermined source separation separation (UBSS) is quite challenging. Moreover, principal component analysis through whitening cannot be employed to solve 'half' of the BSS problem, as seen in section 2.4.2. This is because the rank of the spatial covariance matrix $\mathbf{C_x}$ of the mixture signals is less than that of the source covariance matrix $\mathbf{C_s}$. Therefore, to compensate for the lack of availability of mixture signals, it is common to assume that the sources are sparse. One common approach is to assign each sample of the mixtures at a given instant to one of the sources. However, if more than one source is active over a given period, then a technique that partially assigns each data point of the mixtures to multiple sources is more suitable. The ℓ_1-norm minimisation algorithm operates in this fashion [11]. Mathematically, it can be written as:

$$\min ||\hat{\mathbf{s}}(t)||_1 \quad s.t. \;\; \hat{\mathbf{A}}\hat{\mathbf{s}}(t) = \mathbf{x}(t) \tag{2.6.17}$$

where $||.||_1$ denotes the ℓ_1- norm, and the 'hat' superscript $\hat{(.)}$ denotes the corresponding estimate. It is worth noting that this method forms part of the linear programming techniques. As this algorithm does not depend on any statistics, it can be employed for data of short sample size such as in micro-arrays in genomics [18]. Furthermore, since it does not assume statistical independence of the sources, it can be utilised potentially for correlated sources. Nevertheless, this technique requires the sources to be sufficiently sparse (at most 'm' active sources at a given time instant t) [73]. Additionally, its prerequisite is the availability of the mixing matrix $\mathbf{A}$ or at least its estimate $\hat{\mathbf{A}}$.

Now that ICA and SCA have been reviewed and illustrated with examples of both techniques, performance measures for BSS employed in this thesis are defined next.

2.7 Performance Measures

Like most performance measures, the ones presented herein requires either the mixing matrix or the original source signals. Therefore, it restricts their usage to synthetic simulation, i.e. the sources are mixed synthetically in Monte Carlo trials run in MATLAB®. In cases whereby sources are extracted from *real* recordings, these are assessed against the estimates of convolutive Infomax, which has already been applied in the same context, i.e. TMJ source separation. Furthermore, the spectra of the estimated TMJ sources are compared to the spectra of TMJ sources available from the existing literature, to verify whether or not successful separation has been achieved.

The performance index (PI) is the dominant measure of performance in BSS, however there exist other variants to cater for the need of convolutive

or under-determined BSS. PI and its variants measure the quality of either the estimated separating matrix or the estimated mixing matrix. On the other hand, the classical mean square error (MSE) can be employed to measure the quality of the separation in terms of the estimated sources. A low value of those measures indicates good separation. In the context of audio BSS, Vincent *et al.* proposed recently a signal to interference ratio [74]. In contrast to the other measures, the higher the SIR, the better is the source separation. All these measures are defined in the following subsections. It is noted that in the context of ICA approaches, the statistical independence of the estimated sources can be quantified in terms of the mutual information. Thus, mutual information can also assist in analysing the performance of the ICA approaches.

2.7.1 The performance index for instantaneous BSS

The so-called global mixing-separating matrix is defined as $\mathbf{G} = \mathbf{WA}$. From (2.2.2), it is clear that $\mathbf{G} = \mathbf{P\Lambda}$. Therefore, $\mathbf{G}$ accounts for the scaling and the permutation resulting from the separation procedure. $\mathbf{G}$ can be utilised in a measure, which will be insensitive to permutations and scaling ambiguities. It is noteworthy to say that the minimum value of PI is zero, while its maximum value depends on the normalisation factor. However, PI is mainly used for comparison purposes, and therefore knowledge of these bounds is not crucial in assessing the performance of the BSS approaches. The performance index (PI) is such a measure and can be formulated as follows [12]:

$$\mathrm{PI} = \frac{1}{n}\sum_{i=1}^{n}\left\{\sum_{j=1}^{n}\frac{|g_{ij}|^2}{\max_k |g_{ik}|^2} - 1\right\} + \frac{1}{n}\sum_{j=1}^{n}\left\{\sum_{i=1}^{n}\frac{|g_{ij}|^2}{\max_k |g_{kj}|^2} - 1\right\} \quad (2.7.1)$$

where g_{ij} is the ijth entry of the global matrix. It is assumed that the number of sources equals to the number of mixtures.

2.7.2 The multichannel intersymbol interference for convolutive BSS

In the context of convolutive BSS, PI can be formulated as the multichannel intersymbol interference (MISI). It is defined as follows [75]:

$$MISI = \sum_{i=1}^{m} \frac{\sum_j \sum_p |\Theta_{ijp}| - \max_{j,p} |\Theta_{ijp}|}{\max_{j,p} |\Theta_{ijp}|} + \sum_{j=1}^{n} \frac{\sum_i \sum_p |\Theta_{ijp}| - \max_{i,p} |\Theta_{ijp}|}{\max_{i,p} |\Theta_{ijp}|} \tag{2.7.2}$$

where $\Theta(z) = a(z)w(z)$ denotes the global mixing-separating filter coefficient. Note the strong similarity between this performance measure and the performance index.

2.7.3 The performance measure for underdetermined BSS

The performance measure (PM) provides an indication of the difference between $\mathbf{A}$ and its estimated $\hat{\mathbf{A}}$ in the context of underdetermined BSS [76]. However, PM requires both $\mathbf{A}$ and $\hat{\mathbf{A}}$ to have unit norm columns. This index falls within $0 \leq PM \leq 1$. PM equals to 0 if $\hat{\mathbf{A}} = \mathbf{AP}$ where $\mathbf{P}$ is a permutation matrix. Therefore, the lower the PM, the better is the performance of the UBSS algorithm.

$$PM(\mathbf{A}, \hat{\mathbf{A}}) = 1 - \left(\frac{1}{2n} \sum_{i=1}^{n} \sup_j |\mathbf{A}^T \hat{\mathbf{A}}|_{ij} + \frac{1}{2n} \sum_{j=1}^{n} \sup_i |\mathbf{A}^T \hat{\mathbf{A}}|_{ij}\right) \tag{2.7.3}$$

2.7.4 Mean square error

Due to the scaling ambiguity, different BSS algorithms yield different scaled version of $\mathbf{y}_i$. For fair comparisons, all the estimated ICs are first normalised to unit variance, and then the sign of each source and its corresponding estimated IC is ensured to be the same. Thereafter, the MSE corresponding to the ith source can be computed as:

$$\text{MSE}_i = \frac{1}{T} \sum_{t=1}^{T} (s_i(t) - y_i(t))^2 \tag{2.7.4}$$

2.7.5 Signal to interference ratio

As TMJ sound separation falls under the category of audio BSS, the SIR proposed by Vincent *et al.* [74] is also considered in this thesis. It is defined in dB as [74]:

$$\text{SIR} = 10\log_{10}\frac{||s_{\text{target}}||^2}{||e_{\text{interf}}||^2} \tag{2.7.5}$$

where

$$s_{\text{target}} = \langle \hat{s}_i, s_i \rangle s_i / ||s_i||^2 \tag{2.7.6}$$

$$e_{\text{interf}} = \sum_{i' \neq i} \langle \hat{s}_i, s_{i'} \rangle s_{i'} / ||s_{i'}||^2 \tag{2.7.7}$$

and s_{target}, e_{interf} represent respectively the source of interest and the interference introduced by the other sources. It is understood that $\langle \hat{s}_i, s_j \rangle = \sum_{t=1}^{T} \hat{s}_i(t) s_j(t)$ denotes the inner product between the estimate of the ith source and the jth source. Note that if the sources are mutually orthogonal, $e_{\text{interf}} \ll s_{\text{target}}$ since $\sum_{i' \neq i} \langle \hat{s}_i, s_{i'} \rangle \ll \langle \hat{s}_i, s_i \rangle$. This leads to a large value of SIR, provided good estimates of the sources can be achieved.

2.7.6 Signal to noise ratio

The signal to noise ratio (SNR) is a well-known measure in signal enhancement techniques to assess quantitatively the strength of the restored signals in the presence of noise. SNR with respect to the ith source y_i in the presence of the ith noise v_i can be defined as follows:

$$\text{SNR} = 10\log_{10}\frac{||y_i||^2}{||v_i||^2} \tag{2.7.8}$$

2.8 Conclusions

In this chapter an overview of BSS approaches has been provided. Both the generic mixing and separating models have been deployed in Tables 2.1 and 2.2 to demonstrate the BSS problem. The two main approaches of BSS have been discussed in sections 2.4 and 2.5. In the context of TMJ BSS,

convolutive infomax has already been utilised by Guo *et al.* [1] and therefore, it has been illustrated in section 2.6.1. It has been pointed out that this particular time-domain convolutive BSS algorithm can be improved by adopting a variable tap length concept. This is the subject of the next chapter. It has also been highlighted that although FastICA is a well-known fixed point BSS algorithm, there is a lack of analytic work on the convergence of fixed point algorithms from a fixed point theory perspective. Hence, Chapter four formulates a fixed point algorithm, which coincides with FastICA. Thereafter, an analytic approach (based on fixed point concepts) of the formulated algorithm leads to an improved convergence to the solution. Furthermore, Chapter five extends FastICA to a 2×3 UBSS TMJ scenario and compares its estimates with SCA approaches presented in this chapter. Last but not least, BSS performed on real TMJ recordings is addressed in the last part of this thesis to demonstrate that BSS technique can be powerful. This shows that although an optimum BSS technique is critical, an effective method can be proposed for special scenarios such as that of the temporomandibular joint sound mixing system.

Chapter 3

VARIABLE TAP LENGTH CONVOLUTIVE BLIND SOURCE SEPARATION

3.1 Introduction

The Infomax algorithm of Bell and Sejnowski has enjoyed much success in independent component analysis (ICA) and blind source separation (BSS). The main appeal of Infomax lies in its conceptual simplicity and intuitivity [17]. This algorithm has been extended to the convolutive case by Torkkola, who details his approach in Chapter eight of [17]. Convolutive BSS figures amongst the ongoing trends in the BSS community, as modelling the acoustics of an echoic room still remains a challenge. The common approach to tackle such a problem is to treat the separation task in the frequency domain and thereby employ complex valued instantaneous ICA methods in each frequency bin. This approach suffers from a number of non-trivial issues such as permutation and scaling inconsistencies across the frequency bins. Although there has been much effort to address these issues [77–79], this is still an open problem. However, these ambiguities do not pose any significant problem to the time-domain approaches [17, 80]. Nonetheless, it may be computationally lighter to perform the source separation in frequency domain as the convolutions with long filters in time-domain are translated

to efficient multiplications in the frequency domain [17]. In the convolutive BSS model, the unmixing filter may be of fractional delays. Thus, whenever a fractional delay is encountered, a longer filter length than when there is no fractional delay is required [68]. Theoretically, the fractional delay can be modelled by the convolution of the signal to be delayed with a sinc function, which is infinite in duration. It is reasonable to assume that fractional delays can arise in the convolutive source separation.

Therefore, the objective of this chapter is to determine the optimum tap length of time-domain convolutive Infomax as a means to alleviate such computational complexity or, to avoid under-estimation of the filter length whenever the problem of fractional delays arise in BSS. Finding the optimum tap length can be achieved by adopting the fractional tap (FT) length described by Gong *et al.* in the LMS algorithm [81]. Indeed, the variable tap length concept is a relatively new idea, but has proved to be quite successful [81–84]. In blind signal processing, this idea has been introduced recently in the context of blind deconvolution [85]. This concept has not yet been applied to blind signal separation and therefore constitutes the main contribution of the present work. Moreover, in the LMS technique, the desired signal is known a priori, while the unsupervised nature of BSS makes the latter more challenging. Due to the statistical independence assumption of the sources, the optimum filter length is defined as the minimum filter length that minimises the off-diagonal terms of the covariance matrix of the sources $\mathbf{s}(t)$, which will be discussed in detail later.

The organisation of the chapter is as follows: in section 3.2, the fractional tap length LMS of Gong *et al.* is reviewed [81]. Convolutive Infomax is overviewed in section 3.3. The variable tap length Infomax is formulated in section 3.4. Section 3.5 provides simulation results to support the advantages of the new approach, followed by discussions in section 3.6. Section

3.7 concludes this work.

3.2 Variable Tap Length LMS Algorithm

The LMS algorithm is an old and well established technique. Hence, it is assumed that the reader has sufficient knowledge of this technique, and therefore, the focus is on the variable filter length concept. Otherwise, the reader can refer to [86–88]. The FT variable tap length LMS algorithm aims at finding the *minimum* filter length as the optimum filter length that fulfils the following criterion C [81]:

$$C = \xi_{L-\Delta}^{(L)} - \xi_{L}^{(L)} \leq \epsilon \tag{3.2.1}$$

where $\xi_{\ell}^{(L)}$ denotes the expected squared error, the superscript $(.)^{(L)}$ denotes that the tap length to learn the coefficients is L, whilst the subscript $(.)_{\ell}$ refers to the case when only ℓ of the L learned coefficients are employed. It is understood that Δ is a small positive integer, ϵ is a small positive value determined by the system requirements, and that:

$$1 \leq \Delta < \ell \leq L \tag{3.2.2}$$

Eq. (3.2.1) can be satisfied even in the case of $C < -\epsilon$, which will be explained next. Firstly, let L_{opt} denotes the optimum tap length with the smallest squared error between the desired signal and its estimate. $L_{opt} - (L - \Delta) < L_{opt} - L$ implies that $C < 0$, but if $L_{opt} \ll L$, then $C < -\epsilon$. Hence, the philosophy that C should be a small value is violated. Nevertheless, the definition of optimum L is considered as the *minimum* L satisfying (3.2.1), and therefore the definition of the optimum tap length stands correct.

The next straightforward step is to apply the steepest descent to minimise

C with respect to the tap length l_f. Thus, the following can be obtained as:

$$l_f(k+1) = (l_f(k) - \lambda) - \mu\left[\xi_L^{(L)} - \xi_{L-\Delta}^{(L)}\right] \tag{3.2.3}$$

where λ is the leakage parameter to avoid over-estimation of L_{opt}, and μ is the stepsize of the update equation. It is noteworthy to say that $l_f(k)$ is likely to be fractional referring to the name fractional tap length LMS algorithm [81]. Nevertheless, the tap length L employed in the LMS algorithm is not $l_f(k)$ due to its fractional nature. Instead, L is updated only when a non-trivial change of $l_f(k)$ has occurred. Mathematically, this can be expressed as:

$$L(k+1) = \begin{cases} \lfloor l_f(k) \rfloor, & \text{if } |L(k) - l_f(k)| > \delta \\ L(k), & \text{otherwise} \end{cases} \tag{3.2.4}$$

where $\lfloor . \rfloor$ denotes the floor operator and δ is a small positive integer. This concludes the variable tap length concept. For further details, refer to [81–84]. In the following section, an overview of the convolutive Infomax is given, in the context of speech processing.

3.3 Convolutive Infomax in Speech Processing

Before the convolutive Infomax [17] is introduced, it is important to examine the source signals, since the performance of BSS algorithms relies on the statistics of the sources. The temporal structure of speech does not allow us to make the assumption that the consecutive samples of speech are statistically independent. In fact, Torkkola confirms this by the statement,'*Note that speech signals violate the assumption of samples being independent*' [89]. Therefore, the convolutive Infomax which performs source separation only, and not deconvolution, is considered. Although deconvolution cancels the echoes of the same speech signal from the estimated output of a BSS system, it also results in a whitened output. Hence, all the temporal dependencies

are removed within the whitened output. This interferes with the temporal structure of the speech signal, resulting in its poor estimate. Next, the convolutive Infomax which performs only source separation is explained.

The convolutive Infomax proposed by Torkkola [17] uses a feedback network, and its output $\boldsymbol{y}(t)$ can be expressed as follows:

$$y_1(t) = \sum_{p=0}^{L_{11}} w_{11p} x_1(t-p) + \sum_{p=0}^{L_{12}} w_{12p} y_2(t-p)$$

$$y_2(t) = \sum_{p=0}^{L_{22}} w_{22p} x_2(t-p) + \sum_{p=0}^{L_{21}} w_{21p} y_1(t-p) \qquad (3.3.1)$$

where w_{ijp} denotes the corresponding separating filter coefficient, $y_j(t)$ denotes the jth estimated source at discrete time t, and $x_i(t)$ denotes the ith convolutive mixture. Let z depict the z-transform. As the filter coefficients $w_{11}(z)$ and $w_{22}(z)$ are not summed in the cross-branches of the network, they convey the deconvolution procedure. On the other hand, $w_{12}(z)$ and $w_{21}(z)$ perform source separation by removing the redundancies from the feedback adjacent branch signals $y_1(z)$ and $y_2(z)$ respectively [17]. On this basis, the solution for source separation is

$$w_{11}(z) = 1 \qquad w_{12}(z) = -a_{12}(z)a_{22}(z)^{-1}$$

$$w_{22}(z) = 1 \qquad w_{21}(z) = -a_{21}(z)a_{11}(z)^{-1} \qquad (3.3.2)$$

where $a_{ij}(z)$ denotes the corresponding mixing filter coefficient. This solution leads to $y_1(z) = a_{11}(z)s_1(z)$ and $y_2(z) = a_{22}(z)s_2(z)$ [17]. In the sequel, the cross filter coefficients can be updated as follows:

$$w_{ijp}(k+1) = w_{ijp}(k) - 2\eta(u_i(t)y_j(t-p)) \quad i \neq j \qquad (3.3.3)$$

where $u_i(t) = \tanh(y_i(t))$, tanh(.) is the so-called non-linear function, and η is the learning rate of the update. The derivation of the update is included in Appendix A.2. In the next section, the variable tap length convolutive Infomax is proposed.

3.4 Variable Tap Length Convolutive Infomax

In Eq. (3.2.1), it is assumed that the desired signal for computation of the expected squared error $\xi_\ell^{(L)}$ between the desired signal and its estimate is accessible. In this case, however, the statistical independence assumption requires that the output covariance matrix $C_\mathbf{y} = E\{\mathbf{y}(t)\mathbf{y}(t)^T\}$ tends to be diagonal. It is noted that $\mathbf{y}(t)=[y_1(t)\ y_2(t)]^T$ and $(.)^T$ denotes the transpose operation. In the LMS algorithm, the objective is to minimise $\xi_\ell^{(L)}$, while here, the aim is to minimise the off-diagonal terms of $C_\mathbf{y}$. Thus, the criterion C' is:

$$C' = \kappa_{L-\Delta}^{(L)} - \kappa_L^{(L)} \leq \epsilon \tag{3.4.1}$$

where

$$\kappa = ||C_\mathbf{y} - diag(C_\mathbf{y})||_F \tag{3.4.2}$$

with the same notations as in (3.2.1), $||.||_F$ denotes the Frobenius norm, and $diag(C_\mathbf{y})$ denotes a diagonal matrix with the diagonal elements of $C_\mathbf{y}$. κ can be regarded as the sum of the square of the off-diagonal terms of $C_\mathbf{y}$. To mitigate the computational complexity of updating the covariance matrix $C_\mathbf{y}$, as one go from sample to sample, it is practical to employ the following:

$$C_y(k+1) = \gamma C_y(k) + (1-\gamma)R_\mathbf{y} \tag{3.4.3}$$

where γ is a forgetting factor less than unity and

$$R_\mathbf{y} = \frac{1}{M}\mathbf{y}(t)_{1:M}\mathbf{y}(t)_{1:M}^T \tag{3.4.4}$$

where the subscript $(.)_{1:M}$ denotes the last M new samples of $\mathbf{y}(t)$ of the Infomax algorithm. The resulting update equation for the convolutive Infomax regarding the fractional tap length is as follows:

$$l_f(k+1) = (l_f(k) - \lambda) - \mu\left[\kappa_L^{(L)} - \kappa_{L-\Delta}^{(L)}\right] \tag{3.4.5}$$

Eq. (3.3.1) shows that L_{12} and L_{21} can be potentially of different lengths. Chapter eight of [17] states that Torkkola considered L_{12} and L_{21} to be equal. Likewise, similar assumption is made here. Although this assumption may not be necessarily true, it does not affect the effectiveness of the variable tap length approach. On the other hand, only one criterion has been utilised, and therefore, only 'one' L can be altered. This aspect is subject to further research. Hence, the variable tap convolutive Infomax is summarised in the following résumé.

Résumé of variable tap length convolutive Infomax

1. Initialise L and l_f with the same value.
2. Execute Eq. (3.3.1) for the first segment of M samples, while updating the coefficients $w_{ijp}(k)$ from sample to sample via Eq. (3.3.3).
3. Compute C_y corresponding to both $\kappa_{L-\Delta}^{(L)}$ and $\kappa_L^{(L)}$ and update the fractional length $l_f(k)$ through Eq. (3.4.5)

   For $\iota = M+1, ..., T$
4. Update the coefficient $w_{ijp}(k)$ via Eq. (3.3.3) and estimate the new samples $\mathbf{y}(t) = [y_1(t) y_2(t)]$ from Eq. (3.3.1).
5. Update C_y corresponding to both $\kappa_{L-\Delta}^{(L)}$ and $\kappa_L^{(L)}$ via (3.4.3).
6. Compute the new values of l_f and L respectively via Eqs. (3.4.5) and (3.2.4).

7. if $\iota \neq (T+1)$, where T is the number of samples of the signals, go back to step (4).

3.5 Simulations

In this section, the same simulation as Torkkola's in [17] is considered for comparison purposes. In effect, it is a 2×2 speech separation problem where the mixing matrix is given as follows:

$$a_{11}(z) = 1 - 0.4z^{-25} + 0.2z^{-45}$$

$$a_{12}(z) = 0.4z^{-20} - 0.2z^{-28} + 0.1z^{-36}$$

$$a_{21}(z) = 0.5z^{-10} + 0.3z^{-22} + 0.1z^{-34}$$

$$a_{22}(z) = 1 - 0.3z^{-20} + 0.2z^{-38}$$

The filter length parameters were set as: $L = l_f = 10$, $\delta = 2$, and $\Delta = 3$. As for the performance, the multichannel intersymbol interference (MISI) is used and defined as follows [75]:

$$MISI = \sum_{i=1}^{2} \frac{\sum_j \sum_p |G_{ijp}| - \max_{j,p} |G_{ijp}|}{\max_{j,p} |G_{ijp}|} + \sum_{j=1}^{2} \frac{\sum_i \sum_p |G_{ijp}| - \max_{i,p} |G_{ijp}|}{\max_{i,p} |G_{ijp}|} \tag{3.5.1}$$

where $G(z) = a(z) * w(z)$ denotes the global mixing-separating filter coefficient. Note the strong similarity between this performance measure and the performance index as given in [12]. In Fig. 3.1, the upper plot demonstrates the performance of the proposed method, with convolutive Infomax of different filter lengths. The lower plot shows the sum of the off-terms of covariance of $\mathbf{y}(t)$ for the corresponding filter length of Infomax. In this case, it is only the cross-correlation between $y_1(t)$ and $y_2(t)$. Fig. 3.2 illustrates the resulting global filter coefficients of the proposed approach in the upper

four plots. The lowest plot illustrates the evolution of the tap length of the proposed technique.

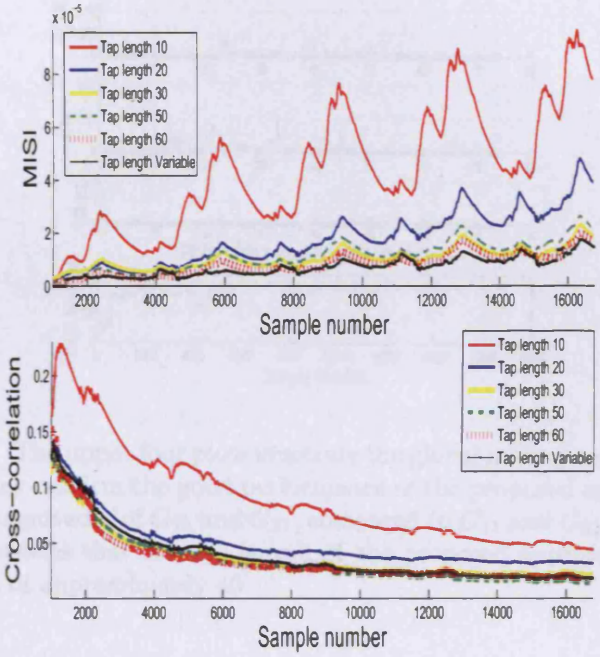

Figure 3.1. The top most plot illustrates the superior performance of the proposed approach, compared to fixed tap length Infomax. The lower plot shows the cross-correlation between the two estimated sources. It is noteworthy to say that although Infomax minimises the cross-correlation between its outputs, MISI increases as the number of iterations increases. Effectively, this implies that increasing the statistical independence does not necessarily improve the performance. It is noted that the final tap length of the proposed approach is 38.

3.6 Discussions

In Fig. 3.1, the lowest MISI achieved using the proposed approach indicates its superiority. Nonetheless, note the strong correlation between the performance of the proposed approach and Infomax of tap length 60, 50 & 30. On the other hand, there is a significant difference in performance between the proposed approach and Infomax of tap length 10 & 20. The lower plot of Fig. 3.1 follows suit, i.e. the variable tap length Infomax estimates have

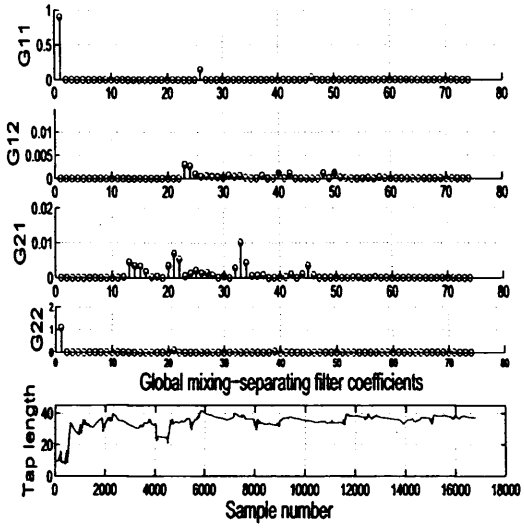

Figure 3.2. The upper four plots illustrate the global mixing-separating matrix G_{ij}. They confirm the good performance of the proposed approach, due to the low magnitudes of G_{12} and G_{21}, compared to G_{11} and G_{22}. The lowest plot demonstrates that the tap length of the proposed approach reaches a steady state of approximately 40.

approximately *similar* cross-correlation as those of Infomax of tap length 60, 50 & 30. It is noteworthy to say that although the cross-correlation between the outputs is minimised, the MISI increases, as the number of iterations increases. Effectively, this implies that increasing the statistical independence does not necessarily improve the performance. Fig. 3.2 confirms the reasonable performance of the proposed approach, due to the low magnitudes of G_{12} and G_{21}, compared to G_{11} and G_{22}. The lowest plot demonstrates that the tap length of the proposed approach reaches a steady state of approximately 40, with its final length being 38. Notice that this graph is not as 'smooth' compared to [81–83]. This 'non-smoothness' arises due to the non-stationarity of the speech signals, while in [81–83], the authors utilised stationary white Gaussian noise, instead of real signal sources. Moreover, the final tap length of 38 coefficients of the proposed approach is much smaller than the tap length of 100 coefficients employed by Torkkola [17]. Although,

the same speech signals as those of Torkkola have not been employed, this difference in tap length is not negligible. Moreover, notice that a tap length of 30 coefficients will be computationally less intensive, while achieving a similar performance as the proposed approach. However, it is highlighted that the problem at hand is 'blind', and therefore a tap length of 38 is reasonable enough. In fact, the separating filter length is expected to be near to that of $a_{12}(z)$, which is indeed the case.

3.7 Conclusions

In this chapter, it has been shown how the concept of variable tap length can be applied to blind signal separation. To the author's knowledge, this work has not yet been introduced in the existing literature. The results of this work look promising. Variable tap length assists in alleviating computational complexity of the time-domain convolutive BSS algorithms. The adaptive property can also prove useful, whenever the delay is fractional, which leads to a longer filter length than expected [68]. This work is still in its infancy, as there remains many issues to be addressed. One of these issues is to devise a technique to allow the tap length of each filter to be of different lengths. For example, another criterion can be used in parallel to the minimisation of the off-diagonal terms of the covariance such as kurtosis of the individual estimated sources. Another venue for future work is to apply the same concept to other convolutive time-domain BSS algorithms such as the natural gradient approach in [80].

Chapter 4

APPLICATION OF FERRANTE'S AFFINE TRANSFORMATION TO IMPROVE THE CONVERGENCE OF ICA FIXED POINT ITERATIONS

4.1 Introduction

Application of Ferrante's affine transformation is proposed in this chapter to improve the convergence properties of fixed point iterations within an ICA algorithm. This ICA algorithm is based upon the generalised Gaussian distribution (GGD) and Ferrante's affine transformation. Ferrante's affine transformation provides the freedom to select the eigenvalues of the Jacobian matrix of the fixed point function and thereby improves the convergence properties of the fixed point iteration. The maximum likelihood estimator of the shape parameter α has also been re-derived for the GGD assumption, subject to a unit variance constraint on the independent component (IC). At each step of the fixed point iteration an estimate of the unmixing vector to

extract one source is first found, thereafter the shape parameter α is updated by maximising the likelihood of the distribution of the estimated source with respect to α. These steps are repeated until convergence. An orthogonalisation procedure is then adopted to allow the extraction procedure to be repeated for the next source. Simulation studies verify that the proposed approach has similar performance to the *efficient* variant of FastICA proposed by Koldovský [53] in a 2×2 scenario in terms of mean-square-error (MSE) of the ICs and the performance index (PI) of the global mixing-separating matrix as applied to both synthetic and biomedical temporomandibular joint sound sources. However, the proposed approach requires only two iterations for convergence. Moreover, in simulation it is found that even when both sources are Gaussian distributed, the proposed approach achieves a good separation performance measured by both MSE and PI.

Exploitation of a fixed point iteration in ICA algorithms is quite common, highlighted by the fact that the optimisation of the well-known FastICA is performed through a form of fixed point iteration. In this study, such iterations within a particular ICA algorithm are analysed. Hence, the objectives herein are to show that 1) the fixed point iteration in such a particular ICA algorithm which exploits the GGD is unstable, 2) Ferrante's transformation can be exploited to formulate an ICA algorithm which overcomes instability at the fixed point, and 3) an associated maximum likelihood (ML) estimator of the shape parameter α can be derived.

The organisation of this chapter is as follows: in the next section,the GGD is briefly described and it is shown that the stability of the fixed point iteration in a particular ICA algorithm relies on the source distribution. Thereafter, Ferrante's transformation is utilised to provide an ICA algorithm with improved convergence properties. Additionally, the global convergence of the resulting ICA algorithm by determining an optimum parameter λ is illustrated. Next, a useful likelihood estimator of the shape parameter α which is exploited within the source extraction procedure is derived. Section 4.6

then provides two sets of simulation results: one with synthetic sources generated from GGDs, and the other one with temporomandibular joint sound (biomedical) sources. As FastICA [7] remains the dominant fixed point iteration in BSS, its performance and that of its *efficient* variant [53] are compared with the performance of the proposed algorithm. Finally, concluding remarks and discussions regarding the Ferrante method are presented in the last section of this chapter.

4.2 A fixed point ICA algorithm

4.2.1 The generalised Gaussian distribution

The generalised Gaussian distribution whose peakedness is steered by a shape parameter α, is a typical model to characterise source signals [90,91]. Exploitation of a GGD model is very popular within the signal processing community, mainly due to its generic nature to model the source signals and its algebraic convenience. Moreover, with the shape parameter $\alpha < 2$ of GGD, super-Gaussian/leptokurtic sources can be generated, while with $\alpha > 2$, sub-Gaussian/playtykurtic distributions can be modelled. In the BSS area, Koldovský *et al.* [53], and Wang *et al.* [54] incorporated independently variants of an adaptive score function based on a GGD assumption within the FastICA algorithm, while Waheed *et al.* [75] employed another variant of the GGD score function in the natural gradient algorithm. However, the focus of this study is not the GGD itself, but to analyse the convergence properties of the fixed point iteration within a particular ICA algorithm which exploits a GGD source model. The GGD family of distributions which is reviewed next, has the following form [92,93]:

$$p(s_i(t), \sigma, \alpha) = \frac{\alpha\sqrt{\beta}}{2\sigma\Gamma(1/\alpha)} e^{-|\sqrt{\beta}s_i(t)/\sigma|^\alpha} \tag{4.2.1}$$

where $s_i(t)$ denotes the i-th source, $\sigma > 0$ is the scale parameter, $\Gamma(.)$ is known as the gamma function, and $\alpha > 0$ is the shape parameter. As for

$\sqrt{\beta} = \sqrt{\frac{\Gamma(3/\alpha)}{\Gamma(1/\alpha)}}$, it is merely a scaling factor which causes $Var(s_i(t)) = \sigma^2$ where $Var(.)$ denotes variance. It is noted that in some GGD models, the factor $\sqrt{\beta}$ is omitted and therefore $Var(s_i(t)) \neq \sigma^2$ [90,91].

4.2.2 Derivation of the fixed point algorithm

In this study, the sources are assumed to be generalised Gaussian distributed. It is noteworthy to say that Mathis and Douglas made the same assumption in their work [92]. In order to derive one of the so-called separating vectors $\mathbf{w}_i$, the generic source $s_i(t)$ is substituted by its estimated one $y_i(t)$ in Eq. (4.2.1) and the natural logarithm (ln) is taken to obtain:

$$L(y_i(t), \sigma_i, \alpha_i) = \ln\left(\frac{\alpha_i\sqrt{\beta_i}}{2\sigma_i\Gamma(1/\alpha_i)}\right) - |\sqrt{\beta_i}y_i(t)/\sigma_i|^{\alpha_i} \tag{4.2.2}$$

It is common practice to impose some constraints on the scale of the estimated sources (p. 238 [47]). This is possible due to the scaling ambiguity prevailing in BSS. Similarly, the scaling parameter σ_i can be replaced by unity. Also, take note that the first term on the right hand side of (4.2.2) is independent of $y_i(t)$, and therefore the first term can be dropped. Since (4.2.2) is maximised with respect to $\mathbf{w}_i$, the following cost function can be minimised:

$$\begin{aligned} C &= \beta_i^{\alpha_i/2}|y_i(t)|^{\alpha_i} \\ &= sgn(y_i(t))^{\alpha_i}\beta_i^{\alpha_i/2}y_i^{\alpha_i}(t) \end{aligned} \tag{4.2.3}$$

using the fact that $|y_i(t)| = sgn(y_i(t))y_i(t)$ where $sgn(.)$ is the signum function. Note that $\alpha_i > 0$ implies $\beta_i^{\alpha_i/2} > 0$. Hence, the latter can be separated from the absolute value $|.|$. If C is differentiated with respect to $\mathbf{w}_i$, the

following is obtained

$$\begin{aligned}\frac{\partial C}{\partial \mathbf{w}_i} &= \alpha_i \beta_i^{\alpha_i/2} sgn(y_i(t))^{\alpha_i} y_i^{\alpha_i - 1}(t)\mathbf{z}(t) \\ &= \alpha_i \beta_i^{\alpha_i/2} sgn(y_i(t))^{2\alpha_i - 1} |y_i|^{\alpha_i - 1}(t)\mathbf{z}(t) \end{aligned} \tag{4.2.4}$$

Since $\mathbf{z}(t) = \boldsymbol{VA}s(t)$ and taking the expectations w.r.t. $y_i(t)$ on both sides to yield,

$$E\{\frac{\partial C}{\partial \mathbf{w}_i}\} = \alpha_i \beta_i^{\alpha_i/2} \mathbf{VA} E\{sgn(y_i(t))^{2\alpha_i - 1} |y_i(t)|^{\alpha_i - 1} s(t)\} \tag{4.2.5}$$

Assuming the statistical independence of the sources,

$$E\{\frac{\partial C}{\partial \mathbf{w}_i}\} = \alpha_i \beta_i^{\alpha_i/2} \mathbf{VA} E\{sgn(y_i(t))^{2\alpha_i - 1} |y_i(t)|^{\alpha_i - 1} s_i(t)\}\mathbf{e}_i \tag{4.2.6}$$

where $\mathbf{e}_j$ is the **element column vector** (of length n) that is 0 everywhere except 1 at the jth location. For example, $\mathbf{e}_3 = [0\ 0\ 1\ 0\ 0]^T$. Since $s_i(t) = sgn(s_i(t))|s_i(t)|$ and $y_i(t) \approx s_i(t)$, (4.2.6) can be simplified to

$$E\{\frac{\partial C}{\partial \mathbf{w}_i}\} = \alpha_i \beta_i^{\alpha_i/2} \mathbf{VA} E\{sgn(y_i(t))^{2\alpha_i} |y_i(t)|^{\alpha_i}\}\mathbf{e}_i \tag{4.2.7}$$

Likewise, the next simplification follows:

$$\begin{aligned} sgn(y_i(t))^{2\alpha_i} |y_i(t)|^{\alpha_i} &= sgn(y_i(t))^{\alpha_i} sgn(y_i(t))^{\alpha_i} |y_i(t)|^{\alpha_i} \\ &= sgn(y_i(t))^{\alpha_i} y_i^{\alpha_i}(t) \\ &= |y_i(t)|^{\alpha_i} \end{aligned} \tag{4.2.8}$$

Hence, the $sgn(.)$ term can be dropped from (4.2.7) and $E\{sgn(y_i(t)^{2\alpha_i})|y_i(t)|^{\alpha_i}\}$ becomes $E\{|y_i(t)|^{\alpha_i}\}$.

Using the formula $E\{|y_i(t)|^{\alpha_i}\} = \frac{\Gamma(\frac{\alpha_i+1}{\alpha_i})}{\Gamma(\frac{1}{\alpha_i})}\beta_i^{-\alpha_i/2}$ [92] along with the prop-

erty $\Gamma(a+1) = a\Gamma(a)$, one can compute:

$$E\{|y_i(t)|^{\alpha_i}\} = \frac{\beta_i^{-\alpha_i/2}}{\alpha_i} \tag{4.2.9}$$

Substituting (4.2.9) into (4.2.7) after the above simplification, one can deduce that

$$E\{\frac{\partial C}{\partial \mathbf{w}_i}\} = \mathbf{VAe}_i = \mathbf{w}_i \tag{4.2.10}$$

It is crucial to see that the gradient yields one of the columns of the whitened mixing matrix **VA**, which in fact corresponds to one of the rows $\mathbf{w}_i^T$ of the separating matrix **W** as stated in pp. 192-194 of [7]. This in turn implies the following fixed point iteration $f(\mathbf{w}_{i,k})$:

$$f(\mathbf{w}_{i,k}) : \mathbf{w}_{i,k+1} \leftarrow \alpha_i \beta_i^{\alpha_i/2} E\{sgn(y_i(t))|y_i(t)|^{\alpha_i - 1}\mathbf{z}(t)\} \tag{4.2.11}$$

with k denoting the kth iteration. To clarify (4.2.11) the expected value of (4.2.4) was computed, exploiting the fact that $|y_i(t)|^{\alpha_i} = sgn(y_i(t))^{2\alpha_i}|y_i(t)|^{\alpha_i}$ from (4.2.8) and that $sgn(y_i(t))^{-1} = sgn(y_i(t))$. In general, this iteration is repeated until $\mathbf{w}_{i,k+1} = \mathbf{w}_{i,k}$, which means that the fixed point $\mathbf{w}_i$ has been reached, for $f(\mathbf{w}_i) = \mathbf{w}_i$. On this basis, it would not be wise to take the steepest descent approach to minimise (4.2.3). This is because in such an optimisation technique, one hopes that the gradient goes to zero, as one approaches the desired point $\mathbf{w}_i$. Equation (4.2.10) shows that this is not the case, i.e. the gradient is unconditionally constant. Hence, the convergence cannot be ensured if the steepest descent is adopted. Note that Eq. (4.2.11) coincides with FastICA as (p. 188 [7]):

$$\mathbf{w}_{i,k+1} \leftarrow E\{g(\mathbf{w}_{i,k}^T\mathbf{z}(t))\mathbf{z}(t)\} \tag{4.2.12}$$

where the non-linear function $g(.)$ is equal to $sgn(y_i(t))|y_i(t)|^{\alpha_i - 1}$.

Next, the stability of the fixed point iteration $f(\mathbf{w}_{i,k})$ at the solution $\mathbf{w}_{i,k} =$

$\mathbf{w}_i$ is examined. To this end, the derivative of (4.2.11) is taken to compute the Jacobian matrix $\mathbf{J}(\mathbf{w}_{i,k}) = \nabla_{\mathbf{w}_{i,k}} f(\mathbf{w}_{i,k})|_{\mathbf{w}_{i,k}=\mathbf{w}_i}$:

$$\begin{aligned}\mathbf{J}(\mathbf{w}_i) &= \alpha_i(\alpha_i - 1)\beta_i^{\alpha_i/2} E\{sgn(y_i^{2\alpha_i}(t))|y_i(t)|^{\alpha_i-2}\mathbf{z}(t)\mathbf{z}(t)^T\} \\ &\quad \alpha_i(\alpha_i - 1)\beta_i^{\alpha_i/2} E\{|y_i(t)|^{\alpha_i-2}\mathbf{z}(t)\mathbf{z}(t)^T\}\end{aligned} \tag{4.2.13}$$

If $diag(\mathbf{e}_i)$ denotes a diagonal matrix with the vector $\mathbf{e}_i$ on the main diagonal, then by simplifying the expected value in (4.2.13), one can obtain

$$E\{|y_i(t)|^{\alpha_i-2}\mathbf{z}(t)\mathbf{z}(t)^T\} = \mathbf{V}\mathbf{A}E\{|y_i(t)|^{\alpha_i}\}diag(\mathbf{e}_i)\mathbf{A}^T\mathbf{V}^T \tag{4.2.14}$$

Utilising (4.2.9), (4.2.13), and (4.2.14), one can deduce that

$$\mathbf{J}(\mathbf{w}_i) = (\alpha_i - 1)\mathbf{V}\mathbf{A}diag(\mathbf{e}_i)\mathbf{A}^T\mathbf{V}^T \tag{4.2.15}$$

$$= (\alpha_i - 1)\mathbf{w}_i\mathbf{w}_i^T \tag{4.2.16}$$

But the matrix $\mathbf{w}_i\mathbf{w}_i^T$ is a rank one matrix with one unit eigenvalue due to the constraint $\mathbf{w}_i^T\mathbf{w}_i = 1$. Hence, there is one eigenvalue of the Jacobian matrix $\mathbf{J}(\mathbf{w}_i)$ of the fixed point iteration $f(\mathbf{w}_i)$ equal to $(\alpha_i - 1)$ and the rest are zeroes. In terms of stability of a fixed point, note that it is asymptotically stable, provided that each eigenvalue of $\mathbf{J}(\mathbf{w}_i)$ is less than unity in magnitude (p. 247 [57]) or see Appendix A.1. However, it is clear that the dominant eigenvalue of $\mathbf{J}(\mathbf{w}_i)$ depends on the value of $(\alpha_i - 1)$ and hence the stability depends on the distribution of the ith source. In order to circumvent this, the affine transformation of the fixed point iteration proposed by Ferrante *et al.* is considered [94].

4.3 Formulation of Ferrante's affine transformation in BSS

The aim of Ferrante's algorithm is to ensure that the fixed point iterations converge to the fixed point $\mathbf{w}_i$, starting from a suitable neighbourhood of $\mathbf{w}_i$ by applying an affine transformation to the fixed point algorithm $f(\mathbf{w}_{i,k})$ in hand [94]. This affine function $\rho(\mathbf{w}_{i,k}, \Lambda)$ is given as:

$$\rho(\mathbf{w}_{i,k}, \Lambda) = \mathbf{w}_{i,k+1} = \mathbf{w}_i + \Lambda(\mathbf{w}_{i,k} - \mathbf{w}_i) \tag{4.3.1}$$

$$= \Lambda\mathbf{w}_{i,k} + (I - \Lambda)\mathbf{w}_i \tag{4.3.2}$$

where $\mathbf{w}_{i,k}$ is the kth iteration solution of $\rho(\mathbf{w}_{i,k-1}, \Lambda)$ and $\mathbf{w}_i$ is the fixed point to be estimated. In (4.3.1), the term $(\mathbf{w}_{i,k} - \mathbf{w}_i)$ can be considered as the error term. Hence, it would be wise to choose Λ to be a diagonal matrix with its diagonal elements less than unity in magnitude such that the error goes to zero, as k tends to infinity.

Thereafter, Ferrante *et al.* derived the first order approximation $q(\mathbf{w}_i)$ of $f(\mathbf{w}_i)$ by employing a Taylor series expansion and substituting $\mathbf{w}_i$ by $\mathbf{w}_{i,k}$. The downfall of this substitution is the initialisation of $\mathbf{w}_{i,0}$ in the 'neighbourhood' of $\mathbf{w}_i$ for convergence. It will be shown later how convergence can be achieved, and this initialisation issue can be overcome. Hence, the final Ferrante algorithm becomes:

$$\begin{aligned}\rho(\mathbf{w}_{i,k}, \Lambda) &= \Lambda\mathbf{w}_{i,k} + (\mathbf{I} - \Lambda)q(\mathbf{w}_{i,k}) \\ &= \Lambda\mathbf{w}_{i,k} + (\mathbf{I} - \Lambda)[\mathbf{I} - \mathbf{J}(\mathbf{w}_{i,k})]^{-1}[f(\mathbf{w}_{i,k}) - \mathbf{J}(\mathbf{w}_{i,k})\mathbf{w}_{i,k}]\end{aligned} \tag{4.3.3}$$

As in [94], it is shown that the first-order 'approximant' $q(\mathbf{w}_{i,k})$ at the point $\mathbf{w}_{i,k} = \mathbf{w}_i$ yields the fixed point. However, the proof derived herein is different from that of Ferrante due to this particular ICA formulation. In this approach, the constraint $\mathbf{w}_i^T\mathbf{w}_i = 1$ is imposed, and the expected value of

$\mathbf{J}(\mathbf{w}_i)$ in (4.2.16) is considered, together with Woodbury's identity. Based on a special case of this identity (p. 535 [47]),

$$(\mathbf{A} + \mathbf{u}\mathbf{v}^T)^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\mathbf{u}\mathbf{v}^T\mathbf{A}^{-1}}{1 + \mathbf{v}^T\mathbf{A}^{-1}\mathbf{u}} \tag{4.3.4}$$

the inversion $[\mathbf{I} - \mathbf{J}(\mathbf{w}_i)]^{-1}$ can be achieved. In the context of the proposed ICA algorithm, note that $\mathbf{J}(\mathbf{w}_i) = (\alpha - 1)\mathbf{w}_i\mathbf{w}_i^T$ from (4.2.16). Thus in (4.3.4), the followings can be replaced: $\mathbf{A}$ with $\mathbf{I}$, $\mathbf{u}\mathbf{v}^T$ with $(\alpha - 1)\mathbf{w}_i\mathbf{w}_i^T$, and $\mathbf{v}^T\mathbf{A}^{-1}\mathbf{u}$ with $(\alpha - 1)\mathbf{w}_i^T\mathbf{w}_i$. Using the constraint $\mathbf{w}_i^T\mathbf{w}_i = 1$ with Woodbury's identity, one can compute

$$[\mathbf{I} - \mathbf{J}(\mathbf{w}_i)]^{-1} = \mathbf{I} + \frac{(\alpha - 1)\mathbf{w}_i\mathbf{w}_i^T}{2 - \alpha} \tag{4.3.5}$$

Employing (4.2.16), and (4.3.5) together with the fact that $f(\mathbf{w}_i) = \mathbf{w}_i$, one can then obtain

$$\begin{aligned} g(\mathbf{w}_i) &= [\mathbf{I} - \mathbf{J}(\mathbf{w}_i)]^{-1}[f(\mathbf{w}_i) - \mathbf{J}(\mathbf{w}_i)\mathbf{w}_i] \\ &= [\mathbf{I} + \frac{(\alpha - 1)\mathbf{w}_i\mathbf{w}_i^T}{2 - \alpha}][\mathbf{w}_i - (\alpha - 1)\mathbf{w}_i\mathbf{w}_i^T\mathbf{w}_i] \\ &= [\mathbf{I} + \frac{(\alpha - 1)\mathbf{w}_i\mathbf{w}_i^T}{2 - \alpha}][(2 - \alpha)\mathbf{w}_i] \\ &= \mathbf{w}_i \end{aligned} \tag{4.3.6}$$

Thus, it is clear that the formulation of Ferrante's algorithm in BSS subject to the constraint $\mathbf{w}_i^T\mathbf{w}_i = 1$ yields a desired separating vector $\mathbf{w}_i$. Now, the properties of Ferrante's algorithm are reviewed.

Theorem 1. The affine function (4.3.3) has the following two properties [94]:

$$\rho(\mathbf{w}_i,\ \Lambda) = \mathbf{w}_i \qquad \forall\ \mathbf{w}_i \tag{4.3.7}$$

$$\left.\frac{\partial \rho}{\partial \mathbf{w}_{i,k}}\right|_{\mathbf{w}_{i,k}=\mathbf{w}_i} = \Lambda \quad \forall\, \mathbf{w}_i \tag{4.3.8}$$

The proof of (4.3.7) is straightforward by using (4.3.1) and replacing $\mathbf{w}_{i,k} = \mathbf{w}_i$. However, the second property can be proved after some algebraic manipulations as in [94]. A trivial yet useful way to prove (4.3.8) is to realise that $\mathbf{w}_i$ is a fixed point, and not a variable. By taking the derivative of (4.3.2) with respect to the variable $\mathbf{w}_{i,k}$, the second property can then be proved.

Property (4.3.8) permits one to set a priori the absolute value of the eigenvalues of Λ to be less than unity so that the fixed point is asymptotically stable (p. 247 [57]). Therefore, Ferrante's algorithm has not only rendered the stability of the fixed point independent of the parameter α_i, as shown in Eq. (4.2.16), but has also given us the freedom to select the eigenvalues of the new $\mathbf{J}(\mathbf{w}_i) = \frac{\partial \rho}{\partial \mathbf{w}_{i,k}}|_{\mathbf{w}_{i,k}=\mathbf{w}_i} = \Lambda$. This paves the way to determine the optimum Λ as follows.

4.4 Global convergence of fixed point Ferrante's algorithm

In this section, the optimum Λ refers to the global convergence to the fixed point $\mathbf{w}_i$. In fixed point theory, the conventional tool to prove convergence to a fixed point manifests itself in the shape of the contraction mapping theorem (CMT) [57]. This is reviewed briefly in the following paragraph.

4.4.1 Contraction mapping theorem

According to CMT, a vector function $F(\mathbf{u})$ is a contraction within $[\mathbf{w}_{lo}, \mathbf{w}_{up}] \in \mathbb{R}^n$ if [57], [2]:

- $\mathbf{u} \in [\mathbf{w}_{lo}, \mathbf{w}_{up}] \in \mathbb{R}^n \Rightarrow F(\mathbf{u}) \in [\mathbf{w}_{lo}, \mathbf{w}_{up}]$
- $\exists \lambda < 1 \in \mathbb{R}^+$ subject to $||F(\mathbf{u}) - F(\mathbf{v})||_2 \leq \lambda ||\mathbf{u} - \mathbf{v}||_2 \; \forall \mathbf{u}, \mathbf{v} \in [\mathbf{w}_{lo}, \mathbf{w}_{up}]$

where $||.||_2$ denotes the Euclidean norm. The theory of CMT is now applied to prove the convergence of Ferrante's algorithm in BSS.

4.4.2 The lower and upper bounds

Consider the closed region $[\mathbf{w}_{i,lo}, \mathbf{w}_{i,up}]$ over which the Ferrrante fixed point algorithm contracts (for further details of the contraction mapping theorem of fixed points, refer to pp. 245-246 [57] & pp. 629-631 [2]). Likewise, the affine function $\rho(\mathbf{w}_i, \Lambda)$ in (4.3.1) needs to be proved to contract the upper and lower bounds of the closed region $[\mathbf{w}_{i,lo}, \mathbf{w}_{i,up}]$. To this end, the following pair of conditions have to be satisfied (p. 119 [57]):

$$\mathbf{w}_{i,lo} - \rho(\mathbf{w}_{i,lo}, \mathrm{A}) \leq \mathbf{0} \tag{4.4.1}$$

$$\mathbf{w}_{i,up} - \rho(\mathbf{w}_{i,up}, \Lambda) \geq \mathbf{0} \tag{4.4.2}$$

In the first place, examine the lower bound condition (4.4.1) inline with (4.3.1) and substitute $\mathbf{w}_i$ with its estimate $\hat{\mathbf{w}}_i$:

$$\begin{aligned} \mathbf{w}_{i,lo} - [\hat{\mathbf{w}}_i + \Lambda(\mathbf{w}_{i,lo} - \hat{\mathbf{w}}_i)] &\leq \mathbf{0} \\ (\mathbf{I} - \Lambda)\mathbf{w}_{i,lo} - (\mathbf{I} - \Lambda)\hat{\mathbf{w}}_i &\leq \mathbf{0} \\ (\mathbf{I} - \Lambda)(\mathbf{w}_{i,lo} - \hat{\mathbf{w}}_i) &\leq \mathbf{0} \end{aligned} \tag{4.4.3}$$

For simplicity, consider the case where $\Lambda = \lambda$ is a scalar, subject to the constraints $(1 - \lambda) > 0$ and $||\mathbf{w}_{i,k}||_2 = 1$. The constraint $(1 - \lambda) > 0$ is a consequence of fixed point stability of theory of an attractive fixed point (p. 247 [57]), while $||\mathbf{w}_{i,k}||_2 = 1$ is quite common in ICA to bound the solution. Taking the Euclidean norm on both sides of (4.4.3), it can be seen that

$$(1 - \lambda)||(\mathbf{w}_{i,lo} - \hat{\mathbf{w}}_i)||_2 \leq 0 \tag{4.4.4}$$

Examine the equality in (4.4.4), i.e.

$$(1-\lambda)||(\mathbf{w}_{i,lo}-\hat{\mathbf{w}}_i)||_2 \approx 0 \tag{4.4.5}$$

For the condition (4.4.5) to be fulfilled, it is required that either $\mathbf{w}_{i,lo} = \hat{\mathbf{w}}_i$ or $(1-\lambda) \rightarrow 0$. Since one does not have control over the lower bound $\mathbf{w}_{i,lo}$, because one does not know a priori the fixed point $\mathbf{w}_i$, therefore λ has to be altered such that $(1-\lambda) \rightarrow 0$ subject to the constraint $(1-\lambda) > 0$. Therefore, it can be deduced that the optimum $\lambda \rightarrow 1$. A consequence of $\lambda \rightarrow 1$ on the upper bound (4.4.2) makes the latter equal to zero. Additionally, $\lambda \rightarrow 1$ slows down the rate of convergence. This is clear from (4.3.1), and the fact that the error term is $(\mathbf{w}_{i,k} - \mathbf{w}_i)$. In practice, λ can be initialised to less but near unity and thereafter λ can be switched to lower values to speed up the convergence in the neighbourhood of $\mathbf{w}_i$. How to determine the neighbourhood of $\mathbf{w}_i$ remains an open problem and is subject to further investigation. Nevertheless, simulation studies show that the proposed approach requires two iterations for convergence. Thus, its rate of convergence does not pose any problem. On the basis of these theoretical aspects of Ferrante's affine transformation, the implementation issues of the proposed algorithm are next addressed.

4.5 Implementation issues of Ferrante's algorithm in BSS

4.5.1 The update equation

Recall that Ferrante's algorithm is described as

$$\rho(\mathbf{w}_{i,k},\ \lambda) = \lambda\mathbf{w}_{i,k} + (1-\lambda)[\mathbf{I} - \mathbf{J}(\mathbf{w}_{i,k})]^{-1}[f(\mathbf{w}_{i,k}) - \mathbf{J}(\mathbf{w}_{i,k})\mathbf{w}_{i,k}] \tag{4.5.1}$$

One may take the expected value of $q(\mathbf{w}_{i,k})$, that is, $[\mathbf{I}-\mathbf{J}(\mathbf{w}_{i,k})]^{-1}[f(\mathbf{w}_{i,k}) - \mathbf{J}(\mathbf{w}_{i,k})\mathbf{w}_{i,k}]$ with respect to the random variables $z(t)$ and $y(t)$ accordingly.

This is due to the dependence of the Jacobian $\mathbf{J}(\mathbf{w}_{i,k})$ and $f(\mathbf{w}_{i,k})$ on these random variables. Consequently, the following might be erroneous to formulate:

$$E\{[\mathbf{I}-\mathbf{J}(\mathbf{w}_{i,k})]^{-1}[f(\mathbf{w}_{i,k})-\mathbf{J}(\mathbf{w}_{i,k})\mathbf{w}_{i,k}]\} = [\mathbf{I}-E\{\mathbf{J}(\mathbf{w}_{i,k})\}]^{-1}[E\{f(\mathbf{w}_{i,k})\}-E\{\mathbf{J}(\mathbf{w}_{i,k})\mathbf{w}_{i,k}\}] \tag{4.5.2}$$

The statistical dependence between $z(t)$ and $y(t)$ involved in $\mathbf{J}(\mathbf{w}_{i,k})$ and $f(\mathbf{w}_{i,k})$ prohibits factorising out the expected values as in the right hand side term of the above expression. However, in demonstrating that $q(\mathbf{w}_i) = \mathbf{w}_i$ in (4.3.6), the expected values of $\mathbf{J}(\mathbf{w}_i)$ and $f(\mathbf{w}_i)$ in (4.2.11) and (4.2.13) were assumed to be computed exactly. Hence, one can compute the right hand side term of (4.5.2) to estimate $\mathbf{w}_i$. To invert $[\mathbf{I} - E\{\mathbf{J}(\mathbf{w}_{i,k})\}]^{-1}$, first consider (4.3.5) and substitute $\mathbf{w}_i$ with $\mathbf{w}_{i,k}$. Then, the term $(\alpha-1)\mathbf{w}_{i,k}\mathbf{w}_{i,k}^T$ is replaced by the expected value of $\mathbf{J}(\mathbf{w}_{i,k})$ w.r.t. $y_i(t)$ in (4.2.13) to obtain

$$[\mathbf{I} - \mathbf{J}(\mathbf{w}_i)]^{-1} = \mathbf{I} + \frac{\alpha_i(\alpha_i - 1)\beta_i^{\alpha_i/2}E\{|y_i(t)|^{\alpha_i-2}\mathbf{z}(t)\mathbf{z}(t)^T\}}{2-\alpha_i} \tag{4.5.3}$$

Eqs. (4.2.11) and (4.2.13) can be utilised to calculate $[E\{f(\mathbf{w}_{i,k})\}-E\{\mathbf{J}(\mathbf{w}_{i,k})\mathbf{w}_{i,k}\}]$, as follows:

$$E\{f(\mathbf{w}_{i,k})\} - E\{\mathbf{J}(\mathbf{w}_{i,k})\mathbf{w}_{i,k}\}$$

$$= \alpha_i\beta_i^{\alpha_i/2}E\{sgn(y_i(t))|y_i(t)|^{\alpha_i-1}\mathbf{z}(t)\} - \alpha_i(\alpha_i-1)\beta_i^{\alpha_i/2}E\{|y_i(t)|^{\alpha_i-2}\mathbf{z}(t)\underbrace{\mathbf{z}(t)^T\mathbf{w}_{i,k}}_{sgn(y_i(t))|y_i(t)|}\}$$

$$= \alpha_i(2-\alpha_i)\beta_i^{\alpha_i/2}E\{sgn(y_i(t))|y_i(t)|^{\alpha_i-1}\mathbf{z}(t)\} \tag{4.5.4}$$

Hence, the weight update equation of the final Ferrante BSS algorithm can be expressed as:

$$\mathbf{w}_{i,k+1} = h(\mathbf{w}_{i,k},\ \lambda) = \lambda \mathbf{w}_{i,k} + \alpha_i(2-\alpha_i)\beta_i^{\alpha_i/2}(1-\lambda)E\{sgn(y_i(t))|y_i(t)|^{\alpha_i-1}\mathbf{z}(t)\}\mathbf{G} \tag{4.5.5}$$

where

$$\mathbf{G} = [\mathbf{I} + \frac{\alpha_i(\alpha_i-1)\beta^{\alpha_i/2}E\{|y_i(t)|^{\alpha_i-2}\mathbf{z}(t)\mathbf{z}(t)^T\}}{2-\alpha_i}] \tag{4.5.6}$$

Until now, it was assumed that the shape parameter α_i can be accessed, whilst in BSS this is not available. Thus, determining α_i blindly constitutes the topic of the next section.

4.5.2 Estimation of the shape parameter α

The update equation (4.5.5) requires the shape parameter α. Much work has already been undertaken to estimate α [91, 93, 95]. Subsequently, the aim of this chapter is not to challenge such work, but to devise a pragmatic and robust estimator of α to fulfil the need of Ferrante's algorithm.

The most classic method is the maximum likelihood (ML) estimator which has been studied thoroughly and advocated by Varanasi [93]. Likewise, Vetterli found experimentally that the ML estimator needs only three iterations for an accuracy of the order 10^{-6} [90]. In addition, the ML estimator still outperforms Song's algorithm [91] and other techniques [93]. Furthermore, as for any ML estimator, it is asymptotically unbiased and efficient. The contribution here is to propose to use an ML estimator with the assumption $\sigma = 1$ (the scaling ambiguity is exploited).

The same approach as Vetterli [90] is taken here to derive the shape parameter α. However, note that, the term $\sqrt{\beta}$ has been included as in (4.2.1) such that $\sigma^2 = Var(.)$, whilst in [90] the author omitted $\sqrt{\beta}$ and therefore $\sigma^2 = Var(.)/\beta$. Hence, σ can be accessed through the estimation of $Var(.)$,

without the a priori knowledge of α. On the other hand, the motivation that lies behind $\sigma = 1$ is to leapfrog its estimation which is irrelevant in BSS. Thus, after each iteration of (4.5.5), the variance of $y_i(t)$ is normalised to unity. Now, the ML estimator is derived.

Firstly, define the likelihood function of the signal vector $\mathbf{u}=[u(1)u(2)...u(t)..u(T)]$ having independent samples as

$$\mathrm{L}(\mathbf{u};\sigma,\alpha) = \log \prod_{j=1}^{T} p(u(j);\sigma,\alpha) \tag{4.5.7}$$

If Eq. (4.5.7) is differentiated with respect to α, then the following can be obtained (4.5.8). The derivation is provided in Appendix B,

$$\frac{\partial \mathrm{L}(\mathbf{u};\sigma,\alpha)}{\partial \alpha} = T\left(\frac{1}{\alpha} + \frac{3}{2\alpha^2}\Psi(1/\alpha) - \frac{3}{2\alpha^2}\Psi(3/\alpha)\right) - \sum_{j=1}^{T} \frac{\partial O_j}{\partial \alpha} \tag{4.5.8}$$

where

$$\frac{\partial O_j}{\partial \alpha} = O_j \gamma_j \tag{4.5.9}$$

$$O_j = \left[\sqrt{\frac{\Gamma(3/\alpha)}{\Gamma(1/\alpha)}}|u(j)|\right]^{\alpha} \tag{4.5.10}$$

$$\gamma_j = \left\{\frac{1}{2}\left[\frac{\Psi(1/\alpha) - 3\Psi(3/\alpha)}{\alpha} + \log\left(\frac{\Gamma(3/\alpha)}{\Gamma(1/\alpha)}\right)\right] + \log|u(j)|\right\} \tag{4.5.11}$$

and $\Psi(i) = \Gamma'(i)/\Gamma(i)$ is the digamma function. Therefore, (4.5.8) is equated to zero to determine α. However, a closed form solution does not exist, therefore consider:

$$\mathrm{h}(\alpha) = \frac{\partial \mathrm{L}(\mathbf{u};\sigma,\alpha)}{\partial \alpha} \tag{4.5.12}$$

Employing the Newton-Raphson optimisation technique as in [90], one can find α iteratively using:

$$\alpha_{k+1} = \alpha_k - \frac{\mathrm{h}(\alpha_k)}{\mathrm{h}'(\alpha_k)} \tag{4.5.13}$$

where

$$\begin{aligned} \mathrm{h}'(\alpha) &= \frac{\partial^2 \mathrm{L}}{\partial \alpha^2} \\ &= T\xi' + \sum_{j=1}^{T}[O_j(\gamma_j^2 + \gamma_j')] \end{aligned} \tag{4.5.14}$$

$$\xi' = \frac{1}{\alpha^3}\Big(-\alpha + 3(\Psi(3/\alpha) - \Psi(1/\alpha)) + \frac{3}{2\alpha}(3\Psi(3/\alpha)' - \Psi(1/\alpha)')\Big) \tag{4.5.15}$$

$$\gamma_j' = \frac{1}{2\alpha^3}\Big(9\Psi(3/\alpha)' - \Psi(1/\alpha)'\Big) \tag{4.5.16}$$

Similarly, $\Psi(.)'$ denotes the trigamma function and the derivation of $\mathrm{h}'(\alpha) = \frac{\partial^2 \mathrm{L}}{\partial \alpha^2}$ is also included in Appendix B. As the issues concerning the implementation of Ferrante's algorithm have now been addressed, the overall algorithm can now be concluded.

4.5.3 Résumé of the proposed approach

The proposed method based on Ferrante's affine transformation is summarised as:

1. Whiten the mixtures $\mathbf{X}$ to yield $\mathbf{Z}$ as shown in Chapter 2.

2. Initialise randomly $\mathbf{w}_i$, and set $\alpha = 1$ at iteration k=0. Furthermore, set the parameter λ in (4.5.5) to 0.999 to satisfy the global convergence condition (4.4.5).

3. Compute the i-th estimated source vector $\mathbf{y}_i = \mathbf{w}_{i,k}^{\mathrm{T}}\mathbf{Z}$.

For the k-th iteration:

4. Update $\mathbf{w}_{i,k}$:

$$\mathbf{w}_{i,k} = \lambda \mathbf{w}_{i,k-1} + (1-\lambda) q(\mathbf{w}_{i,k-1})$$

where

$$q(\mathbf{w}_{i,k}) = \alpha_{i,k}(2-\alpha_{i,k})\beta_{i,k}^{\alpha_{i,k}/2} E\{sgn(y_i(t))|y_i(t)|^{\alpha_{i,k}-1}\mathbf{z}(t)\}\mathbf{G}$$

and

$$\mathbf{G} = [\mathbf{I} + \frac{\alpha_{i,k}(\alpha_{i,k}-1)\beta_{i,k}^{\alpha_{i,k}/2} E\{|y_i(t)|^{\alpha_{i,k}-2}\mathbf{z}(t)\mathbf{z}(t)^T\}}{2-\alpha_{i,k}}]$$

5. For $i \neq 1$, orthogonalize $\mathbf{w}_{i,k}$ with respect to the previously estimated separating vectors $\mathbf{w}_j$, where $j < i$, by:

$$\mathbf{w}_{i,k} = \mathbf{w}_{i,k} - \sum_{j=1}^{i-1} (\mathbf{w}_{i,k}^T \mathbf{w}_j)\mathbf{w}_j \tag{4.5.17}$$

6. Calculate $\mathbf{y}_i$ and normalise $\mathbf{y}_i$ to unit variance and $\mathbf{w}_i$ by $\mathbf{w}_i/||\mathbf{w}_i||_2$.

7. Estimate $\alpha_{i,k}$ by optimising the Newton-Raphson algorithm given in equation (4.5.13).

8. If $\alpha_{i,k} = \alpha_{i,k-1}$, then convergence is attained. Therefore at convergence, the final estimated separating vector of the ith source is $\mathbf{w}_i = \mathbf{w}_{i,k}$. Otherwise, return to step 4.

For step 4, it is understood that the $E\{.\}$ values are replaced by their time-averages of $\mathbf{y}_i$, which denotes the i-th estimated source vector, and is appropriate for ergodic sources. On the other hand, the discrete time-instant $t = k$ is considered. As for step 5, this avoids $\mathbf{w}_i$ converging to the same solution as $\mathbf{w}_j$ [7]. Now that the résumé of the proposed approach has been outlined,

simulation studies are presented in the following section to demonstrate its efficacy and convergence.

4.6 Simulations

Ferrante's algorithm is quite robust only for a 2 × 2 BSS scenario, i.e. two sources and two mixtures. Although the 2 × 2 BSS scenario might seem trivial, Ferrante's algorithm has similar performance to *efficient* FastICA [53], and outperforms FastICA [7] in most cases. Furthermore, simulation studies suggest that the proposed scheme has attractive convergence properties in terms of convergence rate, and its stability at convergence. An example of a 2 × 2 noiseless BSS scenario by simulating BSS of temporomandibular joint (TMJ) sounds is provided. Consequently, two scenarios are presented: 1) two white noise sources synthetically generated from GGD, and 2) real TMJ sources were considered. 50 Monte Carlo simulations were run. For each scenario, a different mixing matrix $\mathbf{A}$ was generated randomly (elements from a standardised normal distribution). It can be claimed that due to the equivariance property of the noiseless BSS, i.e. the mixing matrix should have no effect on the resulting separation (p. 248 [7]). Thus, the synthetic sources for each Monte Carlo simulation were re-generated randomly. Both the mixing matrix and the synthetic sources were randomly re-generated in scenario one. For comparison purposes, the FastICA [7], which is the most well-known fixed point iteration for ICA, and its efficient variant [53] are taken as benchmarks. FastICA achieves statistical independence of the outputs by maximising the non-Gaussianity of the ICs. On the other hand, its efficient variant [53] employs FastICA as a preprocessing step to gain estimates of the sources, and thereafter applies adaptive non-linear functions on each estimated IC and finally employing a refinement step to improve the results of the FastICA.

In terms of the performance measures for all three algorithms, the well-

established mean-square-error (MSE) of the estimated ICs $\mathbf{y}_i$ and the performance index (PI) of the global mixing-separating matrix are considered. To monitor the convergence of each algorithm, $\alpha_{i,k}$ of $\mathbf{y}_i$ is estimated after each update of $\mathbf{w}_{i,k}$ in each scenario. This is illustrated in Fig. 4.1 for scenario one, and Figs. 4.2 & 4.3 for scenario two.

4.6.1 Scenario one: Sources generated from GGD

Table 4.1 summarises the results in terms of MSE when the sources were generated from the GGD, while Table 4.2 provides the results for the same corresponding experiments, but with PI as the performance measure. Furthermore, a graph monitoring the values of α_i for each algorithm is included for convergence analysis in Fig. 4.1. This graph corresponds to experiment one of Table 4.1 & 4.2.

4.6.2 Scenario two: The temporomandibular joint BSS

The scenario whereby TMJ sources are mixed linearly is simulated. A hard TMJ source normally refers to a mature stage of TMD of the corresponding joint, while the joint generating a soft TMJ sound source is not severely affected. In this scenario, such a situation is considered, i.e. when a hard TMJ source is mixed with a soft one. Table 4.3 provides two sets of results: one when click TMJ sources were considered, and the other one illustrating the crepitus BSS scenario in terms of MSE. Likewise, Table 4.4 provides the results for the same corresponding experiments as Table 4.3 with PI as the performance measure. Figs. 4.2 & 4.3 illustrate the progression of α_i of click and crepitus ICs respectively.

Based on these simulation results, the performance of these algorithms is discussed next. In the sequel, the overall work based on Ferrante's affine transformation is concluded.

Table 4.1. Summary of the results when the sources were generated from the GGD. MSE stands for mean square error, while Super refers to super-Gaussian source, Gaus denotes Gaussian source, and Sub stands for sub-Gaussian source.

Expt no	Sources	α_i	MSE $\times 10^{-3}$ Fast	MSE $\times 10^{-4}$ Efficient Fast	MSE $\times 10^{-5}$ Ferrante
1	Super	0.5	0.208	0.020	0.4175
	Sub	5	0.237	0.127	0.4175
2	Super	0.5	0.146	0.026	0.420
	Gaus	2	0.149	0.171	0.420
3	Super	0.5	0.192	0.019	0.681
	Super	1.2	0.194	0.172	0.972
4	Sub	3	0.067	0.282	0.446
	Sub	5	0.7539	0.184	0.446
5	Sub	5	0.223	0.247	0.468
	Gaus	2	0.253	0.400	0.468
6	Gaus	2	-	1680	2.73
	Gaus	2	-	1680	2.64

Table 4.2. Summary of the results when the sources were generated from the GGD. PI stands for Performance Index, while Super refers to super-Gaussian source, Gaus denotes Gaussian source, and Sub stands for sub-Gaussian source.

Expt no	Sources	PI $\times 10^{-4}$ Fast	PI $\times 10^{-5}$ Efficient Fast	PI $\times 10^{-6}$ Ferrante
1	Super Sub	4.45	1.46	8.35
2	Super Gaus	2.96	1.96	8.40
3	Super Super	3.86	1.91	16.5
4	Sub Sub	1.43	4.66	8.92
5	Sub Gaus	4.78	6.47	9.36
6	Gaus Gaus	-	45250	9.33

Table 4.3. Summary of the results when TMJ sources were considered. MSE denotes mean square error, with H clk as hard click, S clk as soft click, H cre as hard crepitus, and S cre as soft crepitus.

Expt no	Sources	α_i	MSE Fast	MSE Efficient Fast	MSE Ferrante
1	H clk	0.11	0.837×10^{-6}	0.890×10^{-6}	0.834×10^{-6}
	S clk	0.25	0.293×10^{-6}	0.461×10^{-6}	0.293×10^{-6}
2	H cre	0.68	0.326×10^{-3}	0.326×10^{-3}	0.181×10^{-4}
	S cre	0.57	0.155×10^{-3}	0.010×10^{-3}	0.230×10^{-4}

Table 4.4. Summary of the results when TMJ sources were considered. PI denotes performance index, with H clk as hard click, S clk as soft click, H cre as hard crepitus, and S cre as soft crepitus.

Expt no	Sources	PI Fast	PI Efficient Fast	PI Ferrante
1	H clk			
	S clk	1.11×10^{-8}	1.97×10^{-7}	8.99×10^{-9}
2	H cre			
	S cre	4.75×10^{-4}	3.29×10^{-4}	3.50×10^{-5}

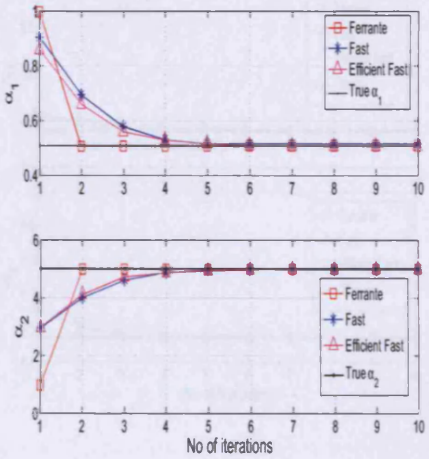

Figure 4.1. From top to bottom: progression of the shape parameter α of the super-Gaussian IC, followed by that of the sub-Gaussian source from experiment one of table 4.1. Both ICs were synthetically generated from GGD. Note that Ferrante's algorithm requires two iterations for convergence, whilst efficient FastICA and FastICA have similar convergence with more number of iterations. Also, notice the stability of all three algorithms at convergence, i.e. there are no fluctuations at the steady state.

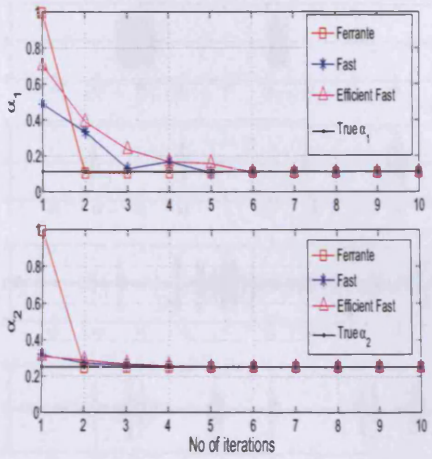

Figure 4.2. From top to bottom: progression of the shape parameter α of the hard click, followed by that of the soft click from experiment one of table 4.2. Note that Ferrante's algorithm requires two iterations for convergence, whilst efficient FastICA and FastICA have similar convergence with more number of iterations. Regarding α_1, FastICA suffers from a slight instability between iteration 3 & 5. The closeness of α_1 to zero might explain this observation.

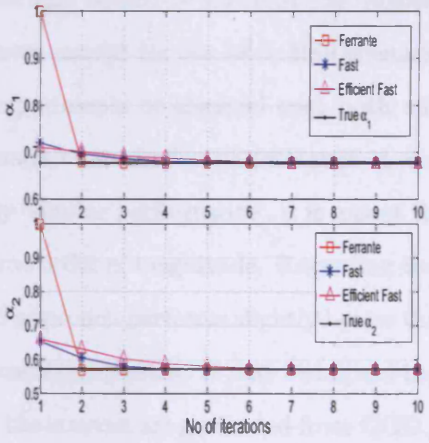

Figure 4.3. From top to bottom: progression of the shape parameter α of the hard crepitus, followed by that of the soft crepitus from experiment one of table 4.2. Notice that α_i of Ferrante's algorithm has always been initialised to unity, while the α_i of estimates of both FastICA and its efficient variant in the first iteration is much closer to the true value. Nonetheless, note that Ferrante's algorithm requires two iterations for convergence, whilst the efficient FastICA and FastICA have similar convergence with more number of iterations.

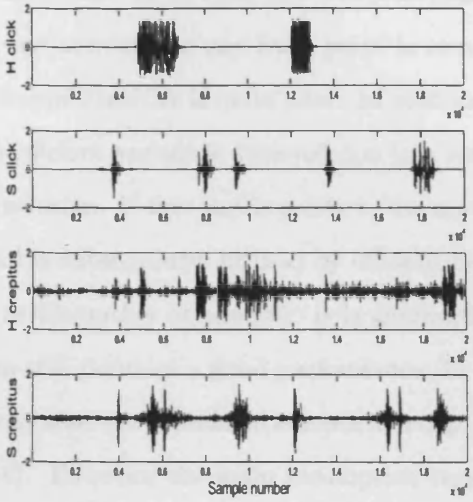

Figure 4.4. From top to bottom: hard click, soft click, hard crepitus and soft crepitus. Note that the click is active for short and *distinct* periods, while the crepitus is a more noise-like signal.

4.7 Discussion and concluding remarks

As expected, the efficient variant of FastICA has outperformed the original FastICA in most cases except for the click BSS scenario in terms of MSE. In the first three experiments of scenario one, both efficient FastICA and the proposed approach have similar performance in extracting the super-Gaussian source. By 'similar performance', it is meant that the performance measure is of the same order of magnitude. Regarding the same three experiments, the proposed approach performs slightly better than efficient FastICA in extracting the second component. It may be argued that this algorithm already assumes that the sources are generated from GGD. Nonetheless, in the case of the second scenario (table 4.4), the proposed algorithm still performs slightly better than efficient FastICA. On the other hand, consider the sixth experiment (expt no 6) of the first scenario in Table 4.1, where both sources are Gaussian. The results of FastICA for this particular experiment have been omitted because in some Monte Carlo runs, FastICA did not converge to any fixed point. In ICA, due to the statistical independence assumption,

it is known that there must be *at most* one Gaussian source. This explains why FastICA did not converge to any fixed point in some cases, while the performance of efficient FastICA is quite poor. In contrast to FastICA, the convergence of its efficient variant is achieved due to a test of saddle points on the estimated solution. If this test is positive, the separating matrix $\mathbf{W}$ is then rotated and is subsequently utilised by efficient FastICA to find the 'correct' solution as the author argues [53]. It is noteworthy to say that the proposed approach still provides a good performance. In deriving the algorithm, the only steps where it is assumed statistical independence are in Eqs. (4.2.6) and (4.2.14). However, the main assumption regarding the sources is their GGDs as in [92]. This assumption might explain why the proposed algorithm is still quite robust when both sources are Gaussian. Moreover, experiment six in Table 4.1 highlights the fact that even when both sources have similar distribution (i.e. they both have equal α_i), the proposed algorithm is quite robust. It can be shown that it has similar performance for other values of α_i when $\alpha_1 = \alpha_2$. In the case of click, note that all three algorithms have excellent performances. One of the main reason for this is due to the fact the click itself is sparse, i.e. its magnitude is approximately zero most of the time, which means that there is no effective mixing of the two clicks when both clicks are temporally 'off' (i.e. zero magnitude). For this particular case, note that MSE suggests that all three algorithms have similar performance, but PI suggests otherwise as shown in tables 4.3 & 4.4 respectively. MSE might be misleading, because the clicks are highly sparse as seen in Fig. 4.4. Therefore, the active periods of the clicks accounts for most of the MSE, while the PI is independent of the nature of the signal and hence it provides a better indication as a performance measure. Additionally, the TMD BSS scenario demonstrates that even though the sources are not generated from GGD, the proposed algorithm has still a reasonable performance. The main drawback of the algorithm is its robustness only for a 2×2 BSS scenario. This is open to discussion and therefore needs further

investigation.

Nevertheless, the main focus of this study was not to derive a better BSS algorithm than FastICA or its efficient variant, but to demonstrate that by analysing the convergence properties of a fixed point ICA algorithm, one can render an unstable algorithm stable. Ferrante's affine transform offers certainly one of the solutions. In fact, it is confirmed that the proposed method requires only two iterations for convergence for all the presented simulations (as seen in Figs. 4.1, 4.2, and 4.3). A slow convergence was expected because the rate of convergence decreases as $\lambda \rightarrow 1$ [94]. The λ was set to 0.999. In Ferrante's work, $\mathbf{w}_i$ does not depend on any other variables. But in the proposed algorithm, $\mathbf{w}_i$ implicitly relies on y_i whose distribution has been exploited in the form of $E\{\mathbf{J}(\mathbf{w}_i)\}$ and $E\{f(\mathbf{w}_i)\}$. This extra information might explain why the approach converges in only two iterations. Furthermore, Ferrante assumed that $\mathbf{w}_{i,0}$ was initialised in the neighbourhood of $\mathbf{w}_i$ for convergence. In section 4.4.2, this difficulty was circumvented, i.e. the a priori knowledge of the neighbourhood of $\mathbf{w}_i$ by selecting an appropriate λ to satisfy the convergence condition (4.4.5). From Figs. 4.1, 4.2, and 4.3, also note the stability of all three algorithms at convergence, i.e. there are no fluctuations of α_i at the steady state.

In conclusion, Ferrante's algorithm can offer an alternative to overcome the instability of fixed point ICA algorithms. The fixed point algorithm in (4.2.11) is very similar to that of FastICA as in (4.2.12). Likewise, it was shown that the stability of (4.2.11) relies on the distribution of the sources as in (4.2.16), while the author of FastICA found the instability of FastICA in the form of (4.2.12) (p. 188 [7]). Subsequently, he applied a Lagrangian approach to overcome this instability. In contrast to his approach, the contraction mapping and concepts such as attractive fixed points (p. 247 [57]) have been applied to analyse this instability. Last but not least, the ML estimator of the shape parameter α_i has been re-derived, which might be useful for other BSS algorithms where this parameter is required.

Chapter 5

UNDERDETERMINED BLIND IDENTIFICATION OF TMJ SOURCES

5.1 Introduction

In this chapter, a particular 2×3 underdetermined blind source separation problem using a filtering approach is addressed. More specifically, the case where a pair of temporomandibular joint (TMJ) *sparse* sound sources prevail in the presence of a third *non-sparse* source modelled as noise is considered. If the number of mixtures is less than the number of sources, the problem is termed underdetermined BSS (UBSS). Firstly, it is required to *identify* the mixing matrix $\mathbf{A}$ of size $m \times n$, and thereafter source extraction is performed to solve fully UBSS. However, UBSS is quite a challenging task since 1) $n > m$ implies that there are less number of equations than variables, and therefore an UBSS is an ill-posed problem, and 2) no explicit a priori knowledge of $\mathbf{A}$ and the sources is available. Therefore, many researchers have focused solely on the identification of $\mathbf{A}$ as in [96], [97], [98]. In this chapter, an extension of FastICA algorithm, which exploits the disparity in the kurtoses of the underlying super-Gaussian sources to estimate the mixing matrix is proposed. This algorithm is coined as UBSS FastICA. Furthermore, it is demonstrated how effective such source extraction approach

can be.

Due to the empirical formulation of UBSS FastICA, it will be examined why such filtering approach would lead to estimation of the non-sparse source. To this end, the objective of the second part of this chapter is to provide an extensive set of simulations to demonstrate why this filtering approach fully solve this particular underdetermined blind identification. The shape parameter α of the generalised Gaussian distribution (GGD) is employed as a measure of sparseness and Gaussianity. This shape parameter α was also utilised to illustrate the convergence of this filtering approach and the sub-Gaussian effect of the filter on the mixtures. Moreover, the cases where the noise source is modelled as either sub-Gaussian or Gaussian are also considered as an extension of the first part of this chapter. Simulation studies show that this filtering approach is robust and performs well in this particular TMJ UBSS application.

Hence, the objectives of this study are to 1) address the 2×3 underdetermined blind source separation in the context of TMJ signals 2) demonstrate why linear filtering does not alter the structure of $\mathbf{A}$ prior to ICA, 3) investigate the effect of considering different degrees of Gaussianity of the non-sparse source on identification of $\mathbf{A}$, 4) demonstrate the relationship between sparsity and the degree of Gaussianity of the sources, and 5) illustrate how the pre-filtering approach leads to the identification of the column of $\mathbf{A}$ corresponding to the non-sparse source, which thereby results in full identification of the underdetermined mixing matrix.

Likewise, the organisation of this chapter is as follows; section 5.2 defines and provides discussions regarding the following concepts exploited in this study: sparsity, sparseness, and super-Gaussianity. Then, the assumptions made in this work are listed, and their explanations follow suit. Similarly,

section 5.4 demonstrates why linear filtering is possible prior to ICA. Section 5.5 outlines the proposed UBSS FastICA approach. This is followed by some simulation studies in section 5.6. The first part of this section assesses the performance of UBSS FastICA algorithm against two other algorithms, namely the k-means clustering algorithm [8, 72] and the algorithm of Li *et al.* [42] for the identification of the mixing matrix. As for the source reconstruction stage, the ℓ_1-norm minimisation algorithm [42] is compared with the proposed approach. The second part of the simulation section comprises of a set of extensive simulations to demonstrate the robustness of UBSS FastICA, subject to different conditions. Subsequently, discussions regarding the simulation results are provided in section 5.7 and this chapter is concluded in section 5.8.

5.2 Sparsity, sparseness and super-Gaussianity

Sparsity (or disjointness) in this work refers to the situation where a relatively small number of source signals are active over any particular time interval. For the case of a single active source, sparsity can be mathematically described as

$$\{s_i(t);\ i = 1, ..., n\}$$

$$\text{where } \forall\ t\ \exists\ k \in 1, ..., n \text{ where } |s_k(t)| >> |s_j(t)| \qquad (5.2.1)$$

$$\text{and for } j \neq k\ s_j(t) \approx 0$$

where $s_k(t)$ is a given source signal and $s_j(t)$ is another source signal.

On the other hand, the degree of sparseness of a source signal depends on the number of occurrences of its samples being zero or approximately zero. In [62], the authors refer to a sparse signal if the magnitude of most of its samples is zero or near zero, with only a few sample entries taking significant values. It is noteworthy to highlight that this statement strongly correlates

with the nature of an impulsive signal as described in [47]. This is because an impulsive signal consists of only a few high peaks of short duration. Cichocki classified a signal as impulsive if $0 < \alpha < 1$ (p. 245 [47]), where α is the shape parameter of the generalised Gaussian distribution (GGD). On the other hand, He *et al.* categorised a signal to be sparse if its corresponding shape parameter α is less than two [99]. For the purpose of clarity, GGD is re-defined as follows [92, 93]:

$$p(s_i(t), \sigma, \alpha) = \frac{\alpha\sqrt{\beta}}{2\sigma\Gamma(1/\alpha)} e^{-|\sqrt{\beta}s_i(t)/\sigma|^\alpha} \tag{5.2.2}$$

where $s_i(t)$ denotes the i-th source, $\sigma > 0$ is the scale parameter, $\Gamma(.)$ is known as the gamma function, and $\alpha > 0$ is the shape parameter. As for $\sqrt{\beta} = \sqrt{\frac{\Gamma(3/\alpha)}{\Gamma(1/\alpha)}}$, it is merely a scaling factor which causes $Var(s_i(t)) = \sigma^2$ where $Var(.)$ denotes variance.

Recall that a signal is said to be Gaussian distributed if $\alpha = 2$. However when the shape parameter α of a signal is less than two, it has a super-Gaussian or leptokurtic distribution, while with $\alpha > 2$, it has a sub-Gaussian/playtykurtic distribution [7]. Hence, *He et al.* suggested the equivalence between a super-Gaussian signal and a sparse signal. However, the strong similarity between the impulsiveness definition and that of sparseness given by Pearlmutter [62] suggests that the latter was more restrictive in his definition of sparseness, i.e. $0 < \alpha < 1$. In effect, for the sparsity condition (5.2.1) to be fulfilled, the sparseness definition of Pearlmutter seems more appropriate. This is because as α tends to zero, the probability of the signal to be of zero magnitude increases as was shown in [99]. Therefore, a signal is referred to as sparse if $0 < \alpha < 1$. Next, the assumptions made in this work are explained.

5.3 Assumptions made in the proposed UBSS FastICA approach

In this section, the hypotheses required for the underdetermined blind identification of **A** are listed below. Moreover, these assumptions are supported by sufficient explanations for the purpose of clarity.

A1) All source signals are statistically independent and super-Gaussian.

A2) The columns of the mixing matrix **A** are pairwise linearly independent.

A3) Considering the three sources as $s_1(t)$, $s_2(t)$, and $s_3(t)$,

$kurt(s_1(t))$, $kurt(s_2(t)) \gg kurt(s_3(t))$ where $kurt(.)$ denotes kurtosis.

A4) The two highly super-Gaussian source signals $s_1(t)$ and $s_2(t)$ are sparse.

The statistical independence assumption made by **A1** is simply the cornerstone of ICA utilised in this work. Based on pp. 306-313 in [100], this section explains and justifies assumptions **A1** and **A2**. The non-Gaussianity assumption in **A1** and **A2**, along with the implicit assumption that the number of sources is known, guarantees the uniqueness of the model $\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t)$ [96]. By the term 'uniqueness', it is meant that $\mathbf{x}(t)$ does not have two non-equivalent representations [100]. Two representations, i.e. $\mathbf{x}(t)=\mathbf{A}\mathbf{s}(t)=\mathbf{B}\mathbf{q}(t)$ are referred to as non-equivalent if every column of **A** is not proportional to any column **B** and *vice-versa*. Further to this, **lemma 10.2.4** of [100] is given as follows:

Consider $\mathbf{x}(t)$ to be a two-dimensional random vector, $\mathbf{x}(t)=[x_1(t)\ x_2(t)]$ with two representations:

$$\begin{aligned} x_1(t) &= a_{11}s_1(t) + ... + a_{1j}s_n(t) \\ x_2(t) &= a_{21}s_1(t) + ... + a_{2j}s_n(t) \end{aligned} \tag{5.3.1}$$

$$\begin{aligned} x_1(t) &= b_{11}q_1(t) + ... + b_{1j}q_k(t) \\ x_2(t) &= b_{21}q_1(t) + ... + b_{2j}q_k(t) \end{aligned} \tag{5.3.2}$$

where $s_1(t)$,...,$s_n(t)$ and $q_1(t)$,...,$q_k(t)$ are sets of independent random variables (r.vs.). If the jth column of $\mathbf{A}$ is not proportional to any other of its ith column ($j \neq i$) or to any column of $\mathbf{B}$, then $s_j(t)$ is Gaussian distributed. Its proof can be found in p. 309 of [100].

Likewise, **theorem 10.3.5** in [100] states that:
If $s_1(t)$,...,$s_n(t)$ are non-Gaussian, then $\mathbf{x}(t)$ has a unique structure with respect to the given number of variables $\mathbf{s}(t)$, i.e. if $\mathbf{x}(t)$=$\mathbf{Bq}(t)$, where the order/size of $\mathbf{B}$ is the same as $\mathbf{A}$, then $\mathbf{A}$ and $\mathbf{B}$ are equivalent. As for the **proof**, assume there are two nonequivalent representations with the same number of variables, $\mathbf{x}(t)$=$\mathbf{As}(t)$=$\mathbf{Bq}(t)$. Then by **lemma 10.2.4**, some of the variables are Gaussian, which contradicts the non-Gaussian assumption.

Moreover, **theorem 10.3.8** [100] states that, provided the variables $\mathbf{s}(t)$ are non-Gaussian and the columns of $\mathbf{A}$ are linearly independent, then the model $\mathbf{x}(t)$=$\mathbf{As}(t)$ is unique for the specified number of variables. **Proof**: By **theorem 10.3.5**, when $\mathbf{s}(t)$ are non-Gaussian, it is deduced from the Eq. $\mathbf{x}(t)$=$\mathbf{As}(t)$=$\mathbf{Bq}(t)$ that $\mathbf{A}$ and $\mathbf{B}$ are equivalent. In the sequel, $\mathbf{x}(t)$=$\mathbf{As}(t)$ and $\mathbf{x}(t)$=$\mathbf{Bq}(t)$ can be two representations of $\mathbf{x}(t)$. Due to the linear independence of the columns of $\mathbf{A}$, $(\mathbf{A}^T\mathbf{A})$ is non-singular and therefore invertible. Hence, $\mathbf{s}(t) = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{x}(t)$ and $\mathbf{q}(t) = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{x}(t)$ are equal, which concludes the proof.

Theorems 10.3.5 and **10.3.8** which imply the non-Gaussianity assumption **A1** together with the implicit knowledge of the number of sources,

and the linear independence of the columns of **A** guarantees the uniqueness of the model $\mathbf{x}(t)$=$\mathbf{A}\mathbf{s}(t)$. However, it should be stressed that the model $\mathbf{x}(t)$=$\mathbf{A}\mathbf{s}(t)$ is still unique if there is at most one Gaussian source. This statement follows from the corollary of **theorem 10.3.6** [100] and from **theorem 5** of [101].

As FastICA maximises the non-Gaussianity of the estimated ICs, it is found that FastICA focuses on the highly non-Gaussian ICs, and not the IC whose distribution is closer to Gaussian, whenever assumption **A3** is satisfied. This solves partially the identification of the mixing matrix. Assumption **A4** ensures that the filtering preprocessing step of UBSS FastICA assists in the identification of the weakly non-Gaussian source, thereby completing the full identification of **A**. Additionally, it is noted that **A3** implicitly implies **A4** which will be explained as follows. The kurtosis $kurt(s_i(t))$ which is a measure of the 'peakedness' of the probability distribution of $s_i(t)$ can be defined as [102]:

$$kurt(s_i(t)) = \frac{E\{s_i^4(t)\}}{E\{s_i^2(t)\}^2} - 3 \tag{5.3.3}$$

where E{.} stands for the expected value. If $kurt(s_i(t)) = 0$, then $s_i(t)$ is referred to as Gaussian, while $kurt(s_i(t)) > 0$ implies that $s_i(t)$ has a super-Gaussian distribution. Otherwise, $s_i(t)$ is known as sub-Gaussian. As the rth moment of a signal with GGD can be expressed as follows [92]:

$$E\{|s_i(t)|^r\} = \frac{\Gamma(\frac{r+1}{\alpha})}{\Gamma(\frac{1}{\alpha})}\beta^{-r/2}\sigma \tag{5.3.4}$$

Then, Eq. (5.3.3) can be formulated in terms of α using (5.3.4) as:

$$kurt(s_i(t)) = \frac{\Gamma(\frac{5}{\alpha})\Gamma(\frac{1}{\alpha})}{\Gamma^2(\frac{3}{\alpha})} - 3 \tag{5.3.5}$$

Thus, $kurt(s_i(t))$ can be plotted as a function of α as shown in Fig. 5.1. Notice that the kurtosis has a large value when $0 < \alpha < 1$. This in turn

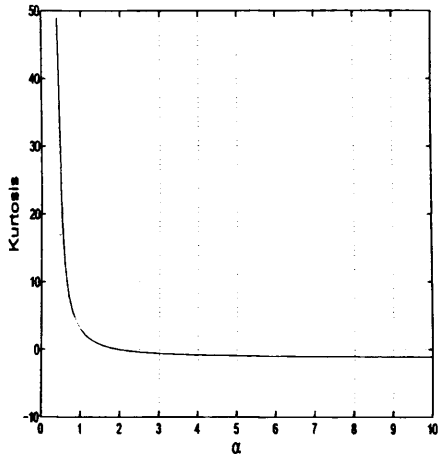

Figure 5.1. Kurtosis as a function of α. Note that when $\alpha > 2$, kurtosis < 0 for the sub-Gaussian case and the rate of change of kurtosis is much lower for the sub-Gaussian case than for of the super-Gaussian case ($\alpha < 2$).

implies the highly super-Gaussianity of $s_i(t)$ for $0 < \alpha < 1$. Recall that a source signal is referred to as sparse if $0 < \alpha < 1$. Thus, as **A3** states that there are two highly super-Gaussian sources, it implies their sparseness (**A4**) as well. Hence, it is stressed that the equivalence between **A3** and **A4** is not straightforward without the formulation of equation (5.3.5) plotted in Fig. 5.1. As mentioned earlier, UBSS FastICA consists of a preprocessing filtering step, prior to ICA. In the following section, it will be shown why ICA model is still valid, even after filtering.

5.4 Why is moving average filtering possible prior to ICA

This section is based on pp. 264-265 in [7] to demonstrate that the ICA model in terms of the structure of $\mathbf{A}$ still stands after applying moving average filtering elementwise on the mixtures $\mathbf{x}(t)$. This is clear from the following equation:

$$\mathbf{X}' = \mathbf{XF} = \mathbf{ASF} = \mathbf{AS}' \tag{5.4.1}$$

where $\mathbf{X}' = [\mathbf{x}'(1) \cdots \mathbf{x}'(t) \cdots \mathbf{x}'(T)]$, $\mathbf{S}' = [\mathbf{s}'(1) \cdots \mathbf{s}'(t) \cdots \mathbf{s}'(T)]$ and $\mathbf{F}$, which conveys the filtering operation matrixwise, will have the following

form if the filter length M is 3 [7]:

$$\mathbf{F} = \frac{1}{M} \begin{pmatrix} & & & & \cdot & & & & & \\ & & & & \cdot & & & & & \\ & & & & \cdot & & & & & \\ \ldots & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & \ldots \\ \ldots & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & \ldots \\ \ldots & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & \ldots \\ \ldots & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & \ldots \\ & & & & \cdot & & & & & \\ & & & & \cdot & & & & & \\ & & & & \cdot & & & & & \end{pmatrix} \tag{5.4.2}$$

Since the independent sources $s_i(t)$ are filtered elementwise by $\mathbf{F}$, the resulting $s_i(t)'$ are not linear mixtures of $s_i(t)$. Thus, Eq. (5.4.1) demonstrates that filtering does not have any effect on $\mathbf{A}$, while the elementwise nature of Eq. (5.4.2) shows that the filtered sources $s_i(t)'$ still enjoy statistical independence. In the light of the above background, the UBSS FastICA can now be outlined as follows.

5.5 Development of UBSS FastICA algorithm

1. Firstly, the two columns of the mixing matrix $\mathbf{A}$ corresponding to the two highly super-Gaussian sources are estimated by employing FastICA [7]. This is achieved by maximising the negentropy (Neg(.)) of the linear combination of the pair of mixtures:

$$\arg \sup_{w_1, w_2} \Big(Neg(w_1 x_1(t) + w_2 x_2(t)) \Big) \tag{5.5.1}$$

where $\mathbf{w} = [w_1, w_2]$, which is one of the rows of the separating matrix $\mathbf{W}$. The second row of $\mathbf{W}$ is estimated by (5.5.1), but followed by a deflationary orthogonalisation. The key observation is that in UBSS

FastICA will focus on the high kurtosis sources (assumption A3). The two columns of $\mathbf{A}$ are estimated by inverting $\mathbf{W}$. The two TMJ sources can be estimated by $\mathbf{y}(t) = \mathbf{W}\mathbf{x}(t)$. Prior to using FastICA, preprocessing the mixtures with *median* filtering with a small window of 3 samples to yield $\mathbf{f}(t)=[\mathrm{f}_1(t)\ \mathrm{f}_2(t)]^T$ assists in attenuating the effects of the outliers to which FastICA is sensitive.

2. Next, the task is to estimate the third column corresponding to the non-sparse source. Temporally mean filter $\mathrm{f}_i(t)$ with a window size greater than twice the duration of the maximum period P_{max} during which the two highly super-Gaussian sources are active. Fig. 5.3 shows these active periods. Although the window size is not known a priori, a large window can be used (assumption A4). The mean filtering mitigates the two highly super-Gaussian source signals in both mixtures. Mean filtering of the signal $\mathrm{f}_i(t)$ to yield $\mathrm{g}_i(t)$ is achieved as follows:

$$\mathrm{g}_i(t) = \frac{1}{\mathrm{M}} \sum_{k=t-\mathrm{M}+1}^{t} \mathrm{f}_i(k) \quad i = 1, 2 \tag{5.5.2}$$

M denotes the window size. Consider the periods when the two highly super-Gaussian source signals are active. Note the predominance of these two source signals in the mixture signals. Similarly, the converse is true when these highly super-Gaussian sources are not active. Hence, the averaging operation widens the active periods of the two highly super-Gaussian sources (in the mixture signals), while suppressing their amplitudes. Note the *absence* of high amplitudes (with respect to the TMJ sources) in the noise signal, resulting in a lower effect of the mean filtering on the noise signal. Now, the noise source predominates in the mixture signals.

3. Apply FastICA to estimate the third column of **A** pertaining to the weakly super-Gaussian source signal. However, FastICA may fail to estimate the column of **A** corresponding to the weakly super-Gaussian source because of its equivalence to projection pursuit (PP) [7]. Thus, two ICs are estimated, instead of one. The column corresponding to the IC with the minimum kurtosis is selected on the basis of A3 to form the last column of the estimated **A**, i.e. $\hat{\mathbf{A}}$.

4. Apply the ℓ_1-norm algorithm [73] to extract the source signals by minimising the following cost function:

$$min \sum_{i=1}^{n} |s_i(t)| \text{ s.t. } \hat{\mathbf{A}}\mathbf{s}(\mathrm{t}) = \mathbf{x}(\mathrm{t}) \tag{5.5.3}$$

5. Alternatively, FastICA can be utilised to perform source extraction. Begin by *eliminating* the mean-filtering effect (5.5.4) from the IC with the minimum kurtosis (from step 3) to obtain the estimate of the least super-Gaussian source. However, there is still a significant contribution of the most super-Gaussian source as seen in the first noise estimate in Fig. 5.6.

$$\mathrm{f}_i(t) = \mathrm{Mg}_i(t) - \sum_{k=t-\mathrm{M}+1}^{t-1} \mathrm{f}_i(k) \quad i = 1, 2 \tag{5.5.4}$$

Note that (5.5.4) does the opposite of (5.5.2), with the same notations for the purpose of clarity. To strengthen the presence of the source with the maximum kurtosis as the independent component, any two distinct scaled version of the source with the maximum kurtosis can be added. Applying FastICA to the two new mixtures improves the estimate of the weakly super-Gaussian source. See the first two plots of Fig. 5.6. This step is optional if one desires to fully solve the UBSS. However, the prime objective of this study is to extract the TMJ sources. The

highly non-Gaussian sources are estimated from step 1.

For the purpose of clarity, the overall UBSS FastICA is summarised in the following flowchart in Fig. 5.2.

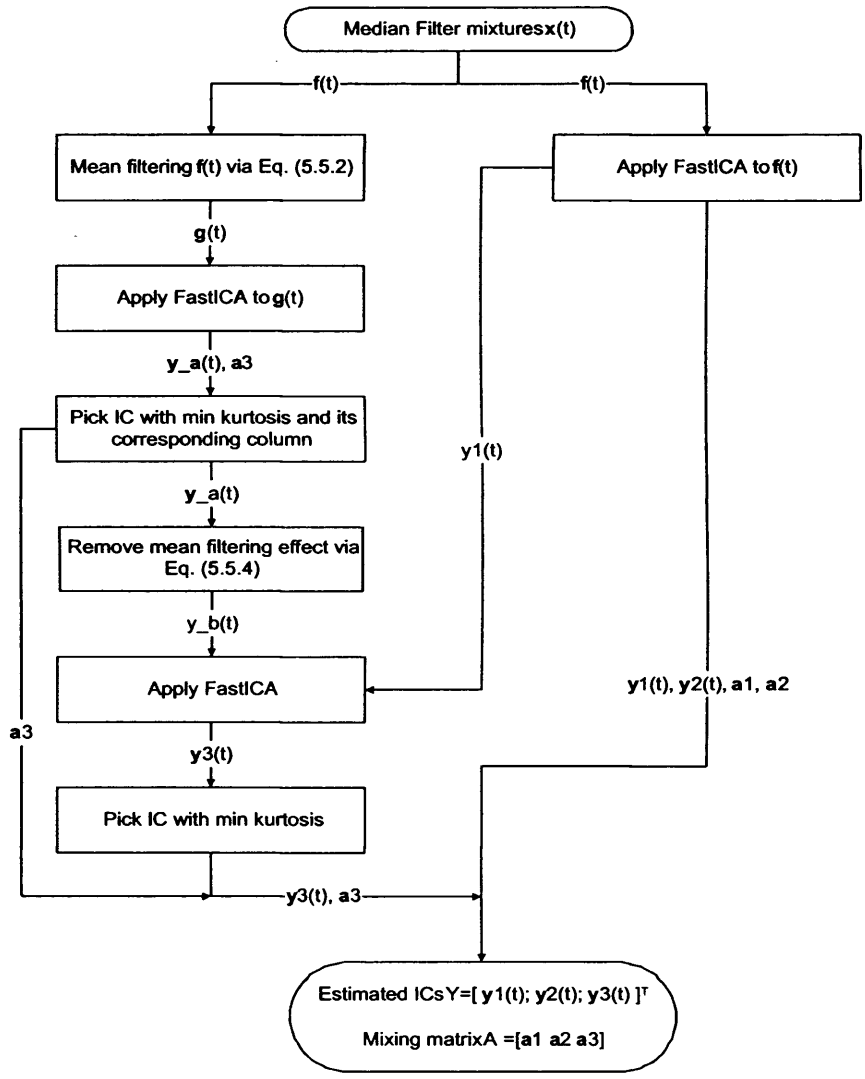

Figure 5.2. Flowchart describing UBSS FastICA procedure.

5.6 Simulations

The performance measure (PM) for the underdetermined BSS as described in Chapter two, is employed here. The scenario where TMJ sounds are the mixtures of click, crepitus, and a non-sparse source modelled as noise

is considered. Fig. 5.3 shows the sources: the weakly super-Gaussian noise, click and crepitus from top to bottom. The kurtoses of click, and crepitus (recorded from the ipsi side of the TMD infected joint) are respectively 23.7 and 14.4.

5.6.1 Underdetermined blind source separation

In this section, the non-sparse source is modelled with Laplacian distribution

$$P(s) = \frac{1}{2\lambda} exp\{-\frac{|s-\theta|}{\lambda}\} \tag{5.6.1}$$

where the variance $\sigma^2 = 2\lambda^2$ and the mean $\mu = \theta$. The $\mathbf{A}$ matrix (generated randomly from a standardised normal distribution) used in this study is given below:

$$\begin{pmatrix} 0.8999 & -0.9158 & 0.5984 \\ -0.4360 & 0.4017 & -0.8012 \end{pmatrix}$$

The performance of the extended FastICA algorithm was measured against

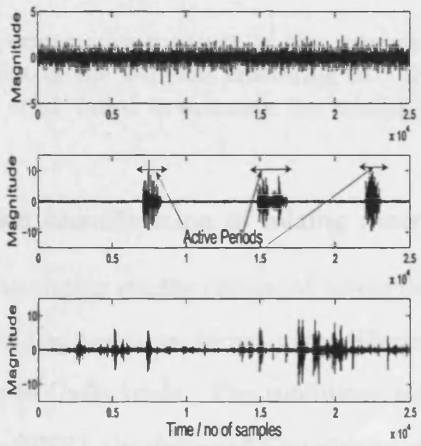

Figure 5.3. The three source signals namely: the super-Gaussian noise, the click and the crepitus. The SNR ratio is 6 dB. Note the sparsity of the click (middle plot) where the click occurs in the three excitation regions. The same observation can be made for the crepitus signal (last plot).

the K-means clustering [8, 72](which relies on minimisation of the distance

between the data points and the assigned centres of the clusters) and the algorithm of Li *et al.* [42] (that scans for the sparse regions of the signals and clusters these regions to estimate the mixing matrix) in terms of SNR. This is shown in Fig. 5.4. Fig. 5.5 illustrates the source estimates when the ℓ_1-norm minimisation algorithm in Eq. (5.5.3) was employed for source reconstruction. Also, Fig. 5.6 visually demonstrates the potential of the extended FastICA algorithm to extract the sources.

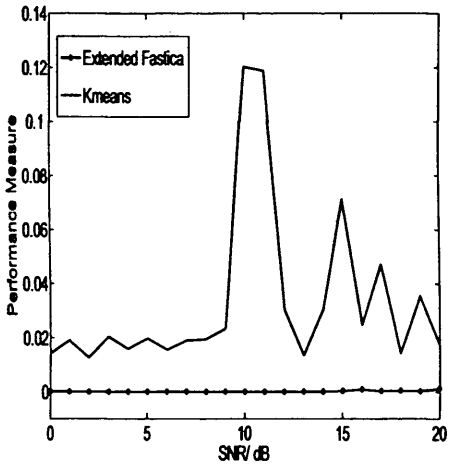

Figure 5.4. The performance measure of the extended FastICA scheme plotted in '.-' and that of the k-means clustering in '-' against SNR in dB. The algorithm of Li *et al.* failed to estimate the mixing matrix A.

5.6.2 Robust blind identification of mixing matrix

In this section, the simulation studies presented herein demonstrate how the filtering UBSS FastICA approach performs under different conditions via an extensive set of Monte Carlo trials. The conditions considered were: the signal to noise Ratio (SNR), the degree of Gaussianity of the noise, and the filter length. Likewise, the three distributions utilised in this study were Laplace, normal, and uniform to convey respectively the super-Gaussianity, Gaussianity and sub-Gaussianity nature of the non-sparse source. Furthermore, the evolution of α was monitored to assess the convergence of UBSS

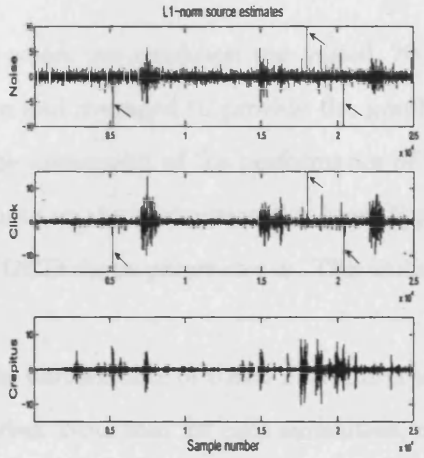

Figure 5.5. ℓ_1-norm estimates of the three source signals. From top to bottom: Estimate of noise, click, and crepitus. The SNR ratio is 6 dB. Note the prominent artifacts pointed by the arrows in both click and noise estimates.

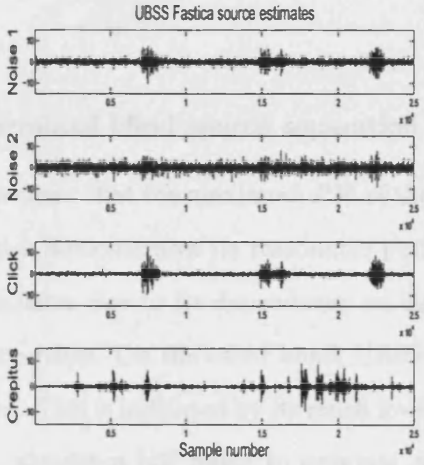

Figure 5.6. UBSS FastICA estimates of the three source signals. From top to bottom: Estimate of noise of step 4, final estimate of noise, click, and crepitus. The SNR ratio is 6 dB. Note the significant presence of the click (most super-Gaussian source) in the first estimate of the noise.

FastICA.

For each simulation where one condition was varied, 20 independent Monte Carlo trials were run and averaged to provide the graphs in Figs. 5.8-5.12. However, prior to the assessment of the performance of UBSS FastICA, the effect of the filter length on the mixtures $\mathbf{x}(t)$ is investigated in terms of their Gaussianity via the GGD shape parameter α. This is illustrated in Fig. 5.7.

In Figs. 5.8 & 5.9, the performance of UBSS FastICA is assessed, as SNR and filter length were varied. Note that for each simulation, noise was considered as the non-sparse source whose distribution ranged from super-Gaussian to sub-Gaussian.

Moreover Figs. 5.10 - 5.12 illustrate the convergence of UBSS FastICA via the evolution of α_i of the estimated ICs at SNR= 0dB. The true values α_i of the ICs are also included for comparison purposes.

5.7 Discussion

5.7.1 Underdetermined blind source separation

From Fig. 5.4, it is evident that the maximum *PM* of the k-means algorithm is less than 0.14 which demonstrates its reasonable performance. However, its performance fluctuates due to its dependency on its initialisation upon which its convergence relies. On the other hand, UBSS FastICA algorithm outperforms k-means. This is indicated by its much lower performance measure. The Li *et al.* algorithm [42] failed to estimate $\mathbf{A}$, probably because the super-Gaussian noise source is always active. One of the assumptions of Li *et al.* is sparsity with respect to *all* the three signals. As for the source reconstruction stage, ℓ_1-norm minimisation algorithm does yield reasonable source estimates, especially the problem at hand is an *underdetermined* one.

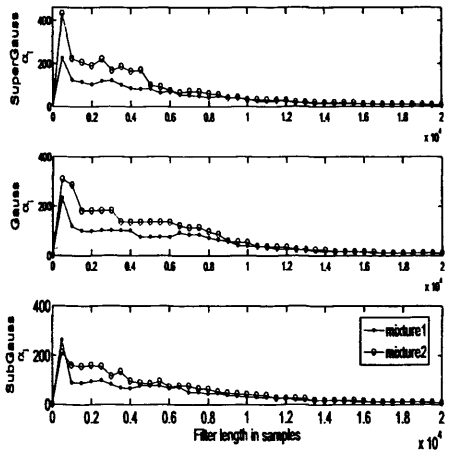

Figure 5.7. Effect of filter length on the degree of Gaussianity of the mixtures for super-Gaussian, Gaussian, and sub-Gaussian noise (from top to bottom) at 0 dB. It is noteworthy to say that at the maximum filter length of M = 20,000 samples, $\alpha > 5$. In other words, the mixtures are still sub-Gaussian. However, prior to pre-filtering of the mixtures, $\alpha < 1$. This explains why without filtering, FastICA focuses on the TMJ sources, while pre-filtering leads to the non-sparseness of the mixtures and consequently estimate the non-sparse noise instead of the sparse TMJ sources.

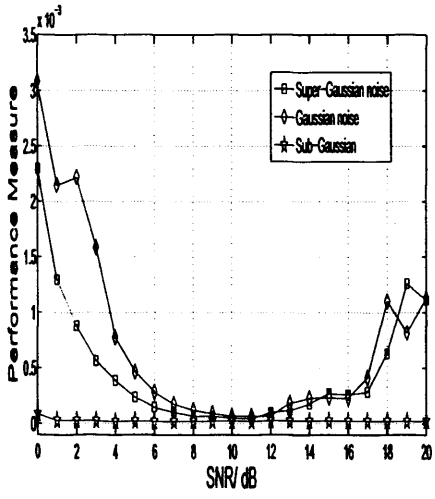

Figure 5.8. Performance measure versus signal to noise ratio (SNR) in dB when super-Gaussian, Gaussian and sub-Gaussian noises were considered. Note the much better performance measure of the sub-Gaussian noise case. This is because pre-filtering leads to the non-sparseness/sub-Gaussianity of the mixtures as seen in Fig. 5.7. Therefore their distributions are much closer to that of the sub-Gaussian noise.

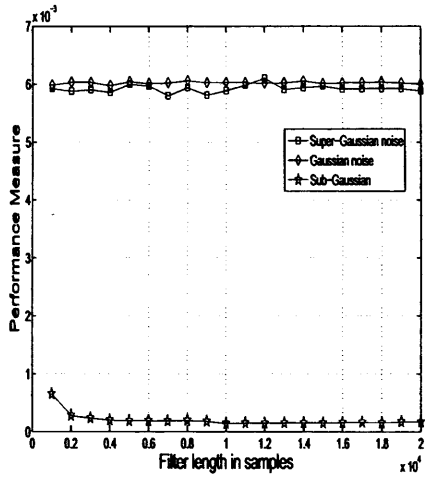

Figure 5.9. Performance measure as a function of filter length at 0 dB when super-Gaussian, Gaussian and sub-Gaussian noises were considered. Note the much better performance measure of the sub-Gaussian noise case. In fact, the moving average pre-filtering leads to the non-sparseness/sub-Gaussianity of the mixtures as seen in Fig. 5.7 . Therefore the distributions of the mixtures are much closer to the sub-Gaussian noise.

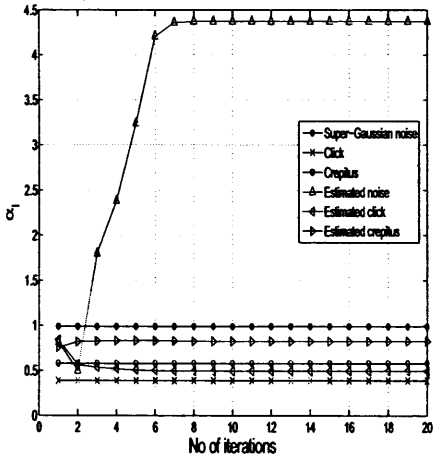

Figure 5.10. Convergence graph: Evolution of α at 0 dB when super-Gaussian noise was considered.

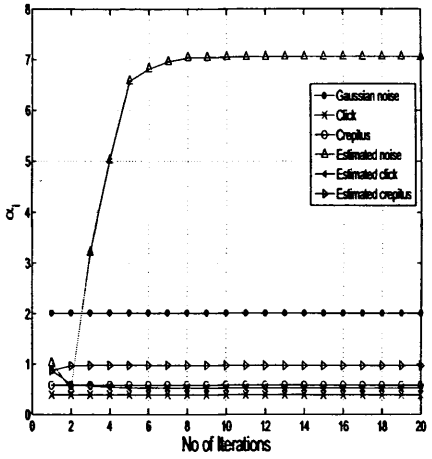

Figure 5.11. Convergence graph: Evolution of α at 0 dB when Gaussian noise was considered.

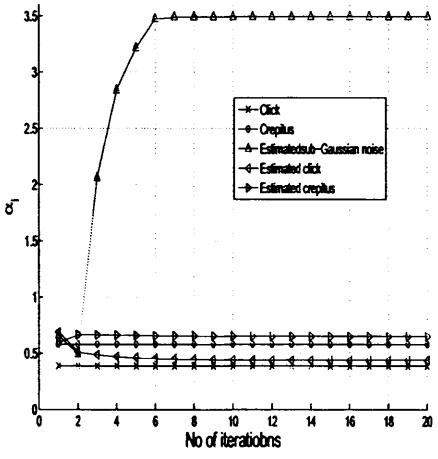

Figure 5.12. Convergence graph: Evolution of α at 0 dB when sub-Gaussian noise was considered. Note the closeness of the α_i of the estimated TMJ sources with those of the original TMJ, compared to the super-Gaussian and Gaussian noise cases in Fig. 5.10 & 5.11 respectively.

Nevertheless, there remain some significant artifacts (as shown in Fig. 5.5). On the other hand, Fig. 5.6 shows that UBSS FastICA estimates do not contain such prominent artifacts. It is noteworthy to say that both TMJ estimates of ℓ_1-norm minimisation algorithm and UBSS FastICA contain artifacts of the noise source. However, it is possible to annihilate significantly the noise component, by employing a Wiener filter (which requires a priori knowledge of SNR) as proposed in [103]. However, noise cancellation using this signal enhancement approach is not the subject of this thesis and therefore will not be discussed herein.

5.7.2 Robust blind identification of mixing matrix

In the first place, note the sub-Gaussianity effect on the mixtures $\mathbf{x}(t)$ by moving average filtering in Fig. 5.7. However prior to filtering, α of the mixtures was less than unity and therefore the signals were highly super-Gaussian and sparse. It was intuitively stated that the filtering process suppressed the two highly super-Gaussian sources in the mixtures and therefore the *non-sparse* noise was more prominent in $\mathbf{x}(t)$. In section 5.2, it was deduced that when $\alpha \geq 1$, the signal is non-sparse. On this basis, it is reaffirmed that the filtering suppresses the two highly super-Gaussian *sparse* sources. This is because the moving average has altered the nature of the mixtures from highly super-Gaussian to sub-Gaussian or equivalently from sparse to non-sparse. On the other hand, it is clear from Fig. 5.7 that as the filter length M increases, both mixtures tend to have Gaussian distribution due to the central limit theorem. Nonetheless, it is worth noting that at the maximum filter length of M = 20,000 samples, $\alpha > 5$ indicates that the mixtures are still sub-Gaussian.

With regard to the *PM* against SNR shown in Fig. 5.8, notice the maximum value of PM is of order 10^{-3}. This demonstrates the good performance of UBSS FastICA in all the three scenarios. Observe the much better perfor-

mance of the sub-Gaussian case. Further to the discussion in the previous paragraph, this was expected, due to the sub-Gaussian/ non-sparse effect of the filter on the mixtures. Additionally, all three curves have a minimum at SNR=10 dB. However, this is not obvious for the sub-Gaussian case from Fig. 5.8 due to its much lower performance measure (of magnitude of order 10^{-5}). The minimum at 10 dB arises due to the following: the higher the SNR, the better the estimation of the two columns of $\mathbf{A}$ pertaining to the two TMJ sources and *vice-versa* for the noise. These explain that the best performance is mid-way between 0 dB and 20 dB.

Next, the performance of UBSS FastICA in terms of the filter length in Fig. 5.9 is discussed. However, prior to this discussion, examine Fig. 5.7. This figure might be misleading due to the *high* magnitude of α in the sense of the degree of sub-Gaussianity, which can be explained as follows. Fig. 5.1 demonstrates that for the sub-Gaussian case (i.e. $\alpha > 2$) the kurtosis does not change exponentially as it does for $\alpha < 2$. For example, the difference between the kurtoses corresponding to $\alpha = 3$ and $\alpha = 400$ is not significant (i.e. difference ≈ 0.6). Hence, the degree of sub-Gaussianity from $\alpha > 2$ to $\alpha = 400$ does not change significantly. In other words, the filter length does not alter the degree of sub-Gaussianity of the mixtures significantly. From Fig. 5.9, note that the filter length does not have any significant effect on the performance measure (all three curves are approximately unvaried). Based on the previous discussion, such trend was expected. Moreover, the superior performance of the sub-Gaussian case is again highlighted.

Last but not least, the convergence of UBSS FastICA in terms of α is examined, when super-Gaussian, Gaussian and sub-Gaussian noises are considered in Figs. 5.10, 5.11, & 5.12 respectively. From these plots, the most striking curve (topmost) belongs to the estimated noise. From Figs. 5.10 & 5.11, the estimated noise converges to sub-Gaussianity as $\alpha > 2$. This

contradicts the super-Gaussianity and Gaussianity nature of the noise considered. Nevertheless from Eq. (5.4.1), it can be deduced that the filtered ICs $\mathbf{S}'$ are estimated, instead of the original $\mathbf{S}$. As discussed previously, the moving average filtering process in this particular TMJ UBSS accounts for this sub-Gaussian/ non-sparse effect. On the other hand, if the convergence of the estimated TMJ ICs (i.e. click and crepitus) is inspected, it is clear that in all cases, they converge to $\alpha \leq 1$ (i.e. highly super-Gaussianity). The reason why they do not converge to the true values is probably due to the low SNR of 0 dB. Again, it is emphasized the much better performance of sub-Gaussian case in Fig. 5.12. This is illustrated by the closeness of the α_i of the estimated TMJ ICs to those of the true ones in Fig. 5.12, compared to those in Figs. 5.10 & 5.11.

5.8 Conclusions

This study has shown how filtering can assist in solving the underdetermined blind source separation in the context of TMJ sounds. The common approach is to exploit the structure of the mixtures as a result of sparsity of the sources, such as the k-means and the algorithm of Li [42]. This observation was also noted in the literature survey on sparse and non-sparse BSS [11]. On the other hand, UBSS FastICA algorithm takes advantage of both the sparsity and the statistical properties of the source signals. In that respect, UBSS FastICA algorithm is more efficient in its solutions than the k-means and Li's algorithms. The main idea of this study is that provided a linear transform is applied to the mixture signals to suppress certain source signals within the mixtures, the prevalence of others can be accentuated. However, the challenge remains in finding more optimum linear transforms that perform source signals attenuation within the mixture signals.

Furthermore, the role of sparsity in this empirical study was investigated

via the shape parameter α of the generalised Gaussian distribution. A close relationship between this parameter, the sparseness of a signal, and consequently the sparsity situation in this particular TMJ UBSS was noted. It was implicitly explained the subtle difference between sparsity and sparseness. It was deduced that a highly super-Gaussian signal, i.e. $\alpha < 1$ is likely to be sparse in section 5.2. Furthermore, it was demonstrated why the ICA model still stands after pre-filtering in equation (5.4.1) in terms of the statistical independence of the new IC $\mathbf{S}'$ and the unaltered structure of $\mathbf{A}$.

In the simulation studies, it was deduced that the moving average filter has a sub-Gaussian/non-sparse effect on these particular TMJ mixtures $\mathbf{x}(t)$. Consequently, the sub-Gaussian noise TMJ UBSS outperformed the other scenarios, i.e. when super-Gaussian and Gaussian noises were considered. However, the performances in all three scenarios were good due to their performance measure being of order 10^{-3} or less. Eq. (5.4.1) demonstrates why the estimated noise does not converge to the original ones in Figs. 5.10 & 5.11. It was intuitively stated that the filter suppressed the two *sparse* TMJ sources. In the sequel, the *non-sparse* noise pre-dominates in the filtered mixtures. This is evident from the *sub-Gaussian/non-sparse* effect of the filter in Fig. 5.7.

Chapter 6

DELAY ESTIMATION FOR SOURCE SEPARATION AND LOCALISATION OF TMJ SOUNDS

6.1 Introduction

In this chapter, the mixing system within the head is modelled as anechoic. The legitimacy of this model will be supported by a survey on the literature related to the acoustic properties of the head. The focus of the first part of this chapter rests upon simulated anechoic mixing of crepitus signals, while the second study investigates TMJ source separation on bilaterally recorded TMJ sounds, whereby the mixing process reflects that of the human head.

As mentioned above, the first part comprises of the synthetic anechoic mixing of the ipsi TMJ sounds to simulate the TMJ BSS scenario. It is shown how the inherent fractional delay between a pair of TMJ sound sources can be estimated by time *localised* sparse component analysis. Likewise, *short* active periods of only *one* source can be blindly tracked within the mixtures using mutual information (MI). In effect, these active periods will assist in estimating the delay by employing the modified versions of either the max-

imum likelihood (ML) delay estimator or the mixed modulated Lagrange explicit time delay estimator (MMLETDE). Thereafter, a strategy based on delaying the mixtures to solve the blind source separation of TMJ sources is considered. Simulation studies support the improvement of the proposed approach when applied to these artificially mixed TMJ sound signals. In particular, in terms of the orthogonally projected signal-to-interference ratio (SIR) defined by Vincent *et al.* [74], there is at least 10 dB improvement over the methods of Parra and Yilmaz & Rickard [4, 5].

The second part of this chapter in section 6.4 deals more specifically with the real TMJ recordings. In particular, patients with only one TMD affected joint generating 'clicks' are examined. The possibility that the TMJ recordings are in fact mixtures of the click source (generated from the TMD joint) and the TMJ sound source (produced by the other healthy/normal TMJ) is considered. The non-stationary nature of the TMJ signals is exploited by employing the DUET (*degenerate un-mixing estimation technique*) algorithm [5] as a time-frequency approach to separate the sources. As the DUET algorithm requires the sensors to be closely spaced, which is not satisfied by our recording setup, the delay between the recorded TMJ sounds has to be estimated to perform an alignment of the mixtures. Thus, the proposed extension of DUET enables for arbitrary separation of the sensors. It is also shown that DUET outperforms the convolutive Infomax algorithm in this particular TMJ source separation scenario. The spectra of both separated TMJ sources with the proposed method are comparable to those available in the existing literature. In addition, examination of both spectra suggests that the click source has a better audible prominence than the healthy TMJ source. Furthermore, the problem of source localisation is addressed. This can be achieved automatically by detecting the sign of the proposed mutual information estimator which exhibits a maximum at the delay between the two mixtures.

This chapter is concluded in section 6.5. As a result, the localised separated TMJ sources can be of great clinical value to dental specialists.

6.2 The anechoic model

Mathematically, the model of the observed sound measurements is represented as:

$$x_i(t) = \sum_{j=1}^{2} a_{ij} s_j(t - \delta_{ij}) + v_i(t) \tag{6.2.1}$$

where $x_i(t)$ is the ith TMJ mixture signal and $v_i(t)$ represents additive zero mean white Gaussian noise at discrete time t for $i = 1, 2$. The parameters a_{ij} are the attenuation coefficients and δ_{ij} are generally the non-integer time delays associated with the path from the jth source to the ith sensor (microphone). Likewise, the mixing matrix $\mathbf{A}$ can be defined as:

$$\mathbf{A} = \begin{pmatrix} a_{11}z^{-\delta_{11}} & a_{12}z^{-\delta_{12}} \\ a_{21}z^{-\delta_{21}} & a_{22}z^{-\delta_{22}} \end{pmatrix} \tag{6.2.2}$$

where z^{-1} denotes the unit delay. These delays are in terms of samples. For accuracy in modelling, δ_{ij} is considered to be fractional since the *exact* head size, and speed of sound in the tissue differ from person to person.

Modelling the acoustic properties of the human head remains an open problem. The geometrical structure of the skull, coupled with the fact that the human head comprises soft tissue, layered bone, and brain tissue, has made it impossible to date to achieve an analytical solution of the acoustic properties of the brain. Also, due to ethical reasons, direct measurements in a living human being is hardly possible [104]. Guo *et al.* made the assumption that the acoustic propagation model of the head was convolutive, in the context of TMJ source separation [1]. This *intuitive* assumption did not consider

any physiological aspects of the human head and no existing literature review supports this convolutive model. The existence of acoustic multipaths within the human head is plausible, yet the acoustic attenuation within the brain reported in the literature suggests that these multipaths from one side of the head to the opposite side are negligible. Hence, an anechoic model of the head is reasonable due to the following arguments:

- Widmalm *et al.* demonstrated in their study that a contra (opposite side of the TMD joint) TMJ sound was a delayed version of its ipsi (same side of the TMD joint) TMJ sound by only one lag [33]. This supports the fact that there is at most one 'effective' acoustic path across the skull.

- Furthermore, the higher frequencies (i.e. greater than 1200Hz) of the ipsi TMJ sound were found to be severely attenuated [27] when it propagated to the contra side, which corroborates with the findings in [104]. Moreover, the spectrum of TMJ sounds has a bandwidth of 20-3200 Hz [27], [40]. Hence, most of the energy content of the propagated TMJ sound is severely attenuated. This is clear by comparing the spectrum of the contra and the ipsi TMJ sounds in Fig. 6 in [27]. This remark, regarding the significant loss across the human head, is also in agreement with the study of O'brien *et al.*, which indicated an acoustic loss of approximately 33dB [105]. On this basis, the assumption of multipaths of the ipsi TMJ source to the contra side is questionable due to the significant energy loss of the ipsi TMJ source.

- In the illustrations of [1], there is no evidence of reverberations. In fact, Figs. 1, 6, & 7 of [1] suggest that there is only one lag. Furthermore, Guo *et al.* stated in section 1.3 of their paper that "When two channels show similar waveforms, with one lagging and attenuated to some degree, it can be concluded that the lagging signal is in fact the propagated version of the other signal". This supports the idea of Guo

et al. that the TMJ sound recorded from the contra side constitutes one lagged version of the ipsi TMJ source.

In the sequel, the reasonable assumption that there is ultimately one 'effective' acoustic path from one side of the head to the other side, considering the other possible paths to be negligible is made herein. Furthermore, simulation studies on *non*-synthetically mixed TMJ sounds verify the validity of the anechoic model in the second part of this chapter.

6.3 Delay estimation in time localised sparse component analysis of TMJ sounds

In the anechoic model (6.2.2), $\delta_{11} = \delta_{22} = d$ and $\delta_{21} = \delta_{12} = d + \mathrm{D}$, due to the symmetric geometry of the human head; d is the time for a TMJ source to reach to its ipsi (same side of the head) sensor, D is the *differential* time interval between the two contra (opposite sides of the head) sources. In this section, time *localised* sparse component analysis is employed to determine the *differential* delay D in the synthetic anechoic mixtures. Note that d is not required in estimation of the sources as will be shown in section 6.3.5 and therefore the estimation of d is not considered in this work. The main hypothesis made is that there is at least one time interval during which *only* one source is active.

The organisation of the study on synthetically mixed anechoic TMJ sounds is as follows; in the next section the MMLETDE and ML delay estimators are briefly reviewed and their modified versions are proposed. In addition, sparse component analysis is employed here not to solve the underdetermined BSS, but to assist in the estimation of the differential delay D. Section 6.3.3 illustrates how mutual information (MI) can assist in identifying blindly the active period of a source from the mixture signals. The scaling and the sign ambiguities of the sources are tackled in order to estimate the delay. Prior

to simulation studies, an outline of a procedure to reconstruct the sources is given. Section 6.3.6 compares the performance of the proposed algorithm against algorithms such as that of Parra [4], which exploits the nonstationarity of the sources (convolutive modelling), and DUET of Yilmaz and Rickard [5], which performs time-frequency masking (anechoic modelling). As Vincent *et al.* [74], the *orthogonally projected* SIR is employed as a measure of audio source separation performance. Finally, the conclusions are made in section 6.3.7.

6.3.1 Background on fractional delay estimators

For discrete time signals, whenever a signal $s(t)$ is delayed by a *non-integer* delay D, the computation of the subsample between $s(t\text{-}\lfloor \mathrm{D} \rfloor)$ and $s(t\text{-}\lfloor \mathrm{D} \rfloor\text{-}1)$, where $\lfloor . \rfloor$ denotes the floor operation, is required. To *approximate* this non-integer delay, $s(t\text{-D})$ can be fomulated as [3]:

$$s(t-\mathrm{D}) = \sum_{k=-\infty}^{\infty} \mathrm{sinc}(k\text{-D}) s(t-k) \tag{6.3.1}$$

Appendix C.1 demonstrates why $s(t)$ is convolved with a delayed sinc function to yield $s(t-\mathrm{D})$. However, Eq. (6.3.1) implies an infinite number of samples to achieve this non-integer delay D. The well-known Lagrange interpolation finite impulse response (FIR) filter, $h(k)$, which approximates this sinc function is defined as:

$$h(k) = \prod_{i=-\mathrm{M}_a}^{\mathrm{M}_b} \frac{\mathrm{D}-i}{k-i} \quad i \neq k \tag{6.3.2}$$

where either $\mathrm{M}_a = \mathrm{M}_b = \mathrm{L}/2$ when L is even, or $\mathrm{M}_a = (L-1)/2$, $\mathrm{M}_b = (L+1)/2$ when L is odd, and $\mathrm{L} = \mathrm{round}(2\mathrm{D})$ is the order of the filter, where $\mathrm{round}(x)$ denotes rounding x to the nearest integer [3]. Appendix C.1 also illustrates how Lagrange interpolation formula (6.3.2) can approximate the ideal transfer function (i.e. sinc function) of a delay system.

Based on the least mean square error principle, the MMLETDE can be summarised as follows [3]:

$$\text{error}(l) = s(t - \text{D}) - \sum_{k=-M_a}^{M_b} \left(\hat{h}(k) e^{j\omega\upsilon} \right) s(t - k) \tag{6.3.3}$$

where

$$\nu = k - \hat{D}(l)$$

and

$$\hat{D}(l+1) = \hat{D}(l) - 2\mu Re\left\{ \text{error}(l) \sum_{k=-M_a}^{k=M_b} g(\nu) s(t-k) \right\} \tag{6.3.4}$$

where

$$g(\nu) = e^{j\omega\nu} \left(\frac{\cos(\pi\nu) - \text{sinc}(\nu)}{\nu} - j\omega \text{sinc}(\nu) \right)$$

and

$$j \triangleq \sqrt{-1}$$

Note that $\hat{h}(k)$ is the estimate of (6.3.2) when D is substituted with $\hat{D}$ and l denotes the lth iteration of the algorithm. In practice, an initial value $\hat{D}(0)$ is chosen to be zero and, (6.3.3) & (6.3.4) are repeated until convergence. From (6.3.3), it can be seen that Lagrange interpolation FIR filter coefficients $\hat{\text{h}}(k)$ are modulated by $e^{j\omega\upsilon}$ where ω is an arbitrary angular frequency. It is reasonable to remove $e^{j\omega\upsilon}$ from (6.3.3) with regard to its verification in the appendix C.2 [3]. The performances of the original MMLETDE and its modified version in Fig. 6.1 are compared by using second order auto-regressive filtered white noises (to approximate the TMJ signals) with the delay D = 0.83. From this simulation (left plot of Fig. 6.1), it is clear that the original MMLETDE reaches its steady state at around 1000th iteration, while the modified one converges at about 1500th iteration. However, the original MMLETDE results in $\hat{D} = 0.8141$, whereas the modified one leads

to $\hat{D} = 0.83$. The right plot of Fig. 6.1 shows similar convergence of both algorithms when the step-size of the modified MMLETDE is $\mu = 0.1$, while for the original MMLETDE it is $\mu = 0.05$. It should be noted that both algorithms perform similarly for other values of D.

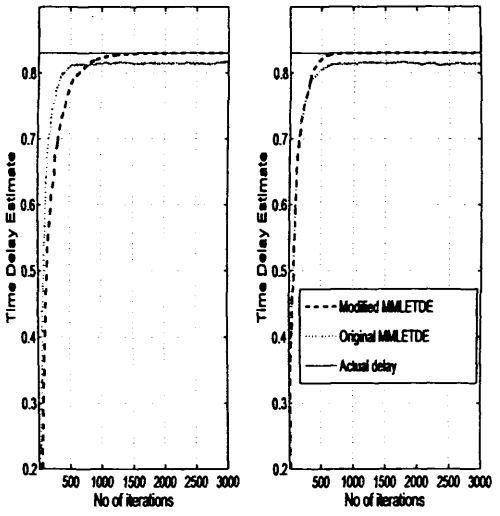

Figure 6.1. The left plot compares the performance of the MMLETDE [3] and the modified version when d = 0.83 and $\mu = 0.05$ for both algorithms. Note that the original MMLETDE has a quicker convergence to the steady state, but has a bias ≈ 0.0159. The right plot demonstrates similar convergence of both algorithms, when $\mu = 0.05$ is utilised for the original MMLETDE and $\mu = 0.1$ for the modified one.

The maximum likelihood delay estimator derived in [106] leads to a delay estimation of the form:

$$\hat{D} = \arg\max_{\hat{D}} \langle s(t - \mathrm{D}),\ s(t - \hat{D}) \rangle \tag{6.3.5}$$

where $\max\langle s(t-\mathrm{D}),\ s(t-\hat{D})\rangle$ corresponds to maximising the cross-correlation between $s(t - \mathrm{D})$ and $s(t - \hat{D})$. Note that in this case the delay must be integer and therefore, this estimator cannot be used in its present form for the estimation of fractional delay. Therefore, it is required to substitute $s(t - \hat{D})$ with $\sum_{k=-M_a}^{k=M_b} \hat{h}(k)\ s(t - k)$ and the MI between $s(t - \mathrm{D})$ and $s(t - \hat{D})$ is

maximised instead of the cross-correlation, i.e.:

$$\hat{D} = \arg\max_{\hat{D}} \left\langle I\left(s(t-\mathrm{D}), \sum_{k=-M_a}^{k=M_b} \hat{h}(k)s(t-k)\right)\right\rangle \tag{6.3.6}$$

where $I(.)$ is the mathematical notation of MI. The correlation delay estimator method is compared with that of MI in Fig. 6.2 on the basis of artificially constructed second order autoregressive-filtered white noise signals (to approximate the TMJ signals) with known inter-signal delay. For a fair comparison, the correlation delay estimator was also modified to be able to estimate the fractional delay by substituting $s(t-\hat{D})$ with $\sum_{k=-M_a}^{k=M_b} \hat{h}(k)\ s(t-k)$ in Eq. (6.3.5). The absolute error between the delay estimates and the true value of D=0.83 was obtained for each signal-to-noise ratio (SNR) value by averaging over 20 independent Monte Carlo simulation runs. Fig. 6.2 shows that the maximization of the MI algorithm yields a consistently better estimate of the fractional delay. From this point, whenever the ML estimator is

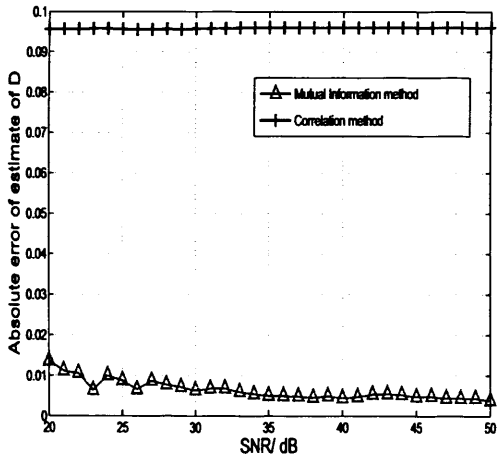

Figure 6.2. Performance of the cross-correlation maximization algorithm and that of the mutual information maximization in terms of their absolute error as a function of SNR when D = 0.83.

referred to, the MI maximisation method is implied. Likewise, MMLEDTE is used to depict the simplified/modified MMLEDTE. Having introduced

the background on fractional delay estimation, the degree of sparsity of the TMJ sounds are next examined.

6.3.2 Sparsity of TMJ sources

Sparsity (or disjointness) in this work refers to the situation where a relatively small number of source signals are active over any particular time interval. For the case of a single active source, sparsity can be mathematically described as

$$\{s_i(t);\ i = 1, ..., n\}$$

$$\text{where } \forall\ t\ \exists\ k \in 1, ..., n \text{ where } |s_k(t)| >> |s_j(t)| \tag{6.3.7}$$

$$\text{and for } j \neq k\ s_j(t) \approx 0$$

where $s_k(t)$ is a given source signal and $s_j(t)$ is another source signal. The time-frequency approach in [5] is adopted to demonstrate that clicks are much sparser than crepitus by comparing Fig. 6.3 and Fig. 6.4. From these figures, the sparsity assumption in (6.3.7) might be reasonable for clicks, but it does not hold for crepitus. Therefore, the sparsity condition is satisfied only for a short time interval, i.e. $\mathrm{T}_1 \leq t \leq \mathrm{T}_2$, with regard to the crepitus signals. Hence it is referred to as time *localised* sparse component analysis. In the following section, the ranges of values over which the delays d and D might vary are examined based on the physiological aspect of the brain and the sampling frequency.

The mixing model as given in equation (6.2.1) in the noise free context may be expanded as:

$$x_1(t) = a_{11}s_1(t - d) + a_{12}s_2(t - d - \mathrm{D}) \tag{6.3.8}$$

$$x_2(t) = a_{21}s_1(t - d - \mathrm{D}) + a_{22}s_2(t - d) \tag{6.3.9}$$

and motivate physical constraints on the values of d and D. The mean width

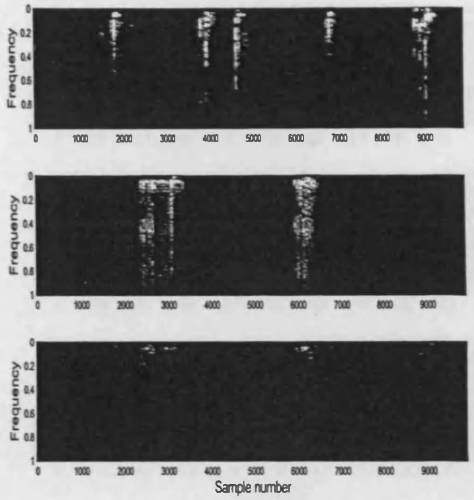

Figure 6.3. The time-frequency (t, f) plots of clicks. From top to bottom: soft click $|s_1(t, f)|$, hard click $|s_2(t, f)|$, and $|s_1(t, f)s_2(t, f)|$. From the last plot, the sparsity of clicks due to the absence of the high magnitude (white regions) is evident.

of the brain is 0.16 m as in [107], while in [108] it is considered as 0.14 m. As a compromise, the mean value of the brain width is taken as 0.15 m. Given the speed of sound within the brain is 1505-1612 ms^{-1} (p. 19 [109]), the differential lag D corresponds approximately to the range of $0.93-1.00 \times 10^{-4}$ s, while due to the proximity of the microphone to the ipsi TMJ source, d can be considered negligible. In terms of the number of samples, D corresponds to the range of 1.1-1.2 samples, with a sampling frequency of 12 kHz. However, the accuracy for this range of D relies on the exact adjustment of the sampling frequency to the size of the head. Therefore, these assumptions are relaxed to account for the changes in the head size as well as the distance from the sources, and these are considered instead for simulation purposes:

$$0 \leq d \leq 0.5 \qquad 0.8 \leq \text{D} \leq 1.5 \tag{6.3.10}$$

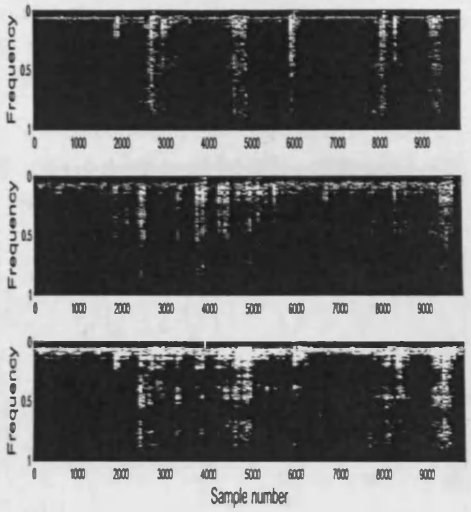

Figure 6.4. The time-frequency (t, f) plots of crepitus. From top to bottom: soft crepitus $|s_1(t,f)|$, hard crepitus $|s_2(t,f)|$, and $|s_1(t,f)s_2(t,f)|$. From the bottom plot, it is obvious that the sparsity of crepitus is not as clear as that of clicks due to the presence of the high magnitudes (white regions). Hence, time localised sparse component analysis will zoom in those regions which satisfy (6.3.7).

Next, the detection of the time interval during which a single sound source is active in the TMJ mixtures is explained.

6.3.3 Blind detection of the active periods of a single source

If only one source prevails in both mixtures during a time interval, the MI between the two mixture signals for that segment is greater than that of the other segments. Since D > 0, it can be deduced that the likelihood function (ML estimator) cannot exhibit a maximum at D=0. Otherwise, it is obvious that during this period ($T_1 \leq t \leq T_2$) both sources exist. Hence, these criteria are employed to blindly detect those regions where the sparsity condition is satisfied. Fig. 6.5 shows the locations (detected by means of these two criteria) where only one source prevails in the mixtures of crepitus. However, even with successful detection of those regions, there are still some issues (i.e. scaling and sign ambiguities inherent to BSS) that

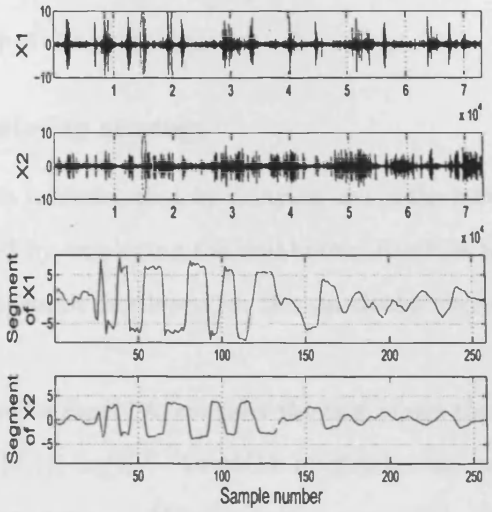

Figure 6.5. The two upper plots show the mixtures (x1 and x2), with the locations of the segments (where the sparsity condition is satisfied) being encircled. In the two lower plots, those segments of the mixtures whose mutual information is maximum are zoomed in. Note that the scale and the sign are different for these segments.

need to be addressed prior to the estimation of the fractional delay.

6.3.4 Scaling and sign ambiguities

To employ MMLEDTE for estimation of the differential delay D, the algorithm requires that both the original signal and its delayed version are accessible. However the mixing process alters the scale and sign of the estimated sources, as shown in the two lower plots of Fig. 6.5. For example, $a_{1j}s_j(t-d)$ and $a_{2j}s_j(t-d-\mathrm{D})$ are available when only the jth source is active within a particular time period in the mixture signals. It is required to solve this scaling problem since the MMLETDE assumes the input to the delay system to be $s(t)$ and the delayed signal to be $s(t-D)$. To tackle the scaling problem, the variances of the segments of the mixture signals (where the sparsity condition (6.3.7) is satisfied) are normalised to unity. On the other hand, if the cross-correlation between the mixtures is negative, the

sign of one of the mixtures is simply altered to ensure both $s(t-d)$ and $s(t-d-D)$ are of the same sign.

6.3.5 The delaying strategy

In this section, it is shown that by delaying one of the mixtures, the sources can be estimated by employing the well-known FastICA algorithm [7]. This is a heuristic approach developed for this particular two-sensor problem.

For a brief review of FastICA, consider the case where there are two mixture signals $\mathbf{x}(t) = [x_1(t)\ x_2(t)]^T$. FastICA maximizes the negentropy of $\mathbf{x}(t)$, more precisely $\arg\max_{w_1,w_2}\Big(\text{Neg}(w_1x_1(t)+w_2x_2(t))\Big)$. It linearly combines the mixtures in order to produce independent components (ICs) in its output, $\mathbf{y}(t)$. The separated jth IC $\mathrm{y}_j(t)$ is expected to resemble one of the sources. Given the differential delay D between the sources, $x_1(t-D)$ is utilised instead of $x_1(t)$ to calculate $y_1(t)$. Therefore, FastICA executes the following expression:

$$\arg\max_{w_1,w_2}\Big(\text{Neg}(w_1x_1(t-\text{D})+w_2x_2(t))\Big) \tag{6.3.11}$$

Expanding the arguments of equation (6.3.11), one can obtain

$$\begin{aligned} y_1(t) &= w_1x_1(t-\text{D})+w_2x_2(t) \\ &= w_1\Big(a_{11}s_1(t-d-\text{D})+a_{12}s_2(t-d-2\text{D})\Big)+ \\ &\quad w_2\Big(a_{21}s_1(t-d-\text{D})+a_{22}s_2(t-d)\Big) \\ &= s_1(t-d-\text{D})\underbrace{\Big(a_{11}w_1+a_{21}w_2\Big)}_{\alpha}+ \\ &\quad \underbrace{w_1a_{12}s_2(t-d-2\text{D})+w_2a_{22}s_2(t-d)}_{\beta s_2(t-d)} \end{aligned} \tag{6.3.12}$$

where $\alpha = a_{11}w_1 + a_{21}w_2$ and $\beta = w_1a_{12} + w_2a_{22}$. It can be shown, through simulation studies, that optimising (6.3.11) yields $\alpha \approx 0$ and that $\beta \not\approx 0$.

Then, $y_1(t) \approx s_2(t\text{-}d\text{-}2\text{D}) + s_2(t\text{-}d)$ (up to a scaling factor). For simplicity, $y_1(t)$ can be considered to be approximately as $\beta s_2(t-d)$ as the maximum value of 2D is three samples, which can be considered to be negligible at a sampling frequency of 12kHz. However, the focus here is to achieve separation of the two TMJ sources, and not deconvolution. On the other hand, Eq. (6.3.11) can be regarded as:

$$\arg \min_{w_1,w_2} \Big(\text{Neg}(\alpha s_1(t-d-\text{D}))\Big) \tag{6.3.13}$$

From Eq. (6.3.12), to have a good *approximation* of $s_2(t-d)$, the most important term is to optimise $\arg\min_{w_1,w_2} \Big(\text{Neg}(\alpha s_1(t-d-\text{D}))\Big)$, which can be shown to be minimised successfully by FastICA through simulation studies.

At this point, recall that $y_1(t) \approx \beta s_2(t-d)$. The second IC $y_2(t)$ is a poor approximation of the second source. FastICA will compute another vector $\mathbf{w}_b$ orthogonal to $\mathbf{w}_a$, while maximising (6.3.11).

$$\begin{aligned} y_2(t) &= \text{s}_1(t-d-\text{D})\Big(a_{11}w_1 + a_{21}w_2\Big) + \\ &\quad a_{12}w_1 s_2(t-d-2\text{D}) + a_{22}w_2 s_2(t-d) \end{aligned} \tag{6.3.14}$$

Recall that the mixture signal $x_1(t)$ was delayed by D samples such that $x_1(t-\text{D})$ was employed as the input to the FastICA algorithm. In other words, $s_1(t-d)$ has also delayed by D samples in $x_1(t)$. This causes the alignment of $s_1(t-d-D)$ in both mixture signals, i.e. $x_1(t)$ and $x_2(t)$. Due to this alignment, FastICA can optimise equation (6.3.13). However, in estimating $s_1(t-d-\text{D})$, FastICA has to perform:

$$\arg \min_{w_1,w_2} \Big(\text{Neg}(w_1 a_{12} s_2(t-d-2\text{D}) + w_2 a_{22} s_2(t-d))\Big) \tag{6.3.15}$$

The misalignment of $s_2(t-d-2\text{D})$ and $s_2(t-d)$ in minimising Neg(.) in (6.3.15) leads to a poor approximation of $s_1(t-d-\text{D})$. The condition that

$\beta \approx 0$ must be satisfied in order to achieve a good estimation of $s_1(t-d-\mathrm{D})$, whilst α reflects the scaling ambiguity of $s_1(t-d-\mathrm{D})$. In short, the misalignment of $s_2(t-d-2\mathrm{D})$ and $s_2(t-d)$ in (6.3.15) will not necessarily satisfy the condition $\beta \approx 0$.

Localisation of the sources: It is desirable to inform the clinician which source corresponds to which TMJ (i.e. left or right). It was just deduced that by delaying $x_1(t)$ by D samples, $s_2(t)$ can be estimated, and not $s_1(t)$. However equation (6.3.8) demonstrates that $s_2(t)$ is the contra (opposite side) source to $x_1(t)$. Therefore, this solves the localisation problem, since it is known which auditory canal corresponds to the sensor for recording $x_1(t)$.

The overall source separation algorithm can be outlined as follows.

For $i = 1, 2$

1. Compute the differential delay D by ML or MMLETDE estimators as in Eq. (6.3.6) and Eqs. (6.3.3) & (6.3.4),
2. Delay $\mathrm{x_i}(t)$ by D,
3. Employ FastICA to maximise $\arg\max_{w_1,w_2}\Big(\mathrm{Neg}(w_i x_i(t-\mathrm{D})+w_j x_j(t))\Big)$, where $i \neq j, j \in \{1,2\}$,
4. If $i = 1$, denote the pair of independent components (ICs) as $y_1(t)$ and $y_2(t)$, otherwise denote the ICs as $y_A(t)$ and $y_B(t)$
5. Estimate the MI between $y_1(t)$ & $y_A(t)$, $y_1(t)$ & $y_B(t)$, $y_2(t)$ & $y_A(t)$, and $y_2(t)$ & $y_B(t)$,
6. Pick the pair with minimum MI as the estimated sources, due to the statistical independence assumption of the sources.

It is understood that the selected IC from the set $\{y_1(t), y_2(t)\}$ corresponds to the same side of the head as $x_2(t)$, and hence the selected IC from the set

$\{y_A(t), y_B(t)\}$ is on the contra side.

6.3.6 Simulations

The following scenario is considered: The patient is suffering from osteoarthrosis [23]. Hence, crepitus is present. It is assumed that one temporomandibular joint is more 'damaged' than the other, thus giving rise to the soft and hard TMJ sources. These TMJ sources (when measured separately) were mixed synthetically by a randomly generated $\mathbf{A}$:

$$\mathbf{A} = \begin{pmatrix} 0.9198z^{-0.15} & -0.2381z^{-1.54} \\ 0.4418z^{-1.54} & 0.9574z^{-0.15} \end{pmatrix}$$

Thus, d=0.15 and D=1.39. In the first place, the improved ML estimator as a function of $\hat{D}$ is illustrated in Fig. 6.6 in the noise-free context. Furthermore, Parra's algorithm [4], the time-frequency approach of Yilmaz and Rickard [5], and the proposed approach are compared in terms of SIR (signal-to-interference ratio) as SNR varies in the upper plot of Fig. 6.7. Also, the corresponding absolute errors between $\hat{D}$ of ML and that of MMLEDTE estimators are monitored. The values of the errors and those of SIRs were obtained by averaging 20 independent simulation runs.

6.3.7 Discussion and concluding remarks

Fig. 6.6 shows that the ML estimator exhibits a maximum near D=1.39 (the *estimated* delay between the sources $\hat{D}$ = 1.38). On the other hand, the upper plot of Fig. 6.7 clearly demonstrates the superiority of the proposed method with the delay estimates of the modified ML and MMLEDTE estimators due to its higher SIR. The much lower performance of Yilmaz & Rickard's algorithm [5] is due to its reliance on k-means clustering (which

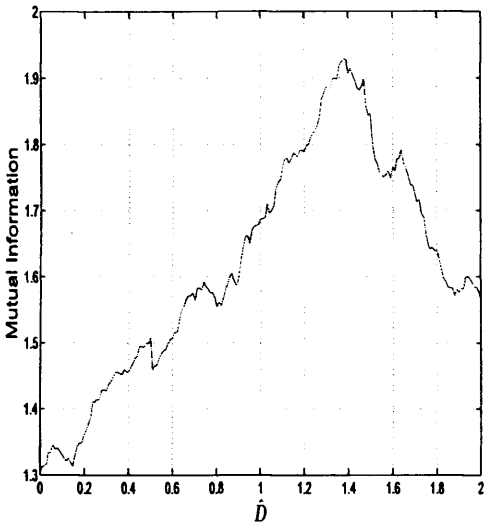

Figure 6.6. Mutual information between the two mixture signals as a function of the delay $\hat{D}$ when sparsity condition (6.3.7) is satisfied. As expected, it has a peak near the true value D=1.39 ($\hat{D} = 1.38$).

in turn depends on the disjointness of the source components in the time-frequency domain and the initialisation of **A**). Parra's algorithm which enjoys a relatively good SIR > 60 dB, uses a long filter length (owing to its frequency approach). This suggests that the filter length of its separating matrix is longer than what is required, which leads to an increase in the level of interference. However notice that due to the orthogonal projections in evaluating SIR proposed by Vincent *et al* [74], it leads to a significant difference ($\approx$ 10 dB) between the proposed method and that of Parra [4]. Also, it is noted that the BSS delay technique with ML estimator $\hat{D}$ enjoys a better performance compared to one with MMLETDE $\hat{D}$. The lower plot of Fig. 6.7 which demonstrates the lower values of errors of ML estimator $\hat{D}$ compared to those of MMLEDTE explains the better performance of the BSS algorithm with ML estimator $\hat{D}$. Finally, the error of the ML estimator $\hat{D}$ in Fig. 6.7 is greater than that in Fig. 6.2. This discrepancy can be due to the fact that the original delayed signal and the undelayed signal are not ac-

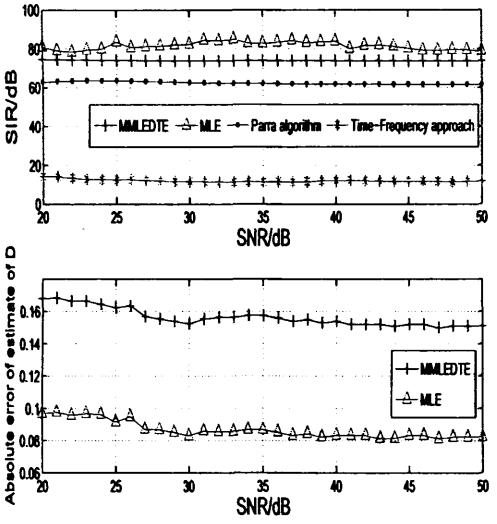

Figure 6.7. The upper plot shows the performances of Parra's algorithm [4], time-frequency approach of Özgur [5], and the proposed method in terms of signal-to-noise (SNR) ratio against signal-to-interference ratio (SIR). The differential delay D=1.39. The lower plot shows the corresponding values of the absolute errors of ML and MMLETDE estimators $\hat{D}$. The error of ML estimator $\hat{D}$ < error of the MMLETDE explains why the SIR of ML estimator > SIR of MMLETDE.

cessible, but the *normalized* segments of the mixtures are utilised. Besides, synthetically generated low pass signals in Fig. 6.2 were used, while here in Fig. 6.7 mixed TMJ signals are considered.

6.4 Separation and localisation of clicks and normal TMJ sound

In contrast to the previous study, TMJ BSS is *not* simulated here via synthetic mixing of TMJ sounds. In other words, source separation on the bilaterally recorded pair of TMJ sounds is attempted. In particular, the case where there is only one defective TMD joint, with the other joint as a healthy/normal one is considered. This scenario has already been addressed by Widmalm *et al.* [32, 33]. However, these researchers regarded the 'echo' recorded on the contra (opposite) side of the TMD joint as the lagged version of the TMD source. In this work, the possibility that this echo can in

fact be a mixture of the TMD source and the sound produced by the normal TMJ is investigated.

In this work, the main assumptions made are 1) all the source signals are statistically non-stationary and sparse in the time-frequency domain; and 2) the mixing model (6.2.1) holds with mixing matrix **A** (6.2.2).

The aim here is to address the problem of source separation of real TMJ sources. In the first part of this chapter, TMJ source separation problem by simulating an anechoic model of the head was addressed; while in Chapter four, the mixing model was an instantaneous and underdetermined one. These studies were however carried out based on a synthetic mixing of the TMJ sources. Guo *et al.* separated the biologically-mixed TMJ sources by modelling the acoustic mixing system of the head as a convolutive one [1]. However, there exists one fundamental issue which has not been addressed by any of these three studies, i.e. the statistical dependence of the TMJ sources. Whenever a person chews, both joints operate in a synchronous fashion. Therefore the statistical independence assumption of the TMJ sources is questionable. In contrast to these approaches, a time-frequency masking approach is adopted to perform source separation on the biologically-mixed TMJ sounds. This approach is more suitable for this particular source separation, as the statistical independence of the sources is not assumed. The extensive literature on the time-frequency analysis of the TMJ signals also supports this approach and it will be discussed in the next section [34], [35], [36], [37], [110].

The organisation of this second part of the chapter is as follows; in the next section an overview of the works on the temporomandibular disorder using time-frequency analysis is given. This is followed by a brief background on the DUET algorithm and in particular, one of its non-trivial constraint

will be addressed and circumvented. This section is closed with the summary of the proposed source separation procedure. Thereafter, experimental results are presented to verify the effectiveness of the source separation technique. Subsequently, concluding remarks on the estimated TMJ sources are provided in section 6.4.5.

6.4.1 Time-Frequency analysis of TMJ sounds

Several time-frequency (T-F) analysis methods have already been performed (with TMJ sounds) mainly for classification purposes [34], [35], [36], [37], [110]. The success of these T-F approaches for classification of TMJ signals stems mainly from the statistical non-stationarity nature of TMJ sounds. Hence, as many researchers argue, these approaches pick up features that are not seen in the waveforms or in conventional power spectra [35], [110], [111]. Many of these approaches have their pros and cons. For example, the reduced interference distribution (RID) of the Cohen's T-F family does not guarantee a non-negative distribution, while suppresssing interference and cross terms [35]. On the other hand, the main appeal of the short time Fourier transform (STFT) is its simplicity to use, but it does not yield a high time-frequency resolution. However, the time-frequency resolution of the STFT can be improved via the evolutionary spectrum proposed by Akan *et al.* [35]. Throughout this study, the short time Fourier transform is utilised to perform the time-frequency analysis of the TMJ sounds due to its simplicity. Nevertheless, it is understood that the focus here is not to discuss the optimum T-F approach, but to demonstrate that the statistical non-stationarity property of the TMJ sources can be exploited to solve BSS. Hence, the well known DUET algorithm which achieves source separation via T-F binary masking is employed in this study. In the following section, an overview of the DUET algorithm together with an explanation of how to accommodate this algorithm in the context of TMJ BSS are given.

6.4.2 The DUET algorithm in the context of TMJ BSS

It is noted that this section is based on the materials provided in [33] and [5]. The main hypothesis implied by the DUET approach is the W-disjoint orthogonality [5]. This concept can be viewed as sparsity which in turn is defined in section 6.3.2. Here, however, sparsity is analysed in the time-frequency domain and the two sources are click sound source, generated from a TMD joint, and a 'normal' TMJ sound from that of the free-TMD joint. Gay *et al.* and Gallo *et al.* found that most of the energy of a 'normal' TMJ sound was centered below 800 Hz [6,39], while TMD joint sounds exhibited peaks greater than 800 Hz, see for instance Fig. 5 in [6]. Furthermore, the time interval between the ipsi and contra sources contributes to the sparse combination of the two TMJ sources. On this basis, the reasonable hypothesis that the TMD source does not overlap the 'TMD-free' source in the time-frequency domain is made. Furthermore, Yilmaz and Rickard demonstrated the robustness of their DUET technique even when the speech sources satisfy a *weakened* version of W-disjoint orthogonality condition. Due to this sparsity assumption of only a single active source in a particular time-frequency (t, f) interval, the estimate of the jth source is

$$\hat{s}_j(t,f) = B_j(t,f)x_1(t,f) \tag{6.4.1}$$

where the binary mask is defined as

$$\begin{aligned} B_j(t,f) &= 1 \quad \text{if } |\hat{s}_j(t,f)| \gg |\hat{s}_k(t,f)| \quad j \neq k \\ &= 0 \quad \text{otherwise} \end{aligned} \tag{6.4.2}$$

To establish which source is active within a particular (t, f) interval or equivalently determination of the binary mask, the following can be performed.

In the time-frequency domain, the mixing model can be expressed as

$$\begin{bmatrix} x_1(t,f) \\ x_2(t,f) \end{bmatrix} = \begin{bmatrix} a_{11}z^{-\delta_{11}} & a_{12}z^{-\delta_{12}} \\ a_{21}z^{-\delta_{21}} & a_{22}z^{-\delta_{22}} \end{bmatrix} \begin{bmatrix} s_1(t,f) \\ s_2(t,f) \end{bmatrix} \tag{6.4.3}$$

Due to the sparsity assumption, there is at most one active source in a particular (t, f) interval. Consider all (t, f) intervals where only the jth source prevails as $U_j := \{(t,f) : B_j(t,f) = 1\}$ and the following ratio

$$R_{21}(t,f) = \frac{x_2(t,f)}{x_1(t,f)} \tag{6.4.4}$$

Regarding all the (t, f) intervals within U_j, it is clear that this ratio is

$$R_{21}(t,f) = \alpha_j e^{-i\delta_j\omega} \tag{6.4.5}$$

where $\alpha_i = a_{2i}/a_{1i}$, $\delta_i = \delta_{2i} - \delta_{1i}$, and $\omega = 2\pi f$. This ratio can then be utilised to determine $\alpha_j = |R_{21}(t,f)|$ and $\delta_j = -(1/\omega)\angle R_{21}(t,f)$, where $|.|$ and $\angle$ denote respectively, the magnitude and the phase. These two features, i.e. α_j and δ_j, computed over the entire time-frequency domain, can then be used to compute two cluster centers corresponding to the two sources. The clustering procedure can be performed by the k-means algorithm [72]. Thus, a (t, f) interval pertaining to a particular active source is equivalent to its membership to the corresponding cluster center. On this basis, the binary mask can be determined to estimate the source via (6.4.1). Next, a non-trivial constraint imposed by the DUET algorithm is examined.

DUET requires the sensor spacing to be less than the distance the TMJ sound travels within one sample, if $Fs = 2f_{max}$, where Fs and f_{max} denote respectively the sampling frequency and the maximum frequency of the source signal. This situation is impossible to implement for when the sensors are placed in the auditory canals. Mathematically, the constraint C can be

expressed as [5]:

$$C = |\omega_{f\max}\tau_{d\max}| < \pi \tag{6.4.6}$$

where $\omega_{f\max} = 2\pi f_{\max}$ and $\tau_{d\max}$ is the maximum time lag determined by the microphone spacing [5]. We can also express (6.4.6) as

$$|\tau_{d\max}| < \frac{1}{2f_{\max}} = \tau_{f\max} \tag{6.4.7}$$

With reference to the work of Widmalm in [33], $\tau_{d\max}$ of the TMJ sounds within the brain was found to be in the range of 0.2-1.2 ms, while $\tau_{\omega\max} = 1/(2 \times 3200) = 0.16$ ms, assuming $f_{\max} = 3200Hz$. As the sampling frequency employed in this study is 12 kHz, $\tau_{d\max}$ corresponds to 2.4 - 14.4 samples and $\tau_{\omega\max} = 1.9$ samples. Likewise, the fact that TMJ signals are oversampled by a factor of approximately 2, i.e. $(Fs/f_{\max}) \approx 2^2$ also explains why the delay between the microphones should be less than two samples. Therefore, the constraint of DUET is not fulfilled in this particular TMJ BSS scenario. Since one does not have access to $\tau_{\omega\max}$, which depends on the maximum frequency $f_{\max}$ present in the sources, one cannot alter its value. However, $\tau_{d\max}$, which is governed by the 'spacing' between the two sensors, can be made to be approximately to zero. Equivalently, the delay introduced by this 'spacing' is simply the delay between the two mixtures. This delay can be cancelled, if one of the mixtures is delayed by τ_{dmax}. As Widmalm showed in his work [33], the TMD joint sound propagates to the contra side in $\tau_{d\max}$ and he considered this contra laterally recorded 'echo' as the delayed version of the TMD joint sound. Subsequently, the mixture recorded in the auditory canal of the TMD side is delayed. The determination of which side of the face corresponds to the TMD joint can be achieved by examining the sign of the delay τ_{dmax} which will be discussed next.

Localisation of the sources: As mentioned before, it is desirable to in-

form the clinician which source corresponds to which TMJ (i.e. left or right). Mutual information (MI) can be utilised as a measure of similarity between one estimated source and a mixture. For example, source $s_1(t)$ is considered to be on the same side as $x_1(t)$ if MI($s_1(t), x_1(t)$)>MI($s_1(t), x_2(t)$), otherwise $s_1(t)$ is located on the adjacent side. Therefore, this solves the localisation problem, since the clinician knows in which auditory canal he has placed the sensor for recording $x_1(t)$ or $x_2(t)$. In the following section, the overall algorithm, which is similar to the source separation procedure given in section 6.3.5 is outlined for the purpose of clarity.

6.4.3 Summary of the delay strategy

1. In the first place, consider for the moment that $x_1(t)$ corresponds to the TMD joint side,

2. Compute the delay D of $x_2(t)$ relative to $x_1(t)$ by utilizing the modified delay estimator ML estimator (6.3.6),

3. If D> 0, then $x_1(t)$ corresponds in fact to the TMD side and denote it as $x_{\text{TMD}}(t)$, otherwise $x_1(t)$ corresponds to the *contra* side of the TMD and label it as $x_{\text{nonTMD}}(t)$.

4. Delay $x_{\text{TMD}}(t)$ by D. This can be achieved via (6.3.1),

5. Apply DUET algorithm to $x_{\text{TMD}}(t\text{-D})$ and $x_{\text{nonTMD}}(t)$ to compute $R_{21}(t, f)$ in Eq. (6.4.5) over the entire time-frequency domain and cluster these (t, f) intervals into two classes of α_j and δ_j. From this clustering procedure, the binary mask $B_j(t, f)$ pertaining to the jth source can be built. As a result, $y_1(t)$ and $y_2(t)$ will be estimated via the binary mask (6.4.1),

6. Estimate the MI between $y_1(t)$ & $x_{\text{nonTMD}}(t)$, and $y_1(t)$ & $x_{\text{TMD}}(t)$,

7. If $\text{MI}\Big(\text{y}_1(t),\ x_{\text{TMD}}(t)\Big) > \text{MI}\Big(\text{y}_1(t),\ x_{\text{nonTMD}}(t)\Big)$, then $y_1(t)$ is the

estimate of the TMD source and $y_2(t)$ is the estimate of the non-TMD source, and vice-versa.

6.4.4 Experimental results

The experimental results presented herein follows from the data recorded on a patient with his left TMJ generating 'clicks on closure', at a sampling frequency of 12 kHz, similar to the sampling frequency used by Sano *et al.* [31]. For comparison purposes, the source estimates of Infomax are also included [1]. In this section, EDUET is used to denote the proposed method, while DUET refers to the original DUET approach. In this section, EDUET is used to denote the proposed method, while DUET refers to the original DUET approach.

Fig. 6.8 illustrates a peak value of mutual information at D=10.3 samples, when the relative delay between the two TMJ sounds was measured. In the upper plot of Fig. 6.9, the mixtures are plotted and the two lower plots zoom in on the prominent peaks of both TMJ sounds. Fig. 6.10 presents the estimates of the proposed approach EDUET in the upper two plots, while the lower two plots portray the estimates of Infomax. Fig. 6.11 illustrates the estimated sources by DUET algorithm. On the other hand, Fig. 6.12 illustrates the spectra of the TMJ mixtures in the left plot and the right plot shows respectively the spectra of the estimated TMJ sources for comparison purposes.

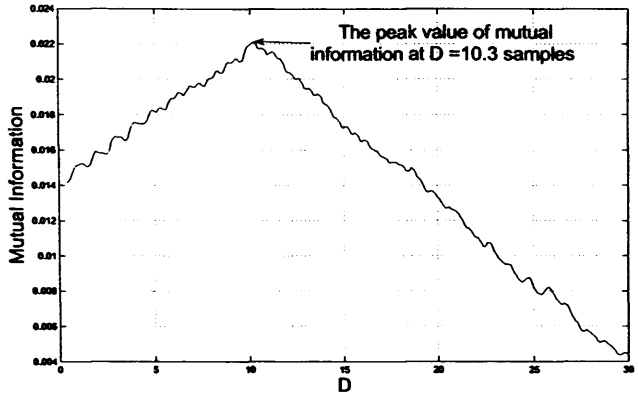

Figure 6.8. Mutual information between the two mixture signals as a function of the delay D. It has a peak at D=10.3 which corresponds to the lag between the prominent peaks of the two mixtures. These peaks are highlighted with the aid of arrows in Fig. 6.9.

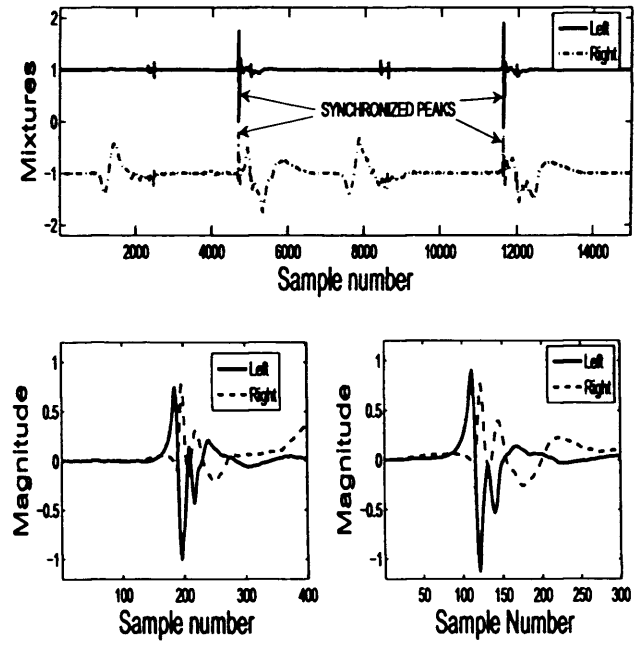

Figure 6.9. The upper plot shows the two TMJ mixtures, one underneath the other one to illustrate their synchronized mechanism. This is clear by the coincidence of the prominent peaks of both TMJ sound mixtures pointed out by the arrows. These peaks are zoomed in the lower two plots. The time difference between the left and the right TMJ mixtures corresponds to approximately 11 samples, which is in agreement with the peak in Fig. 6.8.

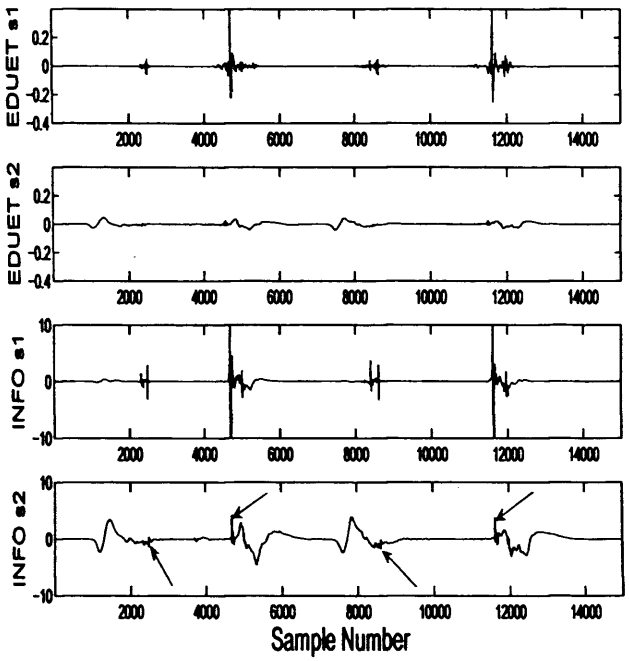

Figure 6.10. The upper two plots show the estimated sources by our proposed approach EDUET. Note the absence of those prominent peaks in $s_2(t)$, which suggests that it corresponds to noise generated by the healthy joint. The lower two plots illustrate the estimates of convolutive Infomax [1]. The arrows point the components present in both estimates of Infomax, indicating the presence of clicks in both estimates.

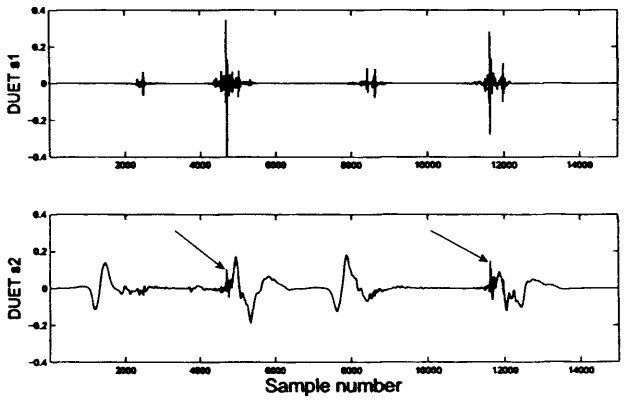

Figure 6.11. Estimated sources using DUET algorithm without alignment of the mixtures; the clicks, pointed out by arrows, can be viewed in the estimated $s_2(t)$ as the normal TMJ sound.

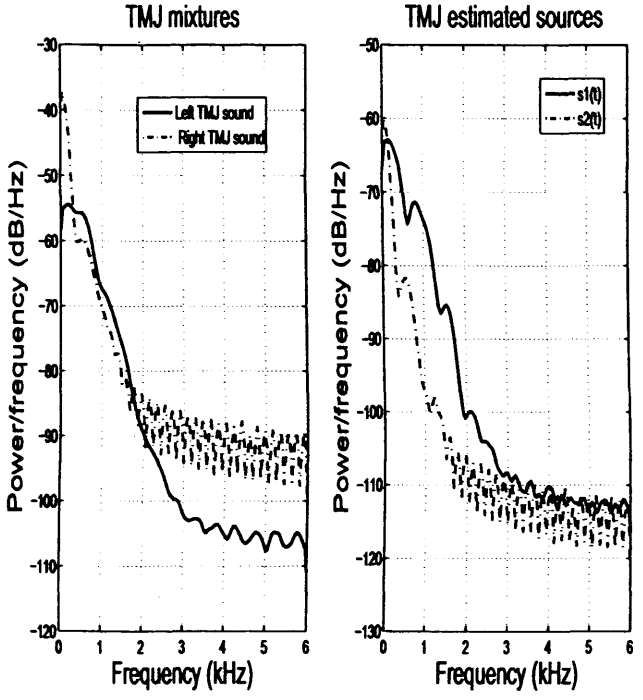

Figure 6.12. The spectra via the Welch power spectral density method of the TMJ mixtures in the left-hand plot and of the estimated EDUET sources in the right-hand plot. Take notice of the closeness of the two TMJ mixtures spectra for the interval from 800 Hz to 1500 Hz, indicating that the click is present in both TMJ sounds. Also, it is noteworthy to say that the spectrum of $s_2(t)$ is severely attenuated for frequencies greater than 800 Hz compared with that of $s_1(t)$, suggesting the successful extraction of the clicks from the right TMJ sound. Note the strong similarity between the spectra of the estimated sources and Fig. 5 in [6] where the authors compared the spectrum of a 'normal' TMJ sound with that of click sounds.

6.4.5 Discussion and concluding remarks

Fig. 6.8 shows that the time delay estimator exhibits a maximum at D=10.3 samples which corresponds to the delay between the two TMJ sounds. This delay is equivalent to 0.86 ms and within the range of 0.2-1.2 ms found by Widmalm *et al.* [33]. From the topmost plot of Fig. 6.9, the synchronicity of two TMJ sounds is evident. If the prominent peaks of the TMJ sounds are zoomed in on the lower two plots, it is clear that the right TMJ sound lags the left TMJ sound. This confirms that it is the left TMJ that generates clicks, as expected. On the other hand, the upper two plots of Fig. 6.10 illustrate the estimated EDUET sources by the proposed approach. The signal $s_1(t)$ is evidently the click source, while the absence of those promiment peaks in $s_2(t)$ suggests that it is in fact the sound produced by the healthy/normal joint. In the same figure, both estimates of convolutive Infomax contain components of click as pointed out by the arrows in the lower two plots. The measured MI between the two TMJ mixtures, the Infomax estimates, the DUET estimates, and the EDUET estimates were 0.594, 0.445, 0.0257, and 0.540 respectively. The lower values of mutual information between the pair of Infomax estimates and between the DUET estimates show that Infomax and DUET achieves a better degree of statistical independence between their respective estimates than the EDUET approach. Nevertheless, both Infomax and DUET estimates still contain components of click as shown in Fig. 6.10 and 6.11. Now, examine the spectrum of the TMJ sound mixtures in the left-hand plot of Fig. 6.12.

Since it is the right TMJ which is 'normal', its spectrum is expected to have a similar shape to that of a 'normal' TMJ illustrated in Fig. 5 of [6] and in the only figure of [39], which is indeed the case. However, note the closeness of the right TMJ spectrum with that of the left TMJ spectrum from the range of about 800 Hz to 1500 Hz. This indicates that the right

TMJ spectrum is in fact contaminated with the clicks from the left TMJ. This was evident in the study of Widmalm *et al.* [32], [33]. Likewise, the spectra of the estimated sources are compared in the right-hand plot of Fig. 6.12. As expected, the spectrum of $s_2(t)$ (considered as the healthy TMJ sound) is severely attenuated above 800 Hz which can also be observed in Fig. 5 of [6] and in the only figure of [39] for the case of a 'normal' TMJ sound. Similarly, the spectrum of $s_1(t)$ in the proximity of 1kHz is much higher than $s_2(t)$, indicating that $s_1(t)$ corresponds to the sound generated by the TMD side. The human ear is most sensitive to the frequency range of 1000 - 3200 Hz. Therefore, Fig. 6.12 indicates that $s_1(t)$ considered as the click source, has a better audible prominence as compared to $s_2(t)$ generated by the 'normal' joint. Therefore, this successful separation and localisation of TMJ sources can be of great clinical value to the dental specialists.

6.5 Conclusions

In the first part of this chapter, sparsity of the crepitus in time domain was exploited to estimate the differential time D between TMJ sources, while in the last part, sparsity of the sources in time-frequency domain was employed for clustering (t, f) intervals pertaining to a particular source. However, both lead to the same objective, i.e. separation of the TMJ sources. In the last part of section 6.3.2, it was deduced that based on speed of sound within the brain, mean width of the brain, and the sampling frequency used, the lag D was approximately 0.1 ms. In contrast to this, the study on the separation of click and 'healthy' TMJ sources indicates that this lag corresponds to 0.86 ms, which is well within the range of 0.2-1.2 ms found by Widmalm *et al.* [33]. This discrepancy from the two studies presented in this chapter does show that there is a difference between theory and practice. However, the lower limit of D found by Widmalm, i.e. 0.2 ms is in close range to the theoretical one, i.e. 0.1 ms. Another major discrepancy between the two studies is

in the performances of DUET algorithm in the two scenarios. In the first part of this chapter, DUET yields a poor performance on the synthetically mixed crepitus sources, while in the second part, DUET separates the click source from the healthy TMJ source. This was expected, as it was already acknowledged in the literature that click and the 'healthy' TMJ source have quite dissimilar spectra [6, 39]. On the other hand, the strong similarity of the soft and hard crepitus sources' spectra explains the poor performance of the DUET algorithm.

Chapter 7

CONCLUSIONS AND FURTHER RESEARCH

7.1 Summary and Conclusions

This study has presented novel methodologies to improve the existing BSS techniques. These can be summarised as follows:

1. A variable tap length for convolutive time domain Infomax algorithm.
2. Application of Ferrante's affine transformation to improve the convergence properties of fixed point iteration for a particular ICA algorithm.
3. Filtering as a preprocessing step to extend the FastICA algorithm to the underdetermined BSS.
4. Exploitation of fractional delays to solve BSS of TMJ sounds.

The first contribution has provided a stepping stone in terms of a novel topic for the BSS community to indulge in. The flexibility of the variable tap length will hopefully attract more researchers to dedicate themselves to time-domain convolutive algorithms. These algorithms are generally dismissed by the BSS community, due to the more efficient frequency domain approaches. The variable tap length concept is likely to have a major impact in applications where thousands of taps of the filters are required.

The second contribution has shed light on how to apply fixed point con-

cepts such as attractive fixed point and contraction mapping theorem to fixed point ICA algorithms. The philosophy behind this contribution is to demonstrate the applicability of fixed point theory in BSS and should be applied more rigorously to BSS algorithms. The independence of fixed point iterations from the learning step-size (to which gradient approaches are vulnerable) further advocates its usage.

The third contribution illustrates how a pre-processing filtering of the mixture signals can prove to be useful in solving underdetermined cases. In this contribution, it was shown how filtering attenuates the presence of certain sources within the mixtures, thus allowing the identification of the other sources. This is analogous to the concept of sparsification, whereby the data are projected in a 'sparse' domain. It is crucial that the sources are sparse within a particular interval in the sparse domain. As a result, underdetermined BSS reduces to a more tractable exactly-determined BSS model. In fact, underdetermined BSS currently figures amongst the hot topics. The enthusiasm shown for this topic is unlikely to wane off, due to the challenge posed by its ill-conditioned nature, and its practical applicability.

The fourth contribution tackles fractional delays arising in source separation. Modelling a fractional delay theoretically implies the convolution of the signal to be delayed with an infinitely long delayed sinc function. Therefore, correct estimation of the fractional delay is crucial in optimising an appropriate filter length, to be as short as possible. This will circumvent thousands of taps often needed to implement the fractional delay.

On the forefront of TMJ source separation, it can be deduced that the mixing model within the human head is closer to an anechoic model than a convolutive model. Although instantaneous TMJ BSS has been simulated, it has been confirmed that the non-stationarity or the non-Gaussianity can

be exploited. In that respect, the simulation of instantaneous BSS has been instructive. In Chapter six, the delay found from the bilateral recordings of TMJ signals corroborates with the delay range reported by Widmalm *et al.* [33]. Additionally, the spectra of the estimated sources are comparable to those reported in the literature [6, 38, 39]. These indicate that successful extraction of the TMJ sources has been achieved. This novel extension to DUET allows for an *arbitrary* spacing of the sensors, which might have other potential applications such as in speech source separation. Furthermore, Yilmaz and Rickard demonstrated that DUET had a reasonable performance in a convolutive scenario where the reverberation was 500 ms and five sources were present (underdetermined source separation) [5]. In contrast to this scenario, firstly, it is unlikely that the reverberation within the head is as long as 500 ms, secondly, it is known a priori that there are only two sources (exactly determined source separation). Hence, the scenario presented herein is 'simpler' and justifies the use of the DUET in the separation of the TMJ sources. In conclusion, the non-stationarity of the TMJ signals has to be exploited in a non-statistical fashion due to the synchronised mechanism of the two joints. Effectively, this means that the statistical independence of the sources might not be a suitable assumption for TMJ source separation.

Last but not least, the human factor is not to be excluded in the consideration of the work undertaken herein. The ethical protocols to record the database on a more diverse scale have limited this study to mostly synthetic simulations. The research fields of dentistry and signal processing are unfortunately two distinct bodies. The lack of mutual cooperation and understanding between the workers in the two disciplines is one of the numerous examples to illustrate why subjects such as telemedicine, or automated prognosis have failed to launch.

7.2 Future works

Many opportunities for future research await to be explored. With regard to the variable tap length, only one tap length is varied in the convolutive Infomax algorithm. This can possibly be extended to multiple variable tap lengths, and achieved through the measurement of a non-Gaussian criterion of each estimated source. Optimising each tap length in accordance with maximisation of the non-Gaussianity of individual estimated sources looks to be an optimistic avenue. Furthermore, the concept of variable tap length is by no means restricted to convolutive Infomax. Indeed, it can be applied to other time domain convolutive algorithms such as the natural gradient approach described in [80].

Ferrante's affine transformation has numerous avenues for development. The algorithm proposed in Chapter four should be further examined and extended to the case of a higher number of sources. Coupled with this, it can be improved to cater for convolutive BSS. This can be undertaken by employing the feedback network, following the approach of Torrkola for the convolutive Infomax algorithm. Moreover, the proposed approach can also be employed in parallel with SCA such that it estimates the active sources over a given interval of time, provided the number of active sources does not exceed the number of mixture signals. It will then potentially have the capability to solve the underdetermined BSS.

The linear filtering UBSS approach in Chapter five has been utilised in the context of TMJ BSS, and therefore leaves room for other applications for which similar conditions arise. However, the main challenge remains in finding more optimum linear transforms that perform source attenuation within the mixture signals in the time domain, thereby reducing the number of active sources. These linear transforms can then compete against sparsifica-

tion methods, which project the data into its 'sparse' domain.

Finally, TMJ source separation on a more wider database looks promising, provided the ethical approval is granted, and not stalled. The *non-invasive* acquisition of TMJ signals should not in theory pose a problem. Research assertiveness should be encouraged, when it comes to those administrative issues. As for blind source separation itself, in my opinion, it has attained its ceiling. Although BSS research has been carried out over two decades, there has not been a significant breakthrough in this field. BSS researchers should start thinking outside the 'box'. The reliance on the intelligence of computers via programming has partly limited BSS research. Nowadays, its main innovations lie in its application rather than on the algorithmic level. As mentioned in the introduction, the capability of the human brain to *signal process* multimodal data, together with its independent learning ability explains why it remains as the most reliable BSS solver. Thus, BSS researchers must comprehend how the brain operates, rather than indulge in formulating mathematically elegant algorithms. For example, the hierarchical temporal memory (HTM) approach [112], which mimics the learning process of the brain can play a significant role in BSS.

Appendix A

CONCEPTS AND DERIVATIONS FOR BLIND SOURCE SEPARATION

A.1 Fixed Point Theorems

The materials presented in this section are based on [57] and [2].

A.1.1 Types of Fixed point

1. A fixed point $\mathbf{u}$ of a function $f(.)$ is said to be **asymptotically stable or attractive**, if

$$\lim_{k \to \infty} f(\mathbf{u}_k) = \mathbf{u} \quad \forall \ \mathbf{u}_k \in \Theta(\mathbf{u})$$

   where $\Theta(\mathbf{u})$ denotes the neighbourhood of the fixed point $\mathbf{u}$. The magnitude of each of the eigenvalues of the Jacobian $\mathbf{J}$ of $f(.)$ at $\mathbf{u}$ is less than unity.

2. A fixed point $\mathbf{u}$ of a function $f(.)$ is said to be **repulsive**, if eigenvalues of $\mathbf{J}$ of $f(.)$ are less than unity in magnitude at $\mathbf{u}$.

3. A fixed point $\mathbf{u}$ of a function $f(.)$ is said to be **saddle point**, if some of its eigenvalues of $\mathbf{J}$ of $f(.)$ are less than unity and some are greater than unity in magnitude at $\mathbf{u}$.

A.1.2 Contraction mapping theorem

Theorem **G.1.2.** [57]: Consider M be a closed subsect of $\mathbb{R}^n$ such that

(i) $f(.) : M \rightarrow M$

(ii)$\exists\ \alpha < 1$ such that $||f(\mathbf{u}) - f(\mathbf{v})|| \leq \alpha||\mathbf{u} - \mathbf{v}||$, $\forall \mathbf{u}, \mathbf{v} \in M$

then the fixed point iteration $\mathbf{u}_{k+1} = f(\mathbf{u}_k)$ converges to the unique fixed point $\mathbf{u}$, for any initial value $\mathbf{u}_0 \in M$.

A.2 Derivation of Convolutive Infomax

The relationship between pdfs of the mixture signals $f(\mathbf{x})$ and the outputs of the neural network $f(\mathbf{u})$ is [48]

$$f(\mathbf{u}) = \frac{f(\mathbf{x})}{\det \mathbf{J}} \tag{A.2.1}$$

where det stands for determinant, and $\mathbf{J}$ is the Jacobian matrix of the network, more precisely:

$$\mathbf{J} = \begin{bmatrix} \frac{\partial u_1}{\partial x_1} & \frac{\partial u_1}{\partial x_2} \\ \frac{\partial u_2}{\partial x_1} & \frac{\partial u_2}{\partial x_2} \end{bmatrix} \tag{A.2.2}$$

The determinant of $\mathbf{J}$ can be found as

$$\det \mathbf{J} = \frac{\partial u_1}{\partial x_1}\frac{\partial u_2}{\partial x_2} - \frac{\partial u_1}{\partial x_2}\frac{\partial u_2}{\partial x_1} = u_1' u_2' D \tag{A.2.3}$$

where

$$D = \frac{\partial y_1}{\partial x_1}\frac{\partial y_2}{\partial x_2} - \underbrace{\frac{\partial y_1}{\partial x_2}\frac{\partial y_2}{\partial x_1}}_{0} = w_{110}w_{220} \qquad u_i' = \frac{\partial u_i}{\partial y_i} \quad i = 1, 2$$

It turns out that D is the determinant of the instantaneous separating matrix $\mathbf{W}$. Since it is desired to maximise the output entropy $H(\mathbf{u})$ by employing

(A.2.1), it can be expressed as:

$$H(\mathbf{u}) = E\{\sum_{i=1}^{2} \log |u_i'|\} + E\{\log D\} + H(\mathbf{x}) \tag{A.2.4}$$

Taking the stochastic gradient of (A.2.4), and thereby dropping the expectation operator $E\{.\}$, the adaptation rule for each separating matrix coefficient can be derived. For example, the gradient w.r.t. w_{110}:

$$\nabla w_{110} = \sum_{i=1}^{2} \frac{1}{u_i'} \frac{\partial u_i'}{\partial w_{110}} + \frac{1}{D} \frac{\partial D}{\partial w_{110}} \tag{A.2.5}$$

where

$$\begin{aligned}
\frac{\partial u_1'}{\partial w_{110}} &= \frac{\partial u_1'}{\partial u_1} \frac{\partial u_1}{\partial y_1} \frac{\partial y_1}{\partial w_{110}} = \Phi(y_1) u_1' x_1 \\
\frac{\partial u_2'}{\partial w_{110}} &= \frac{\partial u_2'}{\partial u_2} \frac{\partial u_2}{\partial y_2} \frac{\partial y_2}{\partial w_{110}} = 0 \\
\frac{\partial D}{\partial w_{110}} &= w_{220} \quad \text{and} \quad \Phi(y_i) = \frac{\partial}{\partial u_i} \frac{\partial g(y_i)}{\partial y_i}
\end{aligned}$$

where $u_i = g(y_i)$. Subsequently, the resulting increments can be generalised as

$$\begin{aligned}
\Delta w_{ii0} &\propto \Phi(y_i(t)) x_i(t) + 1/w_{ii0} \\
\Delta w_{iip} &\propto \Phi(y_i(t)) x_i(t-p) \\
\Delta w_{ijp} &\propto \Phi(y_i(t)) y_j(t-p) \quad i \neq j, \quad \forall i, j, p
\end{aligned} \tag{A.2.6}$$

A.3 Equivalence between non-stationarity and super-Gaussianity

Consider the non-stationary signal $x(t)$ as:

$$x(t) = \kappa(t) u(t) \tag{A.3.1}$$

where $u(t)$ is a zero-mean stationary stochastic process, statistically independent from $\kappa(t)$, which is a scaling factor variable over time. The objective is to demonstrate that the kurtosis of the stationary process $u(t)$ is always less than that of the non-stationary $x(t)$, except in the following case. If the scaling factor is generated from a degenerate pdf, i.e. $p(\kappa) = \delta(\kappa - \ell)$, then $kurt(x(t)) = kurt(u(t))$, for any arbitrary constant ρ. Kurtosis $kurt(.)$ which measures the thickness of the distribution can be defined as [102]:

$$kurt(x(t)) = \frac{E\{x^4\}}{E\{x^2\}^2} \tag{A.3.2}$$

Due to the statistical independence between $u(t)$ and $\kappa(t)$, the kurtosis of $x(t)$ can be factorised as follows [8]:

$$kurt(x(t)) = kurt(u(t))\frac{E\{\kappa^4\}}{E\{\kappa^2\}^2} \tag{A.3.3}$$

If the scaling factor $\kappa(t)$ is a positive r.v., then the following holds:

$$\int_0^\infty [\kappa^2 - \ell^2]^2 \, p(\kappa) \;\; d\kappa \geq 0 \tag{A.3.4}$$

where $p(.)$ refers to pdf. Equality holds only when the integrand vanishes everywhere, i.e. *iff* $p(\kappa)$ vanishes, except for $\kappa^2 = \ell^2$. Therefore, the distribution must be degenerate, i.e. $p(\kappa) = \delta(\kappa - \ell)$, if equality holds for (A.3.4). It can be re-written as:

$$\int [\kappa^4 - 2\kappa^2\ell^2 + \ell^4] \, p(\kappa) \;\; \partial\kappa = E\{\kappa^4\} - 2E\{\kappa^2\}\ell^2 + \ell^4 \geq 0 \tag{A.3.5}$$

This has a minimum at $\ell = E\{\kappa^2\}^{1/2}$. Substituting this yields

$$\frac{E\{\kappa^4\}}{E\{\kappa^2\}^2} \geq 1 \tag{A.3.6}$$

From this equation, it is clear that

$$kurt(x(t)) = kurt(u(t))\frac{E\{\kappa^4\}}{E\{\kappa^2\}^2} \geq kurt(u(t)) \qquad \text{(A.3.7)}$$

This concludes the proof of Parra and Spence to demonstrate the super-Gaussianity of a non-stationary process [8].

Appendix B

AN ML ESTIMATOR FOR SHAPE PARAMETER OF THE GENERALISED GAUSSIAN DISTRIBUTION

The same approach as Vetterli [90] is adopted to derive the shape parameter α, but with $\sigma = 1$. Define the likelihood function of the signal vector $\mathbf{u}=[u(1)u(2)...u(t)..u(T)]$ having independent samples as

$$\mathrm{L}(\mathbf{u};\sigma,\alpha) = \log \prod_{j=1}^{T} p(u(j);\sigma,\alpha) \tag{B.0.1}$$

The likelihood $L(.)$ for one sample is derived for simplicity, noting that $\mathrm{L} = \sum_{j=1}^{T} L(j)$. From (4.2.1), one can show that

$$L(u(j);\sigma,\alpha) = \log(\alpha) - \frac{3}{2}\log(\Gamma(1/\alpha)) + \frac{1}{2}\log(\Gamma(3/\alpha)) - \underbrace{\left[\sqrt{\frac{\Gamma(3/\alpha)}{\Gamma(1/\alpha)}}|u(j)|\right]^{\alpha}}_{O_j} - \log(2) \tag{B.0.2}$$

Taking the derivative of (B.0.2), one obtains

$$\frac{\partial L(u(j);\sigma,\alpha)}{\partial \alpha} = \frac{1}{\alpha} + \frac{3}{2\alpha^2}\Psi(1/\alpha) - \frac{3}{2\alpha^2}\Psi(3/\alpha) - \frac{\partial O_j}{\partial \alpha} \tag{B.0.3}$$

where $\Psi(.)$ is the digamma function and

$$\frac{\partial O_j}{\partial \alpha} = \frac{\partial \theta}{\partial \alpha}|u(j)|^{\alpha} + \theta|u(j)|^{\alpha}\log(|u(j)|) \tag{B.0.4}$$

$$\theta = \left[\sqrt{\frac{\Gamma(3/\alpha)}{\Gamma(1/\alpha)}}\right]^{\alpha}$$

To determine $\frac{\partial \theta}{\partial \alpha}$ requires the functional power rule as follows

$$[F^r]' = F^r[F'\frac{r}{F} + r'\log(F)] \tag{B.0.5}$$

where it is differentiated with respect to r and $(.)'$ denotes the derivative of $(.)$ with respect to r. For the case of $\frac{\partial \theta}{\partial \alpha}$, one have $F = \left[\sqrt{\frac{\Gamma(3/\alpha)}{\Gamma(1/\alpha)}}\right]$ and $r = \alpha$:

$$\begin{aligned}
\frac{\partial \theta}{\partial \alpha} &= F^r\left(F'\frac{\alpha}{\sqrt{\frac{\Gamma(3/\alpha)}{\Gamma(1/\alpha)}}} + \log\left(\sqrt{\frac{\Gamma(3/\alpha)}{\Gamma(1/\alpha)}}\right)\right) \\
&= F^r\left\{\frac{\alpha}{2}\left[\frac{\Gamma(3/\alpha)}{\Gamma(1/\alpha)}\right]^{-1}\left[\frac{-\frac{1}{\alpha^2}(3\Gamma(1/\alpha)\Gamma(3/\alpha)' - \Gamma(3/\alpha)\Gamma(1/\alpha)')}{\Gamma(1/\alpha)^2}\right] + \frac{1}{2}\log\left(\frac{\Gamma(3/\alpha)}{\Gamma(1/\alpha)}\right)\right\} \\
&= \frac{1}{2}\left[\frac{\Gamma(3/\alpha)}{\Gamma(1/\alpha)}\right]^{\alpha/2}\left[\frac{\Psi(1/\alpha) - 3\Psi(3/\alpha)}{\alpha} + \log\left(\frac{\Gamma(3/\alpha)}{\Gamma(1/\alpha)}\right)\right]
\end{aligned} \tag{B.0.6}$$

Thus, substituting (B.0.6) into (B.0.4), it is clear that

$$\frac{\partial O_j}{\partial \alpha} = \underbrace{\theta|u(j)|^{\alpha}}_{O_j}\left\{\frac{1}{2}\left[\frac{\Psi(1/\alpha) - 3\Psi(3/\alpha)}{\alpha} + \log\left(\frac{\Gamma(3/\alpha}{\Gamma(1/\alpha)})\right)\right] + \log|u(j)|\right\} \tag{B.0.7}$$

Since differentiation is a linear process, one can then express the derivative (B.0.1) by using (B.0.3):

$$\begin{aligned}\frac{\partial \mathrm{L}(\mathbf{u};\sigma,\alpha)}{\partial\alpha} &= \sum_{j=1}^{T}\frac{\partial L(u(j);\sigma,\alpha)}{\partial\alpha}\\ &= T\underbrace{\left[\frac{1}{\alpha}+\frac{3}{2\alpha^2}\Psi(1/\alpha)-\frac{3}{2\alpha^2}\Psi(3/\alpha)\right]}_{\xi}-\sum_{j=1}^{T}\frac{\partial O_j}{\partial\alpha}\end{aligned} \tag{B.0.8}$$

Now, the second derivative of (B.0.1) can be obtained. From the equation (B.0.8), the derivative of ξ is first determined .

$$\begin{aligned}\frac{\partial\xi}{\partial\alpha} &= -\frac{1}{\alpha^2}-\frac{3}{2}\left(\frac{-2}{\alpha^3}(\Psi(3/\alpha)-\Psi(1/\alpha))-\frac{1}{\alpha^4}(3\Psi(3/\alpha)'-\Psi(1/\alpha)')\right)\\ &= \frac{1}{\alpha^3}\left(-\alpha+3(\Psi(3/\alpha)-\Psi(1/\alpha))+\frac{3}{2\alpha}(3\Psi(3/\alpha)'-\Psi(1/\alpha)')\right)\end{aligned} \tag{B.0.9}$$

Letting $\gamma_j = \left\{\frac{1}{2}\left[\frac{\Psi(1/\alpha)-3\Psi(3/\alpha)}{\alpha}+\log\left(\frac{\Gamma(3/\alpha)}{\Gamma(1/\alpha)}\right)\right]+\log|u(j)|\right\}$, it can then be differentiated to obtain γ_j' as

$$\begin{aligned}\frac{\partial\gamma_j}{\partial\alpha} &= \frac{1}{2}\left(\frac{\frac{\alpha}{\alpha^2}(9\Psi(3/\alpha)'-\Psi(1/\alpha)')-(\Psi(1/\alpha)-3\Psi(3/\alpha))}{\alpha^2}-\frac{1}{\alpha^2}(3\Psi(3/\alpha)-\Psi(1/\alpha))\right)\\ &= \frac{1}{2\alpha^3}\left(9\Psi(3/\alpha)'-\Psi(1/\alpha)'\right)\end{aligned} \tag{B.0.10}$$

As it is known that $\partial O_j/\partial\alpha = O_j\gamma_j$ from (B.0.7), one can then take its second derivative:

$$\begin{aligned}\frac{\partial^2 O_j}{\partial\alpha^2} &= \frac{\partial O_j}{\partial\alpha}\gamma_j+O_j\frac{\partial\gamma_j}{\partial\alpha}\\ &= O_j(\gamma_j^2+\gamma_j')\end{aligned} \tag{B.0.11}$$

Similarly as in deriving (B.0.8), one can find its second derivative

$$\frac{\partial^2 \mathrm{L}}{\partial \alpha^2} = T\xi' + \sum_{j=1}^{T}[O_j(\gamma_j^2 + \gamma_j')] \tag{B.0.12}$$

where ξ', and γ_j' can be obtained from (B.0.9) and (B.0.10) respectively. It is noteworthy to say that computing recurring terms such as γ_i, and O_i once for each iteration together with the implementation of look up tables of $\Gamma(.), \Psi(.)$ and $\Psi(.)'$ functions renders the algorithm computationally realisable.

Appendix C

DELAY ESTIMATION

C.1 Derivation of the fractional FIR filter

For the sake of completeness, this section has been included to firstly justify the convolution of $s(t)$ with a delayed sinc function in Eq. (6.3.1) and secondly to highlight the equivalence between Lagrange interpolation and discrete FIR filter with transfer function $H(e^{j\omega})$ to approximate the ideal transfer function $H_{\text{id}}(e^{j\omega})$ of a delay line with delay D. In the sequel, the materials provided here are based on the work of Vesa Välimäki [68]. Consider $x(t)$ to be input, and $y(t) = x(t - D)$ as the output of a delay system with delay D. Using the time-shifting property of the Fourier transform, it is clear that

$$y(\omega) = e^{-j\omega D}x(\omega) \tag{C.1.1}$$

Hence, the transfer function of the delay system is

$$H_{\text{id}}(\omega) = \frac{y(\omega)}{x(\omega)} = e^{-j\omega D} \tag{C.1.2}$$

By taking the inverse discrete-time Fourier transform of Eq. (C.1.2) and using the identity $2j\sin\theta = e^{j\theta} - e^{-j\theta}$, the delayed sinc function as the transfer function of the delay system is obtained

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-j\omega D}e^{j\omega k}\partial\omega = \text{sinc}(k - D) \tag{C.1.3}$$

After the justification of using the sinc function in Eq. (6.3.1), the equivalence between the fractional delay FIR filter used to approximate the non-integer delay and the Lagrange interpolation formula is now demonstrated. For this section only, the indexing of vectors starts with zero. The error $\xi(e^{jw})$ between $H(e^{jw})$ and $H_{\mathrm{id}}(e^{jw})$ is defined in the frequency domain as follows:

$$\begin{aligned}\xi(e^{j\omega}) &= H(e^{j\omega}) - H_{\mathrm{id}}(e^{j\omega}) \\ &= \sum_{k=0}^{L} h(k)e^{-j\omega k} - e^{-j\omega D}\end{aligned} \tag{C.1.4}$$

The ith derivative of Eq. (C.1.4) w.r.t. ω is then set to zero and is evaluated at $\omega = 0$, for i=0,..., L.

For i=0,

$$\sum_{k=0}^{L} h(k) - 1 = 0 \quad \Leftrightarrow \quad \sum_{k=0}^{L} h(k) = 1 \tag{C.1.5}$$

For i=1,

$$-\sum_{k=0}^{L} jkh(k) + jD = 0 \quad \Leftrightarrow \quad \sum_{k=0}^{L} kh(k) = D \tag{C.1.6}$$

For i=2,

$$-\sum_{k=0}^{L} k^2h(k) + D^2 = 0 \quad \Leftrightarrow \quad \sum_{k=0}^{L} k^2h(k) = D^2 \tag{C.1.7}$$

All the (L+1) equations obtained from the (L+1) derivatives of Eq. (C.1.4) can be summarised as

$$\sum_{k=0}^{L} k^i h(k) = D^i \quad i=0,1,2,...,L \tag{C.1.8}$$

Matrixwise, Eq. (C.1.8) can be expressed as

$$\mathbf{Dh} = \mathbf{d} \tag{C.1.9}$$

where

$$\mathbf{D} = \begin{bmatrix} 0^0 & 1^0 & 2^0 & \dots & L^0 \\ 0^1 & 1^1 & 2^1 & & L^1 \\ 0^2 & 1^2 & 2^2 & & L^2 \\ \vdots & & & \ddots & \vdots \\ 0^L & 1^L & 2^L & \dots & L^L \end{bmatrix} \tag{C.1.10}$$

$$\mathbf{h} = [h(0)\ h(1)\ h(2)\ \dots\ h(L)]^T \tag{C.1.11}$$

$$\mathbf{d} = [1\ D\ D^2\ \dots\ D^L]^T \tag{C.1.12}$$

Note that the matrix $\mathbf{D}$ is a Vandermonde matrix whose determinant $\det(\mathbf{D})$ can be easily determined by

$$\det(\mathbf{D}) = \prod_{1 \le i < j \le L+1} (d_{2,j} - d_{2,i}) \tag{C.1.13}$$

where $d_{i,j}$ is the element of the matrix $\mathbf{D}$ in its ith row and jth column. The kth coefficient $h(k)$ of the FIR filter in the column vector $\mathbf{h}$ can be computed, by applying Cramer's rule. Denote $\mathbf{D}_l$ as the matrix $\mathbf{D}$ but with its lth column replaced by the column vector $\mathbf{d}$ and l=k+1. According to Cramer's rule, $h(k)$ can be determined as

$$h(k) = \frac{\det(\mathbf{D}_l)}{\det(\mathbf{D})} \tag{C.1.14}$$

Similarly as in Eq. (C.1.13), $\det(\mathbf{D}_l)$ can be computed in the same fashion. Both $\det(\mathbf{D})$ and $\det(\mathbf{D}_l)$ have common terms, i.e. $(d_{2,i} - d_{2,j})$ when $i, j \neq k$. These common terms will cancel out in Eq. (C.1.14) and therefore $h(k)$ can

be simplified as

$$h(k) = \prod_{\substack{e=0 \\ e \neq k}}^{L} \frac{D-e}{k-e} \tag{C.1.15}$$

Eq. (C.1.15) demonstrates that the FIR delay filter is in fact the well-known Lagrange interpolation formula.

C.2 Verification of the modified MMLEDTE

From Eq. (6.3.1), the sinc function is replaced by the Lagrange interpolator FIR coeficients $h(k)$ to yield

$$s(t - \mathrm{D}) = \sum_{k=-\mathrm{M}_a}^{\mathrm{M}_b} h(k)s(t-k) \tag{C.2.1}$$

Without loss of generality, assume a complex spectral envelope of $s(t)$, i.e. $s'(t)$:

$$s'(t) = \frac{s(t)}{e^{j\omega t}} \tag{C.2.2}$$

such that

$$s(t) = s'(t)e^{j\omega t} \tag{C.2.3}$$

Thus

$$s(t-D) = s'(t - \mathrm{D})e^{j\omega(t-\mathrm{D})} \tag{C.2.4}$$

Also

$$s'(t-D) = \sum_{k=-\mathrm{M}_a}^{\mathrm{M}_b} h(k)s'(t-k) \tag{C.2.5}$$

Substituting (C.2.5) into (C.2.4), the following can be obtained

$$s(t-D) = \left(\sum_{k=-M_a}^{M_b} h(k)s'(t-k) \right) e^{j\omega(t-D)} \tag{C.2.6}$$

$$= \left(\sum_{k=-M_a}^{M_b} h(k)s'(t-k) \right) e^{j\omega(t-k+k-D)} \tag{C.2.7}$$

$$= \sum_{k=-M_a}^{M_b} h(k) \underbrace{s'(t-k)e^{j\omega(t-k)}}_{s(t-k)} e^{j\omega(k-D)} \tag{C.2.8}$$

Comparing (C.2.1) and (C.2.8), it is clear that

$$\sum_{k=-M_a}^{M_b} \mathrm{h}(k)\mathrm{s}(t-k) \equiv \sum_{k=-M_a}^{M_b} \mathrm{h}(k)\mathrm{s}(t-k)e^{j\omega(k-D)}$$

This shows that the term $e^{j\omega(k-D)}$ can be dropped from Eq. (6.3.3).

Bibliography

[1] Y. Guo, F. Sattar, and C. Koh, "Blind Separation of Temporomandibular Joint Sound Signals," *ICASSP 99 Proceedings, IEEE International Conference, 1999*, vol. 2, pp. 1069–1072, 1999.

[2] T. K. Moon and W. C. Stirling, *Mathematical Methods and Algorithms for Signal Processing.* Prentice Hall, 1999.

[3] Z. Cheng and T. T. Tjhung, "A New Time Delay Estimator Based on ETDE," *IEEE Trans. on Signal Processing*, vol. 51, pp. 1859–1869, Jul 2003.

[4] L. Parra and C. Spence, "Convolutive blind separation of non-stationary sources," *IEEE Trans. on Speech and Audio Processing*, vol. 8, pp. 320–327, May 2000.

[5] O. Yilmaz and S. Rickard, "Blind Separation of Speech Mixtures via Time-Frequency Masking," *IEEE Trans. on Signal Processing*, vol. 52, pp. 1830–1847, Jul 2004.

[6] T. Gay and C. N. Bertolami, "The acoustical characteristics of the normal temporomandibular joint," *Journal of Dental Research.*, vol. 67, pp. 56–60, 1988.

[7] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis.* John Wiley & Sons, INC, 2001.

[8] S. Roberts and R. Everson, *Independent Component Analysis.* Cambridge University Press, 2001.

[9] S. Sanei and J. Chambers, *EEG Signal Processing.* John Wiley, 2007.

[10] T.-W. Lee, *Independent Component Analysis : Theory and Applications.* Boston : Kluwer Academic Publishers, 1998.

[11] P. D. O'Grady, B. A. Pearlmutter, and S. Rickard, "Survey of Sparse and Non-Sparse Methods in Source Separation," *International Journal of Imaging Systems and Technology, Special Issue: Blind Source Separation and De-convolution in Imaging and Image Processing*, vol. 15, no. 1, pp. 18–33, 2005.

[12] M. G. Jafari, *Novel Sequential Algorithms for Blind Source Separation of Instantaneous Mixtures.* PhD thesis, King's College London, 2002.

[13] B. Rivet, L. Girin, and C. Jutten, "Mixing audiovisual speech processing and blind source separation for the extraction of speech signals from convolutive mixtures," *IEEE Transactions on Audio, Speech and Languages Processing*, vol. 15, pp. 96–108, January 2007.

[14] W. Wang, D. Cosker, Y. Hicks, S. Sanei, and J. Chambers, "Video assisted speech source separation," *ICASSP 05 Proceedings, IEEE International Conference*, vol. 5, pp. v/425–428, Mar 2005.

[15] A. Aubrey, J. Lees, Y. Hicks, and J. Chambers, "Using the Bi-Modality of Speech for Convolutive Frequency Domain Blind Speech Separation," *in Proceedings of IMA International Conference on Mathematics in Signal Processing, Cirencester*, Dec 2006.

[16] J. Herrault, C. Jutten, and B. Ans, "Detection de grandeurs primitives dans un message composite par une architecture de calcul neuromimetique un apprentissage non supervise," *Proceedings of GRETSI, Nice, France*, 1985.

[17] S. Haykin, *Unsupervised Adaptive Filtering: Blind Source Separation, Vol 1.* Wiley-Interscience, 2000.

[18] T. Bowles, *Signal Processing Techniques for the Interpretation of Microarrays.* PhD thesis, Cardiff University, 2006.

[19] M. G. Jafari and J. A. Chambers, "Fetal electrocardiogram extraction by sequential source separation in the wavelet domain," *IEEE Trans. on Biomedical Engineering*, vol. 52, pp. 390 – 400, Mar 2005.

[20] J. C. Principe, S. Cerutti, and S. Amari, "Special topic section on advances in statistical signal processing for medicine," *IEEE Trans. in Biomedical Engineering*, vol. 47, pp. 565–566, May 2000.

[21] A. Yeredor, "Blind Source Separation in the Presence of Doppler Frequency Shifts," *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 5, pp. v/277–280, Mar 2005.

[22] E. Be'ery and A. Yeredor, "Blind Separation of Reflections With Relative Spatial Shifts," *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 5, May 2006.

[23] R. J. M. Gray, S. J. Davies, and A. A. Quayle, *Temporomandibular Disorders: A Clinical Approach, 1st Ed.* British Dental Association, 1995.

[24] H. Koh and P. G. Robinson, "Occlusal adjustment for treating and preventing temporomandibular joint disorders.," *Cochrane Database Systematic Rev.*, vol. 1, 2003.

[25] C. L. Pankhurst, "Controversies in the aetiology of temporomandibular disorders - Part 1: Temporomandibular disorders all in the mind.," *Primary Dental Care*, vol. 3, pp. 1–6, 1997.

[26] S. A. Berman, A. Chaudhary, and J. Appelbaum, "Temporomandibular Disorders," *Emedicine: http://www.emedicine.com/neuro/topic366.htm*, June 2006.

[27] S. E. Widmalm, W. J. Williams, D. Djurdjanovic, and D. C. Mckay, "The frequency range of TMJ sounds," *Oral Rehabilitation*, vol. 30, pp. 335–346, 2003.

[28] C. S. Greene, N. D. Mcneill, C. Clark, and G. T. Truelove, "Temporomandibular disorders and science: a response to the critics," *Journal of Prosthetic Dentistry*, vol. 80, p. 214, 1998.

[29] G. A. Toolson and C. Sadowsky, "An evaluation of the relationship between temporomandibular joint sounds and mandibular movements," *Journal of Craniomandibular Disorders: Facial and Oral Pain*, vol. 5, p. 187, 1991.

[30] Y. Sungyub, J. R. Boston, T. E. Rudy, C. M. Greco, and J. K. Leader, "Time-frequency analysis of temporomandibular joint (TMJ) sounds using radially Gaussian kernels," *IEEE Trans. on Biomedical Engineering*, vol. 48, pp. 936 – 939, Aug 2001.

[31] T. Sano, S. E. Widmalm, P. L. Westesson, K. Takahashi, and H. Yoshida, "Amplitude and frequency spectrum of temporomandibular joint sounds from subjects with and without other signs/symptoms of temporomandibualar disorders," *Oral Rehabilitation*, vol. 26, pp. 145–150, 1999.

[32] S. E. Widmalm, W. J. Williams, and K. P. Yang, "False Localization of TMJ sounds to side is an important source of error in TMD diagnosis," *Oral Rehabilitation*, vol. 26, pp. 213–214, 1999.

[33] S. E. Widmalm, W. J. Williams, B. K. Ang, and D. C. Mckay, "Localization of TMJ sounds to side," *Oral Rehabilitation*, vol. 29, pp. 911–917, 2002.

[34] D. Djurdjanovic, S. E. Widmalm, and W. J. Williams, "Computerized classification of the temporomandibular joint sounds," *IEEE Trans. on Biomedical Engineering*, vol. 47, pp. 977–984, August 2000.

[35] A. Akan and R. B. Ünsal, "Time frequency analysis and classification of temporomandibular joint sounds," *Journal of the Franklin Institute*, vol. 337, pp. 437–451, Jul 2000.

[36] A. Akan, , A. Ergin, M. Yildirim, and E. Öztas, "Analysis of temporomandibular joint sounds in orthodontic," *Computers & Electrical Engineering*, vol. 32, pp. 312–321, Jul 2006.

[37] K. P. Yang, D. Djurdjanovic, K. H. Koh, W. J. Williams, and S. E. Widmalm, "Automatic classification of the temporomandibular joint sounds using scale and time-shift invariant representation of their time-frequency distributions," *Proceedings of the IEEE Signal Processing International Symposium*, pp. 249–252, June 1998.

[38] T. Gay and C. N. Bertolami, "The spectral properties of temporomandibular joint sounds," *Journal of Dental Research.*, vol. 66, pp. 1189–1194, 1987.

[39] L. M. Gallo, R. Airoldi, B. Ernst, and S. Palla, "Power spectral analysis of temporomandibular joint sounds in asymptomatic subjects," *Journal of Dental Research.*, vol. 72, pp. 871–875, 1993.

[40] J. K. Leader, J. R. Boston, T. E. Rudy, and C. Greco, "Quantitative description of temporomandibular joint sounds: defining clicking, popping, egg shell crackling and footsteps on gravel," *Blackwell Science Ltd, Journal of Rehabilitation*, vol. 28, pp. 466–478, 2001.

[41] D. M. Watt, "TMJ sounds," *Journal of Dentistry*, vol. 8, pp. 119–123, 1980.

[42] Y. Li, S. Amari, A. Cichocki, and D. W. C. Ho, "Underdetermined Blind Source Separation based on Sparse Representation," *IEEE Trans. on Signal Processing*, vol. 54, pp. 423–437, Feb 2006.

[43] R. Vigario, J. Sarela, V. Jousmiki, M. Hamalainen, and E. Oja, "Independent component approach to the analysis of EEG and MEGrecordings," *IEEE Trans. on Biomedical Engineering*, vol. 47, pp. 589–593, May 2000.

[44] A. Yeredor, "Blind Source Separation with Pure Delay Mixtures," *Proceedings of The 3rd International Workshop on Independent Component Analysis and Blind Source Separation*, Dec 2001.

[45] L. K. Hansen and M. Dyrholm, "A prediction matrix approach to convolutive ICA," *Proceedings of IEEE Workshop on Neural Networks for Signal Processing XIII Toulouse, France*, pp. 249–258, Sept 2003.

[46] K. Torkkola, "Blind source separation of delayed sources based on information maximization," *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 3509–3512, May 1996.

[47] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing-Learning algorithms and Applications.* John Wiley and Sons Ltd, 2005.

[48] A. J. Bell and T. J. Sejnowski, "An information-maximisation approach to blind separation and blind deconvolution," *Neural Computation*, vol. 7, no. 6, pp. 1129–1159, 1995.

[49] P. Bofill and M. Zibulevsky, "Blind separation of more sources than mixtures using the sparsity of short-time fourier transform," *2nd International Workshop on Independent Component Analysis and Blind Signal Separation*, pp. 87–92, June 2000.

[50] J. Cardoso, "Blind signal separation: statistical principles," *Proceedings of the IEEE, Special issue on blind identification and estimation*, vol. 9, no. 10, pp. 2009–2025, 1998.

[51] A. Hyvrinen and E. Oja, "Independent Component Analysis: Algorithms and Applications (A Tutorial)," *http://www.cis.hut.fi/projects/ica/*, pp. 1–31, Apr 1999.

[52] S. Amari, A. Cichocki, and H. H. Yang, "A new learning algorithm for blind signal separation," *Advances in Neural Information Processing Systems - MIT Press*, vol. 8, pp. 757–763, 1996.

[53] Z. Koldovský, P. Tichavský, and E. Oja, "Efficient variant of algorithm FASTICA for Independent Component Analysis attaining the Cramér-Rao lower bound," *IEEE Trans. on Neural Networks*, vol. 17, pp. 1265–1277, Sept 2006.

[54] G. Wang, X. Xu, and D. Hu, "Self-adaptive FASTICA based on generalized Gaussian model," *Advances in Neural Networks - Springer Berlin / Heidelberg*, vol. 3496, pp. 961–966, Sept 2005.

[55] C. W. Hesse and C. J. James, "The FastICA algorithm with Spatial Constraints," *IEEE Signal Processing Letters*, pp. 1–31, Jul 2005.

[56] P. A. Regalia and E. Kofidis, "Monotonic convergence of fixed-point algorithms for ICA," *IEEE Trans. of Neural Networks*, vol. 14, pp. 943– 949, Jul 2003.

[57] D. P. Mandic and J. A. Chambers, *Recurrent Neural Networks for Prediction*. John Wiley and Sons Ltd, 2001.

[58] A. Belouchrani, K. A. Meraim, J. F. Cardoso, and E. Moulines, "A blind source separation technique based on second order statistics," *IEEE Trans. Signal Processing*, vol. 45, pp. 434–444, 1997.

[59] L. Tong, R. W. Liu, V. C. Soon, and Y. F. Huang, "Indeterminacy and identifiability of blind identification," *IEEE Trans. on Circuits and Systems*, vol. 38, pp. 499 – 509, May 1991.

[60] J.-F. Cardoso and A. Souloumiac, "Blind Beamforming for non-gaussian signals," *Proceedings of the IEE*, vol. 140, pp. 362–370, Dec 1993.

[61] J. C. A. Van-Der-Lubbe, *Information Theory*. Cambridge University Press, 1997.

[62] B. A. Pearlmutter and V. K. Potluru, "Sparse Separation: Principles and Tricks," *Proceedings of SPIE, Independent Component Analyses, Wavelets, and Neural Networks*, vol. 5102, pp. 1–4, April 2003.

[63] P. O'Grady and B. Pearlmutter, "Hard-lost: Modified k-means for oriented lines," *Proceedings of the Irish Signals and Systems Conference*, June-July 2004.

[64] Y. Luo, J. Chambers, S. Lambotharan, and I. Proudler, "Exploitation of Source Non-Stationary in Underdetermined Blind Source Separation with Advanced Clustering Techniques," *IEEE Trans. in signal processing*, vol. 54, pp. 2198– 2212, June 2006.

[65] D. L. Donoho and Y. Tsaig, "Recent Advances in Sparsity-driven Signal Recovery," *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP2005)*, pp. 713–716, 2005.

[66] I. Takigawa, M. Kudo, A. Nakamura, and J. Toyama, "On the Minimum $\ell 1-$Norm Signal Recovery in Underdetermined Source Separation," *Independent Component Analysis and Blind Source Separation: Fifth International Conference*, vol. 5, pp. 193–200, 2004.

[67] P. Georgiev, F. Theis, and A. Cichocki, "Sparse Component Analysis and Blind Source Separation," *IEEE Trans. on Neural Networks*, vol. 16, no. 16, pp. 992–996, 2005.

[68] V. Välimäki, *Discrete-Time Modeling of Acoustic Tubes Using Fractional Delay Filters*. PhD thesis, Helsinki University of Technology, 1995.

[69] W. Liebermeister, "Linear modes of gene expression determined by independent component analysis," *Bioinfomatics*, vol. 18, pp. 51–60, 2002.

[70] T. Aaoyagi, H. Tokutaka, K. Fujimura, and Y. Maniwa, "Application of FastICA to Pulse Wave," *Proceedings of IEEE International Conference on Neural Information Processing*, vol. 2, pp. 769–772, Nov 2002.

[71] B. Kamousi, Z. Liu, and B. He, "Classification of motor imagery tasks for brain-computer interface applications by means of two equivalent dipoles analysis," *IEEE Trans. on Neural Systems and Rehabilitation Engineering*, vol. 13, pp. 166–171, Jul 2005.

[72] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification (2nd Edition)*. Wiley-Interscience, 2000.

[73] Y. Li, A. Cichocki, and S. Amari, "Analysis of Sparse Representation and Blind Source Separation," *Neural Computation*, vol. 16, pp. 1193–1234, 2004.

[74] E. Vincent, C. Fevotte, and R. Gribonval, "Performance measurement in Blind Audio Source Separation," *IEEE Trans. Speech and Audio Processing*, vol. 14, pp. 1462 – 1469, Jul 2006.

[75] K. Waheed and F. Salem, "Blind source recovery using an adaptive generalized Gaussian score function," *45th IEEE International Midwest Symposium on Circuits and Systems, Tulsa, Oklahoma*, vol. 2, pp. 418–421, 2002.

[76] F. C. Meinecke, S. Harmeling, and K. R. Muller, "Robust ICA for Super-Gaussian Sources," *Independent Component Analysis and Blind Source Separation: Fifth International Conference*, vol. 5, pp. 217–224, 2004.

[77] S. Sanei, W. Wang, and J. Chambers, "A coupled HMM for solving the permutation problem in frequency domain BSS," *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. V–565–568, May 2004.

[78] W. Wang, J. Chambers, and S. Sanei, "A Novel Hybrid Approach to the Permutation Problem of Frequency Domain Blind Source Separation," *5th International Conference on Independent Component Analysis and Blind Signal Separation*, pp. 532–539, Sept 2004.

[79] H. Sawada, R. Mukai, S. Araki, and S. Makino, "Robust and precise method for solving permutation problem of frequency domain blind source separation," *IEEE Trans. in Speech Audio Processing*, vol. 12, pp. 530–538, Sept 2004.

[80] S. Amari, S. Douglas, A. Cichocki, and H. Yang, "Multichannel blind deconvolution and equalization using the natural gradient," *Proceedings of Signal Processing Advance in Wireless Communication Workshop*, pp. 101–104, Apr 1997.

[81] Y. Gong and C. F. N. Cowan, "An LMS style variable tap-length algorithm for structure adaptation," *IEEE Trans. on Signal Processing*, vol. 53, pp. 2400– 2407, Jul 2005.

[82] Y. Zhang and J. Chambers, "Convex Combination of Adaptive Filters for Variable Tap-Length LMS Algorithm," *IEEE Signal Processing Letters*, vol. 13, pp. 628–631, Oct 2006.

[83] Y. Zhang, J. Chambers, S. Sanei, P. Kendrick, and T. J. Cox, "A New Variable Tap-Length LMS Algorithm to Model an Exponential Decay Impulse Response," *IEEE Signal Processing Letters*, vol. 14, pp. 263–266, Apr 2007.

[84] Y. Gu, K. Tang, and H. Cui, "LMS algorithm with gradient descent filter length," *IEEE Signal Processing Letters*, vol. 11, pp. 305–307, Mar 2004.

[85] Y. Zhang and J. Chambers, "A variable tap-length natural gradient blind deconvolution/equalization algorithm," *Electronics Letter, accepted for publication*, May 2007.

[86] S. Haykin, *Adaptive Filter Theory (4th Edition)*. Prentice Hall, 2001.

[87] A. H. Sayed, *Fundamentals of Adaptive Filtering*. Wiley IEEE Press, 2003.

[88] B. Farhang-Boroujeny, *Adaptive Filters Theory and Applications.* Wiley, 1999.

[89] K. Torkkola, "Blind deconvolution, information maximization, and recursive filters," *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 4, pp. 3301–3304, Apr 1997.

[90] M. Do and M. Vetterli, "Wavelet-based texture retrieval using Generalized Gaussian Density and Kullback-Leibler distance," *IEEE Trans. on Image Processing*, vol. 11, pp. 146–158, Feb 2002.

[91] K. Song, "A globally convergent and consistent method for estimating the shape parameter of a generalized Gaussian distribution," *IEEE Trans. on Information Theory*, vol. 52, pp. 510–527, Feb 2006.

[92] H. Mathis and S. C. Douglas, "On the existence of universal nonlinearities for blind source separation," *IEEE Trans. on Signal Processing*, vol. 50, pp. 1007–1016, May 2004.

[93] M. Varanasi and B. Aazhang, "Parametric Generalized Gaussian Density estimation," *J. Acoustic Soc. Amer.*, vol. 86, pp. 1404–1414, 1989.

[94] A. Ferrante, A. Lepschy, and U. Viaro, "Convergence analysis of a fixed-point algorithm," *Italian journal of pure and applied mathematics*, vol. 9, pp. 179–186, 2001.

[95] A. M. Djafari, "Statistical Methods for Inverses Problems in Signal and Image Processing," *Laboratoire des Signaux et Systèmes, CNRS- SUPELEC-UPS.*

[96] P. Comon and M. Rajih, "Blind Identification of Under-Determined Mixtures Based on The Characteristic Function," *Elsevier Signal Processing*, vol. 86, pp. 2271–2281, Oct 2006.

[97] J. F. Cardoso, "Super-symmetric decomposition of the fourth-order cumulant tensor. Blind identification of more sources than sensors," *Proceedings of the ICASSP, Toronto*, pp. 3109–3112, 1991.

[98] A. Taleb, "An algorithm for the blind identification of N independent signals with 2 sensors," *IEEE International Symposium on Signal Processing and its Applications*, pp. 5–8, August 2001.

[99] Z. He, S. Xie, and Y. Fu, "Sparseness Measure of Signal," *IEEE International Conference on Neural Networks and Brain, ICNN&B '05*, vol. 3, pp. 1931 – 1936, Oct 2005.

[100] A. M. Kagan, I. Linnik, and C. R. Rao, *Characterization Problems in Mathematical Statistics*. John Wiley & Sons Inc, 1973.

[101] X.-R. Cao and R.-W. Liu, "General Approach to Blind Source Separation," *IEEE Transactions on Signal Processing*, vol. 44, pp. 562–571, Mar 1996.

[102] M. Kendall, A. Stuart, and K. Ord, *Kendall's Advanced Theory of Statistics: Distribution Theory*. Hodder Arnold; 6th Revised Edition, 1994.

[103] P. Scalart and J. V. Filho, "Speech enhancement based on a priori signal to noise estimation," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, pp. 629 – 632, May 1996.

[104] S. Stenfelt and B. Hakansson, "Vibration characteristics of bone conducted sound in vitro," *Acoustical Society of America*, vol. 107, pp. 422–431, Jan 2000.

[105] W. D. O'Brien and Y. Liu, "Evaluation of Acoustic Propagation Paths into the Human Head," *NATO Research and Technology Organisation: New Directions for Improving Audio Effectiveness*, pp. 15–1 – 15–24, Apr 2005.

[106] S. M. Kay, *Fundamentals of Statistical Signal Processing- Estimation Theory*. Prentice Hall Signal Processing Series, 1993.

[107] A. Lee, H. Choi, H. Lee, and J. Pack, "Human Head Size and SAR Characteristics for Handset Exposure," *ETRI Journal*, vol. 24, pp. 176–179, April 2002.

[108] E. H. Chudler, "Brain Facts and Figures," *http://faculty.washington.edu/chudler/facts.html*, Oct 2005.

[109] J. A. Jensen, *Estimation of Blood Velocities using Ultrasound: A Signal Processing Approach.* Cambridge University Press, 1996.

[110] Y. Sungyub, J. R. Boston, T. E. Rudy, C. M. Greco, and J. K. Leader, "Time-frequency analysis of temporomandibular joint (TMJ) sounds," *Proceedings of the IEEE SP international symposium on time-frequency and time-scale analysis*, pp. 289 – 292, Oct 1998.

[111] C. Zheng, S. E. Widmalm, and W. J. Williams, "New time-frequency analyses of EMG and TMJ sound signals," *Proceedings of the Annual International Conference of the IEEE Engineering in Engineering in Medicine and Biology Society*, vol. 2, pp. 741 – 742, 1989.

[112] J. Hawkins, "Learn like a human," *IEEE Spectrum Magazine*, pp. 17–22, April 2007.