UNIVERSITY OF SOUTHAMPTON

Faculty of Engineering and Environment

Institute of Sound and Vibration Research (ISVR)

Benefits Of Multichannel Recording of Auditory Late Response (ALR)

Siavash Mirahmadizoghi

Supervisors: Dr Steven Bell & Prof David Simpson

Thesis for the degree of Doctor of Philosophy

March 2015

University of Southampton

Abstract

Faculty of Engineering and Environment

Institute of Sound and Vibration Research (ISVR)

Doctor of Philosophy

Benefits Of Multichannel Recording of Auditory Late Response (ALR)

By Siavash Mirahmadizoghi

The main purpose of this work is to explore whether and how much multichannel signal processing strategies can be beneficial for improving the detection procedure for auditory late response (ALR) in clinical applications in comparison with single channel recording. To achieve this target, four multichannel noise reduction methods based on independent component analysis (ICA) were proposed for noise reduction for multichannel recording of ALR. The four alternative component selection strategies introduced in this work are: Magnitude Squared Coherence (MSC) [based on coherency of the ICs with an evoking stimulus], the maximum Signal to Noise Ratio (Max-SNR) of ICs over a particular interval, the kurtosis (maximum non-Gaussianity of the ICs), and minimum entropy of the ICs. The proposed methods are applied for the noise reduction of auditory late response (ALR) data captured using 63 channel EEG from 10 normal hearing participants. The performances of the proposed methods for improving signal quality were compared with each other and also with the single channel alternatives. All automated component selection approaches produced high SNR for multichannel ALR data. MSC-ICs produced significantly higher SNR than Max-Kurt-ICs or Min-Entropy-ICs. However the performance of MSC-ICs and Max-Fmp-ICs were not significantly different. Therefore, the MSC-ICs approach was selected for further work. MSC-ICs were used for three different clinical applications: Finding hearing threshold level, exploring the effect of attention and exploring inter- and intra- subject variability. The results for MSC-ICs were compared to the single channel signal processing alternative of weighted averaging. The results confirm that the multichannel signal processing can significantly improve the detection procedure for threshold measurement and for measuring the effects of attention. However, no significant enhancement was found for detecting inter- and intra- subject variability with multichannel processing over single channel alternative.

The MSC-ICs method was also used in an application for removing cardiac artifact from the ALR recordings and the results was compared with an existing artifact rejection platform based on constraint ICA (cICA). The results of this comparison show that the proposed method can significantly improve the quality of cardiac artifact rejection from ALR data.

Finally, the use of MSC-ICs was explored for reducing the required time for recording ALR. Time reduction was investigated in two senses: 1. reducing the number of stimulus repetitions. 2. Optimizing the position and the number of the recording electrodes in multichannel recordings (potentially saving the time required to place many electrodes on the scalp). The results show that using multichannel processing can significantly reduce the number of stimulus repetitions and consequently the time of recording in comparison with the single channel alternative. Minimum required number of stimulus repetition (average

over10 subjects) for having SNR equal to single channel processing at Cz was found to be 74 for un-weighted averaging and 85 for weighted averaging. Moreover, the results of optimal electrode placement procedure confirm that, the ALR can be recorded form the vertex (with the same SNR as when ALR is recorded using 63 channels) by using fewer numbers of electrodes. For the data set of this study (10 normal hearing adults) the same SNR as with 63 channels was achieved by using 40 channels. Placing 40 electrodes (instead of 63) on the scalp decreases the required time for recording ALR considerably, i.e. 53% improved.

List of Tables

List of Figures

Declaration of Authorship

I, Siavash Mirahmadizoghi declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

Benefits of Multichannel Processing for Recording Auditory Late Response (ALR)

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. [Delete as appropriate] None of this work has been published before submission [or] Parts of this work have been published as: [please list references below]:

Signed:.....

Date:.....

Acknowledgements

It is with immense gratitude that I acknowledge the support and help of my supervisors, Prof David Simpson and Dr Steven Bell. Apart from supporting me throughout my PhD with his kindness and knowledge, our meetings were always enlightening the problem domain for me, and proving me with sufficient guidance to investigate the best possible solution.

Department of psychology of University of Southampton for their aid and advice in the experiment arrangement.

This research would not have been possible without the financial support of Rayleigh scholarship. Furthermore, in my daily work I have been blessed with a friendly and cheerful group of fellow colleges in SPCG research group. Also, I would like to thank my friends Dr. Hesam Kouchakpour and Mr Andrew Causon for their supports.

In the end I cannot find words to express my gratitude to my mother (Mehraein Tayebi) and my brother (Arash Mirahmadizoghi) for their continuous support and encouragement.

Definitions and Abbreviations

AER	Auditory Evoked Response
ALR	Auditory Late Response
SNR	Signal to noise Ratio
EEG	Electroencephalograph
BSS	Blind Source Separation
ICA	Independent Component Analysis
PCA	Principal Component Analysis
dB nHL	decibel normal Hearing Level
ASSR	Auditory Steady State Response
AMLR	Auditory Middle Latency Response
ADP	Auditory Processing Disorder
Cz	Channel located at vertex
ECochG	Electrocochleography
ABR	Auditory Brainstem Response
P1	First peak of ALR
N1	First trough of ALR
P2	Second Peak of ALR
MSC	Magnitude Squared Coherence
DFT	Discrete Fourier Transform
Pc	Phase coherence
A/D	Analogue to digital
N_{ave}	Averaged noise
var	Variance
PCs	Principal components
ICs	Independent Components
MSE	Mean Square Error
CLT	According to the central limit theorem
Kurt	kurtosis
Fsp	Fixed single point
Fmp	Fixed multiple points
Cov	Covariance
HL	Hearing level
SPL	Sound Pressure Level
pe SPL	peak equivalent Sound Pressure Level

Max-Kurt-ICs	Component Selection by Maximum Kurtosis of ICs
Min-Entropy-ICs	Component Selection Using Minimum Entropy
Max-Fmp-ICs	Component Selection by Max-Fmp of ICs
MSC-ICs	Component Selection Using Magnitude Squared Coherence of ICs
Weighted Ave	weighted averaging
Un-weighted Ave	un-weighted averaging
Eigen PCs	Eigenvalue ordering of Principal components
ERP	Event Related Potentials
TBI	Traumatic Brain Injury
ADHD	Attention Deficit Hyperactivity Disorder
MSC-Spatial dist.	MSC-Spatial distribution
MSC-PCS	MSC on principal components
OA	Ocular Artifact
cICA	Constraint ICA
R_p	Rejection ratio
$\gamma_{sd}^2(f)$	Coherence

Introduction

Monitoring the functionality of hearing pathway has been the subject of much research. Even a small deficiency in hearing ability can cause big problems and reduce quality of life. A large number of behavioural problems can be traced to hearing issues that individuals faced in their childhood which might have been treatable if detected early (Gelfand 2007). There various strategies by which the functionality of different parts of hearing pathway can be checked. These methods are generally classified into two main groups, subjective tests and objective tests. Subjective tests, such as pure tone audiometry, demand a voluntary response from the subject (Hall 2007). On the other hand objective tests are those that do not need a subject's voluntary contribution. Objective tests are used for cases in which the subject is hard to test, e.g. they are too young to respond to behavioural audiometry, or to provide differential diagnosis of specific dysfunctions in the hearing system (Luck 2005, Hall 2007). The hearing system consists of two main parts, i.e. peripheral (outer ear, middle ear, inner ear) and central (brain and spinal cord) (Gelfand 2007) which need to be evaluated using testing strategies.

Basically in all the objective methods, the response of the brain to a repetitive stimulus, in this thesis an acoustic stimulus, is recorded for further analysis and investigation (Kurtzberg 1989, Luck 2005). An Auditory Evoked Response (AER) is a terminology for a group of objective tests that analyse the response of the brain to an acoustic stimulus and can be recorded from the scalp using electroencephalograph (EEG). These responses can be captured by either single or multichannel recording strategies. Since the responses are generated in the brain but are recorded from the scalp, their amplitude is small in comparison with spontaneous, on-going EEG activity and may also be contaminated with

noise from other sources. For the purposes of this thesis, any activity which is not a response of the brain to the acoustic stimulus is considered as noise. Signal processing methods by which the noise is reduced, and consequently the signal quality is enhanced, are needed. There are various signal processing methods to improve signal quality and the most appropriate choice depends on the recording approach.

The main purpose of this work is to explore whether and how much multichannel signal processing can be beneficial for improving the quality of and AERs and consequently improve their diagnostic analysis. Various algorithms for single and multichannel signal processing are explained in detail in the first chapter. Blind source separation (BSS) is one of the most popular methods which can be used in multichannel signal processing algorithms (Goldstein 1984, Comon 1994, Makeig, Jung et al. 1997, James and Hesse 2005). However, all the BSS based noise reduction algorithms have a common drawback, which is how to select or reject a component. Component selection is usually carried out by visual inspection. Although, some semi or fully automatic methods have been proposed and can be found in the literature, component selection is an issue which is still unresolved in this application.

The main contribution of this study is to develop a multichannel signal enhancement strategy based on automatic component selection in BSS based noise reduction methods and to explore whether using multichannel processing is beneficial compared to using single channel processing.

Generally, analysis of large data with prolonged recording procedures are limitations in objective hearing assessment which limits its use in clinical applications (Dun 2008). Specially, in multichannel recording placing a large number of electrodes, e.g. 32, 64, 128 and etc. is a time consuming procedure and may be unacceptable in some cases, e.g. with elderly participants or small children. Therefore, one aim of this project is to explore if the introduced method can be helpful for reducing the time required for AER recording.

First chapter of this work is a literature review which provides information about the structure and function of the auditory system and alternative methods for assessing functionality of hearing pathways. Conventional methods for AER recording are also explained in the first chapter. This is followed by explanations about single and multichannel recordings of AERs. In addition, two well

established methods for quantifying signal quality are explained. Independent Component Analysis (ICA) and Principal Component Analysis (PCA) are outlined as they are used for solving the source separation problem to provide a method for signal quality enhancement and noise reduction in multichannel recordings in the third chapter.

Second chapter is devoted to the design of an experimental protocol to acquire data from normal hearing subjects, using Electroencephalogram (EEG) and AERs. In this work the focus is on auditory late response (ALR).

In chapter 3, four novel methods of component selection for an ICA based noise reduction methods are proposed. The performances of the proposed methods for multichannel signal enhancement are compared with each other and with PCA on simulated data.

In chapter 4, the multichannel noise reduction methods proposed in the third chapter are applied to recorded data, i.e. captured form normal hearing subjects by using the protocol which was described in chapter two. Additionally, two well established single channel noise reduction methods, un-weighted and weighted averaging, are employed for noise reduction. The performance of the methods (both single and multichannel) for noise reduction in terms of SNR improvement are compared, to assess the significance of any benefit. The method which offers the best performance between multichannel methods in terms of SNR enhancement is selected for further assessment in the fifth chapter.

In the fifth chapter, this multichannel processing method is used in three applications relevant in the clinical context: measuring the hearing threshold, exploring the effect of attention on AER, and inter and intra subject variability in the ALR. Performance of multichannel processing is compared with a conventional single channel method (weighted averaging) to explore if multichannel signal processing provides significant improvement in ALR analysis. In other words, in aforementioned applications, is multichannel processing helpful to find the hearing threshold level of normal hearing individuals closer to zero dB nHL in comparison with when single channel is used; or, is multichannel processing advantageous over single channel processing for detecting the effect of attention or inter/intra subject variability on the ALR.

In the sixth chapter, an alternative application of the proposed approach is investigated, namely its use for cardiac artifact rejection from EEG signals. The results is compared with that for established methods.

In the seventh chapter the impact of the proposed method for reducing the recording time required for ALR assessments is investigated, as is optimal electrode placement. Time reduction can be achieved in two ways: 1. reducing the number of stimulus repetitions. 2. Optimizing the position and the number of the recording electrodes in multichannel recordings, thus saving on time required to place many electrodes on the scalp. Since the aim of this work is to know if multichannel signal processing is beneficial over single channel alternatives, the point of reference is a single channel recording.

Chapter 1

Background

1.1. Introduction

Hearing is one of the five senses by which humans and other vertebrates can communicate with each other and sense the environment. Hearing impairment can cause problems and reduce the quality of life. For example the risk of crossing the road for a person who cannot properly hear the sounds from vehicles increases. According to statistics provided by *Action on hearing loss*, one in six of the UK population are suffering from hearing impairment, ranging from mild to profound. This equates to 10 million people in 2011 and it is predicted to be 14.5 million people by 2031[*]. Moreover, according to Deafness Research UK[**], annually 840 babies are born with hearing problem in both ears in the UK (Mandal 2011). A large group of behavioural problems can be traced to hearing problems, which are treatable if recognised early (Gelfand 2007). The importance of hearing to human wellbeing motivates the current work. This chapter consist of two main parts. The first part covers the anatomical and physiological background in which the structure and function of the hearing system is briefly explained. The second part covers technical background, in which includes various methods by which the hearing system is assessed, are reviewed. In brief, the objective of this chapter is to provide information about the hearing pathway and explain methods of assessing the functionality of hearing pathways.

[*] http://www.actiononhearingloss.org.uk/your-hearing/about-deafness-and-hearing loss/statistics.aspx

[**] Deafness Research UK is the only national medical research charity dedicated to helping people with deafness, tinnitus or other hearing problems.

1.2. Physiological Background

1.2.1. Auditory system

The auditory system can be conceptualised into two main parts, the ear and the central auditory system. The ear consists of three main parts, outer ear, middle ear and inner ear which are shown in Figure 1.1. The outer ear contains the folds of cartilage surrounding the ear canal called the pinna and external auditory canal and ends to the tympanic membrane which is the beginning of the middle ear. The region from the tympanic membrane to the cochlea is called the middle ear. The inner ear can be divided into three main parts: the cochlea, the vestibular system and semi-circular canals (Gelfand 2007).

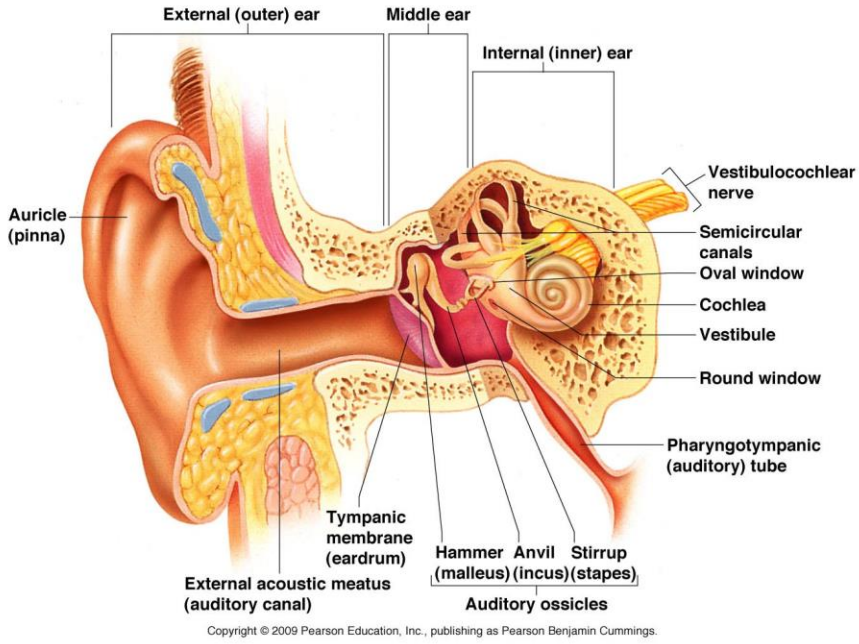

Figure1.1 : Different parts of the ear. Picture is taken from (Pearson 2009).

Sound waves are collected, reflected and attenuated by the pinna and after passing the external auditory canal reach the tympanic membrane. The tympanic membrane is vibrated by the waves and this excites the small bones called *malleus*, *incus* and *stapes*. The tympanic membrane moves the malleus; and this movement is transferred to the incus and from there to the stapes. Stapes movement excites the cochlea via the oval window and generates a travelling wave on the basilar membrane. The basilar membrane is a layer that separates two liquid-filled tubes, the

scala media and the scala tympani, within the coil of the cochlea. A simple schematic uncoiled cutaway cochlea is shown in Figure 1.2. As it is illustrated by the figure, different regions of the basilar membrane can be excited by different frequencies.

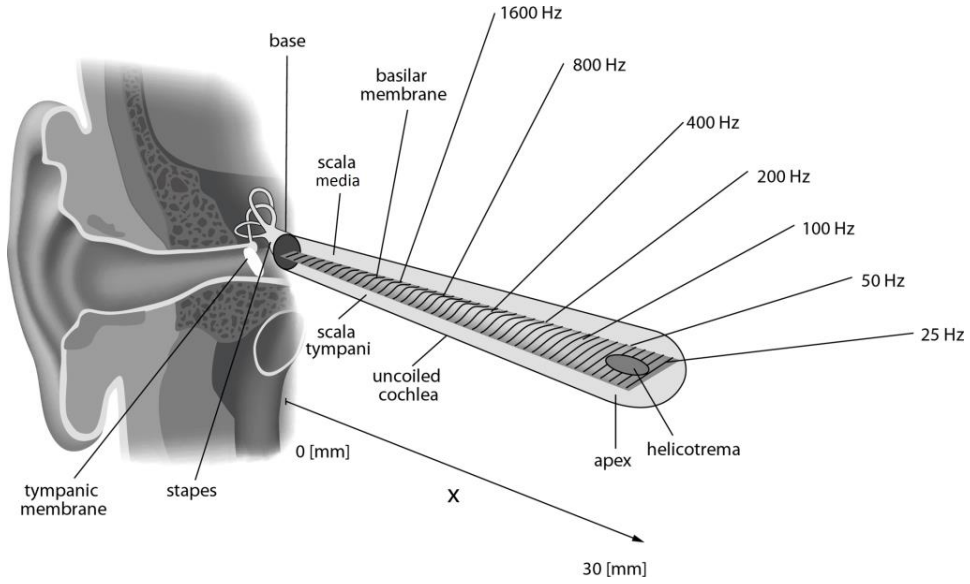

Figure 1.2: Uncoiled cutaway cochlea and frequency dispersion of the basilar membrane. Low frequency sounds excite the apex and high frequency sounds excite the base (Kern, Heid et al. 2008)

The vibration on the basilar membrane moves the hair cells. An electric signal is generated from this movement and transferred to the brainstem and the midbrain auditory system via auditory nerves attached to the cochlea (Hyman 1992). The information eventually reaches the thalamus, and from the thalamus it travels to the cerebral cortex. In the human brain, the primary auditory cortex is located in the temporal lobe (Dun 2008). This sequence is shown in Figure 1.3. Even a small defect in one of the mentioned parts of the auditory system can be the cause of a hearing problem. For example a person who suffers from deafness may have a completely functional outer, middle and inner ear, with impairment further along the hearing pathway.

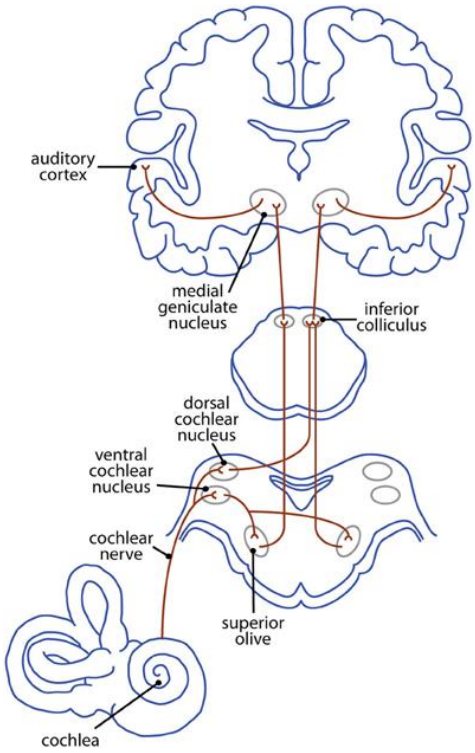

Figure1.3*:* Auditory pathway from the cochlea and through the brain (Butler and Lomber 2013).

1.3. Assessing the hearing system

There exist various strategies by which the functionality of the hearing system can be examined. These methods can be classified into two main groups. The first group is known as behavioural testing, which requires a response from the participant; and second group is made of objective tests which do not need a voluntary (and hence subjective) response from the participant. *Conventional pure tone audiometry* is an example of a behavioural method for checking the hearing system and *cerebral electro-physiological recordings* are examples for the objective assessment of hearing system functionality. Audiometry refers to a group of tests of the function of the hearing system. This includes tests of mechanical sound transmission (middle ear function), neural sound transmission (cochlear function), and speech discrimination ability (central integration). A complete evaluation of a patient's hearing must be done by trained personnel using instruments designed specifically for this purpose (Walker et al. 1990)

Behavioural tests are most commonly used and can help in diagnosing hearing problems effectively when the subject is able to participate in the audiometry test. For

the cases in which the subject is not able to take part in the test, e.g. a subject in coma or too young/old to respond reliably to behavioural audiometry, and also for the cases in which the part of the hearing system that needs to be assessed is not easily accessible. Objective methods are also used in differential diagnosis of hearing impairments (Hall 2007). Auditory evoked responses (AERs) are the best known group of objective tests for checking hearing pathway and are explained in the next subsection (Coats 1978; Goldstein 1984; Hall 2007).

1.3.1. Auditory Evoked Responses (AERs)

Auditory evoked response (AER) is a term used for a group of signals which reflect the function of different parts of the auditory system. They are obtained as a response of the auditory system to the presentation of an acoustic stimulus (Burkard 2006, Hall 2007). The first observation of an AER was carried out by Charls W. Bary and Ernest Glen Waver in 1930 by placing an electrode on a cat's auditory nerve close to the medulla. A ground electrode was located on the body and the action current lead through an amplifier to a telephone receiver. They found that the sound which was presented to the animal's ear was reproduced in the receiver with a high fidelity. They termed this *Auditory Nerve Impulses* (Hall 2007).

There are many microelectrode studies available on the electrical activity of the cochlea and auditory nerve (Davis, Fernandez et al. 1950). However, there was little research on auditory evoked potentials until the development of the averager (Dawson 1954). The electronic averager extracts auditory evoked responses from the much larger EEG recordings, by averaging the responses to many stimuli, given the assumption that the response is invariant between successive stimuli.

In the early to mid 1980s, methods using multichannel time series measurements became popular. Consequently, multichannel analysis tools like independent component analysis opened a new trend in biomedical signal recordings and analysis (James and Hesse 2005).

AERs can be captured by using single or multichannel recording methods and which is preferred, depends on the application (Hall 2007). Based on latencies of the peaks in the recorded waveform, auditory evoked responses can be categorised into three main groups. The first group are the short latency responses which occur within 10ms after stimulus, and include Electrocochleography (ECochG) and the Auditory

Brainstem Response (ABR). It is worth mentioning that the AERs can also be classified based on the region in which the response is generated. For example, ABR is a terminology for the wave which is mainly generated in the brain stem. The Auditory Middle Latency Response (AMLR) arise between 10 ms to 50 ms following stimulation and the Auditory Late Response (ALR) appears after 50 ms (Burkard 2006; Hall 2007). There are other types of AERs, such as Auditory Steady State Response (ASSR) and P300, which are not the subject of this study and hence will not be further discussed. Selecting a suitable AER test depends on the application. Clinically, the AMLR is mostly used for diagnosis and investigating the functionality of auditory pathways and for evaluating auditory sensitivity (Roeser et al. 2000; Hall 2007). Additionally, it is helpful in studying central auditory function in patients with language, speech and learning disabilities and with Auditory Processing Disorder (APD). On the other hand, it has been reported (Thornton et al. 2007) that attention has a greater effect on the ALR than on earlier responses. Moreover, the ALR is a better option for hearing threshold measurement as it has a larger amplitude than short latency responses (Lightfoot 2006), making it easier to detect reliably. As one of the objectives of the project is to explore the effect of attention on the AERs and also measuring the hearing threshold through objective methods, the focus of this work will be on the ALR.

1.3.2. The Auditory Late response (ALR)

The ALR is generated in the cerebral cortex and can be reliably recorded through a non-inverting electrode located in the frontal portion of the scalp of the head. The maximum amplitude of the responses can be recorded by electrodes close to the vertex (Hall 2007). Since the late responses change with changes in the psychological significance of the stimuli they are particularly employed for cognitive neuroscience (Scherg et al. 1989).

A typical waveform of the ALR is shown in Figure 1.4. The first negative voltage component (N1) occurs in the range from 90 ms to 150 ms after stimulus onset. The amplitude of this trough is usually in the range 4 µV to 8 µV. The second major component of the ALR is a positive component (P2) which arises between 160 ms to 200 ms after stimulus onset. This peak amplitude is usually in the range of 3 µV to 6 µV. There also exists an earlier positive component (P1) that is seen around

40 ms to 50 ms after the stimulus onset. P1 occurs less consistently than N1 and P2 and has an amplitude range of 0.5 µV to 2 µV. (Roeser et al. 2000, Hall 2007).

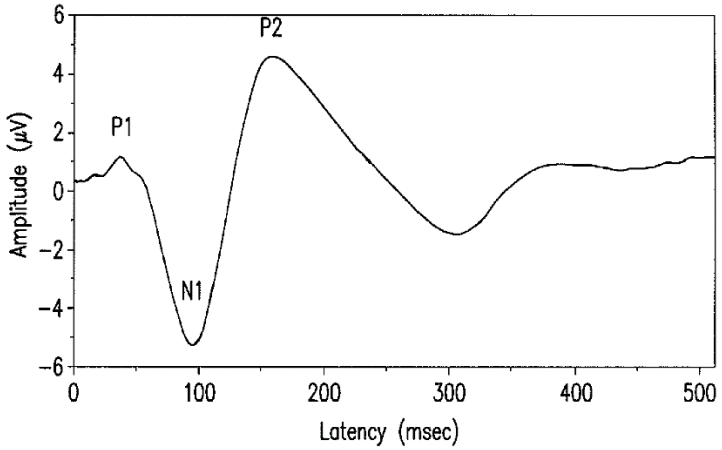

Figure1. 4: The major peaks and latencies of the ALR recorded from the vertex. This picture is adapted from (Durrant, Tlumak et al. 2012).

Stimulus parameters **-**Various types of stimuli can be used for recording the ALR. Depending on the application, speech e.g. full sentence or phonemes such as /da/ and /pa/ (Naatanen and Picton 1987, Kurtzberg 1989, Sussman, Ceponiene et al. 2001, Hall 2007), clicks (Ventura et al. 2009), tone bursts (Roeser et al. 2000) and others can be used as the stimulus. It has been reported (Hall 2007) that the amplitude of the major peaks N1 and P2 are larger for speech stimuli than single frequency tone bursts. However, the latencies of the peaks are earlier when a tone burst is used. Typically a suitable stimulus for ALR recording is a 500 Hz or 1 kHz tone burst with rise/fall times greater than 10 ms and long plateau of about 50 ms (Hall 2007). Intensity around 70 dB nHL is suitable for conventional ALR recording. The amplitudes of N1 and P2 components are larger when the stimulus rate is rather slow. For example 0.7 stimuli per second ($0.7\ s^{-1}$) is a suitable choice for ALR recording (Hall 2007). Typically, 150 to 200 epochs are used for ALR recording (Roeser et al. 2000).

Acquisition parameters-The analysis time should be typically selected long enough to encompass the later components of the ALR. Analysis time for ALR is about 500ms from the stimulus onset (Roeser et al. 2000, Hall 2007). Since the ALR is generated in the brain but recorded from the scalp, the amplitude of the recorded signal is very small. Hence, the data is amplified, typically by factor of 50000-75000,

during the recording. This number can be lower for larger responses (Roeser et al. 2000, Hall 2007). ALR is a low frequency wave and a band pass filter with pass band from 0.1 Hz to 100 Hz is used for ALR recordings. Furthermore, supraaural earphones can be used for ALR recording. For longer AERs recording sessions, insert earphones such as ER-3As are more comfortable. Insert earphones also attenuate the background sound in the test setting (Hall 2007).

Unlike the early AERs such as ECochG, ABR and AMLR which are remarkably consistent from one subject to the next and within subjects for variations of stimulus characteristics, the ALR waveform strongly depends on subject factors such as state of arousal or attention. Therefore, depending on the conditions in which the subject is tested, two recorded ALR waves from one subject could be quite different (Scherg and Voncramon 1985, Hall 2007).

There are important factors influencing AERs recordings which can be categorised into three main groups.

1) Subject factors such as age, gender, state of arousal affects the amplitude and latencies of the major components in AER waveforms. For example, the more the subject pays attention to the test the larger the wave (amplitude) that is recordable (Hall 2007, Thornton et al. 2007). In the middle latency response sex effects the latency of the first peak which is longer in male subjects than in female subjects and also age has statistically significant effects on the amplitude of the first trough (Amenedo and Diaz 1998).

2) Stimulus parameters which should be selected accurately as they have a large effect on the AER waveforms. For example the optimal stimulus rate for an AER is directly related to the latency of the response. Therefore, for recording short latency responses a higher stimulus rate is possible, than for slower responses. In addition, as a general principal, AER latency decreases and amplitude increases as the stimulus intensity increases (Hall 2007).

3) Acquisition factors also have a great impact on AER recordings quality. The quality of the recorded signal, in terms of signal to noise ratio (SNR), can be enhanced by selecting suitable filters, amplifier gain, electrode placement methods etc. For instance in recording slow responses high frequency components of brain activities do not carry any useful information. By filtering the undesired frequency content from the AER signal during the recording, the captured signal will be clearer

(Hall 2007). Clearly the higher the quality of the recordings, the more accurate the diagnosis is.

The quality of the AER recordings can be improved by selecting suitable stimulus parameters and acquisition factors. However, there are limitations that make these methods inefficient. For instance exposing a participant to a high intensity stimulus for a long time can be uncomfortable and potentially harmful. The excessive duration of recordings can lead to habituation (reduction in signal amplitude due to the stimulus repetition) and also loss of signal quality, as patients become fatigued or restless. Signal processing methods which can improve the quality of the AER recordings are thus used for signal quality enhancement and some of these will be explained in the next section.

1.4. Response Detection

Response detection is an essential pre-processing stage to many of data processing algorithm. When comparing the results of applying two or more processing methods on the same data set, a response is needed for a meaningful comparison. Even if one of the methods is significantly better than the other, this significance cannot be confirmed if there is no response present. Multichannel recording of AERs will return an output which may or may not contain a response in some of the channels. A channel may not contain the brain response for many reasons. For instance weak stimuli elicit small responses. Absence of the neural activity in some regions of the brain can indicate disease or injury. Also, technical problems such as poor quality of recording electrodes on the scalp, broken wires etc. can be other reasons for the absence of the response in one or more channels. Therefore, methods that can identify the presence of the response in the recorded data are needed.

-***Magnitude squared coherence (MSC)***: The MSC is a function of frequency that estimates how well a signal, e.g. an AER, linearly relates to another signal, e.g. an EEG to an acoustic stimulus, and is selective by frequency. For a linear system, the MSC is an estimate of the fraction of power in the output that can be explained by the input. MSC denoted as $\gamma_{sd}^2(f)$ in the equation (1.1) is a normalized cross-spectral density function and measures the strength of association and relative linearity between two stationary stochastic processes on a scale from zero to one. When the

processes are independent, the coherence is zero (Barkat 1991, Dun 2008), but zero coherence does not necessarily imply that processes are independent. An estimated coherence value between the EEG and stimulus above the appropriate critical value provides a statistical criterion for concluding that a response is present (Simpson et al. 2007; Santamaria et al. 2007; Ramirez et al. 2008).

$$\gamma_{sd}^2(f) = \frac{|S_{sd}(f)|^2}{S_{ss}(f)S_{dd}(f)} \tag{1.1}$$

where, $S_{ss}(f), S_{dd}(f)$ are the auto spectra of stimulus ***s*** and response ***d***, and $S_{sd}(f)$ is the cross power spectrum of ***s*** and ***d***. Averaged periodogram is used to estimate $S_{ss}(f)$ as follows:

$$S_{ss}(f) = \frac{1}{N}\sum_{i=1}^{N} |S_i(f)|^2 \tag{1.2}$$

In which $S_i(f)$ is the Fourier transform of the i^{th} epoch and *N* is the number of the epochs. Moreover, the cross power spectra can be calculated from (1 .3).

$$S_{sd}(f) = \frac{1}{N}\sum_{i=1}^{N} S_i(f)^*.D_i(f) \tag{1.3}$$

Where, $D_i(f)$ is the Fourier transform of i^{th} epoch of signal ***d***. When the input is an acoustic stimulus and the output is the recorded EEG, if the recorded signal is coherent with that stimulus, it means that the signal can be considered as a response of the brain to the stimulus, unless there is noise that is also coherent with the stimulus. The MSC uses both phase and magnitude information of the signal (by using the complex components of the DFT). To determine whether or not a response is present in a recorded data set, the coherence calculated by (1.1) will be compare to a predefined threshold, provided by statistical analysis. One possible way of estimating this threshold will be presented in chapter 4.

-Phase coherence (Pc): In contrast with the MSC which uses both magnitude and phase information of the signal, the Pc only takes the phase information (θ_i) of each epoch into account. Higher values of phase coherence imply that the phase of the signal was changing less randomly over *N* epochs (Picton et al. 2001).

$$R_\theta(f) = \frac{1}{N}\sqrt{\left(\sum_{i=1}^{N}\cos\left(\theta_i(f)\right)\right)^2 + \left(\sum_{i=1}^{N}\sin(\theta_i(f)\right)^2} \tag{1.4}$$

It has been shown by Dobie and Wilson 1994, 1996 that these methods of response detection have essentially the same performance. Simpson et al. (2001) claimed that the MSC improves the signal detection in comparison with the phase coherence. However this improvement was reported to be small.

There exists other methods for detecting the evoked response such as F-test, t-test and Hotelling T^2 test which can be found in the literature (Dobie and Wilson 1996; Picton 2003). Depending on the data (observed), there exist simpler methods for checking the presence of the response. For instance, for the cases that the response is repetitive, its presence can be confirmed by comparing the average over the first half of the data with the average over the second half of the data. The response is presented in the data if the two aforementioned averages are highly correlated.

1.5. Noise Reduction and Signal Enhancement

Noise reduction strategies are mainly dependent on the protocol under which the AERs are recorded. AERs can be recorded by using either single channel or multichannel methods. Since the signals are generated in the brain but recorded from the scalp, i.e. though hair, skin, the skull and a layer of cerebrospinal fluid and brain matter surrounding the electrical source, their amplitudes are small and consequently, the evoked response recordings are highly contaminated by noise, i.e. the SNR can be as poor as -5dB (Aydin 2005). In this case, *noise* refers to any undesired signals such as muscle activities and also spontaneous brain activities in the measurements. The latter are activities that the brain normally carries out, which are not the response of the brain to the acoustic stimuli. Eye blinks, heart beats and environmental noise (e.g. electrical interference from mains or other electrical devices) are other contaminating factors that can disturb the signal severely. These signals are called artifacts and they have different waveforms from auditory responses and may have different frequency content. They generally have large amplitudes (larger than the response amplitude by several orders of magnitude) and even if their distribution in the recorded EEG is sparse, they can bias analysis methods which are sensitive to outliers, e.g. coherent un-weighted averaging and kurtosis based methods (Delorme et al. 2001). Kurtosis and coherent averaging based signal processing methods are explained in detail in the subsequent sections. It is generally accepted that this noise is not white and more

importantly, it is not coherent with the stimulus (Hall 2007). This is the key feature exploited in removing it from the AER.

Therefore, methods which improve the signal quality and make the signal clearer for diagnostic analysis are needed. Signal quality enhancement can be carried out through either increasing the power of the signal or reducing the power of the noise. Selecting suitable stimulus parameters, i.e. intensity, rate etc., can elicit larger responses (higher signal power) which improves the SNR (Hall 2007). Moreover, acquisition settings such as amplification, filtering, (removing undesired frequency components from the signal), fixed artifact rejection algorithms (removing the blocks which contain artifacts), can be helpful in reducing the noise power and consequently obtaining signals with higher quality. However, apart from acquisition factors, there are various signal processing techniques by which the SNR is improved by reducing the power of noise.

Noise reduction methods are strongly depending on the strategy under which the AERs are recorded. Moreover, noise reduction methods can also be classified into closed form and real-time. In closed form strategies, first the data is recorded and then signal processing methods are performed on the data in a separate session. On the other hand, online noise reduction methods, i.e. it is also called real time noise reduction, are the methods in which the noise is removed from the ALR simultaneous with recording (Rongen et al. 2006, Breuer et al. 2014). In this work, real-time implementation and their challenges are beyond the need and scope of the current work. In the next sections, single and multichannel methods of ALR recording and processing are outlined.

1.5.1 Single Channel Recording and Pre-processing

As illustrated in Figure 1.5, a single channel system for recording AERs requires firstly a stimulus generator able to produce stimulus corresponding to the type of the AER. This device contains a sound card with audio bandwidth. The stimulus is then presented to subject's ear via an insert earphone. The AER as response of the brain to the acoustic stimuli (and noise and background EEG activity) are recorded via electrodes placed on the scalp (for electrode positioning, see next section). Electrodes placed in pairs on the scalp can pick up variations in electrical potential that derive from this underlying cortical or sub cortical activity. One further

electrode is connected to the scalp as the reference ground, and voltages are measured with respect to this point which is usually on the forehead or nose. In Figure 1.5, the difference between the voltage of the electrode placed at the vertex and the reference is amplified by the differential amplifier. The signal is then filtered and amplified and then displayed in the monitoring unit (Larsby et al. 2000). The A/D and the filter unit have not been shown in the figure for simplicity.

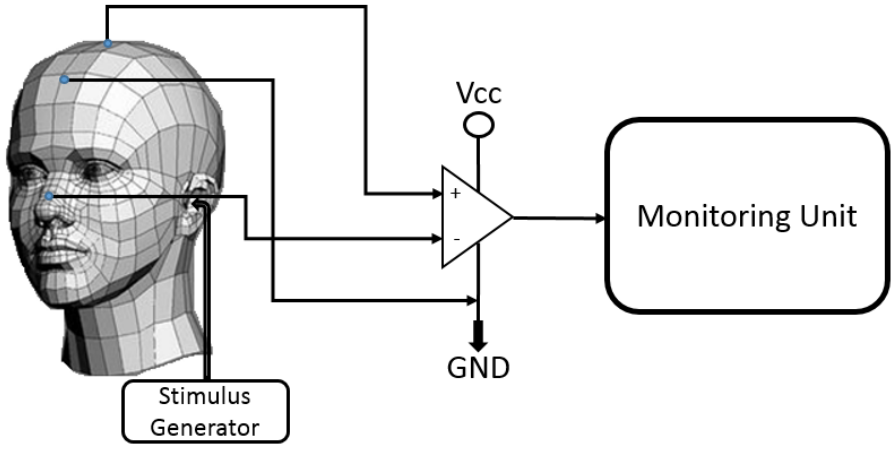

Figure 1.5: Electrode placement for ALR recording. The ground is set on the forehead and the electrode placed on the nose shows the reference. Stimuli are presented to the ear of the subject using an insert earphone. Voltages between vertex and the reference are amplified via the operational amplifier and can be monitored through the monitoring unit.

1.5.1.1. Signal Enhancement for Single Channel Recording

There exist various methods for reducing the noise in single channel recordings. Traditionally, noise is reduced in single channel AER recording by ensemble averaging over a large number of epochs (Aunon et al. 1981; Hoke et al. 1984; Elberling and Wahlgreen 1985; Davila and Srebro 2000). In this method all the epochs have the same contribution to the final average signal. Averaging can also be carried out by assigning smaller weights to the epochs which are highly contaminated by noise (Hoke et al. 1984; Elberling and Wahlgreen 1985; Davila and Mobin 1992; John et al. 2001). Artifact rejection is an alternative method of signal quality enhancement in which the epochs which are highly contaminated by noise are removed from the data in prior to averaging stage. Artifact rejection can be considered an extreme form of weighted averaging with zero weight given to the epochs with large amount of noise. However, it has been reported that weighted

averaging is a better choice over un-weighted averaging or artifact rejection for SNR improvement for evoked response recordings (Dobie and Wilson 1994; John, Dimitrijevic et al. 2001). Other alternative methods such as adaptive filtering (Laguna et al. 1992; Tang and Norcia 1995; Beigi et al. 1998), Wiener filtering (Doyle 1977; Cichocki et al. 2001; Paul et al. 2001), independent component analysis (ICA) based methods (James and Hesse 2005; Davies and James 2007) and wavelet based methods (Heinrich et al. 1999; Effern et al. 2000; Bradley and Wilson 2004) are also used for noise reduction for single channel recordings. Conventionally, *un-weighted averaging* and *weighted averaging* the most frequently used methods for noise reduction for single channel recording of AERs in clinical applications.

Un-Weighted Averaging: the main assumptions for using un-weighted averaging for noise reduction are: 1. the evoked response to the acoustic stimulus is unchanged across the epochs and the noise is stationary and independent between epochs. 2. The signal and noise are uncorrelated. 3. Noise is zero mean with constant variance. Under these assumptions, the average over *M* epochs for the noise (Dawson 1951; Dawson 1954; Hall 2007) , $N_{ave}(n)$, can be written as:

$$N_{ave}(n) = \frac{1}{M}\sum_{i=1}^{M} L_i \tag{1.5}$$

where, *n* is the sample number (within each epoch) and *M* is the number of epochs, *L* is the time series data from each epoch and N_{ave} is averaged noise over *M* epochs. Assuming that the $N_{ave}(n)$ is zero mean, its variance is:

$$var\{N_{ave}(n)\} = E\{N_{ave}(n)^2\} \tag{1.6}$$

where $E\{.\}$ is expectation operator. Equation (1.6), can be re-written as:

$$var\big(N_{ave}(n)\big) = E\left\{\left(\frac{1}{M}\sum_{i=1}^{M} L_i\right)^2\right\} \tag{1.7}$$

Since expectation is a linear operator (1.7) can be written in form of (1.8)

$$\frac{1}{M^2}\sum_{i=1}^{M} \underbrace{E((L_i)^2)}_{var(L_i)} \tag{1.8}$$

Since the noise is stationary $var(L_i) = var\big(L_j\big) = var(N)$. Therefore, the equation (1.8) is equal to:

$$\frac{1}{M^2}\left(M \times var(L_i)\right) = \frac{var(N)}{M} \quad (1.9)$$

Thus, un-weighted averaging over the epochs reduces the noise variance (power) by a factor of $\frac{1}{M}$. As a result, the SNR is improved by factor of *M* (Dawson 1954; Burkard 2006). However, this approach has some restrictions. In this approach, the response is assumed to be unchanged over the epochs and also phase locked to the stimulus; the noise is considered to be stationary with no phase locking to the stimulus. Furthermore, it is assumed that the amount of noise is equal over the epochs. In working with recorded data however, due to the participant's change in position and due to artifacts, the level of noise is higher in some of the epochs, but this is not taken into account in the averaging process. This drawback can be addressed by using weighted averaging (Hoke, Ross et al. 1984).

-***Weighted Averaging***: in this approach, epochs which are highly contaminated by noise have smaller contribution in the averaging than epochs which are less noisy (Davila and Mobin 1992). The epochs are weighted according to their variance prior to summation and then divided by the sum of the weights. The epochs can be weighted using the method proposed by (Ross et al. 1984). By using the formula for weighted averaging we have:

$$S_{ave}(n) = \frac{\sum_{j=1}^{M} w_j L_j(n)}{\sum_{j=1}^{M} w_j} \quad (1.10)$$

where, $S_{ave}(n)$is the weighted average waveform, *M* is the number of the epochs being summed together and w_j is the weighting factor for the j^{th} epoch. w_j is defined as the inverse of the variance of each epoch. It should be noted that other ways of weighting the epochs exist such as percentage-rejection and noise weighting (Dobie and Wilson, 1994) which can be found in the literature (Dimitrijevic et al. 2001). Moreover, weighted averaging can be carried out by weighting the sweeps (blocks of epochs). In this method, each epoch is divided by the variance of the sweep which the epoch belongs to (Silva 2009).

1.5.2 Multichannel Recording of AERs and Pre-processing

The strength of the response in each recording is a function of the distance of the recording electrode to the actual position of the generating source in the brain, i.e. the shorter the distance the stronger the signal. Although, in single channel recording of ALR, highest SNR can be usually obtained from the vertex (Hall 2007), the electrode position which produces the highest SNR is subject dependent (Hall 2007, Breuer, Dammers et al. 2014). Since the number and the position of generator sources are unknown and subject dependent, spreading the electrodes on the scalp increases the chance of one electrode being close to the actual position of the sources. Consequently, the chance of recording a strong response is increased. Therefore, it can be predicted that, by multichannel recording of AERs as an alternative to single channel AER recording, the quality of the AERs may be improved. The number of recording channels depends on the study. Multiple electrodes are usually placed on the head using a scalp cap, as illustrated in Figure 1.6. For electrode placement usually either the 10-20 or equidistant systems are used (Saunders 1990).

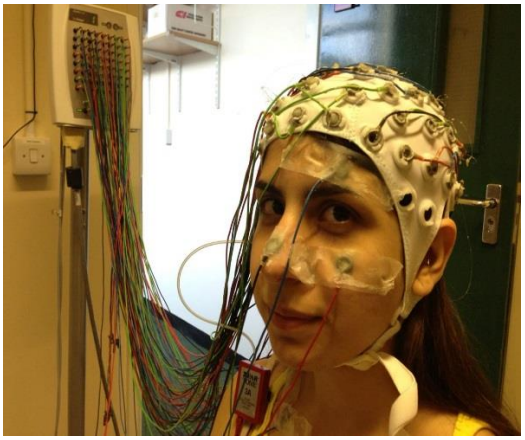

Figure 1.6: Multichannel recording of ALR using 10-20 electrode placement. In this case the reference is set on the nose and the stimulus is presented in the right ear. Signals are amplified by using *SynAmps RT, Neuroscan System*.

As was mentioned before, auditory evoked signals are not usually recognizable in the raw data. To facilitate the diagnostic analysis of the evoked signals, the noise of the recorded signals should be removed. However removing the noise completely is not achievable but it can be reduced considerably.

Multichannel signal processing algorithms have been the topic of much research. Various methods such as adaptive filtering (Yuexian et al. 2011; Acinodotr 2013; Acir 2013), wavelet analysis (Borodina and Aliev 2013; Lee et al. 2013; Yong et al. 2013), spectral analysis (Valenti et al. 2002; Cao et al. 2014; Teng et al. 2014) and blind source separation (BSS) methods (James and Lowe 2003; James and Hesse 2005; Davies and James 2007; Wang et al. 2012; Yuan and Zhang 2014) can be employed in multichannel processing algorithms to reduce the noise from the ALR recordings and improve the signal quality in the terms of increasing the SNR. Among the multichannel processing the BSS methods Principal Component Analysis (PCA) and especially Independent Component Analysis (ICA) have become the most popular (James and Hesse 2005) and are at the core of the current work. The next two sections are therefore devoted to explaining these algorithms.

1.6. Blind Source Separation (BSS) and Noise Reduction

Assuming that the response of the brain to the acoustic stimulus is generated in distinct centres of the brain (sources), it can be said that the recorded signal from each channel, i.e. the output signal of each electrode, would be a combination of signals generated by these sources. Suppose that source i generates signal $\boldsymbol{S}_i$ as illustrated in Figure 1.7. In this case it is assumed that the spreading of sources is instantaneous, an assumption considered reasonable when the spreading is due to physical effects of conducting layers, rather than any spread of activity due to neuronal communication.

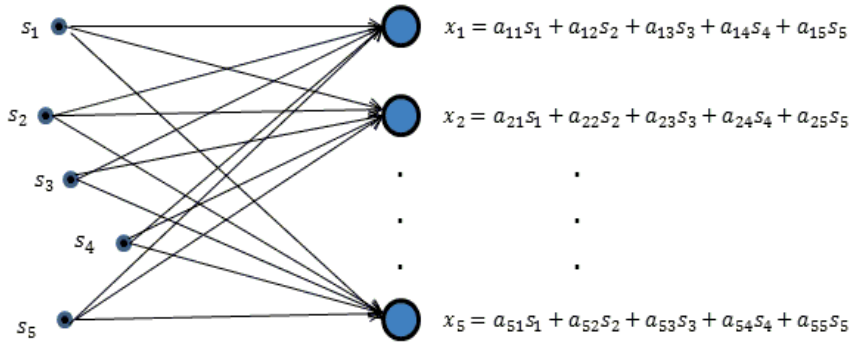

Figure 1.7: Observed output from each channel consists of $\boldsymbol{S}_1 \ldots \boldsymbol{S}_5$ which are generated in the five given sources. a_{ij} are the coefficients which indicate the contribution of a source in an observation

The distance between source i and the j^{th} recording electrode, and the physical properties of the medium determine the coefficients a_{ij}. The longer the distance, the smaller the coefficient (Hyvärinen et al. 2001; Bylund 2001). The fundamental principle of noise reduction with PCA and ICA is that some of the sources represent the desired signal, and some noise. By removing the contribution of the noise sources, the signal quality is then improved. It may also be noted that this process corresponds to data compression, since the signals are now approximated by a smaller number of sources. For the case in which it is assumed that the number of channels is equal to the number of underlying generator sources (*k* sources and *k* channels), the relation between observed signals and source signals can be written in matrix form as follows:

$$\begin{pmatrix} \boldsymbol{x}_1 \\ \vdots \\ \boldsymbol{x}_k \end{pmatrix} = \underbrace{\begin{bmatrix} a_{11} & \cdots & a_{1k} \\ \vdots & \ddots & \vdots \\ a_{k1} & \cdots & a_{kk} \end{bmatrix}}_{\boldsymbol{A}} \begin{pmatrix} \boldsymbol{s}_1 \\ \vdots \\ \boldsymbol{s}_k \end{pmatrix}$$

$$\boldsymbol{X} = \boldsymbol{AS} \tag{1.11}$$

where, matrix $\boldsymbol{X}$ is the observation matrix. Rows of this matrix are recordings from the electrodes. Moreover, $\boldsymbol{A}$ is the mixing matrix and $\boldsymbol{S}$ is the matrix of generator sources, i.e. each row is a source. The aim is to determine how much each of the sources contributes in the observed signals, i.e. finding the coefficients (a_{ij}) in equation (1.11). From these, it is possible to find the inverse of matrix $\boldsymbol{A}$, from which the sources can be calculated:

$$\boldsymbol{S} = \boldsymbol{A^{-1}X} \tag{1.12}$$

Where $\boldsymbol{A^{-1}}$ is the inverse of $\boldsymbol{A}$. Cases in which the generating sources of a recorded signal are to be found without prior information about these sources, are known as blind source separation (BSS) problems. Independent component analysis (ICA) and principal component analysis (PCA) are two methods that give a solution to the BSS problem. Given a set of multivariate measurements, the purpose is to find a smaller set of variables (sources $\boldsymbol{S}$) with less redundancy that provide a good approximation to the recorded signals ($\boldsymbol{X}$). In PCA the redundancy is measured by correlations between variables ($\boldsymbol{S}$), while in ICA stronger concept of independence is

used, and in ICA the reduction of the number of variables is given less emphasis (Hyvärinen et al. 2001). Using only the correlations as in PCA has the advantage that the analysis can be based on second-order statistics only. In connection with ICA, PCA is a useful pre-processing step.

1.6.1. PCA and Noise Reduction

Assume random matrix $\boldsymbol{X}$ with m elements (rows), corresponding to each of m observations. The available samples are $x_i(1), x_i(2), \ldots, x_i(T)$ for i^{th} row. No explicit assumptions on the probability density of the vectors are made in PCA, as long as the first-and second order statistics are known or can be estimated from the data (Hyvärinen et al. 2001).

In the PCA, the matrix $\boldsymbol{X}$ is first centred by subtracting its mean. Next, $\boldsymbol{X}$ is linearly transformed to another vector $\boldsymbol{Y}$ with m rows, $m<k$, so that the redundancy induced by the correlations is removed. This is done by finding a rotated orthogonal coordinate system such that the elements of $\boldsymbol{X}$ in the new coordinates become uncorrelated (Hyvärinen and E. Oja 2000; Hyvärinen et al. 2001). At the same time, the variances of the projections of $\boldsymbol{X}$ on the new coordinate axes are maximized so that the first axis has maximal variance; the second axis corresponds to the maximal variance in a direction orthogonal to the first axis, and so on. For example, assume $\boldsymbol{s_1}$ and $\boldsymbol{s_2}$ are two independent and uniformly distributed vectors. By mixing $\boldsymbol{s_1}$ and $\boldsymbol{s_2}$ using mixing matrix $\boldsymbol{A}=\begin{pmatrix}5 & 2\\ 2 & 5\end{pmatrix}$, two correlated vectors $\boldsymbol{x_1}$ and $\boldsymbol{x_2}$ are obtained. Figure1.8 depicts how PCA rotates the data in the direction in which the variance is maximal. Rotation of the axes can be interpreted as rotation of the data in a new directions (Hyvärinen et al. 2001).

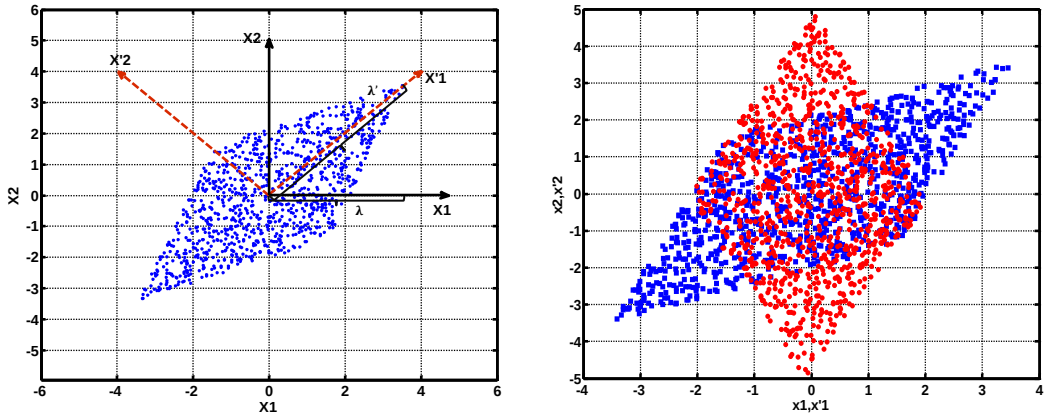

Figure 1.8: Left, $\mathbf{x}_1$ and $\mathbf{x}_2$ are the original axes and $\boldsymbol{\lambda}$ is the standard deviation along the latter. $\mathbf{x}'_1$ and $\mathbf{x}'_2$ are the rotated axes, and $\boldsymbol{\lambda}'$ is the standard deviation in the new axes. Note that $\boldsymbol{\lambda}'$ is larger than $\boldsymbol{\lambda}$. Right, rotation of the axes is equivalent to rotation of data.

Rotation applied by PCA eliminates the correlation information between $\boldsymbol{x_1}$ and $\boldsymbol{x_2}$. The rotated data are called the principal components of the original data. This fact is illustrated in Figure 1.9.

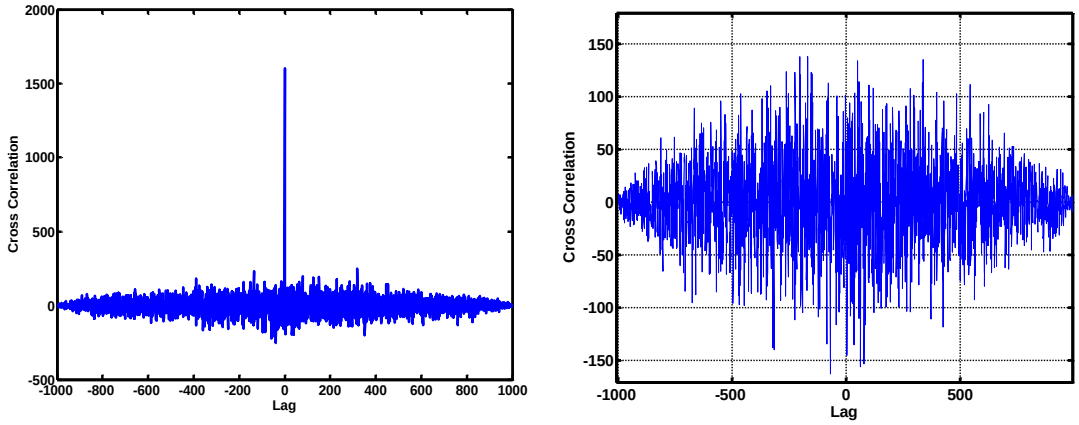

Figure 1.9: Left, cross correlation between $\boldsymbol{x_1}$ and $\boldsymbol{x_2}$ before PCA applied. Right, cross correlation between principal components (data after rotation).

Moreover, correlation coefficient between the two correlated signal $\boldsymbol{x_1}$ and $\boldsymbol{x_2}$ was found to be 0.68 with p-value = 0. This implies a significant correlation between $\boldsymbol{x_1}$ and $\boldsymbol{x_2}$, since the null hypothesis for data being uncorrelated is rejected with p-value <0.05. Whereas, this values found to be 0 with p-value=1 for the principal components which confirms that the two rotated data are uncorrelated. By fitting a line to the scatter diagram of $\boldsymbol{x_1}$ and $\boldsymbol{x_2}$, the linear relationship between correlated data can be illustrated. Whereas, fitting a line to the scatter diagram of two un-correlated

data shows that there is no linear relationship between the two un-correlated data. This fact is illustrated in 1.10.

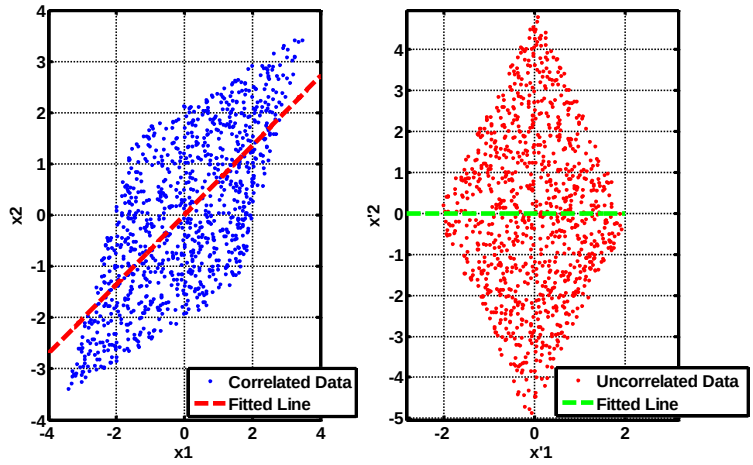

Figure 1.10: left, correlated data and linear curve fitting. Right, un-correlated data and linear curve fitting. Linear relation between data is eliminated by de-correlating the data by applying PCA.

As it can be seen in Figure 1.10, although the linear relation between data is removed by un-correlatedness, there still exists a nonlinear relation between the data. For instance, selecting $\boldsymbol{x_1}$ close to its maximum value restricts the choice for $\boldsymbol{x_2}$. In addition to achieving uncorrelated components, In most practical applications the variances of the components (projections) are very different, and a considerable number of the variances so small that the corresponding components can be discarded altogether. Those components that are left constitute the vector $\boldsymbol{y}$ which is made up of the main components of $\boldsymbol{x}$. PCA is a linear technique, so computing $\boldsymbol{y}$ from $\boldsymbol{x}$ is not computationally heavy, which makes real time processing possible (Hyvärinen and Oja 2000). There are different methods for finding the principal components from a set of measurements. Maximization of the variance and minimization of the mean square error are two famous methods which are explained in the next subsection.

1.6.1.1 PCA by Variance Maximization

The aim is to find a matrix, $\boldsymbol{W}$ ($\boldsymbol{A}^{-1}$ in equation 1.12), such that the product of this matrix and the data matrix provides the principal components of the original matrix so the variance of each component is maximally large.

$$\boldsymbol{S} = \boldsymbol{WX} \tag{1.13}$$

So the first principal component can be calculated as:

$$\boldsymbol{s_1} = \sum_{k=1}^{n} w_{k1}\, x_k = \boldsymbol{w_1}^T \boldsymbol{X} \tag{1.14}$$

Where $\boldsymbol{s_1}$ is the first principal component, x_k with $(k = 1,2,\dots,n)$ are n elements of matrix $\boldsymbol{X}$, and w_{k1} denotes the coefficients (weights) of the n elements of vector $\boldsymbol{x}$. ${\boldsymbol{w}_1}^T$ is the transposed vector of weights for the first principal component. Since the variance of $\boldsymbol{s_1}$ depends on both the norm and orientation of the weight vector $\boldsymbol{w_1}$ and grows without limits as the norm grows, a constraint that the norm of $\boldsymbol{w_1}$ is constant (equal to 1) is imposed. For zero mean data the variance can be calculated using equation (1.15).

$$var(\boldsymbol{y}_1) = E[\boldsymbol{y}_1 \boldsymbol{y}_1^{\boldsymbol{T}}] = E[\boldsymbol{w}_1{}^{\boldsymbol{T}} \boldsymbol{X}\boldsymbol{X}^{\boldsymbol{T}} \boldsymbol{w}_1] = \boldsymbol{w}_1{}^T E[\boldsymbol{X}\boldsymbol{X}^{\boldsymbol{T}}]\boldsymbol{w}_1 = \boldsymbol{w}_1{}^{\boldsymbol{T}} \boldsymbol{C_x} \boldsymbol{w}_1 \tag{1.15}$$

If $\boldsymbol{s}_1$ is one of the principal components (PCs), it should satisfy the maximum variance condition. Therefore, $\boldsymbol{w}_1$ should be found maximizing the equation (1.15) with the constraint $\|\boldsymbol{w}_1\|=1$. In equation (1.15) $\boldsymbol{C_x}$ is the $n \times n$ covariance matrix for zero mean data $\boldsymbol{X}$ (Original data).

According to basic linear algebra the $\boldsymbol{W}$ which maximizes equation (1.15), is given in terms of the unit-length eigenvectors $\boldsymbol{e}_1 \dots \boldsymbol{e}_n$ of covariance matrix $\boldsymbol{C}_x$ (see Appendix I). Therefore, $\boldsymbol{w}_k = \boldsymbol{e}_k$ and the k^{th} principal component is given by $\boldsymbol{y}_k = \boldsymbol{e}_k^T \boldsymbol{x}_k$ (Hyvärinen and E. Oja 2000; Hyvärinen et al. 2001). In brief, it can be stated that the columns of $\boldsymbol{W}$ are the eigenvectors of the covariance matrix of $\boldsymbol{X}$.

1.6.1.2. PCA by Minimum Mean-Square Error Compression

In this method, the target is to find a set of n orthonormal basis vectors to linearly project $\boldsymbol{X}$ into $\widehat{\boldsymbol{X}}$, such that the mean square error between $\boldsymbol{X}$ and its projection on the subspace $\widehat{\boldsymbol{X}}$ is minimal. Denoting the m vectors $\boldsymbol{w}_1 \dots \boldsymbol{w}_m$, projection of $\boldsymbol{X}$ would be:

$$MSE = E[\left\|\boldsymbol{X} - \widehat{\boldsymbol{X}}\right\|^2] = E[||\boldsymbol{X} - \sum_{i=1}^{m} (\boldsymbol{w}_i^T \boldsymbol{X})\boldsymbol{w}_i\,||^2] \tag{1.16}$$

It can be shown (see Hyvärinen and Oja 2000) that (1. 16) is equal to:

$$trace(\boldsymbol{C}_x) - \sum_{i=1}^{m} \boldsymbol{w}_i^T \boldsymbol{C}_x \boldsymbol{w}_i \tag{1.17}$$

By taking the orthonormality condition into account, the minimum of equation (1.17) is given by any orthonormal basis of the PCA, i.e. the *m* first eigenvectors $\boldsymbol{e}_1 \dots \boldsymbol{e}_m$. However, the criterion does not specify the basis of this subspace at all, i.e. any *m* dimensional vector can be considered as a basis of the subspace. This can be considered as a disadvantage of this method. It is worthwhile to say that there are some methods by which a certain basis in the PCA subspace is to be preferred over others (Hyvärinen and Oja 2000; Hyvärinen et al. 2001). More information about selecting subspace in PCA is provided in the third chapter.

1.6.1.3. How to Reduce Noise Using PCA

According to the results in 1.6.1.2, the principal components of observation matrix, $\boldsymbol{X}$, are the eigenvectors of its covariance matrix, $\boldsymbol{C_x}$. Since $\boldsymbol{w}_k = \boldsymbol{e}_k$ from equation (1.15) each principal component has the variance which follows that

$$E[\boldsymbol{y}_m \boldsymbol{y}_m{}^T] = E[\boldsymbol{e}_m^T \boldsymbol{X}\boldsymbol{X}^T \boldsymbol{e}_m] = \boldsymbol{e}_m^T \boldsymbol{C}_x \boldsymbol{e}_m = d_m \qquad (1.18)$$

where $\boldsymbol{y_m}$ is m^{th} principal component $\boldsymbol{e}_m$ is the m^{th} eigenvector and d_m is the variance of m^{th} principal component. From (1.18) it can be seen that the variances of the principal components are thus directly given by the eigenvalues of $\boldsymbol{C}_x$. Since the principal components have zero means, a small eigenvalue (a small variance) d_m indicates that the value of the corresponding principal component $\boldsymbol{y}_m$ is mostly close to zero. The eigenvalue sequence of a covariance matrix for real-world measurements data usually sharply decreases, and it is possible to set a limit below which the eigenvalues, hence principal components, are insignificantly small. This limit determines how many principal components are used and the remaining components can be considered as noise. For the signal-noise model given by equation (1.19):

$$\boldsymbol{X} = \sum_{i=1}^{n} a_i \, \boldsymbol{s}_i + \boldsymbol{N} \qquad (1.19)$$

where $\boldsymbol{a_i}$ are some fixed coefficients, $\boldsymbol{S}_i$ are uncorrelated vectors which have zero mean and $\boldsymbol{N}$ is the noise for which $E[\boldsymbol{N}\boldsymbol{N}^T] = \sigma^2 \boldsymbol{I}$. By assuming that noise is not correlated to the signal, the covariance of *X* can be calculated from:

$$C_x = E[\boldsymbol{X}\boldsymbol{X}^T] = E[\sum_{i=1}^{n} a_i \boldsymbol{s}_i \boldsymbol{s}_i^T a_i^T + a_i \boldsymbol{s}_i \boldsymbol{N}^T + \boldsymbol{N}\boldsymbol{s}_i^T a_i^T + \boldsymbol{N}\boldsymbol{N}^T] =$$

$$\sum_{i=1}^{n} a_i \underbrace{E[\boldsymbol{s}_i \boldsymbol{s}_i^T]}_{\boldsymbol{I}} a_i^T + \underbrace{a_i E[\boldsymbol{s}_i \boldsymbol{N}^T]}_{0} + \underbrace{E[\boldsymbol{N}\boldsymbol{s}_i^T] a_i^T}_{0} + \underbrace{E[\boldsymbol{N}\boldsymbol{N}^T]}_{\sigma^2 I} =$$

$$\sum_{i=1}^{n} a_i a_i^T + \sigma^2 \boldsymbol{I} \tag{1.20}$$

When the eigenvalues of $\boldsymbol{C}_x$ are computed, the first m form a decreasing sequence and the rest are small constants, around σ^2 noise variance; These m eigenvalues span an m dimensional subspace, $\boldsymbol{Y}$, which contains principal components of $\boldsymbol{X}$. The remaining part can be considered as noise and be discarded. If the original signals are reconstructed using only the main components, the noise components do not contribute in signal reconstruction, and thus we expect to have cleaner signals (Hyvärinen and Oja 2000; Hyvärinen et al. 2001).

1.6.2. Whitening

A zero mean random vector $\boldsymbol{z}=(z_1,z_2,\ldots,z_n)$ is called white, if its elements are uncorrelated and it is further assumed to have unit variance (Hyvärinen and Oja 2000; Hyvärinen et al. 2001). Basically whitening is the same as PCA with an extra step, i.e. make the data unit variance. In terms of covariance matrix whitened data means $E[\boldsymbol{z}\boldsymbol{z}^T] = \boldsymbol{I}$ with I the unit matrix. This is illustrated in Figure 1.11.

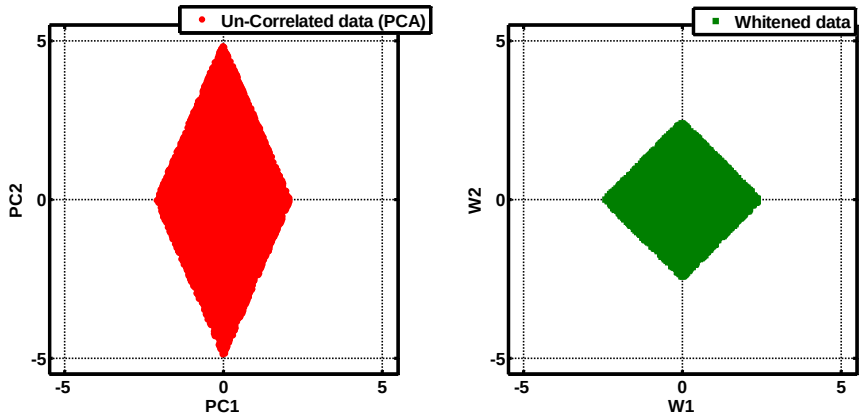

Figure 1.11: Left, un-correlated data by PCA. Right, whitened data. Whitened data are uncorrelated and have unit variance.

In brief, whitening can be interpreted as linearly transform an observed matrix ($\boldsymbol{X}$) into another matrix such that the rows of the new matrix are uncorrelated and also have unit variance. Matrix equation for whitening is shown by equation (1.21).

$$\boldsymbol{Z=VX} \tag{1.21}$$

Where, $\boldsymbol{Z}$ is the whitened data, $\boldsymbol{V}$ is whitening matrix and $\boldsymbol{X}$ is the observation matrix (correlated data). Although, data is uncorrelated and centred by whitening, whitening is not enough for finding for independent components. For any orthogonal matrix $\boldsymbol{U}$, the product of $\boldsymbol{UZ}$ is also a white matrix. Covariance matrix $\boldsymbol{UZ}$ can be written as:

$$E\{\boldsymbol{UZZ^TU^T}\} = \boldsymbol{U}E\{\boldsymbol{ZZ^T}\}\boldsymbol{U^T} = \boldsymbol{UIU^T} = \boldsymbol{I} \tag{1.22}$$

Therefore, it is not possible to tell if the independent components are given by *Z* or *UZ* and further processing is needed to find the ICs. However, whitening is useful as a pre-processing stage for ICA (Hyvärinen and Oja 2000, Hyvärinen et al. 2001).

1.6.3. Independent Component Analysis and Noise Reduction

Using independence instead of un-correlatedness provides a more powerful solution for the source separation problem (Hyvärinen et al. 2001). When two signals are uncorrelated, it means that there is no linear relation between these two signals; but there, possibly, exists a non-linear function which can explain the relationship between them. When two signals are independent, there is neither a linear or non-linear relation between two signals. In the other words, for two independent signals A and B, signal A does not give any information about signal B and vice versa (Hyvärinen et al. 2001). In mathematical terms, two signals are uncorrelated if the covariance between them equals to zero:

$$\boldsymbol{C}_{xy} = E[(\boldsymbol{x} - \mu_x)\big(\boldsymbol{y} - \mu_y\big)^T] = 0 \tag{1.23}$$

In which μ_x is the mean of *x* and μ_y is the mean of *y* .

Equivalently, two signals are uncorrelated if their cross correlation is:

$$\boldsymbol{R}_{xy} = E[(\boldsymbol{x})(\boldsymbol{y})^T] = E[\boldsymbol{x}]E[\boldsymbol{y}^T] \tag{1.24}$$

where, $\boldsymbol{R}_{xy}$ denotes the cross correlation between variables $\boldsymbol{x}$ and $\boldsymbol{y}$**.** Obviously for zero mean variables the right side of equation (1.24) will be 0. Equation (1.24) confirms the expression which says there is no linear relation between two variables. Moreover, statistical independence is defined as:

$$p_{x,y}(\boldsymbol{x},\boldsymbol{y}) = p_x(\boldsymbol{x})p_y(\boldsymbol{y}) \tag{1.25}$$

Where $p_{x,y}(\boldsymbol{x},\boldsymbol{y})$ is joint probability density function of variables $\boldsymbol{x}$ and $\boldsymbol{y}$, $p_x(x)$ and $p_y(\boldsymbol{y})$ are the probability density functions of *x* and *y* respectively. Equation (1.25) can be written in a more general form:

$$E[h(x)g(y)] = E[h(x)]E[g(y)] \tag{1.26}$$

In which *h* and *g* are any absolutely integrable functions of *x* and *y* respectively (Soong 2004). This fact is illustrated through an example in Figure 1.12. In this example three independent and uniformly distributed signals $\boldsymbol{s_1}$, $\boldsymbol{s_2}$ and $\boldsymbol{s_3}$ were mixed using mixing matrix $\boldsymbol{A}$ given by equation (1.27) and return observation matrix $\boldsymbol{X}$ with $\boldsymbol{x_1}$, $\boldsymbol{x_2}$ and $\boldsymbol{x_3}$ as rows of the matrix.

$$\boldsymbol{A}=\begin{pmatrix} 5 & 2 & 0.5 \\ 2 & 5 & 0.03 \\ 0.5 & 0.03 & 7 \end{pmatrix} \tag{1.27}$$

As was mentioned in the previous section, for both cases un-correlated and whitened data, selecting the data close to the maximum value of one of the variables limits the selection for the other variables. However, as it can be seen from the figure, for independent components, having information about one of the independent components, e.g. $\boldsymbol{s_1}=\boldsymbol{IC_1}$, does not provide any information about the other two components ($\boldsymbol{s_2}=\boldsymbol{IC_2}$ and $\boldsymbol{s_3}=\boldsymbol{IC_3}$).

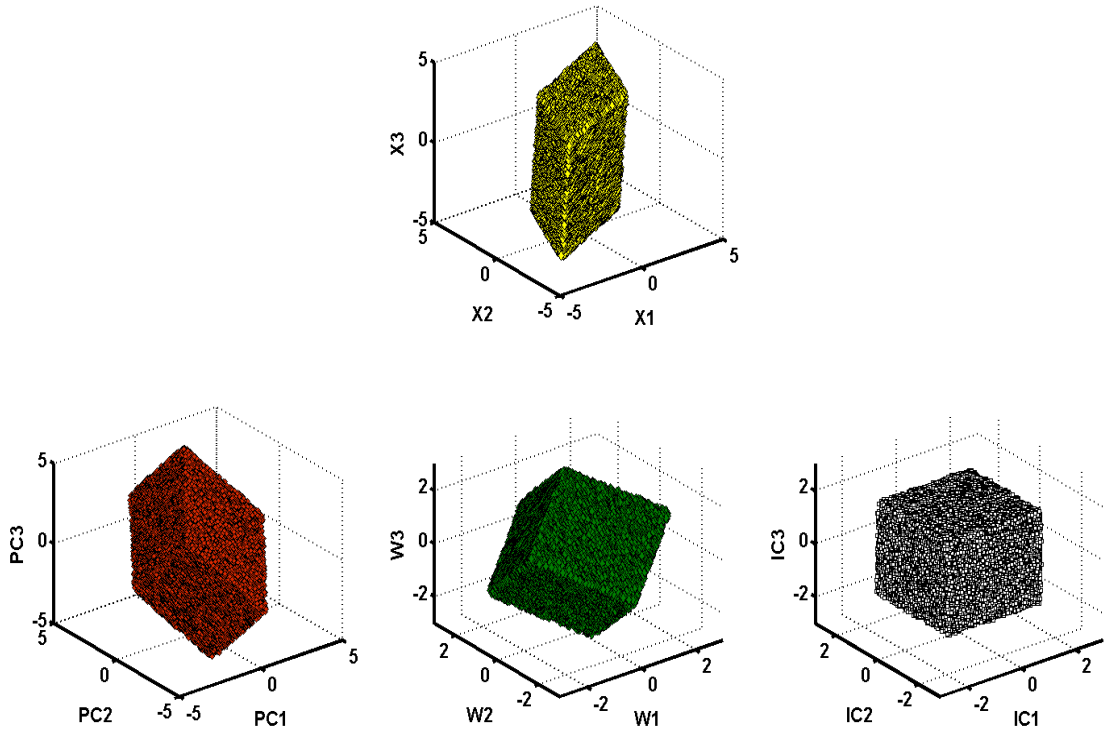

Figure 1.12: Top, three correlated variables ($\boldsymbol{x_1}$, $\boldsymbol{x_2}$ and $\boldsymbol{x_3}$) made from $\boldsymbol{s_1}$, $\boldsymbol{s_2}$ and $\boldsymbol{s_3}$ by using mixing matrix $\boldsymbol{A}$. Down-left, uncorrelated data (PCA). Down-middle, whitened data. Down-Right, independent data (ICA).

Figure 1.12 and equation (1.26) reveal that statistical independence is a much stronger property than un-correlatedness and provides a stronger tool for finding the main components of a signal and source separation (Soong 2004). It is worth mentioning that, for variables with Gaussian distribution, un-correlatedness is equivalent to independence. However, uncorrelated non-Gaussian variables are not necessarily independent from each other (Hyvärinen et al. 2001). From the equation (1.11), for the case of ALR recording an observation (i^{th} row of matrix $\boldsymbol{X}$) can be can be written in the form:

$$\boldsymbol{x}_i = \sum_{j=1}^{n} a_{ij}\boldsymbol{s}_i \tag{1.28}$$

where, $\boldsymbol{x}_i$ is i^{th} observation and a_{ij} are coefficients and S_i are sources. If we find (estimate) a_{ij} and $\boldsymbol{s}_i$ $i = 1, \dots, n$ ($\boldsymbol{s}_i$ are rows of matrix $\boldsymbol{S}$), in a way that $\boldsymbol{s}_i$ are mutually independent, it can be said that we have separated the observed signal into generator components. In other words, we are looking for matrix $\boldsymbol{W}$ in equation (1.13) in a way that the product of $\boldsymbol{W}$ and $\boldsymbol{X}$ provides independent components (ICs) of $\boldsymbol{X}$.

ICA can be interpreted as a further rotation to whitened data which makes the data independent.

1.6.3.1. Ambiguities of ICA

One of the ambiguities in ICA is that the variances (or powers), of the components are not clearly defined, which is in contrast to PCA. Both matrices $\boldsymbol{A}$ and $\boldsymbol{S}$ are unknown, and any scalar multiplier in one of the sources $\boldsymbol{s}_i$ could always be cancelled by dividing the corresponding column $a_i \, of \, \boldsymbol{A}$ by the same scalar. According to (Hyvärinen et al. 2001) the most conventional way of addressing this ambiguity is to normalise the sources to have unit variance. The columns of the mixing matrix $\boldsymbol{A}$ indicate the power of each component across the measurement space. ICs can be normalized by dividing the columns of mixing matrix by its norm. This fixes the problem of variances but still leaves an ambiguity of the sign, i.e. multiplying an IC by -1 does not affect the model.

Another issue of using ICA is determining the order of the independent components. Since $\boldsymbol{S}$ and $\boldsymbol{A}$ are unknown, correct determination of the number of sources is a problem in using ICA. The order of the ICs can freely change and any of the independent components can be called the first one (Hyvärinen et al. 2001). It is possible to calculate quantities such as *rms* power for each source from the columns of the mixing matrix which will then allow ranking of the sources according to their *rms* power. Adapting the formula from (James and Hesse 2005) the *rms* power for each source can be calculated from:

$$p_j = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(a_i^j)^2} \tag{1.29}$$

where, p is the *rms* power for the j^{th} source, m is the number of coefficient and a_i is the corresponding column to the source in the mixing matrix.

This issue is more problematic especially when the number of channels exceeds the number of sources, as is the case in many biomedical applications such as high density EEG/MEG. Knowing how many sources to estimate can have major impact on the quality and accuracy of the ICA solution. Using PCA is one of the common approaches to estimate the number of underlying sources in the data. In the simplest case, the number of sources is equal to the number of dominant eigenvalues. Number

of dominant eigenvalues is defined to be the number of eigenvalues which account for high proportion of the total observed variance (e.g. 95% or 99%). Alternatively, it can be the number of eigenvalues whose individual contribution to the total variance is greater than some minimum amount (e.g. 5% or 1%) (James and Hesse 2005). This approach has many drawbacks which makes is inefficient for many applications. There is no a priori reason to suppose that the sources of interest are contained in the signal subspace spanned by the dominant principal components (e.g. if there is a lot of noise, or the sources of interest are relatively weak compared with other artifactual sources). Moreover, the variance proportion threshold for considering a source as a response is quite arbitrary. In addition, in the presence of noise (especially Gaussian), the number of sources that will be estimated using a cumulative variance criterion varies with the number of sensors, since the proportion of total variance due to the sensor noise increases with the number of sensors. This results in overestimation of the number of sources, which in turn can lead to over-fitting (James and Hesse 2005).There are a number of source enumeration methods which use statistical or information theoretic measures to determine the number of underlying sources which are explained in the third chapter.

1.6.3.2. Restrictions of ICA

There are some basic assumptions in using ICA which facilitate the implementation in different applications. It is worth noting that without making these assumptions, the BSS problem remains intractable. In addition to linear mixing, i.e. the observations are a linear mixture of the underlying which was mentioned before sources, for simplicity it is assumed that the mixing matrix is square. However, there exist methods for finding ICs for the cases in which the number of ICs is less than the number of observations, i.e. mixing matrix is not square, (Hyvärinen and Oja 2000) but such methods are not the subjects of this study.

1.6.3.3. ICA Approaches

Blind separation of statistically independent sources are commonly based on techniques involving higher-order statistics (HOS) and several such implementations can be found in the literature (Bell and Sejnowski 1995; Hyvarinen and Oja 1997; Makeig et al. 1997; Lee et al. 1999; James and Hesse 2005). In this implementation

the observed signal (recorded via sensor) is assumed to be a linear mixture of statistically independent sources. With the assumption that independent components are not normally distributed (this is a valid assumption in many applications such as biomedical data) one practical approaches for estimating the independent components, is to make the estimates as non-Gaussian as possible (Hyvärinen et al. 2001; James and Hesse 2005). This approach is explained in detail in the next section. There also exist alternative strategies for ICA implementation that are valid for separating ICs regardless of their distributions based on their temporal structure (James and Hesse 2005). In this approach the assumption of the independence of the sources has a very important and useful consequence. The source waveforms have no spatial temporal or spatial time–frequency correlations (Ziehe and Muller 1998; James and Hesse 2005). This approach is not the focus of the current work. In the current work, an implementation based on non-Gaussianity of the components is selected for BSS. Two of the most famous ICA implementations by finding maximum non-Gaussianity are explained in the next sections.

1.6.3.4. ICA by Non-Gaussianity

According to the central limit theorem (CLT), the sum of sufficiently large number of independent random variables, even with different distributions which have finite mean and variance, will be approximately normally distributed (Hyvärinen et al. 2001; Soong 2004). Obviously, the central limit theorem is also applies for variables with a Gaussian distribution. Therefore, $\boldsymbol{x}_i$ in (1.17) which is a mixture of non-Gaussian signals $\boldsymbol{s}_i$ has a distribution closer to Gaussian than the source signals. Since we assumed the ICs have non-Gaussian distribution in (1.17) we can select the components with maximum non-Gaussianity as ICs. Central limit theorem is not useful if ICs have a Gaussian distribution. For a signal such as $y = a_1\boldsymbol{s}_1 + a_2\boldsymbol{s}_2$, assuming $a_1^2 + a_2^2 = 1$ and considering the properties of kurtosis it can be shown that absolute value of *Kurt(**y**)* will reach its maximum, when $\boldsymbol{y} = \pm\boldsymbol{s}_i$ (for proof see A.Hyvärinen and E. Oja 2000).

1.6.3.5. Non-Gaussianity by Kurtosis

Kurtosis is the name given to the fourth-order cumulant of a random variable (Ding and Nguyen 2000, A. Hyvärinen, J. Karhunen et al. 2001). Cumulant of a

random variable can be calculated from the second characteristic function of the variable as following:

$$K_i = (-j)^i \frac{d^i(\emptyset(\omega))}{d\omega^k}\Big|_{\omega=0} \tag{1.30}$$

K_i is the i^{th} cumulant, $\emptyset(\omega)$ is the second characteristic function of the random variable ω which is calculated as:

$$\emptyset(\omega) = \ln \varphi(\omega) \tag{1.31}$$

$\varphi(\omega)$ is the first characteristic function and can be calculated from (1.31).

$$\varphi(\omega) = \int_{-\infty}^{+\infty} e^{j\omega x} p_x(x) dx \tag{1.32}$$

From (1.30), (1.31) and (1.32) and for a zero mean variable, kurtosis can be formulated as:

$$Kurt(x) = E[x^4] - 3(E[x^2]^2) \tag{1.33}$$

The second term of the right side of (1.33) is a function of the variance of the variable and if we assume that the variance equals to one, (1.33) can be rewritten as:

$$Kurt(x) = E[x^4] - 3 \tag{1.34}$$

For Gaussian variables, $Kurt(x) = 0$. Variables with $Kurt(x)$>0 are called super-Gaussian, e.g. the Cauchy distribution, and variables with $Kurt(x)$<0 are called sub-Gaussian, e.g. the uniform distribution. Typically non-Gaussianity is measured by the absolute value of the kurtosis in which case by maximizing the absolute value of $Kurt(x)$ we can find the independent components. Two different methods for maximizing kurtosis are explained in the next subsection (Soong 2004).

1. Maximizing kurtosis by Gradient ascent: gradient of a function is a direction in which the function is growing most strongly. By taking this fact into account, the maximum of a cost function is computed iteratively by starting from some initial point *w*, calculating the gradient of the cost function at this point and then moving in the direction of the gradient with a suitable step size. In maximizing the cost function *J(w)* by gradient ascent the initial point is updated by the new point iteratively. The update rule is given by equation (1.36).

$$w_n = w_{n-1} + \alpha(t)\frac{\partial J(w)}{\partial w}\,|_{w=w_{n-1}} \tag{1.35}$$

In which, α is the step size which is usually a function of time and *n=2,3,…* , w_n is the new point which is computed from previous point w_{n-1} via the update rule given in equation (1.35). The term $\frac{\partial J(w)}{\partial w}\,|_{w=w_{n-1}}$ is the gradient of the cost function *J(w)* at point w_{n-1}. The iteration is continued until the new point converges to the optimum (maximum in this case) point. In practice, convergence is when the Euclidean distance between $w_n$and $w_{n-1}$is below a specific value (Hyvärinen and Oja 2000; Hyvärinen et al. 2001). Equation (1.35) can be re-written as:

$$\Delta w = w_n - w_{n-1} = \alpha(t)\frac{\partial J(w)}{\partial w}\,|_{w=w_{n-1}} \tag{1.36}$$

Where Δw indicates the change in *w*. Therefore,

$$\Delta w \propto \frac{\partial J(w)}{\partial w}\,|_{w=w_{n-1}} \tag{1.37}$$

Equation (1.37) shows that the change in *w* is proportional to the gradient direction. Hence, in optimization by gradient descent we are only interested in gradient direction. Similar fashion is employed for maximizing a cost function for vectors.

According to equation (1.21) in order to find the independent component, the whitened data (matrix $\boldsymbol{Z}$) is linearly transformed to another matrix ($\boldsymbol{Y}$) by multiplying it with matrix ($\boldsymbol{W}$); with rows of matrix $\boldsymbol{Y}$ to be the independent components, i.e. $\boldsymbol{Z}$ is whitened data, so the product of $\boldsymbol{W}$ and $\boldsymbol{Z}$ provides $\boldsymbol{Y}$ which matrix of ICs. Therefore, matrix $\boldsymbol{W}$ should be estimated in a way that product of $\boldsymbol{W}$ and $\mathbf{Z}$ results in a matrix that the kurtosis of its rows are maximum. For one of the independent components *y* (one row of matrix $\boldsymbol{Y}$), start from a vector "$\boldsymbol{w}$" ($\boldsymbol{w}$ is one column of matrix $\boldsymbol{W}$ given by (1.21)) and calculate the $\boldsymbol{y} = \boldsymbol{w}^T\boldsymbol{Z}$ in the direction that kurtosis is growing most strongly. Therefore, first we need to find the gradient of the absolute value of the kurtosis, then by moving in this direction maximum absolute value of kurtosis will be found.

Gradient of the absolute kurtosis can be calculated as follows:

From the equation (1.33) and also considering that for whitened data $E[(\boldsymbol{w}^T\boldsymbol{Z})^2]=\|\boldsymbol{w}\|^2$, the gradient of absolute kurtosis is given by (1.38).

$$\frac{\partial|kurt(\boldsymbol{w}^T\boldsymbol{z})|}{\partial w} = 4sign\big(Kurt(\boldsymbol{w}^T\boldsymbol{Z})\big)[E[\boldsymbol{z}(\boldsymbol{w}^T\boldsymbol{Z})^3] - 3\boldsymbol{w}\|\boldsymbol{w}\|^2] \qquad (1.38)$$

Since the function is optimized on unit sphere $\|\boldsymbol{w}\|^2$=1, $\boldsymbol{w}$ should be normalised (divided by its norm) after every step. Since $3\boldsymbol{w}\|\boldsymbol{w}\|^2$ in (1.35) is only changing the norm of the $\boldsymbol{w}$, the latter term in brackets in (1.38) can be omitted and change in $\boldsymbol{w}$ can be written as (1.39).

$$\Delta w \propto sign\big(Kurt(\boldsymbol{w}^T\boldsymbol{Z})\big)E[\boldsymbol{z}(\boldsymbol{w}^T\boldsymbol{Z})^3]$$
$$\boldsymbol{w} \leftarrow \frac{\boldsymbol{w}}{\|\boldsymbol{w}\|} \qquad (1.39)$$

In equation (1.39), the expression $\boldsymbol{w} \leftarrow \frac{\boldsymbol{w}}{\|\boldsymbol{w}\|}$ indicates that in every step the $\boldsymbol{w}$ is substituted by its normalised version. An advantage of using this method is that maximizing by gradient ascent is easy to implement. A drawback, however, is that the convergence is slow, and depends on a good choice of the step size. Moreover, for the cases in which the cost function is not simple, there is a chance that gradient algorithm points at a local extrema instead of the global extrema. Therefore, selecting good initial values have an important effect on the algorithm performance (Hyvärinen et al. 2001)

Fast-fixed point algorithm using kurtosis: this method is an alternative methods for maximizing kurtosis and makes the learning radically faster and more reliable (Hyvarinen and Oja 1997). In this method the problem of slow convergence is solved by selecting the initial vector in the gradient direction.

Therefore we will have:

$$\boldsymbol{w} \propto E[\boldsymbol{z}(\boldsymbol{w}^T\boldsymbol{z})^3] - 3\boldsymbol{w}\|\boldsymbol{w}\|^2\} \qquad (1.40)$$

This equation suggests a fixed-point algorithm. Initially, the right-hand side of the equation (1.40) is computed, and the value is given to $\boldsymbol{w}$:

$$\boldsymbol{w}_n = E[\boldsymbol{z}(\boldsymbol{w}_{n-1}^T\boldsymbol{z})^3] - 3\boldsymbol{w}_{n-1} \qquad (1.41)$$

where, *n= (2, 3, 4,…)*, $\boldsymbol{w}_n$ is the new vector which is computed from $\boldsymbol{w}_{n-1}$via the equation (1.41) and $\boldsymbol{z}$ is the whitened. Similar to gradient ascent, to apply the constraint $\|\boldsymbol{w}\|$=1 after any fixed-point iteration, $\boldsymbol{w}$ is normalized by its norm. The final vector $\boldsymbol{w}$ gives one of the independent components as the linear combination

$\boldsymbol{w}^T\boldsymbol{z}$ (Hyvarinen and Oja 1997; Hyvarinen 1999; Hyvärinen et al. 2001). Note that in equation (1.40) expectation is estimated by an average over large sample of $\boldsymbol{z}$ and convergence happens when the old and new values of $\boldsymbol{w}$ point in the same direction, i.e., if $|\boldsymbol{w}_n^T\boldsymbol{w}_{n-1}|$ is close to one. This algorithm is called Fast ICA. It has been shown by (Hyvarinen and Oja 1997) that such an algorithm converges very quickly and faster than gradient ascent algorithm. Moreover, contrary to gradient-based algorithms, there is no step size or other adjustable parameters in the algorithm, which makes it easy to use, and more reliable. (Hyvarinen and Oja 1997; Hyvarinen 1999; Hyvärinen and Oja 2000).

The main problem of using kurtosis is it is sensitive to outliers. This means that large observations in the tail, i.e. large samples which occur infrequently in the distribution can affect the kurtosis considerably. Therefore kurtosis is not a robust measurement for non-Gaussianity (Hyvärinen et al. 2001). Alternatively, non-Gaussianity can be quantified by negentropy which is the topic of the next subsection.

1.6.3.6. Non-Gaussianity by Negentropy

As a solution for the drawback of kurtosis in non Gaussianity measurement, negentropy which is more robust than kurtosis can be employed. But negentropy is computationally complicated. In information theory and statistics, negentropy is used as a measure of distance to normality (A. Hyvärinen, J. Karhunen et al. 2001). The entropy of a random variable is related to the information that the observation of the variable gives. Entropy can be considered as a measurement of randomness of a signal. The higher the signal's randomness the larger its entropy. Entropy for a discrete random variable $\boldsymbol{x}$ is given by equation (1.42).

$$H(\boldsymbol{x}) = -\sum_i P(x = a_i)\log(P(x = a_i)) \tag{1.42}$$

Entropy for continuous random variable $\boldsymbol{x}$ with pdf (probability density function) p_x is often called differential entropy and it is defined as:

$$H(x) = -\int p_x \log(p_x)\ dx = \int f(p_x)\, dx \tag{1.43}$$

In which

$$f(p) = -p\log p \tag{1.44}$$

Since p_x can be larger than one, the differential entropy can be negative. However, for discrete variables entropy is positive, since the probabilities necessarily stay in interval [0,1]. Hence, "small differential entropy", may be a negative number and have a large absolute value. Signals whose probability densities take large values have small entropy, since these give strong negative contributions to the integral in (1.43). High values of p_x implies that certain intervals are quite probable. Using (1.43), it can be shown that for a given mean and standard deviation, the maximum entropy occurs for a Gaussian distribution (Hyvärinen et al. 2001). This means that entropy could be used as a measure of non-Gaussianity. It can be shown that differential entropy is small for the variables which are concentrated on particular intervals. Considering the fact that the Gaussian distribution has maximum entropy among all distributions with a given covariance matrix, negentropy can be defined as a measure that is zero for a Gaussian variable and it is always nonnegative for other kinds of distributions. Negentropy *J* for vector $\boldsymbol{x}$ can be simply obtained from differential entropy as follows:

$$J(\boldsymbol{x}) = H(\boldsymbol{x}_{gauss}) - H(\boldsymbol{x}) \tag{1.44}$$

$\boldsymbol{x}_{gauss}$ is a Gaussian random vector of the same correlation (and covariance) matrix as $\boldsymbol{x}$. The advantage of using negentropy is that the negentropy is the optimal estimator of non-Gaussianity (Hyvärinen and Oja 2000; Hyvärinen et al. 2001). The problem of using negentropy is, however, that it is computationally very difficult and can only be approximated for practical applications. Conventionally negentropy is approximated by using higher order cumulants. Equation (1.46) gives an approximation of negentropy by higher order cumulants (Hyvärinen et al. 2001).

$$J(\boldsymbol{x}) \approx \frac{1}{12} E\{\boldsymbol{x}^3\}^2 + \frac{1}{48}\, Kurt(\boldsymbol{x})^2 \tag{1.45}$$

The random variable $\boldsymbol{x}$ is assumed to be of zero mean and unit variance. For a zero mean random variable $\boldsymbol{x}$ which is symmetrically distributed the first term of the right hand side in equation (1.45) is zero (skewness is zero). For such cases, maximizing $J(\boldsymbol{x})$ will be the same as maximizing the square of kurtosis. It is evident that, maximizing the square of the kurtosis is equivalent to maximizing the absolute of kurtosis. Therefore, the approximation (1.45) suffers from the same non-robustness

problem of the maximizing the absolute kurtosis. It has been shown by (Hyvärinen and Oja 2000; Hyvärinen et al. 2001) that a better approximation of negentropy is given by equation (1.46).

$$J(x) \approx \sum_{i=1}^{n} c_i[E[G_i(\boldsymbol{x})] - E[G_i(\boldsymbol{v})]]^2 \tag{1.46}$$

where, c_i are scalar coefficients, $\boldsymbol{v}$ is a zero mean and unit variance Gaussian distributed variable, and n is the number of G_i which are non-quadratic examples of which are:

$$G_1(u) = \frac{1}{a_1} log\, cosh\, a_1 u \qquad G_2(u) = -exp(-u^2/2) \tag{1.47}$$

where $1 \leq a_1 \leq 2$ is some suitable constant. For the case in which, only one non-quadratic function is used equation (1.47) can be written as:

$$J(x) \approx c[E\{G(\boldsymbol{x})\} - E\{G(\boldsymbol{v})\}\,]^2 \tag{1.48}$$

This approximation is computationally simple, fast and robust. To maximize negentropy given by equation (1.48) gradient ascent can be employed in a similar fashion as was used for kurtosis. Using gradient ascent algorithm for maximizing negentropy as a measurement of non-Gaussianity gives:

$$\Delta \boldsymbol{w} \propto \gamma E[\boldsymbol{z} g(\boldsymbol{w}^T \boldsymbol{z})]$$
$$\boldsymbol{w} \leftarrow \frac{\boldsymbol{w}}{\|\boldsymbol{w}\|} \tag{1.49}$$

and $\gamma = E[G(\boldsymbol{w}^T \boldsymbol{z})] - E[G(\boldsymbol{v})]$, $\boldsymbol{z}$ is whitened data, g is derivative of the non-quadratic function G and $\boldsymbol{w}$ is the vectors that $\boldsymbol{w}^T \boldsymbol{z}$ gives an IC.

Maximizing negentropy can also be carried out by using fast-fixed point algorithm. It is shown by (Hyvärinen and Oja 2000; Hyvärinen et al. 2001) that a fixed point iteration given by equation (1.50) can be used for maximizing negentropy.

$$\boldsymbol{w} \leftarrow E[\boldsymbol{z} g(\boldsymbol{w}^T \boldsymbol{z})] - E[g'(\boldsymbol{w}^T \boldsymbol{z})]\boldsymbol{w} \tag{1.50}$$

In which, g and g' are the first and second derivatives of non-quadratic function G given by equation (1.47) and can be calculated from (1.51).

$$g'_1(x) = a_1(1 - tanh^2(a_1 x))$$

$$g'_2(x) = (1 - x^2)exp(-\frac{x^2}{2}) \tag{1.51}$$

It is worth mentioning that the expectations in formulas above are estimated in practice as an average over the available data samples. What was explained about ICA by non-Gaussianity has been summarized in figure below.

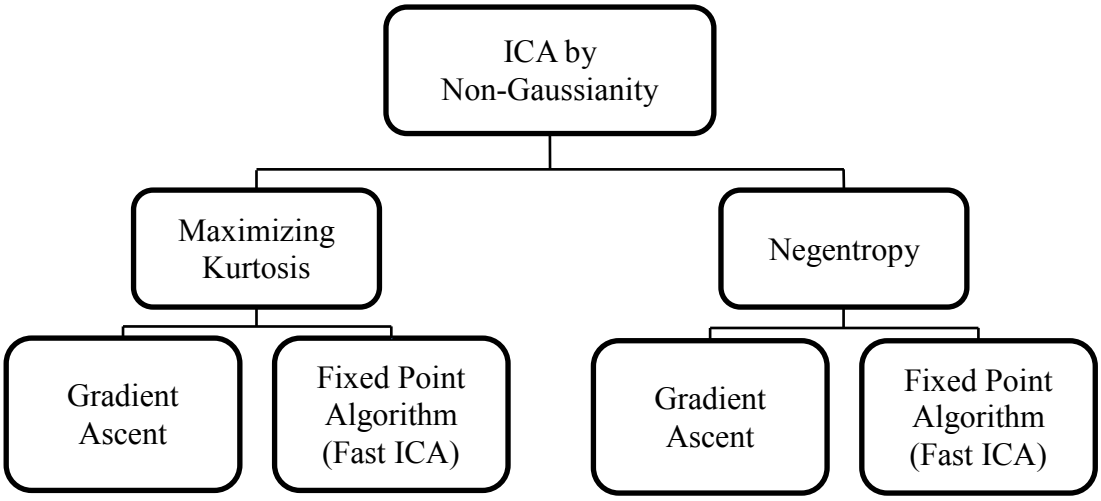

Figure 1.13: Two different methods of finding ICs based on maximum non-Gaussianity

To sum up, for an $m \times T$ observed matrix $\boldsymbol{X}$, PCA and ICA are the methods which find another matrix, with different dimensions $L \times T$, $L < m$, which only contains the main components of the original matrix from which the original matrix could be reconstructed. For cases in which data have Gaussian distribution, whitening has the same performance as ICA for separating the independent components, since for data with Gaussian distribution un-correlatedness implies independence. But for the data with other distributions, ICA is stronger than PCA for decomposing the data into its main components.

In reality the mixing matrix is not necessarily linear. Moreover, in most of recorded bio-signals the noise is not negligible. Thus, the BSS model for the observation signal can be rewritten as:

$$\boldsymbol{X}(t) = f\{\boldsymbol{S}(t)\} + \boldsymbol{n} \tag{1.52}$$

In which ***X(t)*** is one observation matrix, ***S(t)*** is the matrix of independent components, f is any (linear or non-linear) function of ***S(t)*** and ***n*** is additive noise.

Noise is generally assumed to be independent and identically distributed (i.i.d.), spatially and temporally.

In ICA and PCA methods for noise reduction the general idea is to classify the main components of the observed data into signal of interest and noise parts. Then, by discarding the noise part and reconstructing the signal using only the response components a clearer signal will be obtained, i.e. with a higher SNR.

ICA can also be used in artifact rejection by identifying ICs as either signal or noise components; and then reconstructing the signal with only good components. Existing noise/artifact identification methods using ICA mostly follow two conventional algorithms, i.e. by visual inspection or alternatively by constraint ICA (cICA). Firstly, undesired ICs are identified following visual analysis and usually using spatial distribution of the components. Using the information obtained from spatial distribution allows incorporation of additional prior knowledge into the artifact rejection problem. It indicates the region on the head where the IC originate, which can aid in deciding if the ICA represents an artefact or not. Alternatively, in some cases constraint ICA (cICA) (James and Hesse 2005) can be used. In cICA, independent components are found by satisfying the constraint of being close, e.g. mean square error, to a reference signal that is linked to the artefact, e.g. the heart-beat as observed from the ECG, or an eye-blink as observed from electrodes around the eye. This makes the algorithm able to look for a particular component, which is likely to be an artifact component. The artifact component is subtracted from the signal after being projected to the measurement space (James and Hesse 2005; Rajapakse and Wei 2006).

For the case of ALR recording, considering the fact that if the recording electrode is placed close to the response generator sources in the brain, the amplitude of the response will generally be larger and the recorded signal would be less noisy. Increasing the number of electrodes increases the chance of being close to the generator sources. However, if we use too many electrodes, i.e. if the distance between the electrodes is not bigger than a minimum value, neighbouring electrodes will record (almost) the same signal which would not be useful in analysis. This would increase the number of electrodes, and hence the time needed in putting on the scalp cap. Highly redundant recordings may also lead to less robust calculation of the independent components. According to empirical evidence, for the case of

multichannel recording conventional electrode placement, i.e. spreading the electrodes on the scalp rather than concentrating the electrodes on a particular part of the brain will provide the largest signal amplitude for evoked potentials. Moreover, for the cases in which we are interested in investigating the contribution of the different regions of the cortex in an event related brain response, e.g. does the frontal part of the brain contribute in auditory late responses or not, multichannel recording will be a stronger tool than single channel recording. Signal components should be selected carefully. If the signal component is discarded by mistake or alternatively if the noise component kept as the response the quality of the reconstructed signal will be poor and it can be said that the noise reduction method is not reliable. Different methods of component selection are explained and compared in the third and fourth chapter.

1.7. SNR Calculation

SNR as a measure of quality of the signal can be selected as a criterion for comparing the performance of noise reduction from different methods. There are different methods available for calculating the SNR. In the next section two well established methods for calculating SNR are explained.

1.7.1. SNR estimation using fixed-single-point (Fsp)

For a signal model of:

$$\boldsymbol{X}=\boldsymbol{S}+\boldsymbol{N} \tag{1.53}$$

Where $\boldsymbol{X}$ is matrix output signal, i.e. rows of the matrix are observed signal on each electrode, $\boldsymbol{S}$ is matrix of independent components and $\boldsymbol{N}$ is background noise. For averaged signal, since $\boldsymbol{S}$ is assumed to be deterministic, i.e. does not change in amplitude, latency and morphology over *M* epochs, equation (1.53) can be re-written in the form shown in equation (1.54).

$$\bar{\boldsymbol{X}}=\mathbf{S}+\bar{\boldsymbol{N}} \tag{1.54}$$

Where $\bar{\boldsymbol{X}}$ is the averaged observed signal, $\bar{\boldsymbol{N}}$ is the averaged noise and $\boldsymbol{S}$ is the averaged response. From (1.52) SNR can be formulated:

$$SNR = \frac{rms(\boldsymbol{S})}{rms(\bar{\boldsymbol{N}})}$$

$$SNR^2 \equiv \frac{var(\boldsymbol{S})}{var(\bar{\boldsymbol{N}})} \tag{1.55}$$

where *rms* denotes root mean square of the signal and *var* denotes the variance. Calculating the variance of both side of (1.54) we have:

$$var(\bar{\boldsymbol{X}}) = var(\boldsymbol{S}) + var(\bar{\boldsymbol{N}}) + 2cov(\boldsymbol{S}, \bar{\boldsymbol{N}}) \tag{1.56}$$

Equation (1.56) can be written as:

$$var(\bar{\boldsymbol{X}}) = \left[\frac{var(\boldsymbol{S})}{var(\bar{\boldsymbol{N}})} + \frac{var(\bar{\boldsymbol{N}})}{var(\bar{\boldsymbol{N}})} + \frac{2cov(\boldsymbol{S},\bar{\boldsymbol{N}})}{var(\bar{\boldsymbol{N}})}\right] var(\bar{\boldsymbol{N}}) \tag{1.57}$$

Assuming that the noise and signal are uncorrelated, the third term in equation (1.56) is zero. This method (*Fsp*) with the assumption of having single noise source, suggests that the background noise distribution can be approximated by collecting the values in one single point of each individual epoch. With increasing number of epochs the sample of single point values converges to the distribution of the real background noise. After a few hundred epochs, the variance of the single point samples, $var(SP)$, will be a rather accurate measure of the variance of the true background noise, $var(N)$ (Elberling and Don 1984). It has been shown by (Elberling and Don 1984) that the value of *Fsp* is estimated by equation (1.58).

$$Fsp = \frac{var(\bar{\boldsymbol{S}})}{var(\overline{\boldsymbol{SP}})} \tag{1.58}$$

where, $var(\bar{\boldsymbol{S}})$ denotes variance of the averaged signal and $var(\overline{SP})$ denotes the variance of the single points selected in each epoch. It is worth noting that the *Fsp* is the an abbreviation of *F-ratio single point* (Hood 1998).

1.7.2. SNR Calculation by Fixed multiple points (*Fmp*)

Alternatively, by assuming that the noise has multiple generating sources, SNR as a measure of signal quality and for the zero mean data can be estimated by adopting the formula from (Silva 2009):

$$SNR = \frac{\sigma^2_{Signal}}{\sigma^2_{Nres}} = \frac{\sigma^2_{Save}}{\underbrace{\sigma^2_{Nres}}_{Fmp}} - 1 \tag{1.59}$$

where σ^2_{Save} is the variance of the averaged trials and σ^2_{Nres} is the residual noise power. σ^2_{Nres}, in equation (1.59) can be estimated by segmenting the trials into stationary regions and estimating noise variances within the regions. The noise variance can be estimated by selecting *L* fixed points with respect to the stimulus onset and measuring the variability of these points across *M* trials. By using the formula suggested by (Elberling and Don 1984), the locally stationary noise source power $\sigma^2_{\eta i}$ can be estimated as:

$$\sigma^2_{\eta_i} = \frac{1}{L}\sum_{n=0}^{L-1}\frac{1}{M_i-1}\sum_{m=1}^{M_i}\left(x(n,m) - \overline{x(n)}\right)^2 \tag{1.60}$$

Where, *L* is the number of the fixed points, $\sigma^2_{\eta_i}$ is the noise source power and $\overline{x(n)}$ is the average of a fixed point over *M* trials and M_i is the i^{th} epoch. Therefore, σ^2_{Nres} can be calculated in the averaged signal from:

$$\sigma^2_{Nres} = \frac{1}{M^2}\sum_{i=1}^{m} M_i\sigma^2_{\eta_i} \text{ with } \sum_{i=1}^{N} M_i = M \tag{1.61}$$

Computer simulations of averaging under non-stationary noise sources done by (Silva 2009) showed that *Fsp* has a higher mean square error (*MSE*) than *Fmp* for estimating residual noise, i.e. *MSE* increases as the number of noise sources decrease.

Chapter 2

Test Protocol and Data Acquisition

2.1. Introduction

There is a misperception that by following a fixed test protocol any tester, even with minimal technical skills, can obtain reliable, valid and clinically useful AERs. Test protocols are strongly dependent on applications and often need to be modified during testing, i.e. reason for the assessment and unpredictable environmental and subject variables. Therefore, it is not possible to design and follow a fixed and inflexible protocol and assess many patients effectively and efficiently. A clinician is frequently required to revise his or her strategy (Hall 2007).

This chapter describes the design of a test protocol for recording brain auditory late response (ALR) using 63 channels of EEG for three clinical applications. Exploring the effect of attention on the ALR waveform, measuring inter/intra subject variability and measuring hearing threshold sensitivity. A key question is whether multichannel signal processing can be more sensitive compared to single channel signal processing alternatives and find a the hearing threshold closer to 0 dB nHL in normal hearing participants. By definition, on average a stimulus can be heard at 0 dB nHL (Gelfand 2007), but it elicits a brain response at

low level (small amplitude), and so may not be detected with evoked potentials (Beagley 1967).

The main objective of this research work is to explore whether and by how much, multichannel signal processing strategies can improve the quality of the AER signals, and consequently their diagnostic utility for different clinical applications. In this work signal to noise ratio (SNR) is used as a measure of signal quality and it is measured by using Fixed-multiple-point (*Fmp*) (Silva 2009). In addition, we want to see whether we can reduce acquisition time as we may be able to achieve a given SNR in less time using multichannel analysis.

In the first part of this chapter, a method of selecting adequate number of participants is explained. Afterward, subject preparation for multichannel recording of ALR and calibration of stimulus presenter are explained. Eventually, testing protocol (including stimulus factor and recording settings) for multichannel recording of ALR is presented along with a brief explanation of data acquisition for three clinical applications (hearing threshold measurement, monitoring the effect of attention and inter/intra subject variability). For the lab work risk assessments were carried out and ethic approval obtained (Ethics approval number: 1204).

2.2. Subject Selection

When planning an experimental trial, it is very important to consider how many participants are needed to reliably answer the clinical questions. Too many participants are a needless waste of resources and too few participants will not produce a precise, reliable and definitive answer. Both cases are considered unethical. Under this common scenario, patients might be denied a useful treatment because the trials were underpowered (i.e. too small to detect a treatment effect) – this can also result in further studies being cancelled without good reason (Woodward 2005). The objective of sample size calculation is to calculate and fix a sample size which is adequate for it to be highly probable that the study will detect a real effect as statistically significant. Conversely, there must be adequate confidence that this effect is genuinely absent if it is not detected in the study (Röhrig 2010). The sample size can be estimated by using prediction of the means of SNRs calculated for each of the methods (single channel versus multichannel) with the assumption of having equal variances from both methods.

IBM SPSS Sample power 3 was employed to estimate the sample size. In this calculation the first important step is to know how big of a difference is expected to be observed. This can be calculated from pilot studies, previous works or an educated guess. From the results of pilot studies, multichannel processing improved SNR by factor 2.75 in comparison with single channel processing. For a significance level α=0.05 and power level 95% the adequate number subjects was found to be 10 people. But, as each of the channels in multichannel recording is itself a single channel recording of data each of the channels of multichannel recording can be used as the single channel sample for each subject. Note that the test is a repeated measure design in which single channel data quality is compared to multichannel data quality in the same group of subjects.

2.3. Subject Preparation

To make sure that the subjects had normal middle ear function, the audiometric tests of tympanometry and audiometry were performed. Tympanometry is a measure of how well sound is transmitted by the eardrum when negative or positive pressure is applied to the eardrum. A small rubber tip placed in the ear to seal the ear canal and a pump alters pressure. At the same time sound is emitted and the return signal is measured. As the eardrum stiffens it is pushed in or out. The eardrum acts like a reflector sending the echo back to the tympanometer. The degree of compliance (transmission of sound through the eardrum) against the pressure applied is measured and the result is what is shown as a tympanogram. Lack of movement of the eardrum can indicate a build-up of fluid behind the middle ear. This can be caused by an infection or it may indicate blockage of the eustachian tube. If these conditions exist, a flat tympanogram score appears on the report. If tympanometry did not show any hearing problem with the subjects, a pure tone audiometry test was then performed for finding hearing range threshold level in the frequencies, 1 kHz, 2 kHz, 4 kHz, 8 kHz, 500 Hz and 250 Hz (Gelfand 2007).The normal hearing range in adults is from 0 dB hearing level to 20 dB hearing level.

An audiogram is a graph that shows the audible threshold for standardized frequencies as measured by an audiometer. The Y axis represents intensity measured in decibels and the X axis represents frequency measured in Hertz (Gelfand 2007). Most audiograms cover a limited range of frequencies from typically 100 Hz to 8 kHz,

because this range includes the main frequencies of sounds in speech. The threshold of hearing is plotted relative to a standardized curve that represents 'normal' hearing, in dB hearing level (HL). An Audiogram for a normal hearing person is shown in Figure 2.1.

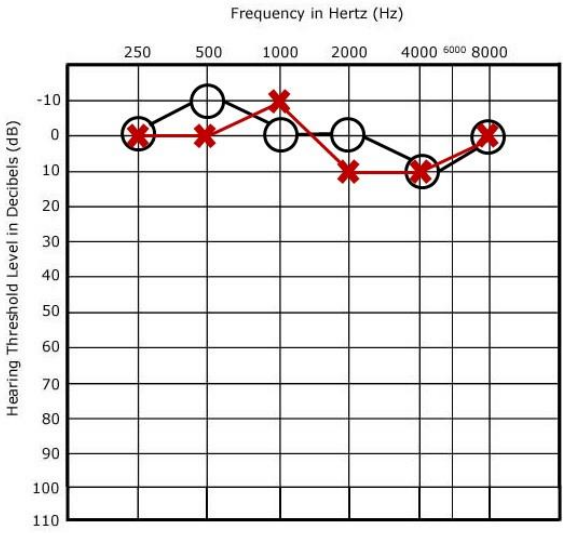

Figure 2.1: Audiogram for a normal hearing subject. Hearing threshold at each frequency is shown by circle for the right ear and cross for the left ear. People with hearing thresholds within the range of -10 to 20 dB HL are considered as normal hearing subjects.

For recording AERs, a scalp cap of 63 channels was applied. In order to have good electrode contact with the skin and have the recorded signal as clean as possible, chemicals such as hair gels and hair mousses should be washed from the hair (e.g. with warm water and baby shampoo) to remove any possible layer which might prevent the connection between electrodes of the scalp cap and head skin. Following this the scalp cap is put on the subject's head. In this work the 10-20 electrode system is used for electrode placement. The electrode positions are constructed by dividing the line between the nasion and inion into 10% or 20% intervals. The line between the preauricular points is divided similarly (Teplan 2002). The 10-20 electrode placement system schematic is shown in Figure 2.2.

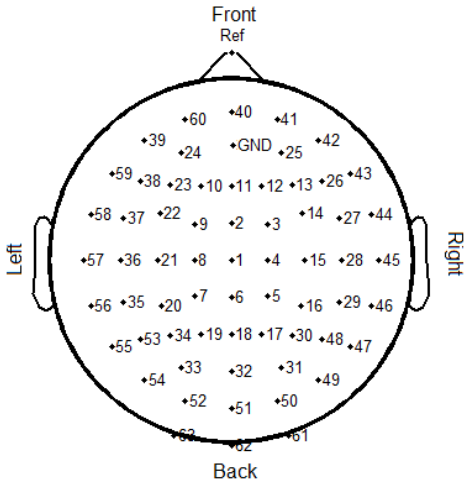

Figure 2.2: 10-20 electrode placement system.

Since the signals generated in the cortex but recorded from the scalp, the signal has to pass through skull, skin and hair. This pathway has an impedance that causes a drop in the signal amplitude. The impedance between the brain and the scalp is not controllable but the impedance between the electrode and the scalp can be controlled and should be as small as possible, i.e. usually less than 5 kΩ. A conductive gel should be injected to the electro cups (i.e. the syringe should be prepared with Abralyt HiCl gel, avoiding the introduction of air bubbles in to the syringe; a big bubble of air introduced in an electro cup usually results in poor impedance) which allows a continuous contact between the skin and the electrodes, despite the small movements of the subject. Putting on a scalp cap with 63 electrodes and reducing the impedances, between electrode and scalp, takes about 45 to 60 minutes (Tyner et al. 1989; Teplan 2002). These steps prepare the subject for testing.

2.4. Calibration and Finding Zero dB nHL

The reference for the decibel used to express deviation from normal hearing sensitivity for nonstandard AEP stimuli, such as clicks, is dB nHL (dB normal Hearing Level) (for standard pure tones the dB HL scale can be applied). According to (Gelfand 2007), the ear is not equally sensitive to all frequencies, i.e. hearing sensitivity changes as a function of the frequency of the sound; therefore 0 dB nHL represents normal hearing level at each frequency. For example 25 dB nHL is 25 dB above the threshold for a group of normal hearing people at that frequency. Audiometric zero dB can be established by testing a group of normal hearing people, 18 to 30 year olds as the reference. For this study, five persons (ten ears) were tested

to find 0 dB nHL for 1 kHz. For this group of five normal hearing subjects the zero was found at 15 dB SPL.

Since the stimulus is presented through a stimulus generator (*REM Fireface 400 sound card*) the stimulus generator needs to be calibrated to make sure that the sound card generates the stimulus at different levels accurately. To calibrate the sound card, a sound generator which produces a pure tone at 94 dB SPL was used as a reference for the calibration. The calibrator was used to calibrate an occluded ear simulator which contains a microphone that converts pressure to voltage. The peak value of the sound was measured to be 0.37 V by using an oscilloscope. The peak equivalent of the stimulus without any attenuation was also measured and it was found to be 8.32 V. The intensity of the stimulus in dB Peak equivalent Sound Pressure Level can be calculated form:

$$94 + 20\log(\frac{V_{St}}{V_{ref}}) = 94 + 20\log(\frac{8.32}{0.37}) = 121.04 \approx 121\ dB\ pe\ SPL \quad (2.1)$$

where, V_{St} is peak value of the stimulus and V_{ref} is the peak value of the reference.

To make sure that any decrement in the stimulus intensity applied by the sound card can be precisely transfer to the earphones, the intensity of the stimulus was decreased by 10, 20 and 40 dB steps via the sound card; and the peak of the sound signal was measured using the oscilloscope at the output. Then peak equivalent sound pressure level (pe SPL) was measured at each stage using equation (2.1). Table 2.1 shows the calibration procedure for the reference 0.37V.

	Without Attenuation		10dB Attention		20dB Attention		40dB Attention	
	Peak Volt	dB pe SPL	Peak Volt	dB pe SPL	Peak Volt	dB pe SPL	Peak Volt	dB pe SPL
ALR Stimulus*	8.32 V	121 dB	2.48 V	110 dB	0.48 V	98 dB	60m V	79 dB

* Suitable stimulus for ALR extraction is explained in the next section.

Table 2.1: Calibration table for the ALR stimulus for the reference voltage equal to 0.37V.

It can be seen form Table 2.1 that any attenuation applied by the sound card can be observed through the measurement. For the ideal case any change in the input should be transmitted to the output without any change. For this case, calibration curve given in Figure 2.3 shows that, 10 dB attenuation in the input (via the sound

card) results almost 10 dB attenuation at the output (earphone). Mean squared error for this calibration was found to be 2.38 dB.

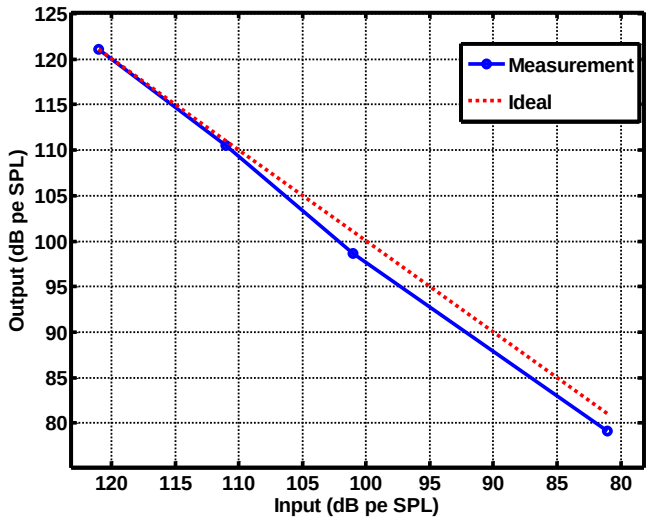

Figure 2.3: Calibration curve for calibrating the equipment for frequency 1 kHz. Any attenuation in input is transmitted to the output and the mean square error of the calibration is equal to 2.38 dB.

2.5. Test Protocol for ALR Recording

Ten normal hearing subjects (5 male, 5 female) aged between 18 and 30 years old volunteered to participate in to this study. All the subjects were checked to have normal hearing, i.e. normal results on tympanometry and behavioural pure tone audiometry. Moreover, the testing procedure was explained to the subjects and a consent form was signed by each of them. As was suggested in (Hall 2007), the stimulus for recording the ALR was selected to be a 1 kHz tone burst of 70 ms duration, with a 10 ms rise/fall time which was generated by a PC through a *REM Fireface 400* soundcard by using software called "*Presentation 14.0.1*". The stimulus was presented 160 times at 60 dB nHL through an insert earphone to the right ear of the participants. This is a moderately loud sound which should elicit good quality signals. The EEG was recorded from the participants' scalp though a 63 channels scalp cap and an EEG system called *Neuroscan system (scan 4)* with the reference set on the nose. Since the ALR signal has low frequency content, it is low pass filtered by 100 Hz at the recording stage (Hall 2007). In addition, two more electrodes were placed on the chest and the cheek of each subject to record the heart beat and eye blink signals, respectively. The time required for recording the ALR was 4 min;

additionally, a 4 min period of no-stimulus resting EEG was also recorded. The resting EEG (without stimulation) was recorded for estimating the noise distribution.

The ALR was recorded from the participant's brain by using stimulus factors given in Table 2.2. This data set was used for three clinical applications hearing sensitivity threshold measurement, checking the effect of attention and inter/intra subject variability) in the chapter 5. Data acquisition strategies for the aforementioned applications are briefly presented here, but they are explained in more details in chapter 5.

Stimulus parameter:	Acquisition Factors
Type: tone burst	Amplification:×75,000
Duration:70 ms, (rise/fall 10 ms, plateau 50 ms)	Number of sweeps:160
Rate: 0.7/sec	Analysis period:250ms
Intensity: 60 dB nHL	Polarity: rarefaction
Frequency: 1 kHz	Transducer: insert, ER-3A
	Filter setting: $\begin{cases} 1-100\ \text{Hz} \\ \text{slope: } 48\ \text{dB/octave} \\ \text{sampling rate: } 4\ \text{kHz} \end{cases}$
	Recording time: 55-60 min

Table 2.2: Stimulus parameters and acquisition factors of recording ALR (Hall 2007).

This experiment is divided into 3 phases as following:

Threshold measurement: For hearing threshold measurement first the stimulus given by Table 2.2, is performed at intensity of 60 dB nHL and then reduced to 0 dB nHL in eight steps. In the first step, the intensity is reduced from 60 dB nHL to 40 dB nHL and then reduce to 30 dB nHL, 20 dB nHL, 15 dB nHL, 10 dB nHL, 5 dB nHL and 0 dB nHL (Hall 2007, Lightfoot 2010). Recording the ALR at each intensity level takes approximately 4 minutes (160 stimuli, 0.7 stimuli/s), resulting in a total testing time of 32 minutes.

Inter/Intra subject variability*:* To measure inter/intra subject variability the ALR is recorded three times at the intensity of 60 dB nHL for each subject. Thus, required time for this part is 12 min.

Attention condition*:* for this the subject is asked to count the number of stimuli. Recording time required for this task is about 4 minutes.

Ignore condition*:* that subjects are asked to read something and not pay attention to the test. Recording time required for this task was also 4 minutes. This is a conventional method for ignoring the stimulus which has been used by many research (Hillyard, Hink et al. 1973, Picton and Hillyard 1974, Thornton, Harmer et al. 2007).

The overall required time for ALR testing was 52 min. Moreover, a 4 minutes EEG without stimulation was also recorded for noise estimation. Therefore, the overall recording time for the test was around 1 hour. Including the time for putting the scalp cap into account, the total time needed for a session was 2 hours. All the steps of the test are summarized in Table 2.3.

Finding Hearing threshold	Attention Effect	Inter/Intra subject variability
60 dB nHL	60 dB nHL (Counting stimulus	60 dB nHL
40 dB nHL	for attention condition)	60 dB nHL
30 dB nHL	60 dB nHL (Reading a text	60 dB nHL
20 dB nHL	book to ignore the stimulus)	
15 dB nHL		
10 dB nHL		
5 dB nHL		
0 dB nHL		
No Stimulus EEG		

Table 2.3: Entire test protocol for the ALR recording for clinical applications, hearing threshold measurement, effect of attention and inter/intra subject variability.

Since the amplitude of the response is very small in comparison with the amplitude of undesired factors, a group of signal processing methods is needed to extract the response from the noise. Different signal processing strategies for noise reduction for multichannel signal recording are demonstrated in the next chapter.

2.6. Summary

In this chapter the test protocol for recording ALR was presented. Power calculation, subject preparation and equipment calibration were also explained. In addition, test design for recording the ALR for three clinical applications was briefly explained.

Chapter 3

Signal Processing Methods

3.1. Introduction

As was mentioned in the first chapter, there exist various approaches to improve the signal to noise ratio (SNR) in AER recordings such as increasing the intensity of the stimulus and amplifying, filtering etc. (Hall 2007). However, the limitations that these methods have make them inefficient. For example the intensity of the stimulus cannot be increased in an unlimited way. Sounds with intensities above a specific value (usually above 85 to 95 dB nHL) are irritating for the subjects (Hall 2007). Exposing the subjects to high intensity stimuli for long time may cause damage in their hearing system (Gelfand 2007; Hall 2007). Therefore, an efficient noise reduction method is essential for extracting information from the recorded raw signals.

In this chapter, four novel automatic ICA component selection strategies for noise reduction in the special case of event related activities are proposed and preliminary tests are presented using illustrative data. Furthermore, to explore the impact of ICA on the quality of noise reduction for biomedical signals, results are compared with equivalent procedures using PCA using simulated data. In the next chapter, the proposed methods are used for noise reduction in the multichannel recording of the ALR, captured from 10 normal hearing subjects, with the aim of

assessing the ability of these methods to improve the quality of the estimated evoked response. The performances of the new algorithms will be compared with each other and also to that of single channel estimates of the ALR, for both weighted and un-weighted averaging.

3.2. Algorithms and Simulation

To model an event related activity in the brain, two signals *x* and *y* are used as shown in Figure 3.4. It is assumed that the signal *r* is the response of the brain to the acoustic stimulus *x, n* is noise (spontaneous background EEG activities and other undesired signals) and *y* is observed signal at each electrode:

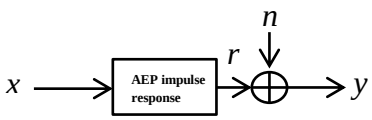

Figure 3.1: Model for an event related activity. Signal *x* is the input to the system and signal *r* is the response of the system to the input which is contaminated by noise, *n,* and *y* is the observed (recorded) signal.

This model is used for simulating the evoked potentials throughout this chapter. Two type of algorithms are presented in this section. The first two which are based on kurtosis and entropy respectively, are designed for situations in which there is no prior information available about the recorded signal. Whereas, for the second two methods which are based on SNR of the ICs and coherence of the ICs with a stimulus respectively, the onset time of the stimulus repetitions is needed as a prior information. Hence, two different sets of simulated data were used for testing the performance of the algorithms. However, to be able to compare the performance of the methods, all the four methods were also applied on the same data set.

3.2.1. Component Selection by Maximum Kurtosis of ICs (Max-Kurt-ICs)

It was shown in chapter 1, that the absolute kurtosis of each component can be used as the measure of non-Gaussianity of the component. For a random variable ***x*** kurtosis can be calculated by using the formula given in equation (1.34). Equation (1.34) can be modified and written in the form of equation (3.1).

$$k = |(\frac{E[(\boldsymbol{x}-\mu)^4]}{(E[(\boldsymbol{x}-\mu)^2])^2} - 3)| \quad (3.1)$$

where, μ is the mean of the signal $\boldsymbol{x}$, i.e. ICs for this case. (Hyvärinen et al. 2001). An observed signal can be conceptualised as a linear mixture of the activity of many sources (response and noise) in the brain. It has been shown by (Mauricio et al. 2009) that that distribution of background EEG (noise) is closer to Gaussian than the distribution of evoked potentials. Therefore, large non-Gaussianity of an IC may imply that the IC is not noise.

In this method, first absolute kurtosis was calculated for all the ICs by using equation (3.1). The expectation operator in equation (3.1) was estimated by an average over the data. Components whose absolute kurtosis is below a predefined value (threshold) are considered to be noise components and should be discarded. Selecting this threshold has an important impact on the quality of noise reduction and should be carried out carefully. If the value is selected too large, response components will be rejected as noise, i.e. components with absolute kurtosis below the threshold will be rejected, and selecting the value too small will include noise components and will lead to a noise reconstruction. In practice, this value is usually estimated through pilot studies and extra measurements. One methods of estimating this value via extra measurement is explained in the next chapter. The main drawback of this method is that the kurtosis is sensitive to outliers. For instance in a case that the response is not present in the signal and the signal is almost normally distributed, a large peak, e.g. due to muscle activity, can affect kurtosis considerably; and the IC can then erroneously be selected as a response IC. Considering that spiky signals have large kurtosis, the kurtosis method cannot thus be a suitable choice for component selection. Also it is worth noting that there exist distributions that have zero kurtosis while being far from Gaussian (Karlisnikov 2013).

To illustrate how the method works an example was run using simulated data. In this simulation, four signals whose distributions are not normal were used to evaluate the performance of the kurtosis method in solving the automatic component selection problem of the ICA. Moreover, the condition under which the kurtosis method is not a suitable choice for component selection is also illustrated. The four signals which are used in this example are shown in the Figure 3.2 left. A *sine wave* and an *impulse train* were used to simulate the response signals; and the other two

signals represent the noise factors which are random signals with non-normal distributions. In this simulation, the signals which are generated by different sources in the brain were modelled by the sine wave and the impulse train. In working with recorded data from the brain it is possible that one of the electrodes has been placed in a wrong position or it has not been fixed on the scalp properly. Consequently, the electrode either does not record a brain response to the acoustic stimulus or records a rather weak signal; while it records the other stronger signals such as muscle activity or the heart beats. Therefore, the recorded data will only contain the muscle activity or the heart beat and practically does not contain any useful information about the brain response to the acoustic stimulus. Thus, the main activity (IC) in such signals will be the muscle activity or the heart beat which both have large kurtosis, and will be selected as response component erroneously. To simulate this condition one of the noise signals is designed to have large amplitude at only one of its samples. The four signals in Figure 3.1 left are combined by using a 4×4 mixing matrix, $\boldsymbol{A}$, given by equation (3.2), and applied to the signals as given in equation (3.3).

$$\boldsymbol{A} = \begin{pmatrix} 2 & 0.5 & 10 & 1 \\ 0.4 & 20 & 0.7 & 3 \\ 15 & 5 & 0.8 & 0.2 \\ 0.1 & 5 & 0.6 & 14 \end{pmatrix} \tag{3.2}$$

$$\boldsymbol{Y} = \boldsymbol{AX} \tag{3.3}$$

where, $\boldsymbol{Y}$ (Observed signals) is a matrix that contains the linear mixture of the response signals $\boldsymbol{X}$ (Sources) whose rows are the signals shown on the left in Figure 3.1.

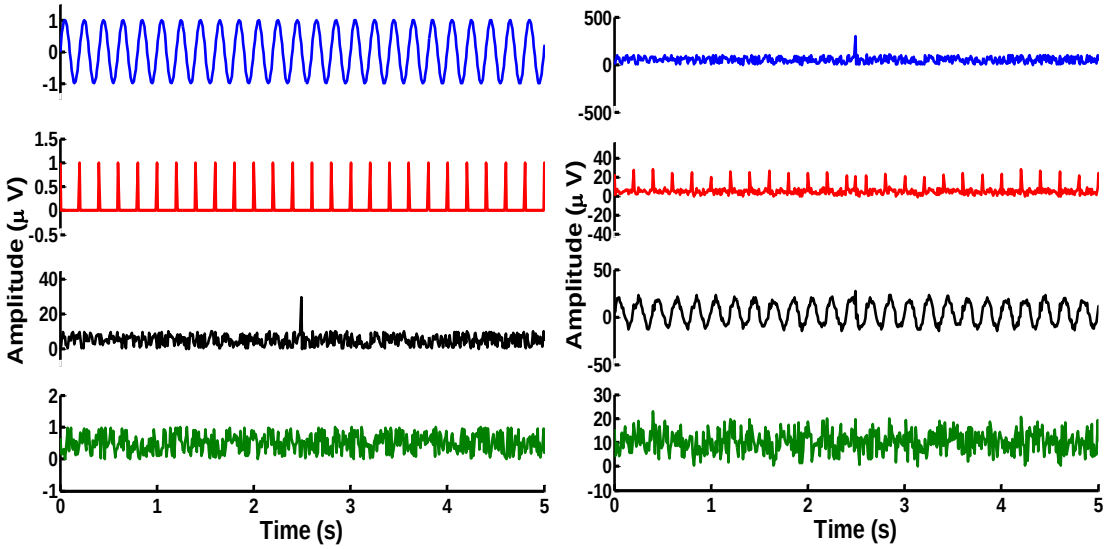

Figure 3.2: Left: The source signals which are a sine wave, an impulse train (of the same period as the sine-wave), a random signal with a large value at only one of its samples (high kurtosis) and a random signal which is uniformly distributed. Right: the mixture of the four signals using the mixing matrix ***A***.

The mixture of the signals is shown on the right in the Figure 3.1. The mixing matrix ***A*** was designed to make one signal dominant at each observation. For example the third signal is dominant in the first observation and the second signal is dominant in the second observation. The aim of this simulation was to explore if this method was able to recover the response signals (the sine wave and the impulse train).

A MATLAB implementation of FastICA (Hyvärinen et al. 2005) was used to separate the main components of the mixed signals. Considering that the observed signals are mixtures of the sine wave, pulse train and the two noise signals, it is known a prior that two of the ICs should be discarded, i.e. there are only two noise factors. Following the Max-Kurt-ICs method algorithm, the two ICs whose kurtosis are largest will be kept as the signal ICs and the other two ICs will be discarded as noise. Then the signals will be reconstructed by only using the response components. This attempted signal reconstruction is shown on the left in the Figure 3.3. The single sample with large amplitude, in the third signal from the top in Figure 3.2 left, increases the kurtosis of the corresponding IC. The absolute kurtosis of the sine wave was 1.5 and the kurtosis of the noise and the signal with one large sample were 1.2 and 2.98 respectively. Consequently, the noise (the fourth signal from the top in Figure 3.1 left) and the sine wave were removed as noise components. Additionally, to show the noise less condition, the sine wave and the impulse train were combined

and the result is shown on the right in the Figure 3.3. This illustrates the situation in which the noise is completely removed from the observed signals (ideal noise reduction).

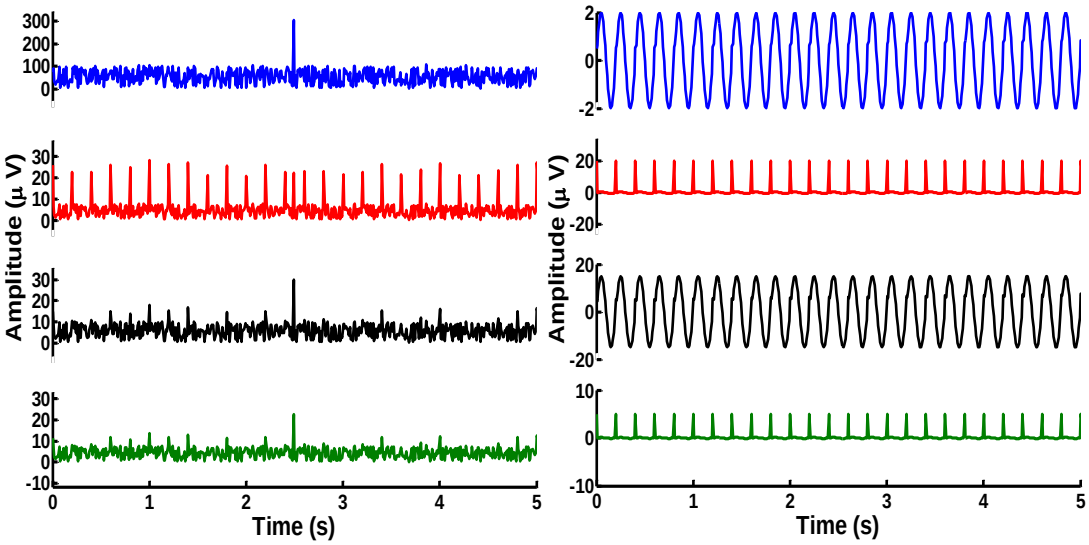

Figure 3.3: Left, reconstruction after discarding the ICs with small kurtosis. Right, ideal noise reduction by excluding the noise components in the mixtures.

As can be seen in Figure 3.3, this method does not appear to be an effective noise reduction method as the reconstructed signal was quite different from the ideal noise reduced signal. The signal to noise ratio (SNR) can be calculated for each reconstructed signal using the equation (3.4).

$$SNR = \frac{\sigma^2_{Signal}}{\sigma^2_{Noise}} \tag{3.4}$$

where, $\sigma^2_{signals}$ is the variance obtained, when excluding the noise (setting the last two columns of $\boldsymbol{A}$ in equation (3.2) to zero). Thus, the noise can be obtained by subtracting the signals which are shown on the right side of Figure 3.2 from the reconstructed signals (Figure 3.3 left). The SNRs for the reconstructed signals shown in the Figure 3.3 left (from the top) were found 0.002, 5.61, 0.97 and 1.66 – confirming the poor SNR visually evident. This result illustrates that even though the Max-Kurt-ICs has the potential to recover the signals from the noisy mixture, it fails when the contaminating component has a larger kurtosis than the component of interest.

3.2.2. Component Selection Using Minimum Entropy (Min-Entropy-ICs)

In chapter 1 the entropy of a random variable is related to the information that the observation of the variable conveys (Hyvärinen and Oja 2000). Entropy can be considered as a measurement of randomness of a signal, i.e. the more signal randomness the larger its entropy.

In this method, entropies are calculated for all the ICs by using equation (1.43) in which the expectations in the formula are approximated by averaging. ICs with small entropies are kept as the response part and the rest are discarded as the noise components. For this synthetic example the number of undesired signals is already known (two noise signals). Therefore, the two components with the largest entropies are discarded and the data are reconstructed only using the response components. However, in working with recorded data number of underlying components is unknown and for discarding the undesired components a threshold is needed. In practice, finding this threshold is not simple and should be estimated. One method of estimating this threshold for working with real data will be explained in the next chapter. Since the Min-Entropy-ICs method classifies the ICs based on the randomness, i.e. how random the signal is distributed, it is less sensitive to outliers. This method of IC selection thus addresses the major drawback of outliers for kurtosis based methods. For instance, in the previous example for the case of the signal with a single large sample, the kurtosis of the IC is increased due to the large sample; while, the signal's randomness is still large and consequently the IC will be discarded as noise.

In order to illustrate how the Min-Entropy-ICs method can address the drawback of using the Max-Kurt-ICs for noise reduction for data with outlier, the same illustrative data set (the same signals and mixing matrix) that were used in the previous section were processed using the Min-Entropy-ICs method. The results are shown on the left in Figure 3.4. By comparing these results with the results of ideal noise reduction which is shown in the Figure 3.4 on the right, it can be seen that the drawback of the Max-Kurt-ICs approach is addressed.

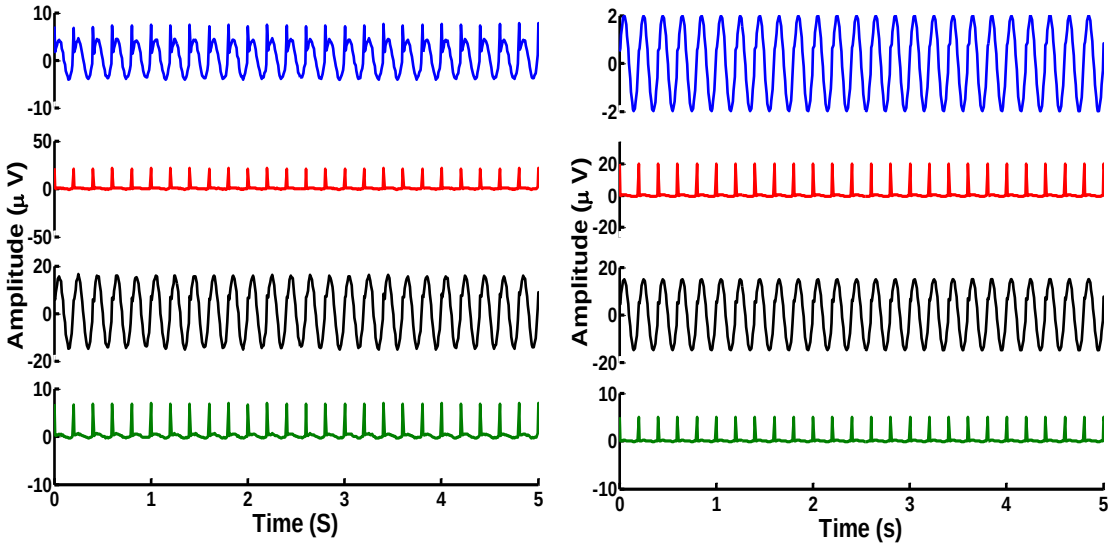

Figure 3.4: Left, noise reduction by using Min-Entropy-ICs for component selection. Right, ideal noise reduction by excluding the noise signals (the noise signals (***X***) in equation 3.7) in the mixture.

To compare the signals obtained from Min-Entropy-ICs reconstruction and the ideal noise reduction, the SNRs were calculated for each of signals. The SNR values for the signals shown in Figure 3.4 (from the top) were found to be 2.89, 562.85, 1055 and 7.35, respectively, for the four reconstructed signals, which are considerably higher than the SNRs obtained by the Max-Kurt-ICs. It may be noted that the SNR is poor for the first signal, which is dominated by noise while the other channels especially channels 2 and 3 (see mixing matrix ***A*** in equation 3.2) are dominated by signal components.

3.2.3. Component Selection by Max-Fmp of ICs (Max-Fmp-ICs)

A priori information about a signal can be helpful in noise reduction. For instance, if the signal of interest is an event related signal, this can be exploited. Two novel methods which are based on using a prior information, are proposed in the next two subsections.

Depending on the length of the response (*r*), the time interval over which search for the response should be carried out can be selected. For instance, if *r* occurs within 100 ms after stimulus (*x*) onset, we can use the onset as a reference moment and investigate100 ms after that onset in the signal to assess the response. Therefore the problem has been changed into signal detection in particular time intervals. High

SNR over intervals related to the input implies the response is present over the interval.

In this method ICs that have large SNRs over the defined intervals following the stimulus, will be selected as signal components and the rest are discarded as noise components. By setting a threshold for minimum acceptable SNR the signal components can be separated from the noise components, i.e. ICs whose SNR are above the threshold will be kept and the rest will be discarded. For simulating this condition, an impulse train as given by equation (3.6) was selected as the stimulus.

$$F(n) = \sum_{k=0}^{L} \delta(n - Mk) \tag{3.6}$$

where, δ(.) is the Kronecker delta function, *n* are samples, *M* is the delay between impulses, and *k* identifies each impulse (*L*+1 in total). In the current simulation, *M*=100 and *L*=10. Additionally, four periodic signals *P(n), Q(n),R(n)* and *S(n),* which are described by equations (3.7) to (3.10) and also a random signal with uniform distribution (to simulate the noise) were used in this simulation. The signals *Q(n)* and *R(n)* are selected to be periodic with the same period as the stimulus, given by the equation (3.6), to model the activities which are synchronized with the stimulus; while *P(n)* and *S(n)* were used to model the other components which are periodic but they are not the response to the stimulus, e.g. eye-blink or hear beats. These signals are shown on the left hand side of Figure 3.5

$$p(n) = \begin{cases} x_1(n) & 0 \le n < 200 \\ 0 & otherwise \end{cases}$$
$$P(n) = \sum_{k=0}^{4} p(n).\delta(n - 200k) \tag{3.7}$$

$$q(n) = \begin{cases} x_2(n) & 0 \le n < 100 \\ 0 & otherwise \end{cases}$$
$$Q(n) = \sum_{k=0}^{9} q(n).\delta(n - 100k) \tag{3.8}$$

$$r(n) = \begin{cases} x_3(n) & 0 \le n < 100 \\ 0 & otherwise \end{cases} \tag{3.9}$$

$$R(n) = \sum_{k=0}^{9} r(n).\delta(n - 100k)$$

$$s(n) = \begin{cases} x_4(n) & 0 \le n < 127 \\ 0 & otherwise \end{cases} \qquad (3.10)$$

$$S(n) = \sum_{k=0}^{7} s(n).\delta(n - 127k)$$

It can be deduced from the equations (3.8) and (3.9) that *Q(n)* and *R(n)* are periodic with a period of 100 samples; while from (3.7) and (3.10), it can be seen that the *P(n)* and *S(n)* are periodic with periods of 200 and 127 samples respectively. Using the mixing matrix $\boldsymbol{A}$ given by the equation (3.11) the four signals and noise in Figure 3.5 (left) are mixed and the result is shown in Figure 3.5 (right).

$$\boldsymbol{A} = \begin{pmatrix} 15 & 0.2 & 1 & 0.6 & 3 \\ 0.3 & 12 & 2 & 2 & 4 \\ 0.7 & 3. & 10 & 1.5 & 2 \\ 0.2 & 0.5 & 0.75 & 13 & 3.5 \\ 1 & 1.5 & 2.9 & 3 & 17 \end{pmatrix} \qquad (3.11)$$

This mixture can be written in the matrix form as:

$$\boldsymbol{Y} = \boldsymbol{AX} \qquad (3.12)$$

where $\boldsymbol{A}$ is the mixing matrix, $\boldsymbol{Y}$ is a matrix which includes the signals and the noise (source matrix) and the $\boldsymbol{X}$ is a matrix which contains the linear mixture of the signals and noise (observation matrix). It can be seen in equation (3.11), that the elements of the rows of the mixing matrix $\boldsymbol{A}$ are selected in a way that one signal is dominant, i.e. it has larger weight, in the mixture at each time. For instance in the first row of the matrix $\boldsymbol{Y}$, the first signal has weigh 15 and it is dominant in that mixture.

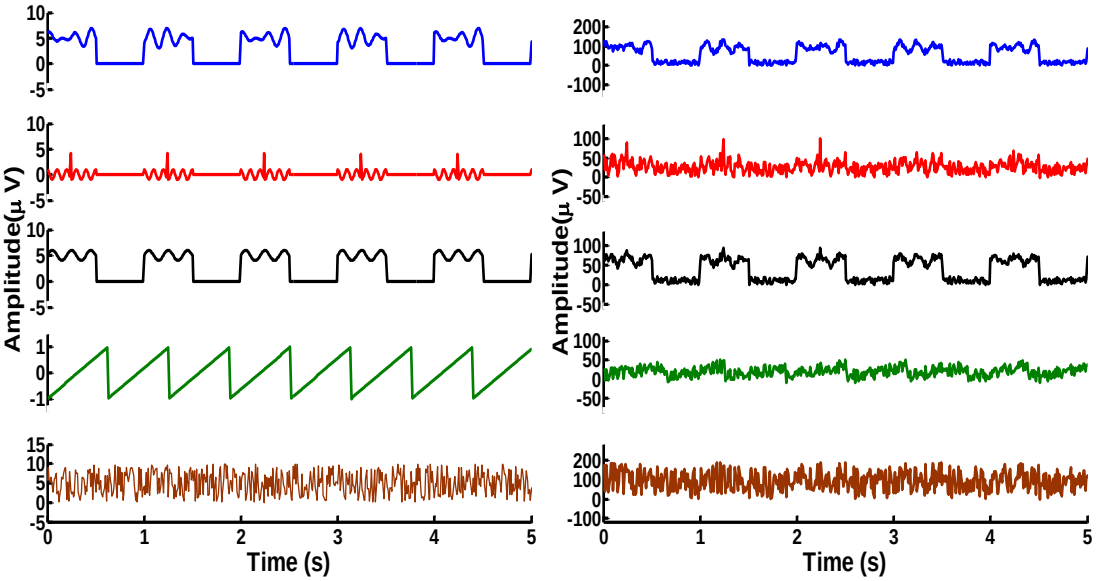

Figure 3.5: Left, four signals *P(n),Q(n), R(n)* and *S(n),* given by equations (3.9) to (3.12) plus noise. Notice that the *P(n)* is periodic with 200 samples, *Q(n)* and *R(n)* are periodic with 100 samples and the *S(n)* is periodic with 127. Right, linear mixture of the signals using the mixing matrix ***A***

*FastIC*A was employed to find the main components (ICs) of the matrix ***X***. According to the event related activity model presented in Figure 3.1, the response is expected to occur over the interval between two successive stimuli.

The SNR of each IC is calculated by using *Fmp* which was explained in section 1.7.2. To calculate SNR via *Fmp* for each IC, first the IC is segmented into trials (epochs) with length of 100 samples (period of the stimulus). Then, by selection on 5 fixed points, i.e. for this simulation the 10^{th}, 30^{th}, 50^{th}, 70^{th} and 90^{th} sample, in the trials and computing variation of the signal across the trials for these points and then calculating the mean over the variations, *Fmp* is calculated for each IC. Independent components whose *Fmp* values are higher than a critical threshold are deemed to contain the response and should be kept. If the method works perfectly, only *Q(n)* and *R(n)* will be used in reconstruction. The data was reconstructed by only using the response components and the result is shown in Figure 3.6 (left). Moreover, Figure 3.6 (right) shows the mixture of the response signals, i.e. only mixing the *Q(n)* and the *R(n)*. In another words, the Figure 3.6 (right) shows the mixture of the signals without any undesired factors (noise or the activities which are not following the stimulus). Considering that the SNR calculation is constrained to intervals between each two successive stimuli, the signals that are periodic with a different period will

have small SNR and cannot be considered as response components and will be rejected as an undesired component.

For this simulation it was already known that just two of the signals follow the stimulus, and thus component rejection was easy, i.e. the first two signals with the largest *Fmp* values were selected as response components. This issue is not simple for real-world recorded cases such as the EEG in which there is no certain information about the exact number of underlying sources. Having a priori information about the signal can be helpful to estimate this threshold. One method for finding this threshold is explained in the next chapter.

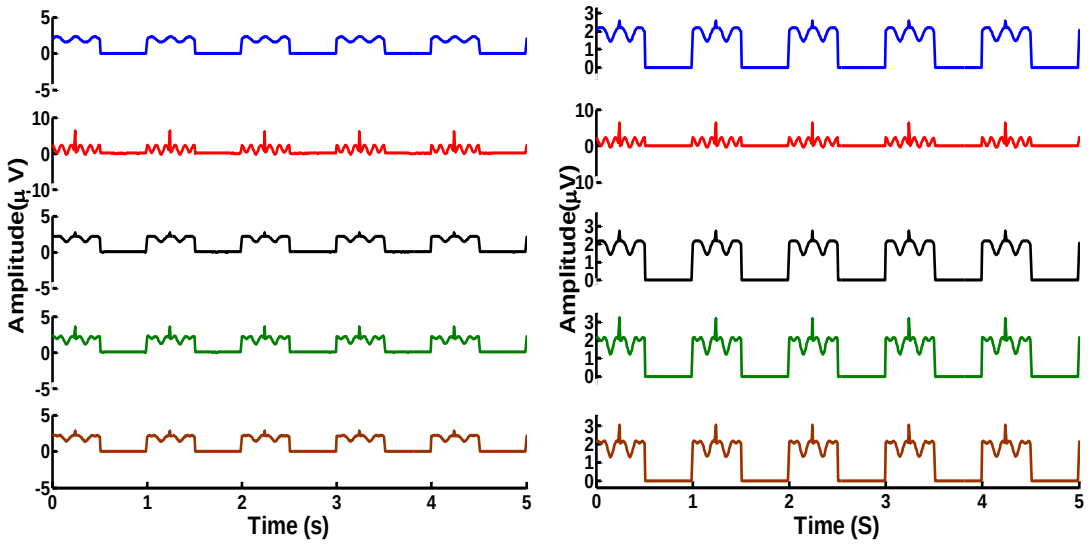

Figure 3.6: Left, reconstructed data only using signal components (the noise components are discarded). Right, mixing only the response signals, the signals which are synchronised with the stimulus, ideal noise reduction

The SNRs were calculated for the reconstructed data. These values were found 0.004, 53.153 162.28 and 4.25 respectively (from the top). In the next subsection, an alternative method of component selection for event related activities based on coherence of the signals and the stimulus is proposed.

3.2.4. Component Selection Using Magnitude Squared Coherence (MSC) of ICs (MSC-ICs)

This method is also designed for noise reduction for event related activity cases. The key idea in this method is to use the coherence between the stimulus and the IC as a criterion for the presence of a response in the IC. If an IC is significantly coherent with the stimulus, the IC is selected as a response component. Otherwise it

will be rejected as a noise component. Magnitude squared coherence (MSC), given by equation (1.1), is used as a measure of coherence between the ICs and the stimulus. For two signals, ***s*** (the stimulus) and ***d*** (IC), the equation (1.1) can be rewritten as:

$$\gamma_{sd}^2(f) = \frac{|S_{sd}(f)|^2}{S_{ss}(f)S_{dd}(f)} = \frac{|\sum_{i=1}^{M} D_i(f)S_i^*(f)|^2}{\sum_{i=1}^{M}|D_i(f)|^2 \sum_{i=1}^{M}|S_i(f)|^2} \tag{3.13}$$

Where, $S_{sd}(f)$ is cross power spectra between *s* and *d* and $S_{ss}(f)$ and $S_{dd}(f)$ are auto spectra of signals *s* and *d* respectively. In right hand side of equation (3.13), $D_i(f)$ and $S_i(f)$ are the samples of the Fourier transform of *s* and *d* and "*" equation denotes conjugate operator. Since the stimulus is constant, i.e. identical for all the epochs, (3.13) will be written as:

$$\gamma_{sd}^2(f) = \frac{|S^*(f)|^2|\sum_{i=1}^{M} D_i(f)|^2}{M|S(f)|^2 \sum_{i=1}^{M}|D_i(f)|^2} \tag{3.14}$$

Therefore for a constant stimulus MSC can be calculated as following.

$$\gamma_{sd}^2(f) = \frac{|\sum_{i=1}^{M} D_i(f)|^2}{M \sum_{i=1}^{M}|D_i(f)|^2} \tag{3.15}$$

It should be noted that the actual stimulus function *S(f)* is irrelevant, as it cancels in the equation (3.14). Equation (3.15) shows that when the stimulus is constant and the response is phase locked to the stimulus, the coherence can be calculated by segmenting signal according to the stimulus onsets and then the equation (3.14) can be simplified to equation (3.15). In order to apply a statistical test for the presence of a response in an IC, the null hypothesis is that the IC and the stimulus are not coherent. The null hypothesis is rejected if the p-value of the estimated MSC is smaller than a user-selected value (typically α=5%), and the IC is then considered as statistically significant. In statistical hypothesis testing, the p-value is the probability of obtaining a test statistic at least as extreme as the one that was actually observed, assuming that the null hypothesis is true. One often "rejects the null hypothesis" when the p-value is less than the significance level α, which is often 0.05 or 0.01. The p-value can be calculated for the MSC based on the F-statistic.

According to (Simpson et al. 2000) coherence estimates for zero coherence between signals ***s*** and ***d*** follow an F-distribution after the following transformation:

$$\frac{(M-1)\gamma_{sd}^2(f)}{1-\gamma_{sd}^2(f)} \sim F_{2,2(M-1)} \tag{3.16}$$

where *N* is the number of the epochs, the $\gamma_{sd}^2(f)$ is the magnitude squared coherence, '$\sim$' denotes 'is distributed as' and $F_{2,2(M-1)}$ is the F-distribution with 2 and 2(M-1) degrees of freedom. Setting an optimal threshold for the p-value is not a trivial issue: too low will remove too many ICs, some of which may include a weak response. Setting the significance value α too high will lead to the inclusion of too many ICs, including some dominated by noise. Both these extremes will reduce the final SNR of the estimated responses. The MSC method was employed for noise reduction in the same signals as used in the section 3.2.3 and the results are shown in Figure 3.7.

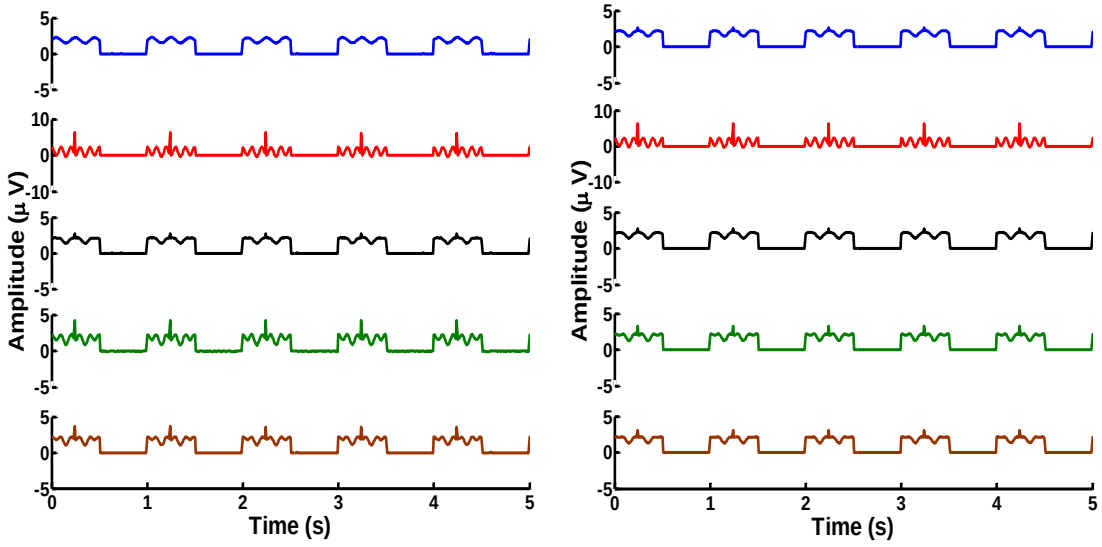

Figure 3.7: Left, de-noised data using the MSC-ICs for component selection. Right, ideal noise reduction (by only mixing *Q(n)* and *R(n)*).

For this example, the threshold for p-value was selected as 0.05. ICs which are significantly coherent with the input (p-value < 0.05) are kept as signal components and the rest are removed as noise components. As a measure of signal quality, SNR was calculated for each of the reconstructed signals shown on the left in Figure 3.7. These values were found 0.0046, 157.86, 283.57 and 5.621 (from the top).

The performance of the proposed methods for noise reduction and SNR improvement are compared by applying all four suggested methods on a same set of data. The results of this comparison are shown in Figure 3.8.

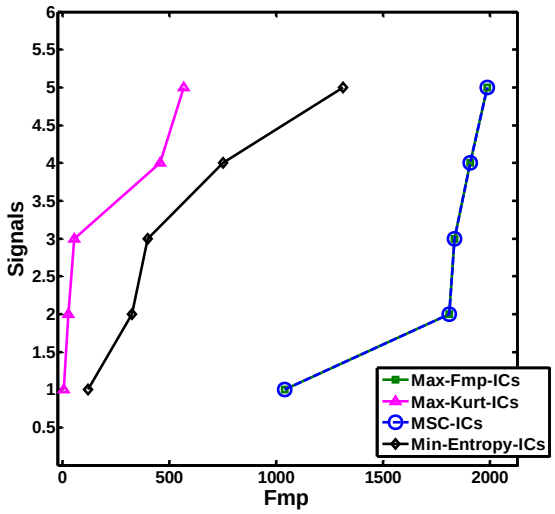

Figure 3.8: Comparing *Fmp*s (as a measure of SNR) obtained by the proposed methods. The Max-Fmp- ICs and the MSC-ICs produce signals with higher *Fmp*s (SNRs) than the Max-Kurt-ICs and Min-Entropy-ICs. For this example MCS-ICs and Max-Fmp-ICs have the same performance (the graphs are overlapping).

It can be stated from the results of Figure 3.8 that in this example the MSC-ICs and the Max-Fmp-ICs have identical performance in noise reduction (the graphs are overlapping). Moreover, it can be seen that for this set of data, the MSC-ICs and the Max-Fmp-ICs provide higher SNRs than the other two methods (Max-Kurt-ICs and Min-Entropy-ICs).

3.2.5. PCA versus ICA for Source Separation

As discussed in section 1.4, when the signal of interest is highly contaminated by noise and the power of signal is small compared to the noise power, PCA cannot entirely separate the signal components from the noise components. Signal sets given by equations (3.17) to (3.18) and three uniformly distributed random signals (noise) were used to demonstrate the drawbacks of PCA based noise reduction methods.

$$s_1(t) = \begin{cases} sin(0.5\pi t) & 0 < t < 100 \\ 0 & otherwise \end{cases}$$
$$S_1(t) = \sum_{k=0}^{9} s_1(t).\delta(t - 100k) \qquad (3.17)$$

$$s_2(t) = \begin{cases} sin(0.3\pi t) & 0 < t < 100 \\ 0 & otherwise \end{cases} \qquad (3.18)$$

$$S_2(t) = \sum_{k=0}^{9} s_1(t).\delta(t-100k)$$

$$s_3(t) = \begin{cases} sin(0.2\pi t) & 0 < t < 100 \\ 0 & otherwise \end{cases} \qquad (3.19)$$

$$S_3(t) = \sum_{k=0}^{9} s_1(t).\delta(t-100k)$$

$$s_4(t) = \begin{cases} sin(\pi t) & 0 < t < 100 \\ 0 & otherwise \end{cases} \qquad (3.20)$$

$$S_4(t) = \sum_{k=0}^{9} s_1(t).\delta(t-100k)$$

The signal set and three noise signals were mixed using mixing matrix $\boldsymbol{A}$ given by equation (3.21) and provide the observation signals.

$$\boldsymbol{A} = \begin{pmatrix} 15 & 0.2 & 1 & 0.6 & 3 & 1 & 0.5 \\ 0.3 & 12 & 2 & 2 & 4 & 1 & 1 \\ 0.7 & 3 & 10 & 1.5 & 2 & 0.4 & 2 \\ 0.2 & 0.5 & 0.75 & 13 & 3.5 & 1.5 & 2 \\ 1 & 1.5 & 2.9 & 3 & 17 & 0.3 & 2.4 \\ 2 & 2.5 & 4.9 & 2 & 0.3 & 25 & 2.4 \\ 4 & 1.1 & 1.9 & 3 & 5 & 0.9 & 15 \end{pmatrix} \qquad (3.21)$$

Matrix $\boldsymbol{A}$ was selected in a similar way as the corresponding matrix $\boldsymbol{A}$ in the previous example to produce signals with one dominant component dominant at each time for each observation. In this simulation, for PCA components with highest variances were selected as the components of interest and MSC-ICs were used for selecting the ICs for ICA. The results of using PCA in noise reduction algorithm was compared with ICA based alternative. Moreover, un-weighted averaging was used as a reference for comparing the performances of the two methods in noise reduction. This comparison was carried out for when the level of noise was increased by the factor of 100 in three stages, i.e. for this example noise variance was changed from 0.483 to 48.3. Initially the noise variance was selected equal to signal variance and then it was increased to 100 times larger than the signal's variance in three stages. The performance of PCA for separating the main components (signal components from noise components) was compared with ICA by comparing the components

found through each of the method form the signal set given by equation (3.17) to (3.18) and the noise signals. The signal set along with the noise signals and also the mixture of the signals (mixing by matrix ***A***) for noise variance equal to the signal's variance (0.438) are shown in Figure 3.9.

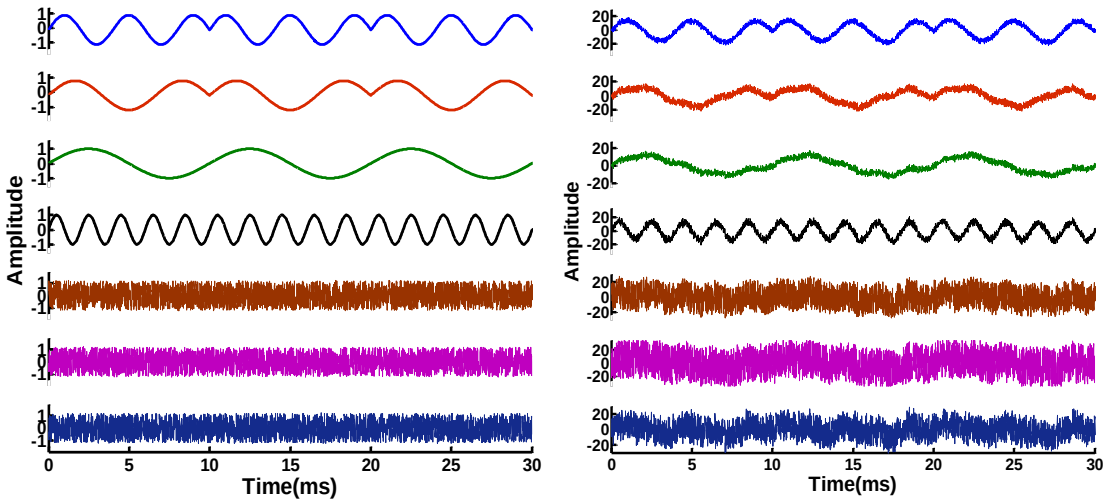

Figure 3.9: Left, Signal set given by equations (3.17) to (3.20) along with the noise signals. Noise signals have equal variances (0.483). Right, signals and noise are mixed by using mixing matrix ***A***.

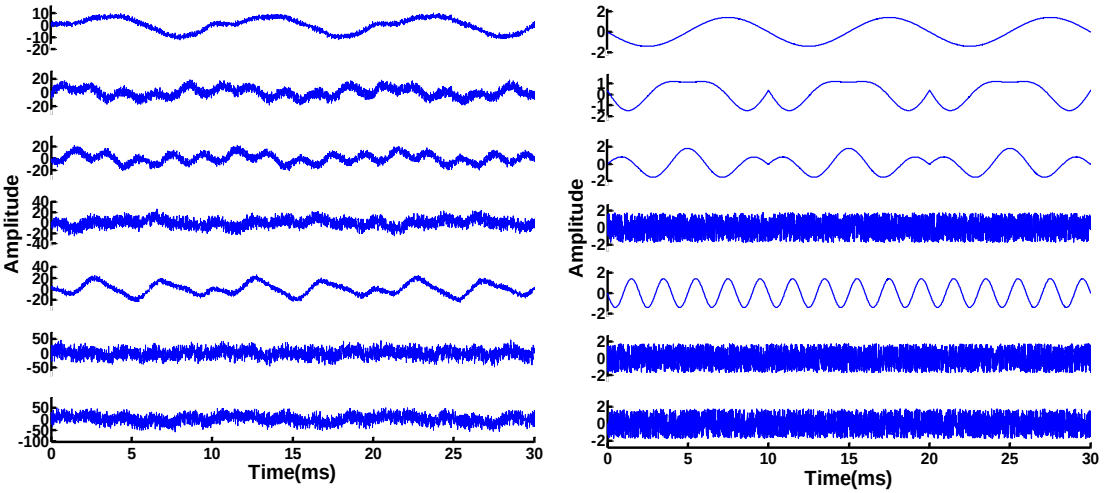

Figure 3.10: Left, main components found from the mixture by PCA. Right, Main component found form the mixture by ICA.

Since the signal and noise have equal powers, as can be seen from the left of Figure 3.10, PCA was not able to find the main components precisely; while, the main components found by ICA were more similar to the actual signals. In this example, PCA fails to find the main components when variance of noise is in the range of signals variance. This result is in accordance with the fact that PCA decompose the data into the components whose variances are maximally large, i.e.

since the noise and the signal have the same variance, the PCA algorithm select noise as a principal component. The time domain de-noised signals were reconstructed by only using the components 1, 2, 3 and 5 shown on the left of Figure 3.9. The ensemble average of the reconstructed signals and also the raw data (Figure 3.9, right) is shown in Figure 3.11. It can be seen from Figure 3.11 that using ICA in noise reduction algorithms has a considerably better performance than the PCA alternative and un-weighted averaging, i.e. signal looks less noisy. These results were confirmed by calculating the SNR of each de-noised signal (using *Fmp*) produced by each of the methods. These SNRs are shown in Table 3.1.

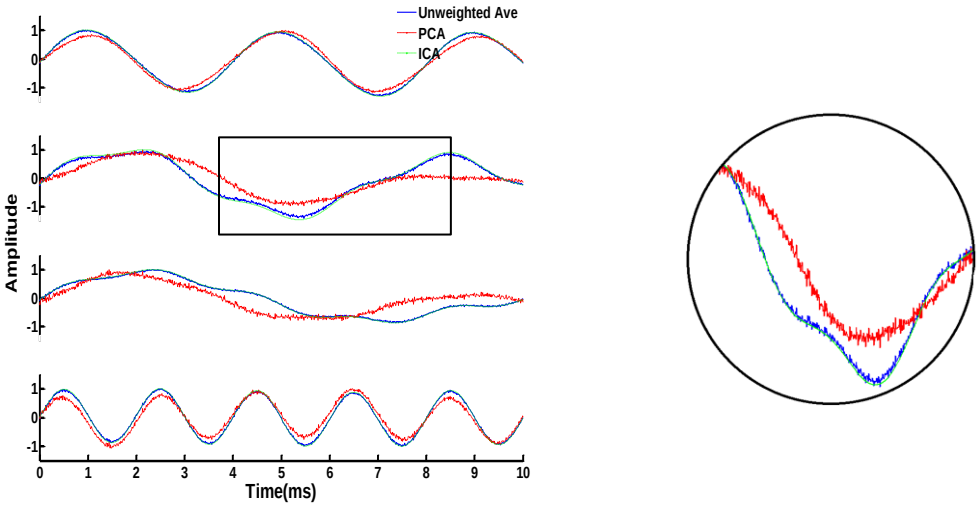

Figure 3.11: SNR improvement using PCA, ICA and un-weighted averaging. Noise variance is equal to signal variance.

Methods	Signal to Noise Ratio(*Fmp*)			
	Sig.1	Sig.2	Sig.3	Sig.4
Un-weighted Ave	2.27×10^3	0.82×10^3	1.45×10^3	10^3
PCA	0.052×10^3	1.77×10^3	0.80×10^3	10^3
ICA	3.25×10^6	1.30×10^7	5.53×10^6	6.21×10^7

Table 3.1: SNR improvement comparison between PCA, ICA and un-weighted averaging when noise variance is equal to signal variance.

This procedure was repeated when the noise level is ten times larger than the signal level and the result is shown in Figure 3.12.

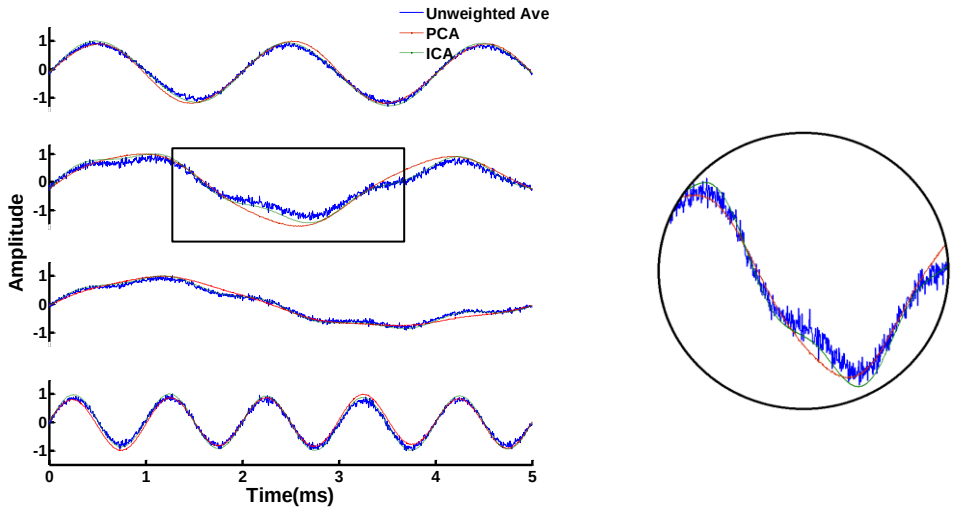

Figure 3.12: SNR improvement comparison between PCA, ICA and un-weighted averaging. The noise variance is now ten times larger than signal variance.

Figure 3.12 illustrates that PCA is able to reduce the noise considerably better than simple un-weighted averaging. Moreover, Table 3.2 shows the SNRs of the noise reduced signals produced by PCA, ICA and un-weighted averaging for the case in which noise variance is ten times larger than signal variance.

Methods	Signal to Noise Ratio(*Fmp*)			
	Sig.1	Sig.2	Sig.3	Sig.4
Un-weighted Ave	218.81	78.36	134.46	92.33
PCA	1.47×10^4	1.39×10^4	8.21×10^3	1.06×10^4
ICA	8.32×10^6	5.89×10^6	6.11×10^5	6.99×10^6

Table 3.2: SNR improvement comparison between PCA, ICA and un-weighted averaging when noise variance is ten times larger than signal variance.

Finally the noise level was selected to be 100 times greater than the signal level. It is shown in Figure 3.13 that PCA had a better performance than un-weighted averaging for noise reduction. However, Table 3.3 shows that ICA still has a better performance than PCA in terms of SNR improvement.

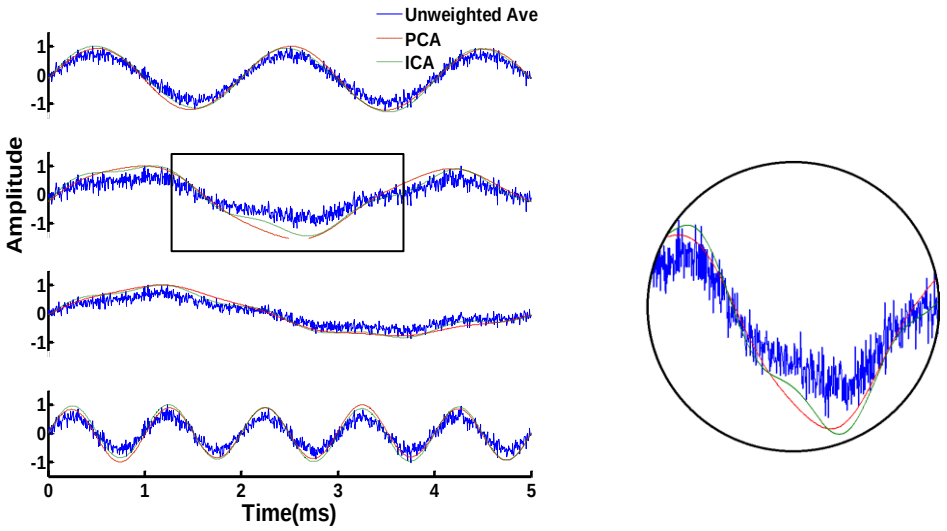

Figure 3.13: SNR improvement using PCA, ICA and un-weighted averaging. Noise variance is ten times larger than signal's variance.

Methods	Signal to Noise Ratio(*Fmp*)			
	Sig.1	Sig.2	Sig.3	Sig.4
Un-weighted Ave	218.81	78.36	134.46	92.33
PCA	1.47×10^4	1.39×10^4	8.21×10^3	1.06×10^4
ICA	8.32×10^6	5.89×10^6	6.11×10^5	6.99×10^6

Table 3.3: SNR improvement comparison between PCA, ICA and simple averaging when noise variance is hundred times larger than signal variance.

Form the results presented by Figures 3.10 to 3.13 and also Tables 3.1 to 3.3, for this example, it can be concluded that the component separation by PCA is susceptible by noise level. Also it was shown that ICA based noise reduction has better performance in noise reduction compared with PCA and un-weighted averaging.

When PCA is employed for source separation in noise reduction algorithms, if the noise variance is smaller than the signal variance of the components with larger variances should be kept as signal components; while for the cases in which noise variance is larger than signal's variance components with large variances should be discarded. This dependency on noise level for component selection is problematic for cases in which there is no information available about the noise level. This fact suggests that MSC may be advantageous for automatic component selection for PCA based noise reduction algorithms in a similar fashion as ICA based noise reduction methods. The ensemble average for noise components found by PCA and ICA when

the noise variance was ten times larger than the signal variance are shown in Figure 3.14.

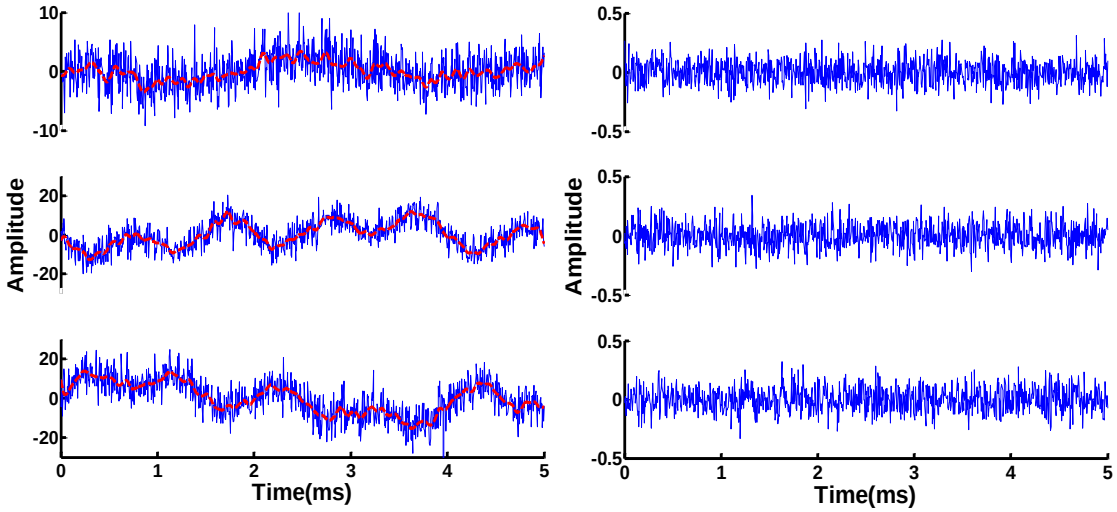

Figure 3.14: Ensemble average over noise components found by PCA (left) and ICA (right). Signal pattern can be seen though the noise found by PCA.

In Figure 3.14(left) the signal pattern is clearly visible in the noise components found by PCA; while the noise components found by ICA do not seem to contain any signal. For this example, it can be deduced from the result that PCA is not able to separate the signal components from noise components entirely. Due to the presence of the signal in the noise, all the components even the noise components are found to be significantly coherent. Consequently, the noise components will be kept as signal components and the reconstruction will be noisy. Although, this example shows that it is unlikely that MSC will be helpful in automatic component selection for PCA based noise reduction algorithms, this issue needs more investigation. Hence, in the next chapter MSC is used for component selection for PCA and ICA when real data is analysed. It is worth noting that, this pattern is also visible when the noise level is hundred times larger than the signal level but noise variance ten times larger than signal's variance has been selected here for simplicity.

3.3. Discussion and Conclusion

For the four noise reduction methods presented for selecting ICA components, it was demonstrated that all the methods are able to reconstruct the response form the noisy signal. Moreover, the results of applying all the four methods on the same set of data show that for this example, the MSC-ICs and the Max-Fmp-ICA have better

performance in terms of SNR improvement than the other two methods. However, in order to show that this improvement is significant, methods should be tested on real data followed by statistical analysis. Statistical comparison of the proposed methods is carried out and explained in detail in the next chapter.

For using the MSC-ICs and the Max-Fmp-ICs, extra information about the signal is needed, including crucially stimulus onsets; while the Max-Kurt-ICs and the Min-Entropy-ICs do not need this prior information. However, the Max-Kurt-ICs is sensitive to the outliers and it is not a suitable choice for the cases in which the undesired signal has large positive kurtosis (super Gaussian distribution). Setting a suitable threshold for rejecting or accepting an IC is another difficulty of the introduced methods.

In biomedical data the variance of the noise is much higher than the variance of the signal of interest, it can be predicted that PCA might be able to enhance the signal quality, i.e. in terms of SNR improvement. However, according to the results of the simulation presented in this chapter, ICA may have a better performance in SNR improvement. Moreover, due to the signal residue in noise components, MSC may not be a suitable method for component selection for PCA based noise reduction methods.

3.4. Summary

In this chapter four novel methods for automatic component selection for ICA based noise reduction algorithms were proposed and tested by using simulated data. Moreover, the performances of the suggested methods were compared in terms of signal to noise ratio improvement. The results showed that the MSC-ICs and Max-Fpm-ICs seem to have a better performance than the other two methods.

Furthermore, the idea of employing PCA instead of ICA in noise reduction algorithms was explored. The results of simulation confirmed that PCA can also be used in multichannel noise reduction algorithms. However, ICA provides signals with higher qualities (larger SNRs). Moreover, it was shown that MSC cannot be beneficial for component selection when PCA is employed as component separation methods in multichannel noise reduction algorithms. However, comparison between the methods should be carried out through statistical analysis over more set of data

and real data. In the next chapter these methods will be statistically compared with each other and also the alternatives.

Chapter 4

A Comparison of Different Noise Reduction Approaches on Recorded Data

4.1. Introduction

The main goals of this chapter are: 1. to quantitatively compare the noise reduction methods proposed in chapter 3 (ICA and PCA based methods) on data recorded on patients in order to assess which of the methods is the best choice for noise reduction (SNR improvement) in multichannel ALR recordings. 2. to explore whether, and to what extent, using multichannel signal processing can be beneficial for SNR improvement in ALR recording in comparison with single channel processing.

In this chapter the four proposed ICA component selection methods described in chapter 3 (Max-Kurt-ICs, Min-Entropy-ICs, Max-Fmp-ICs and MSC-ICs) are used for signal quality enhancement for 63 channel ALRs recorded from 10 normal hearing subjects. The performance of the methods are compared with each other and also with single channel processing alternatives (un-weighted and weighted averaging which are explained in chapter one) in terms of SNR improvement. This comparison

is carried out with regard to two aspects: the highest SNR produced by each method and also the number of the channels that each method provides with an SNR above a predefined critical value. The reason for being interested in having more channels with high SNR is that, in some clinical applications the signal of interest should be extracted from some specific regions of the scalp, e.g. frontal or lateral scalp regions. Therefore, a method which can give a consistently high SNR across several channels may be useful. Furthermore, the performance of the best proposed method is compared with two PCA based multichannel noise reduction alternatives. In the first instance the principal components (PCs) are ordered by eigenvalues of the covariance matrix (Hyvärinen and Oja 2000) and in the second the MSC is used for component selection. At the end statistical comparison between all the alternative noise reduction methods is carried out and the best method in terms of SNR improvement is selected and will be applied to different clinical ALR applications in chapter 5.

4.2. Noise Reduction Methods

4.2.1. Single Channel Processing Methods

In chapter one, single channel recording and the effect of un-weighted coherent averaging was explained. It was highlighted that, noise reduction strategies are depending on the protocol under which the AERs are recorded. The ALR is either recorded by single or multichannel strategies. There exist various noise reduction methods when the ALR is recorded using single channel protocol. Conventional un-weighted coherent averaging and weighted averaging are two methods which are used frequently in clinical applications (Ross et al. 1984; Elberling and Wahlgreen 1985; Davila and Mobin 1992; Dimitrijevic et al. 2001).

For this work, the ALR was recorded using the testing protocol explained in section 2.5 with 155 stimulus repetitions. Recall that, the ALR occurs over 250 ms after stimulus onset. Hence, although inter stimulus interval is 1499 samples we are only interested in 250 samples after each stimulus onset, i.e. 250 samples is equivalent to 250 ms when sampling rate is 1 kHz. Therefore, the recorded signal is segmented in to 155 epochs each of length 250 samples. It is worth mentioning that, this segmentation is used throughout this work. Hence, whenever the term epoch is used, it means a segment of 250 samples after the stimulus onset; unless it is

specifically stated otherwise. Un-weighted averaging was simply implemented by using the equation (1.5). In other words, the epochs were added to each other and divided by 155. For implementing the weighted averaging given by equation (1.10) in this work, the weights were calculated by taking every five epochs as a sweep and calculating the variation of the signal over each sweep by using *Fmp*. In this implementation, for calculating the variation of the signal over each sweep, sample points (10, 30, 50, 70, 90, 110, 130, 140, 160, 180, 200, 220 and 240) were selected in each epochs; and variation of each sample point across the five epochs were calculated. Mean value of signal variation over the aforementioned points was calculated as the weight of each sweep. Then, the epochs were weighted by dividing each epoch by the weight of the sweep that the epoch belonged to. The weighted average ALR was obtained by adding the weighted epochs and dividing the results by addition of all the weights, as in (1.10).

4.2.2. Multichannel Processing Methods

In chapter 3, the general approach for multichannel BSS based noise reduction methods was described as consisting of three steps:

1. Decomposing the recorded signals into its main components, here defined to be independent components (ICs) or principal components (PCs).
2. Classify the main components into two main groups: response components and noise components.
3. Reconstruction the data using only the response components. Stage 2 is often the most challenging, and frequently carried out based on visual inspection and the spatial distribution of components.

In this section, the implementation of the four component selection methods which were introduced in the third chapter are explained. Similar to the single channel processing methods, as a pre-processing step, the data was segmented, i.e. 250 samples after each stimulus onset were kept. Afterwards, *FastICA* was applied on the segmented data to find the ICs. In the Max-Kurt-ICs method, to automate the component selection, kurtosis of each IC was calculated using equation (1.23). The expectation operator in equation (1.23) was estimated by averaging over data. In the next step, the critical value below which the ICs should be rejected as noise components was found from the 4 min no-stimulus EEG (resting EEG) and a

Bootstrap method. Bootstrap was used to estimate distribution of noise kurtosis. For this estimation, kurtosis was calculated for 155 epochs which were randomly selected from the no-stimulus EEG recorded from the vertex and this procedure was repeated 200 times. Then, 95^{th} percentile of the noise kurtosis distribution gives the value above which the signal deemed to be significantly different from noise. Since the heart beat component has also a large kurtosis, in order to not to include the heart beat component in signal reconstruction, the component with largest kurtosis should also be removed. In other words, if an IC has kurtosis above the threshold value and less than the largest kurtosis value, the IC will be kept as the response components and otherwise will be discarded as noise component.

In Min-Entropy-ICs method, entropy of each IC was calculated by using the equation (1.29). In practice, in order to use the equation (1.29), the pdf was approximated by histogram and 256 bins (implemented in a MATLAB function). Setting the threshold for component selection, was carried out in a similar fashion with the previous section but slightly different. Distribution of noise entropy was estimated by bootstrapping; and lower 95^{th} percentile of the distribution was selected as the threshold. ICs with entropy values above this value should be discarded as noise components.

For component selection by Max-Fmp-ICs, *Fmp* of each IC was calculated by selecting 13 points (10, 30, 50, 70, 90, 110, 130, 140, 160, 180, 200, 220 and 240) in each epoch and calculating the mean value of variation of the signal over these points across 155 epochs. This value was compared with the threshold value obtained from rest EEG and bootstrapping. Here, the 95^{th} percentile was selected as the critical value and ICs with *Fmp* above the threshold were selected as response components and the rest were discarded as noise components.

MSC-ICs was also used for selecting response components. MSC was calculated for each IC by following equation (3.15). Since, the ALR waveform is similar to a sinewave over epochs with length 250 samples, highest signal power is expected to be observed at the first harmonic. For the null hypothesis of zero coherence between stimulus and EEG, distribution of (3.15) is given by equation (3.16). For each IC, p-value was calculated from the MSC distribution. An IC was selected to be a response component, if the IC was significantly (p-value$<$ 0.05) coherent with the stimulus at the first harmonic. After selecting the components, for

each of the methods, the data was reconstructed using the mixing matrix (calculated by FastICA) and also assigning the undesired components to zero (the rows of the IC matrix corresponding to undesired components were assigned to zero) and after reconstructing the data, an un-weighted average was taken over 155 epochs.

4.3. Comparison of Methods

SNR as a measure of signal quality was calculated using *Fmp* which produces an estimate of the SNR. To compare the performance of multichannel noise reduction with single channel alternatives, the highest SNR that each method can provide in each subject were compared. Moreover, the number of channels with SNR higher than a critical value for each subject is another criterion for comparing the performances of the methods. In the current work, the threshold *Fmp* value (section 1.7.2) above which a response was deemed present was obtained from the 95^{th} percentile of the *Fmp* distribution of the 63 channels recorded at rest (with no auditory stimulation) by the bootstrap method. The *Fmp* was calculated for each method in an identical manner to the data with stimulation, i.e. the noise reduction methods were applied on resting EEG and then *Fmp*s were calculated. Since there is no stimulus present in the resting EEG signal, there will be no response present, and the resulting distribution represents the null-hypothesis. The distributions of the *Fmp*s were obtained for all the noise reduction methods by the *Bootstrap* method (Simpson et al. 2007; Chernick 2008), by repeatedly (200 times) choosing 155 epochs from the recorded resting EEG at Cz (situated at the vertex), with random starting points. The procedure of finding *Fmp* distribution for noise reduced data was carried out for all ten subjects and the mean value of the thresholds found in each method was selected as the threshold for the corresponding method. Signals with *Fmp* higher than the threshold are thus considered as high quality signals and the channels with SNR lower than this threshold are considered as channels with poor quality signals, and either the response is not present or not detectable. The significance of the noise reduction methods is checked by applying a parametric or non-parametric statistical tests, i.e. depending on the distribution of the data; results with p-value <0.05 are deemed statistically significant.

4.4. Statistical Analysis and Results

The result of applying un-weighted and weighted averaging on the ALR recorded form the Cz channel for one of the subjects is shown in the Figure 4.1.

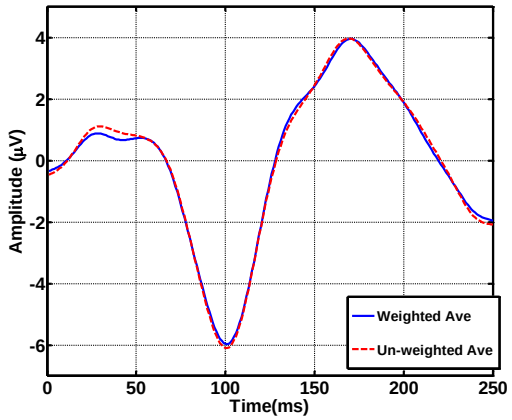

Figure 4.1: The ALR waveform obtained by un-weighted (dashed line) and weighted averaging (solid line).

Fmp threshold found calculated by only using Cz channel was found to be similar with when all 63 channels were used. For un-weighted averaging, mean value of *Fmp* threshold across all ten subjects found to be 1.25 (standard deviation 0.02) when only Cz was used and 1.24 (standard deviation 0.03) when all 63 channels were used. The distribution of the *Fmp*s and the critical values obtained from Cz are shown in Figure 4.2. Signals with *Fmp* above these thresholds are deemed to contain a response.

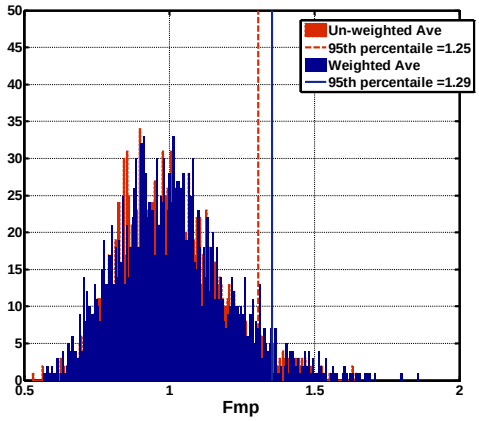

Figure 4.2: Distribution of *Fmp* for the resting EEG for one of the normal hearing participants over the Cz channel, estimated using the bootstrap method for no stimulation data. The dashed lines show the 95% cut offs for un-weighted and weighted averaging. Channels with an *Fmp* above these thresholds can be considered as showing a significant response to auditory stimulation.

The *Fmp* threshold was found at Cz for all ten subjects using the same procedure. The average value of the thresholds found by the un-weighted and weighted averaging were 1.25 and 1.29 with a standard deviation of 0.013 and 0.012 respectively. Therefore, 1.25 and 1.29 are selected as the *Fmp* threshold for these methods. The outcomes of the methods for un-weighted and weighted averaging in all 63 channels (for the same, typical, subject) are shown in Figure 4.3. For this subject the highest SNR found by the weighted averaging was 7.93 while this value found to be 7.19 for the un-weighted averaging. Moreover, numbers of channels that each method could find above their thresholds (1.25 and 1.29) were 43 and 46 for un-weighted and weighted coherent averaging respectively.

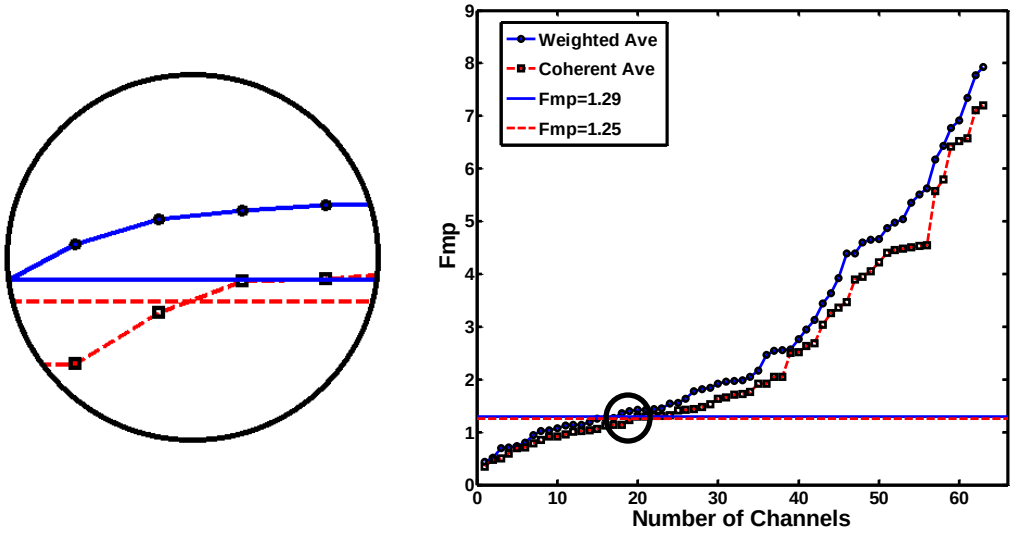

Figure 4.3: The cumulative histogram of the *Fmp* calculated in one subject for both the un-weighted and weighted averaging methods using stimulus. The vertical lines show the critical values of *Fmp*=1.25 and *Fmp*=1.29, for the Coherent averaging and weighted averaging respectively.

This procedure was carried out for all 10 subjects. The highest SNRs and also number of channels above the critical values (1.25 for coherent averaging and 1.29 for weighted averaging) were calculated for each method for each subject. Applying a Shapiro-Wilk test for normality showed that the distribution of the highest SNRs, i.e. a vector which contains maximum SNR in each subject, was not normal. Whereas, the numbers of channels above the predefined thresholds, i.e. a vector consisting of the number of channels above the critical value for each method in each subject, was normally distributed. Therefore, to compare highest SNRs found by the two methods

for the 10 subjects, a Wilcoxon signed rank test was employed. Moreover, a dependent paired t-test was employed for comparing the number of the channels that each method can find above its critical value as the data was normally distributed. The results show that the SNR was found to be significantly higher when weighted averaging was employed instead of coherent averaging for ALR single channel noise reduction (p-value<0.04). However, the number of the channels that each method can provide with SNR above its threshold was not significantly different between the two methods. For the data with a non-normal distribution a boxplot (Figure 4.4) shows the results of *Fmp* for the two approaches. In Figure 4.4, the median (middle line), 75^{th}, 25^{th} percentiles (the edges), maximum, minimum (whiskers) and outliers (stars) are shown.

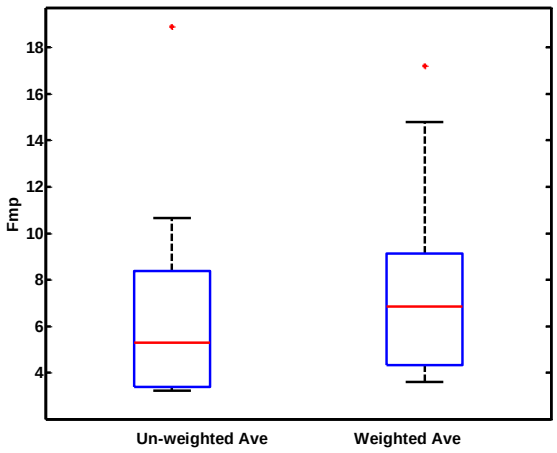

Figure 4.4: Comparing the *Fmp* of weighted averaging with un-weighted averaging for providing higher SNR over 10 normal hearing subjects. In each boxplot, the middle line shows the median, the edges of the boxes show 75^{th} and 25^{th} percentiles, whiskers show the maximum and minimum values and the stars show outliers.

The comparison between the weighted and un-weighted averaging regarding the number of the channels found above the critical values of *Fmp* is shown in Figure 4.5.

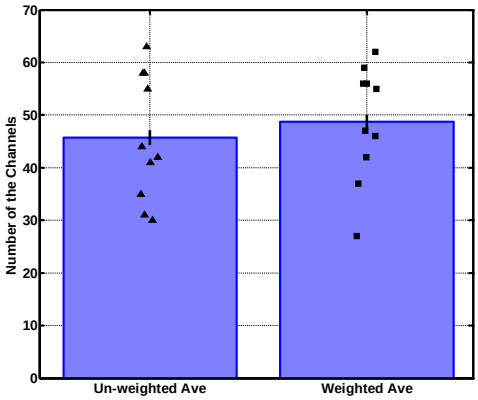

Figure 4.5: Comparing weighted averaging with un-weighted coherent averaging for the number of channels above the critical values of *Fmp*, for 10 normal hearing subjects. The mean value (edge), data scatter (markers) and standard error (error bar), are shown for each bar in the graph.

It can be seen in Figure 4.4 that, although weighted averaging significantly improves the SNR of the ALR in comparison with coherent averaging, this improvement is small, i.e. the median of the highest SNR in ten subjects were found 5.17 for coherent averaging and 6.15 for weighted averaging.

The critical value that ICs are kept/reject respect to, was found for each of the method. According to the result, ICs with absolute kurtosis above 2 should be kept as response. Alternatively, ICs whose entropy are larger than 4.5 are response components. Moreover, critical value for Max-Fmp-ICs was fond to be 1.32.

The 95^{th} percentiles (critical values) were also found for *Fmp* distribution for the four methods introduced in chapter 3, by employing the same aforementioned procedure used in weighted and un-weighted averaging for finding the critical *Fmp* values. These results are summarized in Table 4.1. The critical value for each method is calculated by averaging over the ten values obtained from resting EEG in each subject. The small variances given in the table implies that the critical values are found almost the same in each subject.

Methods	Average of Critical *Fmp*Values over 10 Subjects	Standard deviation of Critical Values over 10 Subjects
MSC-ICs	1.32	0.015
Max-Kurt-ICs	1.32	0.019
Max-Fmp-ICs	1.32	0.015
Min-Entropy-ICs	1.32	0.016
Un-weighted Ave	1.25	0.013
Weighted Ave	1.29	0.012

Table 4.1: Average and variance of *Fmp* thresholds found by each of the methods at the Cz for 10 participants. In each method channels with SNR above the given values are considered as high quality channels.

For each method, channels with SNR above the given value are deemed to have the response present. By applying all the methods on no stimulation EEG and calculating the *Fmp* in each of the channels, it can be shown that none of the methods bias *Fmp* values, i.e. as they all produce low *Fmp* values when there is no stimulus present and consequently there is no response present in the recorded signal. This fact is shown in Figure 4.6.

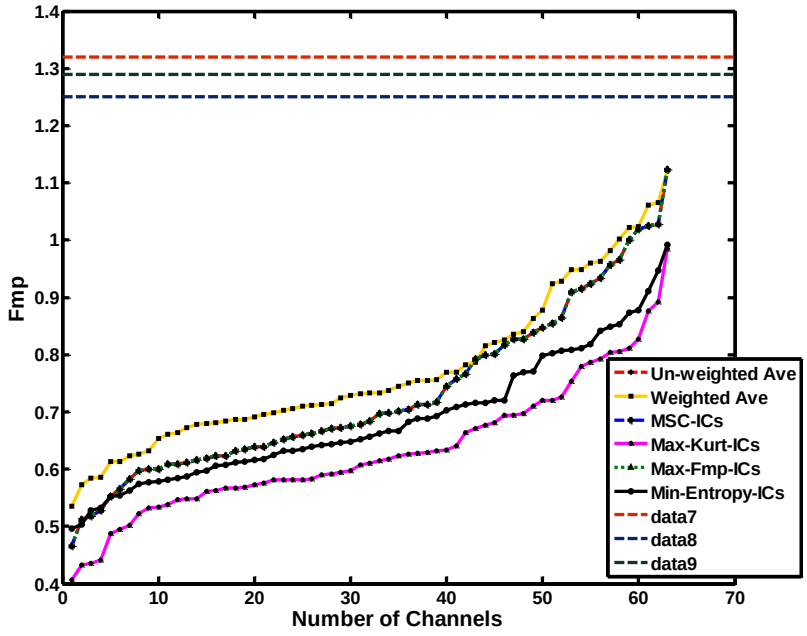

Figure 4.6: Applying the all the noise reduction methods on the no stimulation EEG shows none of the methods is biased to the method of *Fmp* calculation, i.e. all the methods provide signals with low *Fmp*s when there is no response present in the recorded signal. Since none of the ICs are rejected in either of the MSC-ICs and Max-Fmp-ICs method, their graph is averaging with un-weighted averaging graph.

Since the data reconstruction was carried out only if the number of ICs was greater than 4, for the case shown in Figure 4.6 all the SNRs are found below the critical values given in Table 4.1, i.e. there are no false positives. It is worth mentioning that, in MSC-ICs and Max-Fmp-ICs, since none of the components were kept as response components, data reconstruction was carried out by including all the ICs (no change in the original data). Therefore, their graphs are overlapping with un-weighted averaging's graph in Figure 4.6. It can be seen from Figure 4.6 that the graphs are slightly different. However, since these differences are small (even in comparison with what is considered as the minimum *Fmp* value above which the signal considered to have the response present), it can approximately be stated that the methods have similar behaviour in absence of the response.

The performance of the methods (single and multichannel noise reduction methods) for SNR improvement in one subject are compared and shown in Figure 4.7.

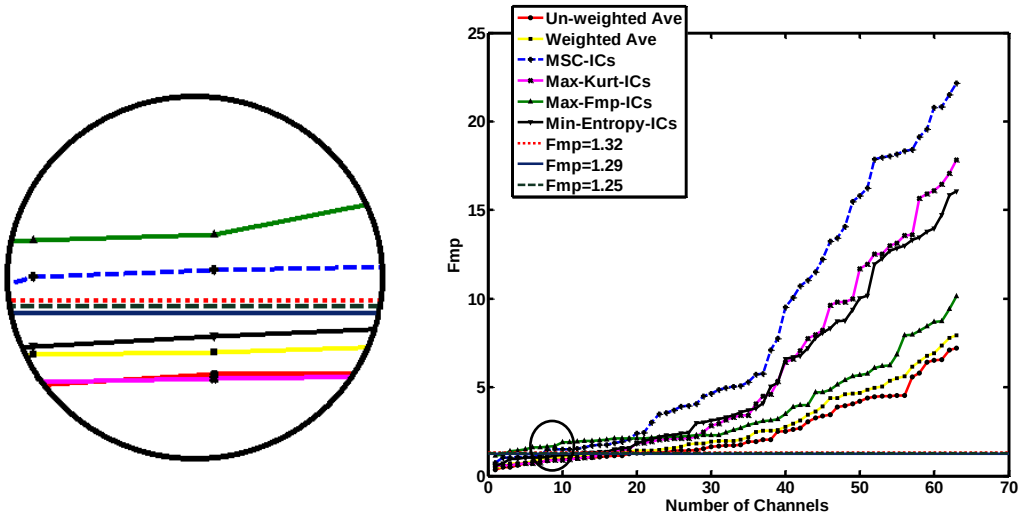

Figure 4.7: Performance of the methods in SNR improvement in 63 EEG channels from one subject using stimulus presented at 60 dB nHL. The dashed lines identify the *Fmp* thresholds value above which the channel can be considered to show a response.

In Figure 4.7, for this subject the highest *Fmp* found by the MSC-ICs was 22.17; and also 56 channels were found to have an *Fmp* higher than 1.32. For Max-Fmp-ICs these values were 10.13 and 61 respectively, i.e. 61 channels with *Fmp* above 1.32. For the Max-Kurt-ICs and the Min-Entropy-ICs methods the highest *Fmp* and the number of the channels with *Fmp* above the threshold were 17.37 and 16.27

for the highest SNR, with 48 and 50 for channels above 1.32. In order to investigate significant differences, these values were calculated for all ten subjects.

The result of comparing the performance of the methods for providing a higher SNR over the ten subjects are shown in Figure 4.8. The boxplot and median has been used for demonstrating the results as the data was not normally distributed. The hypothesis of the data being normally distributed was rejected following a Shapiro-Wilk test. Therefore, the significance of the methods effect was investigated by applying a Friedman test, which gave p-value<0.005. To examine where the differences actually occur, separate Wilcoxon signed-rank tests on the different combinations of methods (on every possible pair) is needed. The results of applying Wilcoxon signed-rank on pair methods are given in Table 4.2. p-values <0.05 shows a significant difference between each pair.

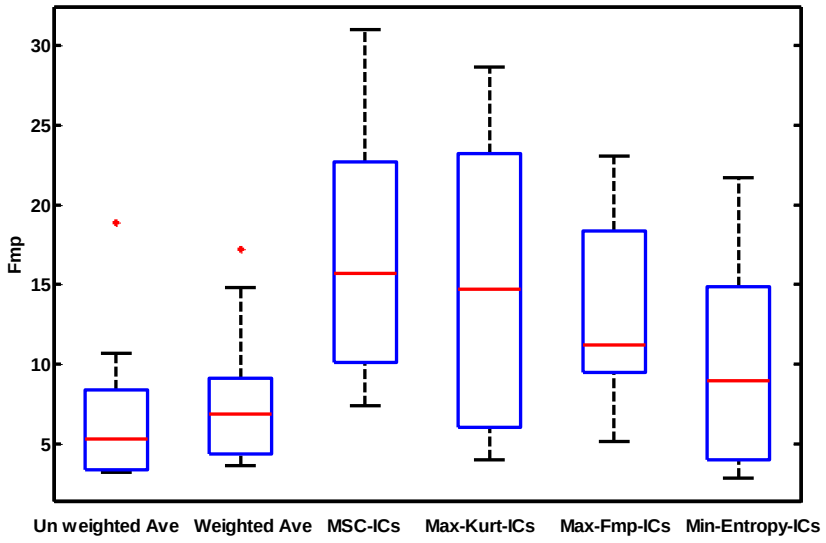

Figure 4.8: Comparison of the methods for providing higher SNR over 10 normal hearing subjects. In each boxplot, the middle line shows the median, the edges show 75th and 25th percentiles, whiskers show the maximum and minimum values and the stars show the outliers.

Methods	Un-weighted Ave	Weighted Ave	MSC-ICs	Max-Kurt-ICs	Max-Fmp-ICs	Min-Entropy-ICs
Un-weighted Ave	-	0.047	0.007	0.005	0.005	0.037
Weighted Ave	0.047	-	0.009	0.007	0.013	NS
MSC-ICs	0.007	0.009	-	0.022	NS	0.017
Max-Kurt-ICs	0.005	0.007	0.022	-	NS	NS
Max-Fmp-ICs	0.005	0.013	NS	NS	-	NS
Min-Entropy-ICs	0.037	NS	0.017	NS	NS	-

[*] NS: Not Significant (p-value>0.05)

[**]Numbers are p-values from a Wilcoxon signed-rank test

Table 4.2: Comparison of single and multichannel noise reduction methods. Results of applying a Wilcoxon signed-rank test on the maximum SNRs, obtained by the methods from the ten subjects a p-value <0.05 implies significant difference between the methods. Prior Friedman tests showed significant differences (p-value <0.005) across all methods.

It can be seen in Table 4.2, that all the four proposed multichannel noise reduction methods are significantly better than coherent averaging for producing higher SNR. However, the results did not show any significant difference between Min-Entropy-ICs and the weighted averaging. Moreover, comparing the multichannel strategies, MSC-ICs shows a better performance than Max-Kurt-ICs and Min-Entropy-ICs while it is not significantly different from Max-Fmp-ICs.

Additionally, this comparison has been carried out for the number of the channels that each method can provide above the critical *Fmp* values given in Table 4.1. Results are given in Figure 4.9.

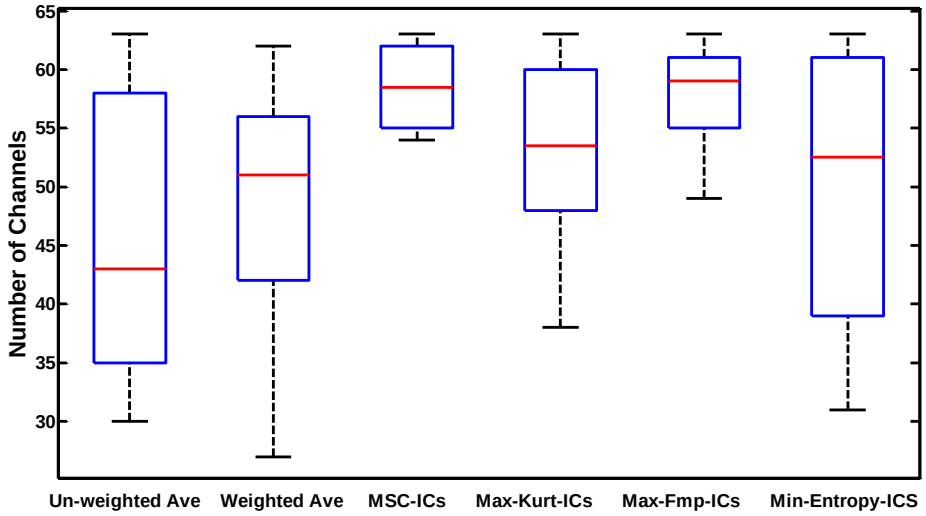

Figure 4.9: Number of channels with *Fmp* above the calculated *Fmp* thresholds given in Table 4.1 by each method for ten participants.

A Friedman test was applied on the obtained data, following a Shapiro-Wilks test indicating that the data was not normally distributed. The result showed a significant difference between the methods (p-value<0.033). To explore which methods differed, Wilcoxon signed-rank tests were applied across all the possible pairs and the results is summarized in Table 4.3.

According to the results given in Table 4.3, all the multichannel noise reduction methods except the Min-Entropy-ICs showed better performance than un-weighted averaging with respect to finding the number of channels above their predefined thresholds. Furthermore, only MSC-ICs and Max-Fmp-ICs were found to be significantly better than weighted averaging. To summarize, by using the information in Table 4.2 and 4.3 it can be stated that, between the ICA based multichannel noise reduction strategies, the MSC-ICs is better than the alternatives for noise reduction of ALR from the recorded EEG. The results of Table 4.3 show that the multichannel noise reduction methods can significantly improve signal quality (SNR) in comparison with when single channel alternatives are employed.

Methods	Un-weighted Ave	Weighted Ave	MSC-ICs	Max-Kurt-ICs	Max-Fmp-ICs	Min-Entropy-ICs
Un-weighted Ave	-	NS	0.012	0.024	0.017	NS
Weighted Ave	NS	-	0.020	NS	0.019	NS
MSC-ICs	0.012	0.020	-	NS	NS	0.034
Max-Kurt-ICs	0.024	NS	NS	-	NS	NS
Max-Fmp-ICs	0.017	0.019	NS	NS	-	0.035
Min-Entropy-ICs	NS	NS	0.017	NS	0.035	-

* NS: Not Significant (p-value>0.05)

** Numbers are p-values from a Wilcoxon signed-rank test

Table 4.3: Wilcoxon signed-rank test results for comparison of single and multichannel noise reduction methods for the number of channel that each method can find above its critical value. P-value <0.05 signifies a significant difference between the methods.

As an alternative to ICA, PCA based noise reduction methods can be employed for noise reduction for ALR. As discussed in section 1.6, the common problem that BSS based methods face, is how to select the components such that maximum signal quality (SNR) can be obtained. To automate the component selection for PCA based noise reduction methods, all the four proposed methods used above on ICA can be employed. However, since the results showed that, here, the MSC-ICs had the best performance between the ICA based component selection methods, only this method is employed for component selection in PCA based noise reduction methods. To be consistent to the terminology used in this work this method is called (MSC-PCs).

This differs from the usual PCA noise-reduction method, where eigenvalue ordering (Hyvärinen and Oja 2000) is usually used. PCA de-correlates the data and orders the components from the highest eigenvalue to the lowest one. For zero mean data, each eigenvalue represents the power (variance) of each component (Hesse 2008). In the case that the variance of the noise is larger than the variance of the signal of interest, components with larger variances are thus likely to be noise and should be discarded.

To compare the PCA based noise reduction methods with the aforementioned noise reduction alternatives, the 95^{th} percentile of the *Fmp* distribution was found using the same procedure used above. For the eigenvalue ordering case this value was

found 1.30 (average over the ten subjects), with a variance of 0.012 and for MSC-PCs this value was found 1.34 with a variance of 0.021. The performance of the PCA based methods for finding the highest SNR over ten subjects is compared with the weighted averaging. A Shapiro-Wilk test indicated that the data was not normally distributed. For non-normal data a box-plot is preferable. Therefore, boxplot has been used to illustrate the results in Figure 4.10. Result of applying a Friedman test on the highest SNRs obtained by the methods shows a significant different between the methods (p-value<0.001). To explore where exactly this difference occurs, a Wilcoxon signed-rank test was applied on each pair of methods. The results show that PCA-Eigenvalue ordering is significantly better both un-weighted (p-value<0.009) and MSC-PCs (p-value<0.007); while it was not found to be significantly different with weighted averaging. Moreover, using MSC for source selection in PCA based noise reduction methods was not found to be significantly different from un-weighted averaging. This result is in accordance with the simulation results in section 3.2.5, where it was found that when PCA was employed for source separation, signal and noise were not entirely separated; and due to the signal residue, all the ICs (including noise components) were found to be significantly coherent with the stimulus.

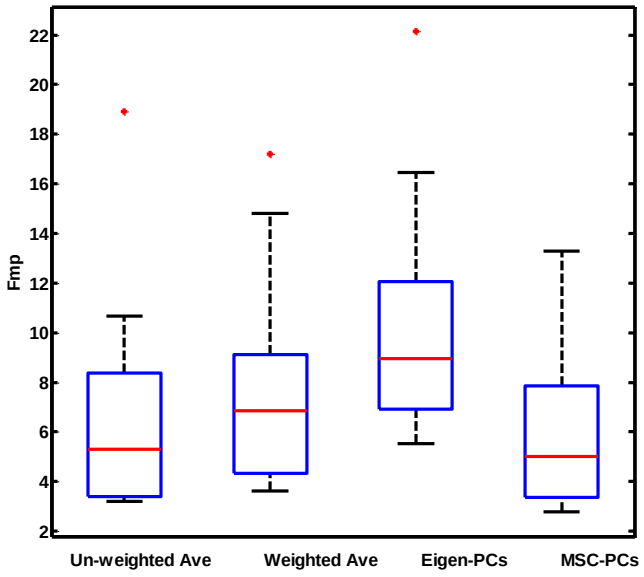

Figure 4.10: Comparison of weighted and un-weighted averaging with PCA based (Eigenvalue ordering and MSC-PCs) for highest SNR provided by each method.

Furthermore, the methods are compared regarding the number of channels that each method provides above that critical *Fmp* value and the result is shown in

Figure 4.11. Since the data (number of the channels found above the critical *Fmp* value by each method across the subjects) were normally distributed a repeated measure ANOVA was employed to explore if there is any significant effect of method. The result showed significant effects between the methods (p-value< 0.006). Additionally, paired-sample t-test were also applied on each pair of methods to understand between which of the pairs the difference were evident. The results showed that PCA-Eigenvalue ordering had significantly better performance than un-weighted averaging (p-value< 0.006), weighted averaging (p-value<0.02) and MSC-PCs (p-value<0.01) for providing more channels above the critical *Fmp* value.

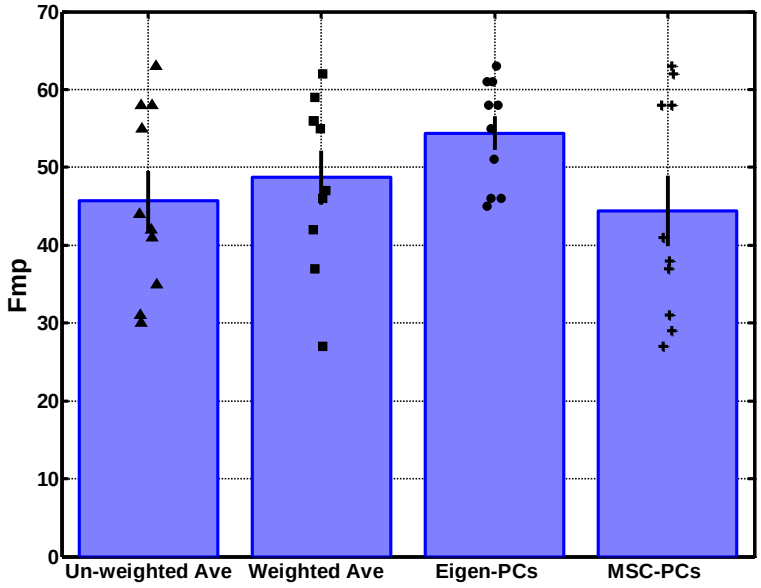

Figure 4.11: Comparison of weighted and un-weighted averaging with PCA based (Eigenvalue ordering and MSC-PCs) for the number of the channels that each of the methods can find above its critical value. In each bar, the edge is the mean, the whisker is standard error and the markers are data scatter.

The last part of this section compares two ICA base noise reduction methods, i.e. when MSC is employed for component selection and also when the components are ordered by their variances (Var-ICs). The result of applying the two noise reduction methods on the ALR recorded from ten normal hearing subjects for obtaining highest SNR is illustrated in Figure 4.12. Bar chart and means have been selected for this comparison as the data have normal distribution.

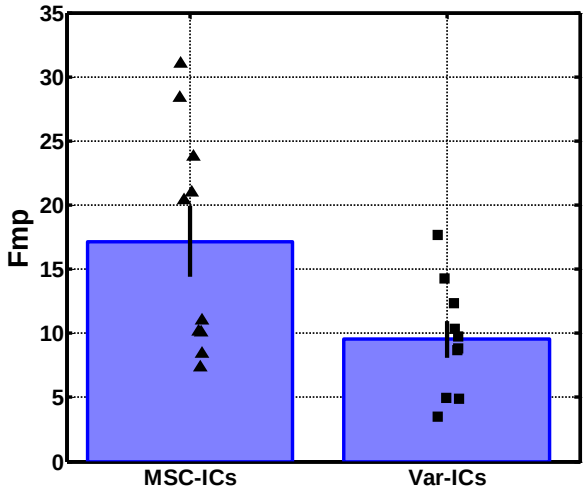

Figure 4.12: Comparing the MSC-ICs with the Var-ICs for providing higher SNR in10 normal hearing subjects. The mean value (edge), data scatter (markers) and standard error (error bar), are shown for each bar in the graph.

Moreover, this comparison is carried out for the number of channels that each method can find above its critical value. The box plot in Figure 4.13 depicts this comparison. The critical value for the Var-ICs was found to be 1.30 with a standard deviation of 0.017 through the same procedure explained in part 4.3.

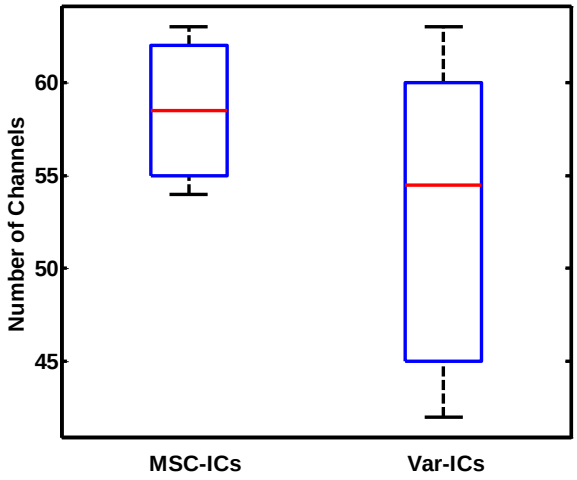

Figure 4.13: Comparison of the MSC-ICs with the Var-ICs for the number of the channels that each of the methods can find above its critical value.

The result of applying a dependent paired sample t-test on the high SNRs obtained by the methods (shown in Figure 4.12) shows that MSC-ICs is significantly better than the Var-ICs (p-vale<0.007) in providing higher SNRs. Moreover, the

MSC-ICs found have a better performance in finding more channel above the critical value than the Var-ICs (p-value<0.03).

4.5. Discussion and Conclusion

From the results it can be stated that weighted averaging significantly improves the SNR for the ALR recording in comparison to coherent averaging. This result is in accordance with (Ross et al. 1984) who showed weighted averaging significantly improved the quality of auditory steady state response (ASSR) signal. According to the results given by Figure 4.10 and 4.11, although PCA-eigenvalue ordering was not found to be significantly different with weighted averaging in providing high SNRs, it can be stated that PCA-Eigenvalue ordering has a somewhat better performance than the weighted averaging since it provides more channels above the critical value. However, as was predicted from the results of the simulation in chapter 3, MSC is not helpful in automatic selection of components when PCA is employed as a source separation tool.

According to the results presented in this chapter, MSC-ICs and MaxFmp-ICs showed the best performance among BSS based multichannel noise reduction methods, and was also to be better than the single-channel methods. Although Figures 4.8 and 4.9 showed difference between MSC-ICs and Max-Fmp-ICs, this difference was not found to be significant. It may be possible to see significance between these two methods for a larger sample size. However, a difference that needs a very large sample size to be detected is probably not of great practical importance, in this case. Additionally, results showed that the MSC-ICs is a better solution to the IC ordering problem than the Var-ICs for ICA based multichannel noise reduction for the ALR.

Although, all of the proposed methods are able to address the automatic selection issue in ICA based noise reduction methods, there still exist drawbacks that make these methods imperfect. For instance, the thresholds for discarding/accepting an IC in each method were estimated from extra recordings (resting EEG). Moreover, ICA is computationally complex and for large data the noise reduction procedure is time consuming. Furthermore, in MSC-ICs only one harmonic was considered, i.e. ICs in which the coherence is strong in other harmonics will be rejected by this method.

Consistent with previous work (Dobie and Wilson 1996; Dimitrijevic et al. 2001), weighted averaging is significantly better for SNR improvement than artifact rejection, i.e. removing the artifact from the data before averaging. On the other hand it has been shown in the current work that the MSC-ICs and Max-Fmp-ICs are significantly better than weighted averaging for both outcomes (obtaining highest SNR and also the number of channels that each method can find above the critical value). Therefore, it can be concluded that they are also better than artifact rejection.

To conclude, according to the results shown it can be stated that multichannel processing of the ALR is significantly a better option than singe channel processing alternatives. Nevertheless, taking the time required for subject preparation and also the noise reduction procedure in to account, single channel processing may still be considerably faster and more practical than multichannel. For the cases in which the signal quality has the first priority, multichannel recording is a better option. In chapter 7 we will discuss ways to obtain a more efficient compromise, using a smaller number of channels.

4.6. Summary

In this chapter first two single channel processing methods for noise reduction of the ALR were presented. The results showed that weighted averaging was a better choice than conventional coherent averaging for extracting the ALR from the noisy recorded signal. Furthermore, the four proposed multichannel noise reduction methods introduced in chapter 3 (MSC-ICs, Max-Fmp-ICs, Max-Kurt-ICs and Min-Entropy-ICs) along with the PCA based noise reduction methods (eigenvalue ordering and MSC-PCs) were applied on the same set of data. The results showed that the MSC-ICs and Max-Fmp-ICs had the best performance and the MSC-PCA had the worst performance in noise reduction of the ALR.

Chapter 5

Benefits of Multichannel Signal Processing for Applications of the ALR

5.1. Introduction

The main purpose of this chapter is to explore whether and how much multichannel signal processing strategies can be advantageous for improving the analysis procedure for applications of auditory late responses (ALR) in comparison with the single channel alternative methods. To achieve this target, the MSC-ICs has been employed as multichannel processing tool and the results were compared with single channel weighted averaging. It was shown in chapter 4 that by using the MSC-ICs as the multichannel noise reduction method, a better signal quality, i.e. higher signal to noise ratio (SNR), can be obtained in comparison with the single channel alternative. Since the quality of the signal is improved and a clearer signal is acquired, an enhancement in detection and diagnostic analysis is expected.

In this chapter, both single and multichannel signal processing strategies are applied on the ALR waveform to investigate the advantages of using multichannel signal processing over single channel in three applications. For each of the applications background, data acquisition and processing strategy are explained.

The first application is using the signal processing methods (single and multiple channel processing) to detect the hearing threshold level from the ALR waveform in ten normal hearing subjects. The aim is to determine if the multichannel processing can find the hearing threshold level, for normal hearing subjects, closer to 0 dB nHL (the threshold measured by pure tone audiometry) than the single channel method. In addition, since the amplitude of evoked response decreases with respect to decrement of stimulus intensity (Beagley and Knight 1967, Hall 2007), by decreasing the stimulus intensity, response detection becomes more difficult. The key question is whether the ICs found in the higher stimulus intensities can be used for noise reduction in lower stimulus intensities or not. In the second case study, the performance of multichannel processing for exploring the effect of inter-intra subject variability on the ALR waveform is compared with single channel alternative. A relative question is to explore the effect of habituation, i.e. the response decrement following repetitive presentation of stimulation (Muenssinger, Stingl et al. 2013), on the ALR amplitude. Finally, the effect of attention on the ALR waveform has also been investigated as the third application using multi and single channel strategies.

5.2. Hearing Threshold Estimation

5.2.1 Background

Although objective diagnostic methods are quite popular in modern medical science, behavioural pure-tone audiometry (PTA) remains the golden standard for identifying hearing threshold levels. Since the 1980's, high quality identification of hearing thresholds within the range of speech frequencies has been achieved by various methods of auditory evoked potentials (Laukli 1988). Using objective methods has made the determination of hearing thresholds much easier for patients in whom PTA cannot be employed, such as infants and those unable or unwilling to participate in traditional behavioural tests. The most commonly used 'objective' clinical method for detecting hearing threshold is auditory brainstem response (ABR) characterized by high wave reproducibility and responses that do not depend on the state of consciousness. Dagna, Canale et al. (2014) used ABR to find the hearing threshold of four different ABR tones at 1 kHz. For normal hearing participants they found the hearing threshold with a mean of 16dB above that of the behaviourally

acquired pure tone threshold. Alternatively, Picton et al. (2003) showed that recording human steady state response at different stimulus intensities can provide an objective assessment of audiometric threshold. Also, Komazec, Lemajic-Komazec et al. (2010) used Auditory Steady State Response (ASSR) to find the hearing threshold at different frequencies, specifically 0.5 1, 2 and 4 kHz. The thresholds measured by ASSR were found to be different with PTA with 4.1, 2.5, 4.4, and 4.2 dB nHL at 0.5 1, 2 and 4 kHz respectively. Moreover, Lightfoot (2006 and 2010) used ALR, N1 (first trough in the ALR waveform) to P2 (second peak in the ALR waveform), waveforms to measure the hearing threshold at 4 kHz. They found the hearing threshold with a mean of 25 dB above the pure tone threshold. Therefore, it can be said that the methods by which Komazecet al. (2010) measured the hearing threshold was superior to the method that Lightfoot (2010) used since the threshold value found by Komazec et al. (2010) was closer to 0 dB nHL at 4 kHz. However, it should be noted that they used different testing strategies (ASSR and ALR). The main purpose of this part is to explore that if multichannel signal processing can find the hearing threshold closer to 0 dB nHL when ALR is used for measuring the hearing threshold level.

5.2.2. Data Acquisition

Ten normal hearing subjects (5 M, 5 F) aged between 18 and 30 years old were asked to participate in to this study. All the subjects had been examined to verify normal hearing, using tympanometry and behavioral pure tone audiometry, before the test started. The test was explained and written informed consent was obtained. The stimulus was a 1 kHz tone burst of 70 ms duration with 10 ms rise/fall time, which was applied to the right ear of the subjects. The test was performed in 8 different levels of stimulus intensity started from 60 dB above the normal hearing threshold (60 dB nHL), down to 0 dB nHL with the following steps: 60 dB, 40 dB, 30 dB, 20 dB, 15 dB, 10 dB, 5 dB and 0 dB. The 0 dB level was found by averaging over the hearing threshold level of 5 normal hearing subjects (different subjects), 10 ears, obtained by behavioral pure tone audiometry. The ALR wave form was recorded from the participants using a 63 channel EEG scalp cap and a sampling rate of 1 kHz with by using Neuroscan system (*Scan 4.03*). Moreover, the data was low-pass filtered at 100 Hz at the recording stage, since. The ALR wave has mainly low

frequencies and removing the high frequency components is helpful for recording signal with high quality. Two more electrodes were placed on the chest and the cheek of each subject to record heart beat and eye blinks respectively. The ALR recording was 4 min at each level of intensity; plus 4 min of no stimulus EEG recording, i.e. rest EEG, which makes the overall required time for the test 37 to 40 min.

5.2.3. Data Analysis

Both single and multichannel processing strategies were employed for extracting the ALR waveform from the recorded EEG. Weighted averaging over 155 sweeps was used for single channel processing and the MSC-ICs was employed for multichannel signal processing. As discussed in section 1.5, the ALR waveform has the best quality (highest SNR) at the vertex (Cz channel) (Hall 2007), so the Cz channel was selected for hearing threshold measurement. The last level of stimulus intensity in which the ALR response is present was used as the estimate of the hearing threshold for each subject. In this work, the critical values given by Table 2 in chapter 4 was used as the values below which the recorded signal was not considered as containing the response. In other words, for each subject the first level of stimulus intensity at which the *Fmp* of the recorded signal in Cz is below 1.32 is considered as the hearing threshold for the subject.

The performance of the multichannel processing and single channel processing were compared to see which of the methods is able to find the hearing threshold closest to 0 dB nHL. As the subjects are normal hearing, the thresholds closest to zero implies a more accurate outcome. Depending on the distribution of the obtained data, a parametric (paired sample t-test) or a non-parametric (Wilcoxon signed rank) statistical test is employed to test the significant difference between the methods.

A further investigation was carried out with the aim of increasing computational efficiency of the method. A study was carried out to test if the mixing matrix found in the ALR recorded at 60 dB nHL can be used for noise reduction in the ALR recorded at lower stimulus intensities. The reason for using the spatial distribution of the ICs instead of order of ICs is that the order of ICs changes in each execution; while the spatial distribution of the ICs is assumed to be unchanged. For example, in two successive executions of FastICA on a set of data, assume the order

of response ICs is 1, 2, 6, 9, 30 and 41. In the second execution this ordering must be like 3, 6, 18, 29, 37 and 39. Thus in the matrix representation of the signal model given by the equation (3.7), the columns of the mixing matrix $\boldsymbol{A}$ give information about the spatial distribution of the ICs. Hence, the inverse of the mixing matrix, $\boldsymbol{A}$, found by FastICA from the ALR recorded at 60 dB nHL was multiplied to the both sides of the equation (3.7) for all the EEGs (ALRs) recorded at lower stimulus intensities as follows:

$$\mathbf{A^{-1}Y} = \mathbf{A^{-1}AX}$$
$$\mathbf{A^{-1}Y} = \mathbf{X} \tag{5.1}$$

where, $\boldsymbol{X}$ is the matrix of independent components (row of $\boldsymbol{X}$ are independent components) and $\boldsymbol{Y}$ is the observation matrix (rows of $\boldsymbol{Y}$ are ALR recording at each electrode). Then, the ICs which were significantly coherent with the stimulus (p-value<0.05) were kept as the response ICs and the data was reconstructed by only using the response ICs, as described previously (in section 4.3). The result of calculating the mixing matrix $\boldsymbol{A}$ only at 60 dB and of repeating the calculation for all intensity levels were compared, and assessed based on a Wilcoxon test.

5.3 Inter and Intra Subject Variability

5.3.1. Background

Intra-subject and inter subject variability has been the topic of much research in brain pathology, brain function and cognitive control (Corr 2008, Leue, Klein et al. 2013). Inter /intra subject variability is generally measured through two main approaches: 1) A performance tests, i.e. analysing the performance of subjects doing an identical task. 2) Event related potentials (ERP) (Leue, Klein et al. 2013). Measuring inter/intra subject variability is helpful when ERP is used to monitor the brain functionality under a particular condition, e.g. following the use of medicines, assessing the state of attention, anaesthesia, etc. Inter-subject variability can be a confounding factor in many studies, including in the effect of attention. Moreover, the performance variability has been reported as a common feature in disorders of frontal and putatively frontal pathology such as Traumatic Brain Injury (TBI), schizophrenia and attention deficit hyperactivity disorder (ADHD) (Bellgrove et al. 2004). Therefore, detecting this variability can be vital for early detection of many disorders.

It has been reported by (Foster, Stevens et al. 2013) that the auditory steady state response (ASSR) changes over the trials. The reason can be the failure of the brain in responding to some of the stimuli. This fact brings up a similar question about the ALR that deserves an answer.

Effect of habituation on auditory late response has been subject of much research (Dorman 1973; Polich 1988; Rosburg 2002; Ritter 2004; Rosburg 2006). The main goal of this section is to compare the performance of multichannel and the single channel processing for detecting the inter/intra subject variability for the ALR. The aim is to address the following questions:

1. Do the ALRs recorded from a subject in different recording sessions change? If so, is multichannel processing advantageous to see this difference?
2. Are the ALR signals the same between different subjects? If not, is multichannel processing beneficial over single channel processing to detect the difference?
3. Does multichannel processing alleviate the problem of habituation, known to affect the ALR?

5.3.2. Data Acquisition

For this application the ALR were acquired under the similar setting as the previous application, i.e. 63 channels EEG, 100Hz low-pass filtering at the recording stage and the same stimulus (1 kHz tone burst of 70 ms length and 10 ms rise/fall time). In contrast with the previous recording protocol, for each subject the ALR was recorded three times (recorded in the same session) at 60 dB nHL stimulus intensity level.

Additionally, for inspecting the effect of habituation on the ALR waveform, a long ALR with 465 stimuli (the same stimulus as in section 5.2) was also acquired. Moreover, a 4min resting EEG without stimulation was recorded for noise estimation.

5.3.3. Data Analysis

In a similar fashion as the previous part, initially both single and multichannel processing methods (MSC-ICs and weighted averaging) were employed for noise reduction for the recorded ALR. After the noise reduction step, since the data recorded from the vertex (Cz) is expected to have the best quality (Hall 2007), the variability of the ALR in different recordings (between subjects or repeated

recordings within a subject) was measured by determining the latency and amplitude of the major peaks of the ALR at Cz in each recording.

To investigate the variability of the ALR over different recording session in one subject, 100 epochs were randomly selected from 155 epochs from each recording and the ensemble average was calculated over the 100 epochs. Then the amplitude of P1, N1 and P2 were measured; and this procedure was repeated ten times for each subject for each session. The measured values of each peak were compared with the values obtained from the other recordings from the same subject. Depending on the distribution of the data this comparison was carried out by applying a parametric or non-parametric statistical test. A similar strategy was employed for measuring the variability of the ALR across the subjects. This time the comparison was carried out between the peaks amplitudes obtained from different subjects.

For inspecting the habituation effect on the ALR waveform recorded from the ten subjects, the aim was to explore if the amplitude of the major peaks (P1, N1, P2 and N2) decreases over time as the stimulus is repeated. Hence, the ALR signal was divided into three parts and the ensemble average was calculated over each part; then the amplitudes of major peaks were measured in each ensemble averaged signal and compared with each other. The maximum value of the averaged ALR wave form over interval 1 to 100 ms, minimum value over 50 ms to 200 and maximum value over 100 ms to 250 ms after stimulus onset gives the amplitudes of P1, N1 and P2 respectively. If a significant decrement in the peaks observed it can be interpreted as a habituation effect. Moreover, the long ALR (465 epochs) was used in a similar fashion to explore if the habituation effect is detectable when stimulation with more repetition is used, i.e. when longer data is recorded.

5.4. Effect of Attention

5.4.1 Background

There exist various methods for investigating the effect of attention on the human auditory system such as a dichotic listening test which is a psychological test commonly used to investigate selective attention within the auditory system and is a subtopic of cognitive psychology and neuroscience. During a standard dichotic listening test, a participant is presented with two different auditory stimuli

simultaneously (usually speech). The different stimuli are directed into different ears over headphones. Research participants are instructed to repeat aloud the words they heard in one ear while a different message is presented to the other ear. A wide range of attention tests are designed based on Dichotic listening test (Hillyard, Hink et al. 1973, Naatanen, Gaillard et al. 1978, Maatta, Paakkonen et al. 2005, Garell, Neelon et al. 2006, Thornton, Harmer et al. 2007)

Alternatively auditory evoked responses can be employed to see the effect of attention on the auditory system. Davis (1964) described the effect of attention on evoked responses as an increase in the N1-P2 peak to peak in the late latency evokes responses. His test was a simple design, and consisted of two sections: reading a book when ignoring the stimuli and counting the stimuli when attending. Davis used bipolar leads and small tone pips as the stimulus. Davis's results (enhancement of evoked responses due to attention) confirmed the results reported by Davis, Fernandez et al. (1950) while they used clicks as the stimulus. It has been reported by (Picton et al. 1971; Picton and Hillyard 1974) that there is no significant effect as function of attention on short latency evoked responses, ABR or AMLR. The effect of attention is expected to be observed in the major peaks and troughs of the ALR: at 100 ms after the stimulus onset, in an enhancement in N1 (Hillyard et al. 1973; Naatanen et al. 1978; Maatta et al. 2005; Garell et al. 2006; Thornton et al. 2007), at 200 ms, with a larger P2 peak (Garell et al. 2006), and at 300 ms, with a deeper N3 trough (Sanders et al. 2006). Alternatively, the N1 to P2 and also P2 to N3 peak to peak amplitude can be investigated to see the effect of the attention on the ALR. It is worthwhile to say that attention does not have a significant effect on the latency of the major components of the ALR waveform (Maatta et al. 2005; Thornton et al. 2007). However, the underlying neural mechanisms of selective attention are still poorly understood (Petkov et al. 2004). In a review, Kreiter (2001) suggested that synchronization of neural populations could be such a mechanism. In ERP terms this would mean that with attention, the individual unit responses would be better synchronised to the stimulus. Thus there would be less latency jitter between single trials, leading to an increase in N1 amplitude in the average of the trials (Thornton et al. 2007).

5.4.2 Data Acquisition

The test consisted of two main parts, attend and ignore. In the ignore part the participants were asked to ignore the stimulus and read a novel, while in the attending part they were asked to count the number of stimuli presented. The stimulus, for both attend and ignore section, was a 1 kHz tone burst which was presented 155 times in their right ear for each subject. The ALR was captured from 10 normal hearing subjects (5 M, 5 F) same subjects and the same system setting as in the previous sections.

Both single channel and multichannel signal processing strategies, i.e. weighted averaging and the MSC-ICs, were employed for noise reduction and ALR detection in the captured signals. The main objective of this section is to explore whether and how much multichannel processing can be advantageous over single channel processing methods for investigating the effect of attention on the ALR waveform.

5.4.3. Data Analysis

The main purpose of this section is compare the performance of multichannel signal processing (MSC-ICs) and single channel processing (weighted averaging) to explore the effect of attention on the ALR. Both MSC-ICs and weighted averaging noise reduction methods were applied on the ALR waveforms captured under attend and ignore conditions. Then, amplitudes of the major peaks of the ALR waveform captured under attention were compared with the peaks of the ALR recorded when instructed to ignore them. Furthermore, N1 to P2 and P2 to N3 peak to peak amplitudes were compared in the attend and the ignore ALR. In order to compare the results from the ignore and attend conditions in both single and multichannel processing, depending on the distribution of the data, a parametric or a non-parametric statistical test were used to investigate significant change in the ALR waveform.

5.5. Results and Statistical Analysis

Hearing threshold measurement-By using the results presented in the Table 4.1, the critical value of *Fmp*, i.e. the *Fmp* value above which a signal is deemed to

contain the response, was selected as 1.32 for the MSC-ICs and 1.29 for weighted averaging.

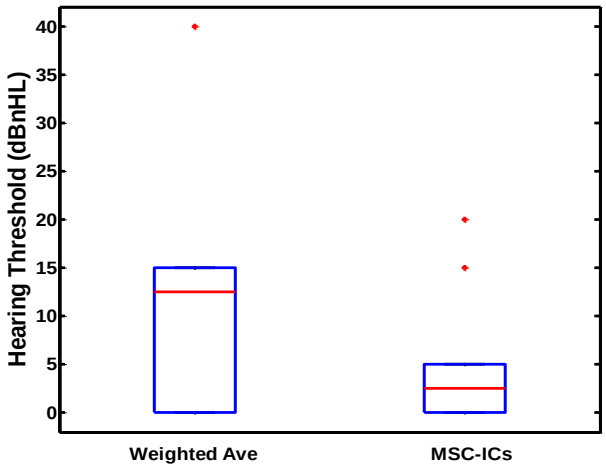

Figure 5.1: Hearing threshold found in ten normal hearing subjects using weighted averaging and the MSC-ICs at Cz. In each boxplot, the middle line shows the median, the edges show 75^{th} and 25^{th} percentiles and the stars show the outliers.

It can be seen from Figure 5.1 that the median of the hearing thresholds found by weighted averaging across the 10 subjects was 12.5 dB nHL while this value was 2.5 dB nHL when MSC-ICs was used for threshold estimation. Results of applying a Wilcoxon signed rank test on the thresholds obtained by the methods shows that the MSC-ICs significantly improves the hearing level detection (p-value< 0.04).

In addition to employing the MSC-ICs to estimate the hearing threshold, the spatial distribution information of ICs (mixing matrix) of the ALR recorded at 60 dB nHL stimulus intensity was used for component separation in the ALR signals recorded at lower levels of stimulus intensity, i.e. 40 dB nHL, 30 dB nHL 20 dB nHL, 15 dB nHL, 10 dB nHL, 5 dB nHL and 0 dB nHL. The SNRs were measured for both methods (MSC-ICs and MSC-Spatial Distribution) at each level of stimulus intensity. The results of applying a Wilcoxon signed rank test on the SNRs produced by the two methods (for each participants) showed no significant difference between the methods. Since the Wilcoxon test was applied on SNR pairs for each subject (ten subjects in total) to correct the false positive rate of repeating a statistical test the significance value was set to α=0.005. Furthermore, the hearing thresholds were measured from both methods and the results were compared with each other using the

boxplot in Figure 5.2. The median of the hearing thresholds for ten subjects were 2.5 dB nHL and 0 dB nHL for MSC-ICs and MSC-Spatial distribution respectively.

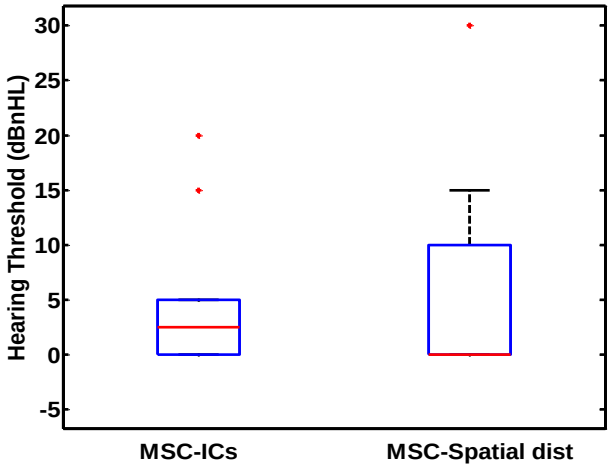

Figure 5.2: MSC-ICs and MSC-Spatial distribution used for finding hearing level thresholds from the ALR recorded at the Cz. In MSC-Spatial distribution the spatial distribution information of the ICs (mixing matrix) recorded at the highest stimulus level was used for the calculation of components in the ALR recorded at lower stimulus levels.

Although in this case the hearing threshold was found to be closer to 0 dB nHL when MSC-Spatial distribution was applied on the ALR, this improvement was not found to be significant – presumably due to the wide scatter of results observed in Figure 5.2. Nevertheless, the MSC-Spatial distribution may be considered a better method for finding hearing threshold, as it is considerably faster than the MSC-ICs which need ICA to be performed on the signals recorded at each level of stimulation.

Inter/Intra subject variability-After applying both MSC-ICs and weighted averaging noise reduction methods on the ALRs recorded from the subjects (three ALR for each subject), the amplitudes of the major peaks (P1, N1 and P2) of the ALRs in different recording sessions from each subject (intra subject variability) were computed as was explained in section 5.3.3. These values are shown for one subject in Figure 5.3 and 5.4 for both methods.

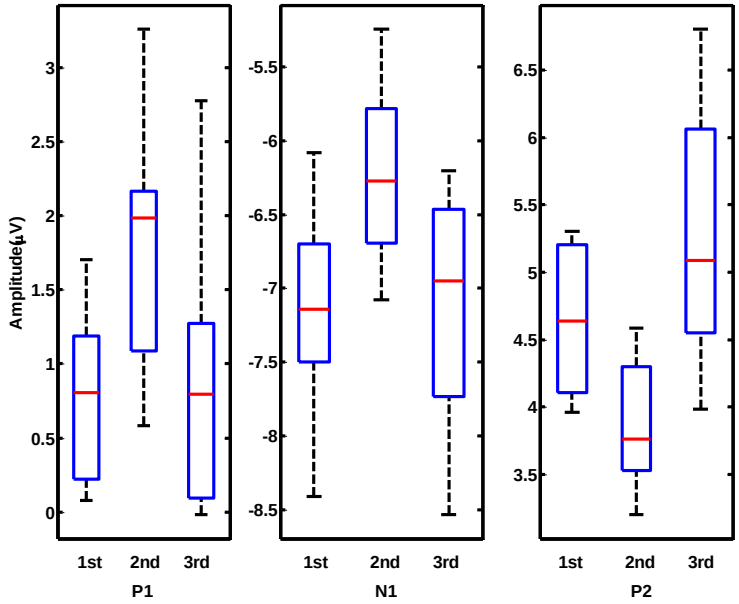

Figure 5.3: Amplitude of P1, N1 and P2 for three recording of ALR for one subject. Values are obtained as explained in section 5.3.3 after noise reduction by weighted averaging.

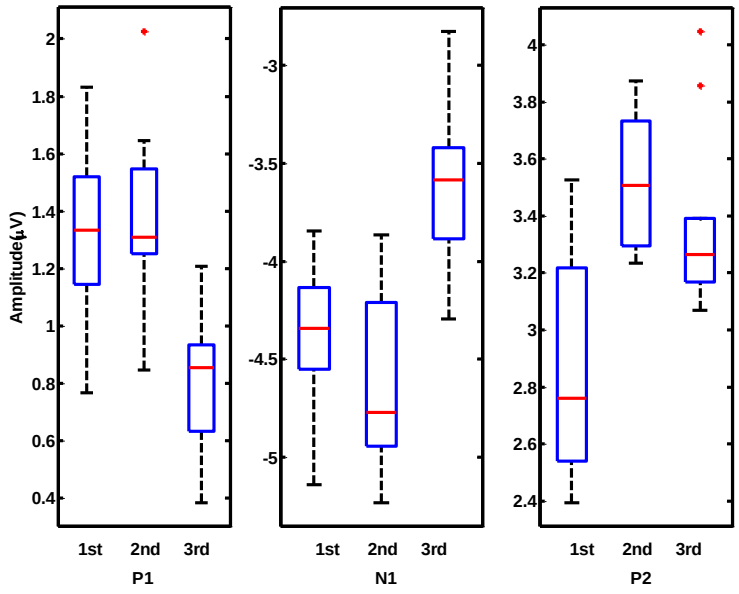

Figure 5.4: Amplitude of P1,N1 and P2 for three recording of ALR for one subject. Values are obtained as explained in section (5.3.3) after noise reduction by MSC-ICs.

Friedman's test was applied on the amplitudes of each major peak measured in different sessions, i.e. three Friedman's test for each subject. The amplitudes of the major peaks of the ALRs recorded form each subject in three different sessions were found to be significantly different (p-value <0.05) for all the subjects. This significance was observed in the results of both weighted averaging and the MSC-ICs

methods. Comparison of the amplitude of major peaks of the ALR was also carried out for the ALRs captured from different subjects (inter subject variability). Results of applying a Friedman test on the measured amplitudes also showed a significant difference between the ALRs recorded form different subjects. Analogous to intra subject variability results, this significance was reported by both the results of both weighted averaging and the MSC-ICs. Note that the positions of the major peaks (latency of the peaks) were not found to be significantly different for inter and intra subject variability. Furthermore the effect of habituation was investigated for both 155 epoch ALR and 465 epoch ALR in ten subjects for both MSC-ICs and weighted averaging noise reduction methods. The amplitudes of the major peaks in the 1st, 2nd and 3rd third of the ALR for both case 155 and 456 epoch ALR were measured. The results from neither of the methods confirmed a consistent decrement in the major peaks or SNR in neither of the recordings (155 and 465 epochs) in any of the subjects. The amplitudes of the major peaks in the 1st, 2nd and 3rd third of the ALR (for 155 epochs ALR) for one subject, after noise reduction by using weighted averaging and MSC-ICs, are shown in Figure 5.5.

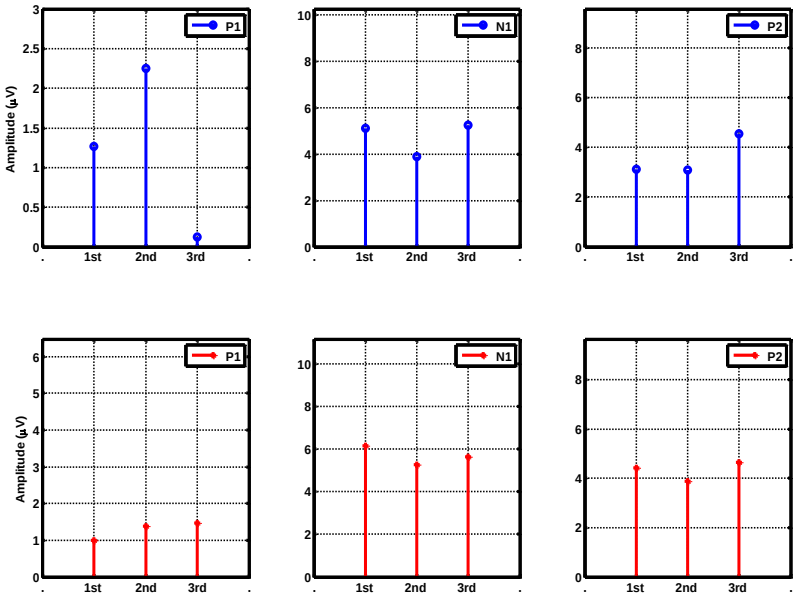

Figure 5.5: Habituation effect on the ALR waveform for one of the subjects. Upper, major peaks amplitude obtained from the in 1st, 2nd and 3rd third of the de-noised ALR by using MSC-ICs. Lower, the major peaks when weighted averaging was used for noise reduction. No significant decrement can be observed in the major peaks obtained by either of the methods.

According to the results presented in previous part, no consistent decrement can be seen in any of the peaks obtained by either of the methods. Moreover, the SNRs were calculated over these intervals for the same subject. As it is shown in Figure 5.6, no consistent decrement in SNR was detected in any of the subjects through any of the methods. These results hold true also for the case in which the 450 epochs of ALR were used. Figure 5.6 shows the SNR over 1st, 2nd and 3rd third of the ALR (155) epochs for one of the subjects.

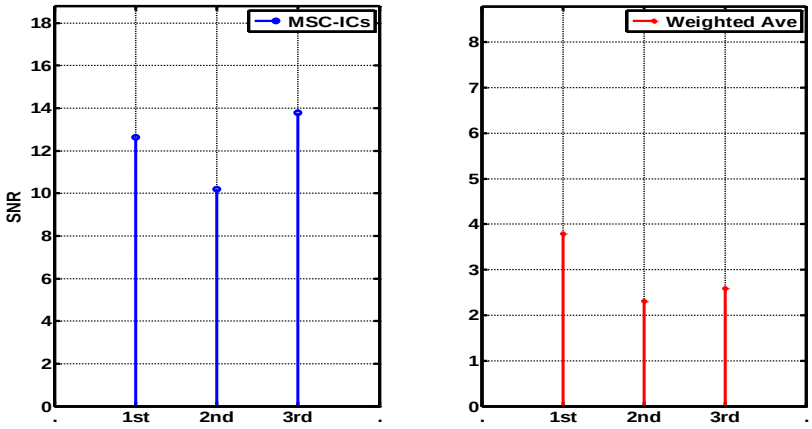

Figure 5.6: The SNRs calculated over the 1st, 2nd and 3rd third of the ALR waveform to investigate the habituation effect. Left, when MSC-ICs was employed for noise reduction. Right, noise reduced by weighted averaging.

Effect of Attention-The effect of attention on amplitudes of the major peaks was investigated in the ALR waves produced by both weighted averaging and MSC-ICs. The ALR waves recorded from one subject under attend and ignore protocols are shown in Figure 5.7. An increment in the amplitude can be observed in the major peaks of the ALR due to the attention. Effect of attention was also investigated on the latency of the major peaks of the ALR and the results did not show a significant change in the peaks latencies due to the attention.

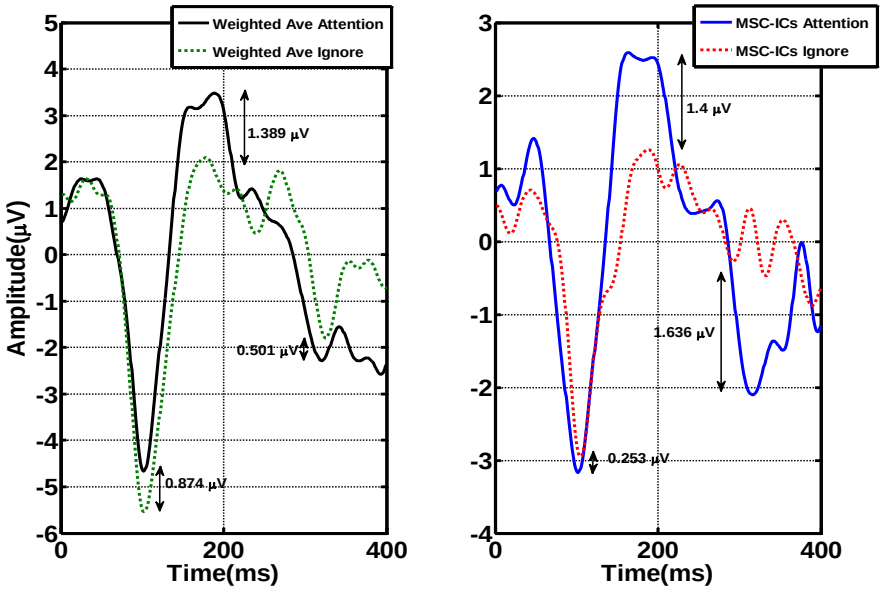

Figure 5.7: The noise reduced ALR by weighted averaging and MSC-ICs for one subject under both attend and ignore conditions. The attention effect is observed as an increment in the amplitude of the major peaks in the ALR.

This comparison was carried out for all the ten participants for both single and multichannel processing methods. The results are shown in Figure 5.8 and 5.9 for weighted averaging and MSC-ICs respectively.

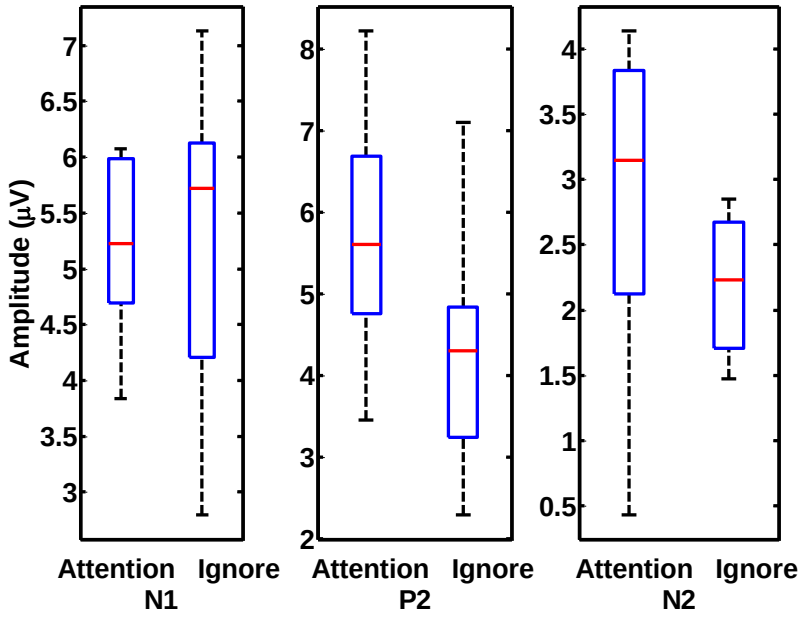

Figure 5.8: Effect of attention on amplitudes of major peaks of the ALR waveform found by weighted averaging. Except N1 the amplitudes of the major peaks become larger. Middle line shows the median, the edges show 25th and 75th percentiles and the whiskers indicate the maximum and minimum extremes.

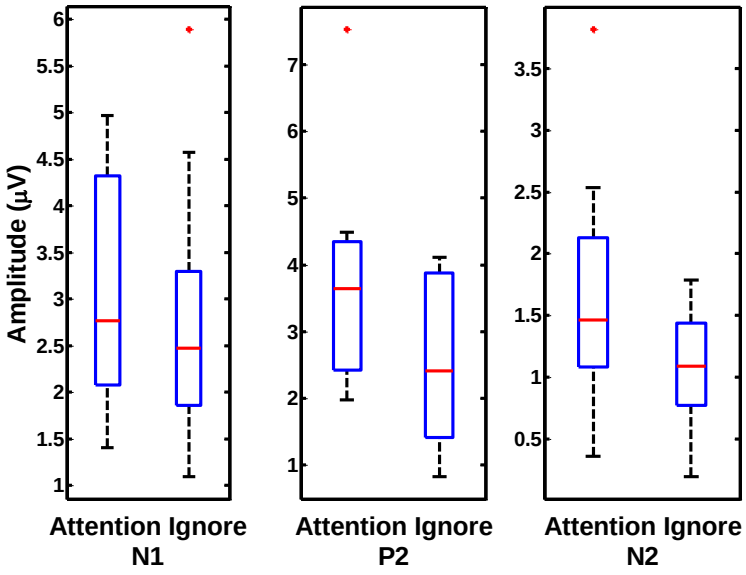

Figure 5.9: Effect of attention on the major peaks of ALR in ten normal hearing subjects found by MSC-ICs. For ten subjects an increment can be seen in all the major peaks. Middle line shows the median, the edges show 25th and 75th percentiles, the whiskers indicate the maximum and minimum extremes and the stars are outliers.

In order to test if the peak enhancement due to attention is significant at each peak, a Wilcoxon signed rank test was applied on each pair of the peak amplitudes shown in Figure 5.6 and 5.7. The results implied a significant enhancement in P2 for both noise reduction methods. In addition, when the MSC-ICs was employed, N2 was found to be significantly enhanced due to the attention. This increment was not found to be significant for the weighted averaging alternative. Although N1 was found to be larger under attention when MSC-ICs was used, there was no significant enhancement detected at N1 by either of the methods.

5.6. Discussion and Conclusion

Although pure tone audiometry is more sensitive than subjective methods, cortical hearing threshold measurement can be useful for the cases in which the subjects are not able to attend to the behavioural audiometry test, e.g. when the subject is in coma. The results presented here show that multichannel signal processing considerably improves the sensitivity of hearing threshold measurement in comparison with the single channel alternative. However, the required time for subject preparation in multichannel recording is much longer than for the single channel set up, and computational time is much greater, as it requires ICA to be

carried out at each stimulus intensity. In regard to the latter, the results suggest that the spatial distribution information of the ICs obtained from the ALR recorded at 60 dB nHL can be used for noise reduction and response detection for the ALRs recoded at lower levels of stimulus intensities. This indicates that one ICA at the highest intensity is sufficient and this computationally expensive step need not be repeated. Although the hearing threshold was estimated closer to 0 dB nHL when the MSC-Spatial dist. was employed instead of MSC-ICs, the results of statistical comparison did not confirm any significant different between the outcomes of the two methods. Significant improvement might have been observed with a larger sample size – such as study is suggested for the continuation of the current work. Nevertheless, it can be stated that the MSC-Spatial distribution might be preferred in finding the hearing threshold level since it is much faster than the MSC-ICs method. The average hearing threshold level found by MSC-ICs and MSC-Spatial dist. for the normal hearing subjects were found 4 dB nHL and 2.5 dB nHL above the pure tone audiometry. This is considerably better than what Lightfoot (2010) reported (25 dB nHL) when ALR was used for hearing threshold measurement.

Both weighted averaging and MSC-ICs found a significant difference in the amplitude of the ALR wave in different recording sessions for a subject and also across the subjects. The results showed that single channel processing can detect the inter and intra subject variability as well as the multichannel alternative and for this application multichannel recording was not appear to be any better than single weighted averaging. However, it is worth noting that this variability can be due to poor signal quality and not because of variability of the ALR. Therefore, it can be stated that for this work multichannel processing was not beneficiary in comparison with single channel processing.

Moreover, the habituation effect was not observed through either of the methods since no consistent decrement was observed in the SNR or amplitudes of the major peaks of the ALR over 1^{st}, 2^{nd} and 3^{rd} third of the ALR waveform. The reason can be traced in the length of the recorded signals, quality of the recorded signals or the method by which the habituation effect was tested. By using better recording equipment or a different testing protocol such the one used by (Zhang 2009), it may be possible to see the effect of habituation on the recorded data.

Both weighted averaging and the MSC-ICs method were applied on the ALR recorded form the normal hearing subjects under attend and ignore conditions and the effect of attention was investigated by comparing the amplitude of the major peaks of the ALR captured under attention and ignore circumstances. Both of the methods showed a significant enhancement in P2 due to the attention. The amplitude of N2 was also found to be significantly increased when MSC-ICs was employed. Contrary to the results reported by (Picton and Hillyard 1974; Thornton et al. 2007), no significant enhancement was observed at the N1. It is worth mentioning that the common factor between all the previous researches was selective attention, whilst in this work attention was not selective. A significant enhancement could be observed at N1, if the data was collected under selective attention (Thornton et al. 2007). Since the results of investigating habituation effect on the ALR did not confirm any significant trend in the amplitudes of the major peaks, the observed significant enhancement in P2 and N2 are due to attention, i.e. this enhancement is not due to inter/intra subject variability as inter and intra subject variability does not follow a consistent increment. Form the results presented in this chapter it can be concluded that detecting the effect of attention on the ALR can be significantly improved by using multichannel processing rather than the single channel alternative.

5.7. Summary

The main objective of this chapter was to compare the performance of a multichannel signal processing tool (MSC-ICs), with the single channel processing alternative (weighted averaging) for three clinical applications (hearing threshold measurement, inter/intra subject variability and effect of attention) of the ALR. The main aim was to explore if the use of multichannel signal processing is advantageous over single channel processing in clinical applications. According to the results, multichannel signal processing is significantly better than single channel processing for hearing threshold measurement and investigating the effect of attention. However, there was no significant difference observed between multichannel and single channel processing for assessing inter/ intra subject variability. Although using multichannel processing improves the quality of technical analysis, more extensive subject preparation with the need to apply many more electrodes limits its clinical use. This limitation will be addressed in chapter 7.

Chapter 6

Multichannel Artifact Rejection

6.1. Introduction

Generally the EEG recorded from individuals during auditory stimulation consists of a mixture of brain response and noise. Artifact is a general terminology to a group of undesired signals originating from eye movement (ocular artifact), muscle activity, heartbeat, environmental noise and other sources. The amplitude of the artifacts are usually several orders larger than the amplitude of the signal of interest (Breuer, Dammers et al. 2014). Therefore, artifact rejection from the data as a prior step to analysis is essential. It has been shown, for example, that artifact rejection prior to source localization can considerably improve the quality of the localization (Breuer, Dammers et al. 2014). Biological artifact can be classified into two categories, repetitive and non-repetitive artifacts. For instance heart beat and eye movement are repetitive artifacts and muscle activity is a non-repetitive artifact, i.e. eye-blink is approximately deterministic, in the sense that each repetition of artifacts has approximately the same shape, whereas the EMG is a random burst, which is always different in details. The focus of this chapter is on removing the heartbeat artifact from the ALR wave, but the approach could also be applied to other deterministic artifacts.

Various methods of artifact removal have been proposed (Barbati et al. 2004; Krishnaveni et al. 2006; Dammers et al. 2008; Kelly et al. 2011). For the particular case of heart beat artifact rejection, two different approaches are widely used. In the first approach which is called *template matching*, the heart beat artifact can be rejected by subtracting a reference signal form the recorded signal. This reference is estimated by calculating the average cardiac activity around the R-peak at each heart beat cycle (Allen et al. 1998; Sijbers et al. 2000). In this method the brain responses which are synchronized to the heart beats, i.e. occur at the same time with the heart beats, are likely to be distorted by this subtraction. In the second approach cardiac activity is removed by using ICA. In ICA-based artifact rejection methods first the data is decomposed in to its independent components; and then artifact rejection is performed discarding the component representing cardiac activity and reconstructing the data from the remaining components. If the signal separation and component selection are properly applied, the ICA-based methods do not distort the signal of interest (Dammers et al. 2008; Breuer et al. 2014).

The major challenge for all kind of ICA based artifact rejection methods is component selection. In ICA based cardiac artifact rejection methods, cardiac component selection stage is carried out either by visual inspection (Jung et al. 2000; Iriarte et al. 2003) or automatic selection (Dammers et al. 2008). When working with high dimensional data, visual inspection become difficult and time consuming (Breuer et al. 2014).

There are various methods to automate the ICA-based artifact rejection algorithms. Croft and colleagues, for example, investigated different regression techniques (Hamalainen et al. 1993; Barbati et al. 2004), with the major drawback that the proposed method can be applied for ocular artifact (OA) removal only. Additionally there are other approaches in which adaptive filter techniques (Escudero et al. 2007; Dammers et al. 2008) including Kalman-filter-based methods (Krishnaveni et al. 2006; Kelly et al. 2011) are employed for artifact rejection, however, all are designed for specific types of artifacts only.

Higher order statistics such as variance, kurtosis, and skewness were employed to identify the artifact components by (Delorme et al. 2001; Barbati, et al. 2004; Dammers et al. 2008; Klados et al. 2010). Also, Shannon's and Ranyi's entropy have been also used to automatically remove the artifact components from

the EEG recording (Greco et al. 2005). However, all amplitude-based methods are limited and less sensitive in cases where the strength of the IC is weak, thus, exhibiting a low signal to noise ratio (Rongen et al. 2006; Dammers et al. 2008). It has also been reported that Renyi's entropy is computationally costly for high-dimensional data due to the kernel density estimation necessary for each component (Dammers et al. 2008).

As an alternative, constrained independent component analysis (cICA) has also been used for artifact rejection from EEG recordings (James and Hesse 2005; Rajapakse and Wei 2006; Chawla 2011; Breuer et al. 2014). In cICA *a priori* information of the underlying source signal is used as a reference to optimize signal decomposition. In brief, cICA returns a component which is the closest component to the reference signal (Rajapakse and Wei 2006). This method is limited where the artifact signal has more than one component. For example, the decomposition of signals containing heart beats artifact often shows multiple ICs (Sander et al. 2002) whose activity is clearly related to heart beats. Since second and third cardiac components usually have small peak amplitudes, they may be identified by visual inspection, but are more difficult to find from its statistical properties though amplitude based methods. This limits the use of amplitude based methods. However, applying cICA in more than one iteration may address this drawback. cICA and artifact rejection is explained in more detail in the next section. There are alternative methods such as wavelets (Kelly et al. 2011) have been used for automatic artifact rejection which are not the main concern of this work and can be found in the literature (Krishnaveni et al. 2006; Hamaneh et al. 2014).

In this chapter cardiac artifact removal from EEG and the ALR is performed by using cICA and a novel artifact rejection method based on the MSC-ICs method which was introduced and explained in the previous chapters. The performance of this artifact rejection methods for removing cardiac artifact is then compared with cICA.

6.2. Cardiac Artifact Rejection

In order to perform and compare cardiac artifact rejection methods (MSC-ICs based and cICA based), the ALR data recorded from 10 normal hearing participants was used. Data acquisition and testing protocol was described in section 2.5. Artifact

rejection methods should be compared in two senses. 1) How much artifact has been removed from the signal 2) How well has the signal of interest been preserved.

6.2.1. Artifact Rejection by Using cICA

The cICA algorithm proposed by Lu and Rajapakse (2000) incorporates prior knowledge of the underlying expected signal in the internal cost function, to perform optimal signal decomposition. This constraint points the ICA algorithm in the direction of finding a particular solution which is optimally close to the reference signal (James and Lowe 2003). Different measures such as mean squared error (MSE) or correlation can be used to measure closeness (Lu and Rajapakse 2000; James and Lowe 2003). In this work MSE is used as the measure of closeness between the component and the reference signal. More information about the constrained optimization in cICA algorithm can be found in (Lu 2005). Heart beat artifact rejection from the ALR by using cICA can be summarized as follows:

1) Generating the reference signal.

2) Estimating the de-mixing matrix that provides the component which is optimally close to the reference. For the matrix form given by equation (6.1).

$$\boldsymbol{y} = \boldsymbol{AX} \tag{6.1}$$

$\boldsymbol{A}$ is the de-mixing matrix which if multiplied to the observation matrix $\boldsymbol{X}$, the output $\boldsymbol{y}$ would be the component (one component) which is optimally close to the reference signal.

3) Projecting the component to the measurement domain and subtracting it from the original signal. This stage removes the contribution of the, artifact component from all the channels.

In this work an extra step has also been added to the aforementioned steps which is repeating the procedure in more iterations until a desired quality of artifact rejection is achieved, i.e. the artifact rejection algorithm was applied on the data repeatedly. In each repetition a new artifact component (closest to the reference) was found and the data was reconstructed without including the artifact components. Assessing the quality of artifact rejection will be explained in the next sections.

The same data set that were used in the chapters 4 and 5 are also used for artifact rejection in this chapter. The reference signal is simply generated from the ECG signal (which is recorded via an electrode placed on the participant's chests) by setting a threshold. Here, the R-peak in the heart beat signal is selected for generating the reference signal, i.e. a threshold is set so that the R-wave peak exceeds it and the rest of the signal is below it and therefore set to zero. Therefore, the reference signal will be an impulse train which is synchronous with the R-Peaks of the heart beats. In this work, a toolbox implemented in MATLAB (Zhang 2008 version 1.0) was used for cICA.

6.2.2. Artifact Rejection by Using MSC-ICs

The MSC-ICs method was used for artifact rejection in a similar fashion to the previous chapter. The only difference in this case is that instead of keeping the ICs based on coherency of the ICs with the stimulus and discarding the rest, ICs which are significantly coherent with the heart beat are discarded and the data is reconstructed from the remaining ICs. Since the MSC uses coherence of the ICs and the heart beats, rather than amplitude information (PDF of amplitudes) of the ICs, it can be predicted that this method will not have the drawback of the amplitude based artifact rejection methods. Coherence of an IC with the heart beats is independent of amplitude and it only depends on the behaviour of the signal. MSC between an IC and the heart beats can be calculated by using equation (3.13). However, since the heart beats (R-peak or Q-peak) do not always happen exactly at the same onset, i.e. the heart beat signal is not exactly periodic; and the interval between two successive heart beats may vary. Hence, (3.13) cannot be simplified to (3.15). An example of three successive heartbeats is shown in Figure. 6.1, illustrating the differing time-intervals between successive heart-beats.

To be able to use the equation (3.13) for calculating the coherence between an IC and the heart beats, first the Q-peaks of the heart beat were found and the data was segmented into blocks of 650 samples, i.e. from 35 samples before the onsets of the Q-peak and 615 samples after. This segmentation is shown for one segment in Figure 6.2.Then equation (3.15) was used to calculate the coherence of the ICs and heart beats. Afterward, ICs which are significantly coherent with the heart beats (p-value

<0.05) are discarded as cardiac artifact components and the rest are projected to measurement domain for the reconstruction of the signal with our artefacts.

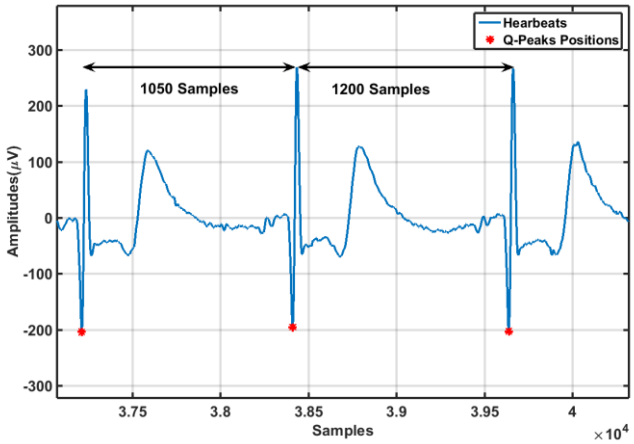

Figure 6.1: Heart beats signal recorded form one participant's chest. Position of Q-peak, onset of heartbeats, the distance between two randomly selected heart beat onsets are shown in the figure.

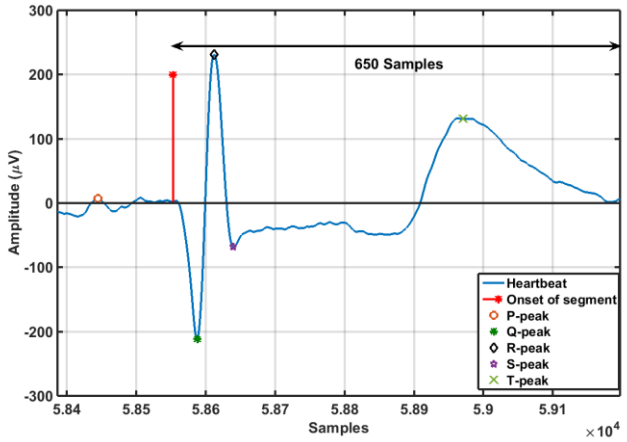

Figure 6.2: Major peaks of a heartbeat along with the onset of the segment. ICs are segmented to blocks with 650 samples.

6.2.3. Quality Measurement for Artifact Rejection

In order to compare the performance of artifact rejection by two methods, the rejection performance quantity R_p, as proposed by (Dammers et al. 2008) given by equation (6.2), was employed.

$$R_p = \frac{r(\boldsymbol{x}_{orig} - \boldsymbol{x}_{art\_rem})}{r(\boldsymbol{x}_{orig})} \quad \text{where } r = \frac{1}{N}\sum_{i=1}^{N}\sqrt{\frac{1}{T}\sum_t(\boldsymbol{x}(t))^2} \qquad (6.2)$$

where, $\boldsymbol{x}_{orig}$ represents the signal before artifact rejection, $\boldsymbol{x}_{art\_rem}$ represents the artifact removed signal, *r* represents the average *root mean square* (rms) value across *N* channels of the EEG recordings. Thus, $r(\boldsymbol{x}_{orig})$ expresses the mean (rms) value before the artifact rejection and $r(\boldsymbol{x}_{orig} - \boldsymbol{x}_{art\_rem})$ represents the mean rms value of the difference signal between the signal before ($\boldsymbol{x}_{orig}$) and after ($\boldsymbol{x}_{art\_rem}$) the artifact rejection respectively. With respect to cardiac artifact rejection, $R_p \rightarrow 0$ can be interpreted as a complete failure of the rejection process. For R_p being close to 1, artifact rejection is maximal. However, this value just shows much data has been rejected by the method and does not say if the artifact rejection has been carried out correctly. For example, if all the components are rejected by mistake, the R_p value will be 1.

For the case of auditory event related activities, the *Fmp* of the artifact removed signal can be considered as a measure of how well the evoked response has been preserved. In brief, a good artifact rejection method is the one which provides R_p close to one and also a signal with high *Fmp*. However, a drawback of this method is that the method is not able to identify which artifact has been removed. For example, if the goal is to remove heartbeats artifact but ocular artifact has been rejected by mistake, the *Fmp* of the artifact signal will still be improved. In order to show that the artifact which was removed by the artifact rejection methods was the heartbeat artifact, the cross correlations of the artifact removed signal with the original signal and also the reference signal which was generated in section 6.2.1 were calculated. High quality artifact rejection is obtained if the correlation between the artifact removed signal and the reference signal is small while correlation between the original signal and artifact removed signal is large.

A simulation was carried out to demonstrate the difference between the artifact rejection quality measurement methods. Four signals (D_1, D_2, D_3 and D_4) were used in this simulation and they are shown in Figure 6.3. Signal D_1 (a sine wave) was selected to simulate the signal of interest and D_2 (square wave) was used to simulate the heartbeat artifact. In addition, signals D_3 and D_4 were used to simulate muscle artifact and background noise respectively.

This simulation consists of two different scenarios. In the first case the artifact rejection was carried out correctly and only D_2 (the square wave which denotes the

heartbeat artifact) was rejected and signals were mixed excluding the square wave. In the second case, D_3 and D_4 and 75% of D_2 were removed. In other words, in the second case the heartbeat artifact was partially removed along with the noise and other artifacts, i.e. anything else than heartbeat like eye blink. Two case of artifact rejection are shown in Figure 6.4.

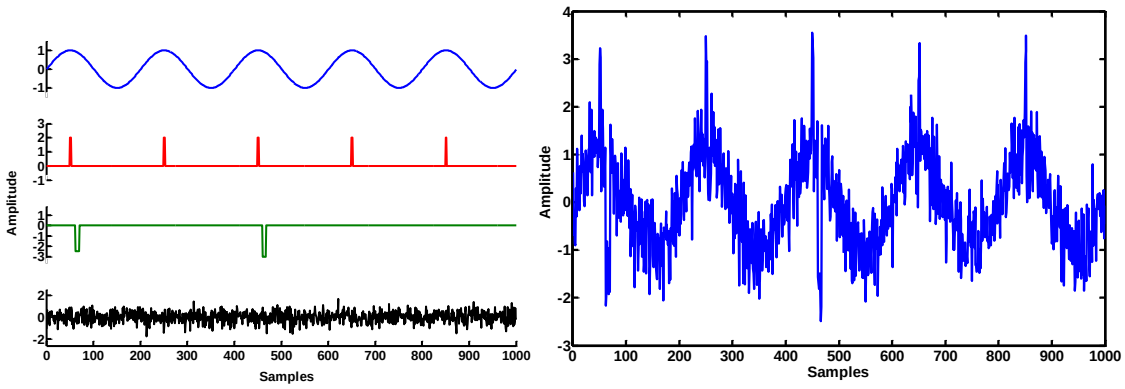

Figure 6.3: left, from the top D_1, D_2, D_3 and D_4 simulating signal, heartbeat artifact, muscle artifact and background noise respectively. Right, signal mixture

Signal to noise ratio (*Fmp*) was calculated for the signal mixture shown in the left side in Figure 6.3 and it was found to be 2.41

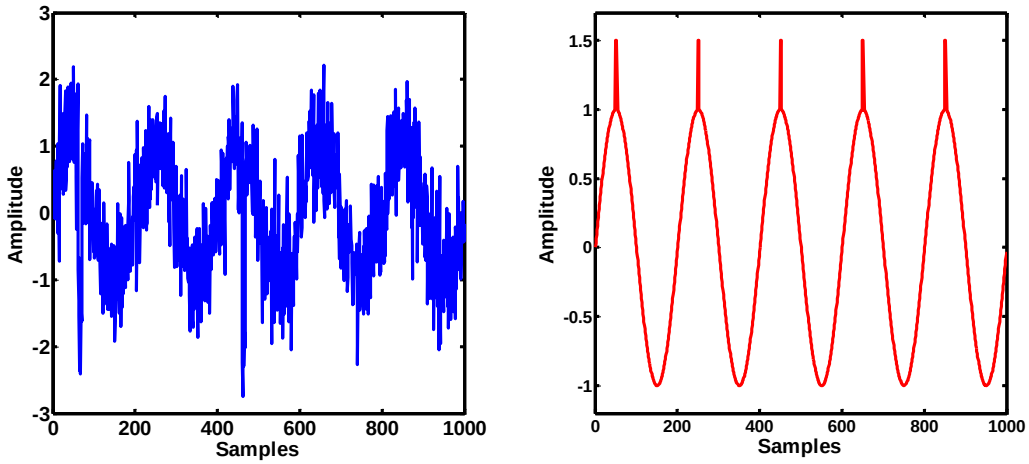

Figure 6.4: Left, correct artifact rejection in which the only removed component is D_2 (the square wave which denotes the heartbeats). Right, wrong artifact rejection in which muscle artifact D_3 and background noise D_4 are rejected and heartbeat artifact is not rejected.

Both *Fmp* and R_p were calculated for the signals shown in Figure 6.4. For the signal which simulates the incorrect artifact rejection and it is shown in the right side of Figure 6.4. R_p and *Fmp* were found to be 0.63 and 4.89 respectively. These values

were found 0.25 and 3.2 for the case in which the artifact rejection was carried out correctly.

On the other hand, correlation coefficient was calculated between the signal mixture on the right side of Figure 6.3 and D_2 (second signal from the top in Figure 6.3 right side) and it was found to be 0.40. Moreover, for the both signals shown in Figure 6.4, correlation coefficient was calculated between the artifact removed signals and signal D_2. In addition, correlation coefficients between the artifact removed signals and the signal mixture shown in the left side of Figure 6.3 were calculated. For the correct artifact rejection the correlation coefficient between the artifact removed signal (Figure 6.4 on the left side) and D_2 was found to be 0.08 while this value found to be 0.15 for the incorrect artifact rejection. Moreover, correlation coefficient between the signal shown on the left side of Figure 6.4 (correct artifact rejection) and right side of Figure 6.3 was found to be 0.96 while this value found to be 0.78 for the incorrect artifact rejection. These results implies that after artifact rejection, in the case of correct artifact rejection the contribution of D_2 in the signal mixture has been reduced more than the incorrect artifact rejection as the correlation coefficient between the artifact removed signal and D_2 (artifact signal) is lower than the other case. Higher correlation coefficient between artifact removed signal and the signal mixture for the correct artifact rejection case implies that the signal is preserved better. Note that the main purpose of this artifact rejection is only removing D_2.

6.3. Results

The result of applying the MSC-ICs method for removing the heart beat artifact from EEG (ALR) for one of the subjects is shown in Figure 6.5. In this figure, the channel which was highly contaminated by the cardiac artifact (channel 59) was selected to illustrate the performance of MSC-ICs for artifact rejection. For the subject whose EEG is shown in Figure 6.5, MSC-ICs found 3 components significantly coherent with the heart beat channel (p-value<0.01). Data was reconstructed after discarding 3 components. It can be seen form Figure 6.3 that the artifact is largely removed from the ALR signal. Moreover Figure 6.6 shows that the signal is mostly kept and only heart beat artifact has been removed. By using

equation (6.2), artifact removal quality indicator for this subject was found to be R_p =0.45.

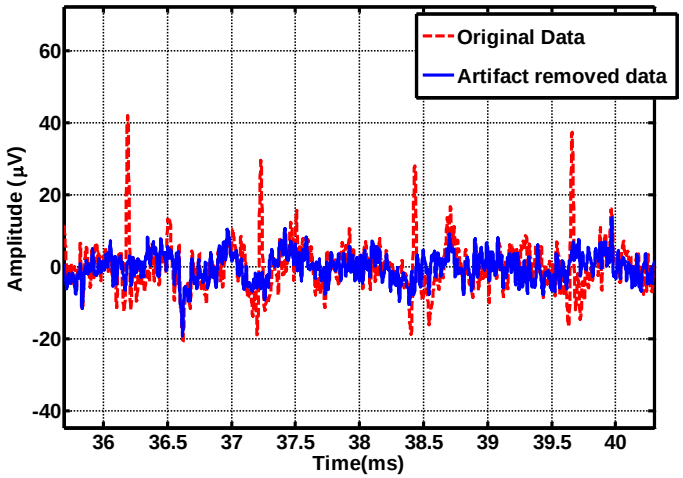

Figure 6.5: Cardiac artifact removal from the ALR signal by using MSC-ICs for one of the subjects. Original data (before artifact rejection), is shown by dashed line and the solid line is the ALR signal after artifact rejection.

Moreover, *Fmp* was calculated for channel 59 for both original signal (signal before artifact rejection) and the artifact corrected signal. For this subject, *Fmp* values were found to be 0.46 and 0.57 for the original data and artifact corrected data respectively. The small improvement in this channel can be explained by the small *Fmp* of the signal. As discussed in section 4.4, channels with *Fmp* below 1.25 are considered as channels with poor signal quality (response is either not present or it has a very small amplitude).

The correlation coefficient between the artifact removed signal and the reference signal (calculated in section 6.2.1) was calculated when MSC-ICs was employed for artifact rejection and the result showed that the artifact removed signal has a very small correlation with the heartbeat reference signal (correlation coefficient= 0.002). Furthermore, the correlation coefficient was also calculated between the artifact removed signal and the original data and the results show that the artifact removed signal and the original signal are highly correlated (correlation coefficient=0.95). These results imply that (for this subject and this channel) the heartbeat artifact has been removed and the rest of the signal is well preserved. This result is in accordance with the graphs shown in Figure 6.7, i.e. heartbeats are removed and the artifact removed signal follows the original signal.

Additionally, for the channels with *Fmp* above 1.25 (when the response is present) comparing the *Fmp*s of the signal before and after the artifact rejection via a paired sample t-test shows the quality of the signals are significantly improved due to artifact rejection. The *Fmp*s of the signal before and after artifact rejection for the channels above 1.25 for this subject is given in Figure 6.7.

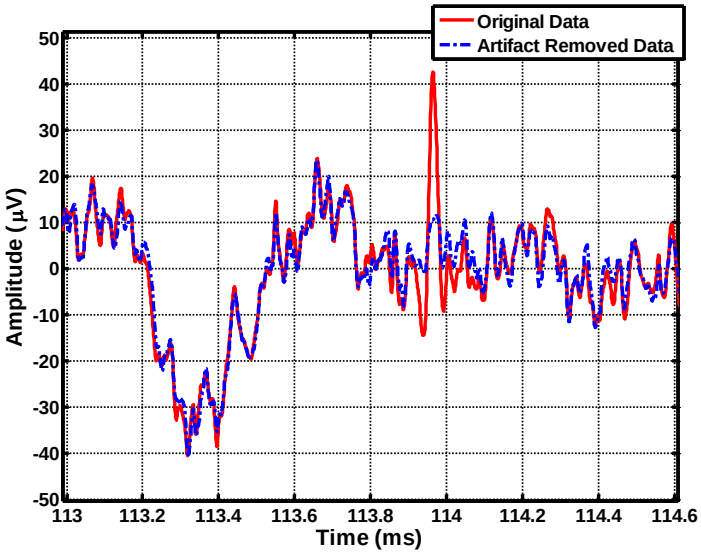

Figure 6.6: Artifact corrected signal is shown by dashed line and data before artifact rejection is shown by solid line. Artifact corrected signal follows the original signal except the heart beat is removed from the signal.

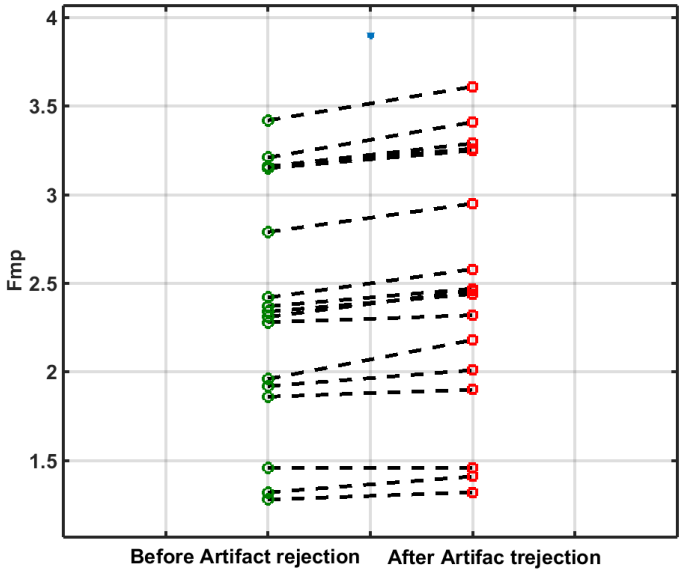

Figure 6.7: Channels with *Fmp* above 1.25 before and after artifact rejection for the same subject. *Fmp* of signal after artifact rejection is always greater or equal to the *Fmp* of signal before artifact rejection. Results of applying a paired sample t-test shows a significant improvement (p-value <0.001), shown by star, due to artifact rejection.

Alternatively, cICA was employed for cardiac artifact removal from the same data set. Figure 6.8, shows the results of artifact rejection by applying cICA on the same subject and in the same channel (channel 59) along with the original signal (before artifact rejection).

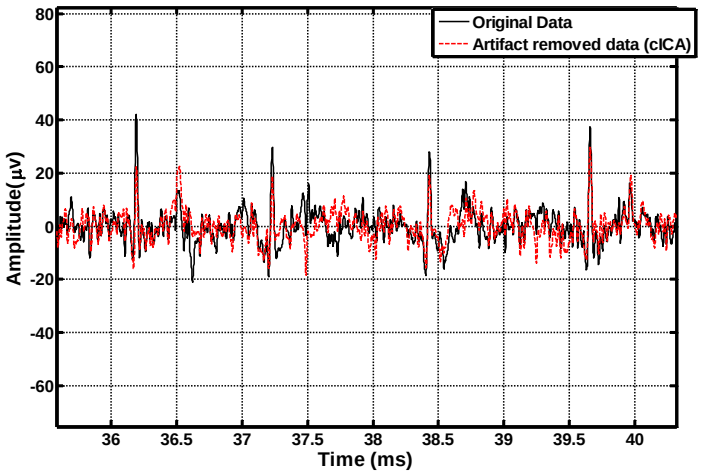

Figure 6.8: Cardiac artifact removal from the ALR signal by using cICA for one of the subjects. Solid line identifies the original data (before artifact rejection) and dashed line shows artifact corrected data for the same data set as the previous part.

Quality of the artifact rejection was measured for this recording. The results showed that using cICA can remove the cardiac artifact with R_p=52% and improve the *Fmp* of the signal (channel 59) from 0.46 to 0.6. Moreover, cICA was employed in 13 iterations for this subject and the results showed that after the first iteration, change in the rejection ratio (R_p) was small (for 13^{th} iteration R_p=55%).

Correlation coefficients were also calculated between the pairs of artifact removed signal and the heartbeats reference, and between artifact removed signal and original data for this subject in channel 59. The result showed that the artifact corrected signal by cICA was still correlated with the signal (correlation coefficient -0.035). Correlation coefficient between the artifact removed signal by cICA and the original data was found to be 0.45. These values are in accordance with the signal shown in Figure 6.8, i.e. heartbeats are not entirely removed from the signal.

As in the previous section, *Fmp* was calculated for the channels with *Fmp* above 1.25 and summarized in Figure 6.9. Results of applying a paired sample t-test on the data obtained from this subject showed that *Fmp* was significantly (p-value <0.001) improved due to artifact rejection by cICA for both one and thirteen

iterations. However, no significant improvement in *Fmp* was found due to increasing the number of iterations (from one to thirteen). Artifact rejection (for the same subject and in the same channel) by cICA is shown in Figure 6.10 for 13 iterations. After the 1st iteration, change in the signals is small.

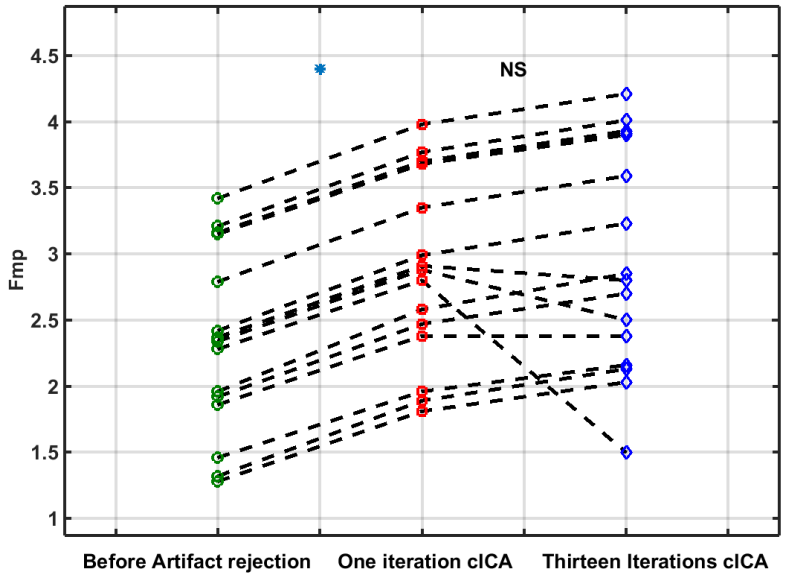

Figure 6.9: Channels with *Fmp* above 1.25 before and after artifact rejection for the same subject. Results of applying a paired sample t-test shows a significant improvement (p-value <0.001) due to artifact rejection. No significant improvement was found in *Fmp* from one iteration to thirteen iterations. In the figure NS means not significant.

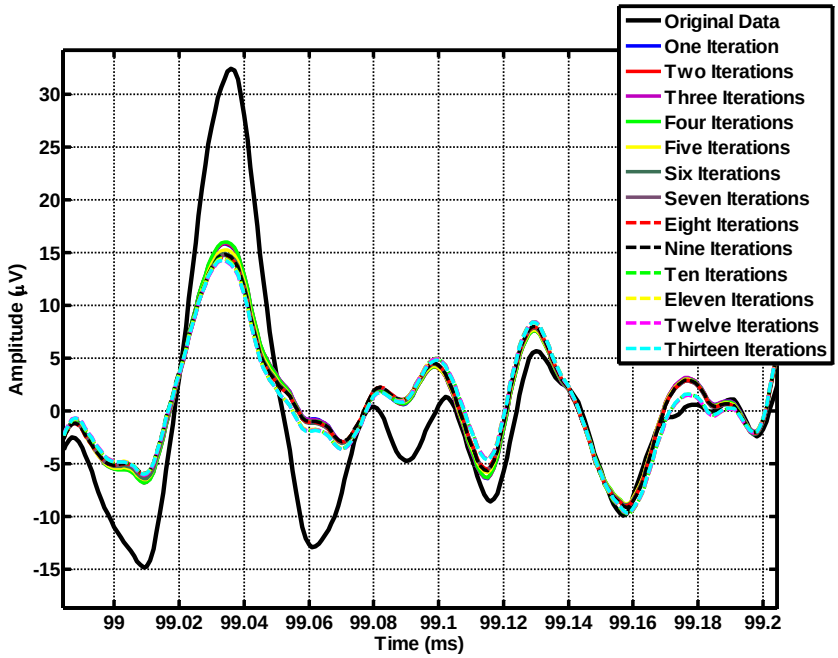

Figure 6.10: Cardiac artifact rejection from ALR by employing cICA in one to thirteen iterations. Change in the signal after the 1st iteration is small.

The whole procedure of artifact rejection was carried out for all the 10 subjects. The results showed that in using MSC-ICs for cardiac artifact rejection, the average number of the components which were significantly coherent with the heart beats across the ten subjects was found to be 5.

The quality of the artifact rejection was also measured for the artifact rejection methods for each of the ten participants and the median value of R_p was found to be 53.14% for artifact rejection by using MSC-ICs and 67.05% for when cICA was employed for cardiac artifact rejection (with one iteration). This result is shown in Figure 6.9. Result of applying a Wilcoxon test showed that rejection ratio is significantly (p-value =0.007) higher when cICA is employed for artifact rejection.

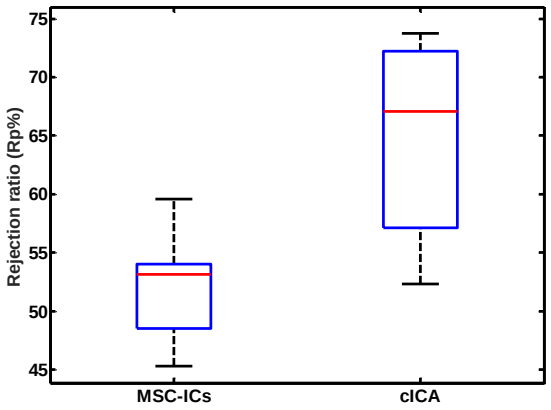

Figure 6.11: Comparing the rejection ratio for two artifact rejection methods. Rejection is significantly higher when cICA is employed.

Furthermore, for each subject the *Fmp* was calculated for the channels above 1.25 before and after artifact rejection (for both artifact rejection methods) and the results were compared via a paired sample t-test. In all the ten subjects a significant improvement (p-value<0.05) in *Fmp* was observed due to artifact rejection (for both cICA based and MSC-ICs based methods).

Correlation coefficients between artifact removed signal and the heartbeat reference were calculated for both artifact rejection methods for all the ten subjects for the channel which was most contaminated by heartbeats artifact. For artifact rejection using MSC-ICs, artifact removed signal was found to be uncorrelated with the heartbeat reference signal for all the ten subjects; while for artifact rejection using cICA, artifact removed signal was found to be correlated to the heartbeat reference

for 4 of the subjects. These results are summarized in Figure 6.5 and 6.7 for artifact rejection using MSC-ICs and cICA respectively.

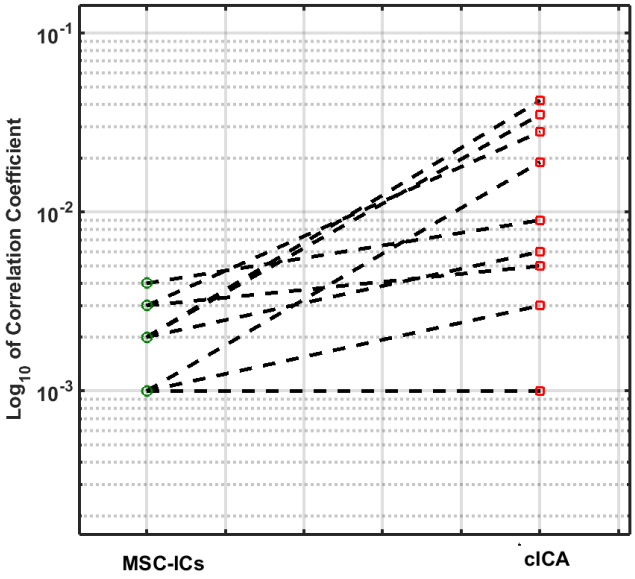

Figure 6.12: Correlation coefficient between artifact removed signal and the heartbeat reference for all the ten subjects for the most contaminated signal by heartbeat artifact for both artifact rejection methods. Correlation between artifact removed signal and heartbeat reference is significantly lower when MSC-ICs was used.

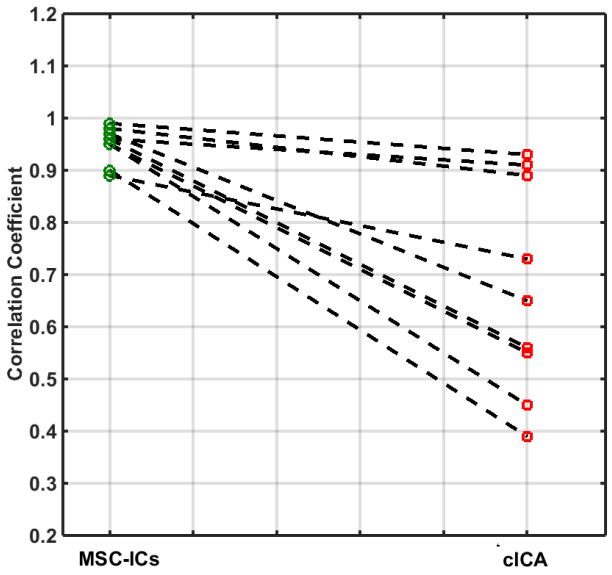

Figure 6.13: Correlation coefficient between artifact removed signal and the original signal for all the ten subjects for the most contaminated signal by heartbeat artifact for both artifact rejection methods. Correlation between artifact removed signal and original signal is significantly higher when MSC-ICs was employed.

Additionally, in order to illustrate that artifact removal by using either of the methods (cICA and MSC-ICs) does not disturb the ALR signal, un-weighted coherent average over 155 epochs was calculated at Cz for one subject and the result is shown in Figure 6.12. The original data, artifact removed signal using MSC-ICs and artifact removed signals using one and thirteen iterations of cICA are shown in Figure 6.12. For this subject, signals from one and thirteen iterations are overlapping since the change in the signal is small after the first iteration.

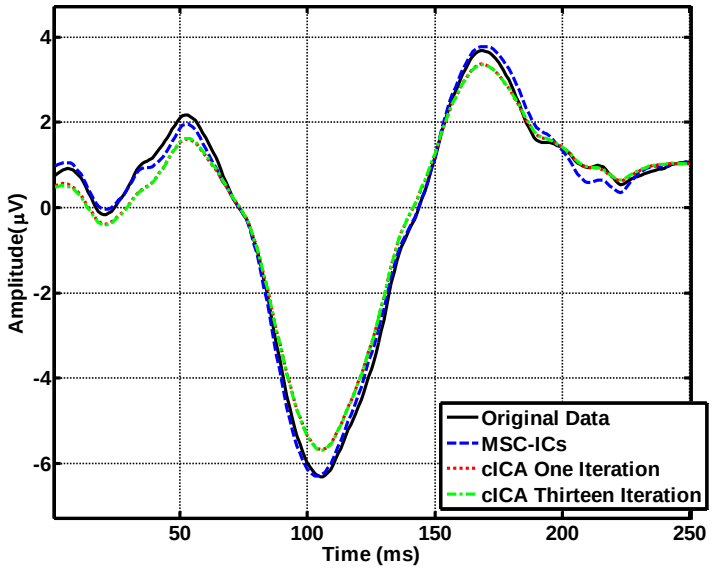

Figure 6.14: Un-weighted averaging over 155 epochs at Cz for one subject before and after cardiac artifact removal. Artifact rejection by using either of the methods does not have a large effect on the quality of the signal at Cz. The averaged ALR from one and thirteen iterations are overlapping.

6.4. Discussion and Conclusion

According to the results which were presented in the chapter, using MSC-ICs can significantly improve the cardiac artifact removal from the ALR recordings in comparison with the existing cICA based method. The average number of cardiac components found by MSC-ICs were 4; while this number was found to be 1 for the cICA based methods. However, possibly by dynamically changing some of the optimization parameters in the cICA artifact rejection algorithm such as the closeness or learning rate, this value and consequently the artifact rejection quality will increase. However, adjusting the parameters during the artifact rejection procedure de-automates the method. Since MSC-ICs removes all the cardiac components at

once, the artifact rejection by this method was considerably faster than going through cICA iterations. Furthermore, the need for selecting optimization parameters and the closeness criterion in cICA based artifact rejection method, is a major drawback of this method.

According to Figure 6.9, for cardiac artifact rejection the rejection ratio (R_p) was found to be significantly (p-value<0.05) higher when cICA was used in comparison with when MSC-ICs was employed. However, by comparing these results with those in Figures 6.10 and 6.11 it can be concluded that, in cICA cardiac artifact is not the only undesired component which was discarded. What has been discarded by cICA from the signal was partly heartbeats and partly other components. From the results presented in Figures 6.12 and 6.13 for this set of data, it can be stated that, using MSC-ICs is a better choice than cICA for cardiac artifact rejection from ECG, i.e. in both senses of removing the heart beat artifact and preserving the signal of interest.

As was mentioned before, using MSC for artifact component selection does not have the drawbacks of amplitude based methods used by (Delorme et al. 2001; Barbati et al. 2004; Dammers et al. 2008; Klados et al. 2010). In addition, the MSC-ICs method can be employed for any type of repetitive artifact, e.g. eye-blink heartbeat. However, the need for a time reference and also setting a suitable threshold for selecting/discarding the components still remains as a drawback for the MSC-ICs method.

According to the results presented in this chapter, artifact rejection had a major impact on signal quality when the artefact was strong. However, in many channels the artifact was weak and the final impact of removing artifact or not on the coherent average was only small. In the current work the focus is on auditory evoked potentials, but the method of artefact rejection has much wider application, for example in spontaneous EEG or separating the maternal ECG from fatal ECG

6.5. Summary

In this chapter a novel ICA based artifact rejection method for cardiac artifact removal from the ALR was introduced. The performance of the proposed method was compared with an alternative method based on constrained ICA. The results of this comparison showed that the novel method was significantly better than the existing

method in terms of speed, time required for the processing procedure, and accuracy (quality of the artifact rejection).

Chapter 7

Time Reduction for the ALR Recording

7.1. Introduction

Increasing the efficiency of AER detection by means of multichannel EEG recordings has been the subject of much research (Reijden et al. 2004; Van Dun et al. 2009; James et al. 2005). As was explained in section 1.6, AERs are assumed to be the outcome of contributions from many sources in the brain in response to a repetitive stimulus. The position of these response generators is not known and may be different from subject to subject. Moreover, the strength of AER signal at each electrode is a function of distance of the generator source to the recording electrode (Bharadwaj 2014). It thus follows that by using an electrode array it is more likely that some electrodes are close to the generator sources and this allows an increase in the efficiency of detecting a response. A multichannel array can also be used to find the electrode configuration which offers the highest SNR values (Dun 2009).

Placing a high density electrode array, e.g. 64, 128 or 256, on the scalp of subjects is one of the most time consuming parts of multichannel recording of AERs (ALR in this case). Experience has shown that placing 64 electrodes on the scalp takes about an hour for a trained technician. Therefore, finding a method by which

the ALR can be recorded with fewer electrodes without a significant decrement in SNR would be highly beneficial for practical applications.

Many research studies have focused on reducing the required time for auditory steady state response (ASSR) measurements by optimizing the ASSR detection. Improvement in detecting ASSR can be achieved via changing the stimulus type (John et al. 2001b; Stürzebecher et al. 2001; John et al. 2002; John et al. 2003; Riquelme et al. 2006; Stürzebecher et al. 2006) or alternatively increasing the efficiency of ASSR detection by using multichannel recordings of ASSR (van der Reijden et al. 2004; Van Dun et al. 2007a, 2007b). A number of authors (John et al., 2001a; Lins and Picton, 1995; Luts and Wouters, 2005) have shown that the average hearing threshold level measured by ASSR for normal hearing adults is lowest, i.e. closest to 0 dB nHL, if the derivation of electrodes is chosen well, typically Cz-Oz or Cz-neck. However, the electrode derivation which produces the highest SNR is subject dependent and cannot be predicted in advance. Hence, using multichannel recording, combined with appropriate multi-channel signal processing, would seem a promising approach to select the "best" channel or combination of channels for as many subjects as possible.

Dun (2009) employed a multichannel ASSR detection strategy based on a statistical method to find the optimal configuration of the electrodes for recoding ASSR with the highest SNR. In their work they used an array of 8 electrodes and calculated the SNR obtained by all the possible configurations. For multichannel recording the highest SNR was achieved when the electrodes were placed on both mastoids and the back of the head. However, the SNR which was achieved by placing a single channel on the back of the head (with weighted averaging as a processing tool) produced the highest SNR and no significant improvement was reported due to multichannel recording. A drawback of Dun's method is that for a high density electrode array the number of possible configurations and consequently number of calculations is very large.

In contrast with the results reported by Dun (2009), Bharadwaj (2014) used 32 channels ASSR and showed that the SNR of ASSR measured at Cz increases by increasing the number of channels. The increment in SNR was reported to be steep at the beginning and then plateaued. The observed plateau in SNR increment can be explained by the incremented number of electrodes: by increasing the number of

electrodes the distance between two adjacent electrodes decreases and two electrodes may carry the same information. Moreover, for the case in which one channel has a good sensitivity to the auditory stimulus (strong response) while another has a poor sensitivity (weak response), when these two channels are combined with similar weights, though the noise may be partially cancelled, the signal may also be diminished by the inclusion of the channel with poor sensitivity. The sensitivity of different channels to the signal also depends on the choice of the reference and the tissue geometry of individual subjects (Bharadwaj 2014). Although Bharadwaj (2014) showed that the SNR increases by increasing the number of the channels, no optimal electrode placement was suggested.

To the best of our knowledge, time reduction for ALR using a multichannel approach has not been investigated previously. Therefore, in this chapter, a multichannel signal processing method is employed for reducing the recording time for the ALR. The same data set used in previous chapters is also used for this application. Time reduction can be carried out in two main senses: 1. reducing the number of stimulus repetitions. 2. reducing the set up time by optimizing the position and the number of the recording electrodes (within a limited region) in multichannel recording. Since the aim is to know if multichannel signal processing is beneficial over single channel alternative, the point of reference is single channel recording. In previous chapters the benefit of multi-channel recordings with 64 channels over single-channel recordings (Cz) was shown. The current chapter will extend this to assess the impact of using fewer channels on the result. Hence, for finding the optimum number of stimulus repetitions for multichannel AER recording, the number of the repetition will be reduced till the same SNR as single channel is obtained at vertex (Cz channel). In similar fashion as previous chapters *Fmp* was used to calculate SNR. On the other hand for optimizing the number and positions of electrodes, the target is to have the same *Fmp* at the Cz as when all 64 channels are used, but with fewer electrodes. The goal is to find an optimal ALR recording protocol, including optimal electrode placement and optimum number of stimulus repetitions. The latter goal will be addressed first.

7.2. Time Reduction by Reducing the Number of Epochs

Chapter 2 detailed how the ALR is generally recorded from 155 stimulus repetitions for a single channel recording. A similar number of stimulus repetitions (155) (at 60dB nHL) was used for multichannel recording of the ALR in each of the 10 subjects. The main goal of this section is to reduce the time of multichannel recording of ALR. In this part the single channel data (Cz) was de-noised by employing un-weighted averaging and weighted averaging methods and the *Fmp* were calculated for each of these methods. The value of *Fmp* at Cz calculated by un-weighted averaging was used as a point of reference for comparison within the multichannel analysis. The epoch reduction procedure (reducing the number of epochs and assessing the change in *Fmp* with progressively lower numbers of epochs) was started with full length data (155 epochs) and reduced to 1 epoch in 154 iterations. In each iteration the MSC-ICs method (which was explained in section 3.2.4) was used for multichannel noise reduction and the *Fmp* was calculated at Cz. The aim was to determine the minimum number of epochs by which the *Fmp* at Cz was found to be identical to the one calculated by un-weighted averaging using all 155 epochs. This procedure was carried out for all the ten normal hearing subjects. The minimum number of epochs was found for each subject for having *Fmp* equal to single channel processing at Cz. To encompass all the possible situations for this data set, the largest number (worst case) was selected as the optimum required number of epochs for multichannel recording of the ALR. Since applying ICA in each step (154 times in total) is very time consuming, the ICs in each step were found through an alternative way. With the assumption that the spatial distribution of the sources (positions of the sources) is unchanged over the scalp, the mixing matrix which was found from applying ICA over all data from each subject (155 epochs) was used for finding the ICs in the next steps. After finding the ICs using this mixing matrix, MSC-ICs was used for in the same fashion as previous chapters and ICs which were significantly (p-value <0.05) coherent with the stimulus were kept as response ICs and the rest were discarded as noise. Data was reconstructed using only the response components. This procedure was carried out for all ten subjects and the minimum number of epochs was calculated for each subject. Finally the largest value (of the minimum epoch numbers calculated for each subject) across the ten subjects was

considered the required number of epochs by which multichannel processing produces signal with the same *Fmp* as single channel processing at Cz when155 epochs are used.

7.3. Time Reduction by Optimal Electrode Placement

The aim of this section is to optimize the number and location of the electrodes in multichannel recording of ALR. The procedure of reducing the number of recording electrodes was carried out in a way that that gives the same result as the single channel approach. The main question of this section is to find the minimum number of channels and their location that achieves the same *Fmp* as single channel weighted averaging at Cz. In this method, first the ALR data captured from the 10 subjects was decomposed into the ICs. Then the response components were selected by using the MSC-ICs method (explained in detail in chapters 3 and 4) for all the subjects. As was mentioned in section 1.6 the ALR observed at each electrode is a linear mixture of independent sources (response and noise) in the brain. In equation (1.12), the un-mixing matrix shows the position of the electrodes which have the main contribution in each IC. Hence, from the un-mixing matrix the electrodes which were the main contributors in response ICs (selected by MSC-ICs) were found for each subject (Jung et al. 2001; Makeig et al. 2002; Zhukov et al. 2000). The spatial distributions of the ICs (selected by MSC-ICs) were found by using the topographic map. In this work a MATLAB toolbox EEGLAB 11-0-0-0b was used for producing the topographic maps. Since the positions of the response sources are subject dependent, it is not possible to suggest a unique electrode placement which can be used for all the subjects. However, it is possible to find a common area of the scalp from which the response components mostly come. For each individual the average over the spatial distribution of the response ICs was calculated. This provides a map from which the response components mostly come from for each subject. Then, the average over the spatial distributions (obtained from the individuals) can be considered as the common area from which the response ICs come from. If the electrodes are placed on the scalp areas best associated with each IC. Response ICs will be the dominant ICs in recordings from these areas, making a reduction of these redundant electrodes possible. In the next stage, first the *Fmp* at Cz for the case in which all 64 channels were used was selected as a basis for comparison for each

subject. Then, the *Fmp* was calculated at Cz for the case in which only the reduced number of channels were placed on the area found in the previous part. Then to explore any significant difference between the *Fmp*s (obtained at Cz) by different electrode placement strategies (optimal electrode placement vs. placing all 64 electrodes), the results were compared with each other using a statistical test.

7.4. Results

Initially, the *Fmp* was calculated at Cz for both un-weighted and weighted averaging methods for 155 epochs for all the subjects. The result of applying the epoch reduction procedure on all the ALRs captured from the subjects showed that on average, the minimum required number of stimulus repetitions that produce an *Fmp* equal to single channel processing at Cz was found to be 70 (with standard deviation equal to 5) for un-weighted averaging and 84 (with variance equal to 4) for weighted averaging. The higher value for the latter reflects the increased SNR and hence *Fmp* expected when using weighted averaging. It may therefore be concluded that by using multichannel processing the required time for recording ALR is reduced by 45% when compared with weighted averaging and 55% for un-weighted averaging in order to achieve the same *Fmp*. In chapter 2, it was stated that recording ALR takes 4 minutes for 155 epochs which can thus be reduced to 2 minutes and 12 seconds, and 1 minute and 48 seconds, respectively, with multichannel. Variation of *Fmp* with number of epochs for one of the subjects is shown in Figure 7.1. In this figure, the *Fmp* values calculated at Cz for both un-weighted an weighted averaging using all 155 epochs are also shown. For this subject, the *Fmp* at Cz produced by MSC-ICs was found to be equal to the *Fmp* after 81 epochs and for un-weighted averaging this occurred after 76 epochs. Note that the MSC-IC line may cross the threshold repeatedly, but the number of epochs considered was the value above which the plot consistently remained above the threshold.

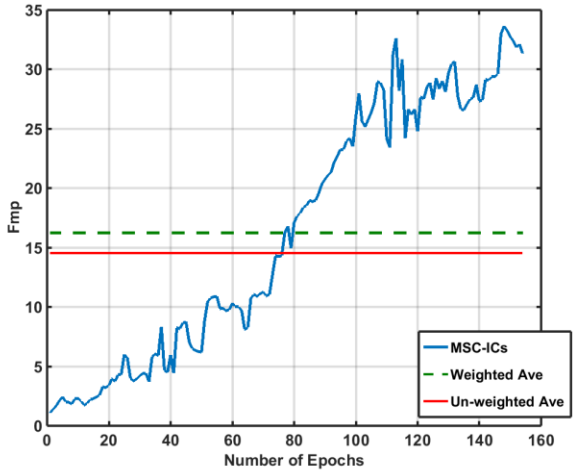

Figure 7.1: *Fmp* at Cz calculated by MSC-ICs (recalculated with increasing number of epochs), un-weighted (calculated for 155 epochs) and weighted (calculated for 155 epochs) averaging for one subject. MSC-ICs produced a signal with the same quality as un-weighted and weighted averaging at Cz with 81 and 75 epochs less respectively.

The averaged ALRs produced by single channel (un-weighted and weighted averaging) and multichannel (MSC-ICs) at Cz for the same subject are shown in Figure 7.2. As it can be deduced from the figure, by using MSC-ICs the ALR can be recovered from noisy recordings at Cz by 81 epochs instead of 155 epochs.

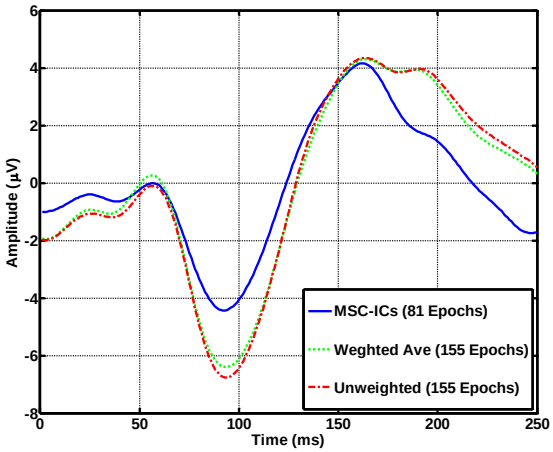

Figure 7.2: Averaged ALR at Cz produced by MSC-ICs (81 epochs) and un-weighted and weighted averaging (155 epochs).

As was mentioned before, the number and position of the response generators is subject dependent. The number of significant (according to the MSC criterion) ICs were found for all the ten subjects and the largest number was 40.

Furthermore, in order to find the areas of the cortex from which the response components come, the spatial distributions of the independent components over the scalp were investigated for all the 10 subjects and shown by a scalp topographic map with the corresponding coherent average (155 epochs) for that IC (Figure 7.4 shows one example). The MSC-ICs was employed and ICS which were significantly coherent with the stimulus (response components) were identified.

The average over spatial distributions of the response ICs selected by MSC-ICs for each individual is shown in Figure 7.3.

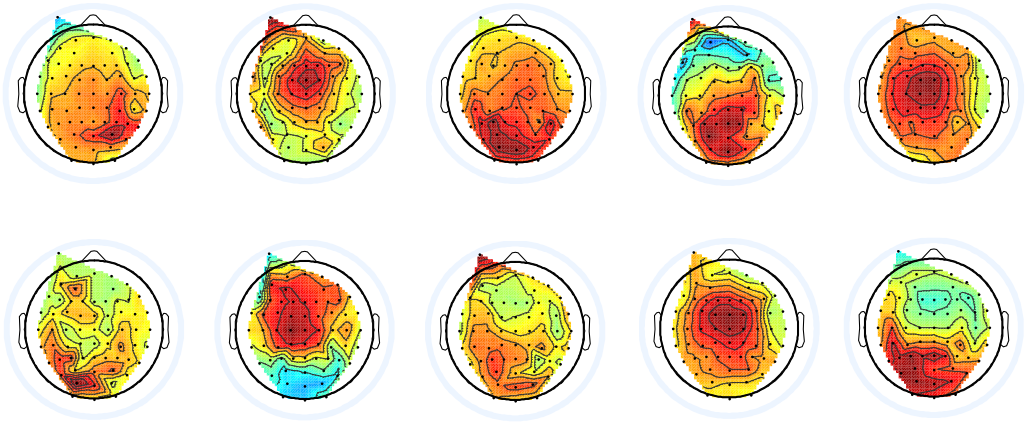

Figure 7.3: The average of spatial distribution of the response ICs selected by MSC-ICs for all the ten subjects

The results of averaging over un-mixing matrices obtained from the subjects showed that, the response components mostly come from the back and top of the head. This result is shown for the channel locations shown in Figure 7.4.

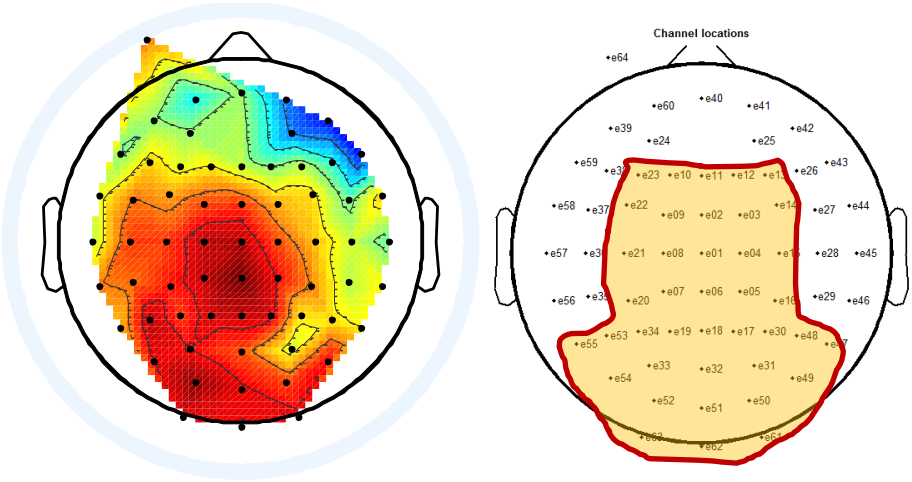

Figure 7.4: Left, the average over spatial distribution of the response ICs obtained from ten subjects. On average for the 10 subjects, the response components mostly come from the back and top of the head (highlighted area on the right).

Forty electrodes were placed on the highlighted area of the scalp shown in Figure 7.4 and the MSC-ICs method was employed using multichannel processing. For optimal electrode placement the *Fmp*s were calculated at Cz for all the subjects and compared to the *Fmp* values obtained when all 64 channel were used. This comparison is shown in Figure 7.5.

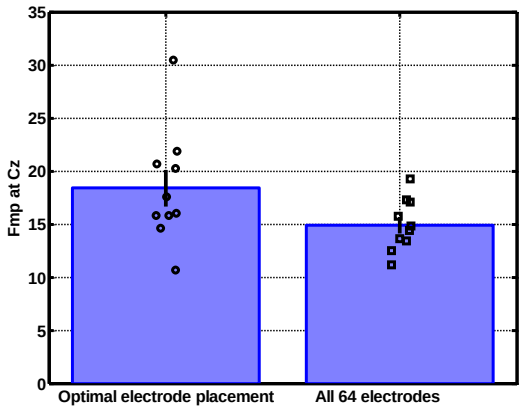

Figure7.5: The *Fmp* at Cz obtained by MSC-ICs for optimally placing 40 channels and when all 64 channels were used.

Although, the *Fmp* seems to be increased by using optimal electrode placement, the result of applying a t-test showed that the difference between *Fmp*s obtained at Cz through the different electrode placement strategies was not significant

(p-value > 0.05). However, reducing the number of electrodes from 64 to 40 (37.5%) can considerably reduce the time required for placing electrodes on subjects' scalps.

7.5. Discussion and Conclusion

Reducing the time of recording for ASSR has been the subject of some researches (John et al. 2001b; John et al. 2002; John et al. 2003; Stürzebecher et al. 2001; Riquelme et al. 2006; Stürzebecher et al. 2006). Despite the methods by which the recording time for ASSR was reduced, the stimulus stayed unchanged in this work. The results of this chapter illustrated that the required time for recording ALR is considerably reduced (without a change in *Fmp*), if multi-channel recording and MSC-ICs are employed. On average, 85 stimuli less were required for the same signal quality (*Fmp*) with multichannel processing compared to un-weighted averaging and 74 less, when comparing with weighted averaging, when for the latter two used 155 epochs. This corresponds to a 55% reduction in time required for recording the ALR. However, placing 64 electrodes on a subjects head is a demanding procedure that adds about an hour to the recording session. Clearly, using multichannel recordings to improve the *Fmp* at Cz will only be effective in terms of time spent on assessment, if the recording sessions are quite long.

Despite previously reported results (Dun 2009) that multichannel recording was not helpful for reducing the time of recording for the ASSR, the results of optimal electrode placement for the ALR showed that by placing 40 electrodes on the area given by Figure 7.4, the *Fmp* at Cz will stay unchanged in comparison with when all the 64 electrodes were used, but fewer electrodes are needed. According to Bharadwaj (2014) *Fmp* increases by increasing the number of channels. It is worth noting that the by reducing the number of channel form 64 to 40, the *Fmp* was slightly increased. However, for this data set this increment was not found to be significant. Given that the selected electrodes were the ones deemed to contribute most strongly to the ALR, the channels removed are probably ones with a weak or absent response, which contribute primarily noise to the analysis (Bharadwaj 2014).

7.6. Summary

The focus of this chapter was reducing the ALR recording time. The time reduction was carried out in two forms: 1) Reducing the number of stimulus repetition required for recording ALR. 2) Reducing the number of recording electrodes in order to reduce the time for subject preparation. Results showed that, by using multichannel processing, the ALR can be recorded from the vertex without loss in signal quality 55% faster (fewer stimulus repetitions) in comparison with the single channel alternatives. Moreover, an optimal electrode configuration was suggested for multichannel recording of ALR. By using the suggested electrode configuration the number of electrodes needed can be reduced by 37.5%, thus reducing time in subject preparation.

Chapter 8

Discussion and Conclusion

For the four noise reduction methods presented based on selecting ICA components, it was demonstrated that all the methods are able to reconstruct the response form the noisy signal. For the data set used in this work the results applying the methods on both simulated data and real data showed that the MSC-ICs and the Max-Fmp-ICs have a significantly better performance than the Max-Kurt-ICs and Min-Entropy-ICs in SNR improvement, which are the more conventional approaches. According to the results, MSC-ICs and Max-Fmp-ICs were not found to be significantly different in terms of SNR improvement. It may be possible to see significance between these two methods with a larger sample size. However, a difference that needs a very large sample size to be detected is probably not of great practical importance, in this case. For using the MSC-ICs and the Max-Fmp-ICs, extra information about the signal is needed, including crucially stimulus onsets; while the Max-Kurt-ICs and the Min-Entropy-ICs do not need this prior information. This additional information included with the former methods is probably the key factor in achieving the improved performance. All amplitude based methods for component selection have the problem that when the response amplitude is small response detection will be either difficult or impossible; while MSC can detect the consistent is less sensitive to the amplitude of the response. Moreover, the Max-Kurt

is sensitive to the outliers and it is not a suitable choice for the cases in which the undesired signal has large kurtosis (Super Gaussian distribution). Setting a suitable threshold for rejecting or accepting an IC is another difficulty of the all the introduced methods.

Since, in biomedical data (especially for brain data) the variance of the noise is usually much higher than the variance of the signal of interest, it can be predicted that PCA might be able to enhance the signal quality, i.e. in terms of SNR improvement. Although PCA is computationally less complex than ICA (and consequently faster), the results of applying PCA on both simulated data and real data showed that for this case ICA based noise reduction method have a better performance than PCA based alternative in SNR improvement.

Moreover, the MSC was not found to be helpful in automatic selection of components when PCA is employed as a source separation tool. In the simulation even the noise signals were found to be significantly coherent with the stimulus. The reason can be the signal residue in noise components (showed via simulation in chapter 3). The results of employing PCA (instead of ICA) showed that the performance of MSC-PCs was poorer than single channel processing (weighted averaging). The results given by Figure 4.10 and 4.11 showed, PCA-eigenvalue ordering was not found to be significantly different with weighted averaging in providing high *Fmp*s. However, it can be stated that PCA-Eigenvalue ordering has a somewhat better performance than the weighted averaging since it provides more channels above the critical value.

According to the results shown it can be stated that multichannel processing of the ALR is a significantly better option than singe channel processing alternatives. Nevertheless, taking the time required for subject preparation and also the noise reduction procedure into account, single channel processing may still be considerably faster and more practical than multichannel alternatives. For the cases in which the signal quality has the first priority, multichannel recording is a better option. It is worth mentioning that, as was shown in previous works (Dobie and Wilson 1996; Dimitrijevic et al. 2001), weighted averaging is significantly better for SNR improvement than artifact rejection, i.e. removing the epochs which are highly contaminated by noise from the data before averaging. Furthermore it has been shown in the current work that the MSC-ICs and Max-Fmp-ICs are significantly better than

weighted averaging both in obtaining highest SNR and a higher number of the channels above the critical value. Therefore, it can be concluded that they are also better than artifact rejection.

While the benefits of the methods have been clearly shown, there are a number of unresolved issues remaining. For instance, in three of the methods (Max-Fmp-ICs, Min-Entropy-ICs and Max-Kurt-ICs) the thresholds for discarding/accepting an IC in each method were estimated from additional recordings (resting EEG). Moreover, ICA is computationally complex and for large data the noise reduction procedure is time consuming. Additionally, the required time for subject preparation in multichannel recording is much longer than for the single channel set up, and this is probably the main issue limiting the performance of the multichannel method.

Although the pure tone audiometry is more sensitive than cortical audiometry, cortical hearing threshold measurement can be useful for the cases in which the subjects are not able to attend to the behavioural audiometry test, e.g. when the subject is in coma. The results presented in this work showed that multichannel signal processing considerably improves the sensitivity of hearing threshold measurement in comparison with the single channel alternative.

The results suggest that the spatial distribution information of the ICs obtained from the ALR recorded at 60 dB nHL can be used for nose reduction and response detection for the ALRs recoded at lower levels of stimulus intensity. This indicates that ICA carried out once at the highest intensity is sufficient (for providing de-mixing matrix) and this computationally expensive step need not be repeated. Although the hearing threshold was estimated closer to 0 dB nHL when the MSC-Spatial dist. method was employed instead of MSC-ICs, the results of statistical comparison did not confirm any significant different between the outcomes of the two methods. Significant improvement might have been observed with a larger sample size – such as the study is suggested below for the continuation of the current work. Nevertheless, it can be stated that the MSC-Spatial distribution might be preferred in finding the hearing threshold level since it is much faster than the MSC-ICs method. The average hearing threshold level found by MSC-ICs and MSC-Spatial dist. for the normal hearing subjects were found 4 dB nHL and 2.5 dB nHL above the pure tone audiometry. This is considerably better than the 25 dB nHL reported by others (Lightfoot 2010) when ALR was used for hearing threshold measurement.

Inter and intra subject variability were detected by multichannel processing as well as the single channel alternative. However, no evidence was found which implies that the variability observed by the methods was because of variability in the ALR waveform i.e. the variability can be because of poor signal quality and too much noise.

Both weighted averaging and MSC-ICs found a significant difference in the amplitude of the ALR wave in different recording sessions for a subject and also across the subjects. Therefore, it can be stated that for this work multichannel processing was not advantageous in comparison with single channel processing.

Moreover, the habituation effect was not observed through either of the methods since no consistent decrement was observed in the *Fmp* or amplitudes of the major peaks of the ALR over 1^{st}, 2^{nd} and 3^{rd} third of the ALR waveform. The reason can be traced in the length of the recorded signals, quality of the recorded signals or the method by which the habituation was tested. By using better recording equipment or a different data testing strategy, e.g. used by (Zhang 2009), it may be possible to see the effect of habituation on the recorded data. The reason that Zhang testing strategy (or generally methods similar to Zhang's method) was not employed for this work was that we wanted to explore if MSC-ICs is able to detect the habituation effect via simple testing strategies (conventional ALR recording).

Both weighted averaging and the MSC-ICs method were applied on the ALR recorded form the normal hearing subjects under attend and ignore conditions and the effect of attention was investigated by comparing the amplitude of the major peaks of the ALR captured under attention and ignore circumstances. Both of the methods showed a significant enhancement in P2 due to the attention. The Amplitude of N2 was also found to be significantly increased when MSC-ICs was employed. Contrary to the results reported by others (Picton and Hillyard 1974; Thornton et al. 2007), no significant enhancement was observed in N1. It is worth mentioning that one common factor in all these studies was the use of selective attention, while in this work attention was not selective. A significant enhancement was observed in N1, when the data was collected under selective attention (Thornton et al. 2007).

Since the results of investigating habituation effect on the ALR did not confirm any consistent trend in the amplitudes of the major peaks, the observed significant enhancement in P2 and N2 are due to the attention, i.e. this enhancement is not due to

inter/intra subject variability as inter and intra subject variability does not follow a consistent increment. Form the results presented in this chapter it can be concluded that detecting the effect of attention on the ALR can be significantly improved by using multichannel processing rather than the single channel alternative.

According to the results, using MSC-ICs can significantly improve the cardiac artifact removal from the ALR recordings in comparison with the existing cICA based method. The average number of cardiac components found by MSC-ICs were 4; while this number was found to be 1 for the cICA based methods. However, possibly by dynamically changing some of the optimization parameters in the cICA artifact rejection algorithm such as the closeness or learning rate, this value and consequently the artifact rejection quality will increase. However, adjusting the parameters during the artifact rejection procedure de-automates the method. Since MSC-ICs removes all the cardiac components at once, the artifact rejection by this method was considerably faster than going through cICA iterations. Furthermore, the need for selecting optimization parameters and the closeness criterion in cICA based artifact rejection method, is a major drawback of that method.

The artifact rejection ratio (R_p) was found to be significantly (p-value<0.05) higher for cICA in comparison with MSC-IC. However, by comparing these results with those in Figures 6.12 and 6.13 it can be concluded that in cICA cardiac artifact is not the only undesired component which was discarded. What has been discarded by cICA from the signal was partly heartbeats and partly other components. From the results presented in Figures 6.12 and 6.13 for this set of data, it can be stated that using MSC-ICs is a better choice than cICA for cardiac artifact rejection from ECG, i.e. in both senses of removing the heart beat artifact and preserving the signal of interest. Using MSC for artifact component selection does not have the drawbacks of amplitude based methods used by others (Delorme et al. 2001; Barbati et al. 2004; Dammers et al. 2008; Klados et al. 2010). In addition, the MSC-ICs method can be employed for any type of repetitive artifact. However, the need for a time reference and also setting a suitable threshold for selecting/discarding the components still remains a drawback for the MSC-ICs method. Artifact rejection had a major impact on signal quality when the artefact was strong. However, in the current study, in many channels the artifact was weak and the final impact of removing artifact or not on the coherent average was only small.

Considering that when MSC-ICs is employed as noise reduction method for the assessment of evoked potentials, all the ICs which are not significantly coherent with the stimulus(including heartbeat components), will be discarded, there is no need for an additional step specifically aimed at rejecting ICs strongly linked to the heart-beats. However, the method of artefact rejection has much wider application in biomedical signal processing, when artefacts specifically linked to a time-marker should be removed, for example separating the maternal ECG from fatal ECG, or in removing eye-movement artefacts in the EEG (based on time-markers from the electrooculogram).

Reducing the time of recording for ASSR has been the subject of some research (John et al. 2001b; John et al. 2002; John et al. 2003; Stürzebecher et al. 2001; Riquelme et al. 2006; Stürzebecher et al. 2006). Despite the methods by which the recording time for ASSR was reduced, the stimulus stayed unchanged in this work. The results illustrated that the required time for recording ALR (including the recording the time for putting the scalp cap on) is considerably reduced (without a change in *Fmp*), if multi-channel recording and MSC-ICs are employed in comparison with single channel processing. On average, 85 stimuli less were required for the same signal quality (Fmp) with multichannel processing compared to un-weighted averaging and 74 less, when comparing with weighted averaging, when for the latter two used 155 epochs. This corresponds to a 55% reduction in time required for recording the ALR. However, placing 64 electrodes on a subjects head is a demanding procedure that adds about an hour to the recording session. Clearly, using multichannel recordings to improve the *Fmp* at Cz will only be effective in terms of time spent on assessment, if the recording sessions are quite long.

In addition, it was shown that by reducing the number of channels form 64 to 40 and placing the electrode on the area given in chapter 7, the *Fmp* remained unchanged. Given that the selected electrodes were the ones deemed to contribute most strongly to the ALR, the channels removed are probably ones with a weak or absent response, and thus contributed primarily noise to the analysis (Bharadwaj 2014). Although the results of this work showed that by placing the electrodes on the highlighted area of the scalp (shown in chapter 7 Figure 7.4) same *Fmp* at Cz can be obtained, the minimum number electrodes and their exact positions in the area was not specified and should be the subject of future work.

Chapter 9

Further Research Suggestions

This chapter presents several suggestions for further research concerning multichannel processing of auditory evoked potentials (AER).

9.1. Finding the Optimal p-value for Rejecting/Keeping the ICs in MSC-ICs

In this work, the p-value for significant coherence between an IC and the stimulus was selected to be 0.05. However, this is not necessarily the optimal value which should be used for all the data sets recorded form every individual. In order to make the MSC-ICs more practical the best p-value should be calculated or alternatively a complimentary method which can calculate the optimal p-vale for each data set should be added to the method.

9.2. Comparing the MSC-ICs with Alternative AER Processing Methods

The performance of MSC-ICs for SNR improvement for AER recordings should be compared with other alternative methods of AER detection such as the t-test, T^2 Hotelling (for time domain based methods), or frequency based methods like

empirical mode decomposition or wavelets (Quiroga et al. 2001; Bradly and Wilson 2004; Cek et al 2010; Kelly et al. 2011).

9.3. Measuring the Hearing Threshold in Other Frequencies

In this work the hearing threshold was measured for a tone burst at 1 kHz. The validity of the MSC-ICs can be evaluated for other frequencies, i.e. 250, 500, 2 kHz, 4 kHz and 8 kHz.

9.4. Using MSC-ICs to Detect the Effect of Selective Attention on the ALR

As was pointed out in the previous chapters, in this work the increment of the N1 peak due to the attention was not detected by the MSC-ICs method. However, applying the MSC-ICs on the data recorded under selective attention condition might detect the effect of attention in N1 as well as P2 and N2.

9.5. Optimizing the Electrode Placement

Further time reduction by reducing the time required for electrode placement. This can be carried out by finding the minimum number of electrodes and their exact positions on the scalp area suggested in chapter 7.

9.6. Evaluating the MSC-ICs for Different AERs

The MSC-ICs was only applied on the ALR. However, the performance of MSC-ICs can be evaluated for other AERs such as auditory middle latency response (AMLR) and confirm whether the benefits of the method observed for ALR can also be seen in other AERs.

9.7. Evaluating the MSC-ICs on the data recorded from the patients

All the proposed methods in this work were only applied on the data recorded from normal hearing subjects. To make the MSC-ICs more practical for clinical applications, it should also be evaluated by being applied on data recorded form subjects with hearing impairment.

9.8. Separating Fetal and Maternal Heartbeats

The artifact rejection method based on the MSC-ICs (proposed in chapter 6) can be employed to separate the maternal heartbeat from fetal heartbeats. In this case the data (which is a mixture of fetal and maternal heartbeats and noise) can be recorded via an array of abdominal electrodes. The maternal heartbeat signal can be selected as the reference. The ICs which are significantly coherent with the reference should be discarded and the data will be reconstructed by using the rest of the ICs.

Appendix I

Lagrange optimization method can be employed to maximize a function like *f(w)*, here $\boldsymbol{w_1}^{\boldsymbol{T}}\boldsymbol{C_x}\boldsymbol{w_1}$, with the constraint *g(w)=c* (here *g(w)*= $||\boldsymbol{w_1}||^2$and c=1). Note that $||\boldsymbol{w_1}||^2=\boldsymbol{w_1}^T\boldsymbol{w_1}$. By re-arranging the constraint equation in form of *g(w)-c*=0 and selecting Lagrange multiplier λ, the objective function Lagrange equation can be written as following:

$$L(W,\lambda) = f(W) - \lambda(g(W) - c) \tag{A1.1}$$

To maximize equation (A1.1), differentiation of the objective function is calculated respect to both arguments (*W* and) and the derivatives are set to zero.

$$\frac{\partial u}{\partial W} = 0 = \frac{\partial f}{\partial w} - \lambda\frac{\partial g}{\partial W} \tag{A1.2}$$

$$\frac{\partial u}{\partial \lambda} = 0 = -(g(w) - c) \tag{A1.3}$$

Equation (A1.3) gives back the constraint equation.

For this problem we have:

$$L(W,\lambda) = \boldsymbol{w_1}^{\boldsymbol{T}}\boldsymbol{C_x}\boldsymbol{w_1} - \lambda(\boldsymbol{w_1}^T\boldsymbol{w_1} - c) \tag{A1.4}$$

$$\frac{\partial u}{\partial W} = 2\boldsymbol{C_x}\boldsymbol{w_1} - \boldsymbol{2}\lambda\boldsymbol{w_1} = 0 \tag{A1.5}$$

$$\boldsymbol{C_x}\boldsymbol{w_1} = \lambda\boldsymbol{w_1} \tag{A1.6}$$

The desired vector $\boldsymbol{w_1}$ is an eigenvalue of the covariance matrix $\boldsymbol{C_x}$ and the maximizing vector will be the one associated with the largest eigenvalue.

References

A. Hyvärinen, J. Karhunen and E. E. Oja (2001). Independent Component Analysis. USA, Wiley-Interscience Publication.

Acinodotr, N. (2013). "Estimation of brainstem auditory evoked potentials using a nonlinear adaptive filtering algorithm." Neural Computing and Applications **22**(6): 1201-1209.

Allen, P. J., G. Polizzi, K. Krakow, D. R. Fish and L. Lemieux (1998). "Identification of EEG events in the MR scanner: The problem of pulse artifact and a method for its subtraction." Neuroimage **8**(3): 229-239.

Amenedo, E. and F. Diaz (1998). "Effects of aging on middle-latency auditory evoked potentials: A cross-sectional study." Biological Psychiatry **43**(3): 210-219.

Anne Z. Saunders, A. V. S., and Nancy Lite Shuster (1990). Clinical Methods: The History, Physical, and Laboratory Examinations, Butterworth-Heinemann

Barbati, G., C. Porcaro, F. Zappasodi, P. M. Rossini and F. Tecchio (2004). "Optimization of an independent component analysis approach for artifact identification and removal in magnetoencephalographic signals." Clinical Neurophysiology **115**(5): 1220-1232.

Barkat, M. (1991). Signal detection and estimation, Artech House Ltd.

Beagley, H. A. and J. J. Knight (1967). "Changes in auditory evoked response with intensity." The Journal of laryngology and otology **81**(8): 861-873.

Bell, A. J. and T. J. Sejnowski (1995). "An iformation maximization approach to blind separation and blind deconvolution." Neural Computation **7**(6): 1129-1159.

Bellgrove, M. A., R. Hester and H. Garavan (2004). "The functional neuroanatomical correlates of response variability: evidence from a response inhibition task." Neuropsychologia **42**(14): 1910-1916.

Bharadwaj, H. M, B. G. Shin-Cunningham,"Rapid acquisition of auditory subcortical steady state responses using multichannel recordings, Clinical Neurophysiology" (2014)

Borodina, U. V. and R. R. Aliev (2013). "Wavelet spectra of visual evoked potentials: Time course of delta, theta, alpha and beta bands." Neurocomputing **121**: 551-555.

Breuer, L., J. Dammers, T. P. L. Roberts and N. J. Shah (2014). "A Constrained ICA Approach for Real-Time Cardiac Artifact Rejection in Magnetoencephalography." Ieee Transactions on Biomedical Engineering **61**(2): 405-414.

Butler, B. E. and S. G. Lomber (2013). "Functional and structural changes throughout the the auditory system following congential and early onset deafness: implications for hearing restoration." Frontiers in Systems Neuroscience **7**.

Bylund, J. (2001). "A humble attemptto solve the "Cocktail-party" proble Using Blind Source Separation." (University of Sydney).

Cao, T., F. Wan, C. M. Wong, J. N. da Cruz and Y. Hu (2014). "Objective evaluation of fatigue by EEG spectral analysis in steady-state visual evoked potential-based brain-computer interfaces." Biomedical Engineering Online **13**: 13.

Chapelle, O. and M. Wu (2010). "Gradient descent optimization of smoothed information retrieval metrics." Information Retrieval **13**(3): 216-235.

Chawla, M. P. S. (2011). "PCA and ICA processing methods for removal of artifacts and noise in electrocardiograms: A survey and comparison." Applied Soft Computing **11**(2): 2216-2226.

Chernick, M. R. (2008). Bootstrap Methods: A Guide for Practitioners and Researchers, Wiley.

Comon, P. (2006). "Ho fast is FastICA?" EUSIPCO.

Corr, P. J., ed. (2008). The Reinforcement Sensitivity Theory of Personality, Cambridge University Press.

Dagna, F., A. Canale, M. Lacilla and R. Albera (2014). "Tone burst stimulus for auditory brainstem responses: Prediction of hearing threshold at 1 kHz." Auris Nasus Larynx **41**(1): 27-30.

Dammers, J., M. Schiek, F. Boers, C. Silex, M. Zvyaginstev, U. Pletrzyk and K. Mathiak (2008). "Integration of Amplitude and Phase Statistics for Complete Artifact Removal in Independent Components of Neuromagnetic Recordings." Ieee Transactions on Biomedical Engineering **55**(10): 2353-2362.

Davies, M. E. and C. J. James (2007). "Source separation using single channel ICA." Signal Processing **87**(8): 1819-1832.

Davila, C. E. and M. S. Mobin (1992). "Weighted Averaging of Evoked Potentials." Ieee Transactions on Biomedical Engineering **39**(4): 338-345.

Davis, H. (1964). "Enhancement of evoked cortical potentials in humans related to a task requiring a decision." Science (New York, N.Y.) **145**(3628): 182-183.

Davis, H., C. Fernandez and D. R. McAuliffe (1950). "The excitatory process in the cochlea." Proceedings of the National Academy of Sciences of the United States of America **36**(10): 580-587.

Dawson, G. D. (1951). "A summation technique for detecting small signals in a large irregular background." The Journal of physiology **115**(1): 2p-3p.

Dawson, G. D. (1954). "A summation technique for the detection of small evoked potentials." Electroencephalography and Clinical Neurophysiology **6**(1): 65-84.

Ding, Z. and T. Nguyen (2000). "Stationary points of a kurtosis maximization algorithm for blind signal separation and antenna beamforming." Ieee Transactions on Signal Processing **48**(6): 1587-1596.

Dobie, R. A. and M. J. Wilson (1994). "Objective detection of 40 Hz auditory evoked potentials: phase coherence vs. magnitude-squared coherence." Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section **92**(5): 405-413.

Dobie, R. A. and M. J. Wilson (1996). "A comparison of t test, F test, and coherence methods of detecting steady-state auditory-evoked potentials, distortion-product otoacoustic emissions, or other sinusoids." The Journal of the Acoustical Society of America **100**(4): 2236-2246.

Don, M. and C. Elberling (1994). "EVALUATING RESIDUAL BACKGROUND-NOISE IN HUMAN AUDITORY BRAIN-STEM RESPONSES." Journal of the Acoustical Society of America **96**(5): 2746-2757.

Dun, B. V. (2008). Improving Auditory Steady State Response Detection Using Multiphase EEG Signal Processing PhD Thesis, Katholieke University.

Durrant, J. D., A. I. Tlumak, R. E. Delgado and J. R. Boston (2012). Steady state measurement and analysis approach to profiling auditory evoked potentials from short-latency to long latency, Google Patents.

Elberling, C. and M. Don (1984). "Quality Estimation Of Averaged Auditory Brain Stem Responses." Scandinavian Audiology **13**(3): 187-197.

Elberling, C. and O. Wahlgreen (1985). "Estimation Of Auditory Brain Stem Responses, ABR, By Means Of Baysian Inference." Scandinavian Audiology **14**(2): 89-96.

Escudero, J., R. Hornero, D. Abasolo, A. Fernandez and M. Lopez-Coronado (2007). "Artifact removal in magnetoencephalogram background activity with independent component analysis." IEEE Trans Biomed Eng **54**(11): 1965-1973.

Foster, M., J. Stevens and S. Brennan (2013). "Intra and intersubject variability in auditory steady-state response amplitude with high modulation rates to 1000 Hz amplitude modulated and tone pip stimuli." International Journal of Audiology **52**(7): 507-512.

Garell, P. C., M. F. Neelon and J. Williams (2006). "The effects of auditory attention measured from human electrocorticograms." Clinical Neurophysiology **117**(3): 504-521.

Gelfand, S. A. (2007). Hearing, an introduction to psychological and physiological acoustics, Informal healthcare USA, Inc.

Greco, A., N. Mammone, F. C. Morabito and M. Versaci (2005). Semi-Automatic Artifact Rejection Procedure based on Kurtosis, Renyi's Entropy and Independent Component Scalp Maps. Proceedings of World Academy of Science, Engineering and Technology, Vol 7. C. Ardil. Canakkale, World Acad Sci, Eng & Tech-Waset. **7:** 22-26.

Hall, J. W. (2007). New Handbook of Auditory Evoked Responses USA, Allyn & Bacon.

Hamalainen, M., R. Hari, R. J. Ilmoniemi, J. Knuutila and O. V. Lounasmaa (1993). "Magnetoencephalography -theory, instrumentation, and applications to noninvasive studies of the working human brain." Reviews of Modern Physics **65**(2): 413-497.

Hamaneh, M. B., N. Chitravas, K. Kaiboriboon, S. D. Lhatoo and K. A. Loparo (2014). "Automated Removal of EKG Artifact From EEG Data Using Independent Component Analysis and Continuous Wavelet Transformation." Ieee Transactions on Biomedical Engineering **61**(6): 1634-1641.

Hesse, C. W. (2008). Model order estimation for blind source separation of multichannel magnetoencephalogram and electroencephalogram signals. Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE.

Hillyard, S. A., R. F. Hink, V. L. Schwent and T. W. Picton (1973). "Electrical Signs of Selective Attention in Human Brain." Science **182**(4108): 177-180.

Hoke, M., B. Ross, R. Wickesberg and B. Lütkenhöner (1984). "Weighted averaging—theory and application to electric response audiometry." Electroencephalography and Clinical Neurophysiology **57**(5): 484-489.

Hood. Linda. j (1998). Clinical application of the auditory brainstem response

Hyman, L. H. (1992). Comparative vertebrate anatomy, University of Chicago.

Hyvarinen, A. (1998). New approximations of differential entropy for independent component analysis and projection pursuit. Advances in Neural Information Processing Systems 10. M. I. K. M. J. S. S. A. Jordan. **10:** 273-279.

Hyvarinen, A. (1999). "Fast and robust fixed-point algorithms for independent component analysis." Ieee Transactions on Neural Networks **10**(3): 626-634.

Hyvarinen, A. and E. E. Oja (1997). "A fast fixed-point algorithm for independent component analysis." Neural Computation **9**(7): 1483-1492.

Hyvärinen, A. and E. E. Oja (2000). Independent Component Analysis:Algorithms and Applications. Helsinki University of Technology, Neural Networks Research Centre.

Iriarte, J., E. Urrestarazu, M. Valencia, M. Alegre, A. Malanda, C. Viteri and J. Artieda (2003). "Independent component analysis as a tool to eliminate artifacts in EEG: A quantitative study." Journal of Clinical Neurophysiology **20**(4): 249-257.

Iyer, D., N. N. Boutros and G. Zouridakis (2002). "Independent component analysis of multichannel auditory evoked potentials." Conference Proceedings. Second Joint EMBS-BMES Conference 2002. 24th Annual International Conference of the Engineering in Medicine and Biology Society. Annual Fall Meeting of the Biomedical Engineering Society (Cat. No.02CH37392): 204-205 vol.215 vol.201.

James, C. J. and C. W. Hesse (2005). "Independent component analysis for biomedical signals." Physiological Measurement **26**(1): R15-R39.

James, C. J. and D. Lowe (2003). "Extracting multisource brain activity from a single electromagnetic channel." Artificial Intelligence in Medicine **28**(1): 89-104.

John, M. S., A. Dimitrijevic and T. W. Picton (2001). "Weighted averaging of steady-state responses." Clinical Neurophysiology **112**(3): 555-562.

Jung, T. P., S. Makeig, C. Humphries, T. W. Lee, M. J. McKeown, V. Iragui and T. J. Sejnowski (2000). "Removing electroencephalographic artifacts by blind source separation." Psychophysiology **37**(2): 163-178.

Kelly, J. W., D. P. Siewiorek, A. Smailagic, J. L. Collinger, D. J. Weber and W. Wang (2011). "Fully automated reduction of ocular artifacts in high-dimensional neural data." IEEE Trans Biomed Eng **58**(3): 598-606.

Kern, A., C. Heid, W.-H. Steeb, N. Stoop and R. Stoop. (2008, 29 Aug). "Uncoiled cochlea with basilar membrane." from http://upload.wikimedia.org/wikipedia/commons/6/65/Uncoiled_cochlea_with_basilar_membrane.png.

Klados, M. A., C. Bratsas, C. Frantzidis, C. L. Papadelis and P. D. Bamidis (2010). A Kurtosis-Based Automatic System Using Naïve Bayesian Classifier to Identify ICA Components Contaminated by EOG or ECG Artifacts. XII Mediterranean Conference on Medical and Biological Engineering and Computing 2010. P. Bamidis and N. Pallikarakis, Springer Berlin Heidelberg. **29:** 49-52.

Komazec, Z., S. Lemajic-Komazec, R. Jovic, C. Nadj, L. Jovancevic and S. Savovic (2010). "Comparison between auditory steady-state responses and pure-tone audiometry." Vojnosanitetski Pregled **67**(9): 761-765.

Kreiter, A. K. (2001). "Functional implications of temporal structure in primate cortical information processing." Zoology **104**(3–4): 241-255.

Krishnaveni, V., S. Jayaraman, L. Anitha and K. Ramadoss (2006). "Removal of ocular artifacts from EEG using adaptive thresholding of wavelet coefficients." Journal of Neural Engineering **3**(4): 338-346.

Kurtzberg, D. (1989). "Cortical evnt-related potentials Assessment of Auditory system function." Seminars in Hearing **10**(3): 252-261.

Larsby, B., M. Hallgren and S. Arlinger (2000). "A system for recording of auditory evoked responses." Technology and health care : official journal of the European Society for Engineering and Medicine **8**(6): 315-326.

Lee, T. W., M. Girolami and T. J. Sejnowski (1999). "Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources." Neural Computation **11**(2): 417-441.

Lee, Y. H., S. S. Kim, S. I. Park and J. H. Park (2013). "Extraction of enhanced evoked potentials using wavelet filtering." Multimedia Tools and Applications **63**(1): 45-61.

Leue, A., C. Klein, S. Lange and A. Beauducel (2013). "Inter-individual and intra-individual variability of the N2 component: On reliability and signal-to-noise ratio." Brain and Cognition **83**(1): 61-71.

Lightfoot, G. (2010). "The N1-P2 Cortical Auditory Evoked Potential in Threshold Estimation." Otometrices.

Lightfoot, G. R. K., V.L., (2006). "Cortical electric response audiometry hearing threshold estimation:

Accuracy, speed and the effects of stimulus presentation features." Ear & Hearing

Lu, W. and J. C. Rajapakse (2000). Constrained independent component analysis. in Advances in Neural Information Processing Systems 13 (NIPS2000, Citeseer.

Luck, S. (2005). An Introduction to the Event-Related Potential Technique (Cognitive Neuroscience), A Bradford Book.

Lv, J., D. M. Simpson and S. L. Bell (2007). "Objective detection of evoked potentials using a bootstrap technique." Medical Engineering & Physics **29**(2): 191-198.

Maatta, S., A. Paakkonen, P. Saavalainen and J. Partanen (2005). "Selective attention event-related potential effects from auditory novel stimuli in children and adults." Clinical Neurophysiology **116**(1): 129-141.

Makeig, S., T. P. Jung, A. J. Bell, D. Ghahremani and T. J. Sejnowski (1997). "Blind separation of auditory event-related brain responses into independent components." Proceedings of the National Academy of Sciences of the United States of America **94**(20): 10979-10984.

Mandal, A. (2011). "Hearing loss - What is hearing loss?".

Mijovic, B., M. De Vos, I. Gligorijevic, J. Taelman and S. Van Huffel (2010). "Source separation from single-channel recordings by combining empirical-mode

decomposition and independent component analysis." Biomedical Engineering, IEEE Transactions on **57**(9): 2188-2196.

Muenssinger, J., K. T. Stingl, T. Matuz, G. Binder, S. Ehehalt and H. Preissl (2013). "Auditory habituation to simple tones: reduced evidence for habituation in children compared to adults." Frontiers in Human Neuroscience **7**: 7.

Naatanen, R., A. W. K. Gaillard and S. Mantysalo (1978). "Early Selective Attention Effect on Evoked Potential Reinterpreted." Acta Psychologica **42**(4): 313-329.

Naatanen, R. and T. Picton (1987). "The N1 wave of the human electric and magnetic response to sound- a review and an analysis of the component structure." Psychophysiology **24**(4): 375-425.

Petkov, C. I., X. Kang, K. Alho, O. Bertrand, E. W. Yund and D. L. Woods (2004). "Attentional modulation of human auditory cortex." Nat Neurosci **7**(6): 658-663.

Picton, T. W., A. Dimitrijevic, M. Sasha John and P. Van Roon (2001). "The use of phase in the detection of auditory steady-state responses." Clinical Neurophysiology **112**(9): 1698-1711.

Picton, T. W. and S. A. Hillyard (1974). "Human Auditory Evoked Potentials.2. Effects of Attention." Electroencephalography and Clinical Neurophysiology **36**(2): 191-199.

Picton, T. W., S. A. Hillyard, Galamobo.R and M. Schiff (1971). "Human auditory attention-central or pripheral process." Science **173**(3994): 351-&.

Quiroga, R. Q., O. A. Rosso, E. Basar, M. Schurmann (2001), " Wavelet Entropy in Event-Related Potentials: A New Methods Shows Ordering of EEG Oscillation", Biol. Cybern. **84**, 291-299

Rajapakse, J. C. and L. Wei (2006). "ICA with Reference." Neurocomputing **69**(16-18): 2244-22572257.

Ramirez, D., J. Via and I. Santamaria (2008). A generalization of the magnitude squared coherence spectrum for more than two signals: definition, properties and estimation. Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on.

Robert F. Burkard, J. J. E., Manuel Don (2006). Auditory evoked potentials: basic principles and clinical application, Lippincott Williams & Wilki.

Röhrig, B. (2010). "Sample Size Calculation in Clinical Trials." Dtsch Arztebl Int 2010; 107(31–32): 552–6.

Rongen, H., V. Hadamschek and M. Schiek (2006). "Real time data acquisition and online signal processing for magnetoencephalography." Ieee Transactions on Nuclear Science **53**(3): 704-708.

Ross J. Roeser, Michael Valente and H. Hosford-Dunn (2000). Audiology: Diagnosis.

Sander, T. H., G. Wubbeler, A. Lueschow, G. Curio and L. Trahms (2002). "Cardiac artifact subspace identification and elimination in cognitive MEG data using time-delayed decorrelation." Ieee Transactions on Biomedical Engineering **49**(4): 345-354.

Sanders, L. D., C. Stevens, D. Coch and H. J. Neville (2006). "Selective auditory attention in 3-to 5-year-old children: An event-related potential study." Neuropsychologia **44**(11): 2126-2138.

Santamaria, I. and J. Via (2007). Estimation of the Magnitude Squared Coherence Spectrum Based on Reduced-Rank Canonical Coordinates. Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on.

Scherg, M., J. Vajsar and T. W. Picton (1989). "A source analysis of the late human auditory evoked potentials." J. Cognitive Neuroscience **1**(4): 336-355.

Scherg, M. and D. Voncramon (1985). "2 Bilateral sources of the late AEP as identied by a spatio -temporal dipole model." Electroencephalography and Clinical Neurophysiology **62**(1): 32-44.

Sijbers, J., J. Van Audekerke, M. Verhoye, A. Van der Linden and D. Van Dyck (2000). "Reduction of ECG and gradient related artifacts in simultaneously recorded human EEG/MRI data." Magnetic Resonance Imaging **18**(7): 881-886.

Silva, I. (2009). "Estimation of Postaverage SNR from Evoked Responses Under Nonstationary Noise." IEEE Transactions on Biomedical Engineering **56**(8): 2123-2130.

Soong, T. T. (2004). Fundamentals of Probability and Statistics for Engineers. USA, John Wiley and Sons Ltd.

Sussman, E., R. Ceponiene, A. Shestakova, R. Näätänen and I. Winkler (2001). "Auditory stream segregation processes operate similarly in school-aged children and adults." Hearing Research **153**(1–2): 108-114.

Teng, C., W. Feng, W. Chi Man, J. N. da Cruz and H. Yong (2014). "Objective evaluation of fatigue by EEG spectral analysis in steady-state visual evoked potential-based brain-computer interfaces." Biomedical Engineering OnLine **13**: 28 (13 pp.)-28 (13 pp.).

Teplan, M. (2002). "Fundamentals of EEG Measurement." Measurement Science Review **2**: 1-11.

Teplan, M., A. Krakovská and S. Štolc (2006). "EEG responses to long-term audio–visual stimulation." International Journal of Psychophysiology **59**(2): 81-90.

Thornton, A. R. D., M. Harmer and B. A. Lavoie (2007). "Selective attention increases the temporal precision of the auditory N-100 event-related potential." Hearing Research **230**(1-2): 73-79.

Tyner, F. S., J. R. Knott and W. B. Mayer (1989). Fundamentals of EEG Technology: Clinical correlates, Raven Press.

Valenti, P., F. Del Piccolo, A. Pellegrini, S. Montagnese, E. De Toni, D. Mapelli, A. Gatta and P. Amodio (2002). "Detection of minimal hepatic encephalopathy: Cognitive evoked potentials vs. EEG spectral analysis." Hepatology **36**(4): 528A-528A.

Ventura, L. M. P., K. d. F. Alvarenga and O. A. Costa Filho (2009). "Protocolo para captação dos potenciais evocados auditivos de longa latência." Brazilian Journal of Otorhinolaryngology **75**: 879-883.

Walker, H., W. Hall and J. Hurst (1990). Clinical Methods: The History, Physical, and Laboratory Examinations., Boston: Butterworths;.

Wang, Y.-X., T.-S. Qiu and R. Liu (2012). "Few-trial Extraction of Single Channel Evoked Potential Based on ICA with Virtual Channels." Chinese Journal of Biomedical Engineering **31**(5): 704-711.

Woodward, M. (2005). Epidemiology: Study Design and Data Analysi, Chapman & Hall/CRC.

Yong Hee, L., K. Soon Seok, P. Sea Il and P. Jong Hyuk (2013). "Extraction of enhanced evoked potentials using wavelet filtering." Multimedia Tools and Applications **63**(1): 45-61.

Yuan, C. and J. Zhang (2014). "Extraction of single-trial evoked potentials with Extended Infomax ICA algorithm and its applications to BCI systems." Proceedings of the 33rd Chinese Control Conference: 7139-7144.

Yuexian, Z., H. Yong and Z. Zhiguo (2011). Fast Extraction of Somatosensory Evoked Potential Based on Robust Adaptive Filtering, InTech.

Zarzoso, V. and P. Comon (2007). Comparative speed analysis of FastICA. Independent Component Analysis and Signal Separation, Proceedings. M. E. Davies, C. J. James, S. A. Abdallah and M. D. Plumbley. Berlin, Springer-Verlag Berlin. **4666:** 293-300.

Ziehe, A. and K. R. Muller (1998). "TDSEP-an efficient algorithm for blind separation using time structure." ICANN 98. Proceedings of the 8th International Conference on Artificial Neural Networks: 675-680 vol.672.