EEG SIGNAL PROCESSING

Saeid Sanei and J. A. Chambers

Centre of Digital Signal Processing
Cardiff University, UK

John Wiley & Sons, Ltd

EEG SIGNAL PROCESSING

EEG SIGNAL PROCESSING

Saeid Sanei and J.A. Chambers

Centre of Digital Signal Processing
Cardiff University, UK

John Wiley & Sons, Ltd

John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester,
West Sussex PO19 8SQ, England

Telephone (+44) 1243 779777

Email (for orders and customer service enquiries): cs-books@wiley.co.uk
Visit our Home Page on www.wileyeurope.com or www.wiley.com

Other Wiley Editorial Offices

John Wiley & Sons Inc., 111 River Street, Hoboken, NJ 07030, USA

Jossey-Bass, 989 Market Street, San Francisco, CA 94103-1741, USA

Wiley-VCH Verlag GmbH, Boschstr. 12, D-69469 Weinheim, Germany

John Wiley & Sons Australia Ltd, 42 McDougall Street, Milton, Queensland 4064, Australia

John Wiley & Sons (Asia) Pte Ltd, 2 Clementi Loop #02-01, Jin Xing Distripark, Singapore 129809

John Wiley & Sons Canada Ltd, 6045 Freemont Blvd, Mississauga, Ontario, L5R 4J3, Canada

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Anniversary Logo Design: Richard J. Pacifico

Library of Congress Cataloging-in-Publication Data

Sanei, Saeid.
EEG signal processing / Saeid Sanei and Jonathon Chambers.
p. ; cm.
Includes bibliographical references.
ISBN 978-0-470-02581-9 (alk. paper)
1. Electroencephalography. 2. Signal processing. I. Chambers, Jonathon. II. Title.
[DNLM: 1. Electroencephalography – methods. 2. Evoked Potentials. 3. Signal Processing, Computer-Assisted. WL 150 S223e 2007]
RC386.6.E43S252 2007
616.8′047547 – dc22

2007019900

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

ISBN-13 978-0-470-02581-9

Typeset in 10/12 Times by Laserwords Private Limited, Chennai, India
Printed and bound in Great Britain by Antony Rowe Ltd, Chippenham, Wiltshire
This book is printed on acid-free paper responsibly manufactured from sustainable forestry in which at least two trees are planted for each one used for paper production.

Contents

Preface

There is ever-increasing global demand for more affordable and effective clinical and healthcare services. New techniques and equipment must therefore be developed to aid in the diagnosis, monitoring, and treatment of abnormalities and diseases of the human body. Biomedical signals (biosignals) in their manifold forms are rich information sources, which when appropriately processed have the potential to facilitate such advancements. In today's technology, such processing is very likely to be digital, as confirmed by the inclusion of digital signal processing concepts as core training in biomedical engineering degrees. Recent advancements in digital signal processing are expected to underpin key aspects of the future progress in biomedical research and technology, and it is the purpose of this research monograph to highlight this trend for the processing of measurements of brain activity, primarily electroencephalograms (EEGs).

Most of the concepts in multichannel EEG digital signal processing have their origin in distinct application areas such as communications engineering, seismics, speech and music signal processing, together with the processing of other physiological signals, such as electrocardiograms (ECGs). The particular topics in digital signal processing first explained in this research monograph include definitions; illustrations; time-domain, frequency-domain, and time-frequency domain processing; signal conditioning; signal transforms; linear and nonlinear filtering; chaos definition, evaluation, and measurement; certain classification algorithms; adaptive systems; independent component analysis; and multivariate autoregressive modelling. In addition, motivated by research in the field over the last two decades, techniques specifically related to EEG processing such as brain source localization, detection and classification of event related potentials, sleep signal analysis, seizure detection and prediction, together with brain–computer interfacing are comprehensively explained and, with the help of suitable graphs and (topographic) images, simulation results are provided to assess the efficacy of the methods.

Chapter 1 of this research monograph is a comprehensive biography of the history and generation of EEG signals, together with a discussion of their significance and diagnostic capability. Chapter 2 provides an in-depth introduction to the mathematical algorithms and tools commonly used in the processing of EEG signals. Most of these algorithms have only been recently developed by experts in the signal processing community and then applied to the analysis of EEG signals for various purposes. In Chapter 3, event-related potentials are explained and the schemes for their detection and classification are explored. Many neurological and psychiatric brain disorders are diagnosed and monitored using these techniques. Chapter 4 complements the previous chapter by specifically looking at the behaviour of EEG signals in patients suffering from epilepsy. Some very recent

methods in seizure prediction are demonstrated. This chapter concludes by opening up a new methodology in joint, or bimodal, EEG–fMRI analysis of epileptic seizure signals. Localization of brain source signals is next covered in Chapter 5. Traditional dipole methods are described and some very recent processing techniques such as blind source separation are briefly reviewed. In Chapter 6, the concepts developed for the analysis and description of EEG sleep recordings are summarized and the important parameters and terminologies are explained. Finally, in Chapter 7, one of the most important applications of the developed mathematical tools for processing of EEG signals, namely brain–computer interfacing, is explored and recent advancements are briefly explained. Results of the application of these algorithms are described.

In the treatment of various topics covered within this research monograph it is assumed that the reader has a background in the fundamentals of digital signal processing and wishes to focus on processing of EEGs. It is hoped that the concepts covered in each chapter provide a foundation for future research and development in the field.

In conclusion, we do wish to stress that in this book there is no attempt to challenge previous clinical or diagnostic knowledge. Instead, the tools and algorithms described in this book can, we believe, potentially enhance the significant clinically related information within EEG signals and thereby aid physicians and ultimately provide more cost-effective and efficient diagnostic tools.

Both authors wish to thank most sincerely our previous and current PhD students who have contributed so much to the material in this work and our understanding of the field. Special thanks to Min Jing, Tracey Lee, Kianoush Nazarpour, Leor Shoker, Loukianous Spyrou, and Wenwu Wang, who contributed to providing some of the illustrations. Finally, this book became truly possible due to spiritual support and encouragement of Maryam Zahabsaniei, Erfan Sanei, and Ideen Sanei.

Saeid Sanei
Jonathon Chambers
January 2007

List of Abbreviations

3D	Three-dimensional
ACT	Adaptive chirplet transform
AD	Alzheimer's disease
ADC	Analogue-to-digital converter
ADD	Attention deficit disorder
ADHD	Attention deficit hyperactivity disorder
AE	Approximate entropy
AEP	Audio evoked potential
Ag–AgCl	Silver–silver chloride
AIC	Akaike information criterion
ALF	Adaptive standardized LORETA/FOCUSS
ALS	Alternating least squares
AMDF	Average magnitude difference function
AMI	Average mutual information
ANN	Artificial neural network
AP	Action potential
APGARCH	Asymmetric power GARCH
AR	Autoregressive modelling
ARMA	Autoregressive moving average
ASDA	American Sleep Disorders Association
BCI	Brain–computer interfacing/interaction
BDS	Brock, Dechert, and Scheinkman
BEM	Boundary element method
BMI	Brain–machine interfacing
BOLD	Blood oxygenation level dependence
BSS	Blind source separation
Ca	Calcium
CANDECOMP	Canonical decomposition
CDR	Current distributed-source reconstruction
CF	Characteristic function
CJD	Creutzfeldt–Jakob disease
Cl	Chloride
CNS	Central nervous system
CSD	Current source density
CT	Computerized tomography

DC	Direct current
DCT	Discrete cosine transform
DLE	Digitally linked ears
DSM	*Diagnostic and Statistical Manual*
DTF	Directed transfer function
DWT	Discrete wavelet transform
ECD	Electric current dipole
ECG	Electrocardiogram/electrocardiography
ECoG	Electrocorticogram
ED	Error distance
EEG	Electroencephalogram/electroencephalography
EGARCH	Exponential GARCH
EGG	Electrogastrography
EKG	Electrocardiogram/electrocardiography
EM	Expectation maximization
EMG	Electromyogram/electromyography
EOG	Electrooculogram
EP	Evoked potential
EPSP	Excitatory postsynaptic potential
ERD	Event-related desynchronization
ERP	Event-related potential
ERS	Event-related synchronization
FA	Factor analysis
FEM	Finite element model
FFNN	Feedforward neural network
FHWA	First half-wave amplitude
FHWD	First half-wave duration
FHWS	First half-wave slope
fICA	Fast independent component analysis
FIR	Finite impulse response
fMRI	Functional magnetic resonance imaging
FOCUSS	Focal underdetermined system solver
FSP	Falsely detected source number percentage
GA	Genetic algorithm
GARCH	Generalized autoregressive conditional heteroskedasticity
GARCH-M	GARCH-in-mean
GFNN	Global false nearest neighbours
GJR-GARCH	Glosten, Jagannathan, and Runkle GARCH
HCI	Human–computer interfacing/interaction
HMM	Hidden Markov model
HOS	Higher-order statistics
IBE	International Bureau for Epilepsy
ICA	Independent component analysis
IIR	Infinite impulse response
ILAE	International League Against Epilepsy

IPSP	Inhibitory postsynaptic potential
IR	Impulse response
ISODATA	Iterative self-organizing data analysis technique algorithm
JADE	Joint approximate diagonalization of eigenmatrices
K	Potassium
KL	Kullback–Laibler
KLT	Karhunen–Loéve transform
KT	Kuhn–Tucker
LD	Linear discriminants
LDA	Linear discriminant analysis
LDA	Long-delta activity
LE	Lyapunov exponent
LEM	Local EEG model
LLE	Largest Lyapunov exponent
LMS	Least mean square
LORETA	Low-resolution electromagnetic tomography algorithm
LP	Lowpass
LRT	Low-resolution tomography
LS	Least squares
LWR	Levinson–Wiggins–Robinson
MA	Moving average
MAF	Multivariate ambiguity function
MAP	Maximum *a posteriori*
MDP	Moving dipole
MEG	Magnetoencephalogram
MI	Mutual information
MIL	Matrix inversion lemma
ML	Maximum likelihood
MLE	Maximum likelihood estimation
MLE	Maximum Lyapunov exponent
MLP	Multilayered perceptron
MMN	Mismatch negativity
MP	Matching pursuits
MRI	Magnetic resonance imaging
MS	Mean square
MS	Multiple sclerosis
MSE	Mean-squared error
MTLE	Mesial temporal lobe epilepsy
MUSIC	Multichannel signal classification
MVAR	Multivariate autoregressive
Na	Sodium
NLMS	Normalized least mean square
NMF	Nonnegative matrix factorization
NN	Neural network
NREM	Nonrapid eye movement

OA	Ocular artefact
OBS	Organic brain syndrome
OP	Oddball paradigm
PARAFAC	Parallel factor
PCA	Principal component analysis
PD	Parkinson's disease
PDF	Probability density function
PET	Positron emission tomography
PLED	Periodic literalized epileptiform discharges
PMBS	Postmovement beta synchronization
PNRD	Persistent nonrhythmic delta activity
POST	Positive occipital sharp transients
PPM	Piecewise Prony method
PSDM	Phase-space dissimilarity measures
PSG	Polysomnography
PWVD	Pseudo Wigner–Ville distribution
QEEG	Quantitative EEG
QGARCH	Quadratic GARCH
QNN	Quantum neural network
QP	Quadratic programming
R&K	Rechtschtschaffen and Kales
RAP	Recursively applied and projected
RBD	REM sleep behaviour disorder
RBF	Radial basis function
REM	Rapid eye movement
RKHS	Reproducing kernel Hilbert space
RLS	Recursive least squares
RV	Residual variance
SAS	Sleep apnea syndrome
SCA	Sparse component analysis
SCP	Slow cortical potential
SCPS	Slow cortical potential shift
SDTF	Short-time DTF
SEM	Structural equation modelling
SHWA	Second half-wave amplitude
SHWD	Second half-wave duration
SHWS	Second half-wave slope
sLORETA	Standardized LORETA
SNNAP	Simulator for neural networks and action potentials
SNR	Signal-to-noise ratio
SOBI	Second-order blind identification
SPET	Single photon emission tomography
SREDA	Subclinical rhythmic EEG discharges of adults
SRNN	Sleep EEG recognition neural network
SSLOFO	Source shrinking LORETA–FOCUSS
SSPE	Subacute sclerosing panencepalities

SSVEP	Steady-state visual-evoked potential
SSVER	Steady-state visual-evoked response
STF	Space–time–frequency
STFD	Spatial time–frequency distribution
STFT	Short-time frequency transform
STL	Short-term largest Lyapunov exponent
SV	Support vector
SVD	Singular-value decomposition
SVM	Support vector machine
SWA	Slow-wave activity
SWDA	Step-wise discriminant analysis
SWS	Slow-wave sleep
TDNN	Time delay neural network
TF	Time–frequency
TGARCH	Threshold GARCH model
TLE	Temporal lobe epilepsy
TNM	Traditional nonlinear method
TTD	Thought translation device
USP	Undetected source number percentage
VEP	Visual evoked potential
WA	Wald tests on amplitudes
WL	Wald test on locations
WMN	Weighted minimum norm
WN	Wavelet network
WT	Wavelet transform
WV	Wigner–Ville

List of Symbols

+eV	Positive charge (electron-volt)
Hz	Hertz; cycles per second
mV	Millivolt
uF/cm^2	Microfarad per squared centimetre
Ca^{++}	Calcium with positive charge
K^+	Potassium with positive charge
Na^+	Sodium with positive charge
$a_{Ca}(.)$	Steady-state activation function
$a_K(.)$	Prediction coefficients
$\mathbf{A}^{\dagger}$	Moore–Penrose pseudo-inverse of $\mathbf{A}$
$\mathbf{A}^{-1}$	Inverse of matrix $\mathbf{A}$
$\hat{\mathbf{A}}$	Estimate of $\mathbf{A}$
A_k	Amplitude of the exponential function
$\mathbf{A}_{\mathrm{k}}$	Matrix with known column vectors
arg	Argument
$\mathbf{A}_{\mathrm{uk}}$	Matrix with unknown column vectors
$A_x(\tau, \nu)$	Ambiguity function
AE	Approximate entropy
$b^2(\omega_1, \omega_2)$	Bicoherency index
(i)	Magnetic field at electrode i
$B(\omega_1, \omega_2)$	Bispectrum
C_e	Number of interneuron cells
C_i	Number of thalamocortical neurons
$C_{ij}(\omega)$	Fourier transform of the cross-correlation coefficients between channels i and j
$C_j(k)$	Wavelet coefficient
$\mathbf{C}_{\mathrm{noise}}$	Covariance of noise, i.e. $\mathbf{C}_{\mathrm{noise}} = \sigma_{\mathrm{n}}^2\mathbf{I}$
$\mathbf{C}_{\mathrm{sig}}$	Covariance of the source signals
$\mathbf{C}_{\mathbf{X}}$	$\mathbf{X}\mathbf{X}^{\mathrm{T}}$
C_1	Maximum subspace correlation
$C_4^{\alpha}(0, 0, 0)$	Fourier coefficients of the fourth-order cyclic cumulant at zero lag
$C_{\boldsymbol{\theta}_n, \boldsymbol{x}_n}$	Cross-covariance matrix between $\boldsymbol{\theta}_n$ and $\boldsymbol{x}_n$

$\mathrm{Coh}_{ij}^2(\omega)$	Coherency of channels i and j
$\mathrm{col}\{\mathbf{F}\}_i$	$[\underbrace{0\cdots0}_{i-1} f_0 \cdots f_M \underbrace{0\cdots0}_{N-M-i}]^T$
Cum	Cumulant
d_E	Embedding dimension
$\frac{\mathrm{d}}{\mathrm{d}t}$	Differentiation with respect to time
$\mathbf{D}$	Scaling matrix
D_a	Attractor dimension
D_I	Information dimension
D_L	Lyapunov dimension
D_r	Correlation dimension
$\mathbf{D}_{ss}(.)$	STFD of the source signals
$\mathbf{D}_{xx}(.)$	STFD of the mixed signals
$\mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_M)$	A diagonal $M \times M$ matrix with the elements $\lambda_1, \lambda_2, \ldots, \lambda_M$
$\mathrm{D}\boldsymbol{w}(k\Delta)$	$\boldsymbol{w}(k\Delta + \Delta) - \boldsymbol{w}(k\Delta)$
$e(n)$	Residual (error)
exp(.)	Exponential
E [.]	Statistical expectation
$\breve{E}_{kb}$	Dispersion measure
μV/cm	Microvolts per centimetre
$\lvert E_n/E_0 \rvert$	Total error amplification factor
E_syn	Synaptic reversal potential
f_k	Discrete-time sinusoidal frequency in samples per second
f_s	Sampling frequency
$\overline{g}$	Maximal conductance
$\overline{g}_\mathrm{K}$	Potassium current
$\overline{g}_\mathrm{syn}$	Maximal synaptic conductance
g'	First derivative of g
g''	Second derivative of g
$\mathbf{G}(.)$	Gain matrix
$G_{\gamma_k}(t, \omega)$	Wigner time–frequency distribution of the kth selected function
GARCH(p,q)	GARCH model with prediction orders p and q
h_e	Excitatory postsynaptic potential
h_i	Inhibitory postsynaptic potential
$H(\omega)$	Fourier transforms of $h(t)$
$H(.)$	Entropy
$\mathbf{I}$	Identity matrix
I_i	Ionic current
$I(s)$	Entropy of s
Im(.)	Imaginary part
j	$\sqrt{-1}$
J_n	Cost function
$\mathbf{J} = [\boldsymbol{j}_1, \boldsymbol{j}_2, \ldots, \boldsymbol{j}_\mathrm{T}]$	Moments of the sources
$\Im$	Function linking the sources to the electrode potentials

$K(.,.)$	Kernel function
KL	Kulback–Laibler distance (divergence)
kurt(.)	Kurtosis
$l(i)$	Latency of the ith ERP component
l_1	Error amplification factor in the context of Lyapunov exponents
$\mathbf{L}_j$	Location of the jth dipole
$\mathbf{L}_k$	Matrix of prediction coefficients
m	Embedding dimension; number of sources
$m_k(.)$	kth moment
$M(\theta, \tau)$	Characteristic function (CF)
$\mathbf{MA}(\theta, \tau)$	Multivariate ambiguity function (MAF)
max	Maximum
min	Minimum
$\min_{\rho,\theta,\mathbf{S}}$	Minimization with respect to ρ, θ, and $\mathbf{S}$
$m \times l$	m by l
n	Discrete time index; iteration index
n_e	Number of electrodes
n_r	Number of remaining mixtures
N	Number of data samples
Neg	Negentropy
p	Prediction order
$p(\mathbf{Y})$	Joint probability distribution
p_x	Probability density function (PDF) of signal $x(n)$
$p_y(y_i(n))$	Marginal distribution of the ith output
$\mathbf{P}$	Permutation matrix
$P(.)$	Probability density function
$\mathbf{P}_\mathbf{H}^\perp$	Projection matrix to the orthogonal complement of the column; space of $\mathbf{H}$, i.e. $\mathbf{P}_\mathbf{H}^\perp = (\mathbf{I} - \mathbf{H}\mathbf{H}^\dagger)$
$P(f, n)$	Signal power at a given time–frequency
$P_\mathrm{ref}(f)$	Average power during some reference time calculated for frequency f
$q(n)$	Two-dimensional Fourier transform of the bispectrum
$r_{xy}(\tau)$	Cross-correlation between two zero mean wide-sense stationary continuous random signals $x(t)$ and $y(t)$
$\hat{r}_x(., m)$	Estimated autocorrelation function for the mth frame
$\mathbf{R}(q)$	Covariance matrix of $\boldsymbol{x}(n)$
$\mathbf{R}_\mathrm{S}(k)$	Covariance of the sources at discrete time lag k
$\mathbf{R}_\mathrm{X}(k)$	Covariance of the mixed signals at discrete time lag k
$\mathbf{R}_\mathrm{Y}(k)$	Covariance of the outputs (estimated sources) at discrete time lag k
$\Re^M$	M-dimensional space
Re(.)	Real part
$\hat{s}(n)$	Best estimate of the actual signal $\boldsymbol{s}(n)$ in the mean-square sense

$S_{xx}(\lambda_k, \boldsymbol{\theta})$	Spectral density of a Gaussian vector process $\boldsymbol{x}$ with Parameters $\boldsymbol{\theta}$
$\mathbf{S}\| \otimes \|\mathbf{D}$	Katri–Rao product
sgn(.)	Sign of (.)
subcorr	Subspace correlation
T	Signal length
T_r	Empirical threshold
tanh	Hyperbolic tangent
$\dot{u}(t)$	$\mathrm{d}u(t)/\mathrm{d}t$
U_s	Spatial prior
U_t	Temporal prior
U_1	Likelihood probability
U_2	Prior probability
$\boldsymbol{v}(n)$	Zero-mean noise
V_{A_1}	Left earlobe reference voltage
V_{A_2}	Right earlobe reference voltage
w_k	Activation variable
$w_\infty(E)$	Steady-state activation function
$\dot{w}(t)$	$\mathrm{d}w(t)/\mathrm{d}t$
w_opt	Optimum $\boldsymbol{w}$
$W(a, b)$	Wavelet transform
x_0	Initial value of x
$\overline{x}_\mathrm{c}$	Cluster centre
$X_i(\omega)$	Fourier transform of the ith segment of one EEG channel
$X_\mathrm{p}(\omega)$	Power spectrum for signal x
$X_\mathrm{WV}(t, \omega)$	Wigner–Ville frequency distribution of a signal $x(t)$
$\breve{X}_\mathrm{WV}(t, \omega)$	Pseudo Wigner–Ville distribution (PWVD)
$X_\mathrm{WV}(t, \omega)$	Spectral distribution of both auto- and cross-correlations
$\boldsymbol{x} \in R^d$	$\boldsymbol{x}$ belongs to the d-dimensional space of real values
Z	Z-transform
Z^{-1}	Inverse Z-transform
α	Alpha brain rhythm
α	Penalty term
α_j and β_j	Nonlinear model coefficients
α_k	Damping factor
$\alpha(E)$	Forward rate function
β	Penalty term
$\beta(E)$	Backward rate function
γ	Gamma brain rhythm
γ	Learning rate
δ	Delta brain rhythm
δ	*A priori* chosen tolerance
δ_1	Minimal error
δ_2	Maximal error
$\Delta w(n)$	$\boldsymbol{w}(n) - \boldsymbol{w}(n-1)$

ζ	Learning rate
$\eta\ (e(n)^2)$	Performance index
θ	Theta brain rhythm
θ_j	Initial phase in radians
κ	Kappa brain activity
λ	Lambda brain activity
λ	Largest Lyapunov exponent; regulation parameter
λ^*	Finite-time exponent
$\lambda(x_0)$	Ljapunov (Lyapunov) exponents of x starting from x_0
λ_i	Lyapunov exponent
λ_j	jth eigenvalue
λ_k	$2\pi k/N$
$\lambda_{\max}$	Maximum eigenvalue
λ_{t}	Temperature-dependent factor
λ_1	Maximum Lyapunov exponent (MLE)
$\hat{\mathbf{\Lambda}}_{\mathrm{e}}$	Diagonal matrix of noise eigenvalues
$\hat{\mathbf{\Lambda}}_{\mathrm{s}}$	Diagonal matrix containing the first m eigenvalues
μ	Mu brain rhythm
μ	Learning rate; convergence parameter; step size; mean
ξ_i	Slack parameter for feature i
ρ_{xy}	Cross-correlation coefficient
σ	Sigma brain activity
σ_i	Width of the ith ERP component
σ_x	Standard deviations of $x(n)$
σ_y	Standard deviations of $y(n)$
σ^2	Variance
σ_{ij}^2	Noise power of the ARMA model
σ_{n}^2	Noise variance
σ_{t}^2	Time-domain variance
σ_{ω}^2	Frequency-domain variance
τ	Tau brain activity
τ	Fractal dimension
τ_i	Latency for the ith template
$\tau_k(E)$	Time-constant function
φ	Phi brain activity
$\phi(t)$	Scaling function
$\varphi(\cdot)$	Nonlinear penalty function
$\phi(\theta, \tau)$	Kernel function acting as a mask to enhance the regions in the TF domain
$\phi(.\ ,\ .)$	Cohen's class
$\mathbf{\Phi}_{\mathrm{e}}$	Noise-only subspace span
$\mathbf{\Phi}_{\mathrm{s}}$	Signal subspace span
$\Phi(\omega)$	Fourier transforms of $\phi(t)$
$\mathbf{\Phi} = \{\varphi_k\}$	Set of orthogonal basis functions
$\hat{\mathbf{\Phi}}_{\mathrm{e}}$	Remaining (noise) eigenvectors

$\hat{\mathbf{\Phi}}_{\mathrm{s}}$	First m left singular vectors
χ_{n-1}	A sequence containing all the past information up to time $n-1$
ω	Angular frequency in radians per second
$\nabla_m(.)$	Approximated gradient with respect to m
$\left.\frac{\partial J(\mathbf{W})}{\partial \mathbf{W}}\right\|_{\mathbf{W}=\mathbf{W}(n)}$	Differentiation with respect to $\mathbf{W}$ evaluated at $\mathbf{W}=\mathbf{W}(n)$
$(.)^{\mathrm{H}}$	Hermitian; complex conjugate transpose
$(.)^{\mathrm{T}}$	Transpose operation
$\|.\|$	Absolute value
$\|.\|_{\mathrm{F}}$	Frobenius norm
$\|.\|_1$	L_1-norm
$\|.\|^2$	Euclidean norm
$\|.\|_2^2$	Squared Euclidean norm
$\langle \cdot, \cdot \rangle$	Inner product
$\prod_{i=1}^{m}$	Multiplication of m components (signals)
$\sum_{i=1}^{N}$	Summation of N components (signals)
$\forall n$	For all n values
$\exists q$	There exists a 'q'
$*$	Linear convolution
$**$	Two-dimensional discrete-time convolution
$\propto$	Proportional to
∞	Infinity
$\geq$	Larger or equal to
$\leq$	Smaller or equal to
$\uparrow_2$	Up-sample by 2
$\downarrow_2$	Down-sample by 2

1

Introduction to EEG

The neural activity of the human brain starts between the 17th and 23rd week of prenatal development. It is believed that from this early stage and throughout life electrical signals generated by the brain represent not only the brain function but also the status of the whole body. This assumption provides the motivation to apply advanced digital signal processing methods to the electroencephalogram (EEG) signals measured from the brain of a human subject, and thereby underpins the later chapters of the book.

Although nowhere in this book do the authors attempt to comment on the physiological aspects of brain activities there are several issues related to the nature of the original sources, their actual patterns, and the characteristics of the medium, that have to be addressed. The medium defines the path from the neurons, as so-called signal sources, to the electrodes, which are the sensors where some form of mixtures of the sources are measured.

Understanding of neuronal functions and neurophysiological properties of the brain together with the mechanisms underlying the generation of signals and their recordings is, however, vital for those who deal with these signals for detection, diagnosis, and treatment of brain disorders and the related diseases. A brief history of EEG measurements is first provided.

1.1 History

Carlo Matteucci (1811–1868) and Emil Du Bois-Reymond (1818–1896) were the first people to register the electrical signals emitted from muscle nerves using a galvanometer and established the concept of neurophysiology [1,2]. However, the concept of *action current* introduced by Hermann Von Helmholz [3] clarified and confirmed the negative variations that occur during muscle contraction.

Richard Caton (1842–1926), a scientist from Liverpool, England, used a galvanometer and placed two electrodes over the scalp of a human subject and thereby first recorded brain activity in the form of electrical signals in 1875. Since then, the concepts of electro-(referring to registration of brain electrical activities) encephalo- (referring to emitting the signals from the head), and gram (or graphy), which means drawing or writing, were combined so that the term EEG was henceforth used to denote electrical neural activity of the brain.

EEG Signal Processing S. Sanei and J. Chambers

Fritsch (1838–1927) and Hitzig (1838–1907) discovered that the human cerebral can be electrically stimulated. Vasili Yakovlevich Danilevsky (1852–1939) followed Caton's work and finished his PhD thesis in the investigation of the physiology of the brain in 1877 [4]. In this work, he investigated the activity of the brain following electrical stimulation as well as spontaneous electrical activity in the brain of animals.

The cerebral electrical activity observed over the visual cortex of different species of animals was reported by Ernst Fleischl von Marxow (1845–1891). Napoleon Cybulski (1854–1919) provided EEG evidence of an epileptic seizure in a dog caused by electrical stimulation.

The idea of the association of epileptic attacks with abnormal electrical discharges was expressed by Kaufman [5]. Pravidch-Neminsky (1879–1952), a Russian physiologist, recorded the EEG from the brain, termed the dura, and the intact skull of a dog in 1912. He observed a 12–14 cycle/s rhythm under normal conditions, which slowed under asphyxia and later called it the *electrocerebrogram*.

The discoverer of the existence of human EEG signals was Hans Berger (1873–1941). He began his study of human EEGs in 1920 [6]. Berger is well known by almost all electroencephalographers. He started working with a string galvanometer in 1910, then migrated to a smaller Edelmann model, and after 1924, to a larger Edelmann model. In 1926, Berger started to use the more powerful Siemens double coil galvanometer (attaining a sensitivity of 130 μV/cm) [7]. His first report of human EEG recordings of one to three minutes duration on photographic paper was in 1929. In this recording he only used a one-channel bipolar method with fronto-occipital leads. Recording of the EEG became popular in 1924. The first report of 1929 by Berger included the alpha rhythm as the major component of the EEG signals, as described later in this chapter, and the alpha blocking response.

During the 1930s the first EEG recording of sleep spindles was undertaken by Berger. He then reported the effect of hypoxia on the human brain, the nature of several diffuse and localized brain disorders, and gave an inkling of epileptic discharges [8]. During this time another group established in Berlin-Buch and led by Kornmüller, provided more precise recording of the EEG [9]. Berger was also interested in cerebral localization and particularly in the localization of brain tumours. He also found some correlation between mental activities and the changes in the EEG signals.

Toennies (1902–1970) from the group in Berlin built the first biological amplifier for the recording of brain potentials. A differential amplifier for recording EEGs was later produced by the Rockefeller foundation in 1932.

The importance of multichannel recordings and using a large number of electrodes to cover a wider brain region was recognized by Kornmüller [10]. The first EEG work focusing on epileptic manifestation and the first demonstration of epileptic spikes were presented by Fischer and Löwenbach [11–13].

In England, W. Gray Walter became the pioneer of clinical electroencephalography. He discovered the foci of slow brain activity (delta waves), which initiated enormous clinical interest in the diagnosis of brain abnormalities. In Brussels, Fredric Bremer (1892–1982) discovered the influence of afferent signals on the state of vigilance [14].

Research activities related to EEGs started in North America in around 1934. In this year, Hallowell Davis illustrated a good alpha rhythm for himself. A cathode ray

oscilloscope was used around this date by the group in St Louis University in Washington, in the study of peripheral nerve potentials. The work on human EEGs started at Harvard in Boston and the University of Iowa in the 1930s. The study of epileptic seizure developed by Fredric Gibbs was the major work on EEGs during these years, as the realm of epileptic seizure disorders was the domain of their greatest effectiveness. Epileptology may be divided historically into two periods [15]: before and after the advent of EEG. Gibbs and Lennox applied the idea of Fischer based on his studies about picrotoxin and its effect on the cortical EEG in animals to human epileptology. Berger [16] showed a few examples of paroxysmal EEG discharges in a case of presumed petit mal attacks and during a focal motor seizure in a patient with general paresis.

As the other great pioneers of electroencephalography in North America, Hallowel and Pauline Davis were the earliest investigators of the nature of EEG during human sleep. A. L. Loomis, E. N. Harvey, and G. A. Hobart were the first who mathematically studied the human sleep EEG patterns and the stages of sleep. At McGill University, H. Jasper studied the related behavioural disorder before he found his niche in basic and clinical epileptology [17].

The American EEG Society was founded in 1947 and the First International EEG Congress was held in London, United Kingdom, around this time. While the EEG studies in Germany were still limited to Berlin, Japan gained attention by the work of Motokawa, a researcher of EEG rhythms [18]. During these years the neurophysiologists demonstrated the thalamocortical relationship through anatomical methods. This led to the development of the concept of centrencephalic epilepsy [19].

Throughout the 1950s the work on EEGs expanded in many different places. During this time surgical operation for removing the epileptic foci became popular and the book entitled *Epilepsy and the Functional Anatomy of the Human Brain* (Penfiled and Jasper) was published. During this time microelectrodes were invented. They were made of metals such as tungsten or glass, filled with electrolytes such as potassium chloride, with diameters of less than 3 μm.

Depth electroencephalography of a human was first obtained with implanted intracerebral electrodes by Mayer and Hayne (1948). Invention of intracellular microelectrode technology revolutionized this method and was used in the spinal cord by Brock *et al*. in 1952 and in the cortex by Phillips in 1961.

Analysis of EEG signals started during the early days of EEG measurement. Berger assisted by Dietch (1932) applied Fourier analysis to EEG sequences, which was rapidly developed during the 1950s. Analysis of sleep disorders with EEGs started its development in the 1950s through the work of Kleitman at the University of Chicago.

In the 1960s analysis of the EEGs of full-term and premature newborns began its development [20]. Investigation of evoked potentials (EPs), especially visual EPs, as commonly used for monitoring mental illnesses, progressed during the 1970s.

The history of EEG, however, has been a continuous process, which started from the early 1300s and has brought daily development of clinical, experimental, and computational studies for discovery, recognition, diagnosis, and treatment of a vast number of neurological and physiological abnormalities of the brain and the rest of the central nervous system (CNS) of human beings. Nowadays, EEGs are recorded invasively and noninvasively using fully computerized systems. The EEG machines are

equipped with many signal processing tools, delicate and accurate measurement electrodes, and enough memory for very long-term recordings of several hours. EEG or MEG (magnetoencephalogram) machines may be integrated with other neuroimaging systems such as functional magnetic resonance imaging (fMRI). Very delicate needle-type electrodes can also be used for recording the EEGs from over the cortex (electrocortiogram), and thereby avoid the attenuation and nonlinearity effects induced by the skull. The nature of neural activities within the human brain will be discribed next.

1.2 Neural Activities

The CNS generally consists of nerve cells and glia cells, which are located between neurons. Each nerve cell consists of axons, dendrites, and cell bodies. Nerve cells respond to stimuli and transmit information over long distances. A nerve cell body has a single nucleus and contains most of the nerve cell metabolism, especially that related to protein synthesis. The proteins created in the cell body are delivered to other parts of the nerve. An axon is a long cylinder, which transmits an electrical impulse and can be several metres long in vertebrates (giraffe axons go from the head to the tip of the spine). In humans the length can be a percentage of a millimetre to more than a metre. An axonal transport system for delivering proteins to the ends of the cell exists and the transport system has 'molecular motors', which ride upon tubulin rails.

Dendrites are connected to either the axons or dendrites of other cells and receive impulses from other nerves or relay the signals to other nerves. In the human brain each nerve is connected to approximately 10,000 other nerves, mostly through dendritic connections.

The activities in the CNS are mainly related to the synaptic currents transferred between the junctions (called synapses) of axons and dendrites, or dendrites and dendrites of cells. A potential of 60–70 mV with negative polarity may be recorded under the membrane of the cell body. This potential changes with variations in synaptic activities. If an action potential travels along the fibre, which ends in an *excitatory* synapse, an excitatory postsynaptic potential (EPSP) occurs in the following neuron. If two action potentials travel along the same fibre over a short distance, there will be a summation of EPSPs producing an action potential on the postsynaptic neuron providing a certain threshold of membrane potential is reached. If the fibre ends in an *inhibitory* synapse, then hyperpolarization will occur, indicating an inhibitory postsynaptic potential (IPSP) [21,22]. Figure 1.1 shows the above activities schematically.

Following the generation of an IPSP, there is an overflow of cations from the nerve cell or an inflow of anions into the nerve cell. This flow ultimately causes a change in potential along the nerve cell membrane. Primary transmembranous currents generate secondary inonal currents along the cell membranes in the intra- and extracellular space. The portion of these currents that flow through the extracellular space is directly responsible for the generation of field potentials. These field potentials, usually with less than 100 Hz frequency, are called EEGs when there are no changes in the signal average and DC if there are slow drifts in the average signals, which may mask the actual EEG signals. A combination of EEG and DC potentials is often observed for some abnormalities in the brain such as seizure (induced by pentylenetetrazol), hypercapnia, and asphyxia [23]. The focus will next be on the nature of active potentials.

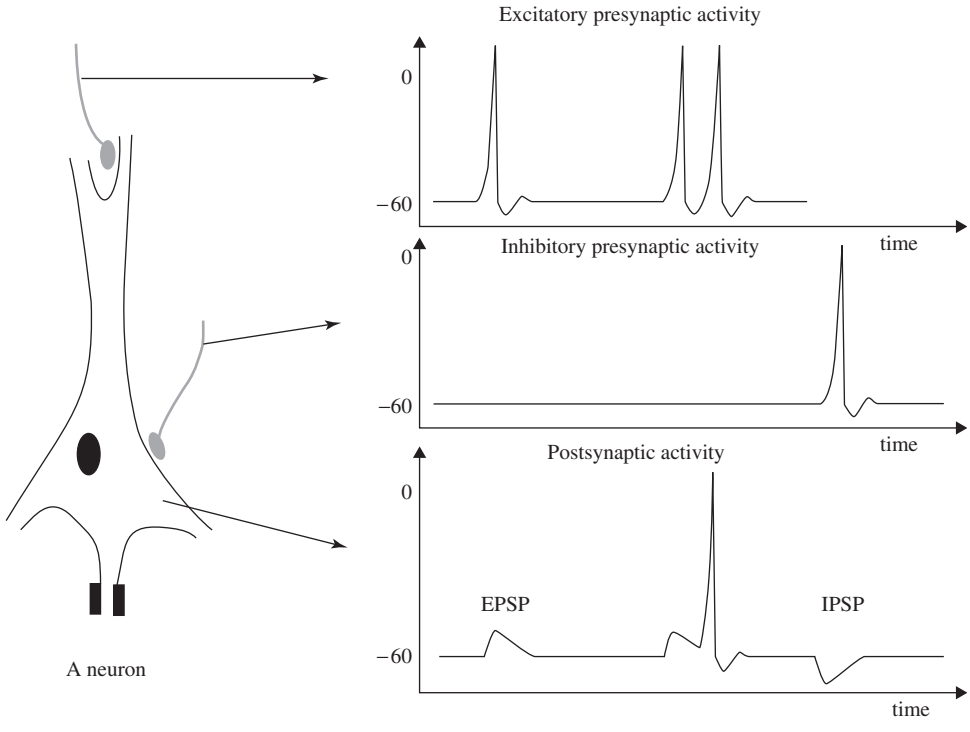

Figure 1.1 The neuron membrane potential changes and current flow during synaptic activation recorded by means of intracellular microelectrodes. Action potentials in the excitatory and inhibitory presynaptic fibre respectively lead to EPSP and IPSP in the postsynaptic neuron

1.3 Action Potentials

The information transmitted by a nerve is called an action potential (AP). APs are caused by an exchange of ions across the neuron membrane and an AP is a temporary change in the membrane potential that is transmitted along the axon. It is usually initiated in the cell body and normally travels in one direction. The membrane potential depolarizes (becomes more positive), producing a spike. After the peak of the spike the membrane repolarizes (becomes more negative). The potential becomes more negative than the resting potential and then returns to normal. The action potentials of most nerves last between 5 and 10 milliseconds. Figure 1.2 shows an example AP.

The conduction velocity of action potentials lies between 1 and 100 m/s. APs are initiated by many different types of stimuli; sensory nerves respond to many types of stimuli, such as chemical, light, electricity, pressure, touch, and stretching. On the other hand, the nerves within the CNS (brain and spinal cord) are mostly stimulated by chemical activity at synapses.

A stimulus must be above a threshold level to set off an AP. Very weak stimuli cause a small local electrical disturbance, but do not produce a transmitted AP. As soon as the stimulus strength goes above the threshold, an action potential appears and travels down the nerve.

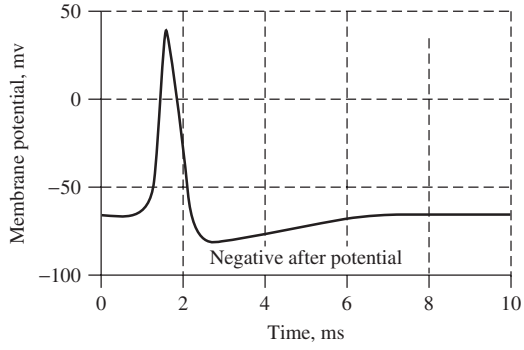

Figure 1.2 An example action potential

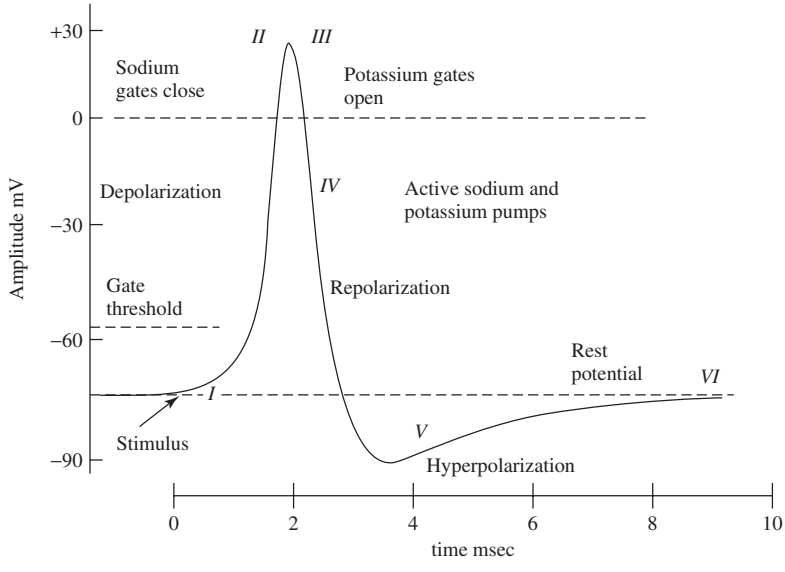

Figure 1.3 Changing the membrane potential for a giant squid by closing the Na channels and opening K channels (adopted from Ka Xiong Charand [24])

The spike of the AP is mainly caused by opening of Na (sodium) channels. The Na pump produces gradients of both Na and K (potassium) ions. Both are used to produce the action potential; Na is high outside the cell and low inside. Excitable cells have special Na and K channels with gates that open and close in response to the membrane voltage (voltage-gated channels). Opening the gates of Na channels allows Na to rush into the cell, carrying positive charge. This makes the membrane potential positive (depolarization), producing the spike. Figure 1.3 shows the stages of the process during evolution of an action potential for a giant squid. For a human being the amplitude of the AP ranges between approximately −60 mV and 10 mV. During this process [24]:

I. When the dendrites of a nerve cell receive the stimulus the Na^+ channels will open. If the opening is sufficient to drive the interior potential from -70 mV up to -55 mV, the process continues.
II. As soon as the action threshold is reached, additional Na^+ channels (sometimes called voltage-gated channels) open. The Na^+ influx drives the interior of the cell membrane up to approximately $+30$ mV. The process to this point is called depolarization.
III. Then Na^+ channels close and the K^+ channels open. Since the K^+ channels are much slower to open, the depolarization has time to be completed. Having both Na^+ and K^+ channels open at the same time would drive the system towards neutrality and prevent the creation of the action potential.
IV. Having the K^+ channels open, the membrane begins to repolarize back towards its rest potential.
V. The repolarization typically overshoots the rest potential to a level of approximately -90 mV. This is called hyperpolarization and would seem to be counterproductive, but it is actually important in the transmission of information. Hyperpolarization prevents the neuron from receiving another stimulus during this time, or at least raises the threshold for any new stimulus. Part of the importance of hyperpolarization is in preventing any stimulus already sent up an axon from triggering another action potential in the opposite direction. In other words, hyperpolarization ensures that the signal is proceeding in one direction.
VI. After hyperpolarization, the Na^+/K^+ pumps eventually bring the membrane back to its resting state of -70 mV.

The nerve requires approximately two milliseconds before another stimulus is presented. During this time no AP can be generated. This is called the refractory period. The generation of EEG signals is next described.

1.4 EEG Generation

An EEG signal is a measurement of currents that flow during synaptic excitations of the dendrites of many pyramidal neurons in the cerebral cortex. When brain cells (neurons) are activated, the synaptic currents are produced within the dendrites. This current generates a magnetic field measurable by electromyogram (EMG) machines and a secondary electrical field over the scalp measurable by EEG systems.

Differences of electrical potentials are caused by summed postsynaptic graded potentials from pyramidal cells that create electrical dipoles between the soma (body of a neuron) and apical dendrites, which branch from neurons (Figure 1.4). The current in the brain is generated mostly by pumping the positive ions of sodium, Na^+, potassium, K^+, calcium, Ca^{++}, and the negative ion of chlorine, Cl^-, through the neuron membranes in the direction governed by the membrane potential [25].

The human head consists of different layers including the scalp, skull, brain (Figure 1.5), and many other thin layers in between. The skull attenuates the signals approximately one hundred times more than the soft tissue. On the other hand, most of the noise is generated either within the brain (internal noise) or over the scalp (system noise or external noise). Therefore, only large populations of active neurons can generate enough potential to be recordable using the scalp electrodes. These signals are later amplified greatly for display

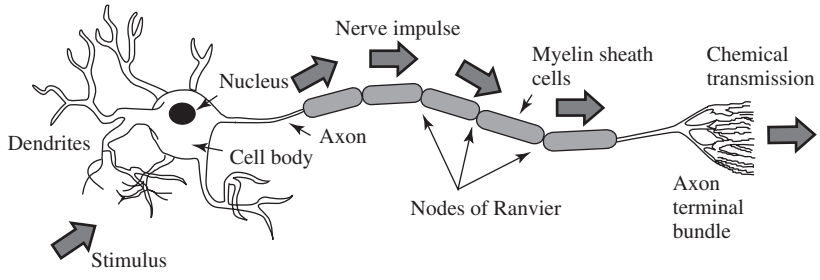

Figure 1.4 Structure of a neuron (adopted from Attwood and MacKay [25])

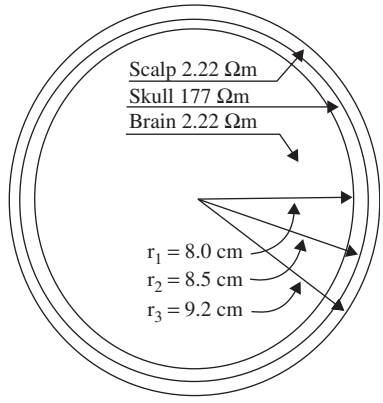

Figure 1.5 The three main layers of the brain including their approximate resistivities and thicknesses (Ω = ohm)

purposes. Approximately 10^{11} neurons are developed at birth when the central nervous system (CNS) becomes complete and functional [26]. This makes an average of 10^4 neurons per cubic mm. Neurons are interconnected into neural nets through synapses. Adults have approximately 5×10^{14} synapses. The number of synapses per neuron increases with age, whereas the number of neurons decreases with age. From an anatomical point of view the brain may be divided into three parts: the cerebrum, cerebellum, and brain stem (Figure 1.6). The cerebrum consists of both left and right lobes of the brain with highly convoluted surface layers called the cerebral cortex.

The cerebrum includes the regions for movement initiation, conscious awareness of sensation, complex analysis, and expression of emotions and behaviour. The cerebellum coordinates voluntary movements of muscles and maintains balance. The brain stem controls involuntary functions such as respiration, heart regulation, biorhythms, and neurohormone and hormone sections [27].

Based on the above section it is clear that the study of EEGs paves the way for diagnosis of many neurological disorders and other abnormalities in the human body. The acquired EEG signals from a human (and also from animals) may, for example, be used for investigation of the following clinical problems [27,28]:

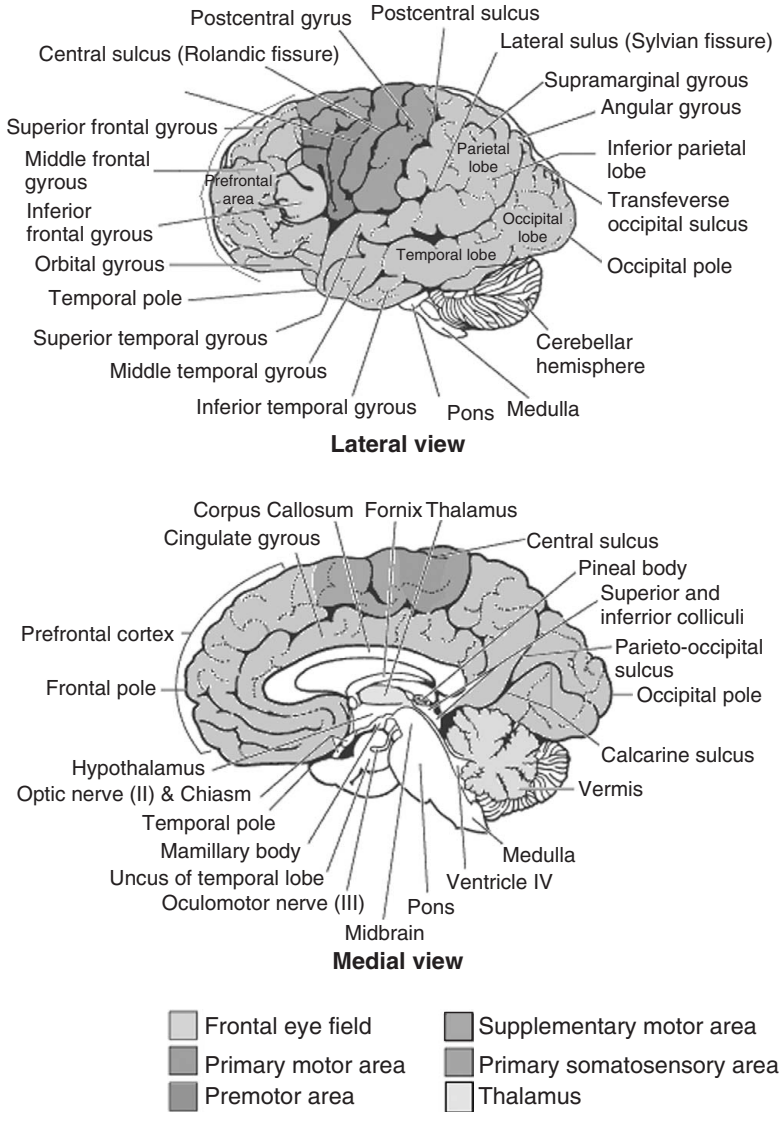

Figure 1.6 Diagrammatic representation of the major parts of the brain

(a) monitoring alertness, coma, and brain death;
(b) locating areas of damage following head injury, stroke, and tumour;
(c) testing afferent pathways (by evoked potentials);
(d) monitoring cognitive engagement (alpha rhythm);
(e) producing biofeedback situations;
(f) controlling anaesthesia depth (servo anaesthesia);
(g) investigating epilepsy and locating seizure origin;

(h) testing epilepsy drug effects;
(i) assisting in experimental cortical excision of epileptic focus;
(j) monitoring the brain development;
(k) testing drugs for convulsive effects;
(l) investigating sleep disorders and physiology;
(m) investigating mental disorders;
(n) providing a hybrid data recording system together with other imaging modalities.

This list confirms the rich potential for EEG analysis and motivates the need for advanced signal processing techniques to aid the clinician in their interpretation. The brain rhythms will next be described, which are expected to be measured within EEG signals.

1.5 Brain Rhythms

Many brain disorders are diagnosed by visual inspection of EEG signals. The clinical experts in the field are familiar with manifestation of brain rhythms in the EEG signals. In healthy adults, the amplitudes and frequencies of such signals change from one state of a human to another, such as wakefulness and sleep. The characteristics of the waves also change with age. There are five major brain waves distinguished by their different frequency ranges. These frequency bands from low to high frequencies respectively are called alpha (α), theta (θ), beta (β), delta (δ), and gamma (γ). The alpha and beta waves were introduced by Berger in 1929. Jasper and Andrews (1938) used the term 'gamma' to refer to the waves of above 30 Hz. The delta rhythm was introduced by Walter (1936) to designate all frequencies below the alpha range. He also introduced theta waves as those having frequencies within the range of 4–7.5 Hz. The notion of a theta wave was introduced by Wolter and Dovey in 1944 [29].

Delta waves lie within the range of 0.5–4 Hz. These waves are primarily associated with deep sleep and may be present in the waking state. It is very easy to confuse artefact signals caused by the large muscles of the neck and jaw with the genuine delta response. This is because the muscles are near the surface of the skin and produce large signals, whereas the signal that is of interest originates from deep within the brain and is severely attenuated in passing through the skull. Nevertheless, by applying simple signal analysis methods to the EEG, it is very easy to see when the response is caused by excessive movement.

Theta waves lie within the range of 4–7.5 Hz. The term theta might be chosen to allude to its presumed thalamic origin. Theta waves appear as consciousness slips towards drowsiness. Theta waves have been associated with access to unconscious material, creative inspiration and deep meditation. A theta wave is often accompanied by other frequencies and seems to be related to the level of arousal. It is known that healers and experienced mediators have an alpha wave that gradually lowers in frequency over long periods of time. The theta wave plays an important role in infancy and childhood. Larger contingents of theta wave activity in the waking adult are abnormal and are caused by various pathological problems. The changes in the rhythm of theta waves are examined for maturational and emotional studies [30].

Alpha waves appear in the posterior half of the head and are usually found over the occipital region of the brain. They can be detected in all parts of posterior lobes of the brain. For alpha waves the frequency lies within the range of 8–13 Hz, and commonly appears as a round or sinusoidal shaped signal. However, in rare cases it may manifest itself as sharp waves. In such cases, the negative component appears to be sharp and the positive component appears to be rounded, similar to the wave morphology of the rolandic mu (μ) rhythm. Alpha waves have been thought to indicate both a relaxed awareness without any attention or concentration. The alpha wave is the most prominent rhythm in the whole realm of brain activity and possibly covers a greater range than has been previously accepted. A peak can regularly be seen in the beta wave range in frequencies even up to 20 Hz, which has the characteristics of an alpha wave state rather than one for a beta wave. Again, very often a response is seen at 75 Hz, which appears in an alpha setting. Most subjects produce some alpha waves with their eyes closed, which is why it has been claimed that it is nothing but a waiting or scanning pattern produced by the visual regions of the brain. It is reduced or eliminated by opening the eyes, by hearing unfamiliar sounds, by anxiety, or mental concentration or attention. Albert Einstein could solve complex mathematical problems while remaining in the alpha state, although generally beta and theta waves are also present. An alpha wave has a higher amplitude over the occipital areas and has an amplitude of normally less than 50 μV. The origin and physiological significance of an alpha wave is still unknown and yet more research has to be undertaken to understand how this phenomenon originates from cortical cells [31].

A beta wave is the electrical activity of the brain varying within the range of 14–26 Hz (though in some literature no upper bound is given). A beta wave is the usual waking rhythm of the brain associated with active thinking, active attention, focus on the outside world, or solving concrete problems, and is found in normal adults. A high-level beta wave may be acquired when a human is in a panic state. Rhythmical beta activity is encountered chiefly over the frontal and central regions. Importantly, a central beta rhythm is related to the rolandic mu rhythm and can be blocked by motor activity or tactile stimulation. The amplitude of beta rhythm is normally under 30 μV. Similar to the mu rhythm, the beta wave may also be enhanced because of a bone defect [29] and also around tumoural regions.

The frequencies above 30 Hz (mainly up to 45 Hz) correspond to the gamma range (sometimes called the fast beta wave). Although the amplitudes of these rhythms are very low and their occurrence is rare, detection of these rhythms can be used for confirmation of certain brain diseases. The regions of high EEG frequencies and highest levels of cerebral blood flow (as well as oxygen and glucose uptake) are located in the frontocentral area. The gamma wave band has also been proved to be a good indication of event-related synchronization (ERS) of the brain and can be used to demonstrate the locus for right and left index finger movement, right toes, and the rather broad and bilateral area for tongue movement [32].

Waves in frequencies much higher than the normal activity range of EEG, mostly in the range of 200–300 Hz, have been found in cerebellar structures of animals, but they have not played any role in clinical neurophysiology [33,34].

Figure 1.7 shows the typical normal brain rhythms with their usual amplitude levels. In general, the EEG signals are the projection of neural activities that are attenuated by

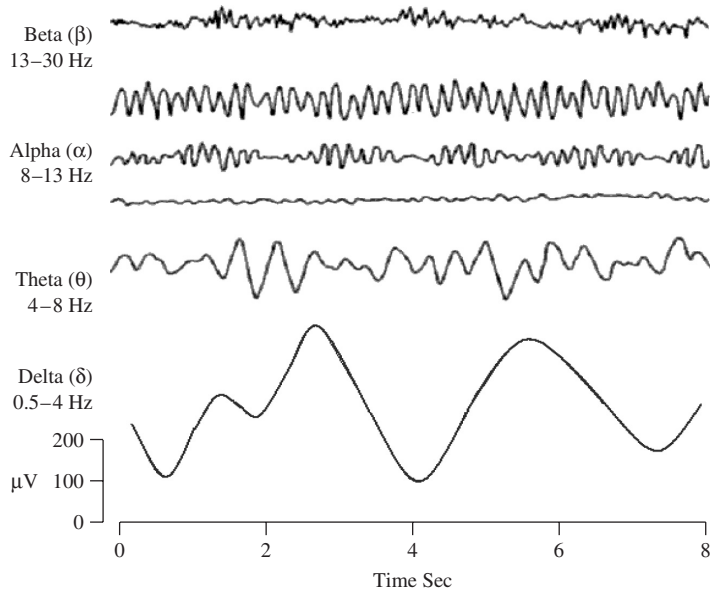

Figure 1.7 Four typical dominant brain normal rhythms, from high to low frequencies. The delta wave is observed in infants and sleeping adults, the theta wave in children and sleeping adults, the alpha wave is detected in the occipital brain region when there is no attention, and the beta wave appears frontally and parietally with low amplitude

leptomeninges, cerebrospinal fluid, dura matter, bone, galea, and the scalp. Cartographic discharges show amplitudes of 0.5–1.5 mV and up to several millivolts for spikes. However, on the scalp the amplitudes commonly lie within 10–100 μV.

The above rhythms may last if the state of the subject does not change and therefore they are approximately cyclic in nature. On the other hand, there are other brain waveforms, which may:

(a) Have a wide frequency range or appear as spiky-type signals, such as K-complexes, vertex waves (which happen during sleep), or a breach rhythm, which is an alpha-type rhythm due to a cranial bone defect [35], which does not respond to movement, and is found mainly over the midtemporal region (under electrodes T3 or T4), and some seizure signals.

(b) Be a transient such as an event-related potential (ERP) and contain positive occipital sharp transient (POST) signals (also called rho (ρ) waves).

(c) Originate from the defective regions of the brain such as tumoural brain lesions.

(d) Be spatially localized and considered as cyclic in nature, but can be easily blocked by physical movement such as mu rhythm. Mu denotes motor and is strongly related to the motor cortex. Rolandic (central) mu is related to posterior alpha in terms of amplitude and frequency. However, the topography and physiological significance are quite different. From the mu rhythm the cortical functioning and the changes in

brain (mostly bilateral) activities subject to physical and imaginary movements can be investigated. The mu rhythm has also been used in feedback training for several purposes such as treatment of epileptic seizure disorder [29].

There are also other rhythms introduced by researchers such as:

(e) Phi (φ) rhythm (less than 4 Hz) occurring within two seconds of eye closure. The phi rhythm was introduced by Daly [36].
(f) Kappa (κ) rhythm, which is an anterior temporal alpha-like rhythm. It is believed to be the result of discrete lateral oscillations of the eyeballs and is considered to be an artefact signal.
(g) The sleep spindles (also called the sigma (σ) activity) within the 11–15 Hz frequency range.
(h) Tau (τ) rhythm, which represents the alpha activity in the temporal region.
(i) Eyelid flutter with closed eyes, which gives rise to frontal artefacts in the alpha band.
(j) Chi (χ) rhythm is a mu-like activity believed to be a specific rolandic pattern of 11–17 Hz. This wave has been observed during the course of Hatha Yoga exercises [37].
(k) Lambda (λ) waves are most prominent in waking patients, but are not very common. They are sharp transients occurring over the occipital region of the head of walking subjects during visual exploration. They are positive and time-locked to saccadic eye movement with varying amplitude, generally below 90 μV [38].

It is often difficult to understand and detect the brain rhythms from the scalp EEGs, even with trained eyes. Application of advanced signal processing tools, however, should enable separation and analysis of the desired waveforms from within the EEGs. Therefore, a definition of foreground and background EEG is very subjective and entirely depends on the abnormalities and applications. Next to consider is the development in the recording and measurement of EEG signals.

1.6 EEG Recording and Measurement

Acquiring signals and images from the human body has become vital for early diagnosis of a variety of diseases. Such data can be in the form of electrobiological signals such as an electrocardiogram (ECG) from the heart, electromyogram (EMG) from muscles, electroencephalogram (EEG) from the brain, magnetoencephalogram (MEG) from the brain, electrogastrogram (EGG) from the stomach, and electroocclugram (or electrooptigram, EOG) from eye nerves. Measurements can also have the form of one type of ultrasound or radiograph such as sonograph (or ultrasound image), computerized tomography (CT), magnetic resonance imaging (MRI) or functional MRI (fMRI), positron emission tomography (PET), and single photon emission tomography (SPET).

Functional and physiological changes within the brain may be registered by either EEG, MEG, or fMRI. Application of fMRI is, however, very limited in comparison with EEG or MEG for a number of important reasons:

(a) The time resolution of fMRI image sequences is very low (for example approximately two frames/s), whereas the complete EEG bandwidth can be viewed using EEG or MEG signals.
(b) Many types of mental activities, brain disorders, and malfunctions of the brain cannot be registered using fMRI since their effect on the level of oxygenated blood is low.
(c) The accessibility to fMRI (and currently to MEG) systems is limited and costly.
(d) The spatial resolution of EEG, however, is limited to the number of recording electrodes (or number of coils for MEG).

The first electrical neural activities were registered using simple galvanometers. In order to magnify very fine variations of the pointer a mirror was used to reflect the light projected to the galvanometer on the wall. The d'Arsonval galvanometer later featured a mirror mounted on a movable coil and the light focused on the mirror was reflected when a current passed the coil. The capillary electrometer was introduced by Lippmann and Marey [39]. The string galvanometer, as a very sensitive and more accurate measuring instrument, was introduced by Einthoven in 1903. This became a standard instrument for a few decades and enabled photographic recording.

More recent EEG systems consist of a number of delicate electrodes, a set of differential amplifiers (one for each channel) followed by filters [27], and needle (pen)-type registers. The multichannel EEGs could be plotted on plane paper or paper with a grid. Soon after this system came to the market, researchers started looking for a computerized system, which could digitize and store the signals. Therefore, to analyse EEG signals it was soon understood that the signals must be in digital form. This required sampling, quantization, and encoding of the signals. As the number of electrodes grows the data volume, in terms of the number of bits, increases. The computerized systems allow variable settings, stimulations, and sampling frequency, and some are equipped with simple or advanced signal processing tools for processing the signals.

The conversion from analogue to digital EEG is performed by means of multichannel analogue-to-digital converters (ADCs). Fortunately, the effective bandwidth for EEG signals is limited to approximately 100 Hz. For many applications this bandwidth may be considered to be even half of this value. Therefore, a minimum frequency of 200 samples/s (to satisfy the Nyquist criterion) is often enough for sampling the EEG signals. In some applications where a higher resolution is required for representation of brain activities in the frequency domain, sampling frequencies of up to 2000 sample/s may be used.

In order to maintain the diagnostic information the quantization of EEG signals is normally very fine. Representation of each signal sample with up to 16 bits is very popular for the EEG recording systems. This makes the necessary memory volume for archiving the signals massive, especially for sleep EEG and epileptic seizure monitoring records. However, in general, the memory size for archiving the radiological images is often much larger than that used for archiving the EEG signals.

A simple calculation shows that for a one hour recording from 128-electrode EEG signals sampled at 500 samples/s a memory size of $128 \times 60 \times 60 \times 500 \times 16 \approx 3.68$ Gbits ≈ 0.45 Gbyte is required. Therefore, for longer recordings of a large number of patients there should be enough storage facilities such as in today's technology Zip disks, CDs,

large removable hard drives, and optical disks. Although the format of reading the EEG data may be different for different EEG machines, these formats are easily convertible to spreadsheets readable by most signal processing software packages such as MATLAB.

The EEG recording electrodes and their proper function are crucial for acquiring high-quality data. Different types of electrodes are often used in the EEG recording systems, such as:

- disposable (gel-less, and pre-gelled types);
- reusable disc electrodes (gold, silver, stainless steel, or tin);
- headbands and electrode caps;
- saline-based electrodes;
- needle electrodes.

For multichannel recordings with a large number of electrodes, electrode caps are often used. Commonly used scalp electrodes consist of Ag–AgCl disks, less than 3 mm in diameter, with long flexible leads that can be plugged into an amplifier. Needle electrodes are those that have to be implanted under the skull with minimal invasive operations. High impedance between the cortex and the electrodes as well as the electrodes with high impedances can lead to distortion, which can even mask the actual EEG signals. Commercial EEG recording systems are often equipped with impedance monitors. To enable a satisfactory recording the electrode impedances should read less than 5 kΩ and be balanced to within 1 kΩ of each other. For more accurate measurement the impedances are checked after each trial.

Due to the layered and spiral structure of the brain, however, distribution of the potentials over the scalp (or cortex) is not uniform [40]. This may affect some of the results of source localization using the EEG signals.

1.6.1 Conventional Electrode Positioning

The International Federation of Societies for Electroencephalography and Clinical Neurophysiology has recommended the conventional electrode setting (also called 10–20) for 21 electrodes (excluding the earlobe electrodes), as depicted in Figure 1.8 [17]. Often the earlobe electrodes called A1 and A2, connected respectively to the left and right earlobes, are used as the reference electrodes. The 10–20 system avoids both eyeball placement and considers some constant distances by using specific anatomic landmarks from which the measurement would be made and then uses 10 or 20 % of that specified distance as the electrode interval. The odd electrodes are on the left and the even ones on the right.

For setting a larger number of electrodes using the above conventional system, the rest of the electrodes are placed in between the above electrodes with equidistance between them. For example, C_1 is placed between C_3 and C_z. Figure 1.9 represents a larger setting for 75 electrodes including the reference electrodes based on the guidelines by the American EEG Society. Extra electrodes are sometimes used for the measurement of EOG, ECG, and EMG of the eyelid and eye surrounding muscles. In some applications such as ERP analysis and brain computer interfacing a single channel may be used. In such applications, however, the position of the corresponding electrode has to be well determined.

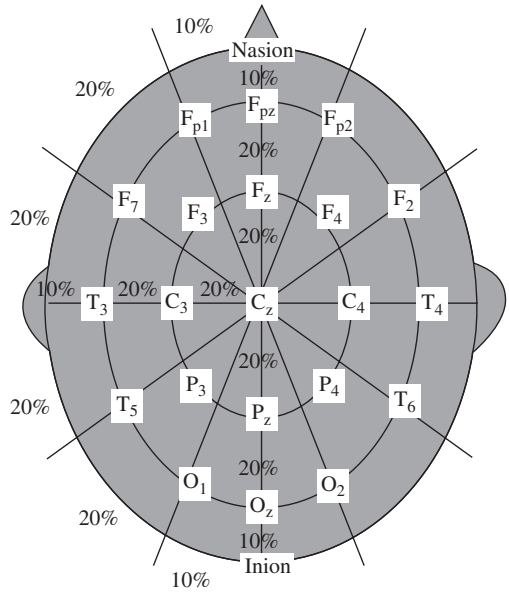

Figure 1.8 Conventional 10–20 EEG electrode positions for the placement of 21 electrodes

For example, C_3 and C_4 can be used to record the right and left finger movement related signals respectively for brain–computer interfacing (BCI) applications. Also F_3, F_4, P_3, and P_4 can be used for recording the ERP P300 signals.

Two different modes of recordings, namely differential and referential, are used. In the differential mode the two inputs to each differential amplifier are from two electrodes. In the referential mode, on the other hand, one or two reference electrodes are used. Several different reference electrode placements can be found in the literature. Physical references can be used as vertex (C_z), linked-ears, linked-mastoids, ipsilateral ear, contralateral ear, C_7, bipolar references, and tip of the nose [28]. There are also reference-free recording techniques, which actually use a common average reference. The choice of reference may produce topographic distortion if the reference is not relatively neutral. In modern instrumentation, however, the choice of a reference does not play an important role in the measurement [41]. In such systems other references such as FP_z, hand, or leg electrodes may be used [42]. The overall setting includes the active electrodes and the references.

In another similar setting, called the Maudsley electrode positioning system, the conventional 10–20 system has been modified to capture better the signals from epileptic foci in epileptic seizure recordings. The only difference between this system and the 10–20 conventional system is that the outer electrodes are slightly lowered to enable better capturing of the required signals. The advantage of this system over the conventional one is that it provides a more extensive coverage of the lower part of the cerebral convexity, increasing the sensitivity for the recording from basal subtemporal structures [43]. Other deviations from the international 10–20 system as used by researchers are found in References [44] and [45].

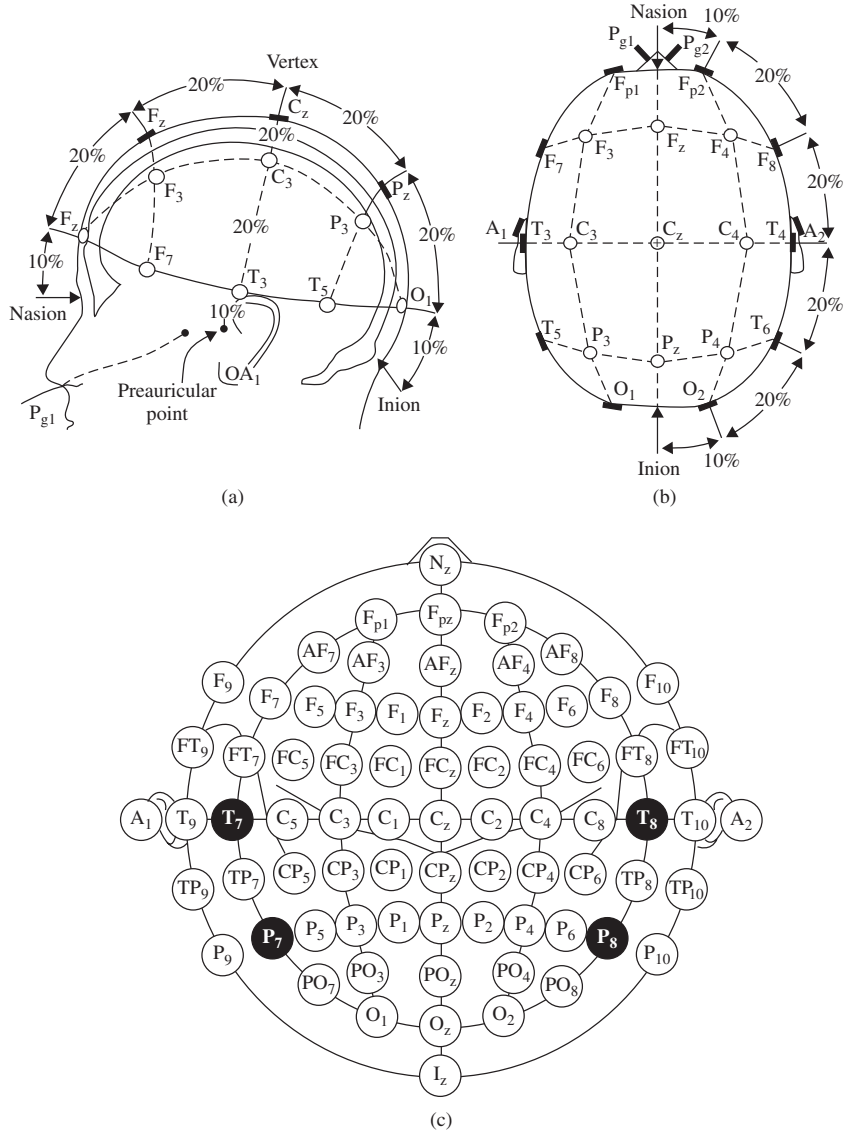

Figure 1.9 A diagrammatic representation of 10–20 electrode settings for 75 electrodes including the reference electrodes: (a) and (b) represent the three-dimensional measures, and (c) indicates a two-dimensional view of the electrode setup configuration

In many applications such as brain–computer interfacing (BCI) and study of mental activity, often a small number of electrodes around the movement-related regions are selected and used from the 10–20 setting system. Figure 1.10 illustrates a typical set of EEG signals during approximately seven seconds of normal adult brain activity.

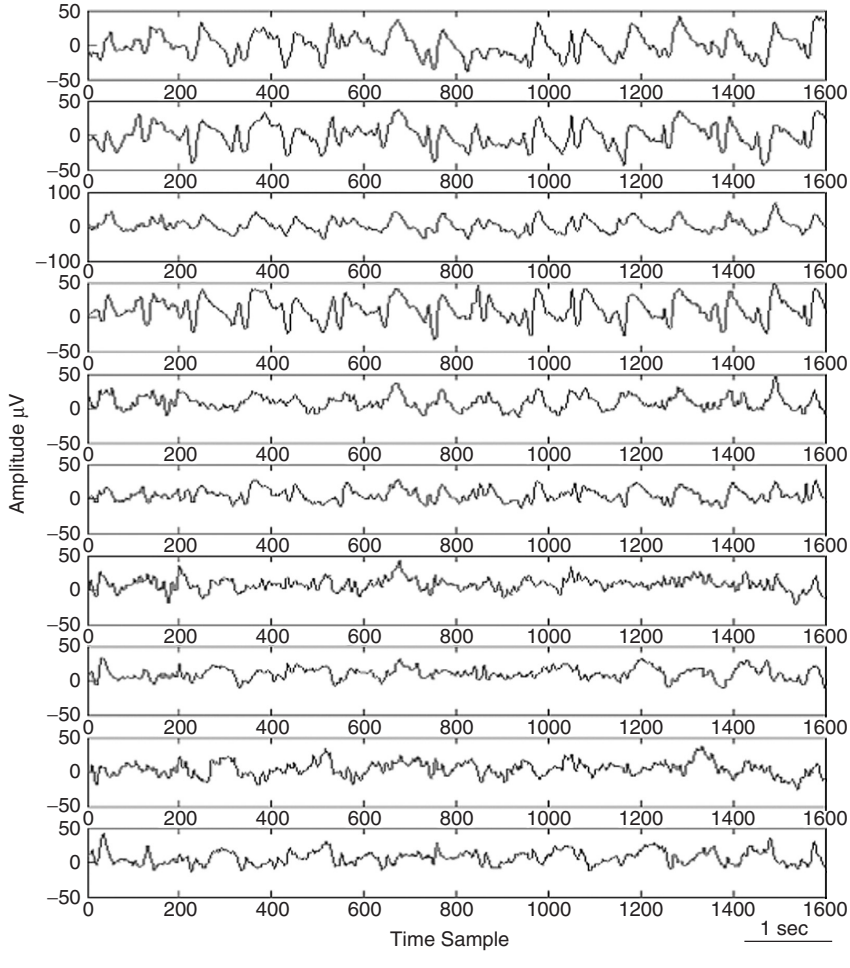

Figure 1.10 A typical set of EEG signals during approximately seven seconds of normal adult brain activity

1.6.2 Conditioning the Signals

The raw EEG signals have amplitudes of the order of μvolts and contain frequency components of up to 300 Hz. To retain the effective information the signals have to be amplified before the ADC and filtered, either before or after the ADC, to reduce the noise and make the signals suitable for processing and visualization. The filters are designed in such a way not to introduce any change or distortion to the signals. Highpass filters with a cut-off frequency of usually less than 0.5 Hz are used to remove the disturbing very low frequency components such as those of breathing. On the other hand, high-frequency noise is mitigated by using lowpass filters with a cut-off frequency of approximately 50–70 Hz. Notch filters with a null frequency of 50 Hz are often necessary to ensure perfect rejection of the strong 50 Hz power supply. In this case the sampling frequency can be as low as twice the bandwidth commonly used by most EEG systems. The commonly used

sampling frequencies for EEG recordings are 100, 250, 500, 1000, and 2000 samples/s. The main artefacts can be divided into patient-related (physiological) and system artefacts. The patient-related or internal artefacts are body movement-related, EMG, ECG (and pulsation), EOG, ballistocardiogram, and sweating. The system artefacts are 50/60 Hz power supply interference, impedance fluctuation, cable defects, electrical noise from the electronic components, and unbalanced impedances of the electrodes. Often in the preprocessing stage these artefacts are highly mitigated and the informative information is restored. Some methods for removing the EEG artefacts will be discussed in the related chapters of this book. Figure 1.11 shows a set of normal EEG signals affected by the eye-blinking artefact. Similarly, Figure 1.12 represents a multichannel EEG set with the clear appearance of ECG signals over the electrodes in the occipital region.

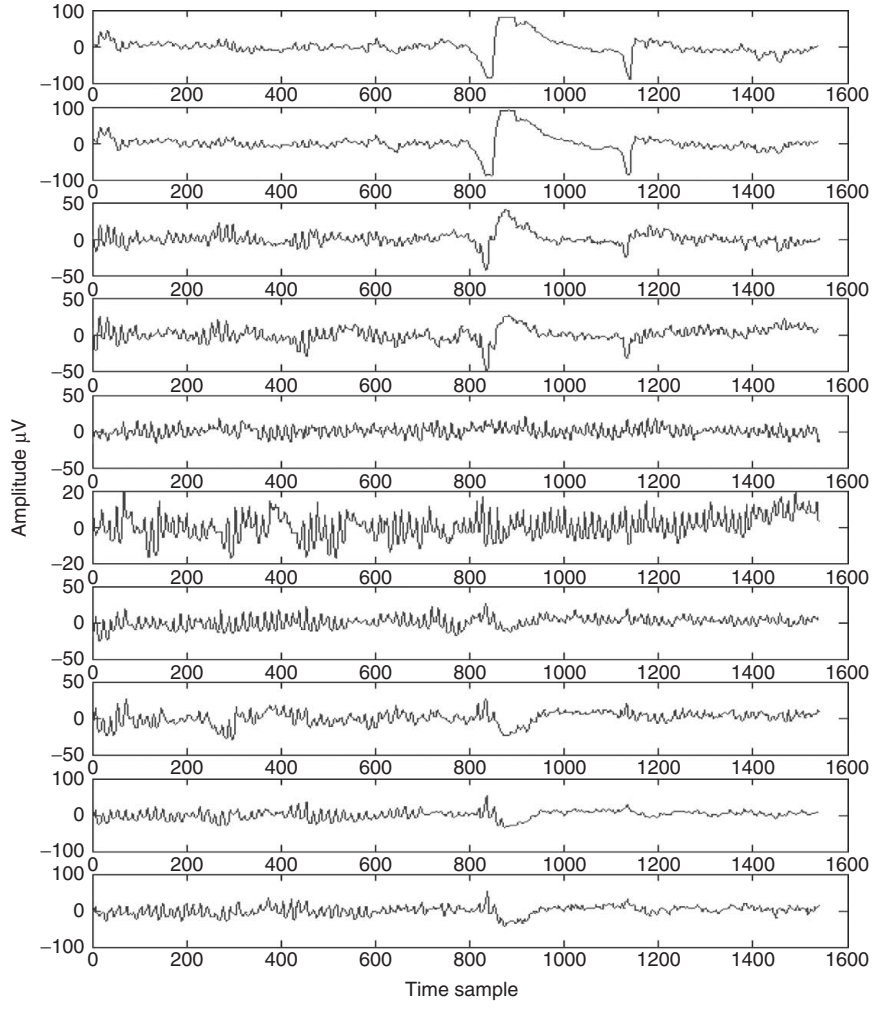

Figure 1.11 A set of normal EEG signals affected by the eye-blinking artefact

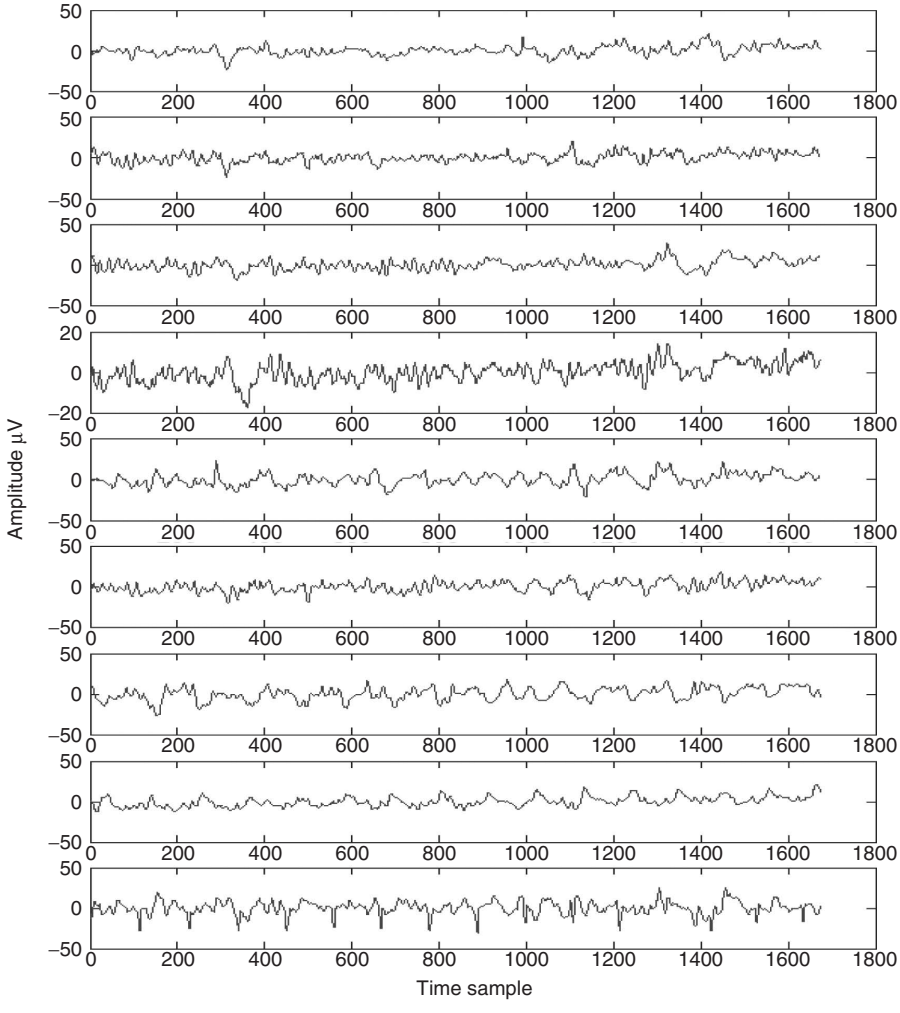

Figure 1.12 A multichannel EEG set with the clear appearance of ECG signals over the electrodes in the occipital region

The next section highlights the changes in EEG measurements that correlate with physiological and mental abnormalities in the brain.

1.7 Abnormal EEG Patterns

Variations in the EEG patterns for certain states of the subject indicate abnormality. This may be due to distortion and the disappearance of abnormal patterns, appearance and increase of abnormal patterns, or disappearance of all patterns. Sharbrough [46] divided the nonspecific abnormalities in the EEGs into three categories: (a) widespread intermittent slow wave abnormalities, often in the delta wave range and associated with brain dysfunction; (b) bilateral persistent EEG, usually associated with impaired conscious

cerebral reactions; and (c) focal persistent EEG usually associated with focal cerebral disturbance.

The first category is a burst-type signal, which is attenuated by alerting the individual and eye opening, and accentuated with eye closure, hyperventilation, or drowsiness. The peak amplitude in adults is usually localized in the frontal region and influenced by age. In children, however, it appears over the occipital or posterior head region. Early findings showed that this abnormal pattern frequently appears with an increased intracranial pressure with tumour or aqueductal stenosis. Also, it correlates with grey matter disease, both in cortical and subcortical locations. However, it can be seen in association with a wide variety of pathological processes varying from systemic toxic or metabolic disturbances to focal intracranial lesions.

Regarding the second category, i.e. bilateral persistent EEG, the phenomenon in different stages of impaired, conscious, purposeful responsiveness are etiologically nonspecific and the mechanisms responsible for their generation are only partially understood. However, the findings in connection with other information concerning etiology and chronicity may be helpful in arriving more quickly at an accurate prognosis concerning the patient's chance of recovering previous conscious life.

As for the third category, i.e. focal persistent EEG, these abnormalities may be in the form of distortion and disappearance of normal patterns, appearance and increase of abnormal patterns, or disappearance of all patterns, but such changes are seldom seen at the cerebral cortex. The focal distortion of normal rhythms may produce an asymmetry of amplitude, frequency, or reactivity of the rhythm. The unilateral loss of reactivity of a physiological rhythm, such as the loss of reactivity of the alpha rhythm to eye opening [47] or to mental alerting [48], may reliably identify the focal side of abnormality. A focal lesion may also distort or eliminate the normal activity of sleep-inducing spindles and vertex waves.

Focal persistent nonrhythmic delta activity (PNRD) may be produced by focal abnormalities. This is one of the most reliable findings of a focal cerebral disturbance. The more persistent, the less reactive, and the more nonrhythmic and polymorphic is such focal slowing, the more reliable an indicator it becomes for the appearance of a focal cerebral disturbance [49–51]. There are other cases such as focal inflammation, trauma, vascular disease, brain tumour, or almost any other cause of focal cortical disturbance, including an asymmetrical onset of CNS degenerative diseases that may result in similar abnormalities in the brain signal patterns.

The scalp EEG amplitude from cerebral cortical generators underlying a skull defect is also likely to increase unless acute or chronic injury has resulted in significant depression of underlying generator activity. The distortions in cerebral activities are because focal abnormalities may alter the interconnections, number, frequency, synchronicity, voltage output, and access orientation of individual neuron generators, as well as the location and amplitude of the source signal itself.

With regards to the three categories of abnormal EEGs, their identification and classification requires a dynamic tool for various neurological conditions and any other available information. A precise characterization of the abnormal patterns leads to a clearer insight into some specific pathophysiologic reactions, such as epilepsy, or specific disease processes, such as subacute sclerosing panencephalitis (SSPE) or Creutzfeldt–Jakob disease (CJD) [46].

Over and above the reasons mentioned above there are many other causes for abnormal EEG patterns. The most common abnormalities are briefly described in the following sections.

1.8 Ageing

The ageing process affects the normal cerebral activity in waking and sleep, and changes the response of the brain to stimuli. The changes stem from reducing the number of neurons and due to a general change in the brain pathology. This pathology indicates that the frontal and temporal lobes of the brain are more affected than the parietal lobes, resulting in shrinkage of large neurons and increasing the number of small neurons and glia [52]. A diminished cortical volume indicates that there is age-related neuronal loss. A general cause for ageing of the brain may be the decrease in cerebral blood flow [52].

A reduction of the alpha frequency is probably the most frequent abnormality in EEG. This often introduces a greater anterior spread to frontal regions in the elderly and reduces the alpha wave blocking response and reactivity. The diminished mental function is somehow related to the degree of bilateral slowing in the theta and delta waves [52].

Although the changes in high-frequency brain rhythms have not been well established, some researchers have reported an increase in beta wave activity. This change in beta wave activity may be considered as an early indication of intellectual loss [52].

As for the sleep EEG pattern, older adults enter into drowsiness with a more gradual decrease in EEG amplitude. Over the age of sixty, the frontocentral waves become slower, the frequency of the temporal rhythms also decreases, frequency lowering with slow eye movements become more prominent, and spindles appear in the wave pattern after the dropout of the alpha rhythm. The amplitudes of both phasic and tonic nonrapid eye movement (NREM) sleep EEG [52] reduce with age. There is also a significant change in rapid eye movement (REM) sleep organization with age; the REM duration decreases during the night and there is a significant increase in sleep disruption [52].

Dementia is the most frequent mental disorder that occurs predominantly in the elderly. Therefore, the prevalence of dementia increases dramatically with ageing of the society. Generally, EEGs are a valuable diagnostic tool in differentiation between organic brain syndromes (OBSs) and functional psychiatric disorders [52], and together with evoked potentials (EPs) play an important role in the assessment of normal and pathological ageing. Ageing is expected to change most neurophysiological parameters. However, the variability of these parameters must exceed the normal degree of spontaneous variability to become a diagnostic factor in acute and chronic disease conditions. Automatic analysis of the EEG during sleep and wakefulness may provide a better contrast in the data and enable a robust diagnostic tool. Next particular and very common mental disorders are described, whose early onset may be diagnosed with EEG measurements.

1.9 Mental Disorders

1.9.1 Dementia

Dementia is a syndrome that consists of a decline in intellectual and cognitive abilities. This consequently affects the normal social activities, mode, and the relationship and interaction with other people [53]. EEG is often used to study the effect of dementia. In

most cases, such as in primary degenerative dementia, e.g. Alzheimer's, and psychiatric disorder, e.g. depression with cognitive impairment, the EEG can be used to detect the abnormality [54].

In Reference [54] dementia is classified into cortical and subcortical forms. The most important cortical dementia is Alzheimer's disease (AD), which accounts for approximately 50 % of the cases. Other known cortical abnormalities are Pick's disease and Creutzfeldt–Jakob diseases (CJD). They are characterized clinically by findings such as aphasia, apraxia, and agnosia. CJD can often be diagnosed using the EEG signals. Figure 1.13 shows a set of EEG signals from a CJD patient. On the other hand, the most common subcortical diseases are Parkinson's disease, Huntington's disease, lacunar state, normal pressure hydrocephalus, and progressive supranuclear palsy. These diseases are characterized by forgetfulness, slowing of thought processes, apathy, and depression. Generally, subcortical dementias introduce less abnormality to the EEG patterns than the cortical ones.

In AD the EEG posterior rhythm (alpha rhythm) slows down and the delta and theta wave activities increase. On the other hand, beta wave activity may decrease. In severe cases epileptiform discharges and triphasic waves can appear. In such cases, cognitive impairment often results. The spectral power also changes; the power increases in delta and theta bands and decreases in beta and alpha bands and also in mean frequency.

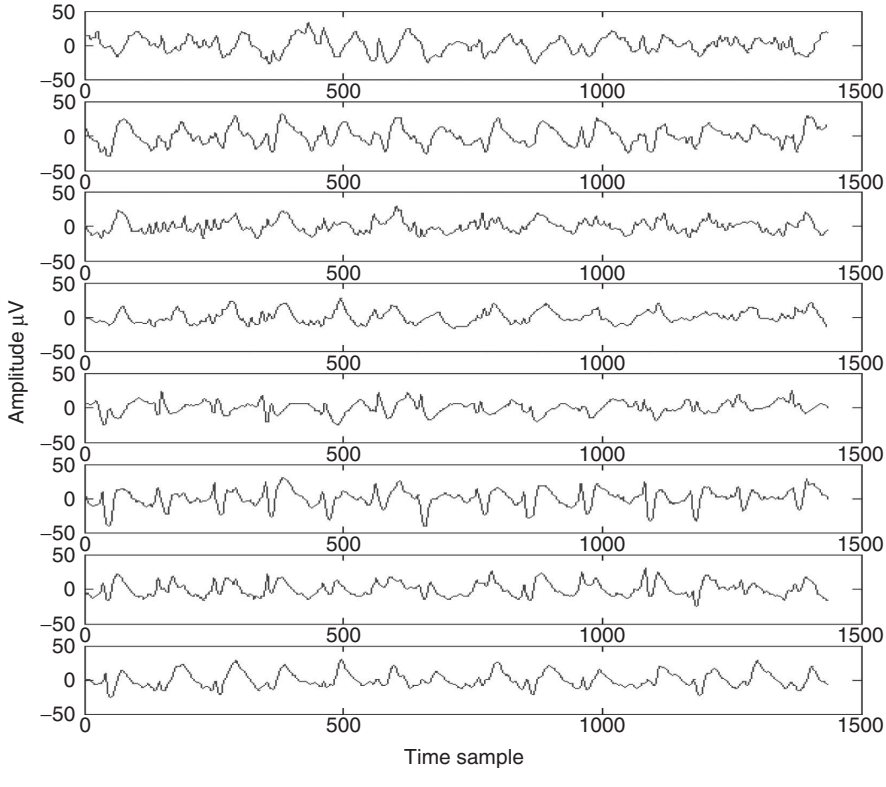

Figure 1.13 A set of multichannel EEG signals from a patient suffering from CJD

The EEG wave morphology is almost the same for AD and Pick's disease. Pick's disease involves the frontal and temporal lobes. An accurate analysis followed by an efficient classification of the cases may discriminate these two diseases. CJD is a mixed cortical and subcortical dementia. This causes slowing of the delta and theta wave activities and, after approximately three months of the onset of the disease, periodic sharp wave complexes are generated that occur almost every second, together with a decrease in the background activity [54]. Parkinson's disease is a subcortical dementia, which causes slowing down of the background activity and an increase of the theta and delta wave activities. Some works have been undertaken using spectral analysis to confirm the above changes [55]. Some other disorders such as depression have a lesser effect on the EEGs and more accurate analysis of the EEGs has to be performed to detect the signal abnormalities for these brain disorders.

Generally, EEG is usually used in the diagnosis and evaluation of many cortical and subcortical dementias. Often it can help to differentiate between a degenerative disorder such as AD and pseudodementia due to psychiatric illness [54]. The EEG may also show whether the process is focal or diffuse (i.e. involves the background delta and theta wave activities). The EEG may also reveal the early CJD-related abnormalities. However, more advanced signal processing and quantitative techniques may be implemented to achieve robust diagnostic and monitoring performance.

1.9.2 Epileptic Seizure and Nonepileptic Attacks

Often the onset of a clinical seizure is characterized by a sudden change of frequency in the EEG measurement. It is normally within the alpha wave frequency band with a slow decrease in frequency (but increase in amplitude) during the seizure period. It may or may not be spiky in shape. Sudden desynchronization of electrical activity is found in electrodecremental seizures. The transition from the preictal to the ictal state, for a focal epileptic seizure, consists of a gradual change from chaotic to ordered waveforms. The amplitude of the spikes does not necessarily represent the severity of the seizure. Rolandic spikes in a child of 4–10 years, for example, are very prominent; however, the seizure disorder is usually quite benign or there may not be clinical seizure [56].

In terms of spatial distribution, in childhood the occipital spikes are very common. Rolandic central–midtemporal–parietal spikes are normally benign, whereas frontal spikes or multifocal spikes are more epileptogenic. The morphology of the spikes varies significantly with age. However, the spikes may occur in any level of awareness including wakefulness and deep sleep.

The distinction of seizure from common artefacts is not difficult. Seizure artefacts within an EEG measurement have a prominent spiky but repetitive (rhythmical) nature, whereas the majority of other artefacts are transients or noise-like in shape. For the case of the ECG, the frequency of occurrence of the QRS waveforms (an element of the ECG) is approximately 1 Hz. These waveforms have a certain shape which is very different from that of seizure signals.

The morphology of an epileptic seizure signal slightly changes from one type to another. The seizure may appear in different frequency ranges. For example, a petit mal discharge often has a slow spike at around 3 Hz, lasting for approximately 70 ms, and normally has its maximum amplitude around the frontal midline. On the other hand, higher frequency spike wave complexes occur for patients over 15 years old. Complexes at 4 Hz and 6 Hz

may appear in the frontal region of the brain of epileptic patients. As for the 6 Hz complex (also called benign EEG variants and patterns), patients with anterior 6 Hz spike waves are more likely to have epileptic seizures and those with posterior discharges tend to have neuroautonomic disturbances [57]. The experiments do not always result in the same conclusion [56]. It was also found that the occipital 6 Hz spikes can be seen and are often drug related (due to hypoanalgetics or barbiturates) and due to withdrawal [58].

Among nonepileptics, the discharges may occur in patients with cerebrovascular disorder, syncopal attacks, and psychiatric problems [56]. Fast and needle-like spike discharges may be seen over the occipital region in most congenitally blind children. These spikes are unrelated to epilepsy and normally disappear in older age patients.

Bursts of 13–16 Hz or 5–7 Hz, as shown in Figure 1.14 (also called 14 and 6 Hz waves), with amplitudes less than 75 μV and arch shapes may be seen over the posterior temporal and the nearby regions of the head during sleep. These waves are positive with respect to the background waves. The 6 and 14 Hz waves may appear independently and

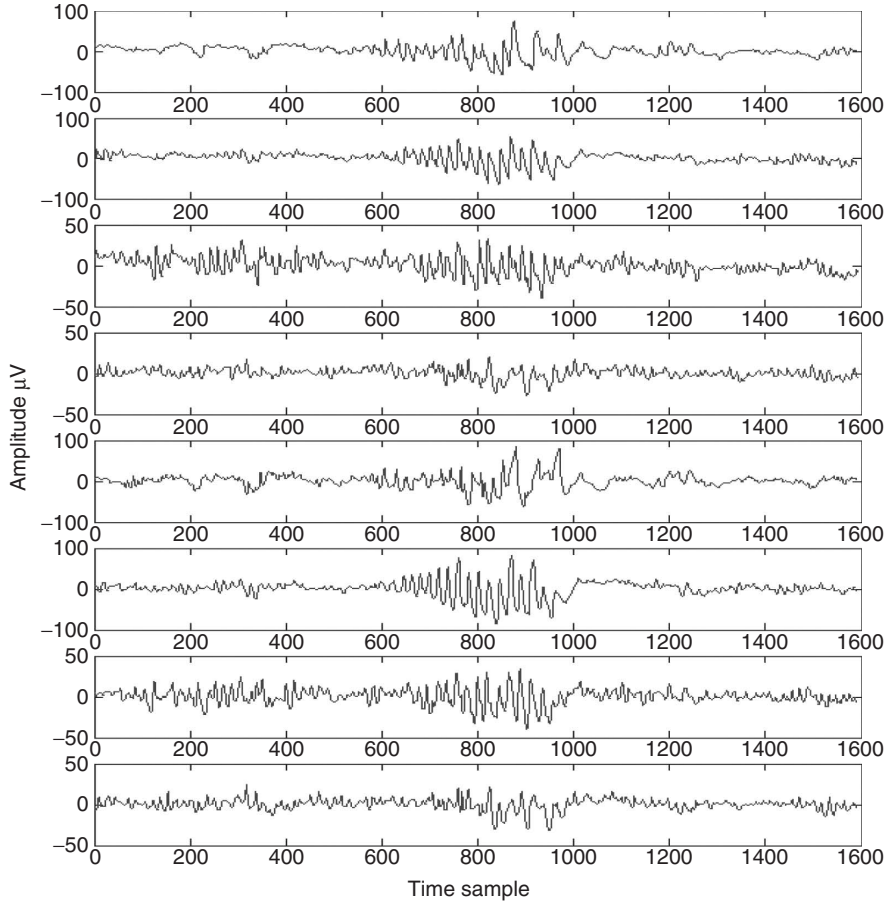

Figure 1.14 Bursts of 3–7 Hz seizure activity in a set of adult EEG signals

be found respectively in younger and older children. These waves may be confined to the regions lying beneath a skull defect. Despite the 6 Hz wave, there are rhythmical theta bursts of wave activities relating to drowsiness around the midtemporal region, with a morphology very similar to ictal patterns. In old age patients other similar patterns may occur, such as *subclinical rhythmic EEG discharges of adults* (SREDA), over the 4–7 Hz frequency band around the centroparietal region, and a wide frequency range (2–120 Hz) *temporal minor sharp transient* and *wicket spikes* over the anterior temporal and midtemporal lobes of the brain. These waves are also nonepileptic but with a seizure-type waveform [56].

The epileptic seizure patterns, called ictal wave patterns, appear during the onset of epilepsy. Although Chapter 4 of this book focuses on an analysis of these waveforms from a signal processing point of view, here a brief explanation of morphology of these waveforms is given. Researchers in signal processing may exploit these concepts in the development of their algorithms. Although these waveform patterns are often highly obscured by muscle movements, they normally maintain certain key characteristics.

Tonic–clonic seizure (also called grand mal) is the most common type of epileptic seizure. It appears in all electrodes but more towards the frontal electrodes (Figure 1.15). It has a rhythmic but spiky pattern in the EEG and occurs within the frequency range of 6–12 Hz. Petit mal is another interictal paroxysmal seizure pattern which occurs at approximately 3 Hz with a generalized synchronous spike wave complex of prolonged bursts. A temporal lobe seizure (also called a psychomotor seizure or complex partial seizure) is presented by bursts of serrated slow waves with a relatively high amplitude of above 60 μV and frequencies of 4–6 Hz. Cortical (focal) seizures have contralateral distribution with rising amplitude and diminishing frequency during the ictal period. The attack is usually initiated by local desynchronization, i.e. very fast and very low voltage spiky activity, which gradually rises in amplitude with diminishing frequency. Myoclonic seizures have concomitant polyspikes, seen clearly in the EEG signals. They can have generalized or bilateral spatial distribution that is more dominant in the frontal region [59]. Tonic seizures occur in patients with the Lennox–Gastaut syndrome [60] and have spikes that repeat with a frequency of approximately 10 Hz. Atonic seizures may appear in the form of a few seconds drop attack or be inhibitory, lasting for a few minutes. They show a few polyspike waves or spike waves with generalized spatial distribution of approximately 10 Hz followed by large slow waves of 1.5–2 Hz [61]. Akinetic seizures are rare and characterized by arrest of all motion, which, however, is not caused by sudden loss of tone as in atonic seizure and the patient is in an absent-like state. They are rhythmic with a frequency of 1–2 Hz. Jackknife seizures, also called salaam attacks, are common in children with hypsarrhythmia (infantile spasms, West syndrome) and are either in the form of sudden generalized flattening desynchronization or have rapid spike discharges [60].

There are generally several varieties of recurring or quasirecurring discharges, which may or may not be related to epileptic seizure. These abnormalities may be due to psychogenic changes, variation in body metabolism, or circulatory insufficiency (which often appears as acute cerebral ischemia). Of these, the most important ones are: periodic or quasiperiodic discharges related to severe CNS diseases; periodic complexes in subacute sclerosing panencephalitis (SSPE); periodic complexes in herpes simplex encephalitis; syncopal attacks; breath holding attacks; hypoglycemia and hyperventilation syndrome due

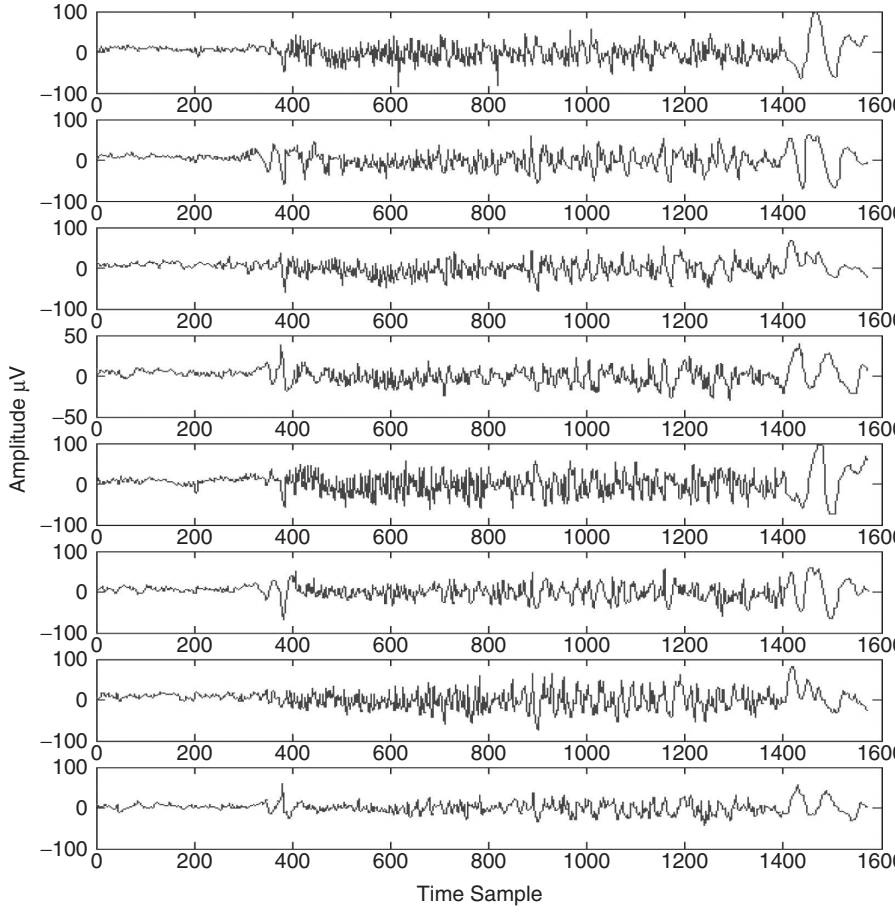

Figure 1.15 Generalized tonic–clonic (grand mal) seizure. The seizure appears in almost all of the electrodes

to sudden changes in blood chemistry [62]; and periodic discharges in Creutzfeldt–Jakob (mad cow) disease [63,64]. The waveforms for this latter abnormality consist of a sharp wave or a sharp triphasic transient signal of 100–300 ms duration, with a frequency of 0.5–2 Hz. The periodic activity usually shows a maximum over the anterior region except for the Heidenhain form, which has a posterior maximum [56]. Other epileptic waveforms include periodic literalized epileptiform discharges (PLED), periodic discharges in acute cerebral anoxia, and periodic discharges of other etiologies.

Despite the above epileptiform signals there are spikes and other paroxysmal discharges in healthy nonepileptic persons. These discharges may be found in healthy individuals without any other symptoms of diseases. However, they are often signs of certain cerebral dysfunctions that may or may not develop into an abnormality. They may appear during periods of particular mental challenge on individuals, such as soldiers in the war front line, pilots, and prisoners.

A comprehensive overview of epileptic seizure disorders and nonepileptic attacks can be found in many books and publications such as References [62] and [65]. In Chapter 5 some recent attempts in application of advanced signal processing techniques to the automatic detection and prediction of epileptic seizures are explained.

1.9.3 Psychiatric Disorders

Not only can functional and certain anatomical brain abnormalities be investigated using EEG signals, pathophysiological brain disorders can also be studied by analysing such signals. According to the *Diagnostic and Statistical Manual (DSM) of Mental Disorders* of the American Psychiatric Association, changes in psychiatric education have evolved considerably since the 1970s. These changes have mainly resulted from physical and neurological laboratory studies based upon EEG signals [66].

There have been evidences from EEG coherence measures suggesting differential patterns of maturation between normal and learning-disabled children [67]. This finding can lead to the establishment of some methodology in monitoring learning disorders. Several psychiatric disorders are diagnosed by analysis of evoked potentials (EPs) achieved by simply averaging a number of consecutive trails having the same stimuli.

A number of pervasive mental disorders cause significant losses in multiple functioning areas [66]. Examples of these are dyslexia, which is a developmental reading disorder; autistic disorder, which is related to abnormal social interaction, communication, and restricted interests and activities, and starts appearing from the age of three; Rett's disorder, characterized by the development of multiple deficits following a period of normal postnatal functioning; and Asperger's disorder, which leads to severe and sustained impairments in social interaction and restricted repetitive patterns of behaviour, interests, and activities.

Attention-deficit hyperactivity disorder (ADHD) and attention-deficit disorder (ADD), conduct disorder, oppositional defiant disorder, and disruptive behaviour disorder have also been under investigation and considered within the DSM. Most of these abnormalities appear during childhood and often prevent children from learning and socializing well. The associated EEG features have been rarely analytically investigated, but the EEG observations are often reported in the literature [68–72]. However, most of such abnormalities tend to disappear with advancing age.

EEG has also been analysed recently for the study of delirium [73,74], dementia [75,76], and many other cognitive disorders [77]. In EEGs, characteristics of delirium include slowing or dropout of the posterior dominant rhythm, generalized theta or delta slow-wave activity, poor organization of the background rhythm, and loss of reactivity of the EEG to eye opening and closing. In parallel with that, the quantitative EEG (QEEG) shows increased absolute and relative slow-wave (theta and delta) power, reduced ratio of fast-to-slow band power, reduced mean frequency, and reduced occipital peak frequency [74].

Dementia includes a group of neurodegenerative diseases that cause acquired cognitive and behavioural impairment of sufficient severity to interfere significantly with social and occupational functioning. Alzheimer disease is the most common of the diseases that cause dementia. At present, the disorder afflicts approximately 5 million people in the United States and more than 30 million people worldwide. A larger number of individuals have lesser levels of cognitive impairment, which frequently evolves into full-blown dementia.

The prevalence of dementia is expected to nearly triple by 2050, since the disorder preferentially affects the elderly, who constitute the fastest-growing age bracket in many countries, especially in industrialized nations [76].

Among other psychiatric and mental disorders, amnestic disorder (or amnesia), mental disorder due to a general medical condition, substance-related disorder, schizophrenia, mood disorder, anxiety disorder, somatoform disorder, dissociative disorder, sexual and gender identity disorder, eating disorders, sleep disorders, impulse-controlled disorder, and personality disorders have often been addressed in the literature [66]. However, the corresponding EEGs have seldom been analysed by means of advanced signal processing tools.

1.9.4 External Effects

EEG signal patterns may significantly change when using drugs for the treatment and suppression of various mental and CNS abnormalities. Variations in EEG patterns may also arise by just looking at the TV screen or listening to music without any attention. However, among the external effects the most significant ones are the pharmacological and drug effects. Therefore, it is important to know the effects of these drugs on the changes of EEG waveforms due to chronic overdosage, and the patterns of overt intoxication [78].

The effect of administration of drugs for anesthesia on EEGs is of interest to clinicians. The related studies attempt to find the correlation between the EEG changes and the stages of anesthesia. It has been shown that in the initial stage of anesthesia a fast frontal activity appears. In deep anesthesia this activity becomes slower with higher amplitudes. In the last stage, a burst-suppression pattern indicates the involvement of brainstem functions, including respiration, and finally the EEG activity ceases [78]. In cases of acute intoxication, the EEG patterns are similar to those of anesthesia [78].

Barbiturate is commonly used as an anticonvulsant and antiepileptic drug. With small dosages of barbiturate the activities within the 25–35 Hz frequency band around the frontal cortex increases. This changes to 15–25 Hz and spreads to the parietal and occipital regions. Dependence and addiction to barbiturates are common. Therefore, after a long-term ingestion of barbiturates, its abrupt withdrawal leads to paroxysmal abnormalities. The major complications are myoclonic jerks, generalized tonic–clonic seizures, and delirium [78].

Many other drugs are used in addition to barbiturates as sleeping pills, such as melatonin and bromides. Very pronounced EEG slowing is found in chronic bromide encephalopathies [78]. Antipsychotic drugs also influence the EEG patterns. For example, neuroleptics increase the alpha wave activity but reduce the duration of beta wave bursts and their average frequency. As another example, clozapine increases the delta, theta, and above 21 Hz beta wave activities. As another antipsychotic drug, tricyclic antidepressants such as imipramine, amitriptyline, doxepin, desipramine, notryptiline, and protriptyline increase the amount of slow and fast activity along with instability of frequency and voltage, and also slow down the alpha wave rhythm. After administration of tricyclic antidepressants the seizure frequency in chronic epileptic patients may increase. With high dosages, this may further lead to single or multiple seizures occurring in nonepileptic patients [78].

During acute intoxication, a widespread, poorly reactive, irregular 8–10 Hz activity and paroxysmal abnormalities including spikes, as well as unspecific coma patterns, are

observed in the EEGs [78]. Lithium is often used in the prophylactic treatment of bipolar mood disorder. The related changes in the EEG pattern consist of slowing of the beta rhythm and of paroxysmal generalized slowing, occasionally accompanied by spikes. Focal slowing also occurs, which is not necessarily a sign of a focal brain lesion. Therefore, the changes in the EEG are markedly abnormal with lithium administration [78]. The beta wave activity is highly activated by using benzodiazepines as an anxiolytic drug. These activities persist in the EEG as long as two weeks after ingestion. Benzodiazepine leads to a decrease in an alpha wave activity and its amplitude, and slightly increases the 4–7 Hz frequency band activity. In acute intoxication the EEG shows prominent fast activity with no response to stimuli [78]. The psychotogentic drugs such as lysergic acid diethylamide and mescaline decrease the amplitude and possibly depress the slow waves [78]. The CNS stimulants increase the alpha and beta wave activities and reduce the amplitude and the amount of slow waves and background EEGs [78].

The effect of many other drugs, especially antiepileptic drugs, is investigated and new achievements are published frequently. One of the significant changes of the EEG of epileptic patients with valproic acid consists of reduction or even disappearance of generalized spikes along with seizure reduction. Lamotrigine is another antiepileptic agent that blocks voltage-gated sodium channels, thereby preventing excitatory transmitter glutamate release. With the intake of lamotrigine a widespread EEG attenuation occurs [78]. Penicillin if administered in high dosage may produce jerks, generalized seizures, or even status epilepticus [78].

1.10 Summary and Conclusions

In this chapter the fundamental concepts in the generation of action potentials and consequently the EEG signals have been briefly explained. The conventional measurement setups for EEG recording and the brain rhythms present in normal or abnormal EEGs have also been described. In addition, the effects of popular brain abnormalities such as mental diseases, ageing, and epileptic and nonepileptic attacks have been pointed out. Despite the known neurological, physiological, pathological, and mental abnormalities of the brain mentioned in this chapter, there are many other brain disorders and dysfunctions that may or may not manifest some kinds of abnormalities in the related EEG signals. Degenerative disorders of the CNS [79], such as a variety of lysosomal disorders, several peroxisomal disorders, a number of mitochondrial disorders, inborn disturbances of the urea cycle, many aminoacidurias, and other metabolic and degenerative diseases, as well as chromosomal aberrations, have to be evaluated and their symptoms correlated with the changes in the EEG patterns. The similarities and differences within the EEGs of these diseases have to be well understood. On the other hand, the developed mathematical algorithms need to take the clinical observations and findings into account in order to enhance the outcome of such processing further. Although a number of technical methods have been well established for the processing of the EEGs with relation to the above abnormalities, there is still a long way to go and many questions to be answered.

The following chapters of this book introduce new digital signal processing techniques employed mainly for analysis of EEG signals followed by a number of examples in the applications of such methods.

References

[1] Caton, R., 'The electric currents of the brain', *Br. Med. J.*, **2**, 1875, 278.

[2] Walter, W. G., 'Slow potential waves in the human brain associated with expectancy, attention and decision', *Arch. Psychiat. Nervenkr.*, **206**, 1964, 309–322.

[3] Cobb, M., 'Exorcizing the animal spirits: Jan Swammerdam on nerve function', *Neuroscience*, **3**, 2002, 395–400.

[4] Danilevsky, V. Y., 'Investigation into the physiology of the brain' [in Russian], Doctoral Thesis, University of Kharkov, 1877, Zit. Nach: Brazier MAB; A history of Neurophysiology in the 19th Century, New York: Raven; 1988, 208.

[5] Brazier, M. A. B., *A History of the Electrical Activity of the Brain; The First Half-Century*, Macmillan, New York, 1961.

[6] Massimo, A., 'In Memoriam Pierre Gloor (1923–2003): an appreciation', *Epilepsia*, **45**(7), July 2004, 882.

[7] Grass, A. M., and Gibbs, F. A., 'A Fourier transform of the electroencephalogram', *J. Neurophysiol.*, **1**, 1938, 521–526.

[8] Haas, L. F., 'Hans Berger (1873–1941), Richard Caton (1842–1926), and electroencephalography', *J. Neurol. Neurosurg. Psychiat.*, **74**, 2003, 9.

[9] Spear, J. H., 'Cumulative change in scientific production: research technologies and the structuring of new knowledge', *Perspectives on Sci.*, **12**(1), 2004, 55–85.

[10] Shipton, H. W., 'EEG analysis: a history and prospectus', *Annual Rev., Univ. of Iowa, USA*, 1975, 1–15.

[11] Fischer, M. H., 'Elektrobiologische Auswirkungen von Krampfgiften am Zentralnervensystem', *Med. Klin.*, **29**, 1933, 15–19.

[12] Fischer, M. H., and Lowenbach, H., 'Aktionsstrome des Zentralnervensystems unter der Einwirkung von Krampfgiften, 1. Mitteilung Strychnin und Pikrotoxin', *Arch. F. Exp. Pathol. und Pharmakol.*, **174**, 1934, 357–382.

[13] Kornmuller, A. E., 'Der Mechanismus des Epileptischen Anfalles auf Grund Bioelektrischer Untersuchungen am Zentralnervensystem', *Fortschr. Neurol. Psychiatry*, **7**, 1935, 391–400; 414–432.

[14] Bremer, F., 'Cerveau isole' et physiologie du sommeil', *C.R. Soc. Biol. (Paris)*, **118**, 1935, 1235–1241.

[15] Niedermeyer, E., 'Historical aspects', Chapter 1, *Electroencephalography, Basic Principles, Clinical Applications, and Related Fields*, Eds E. Niedermeyer and F. Lopes da Silva, 4th edn., Lippincott, Williams and Wilkins, Philadelphia, Pennsylvania, 1999, 1–14.

[16] Berger, H., 'Uber das Elektrenkephalogramm des Menschen', *Arch. Psychiatr. Nervenkr.*, **87**, 1929, 527–580.

[17] Jasper, H., 'Report of committee on methods of clinical exam in EEG', *Electroencephalogr. Clin. Neurophysiol.*, **10**, 1958, 370–375.

[18] Motokawa, K., 'Electroencephalogram of man in the generalization and differentiation of condition reflexes', *Tohoku J. Expl. Medicine*, **50**, 1949, 225.

[19] Niedermeyer, E., 'Common generalized epilepsy. The so-called idiopathic or centrencephalic epilepsy', *Eur. Neurol.*, **9**(3), 1973, 133–156.

[20] Aserinsky, E., and Kleitman, N., 'Regularly occurring periods of eye motility, and concomitant phenomena, during sleep', *Science*, **118**, 1953, 273–274.

[21] Speckmann, E.-J., and Elger, C. E., 'Introduction to the neurophysiological basis of the EEG and DC potentials', in *Electroencephalography Basic Principles, Clinical Applications, and Related Fields*, Eds E. Niedermeyer and F. Lopes da Silva, 4th edn, Lippincott, Williams and Wilkins, Philadelphia, Pennsylvania, 1999.

[22] Shepherd, G. M., *The Synaptic Organization of the Brain*, Oxford University Press, London, 1974.

[23] Caspers, H., Speckmann E.-J., and Lehmenkühler, A., 'DC potentials of the cerebral cortex, seizure activity and changes in gas pressures', *Rev. Physiol., Biochem. Pharmacol.*, **106**, 1986, 127–176.

[24] Ka Xiong Charand, `http://hyperphysics.phy-astr.gsu.edu/hbase/biology/actpot.html`.

[25] Attwood, H. L., and MacKay, W. A., *Essentials of Neurophysiology*, B. C. Decker, Hamilton, Canada, 1989.

[26] Nunez, P. L., *Neocortical Dynamics and Human EEG Rhythms*, Oxford University Press, New York, 1995.

[27] Teplan, M., 'Fundamentals of EEG measurements', *Measmt Sci. Rev.*, **2**(2), 2002.

[28] Bickford, R. D., 'Electroencephalography', in *Encyclopedia of Neuroscience*, Ed. G. Adelman, Birkhauser, Cambridge (USA), 1987, 371–373.

[29] Sterman, M. B., MacDonald, L. R., and Stone, R. K., 'Biofeedback training of sensorimotor EEG in man and its effect on epilepsy', *Epilepsia*, **15**, 1974, 395–416.

[30] Ashwal, S., and Rust, R., 'Child neurology in the 20th century', *Pedia. Res.*, **53**, 2003, 345–361.

[31] Niedermeyer, E., 'The normal EEG of the waking adult', Chapter 10, in *Electroencephalography, Basic Principles, Clinical Applications, and Related Fields*, Eds E. Niedermeyer and F. Lopes da Silva, 4th edn, Lippincott, Williams and Wilkins, Philadelphia, Pennsylvania, 1999, 174–188.

[32] Pfurtscheller, G., Flotzinger, D., and Neuper, C., 'Differentiation between finger, toe and tongue movement in man based on 40 Hz EEG', *Electroencephalogr. Clin. Neurophysiol.*, **90**, 1994, 456–460.

[33] Adrian, E. D., and Mattews, B. H. C., 'The Berger rhythm, potential changes from the occipital lob in man', *Brain*, **57**, 1934, 345–359.

[34] Trabka, J., 'High frequency components in brain waves', *Electroencephalogr. Clin. Neurophysiol.*, **14**, 1963, 453–464.

[35] Cobb, W. A., Guiloff, R. J., and Cast, J., 'Breach rhythm: the EEG related to skull defects', *Electroencephalogr. Clin. Neurophysiol.*, **47**, 1979, 251–271.

[36] Silbert, P. L., Radhakrishnan, K., Johnson, J., and Class, D. W., 'The significance of the phi rhythm', *Electroencephalogr. Clin. Neurophysiol.*, **95**, 1995, 71–76.

[37] Roldan, E., Lepicovska, V., Dostalek, C., and Hrudova, L., 'Mu-like EEG rhythm generation in the course of Hatha-yogi exercises', *Electroencephalogr. Clin. Neurophysiol.*, **52**, 1981, 13.

[38] IFSECN, 'A glossary of terms commonly used by clinical electroencephalographers', *Electroencephalogr. Clin. Neurophysiol.*, **37**, 1974, 538–548.

[39] O'Leary, J. L., and Goldring, S., *Science and Epilepsy*, Raven Press, New York, 1976, pp. 19–152.

[40] Gotman, J., Ives, J. R., and Gloor, R., 'Automatic recognition of interictal epileptic activity in prolonged EEG recordings', *Electroencephalogr. Clin. Neurophysiol.*, **46**, 1979, 510–520.

[41] 'Effects of electrode placement', `http://www.focused-technology.com/electrod.htm`, California.

[42] Collura, T., *A Guide to Electrode Selection, Location, and Application for EEG Biofeedback*, Ohio, BrainMaster Technologies, Inc. 1998.

[43] Nayak, D., Valentin, A., Alarcon, G., Seoane, J. J. G., Brunnhuber, F., Juler, J., Polkey, C. E., and Binnie, C. D., 'Characteristics of scalp electrical fields associated with deep medial temporal epileptiform discharges', *Clin. Neurophysiol.*, **115**, 2004, 1423–1435.

[44] Barrett, G., Blumhardt, L., Halliday, L., Halliday, A. M., and Kriss, A., 'A paradox in the lateralization of the visual evoked responses', *Nature*, **261**, 1976, 253–255.

[45] Halliday, A. M., Evoked potentials in neurological disorders', in *Event-Related Brain Potentials in Man*, Eds E. Calloway, P. Tueting, and S. H. Coslow, Academic Press, New York, 1978, 197–210.

[46] Sharbrough, F. W., 'Nonspecific abnormal EEG patterns', Chapter. 12, in *Electroencephalography, Basic Principles, Clinical Applications, and Related Fields*, Eds E. Niedermeyer and F. Lopes Da Silva, 4th edn., Lippincott, Williams and Wilkins, Philadelphia, Pennsylvania, 1999.

[47] Bancaud, J., Hecaen, H., and Lairy, G. C., 'Modification de la reactivite EEG, troubles des functions symboliques et troubles con fusionels dans les lesions hemispherigues localisees', *Electroencephalogr. Clin. Neurophysiol.*, **7**, 1955, 179.

[48] Westmoreland, B., and Klass, D., 'Asymmetrical attention of alpha activity with arithmetical attention', *Electroencephalogr. Clin. Neurophysiol.*, **31**, 1971, 634–635.

[49] Cobb, W., 'EEG interpretation in clinical medicine', Part B, in *Handbook of Electroencephalography and Clinical Neurophysiology*, Ed. A. Remond, Amsterdam, Vol. 11, Elsevier, 1976.

[50] Hess, R., 'Brain tumors and other space occupying processing', Part C, in *Handbook of Electroencephalography and Clinical Neurophysiology*, Ed. A. Remond, Amsterdam, Vol. 14, Elsevier, 1975.

[51] Klass, D., and Daly, D. (Eds), *Current Practice of Clinical Electroencephalography*, 1st edn. Raven Press, 1979.

[52] Van Sweden, B., Wauquier, A., and Niedermeyer, E., 'Normal aging and transient cognitive disorders in the elderly', Chapter 18, in *Electroencephalography, Basic Principles, Clinical Applications, and Related Fields*, Eds E. Niedermeyer and F. Lopes da Silva, 4th edn, Lippincott, Williams and Wilkins, Philadelphia, Pennsylvania, 1999, 340–348.

[53] America Psychiatric Association, Committee on Nomenclature and Statistics, *Diagnostic and Statistical Manual of Mental Disorder: DSM-IV*, 4th edn., American Psychiatric Association, Washington, DC, 1994.

[54] Brenner, R. P., 'EEG and dementia', Chapter 19, in *Electroencephalography, Basic Principles, Clinical Applications, and Related Fields*, Eds E. Niedermeyer and F. Lopes da Silva, 4th edn., Lippincott, Williams and Wilkins, Philadelphia, Pennsylvania, 1999, 349–359.

[55] Neufeld, M. Y., Bluman, S., Aitkin, I., Parmet, Y., and Korczyn, A. D., 'EEG frequency analysis in demented and nondemented Parkinsonian patients', *Dementia*, **5**, 1994, 23–28.

[56] Niedermeyer, E., 'Abnormal EEG patterns: epileptic and paroxysmal', Chapter 13, in *Electroencephalography, Basic Principles, Clinical Applications, and Related Fields*, Eds E. Niedermeyer and F. Lopes da Silva, 4th edn, Lippincott, Williams and Wilkins, Philadelphia, Pennsylvania, 1999, 235–260.

[57] Hughes, J. R., and Gruener, G. T., 'Small sharp spikes revisited: further data on this controversial pattern', *Electroencephalogr. Clin. Neurophysiol.*, **15**, 1984, 208–213.

[58] Hecker, A., Kocher, R., Ladewig, D., and Scollo-Lavizzari, G., 'Das Minature-Spike-Wave', *Das EEG Labor*, **1**, 1999 51–56.

[59] Geiger, L. R., and Harner, R. N., 'EEG patterns at the time of focal seizure onset', *Arch. Neurol.*, **35**, 1978, 276–286.

[60] Gastaut, H., and Broughton, R., *Epileptic Seizure*, Charles C. Thomas, Springfield, Illinois, 1972.

[61] Oller-Daurella, L., and Oller-Ferrer-Vidal, L., *Atlas de Crisis Epilepticas*, Geigy Division Farmaceut, Spain, Barcelona 1977.

[62] Niedermeyer, E., 'Nonepileptic Attacks', Chapter 28, in *Electroencephalography, Basic Principles, Clinical Applications, and Related Fields*, Eds E. Niedermeyer and F. Lopes da Silva, 4th edn, Lippincott, Williams and Wilkins, Philadelphia, Pennsylvania, 1999, 586–594.

[63] Creutzfeldt, H. G., 'Uber eine Eigenartige Herdformige Erkrankung des Zentralnervensystems', *Z. Ges. Neurol. Psychiatr.*, **57**, 1968, 1, Quoted after W. R. Kirschbaum, 1920.

[64] Jakob, A., 'Uber Eigenartige Erkrankung des Zentralnervensystems mit Bemerkenswerten Anatomischen Befunden (Spastistische Pseudosklerose, Encephalomyelopathie mit Disseminerten Degenerationsbeschwerden)', *Deutsch. Z. Nervenheilk*, **70**, 1968, 132, Quoted after W. R. Kirschbaum, 1921.

[65] Niedermeyer, E., 'Epileptic seizure disorders', Chapter 27, in *Electroencephalography, Basic Principles, Clinical Applications, and Related Fields*, Eds E. Niedermeyer and F. Lopes da Silva, 4th edn, Lippincott, Williams and Wilkins, Philadelphia, Pennsylvania, 1999, 476–585.

[66] Small, J. G., 'Psychiatric disorders and EEG', Chapter 30, in *Electroencephalography, Basic Principles, Clinical Applications, and Related Fields*, Eds E. Niedermeyer and F. Lopes da Silva, 4th ed., Lippincott, Williams and Wilkins, Philadelphia, Pennsylvania, 1999, 235–260.

[67] Marosi, E., Harmony, T., Sanchez, L., Becker, J., Bernal, J., Reyes, A., Diaz de Leon, A. E., Rodriguez, M., and Fernandez, T., 'Maturation of the coherence of EEG activity in normal and learning disabled children', *Electroencephalogr. Clin. Neurophysiol.* **83**, 1992, 350–357.

[68] Linden, M., Habib, T., and Radojevic, V., 'A controlled study of the effects of EEG biofeedback on cognition and behavior of children with attention deficit disorder and learning disabilities', *Biofeedback Self Regul.*, **21**(1), 1996, pp. 35–49.

[69] Hermens, D. F., Soei, E. X., Clarke, S. D., Kohn, M. R., Gordon, E., and Williams, L. M., 'Resting EEG theta activity predicts cognitive performance in attention-deficit hyperactivity disorder', *Pediatr. Neurol.*, **32**(4), 2005, 248–256.

[70] Swartwood, J. N., Swartwood, M. O., Lubar, J. F., and Timmermann, D. L., 'EEG differences in ADHD-combined type during baseline and cognitive tasks', *Pediatr. Neurol.*, **28**(3), 2003, 199–204.

[71] Clarke, A. R., Barry, R. J., McCarthy, R., and Selikowitz, M., 'EEG analysis of children with attention-deficit/hyperactivity disorder and comorbid reading disabilities', *J. Learn. Disabil.*, **35**(3), 2002, 276–285.

[72] Yordanova, J., Heinrich, H., Kolev, V., and Rothenberger, A., 'Increased event-related theta activity as a psychophysiological marker of comorbidity in children with tics and attention-deficit/hyperactivity disorders', *Neuroimage*, **32**(2), 2006, 940–955.

[73] Jacobson, S., and Jerrier, H., 'EEG in delirium', *Semin. Clin. Neuropsychiat.*, **5**(2), 2000, 86–92.

[74] Onoe, S., and Nishigaki, T., 'EEG spectral analysis in children with febrile delirium', *Brain Devel.*, **26**(8), 2004, 513–518.

[75] Brunovsky, M., Matousek, M., Edman, A., Cervena, K., and Krajca, V., 'Objective assessment of the degree of dementia by means of EEG', *Neuropsychobiology*, **48**(1), 2003, 19–26.

[76] Koenig, T., Prichep, L., Dierks, T., Hubl, D., Wahlund, L. O., John, E. R., and Jelic, V., 'Decreased EEG synchronization in Alzheimer's disease and mild cognitive impairment', *Neurobiol. Aging*, **26**(2), 2005, 165–171.

[77] Babiloni, C., Binetti, G., Cassetta, E., Dal Forno, G., Del Percio, C., Ferreri, F., Ferri, R., Frisoni, G., Hirata, K., Lanuzza, B., Miniussi, C., Moretti, D. V., Nobili, F., Rodriguez, G., Romani, G. L., Salinari, S., and Rossini, P. M., 'Sources of cortical rhythms change as a function of cognitive impairment in pathological aging: a multicenter study', *Clin. Neurophysiol.*, **117**(2), 2006, 252–268.

[78] Bauer, G., and Bauer, R., 'EEG, drug effects, and central nervous system poisoning', Chapter 35, in *Electroencephalography, Basic Principles, Clinical Applications, and Related Fields*, Eds E. Niedermeyer and F. Lopes da Silva, 4th edn, Lippincott, Williams and Wilkins, Philadelphia, Pennsylvania, 1999, 671–691.

[79] Naidu, S. and Niedermeyer, E., 'Degenerative disorders of the central nervous system', Chapter 20, in *Electroencephalography, Basic Principles, Clinical Applications, and Related Fields*, Eds E. Niedermeyer and F. Lopes da Silva, 4th edn., Lippincott, Williams and Wilkins, Philadelphia, Pennsylvania, 1999, 360–382.

2

Fundamentals of EEG Signal Processing

EEG signals are the signatures of neural activities. They are captured by multiple-electrode EEG machines either from inside the brain, over the cortex under the skull, or certain locations over the scalp, and can be recorded in different formats. The signals are normally presented in the time domain, but many new EEG machines are capable of applying simple signal processing tools such as the Fourier transform to perform frequency analysis and equipped with some imaging tools to visualize EEG topographies (maps of the brain activities in the spatial domain).

There have been many algorithms developed so far for processing EEG signals. The operations include, but are not limited to, time-domain analysis, frequency-domain analysis, spatial-domain analysis, and multiway processing. Also, several algorithms have been developed to visualize the brain activity from images reconstructed from only the EEGs. Separation of the desired sources from the multisensor EEGs has been another research area. This can later lead to the detection of brain abnormalities such as epilepsy and the sources related to various physical and mental activities. In Chapter 7 of this book it can be seen that the recent works in brain–computer interfacing (BCI) [1] have been focused upon the development of advanced signal processing tools and algorithms for this purpose.

Modelling of neural activities is probably more difficult than modelling the function of any other organ. However, some simple models for generating EEG signals have been proposed. Some of these models have also been extended to include generation of abnormal EEG signals.

Localization of brain signal sources is another very important field of research [2]. In order to provide a reliable algorithm for localization of the sources within the brain sufficient knowledge about both propagation of electromagnetic waves and how the information from the measured signals can be exploited in separation and localization of the sources within the brain is required. The sources might be considered as magnetic dipoles for which the well-known inverse problem has to be solved, or they can be considered as distributed current sources.

EEG Signal Processing S. Sanei and J. Chambers

Patient monitoring and sleep monitoring require real-time processing of (up to a few days) long EEG sequences. The EEG provides important and unique information about the sleeping brain. Major brain activities during sleep can be captured using the developed algorithms [3], such as the method of matching pursuits (MPs) discussed [4] later in this chapter.

Epilepsy monitoring, detection, and prediction have also attracted many researchers. Dynamical analysis of a time series together with the application of blind separation of the signal sources has enabled prediction of focal epilepsies from the scalp EEGs. On the other hand, application of time–frequency-domain analysis for detection of the seizure in neonates has paved the way for further research in this area.

In the following sections most of the tools and algorithms for the above objectives are explained and the mathematical foundations discussed. The application of these algorithms to analysis of the normal and abnormal EEGs, however, will follow in later chapters of this book. The reader should also be aware of the required concepts and definitions borrowed from linear algebra, further details of which can be found in Reference [5]. Throughout this chapter and the reminder of this book continuous time is denoted by t and discrete time, with normalized sampling period $T = 1$, by n.

2.1 EEG Signal Modelling

Most probably the earliest physical model is based on the Hodgkin and Huxley's Nobel Prize winning model for the squid axon published in 1952 [6–8]. A nerve axon may be stimulated and the activated sodium (Na^+) and potassium (K^+) channels produced in the vicinity of the cell membrane may lead to the electrical excitation of the nerve axon. The excitation arises from the effect of the membrane potential on the movement of ions, and from interactions of the membrane potential with the opening and closing of voltage-activated membrane channels. The membrane potential increases when the membrane is polarized with a net negative charge lining the inner surface and an equal but opposite net positive charge on the outer surface. This potential may be simply related to the amount of electrical charge Q, using

$$E = Q/C_{\mathrm{m}} \tag{2.1}$$

where Q is in terms of coulombs/cm^2, C_{m} is the measure of the capacity of the membrane in units of farads/cm^2, and E is in units of volts. In practice, in order to model the action potentials (APs) the amount of charge Q^+ on the inner surface (and Q^- on the outer surface) of the cell membrane has to be mathematically related to the stimulating current I_{stim} flowing into the cell through the stimulating electrodes. The electrical potential (often called the electrical force) E is then calculated using Equation (2.1). The Hodgkin and Huxley model is illustrated in Figure 2.1. In this figure I_{memb} is the result of positive charges flowing out of the cell. This current consists of three currents, namely Na, K, and leak currents. The leak current is due to the fact that the inner and outer Na and K ions are not exactly equal.

Hodgkin and Huxley estimated the activation and inactivation functions for the Na and K currents and derived a mathematical model to describe an AP similar to that of a

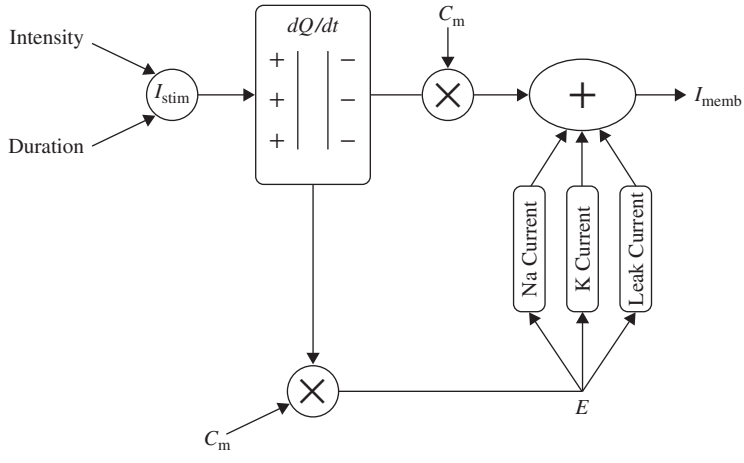

Figure 2.1 The Hodgkin–Huxley excitation model

giant squid. The model is a neuron model that uses voltage-gated channels. The space-clamped version of the Hodgkin–Huxley model may be well described using four ordinary differential equations [9]. This model describes the change in the membrane potential (E) with respect to time and is described in Reference [10]. The overall membrane current is the sum of capacity current and ionic current, i.e.

$$I_{\text{memb}} = C_{\text{m}} \frac{\mathrm{d}E}{\mathrm{d}t} + I_{\text{i}} \tag{2.2}$$

where I_{i} is the ionic current and, as indicated in Figure 2.1, can be considered as the sum of three individual components: Na, K, and leak currents:

$$I_{\text{i}} = I_{\text{Na}} + I_{\text{K}} + I_{\text{leak}} \tag{2.3}$$

I_{Na} can be related to the maximal conductance $\overline{g}_{\text{Na}}$, activation variable a_{Na}, inactivation variable h_{Na}, and a driving force $(E - E_{\text{Na}})$ through

$$I_{\text{Na}} = \overline{g}_{\text{Na}} a_{\text{Na}}^3 h_{\text{Na}} (E - E_{\text{Na}}) \tag{2.4}$$

Similarly, I_{K} can be related to the maximal conductance $\overline{g}_{\text{K}}$, activation variable a_{Na}, inactivation variable a_{K}, and a driving force $(E - E_{\text{K}})$ as

$$I_{\text{K}} = \overline{g}_{\text{K}} a_{\text{K}} (E - E_{\text{K}}) \tag{2.5}$$

and I_{leak} is related to the maximal conductance $\overline{g}_{\text{l}}$ and a driving force $(E - E_{\text{l}})$ as

$$I_{\text{l}} = \overline{g}_{\text{l}} (E - E_{\text{l}}) \tag{2.6}$$

The changes in the variables a_{Na}, a_{K}, and h_{Na} vary from 0 to 1 according to the following equations:

$$\frac{\mathrm{d}a_{\mathrm{Na}}}{\mathrm{d}t} = \lambda_t[\alpha_{\mathrm{Na}}(E)(1 - a_{\mathrm{Na}}) - \beta_{\mathrm{Na}}(E)a_{\mathrm{Na}}] \tag{2.7}$$

$$\frac{\mathrm{d}h_{\mathrm{Na}}}{\mathrm{d}t} = \lambda_t[\alpha_{\mathrm{h}}(E)(1 - h_{\mathrm{Na}}) - \beta_{\mathrm{h}}(E)h_{\mathrm{Na}}] \tag{2.8}$$

$$\frac{\mathrm{d}a_{\mathrm{K}}}{\mathrm{d}t} = \lambda_t[\alpha_{\mathrm{K}}(E)(1 - a_{\mathrm{K}}) - \beta_{\mathrm{K}}(E)a_{\mathrm{K}}] \tag{2.9}$$

where $\alpha(E)$ and $\beta(E)$ are respectively forward and backward rate functions and λ_t is a temperature-dependent factor. The forward and backward parameters depend on voltage and were empirically estimated by Hodgkin and Huxley as

$$\alpha_{\mathrm{Na}}(E) = \frac{3.5 + 0.1E}{1 - \mathrm{e}^{-(3.5+0.1E)}} \tag{2.10}$$

$$\beta_{\mathrm{Na}}(E) = 4\mathrm{e}^{-(E+60)/18} \tag{2.11}$$

$$\alpha_{\mathrm{h}}(E) = 0.07\mathrm{e}^{-(E+60)/20} \tag{2.12}$$

$$\beta_{\mathrm{h}}(E) = \frac{1}{1 + \mathrm{e}^{-(3+0.1E)}} \tag{2.13}$$

$$\alpha_{\mathrm{K}}(E) = \frac{0.5 + 0.01E}{1 - \mathrm{e}^{-(5+0.1E)}} \tag{2.14}$$

$$\beta_{\mathrm{K}}(E) = 0.125\mathrm{e}^{-(E+60)/80} \tag{2.15}$$

As stated in the Simulator for Neural Networks and Action Potentials (SNNAP) literature [9], the $\alpha(E)$ and $\beta(E)$ parameters have been converted from the original Hodgkin–Huxley version to agree with the present physiological practice, where depolarization of the membrane is taken to be positive. In addition, the resting potential has been shifted to −60 mV (from the original 0 mV). These equations are used in the model described in the SNNAP. In Figure 2.2 an AP has been simulated. For this model the parameters are set to $C_{\mathrm{m}} =$ 1.1 uF/cm^2, $\overline{g}_{\mathrm{Na}} = 100$ ms/cm^2, $\overline{g}_{\mathrm{K}} = 35$ ms/cm^2, $\overline{g}_{\mathrm{l}} = 0.35$ ms/cm^2, and $E_{\mathrm{Na}} = 60$ mV.

The simulation can run to generate a series of action potentials, as happens in practice in the case of ERP signals. If the maximal ionic conductance of the potassium current, $\overline{g}_{\mathrm{K}}$, is reduced the model will show a higher resting potential. Also, for $\overline{g}_{\mathrm{K}} = 16$ ms/cm^2, the model will begin to exhibit oscillatory behaviour. Figure 2.3 shows the result of a Hodgkin–Huxley oscillatory model with reduced maximal potassium conductance.

The SNNAP can also model bursting neurons and central pattern generators. This stems from the fact that many neurons show cyclic spiky activities followed by a period of inactivity. Several invertebrate as well as mammalian neurons are bursting cells and exhibit alternating periods of high-frequency spiking behaviour followed by a period of no spiking activity.

A simpler model than that due to Hodgkin–Huxley for simulating spiking neurons is the Morris–Lecar model [11]. This model is a minimal biophysical model, which generally exhibits single action potential. This model considers that the oscillation of a slow

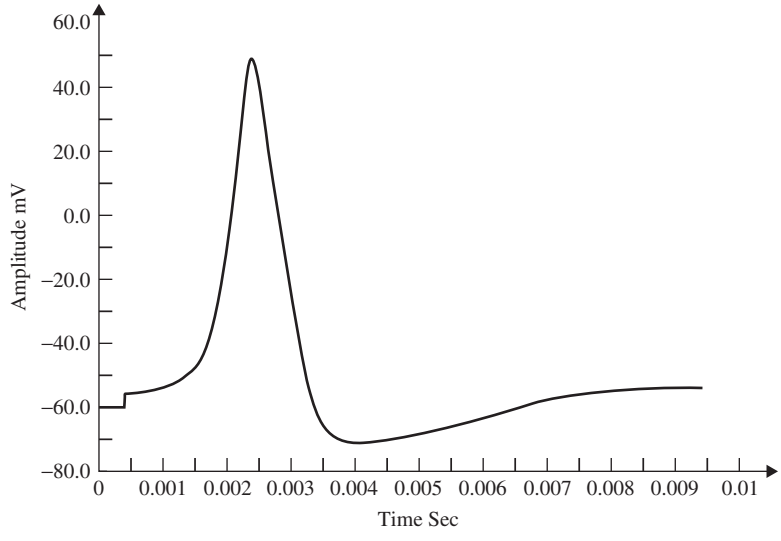

Figure 2.2 A single AP in response to a transient stimulation based on the Hodgkin–Huxley model. The initiated time is at t = 0.4 ms and the injected current is 80 μA/cm^2 for a duration of 0.1 ms. The selected parameters are $C_m = 1.2$ uF/cm^2, $\overline{g}_{Na} = 100$ mS/cm^2, $\overline{g}_K = 35$ ms/cm^2, $\overline{g}_l = 0.35$ ms/cm^2, and $E_{Na} = 60$ mV

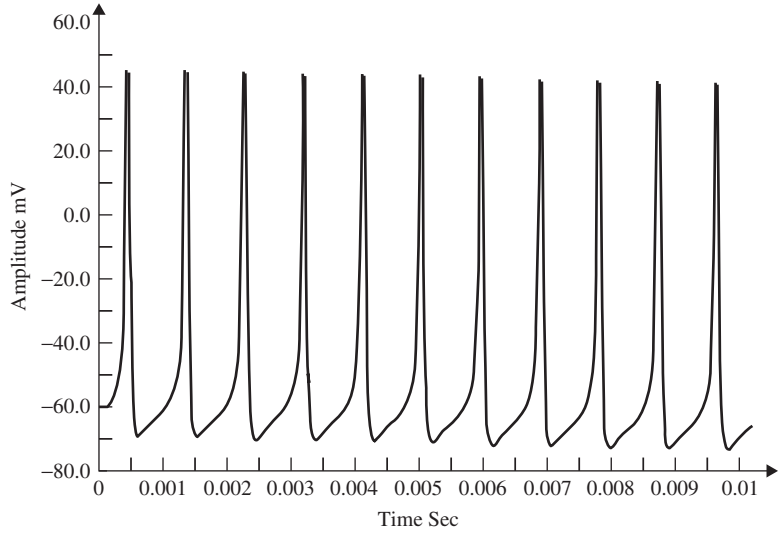

Figure 2.3 The AP from a Hodgkin–Huxley oscillatory model with reduced maximal potassium conductance

calcium wave that depolarizes the membrane leads to a bursting state. The Morris–Lecar model was initially developed to describe the behaviour of barnacle muscle cells. The governing equations relating the membrane potential (E) and potassium activation w_K to the activation parameters are given as

$$C\frac{dE}{dt} = I_i - \overline{g}_{Ca}a_{Ca}(E)(E - E_{Ca}) - \overline{g}_K w_K(E - E_K) - \overline{g}_l(E - E_l) \tag{2.16}$$

$$\frac{dw_K}{dt} = \lambda_t\left(\frac{w_\infty(E) - w_K}{\tau_K(E)}\right) \tag{2.17}$$

where I_i is the combination of three ionic currents, calcium (Ca), potassium (K), and leak (l), and, similar to the Hodgkin–Huxley model, are products of a maximal conductance $\overline{g}$, activation components (in such as a_{Ca}, w_K), and the driving force E. The changes in the potassium activation variable w_K is proportional to a steady-state activation function $w_K(E)$ (a sigmoid curve) and a time-constant function $\tau_K(E)$ (a bell-shaped curve). These functions are respectively defined as

$$w_\infty(E) = \frac{1}{1 + e^{-(E-h_w)/S_w}} \tag{2.18}$$

$$\tau_K(E) = \frac{1}{e^{(E-h_w)/(2S_w)} + e^{-(E-h_w)/(2S_w)}} \tag{2.19}$$

The steady-state activation function $a_{Ca}(E)$, involved in calculation of the calcium current, is defined as

$$a_{Ca}(E) = \frac{1}{1 + e^{-(E-h_{Ca})/s_m}} \tag{2.20}$$

Similar to the sodium current in the Hodgkin–Huxley model, the calcium current is an inward current. Since the calcium activation current is a fast process in comparison with the potassium current, it is modelled as an instantaneous function. This means that for each voltage E, the steady-state function $a_{Ca}(E)$ is calculated. The calcium current does not incorporate any inactivation process. The activation variable w_K here is similar to a_K in the Hodgkin–Huxley model, and finally the leak currents for both models are the same [9]. A simulation of the Morris–Lecar model is presented in Figure 2.4.

Calcium-dependent potassium channels are activated by intracellular calcium; the higher the calcium concentration the higher the channel activation [9]. For the Morris–Lecar model to exhibit bursting behaviour, the two parameters of maximal time constant and the input current have to be changed [9]. Figure 2.5 shows the bursting behaviour of the Morris–Lecar model. The basic characteristics of a bursting neuron are the duration of the spiky activity, the frequency of the action potentials during a burst, and the duration of the quiescence period. The period of an entire bursting event is the sum of both active and quiescence duration [9].

Neurons communicate with each other across synapses through axon–dendrites or dendrites–dendrites connections, which can be excitatory, inhibitory, or electric [9]. By combining a number of the above models a neuronal network can be constructed. The network exhibits oscillatory behaviour due to the synaptic connection between the neurons.

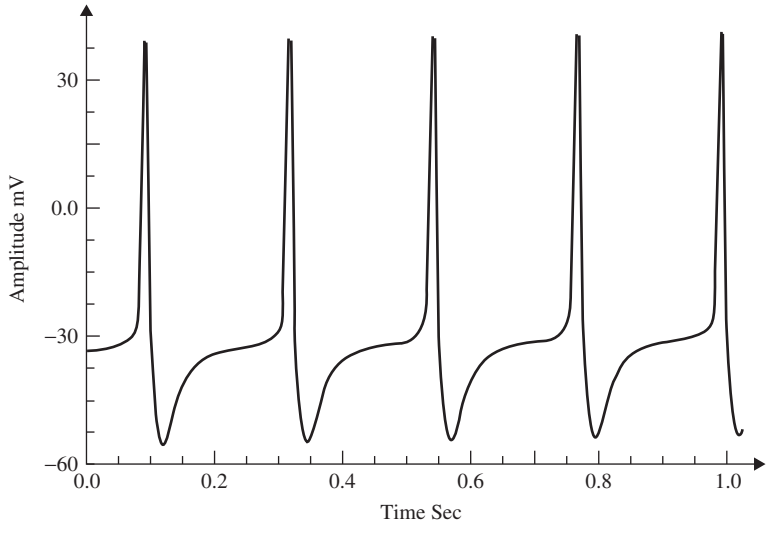

Figure 2.4 Simulation of an AP within the Morris–Lecar model. The model parameters are: $C_m = 22$ uF/cm^2, $\overline{g}_{Ca} = 3.8$ ms/cm^2, $\overline{g}_K = 8.0$ ms/cm^2, $\overline{g}_l = 1.6$ ms/cm^2, $E_{Ca} = 125$ mV, $E_K = -80$ mV, $E_l = -60$ mV, $\lambda_t = 0.06$, $h_{Ca} = -1.2$, and $S_m = 8.8$

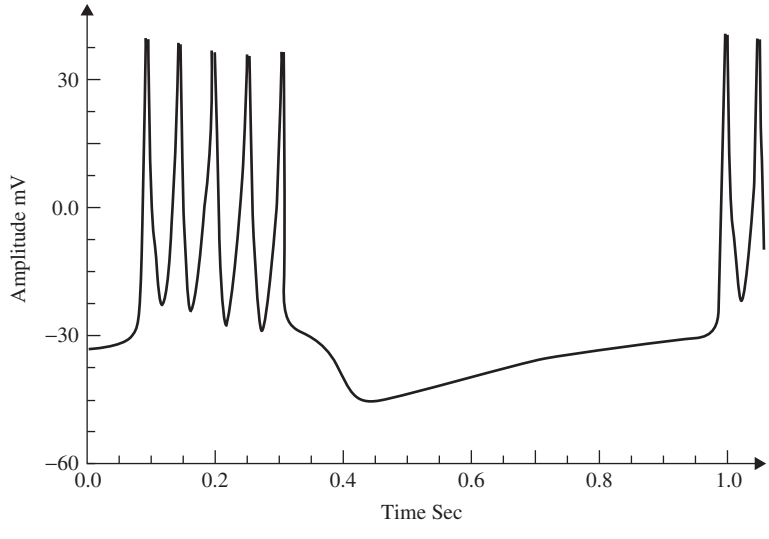

Figure 2.5 An illustration of the bursting behaviour that can be generated by the Morris–Lecar model

A synaptic current is produced as soon as a neuron fires an AP. This current stimulates the connected neuron and may be modelled by an alpha function multiplied by a maximal conductance and a driving force as

$$I_{\text{syn}} = \overline{g}_{\text{syn}}\ g_{\text{syn}}(t)[E(t) - E_{\text{syn}}] \tag{2.21}$$

where

$$g_{\text{syn}}(t) = t\ \mathrm{e}^{(-t/u)} \tag{2.22}$$

and t is the latency or time since the trigger of the synaptic current, u is the time to reach to the peak amplitude, E_{syn} is the synaptic reversal potential, and $\overline{g}_{\text{syn}}$ is the maximal synaptic conductance. The parameter u alters the duration of the current while $\overline{g}_{\text{syn}}$ changes the strength of the current. This concludes the treatment of the modelling of APs.

As the nature of the EEG sources cannot be determined from the electrode signals directly, many researchers have tried to model these processes on the basis of information extracted using signal processing techniques. The method of linear prediction described in the later sections of this chapter is frequently used to extract a parametric description.

2.1.1 Linear Models

2.1.1.1 Prediction Method

The main objective of using prediction methods is to find a set of model parameters that best describe the signal generation system. Such models generally require a noise-type input. In autoregressive (AR) modelling of signals each sample of a single-channel EEG measurement is defined to be linearly related with respect to a number of its previous samples, i.e.

$$y(n) = -\sum_{k=1}^{p} a_k y(n-k) + x(n) \tag{2.23}$$

where $a_k, k = 1, 2, \ldots, p$, are the linear parameters, n denotes the discrete sample time normalized to unity, and $x(n)$ is the noise input. In an autoregressive moving average (ARMA) linear predictive model each sample is obtained based on a number of its previous input and output sample values, i.e.

$$y(n) = -\sum_{k=1}^{p} a_k y(n-k) + \sum_{k=0}^{q} b_k x(n-k) \tag{2.24}$$

where $b_k, k = 1, 2, \ldots, q$, are the additional linear parameters. The parameters p and q are the model orders. The Akaike criterion can be used to determine the order of the appropriate model of a measurement signal by minimizing the following equation [12] with respect to the model order:

$$\text{AIC}(i, j) = N \ln(\sigma_{ij}^2) + 2(i + j) \tag{2.25}$$

where i and j represent the assumed AR and MA (moving average) model prediction orders respectively, N is the number of signal samples, and σ_{ij}^2 is the noise power of the ARMA model at the ith and jth stage. Later in this chapter it will be shown how the model parameters are estimated either directly or by employing some iterative optimization techniques.

In a multivariate AR (MVAR) approach a multichannel scheme is considered. Therefore, each signal sample is defined versus both its previous samples and the previous samples of the other channels, i.e. for channel i,

$$y_i(n) = -\sum_{k=1}^{p} a_{ik} y_i(n-k) - \sum_{\substack{j=1 \\ j\neq i}}^{m} \sum_{k=1}^{p} a_{jk} y_j(n-k) + x_i(n) \tag{2.26}$$

where m represents the number of channels and $x_i(n)$ represents the noise input to channel i. Similarly, the model parameters can be calculated iteratively in order to minimize the error between the actual and predicted values [13].

These linear models will be described further later in this chapter and some of their applications are discussed in other chapters. Different algorithms have been developed to find the model coefficients efficiently. In the maximum likelihood estimation (MLE) method [14–16] the likelihood function is maximized over the system parameters formulated from the assumed real, Gaussian distributed, and sufficiently long input signals of approximately 10–20 seconds (consider a sampling frequency of $f_s = 250$ samples/s as often used for EEG recordings). Using Akaike's method, the gradient of the squared error is minimized using the Newton–Raphson approach applied to the resultant nonlinear equations [16,17]. This is considered as an approximation to the MLE approach. In the Durbin method [18] the Yule–Walker equations, which relate the model coefficients to the autocorrelation of the signals, are iteratively solved. The approach and the results are equivalent to those using a least-squares-based scheme [19]. The MVAR coefficients are often calculated using the Levinson–Wiggins–Robinson (LWR) algorithm [20]. The MVAR model and its application in representation of what is called a direct transfer function (DTF), and its use in the quantification of signal propagation within the brain, will come in the following section. After the parameters are estimated the synthesis filter can be excited with wide-sense stationary noise to generate the EEG signal samples. Figure 2.6 illustrates the simplified system.

2.1.1.2 Prony's Method

Prony's method has been previously used to model evoked potentials (EPs) [21,22]. Based on this model an EP, which is obtained by applying a short audio or visual stimulation to the brain, can be considered as the impulse response (IR) of a linear infinite impulse

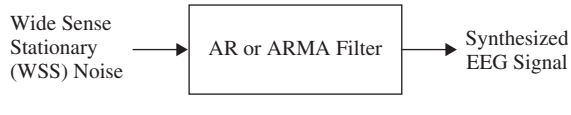

Figure 2.6 A linear model for the generation of EEG signals

response (IIR) system. The original attempt in this area was to fit an exponentially damped sinusoidal model to the data [23]. This method was later modified to model sinusoidal signals [24]. Prony's method is used to calculate the linear prediction (LP) parameters. The angles of the poles in the z plane of the constructed LP filter are then referred to the frequencies of the damped sinusoids of the exponential terms used for modelling the data. Consequently, both the amplitude of the exponentials and the initial phase can be obtained following the methods used for an AR model, as follows.

Based on the original method the output of an AR system with zero excitation can be considered to be related to its IR as

$$y(n) = \sum_{k=1}^{p} a_k y(n-k) = \sum_{j=1}^{p} w_j \sum_{k=1}^{p} a_k r_j^{n-k-1} \tag{2.27}$$

where $y(n)$ represents the exponential data samples, p is the prediction order, $w_j = A_j e^{j\theta_j}$, $r_k = \exp[(\alpha_k + j2\pi f_k)T_s]$, T_s is the sampling period normalized to 1, A_k is the amplitude of the exponential, α_k is the damping factor, f_k is the discrete-time sinusoidal frequency in samples/s, and θ_j is the initial phase in radians. Therefore, the model coefficients are first calculated using one of the methods previously mentioned in this section, i.e. $\boldsymbol{a} = -\mathbf{Y}^{-1}\check{\boldsymbol{y}}$, where

$$\boldsymbol{a} = \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_p \end{bmatrix}, \qquad \mathbf{Y} = \begin{bmatrix} y(p) \dots y(1) \\ y(p-1) \dots y(2) \\ \vdots \\ y(2p-1) \cdots y(p) \end{bmatrix}, \qquad \text{and} \qquad \check{\boldsymbol{y}} = \begin{bmatrix} y(p+1) \\ y(p+2) \\ \vdots \\ y(2p) \end{bmatrix} \tag{2.28}$$

where $a_0 = 1$. The prediction filter output, i.e. on the basis of Equation (2.27), $y(n)$ is calculated as the weighted sum of p past values of $y(n)$, and the parameters f_k and r_k are estimated. Hence, the damping factors are obtained as

$$\alpha_k = \ln |r_k| \tag{2.29}$$

and the resonance frequencies as

$$f_k = \frac{1}{2\pi} \tan^{-1} \left[\frac{\text{Im}(r_k)}{\text{Re}(r_k)} \right] \tag{2.30}$$

where Re(.) and Im(.) denote the real and imaginary parts of a complex quantity respectively. The w_k parameters are calculated using the fact that $y(n) = \sum_{k=1}^{p} w_k r_k^{n-1}$ or

$$\begin{bmatrix} r_1^0 & r_2^0 & \dots & r_p^0 \\ r_1^1 & r_2^1 & \dots & r_p^1 \\ \vdots & \vdots & & \vdots \\ r_1^{p-1} & r_2^{p-1} & \dots & r_p^{p-1} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_p \end{bmatrix} = \begin{bmatrix} y(1) \\ y(2) \\ \vdots \\ y(p) \end{bmatrix} \tag{2.31}$$

In vector form this can be illustrated as $\mathbf{R}\boldsymbol{w} = \boldsymbol{y}$, where $[\mathbf{R}]_{k,l} = r_l^k, k = 0, 1, \ldots, p-1, l = 1, \ldots, p$, denoting the elements of the matrix in the above equation. Therefore, $\boldsymbol{w} = \mathbf{R}^{-1}\boldsymbol{y}$, assuming $\mathbf{R}$ is a full-rank matrix, i.e. there are no repeated poles. Often, this is simply carried out by implementing the Cholesky decomposition algorithm [25]. Finally, using w_k, the amplitude and initial phases of the exponential terms are calculated as follows:

$$A_k = |w_k| \tag{2.32}$$

and

$$\theta_k = \tan^{-1}\left[\frac{\mathrm{Im}(w_k)}{\mathrm{Re}(w_k)}\right] \tag{2.33}$$

In the above solution it was considered that the number of data samples N is equal to $N = 2p$, where p is the prediction order. For cases where $N > 2p$, a least-squares (LS) solution for $\boldsymbol{w}$ can be obtained as

$$\boldsymbol{w} = (\mathbf{R}^{\mathrm{H}}\mathbf{R})^{-1}\mathbf{R}^{\mathrm{H}}\boldsymbol{y} \tag{2.34}$$

where $(.)^{\mathrm{H}}$ denotes the conjugate transpose. This equation can also be solved using the Cholesky decomposition method. For real data such as EEG signals this equation changes to $\boldsymbol{w} = (\mathbf{R}^{\mathrm{T}}\mathbf{R})^{-1}\mathbf{R}^{\mathrm{T}}\boldsymbol{y}$, where $(.)^{\mathrm{T}}$ represents the transpose operation. A similar result can be achieved using principal component analysis (PCA) [15].

In the cases where the data are contaminated with white noise the performance of Prony's method is reasonable. However, for nonwhite noise the noise information is not easily separable from the data and therefore the method may not be sufficiently successful.

In a later chapter of this book it will be seen that, Prony's algorithm has been used in modelling and analysis of audio and visual evoked potentials (AEP and VEP) [26,27].

2.1.2 Nonlinear Modelling

An approach similar to AR or MVAR modelling in which the output samples are nonlinearly related to the previous samples may be followed based on the methods developed for forecasting financial growth in economical studies. In the generalized autoregressive conditional heteroskedasticity (GARCH) method [28] each sample relates to its previous samples through a nonlinear (or sum of nonlinear) function(s). This model was originally introduced for time-varying volatility (honoured with the Nobel Prize in Economic Sciences in 2003). Nonlinearities in the time series are declared with the aid of the McLeod–Li [29] and BDS (Brock, Dechert, and Scheinkman) tests [30]. However, both tests lack the ability to reveal the actual kind of nonlinear dependency.

Generally, it is not possible to discern whether the nonlinearity is deterministic or stochastic in nature, and nor can a distinction be made between multiplicative and additive dependencies. The type of stochastic nonlinearity can be determined on the basis of the Hseih test [31]. Both additive and multiplicative dependencies can be discriminated by using this test. However, the test itself is not used to obtain the model parameters.

Considering the input to a nonlinear system to be $u(n)$ and the generated signal as the output of such a system to be $x(n)$, a restricted class of nonlinear models suitable for the analysis of such a process is given by

$$x(n) = g(u(n-1), u(n-2), \ldots) + u_n \; h(u(n-1), u(n-2), \ldots) \tag{2.35}$$

Multiplicative dependence means nonlinearity in the variance, which requires the function $h(.)$ to be nonlinear; additive dependence, on the other hand, means nonlinearity in the mean, which holds if the function $g(.)$ is nonlinear. The conditional statistical mean and variance are respectively defined as

$$E[x(n)|\chi_{n-1}] = g(u(n-1), u(n-2), \ldots) \tag{2.36}$$

and

$$\text{Var}[x(n)|\chi_{n-1}] = h^2(u(n-1), u(n-2), \ldots) \tag{2.37}$$

where χ_{n-1} contains all the past information up to time $n-1$. The original GARCH(p, q) model, where p and q are the prediction orders, considers a zero mean case, i.e. $g(.) = 0$. If $e(n)$ represents the residual (error) signal using the above nonlinear prediction system, then

$$\text{Var}[e(n)|\chi_{n-1}] = \sigma^2(n) = \alpha_0 + \sum_{j=1}^{q} \alpha_j \text{e}^2(n-j) + \sum_{j=1}^{p} \beta_j \sigma^2(n-1) \tag{2.38}$$

where α_j and β_j are the nonlinear model coefficients. The second term (first sum) in the right-hand side corresponds to a qth-order moving average (MA) dynamical noise term and the third term (second sum) corresponds to an autoregressive (AR) model of order p. It is seen that the current conditional variance of the residual at time sample n depends on both its previous sample values and previous variances.

Although in many practical applications such as forecasting of stock prices the orders p and q are set to small fixed values such as $(p, q) = (1, 1)$, for a more accurate modelling of natural signals such as EEGs the orders have to be determined mathematically. The prediction coefficients for various GARCH models or even the nonlinear functions g and h are estimated recursively as for the linear ARMA models [28,29].

Clearly, such simple GARCH models are only suitable for multiplicative nonlinear dependence. In addition, additive dependencies can be captured by extending the modelling approach to the class of GARCH-M (GARCH-in-mean) models [32].

Another limitation of the above simple GARCH model is failing to accommodate sign asymmetries. This is because the squared residual is used in the update equations. Moreover, the model cannot cope with rapid transitions such as spikes. Considering these shortcomings, numerous extensions to the GARCH model have been proposed. For example, the model has been extended and refined to include the asymmetric effects of positive and negative jumps such as the exponential GARCH (EGARCH) model [33], the Glosten, Jagannathan, and Runkle GARCH (GJR-GARCH) model [34], the threshold GARCH (TGARCH) model [35], the asymmetric power GARCH (APGARCH) model

[36], and the quadratic GARCH (QGARCH) model [37]. In the EGARCH model, for example, the above equation changes to

$$\log[\sigma^2(n)] = \log(\alpha_0) + \sum_{j=1}^{q} \alpha_j \log[e^2(n-j)] + \sum_{j=1}^{p} \beta_j \log[\sigma^2(n-j)] \tag{2.39}$$

where $\log[\cdot]$ denotes natural logarithm. This logarithmic expression has the advantage of preventing the variance from becoming negative.

In these models different functions for $g(.)$ and $h(.)$ are defined. For example, in the EGARCH model proposed by Glosten *et al.* [34] $h(n)$ is iteratively computed as

$$h_{(n)} = b + \alpha_1 u^2(n-1)(1-\eta_{(n-1)}) + \alpha_2 u^2(n-1)\eta_{(n-1)} + \kappa h_{(n-1)} \tag{2.40}$$

where b, α_1, α_2, and κ are constants and $\eta_{(n)}$ is an indicator function that is zero when $u_{(n)}$ is negative and one otherwise.

Despite modelling the signals, the GARCH approach has many other applications. In some recent works [38] the concept of GARCH modelling of covariance is combined with Kalman filtering to provide a more flexible model with respect to space and time for solving the inverse problem. There are several alternatives for solution to the inverse problem. Many approaches fall into the category of constrained least-squares methods employing Tikhonov regularization [39]. Localization of the sources within the brain using the EEG information is as an example. This approach has become known as low-resolution electromagnetic tomography (LORETA) [40]. Among numerous possible choices for the GARCH dynamics, the EGARCH [33] has been used to estimate the variance parameter of the Kalman filter sequentially.

The above methods are used to model the existing data, but to generate the EEG signals accurately a very complex model that exploits the physiological dynamics and various mental activities of the brain has to be constructed. Such a model should also incorporate the changes in the brain signals due to abnormalities and the onset of diseases. The next section considers the interaction among various brain components to establish a more realistic model for generation of the EEG signals.

2.1.3 Generating EEG Signals Based on Modelling the Neuronal Activities

The objective in this section is to introduce some established models for generating normal and some abnormal EEGs. These models are generally nonlinear, some have been proposed [41] for modelling a normal EEG signal and some others for the abnormal EEGs.

A simple distributed model consisting of a set of simulated neurons, thalamocortical relay cells, and interneurons was proposed [42,43] that incorporates the limited physiological and histological data available at that time. The basic assumptions were sufficient to explain the generation of the alpha rhythm, i.e. the EEGs within the frequency range of 8–13 Hz.

A general nonlinear lumped model may take the form shown in Figure 2.7. Although the model is analogue in nature, all the blocks are implemented in a discrete form. This model can take into account the major characteristics of a distributed model and it is easy

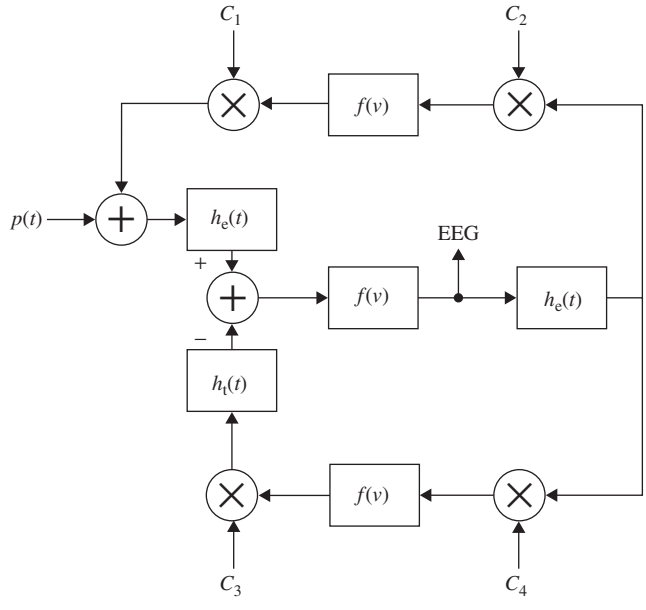

Figure 2.7 A nonlinear lumped model for generating the rhythmic activity of the EEG signals; $h_e(t)$ and $h_i(t)$ are the excitatory and inhibitory postsynaptic potentials, $f(v)$ is normally a simplified nonlinear function, and the C_is are respectively the interaction parameters representing the interneurons and thalamocortical neurons

to investigate the result of changing the range of excitatory and inhibitory influences of thalamocortical relay cells and interneurons.

In this model [42] there is a feedback loop including the inhibitory postsynaptic potentials, the nonlinear function, and the interaction parameters C_3 and C_4. The other feedback includes mainly the excitatory potentials, nonlinear function, and the interaction parameters C_1 and C_2. The role of the excitatory neurons is to excite one or two inhibitory neurons. The latter, in turn, serve to inhibit a collection of excitatory neurons. Thus, the neural circuit forms a feedback system. The input $p(t)$ is considered as a white noise signal. This is a general model; more assumptions are often needed to enable generation of the EEGs for the abnormal cases. Therefore, the function $f(v)$ may change to generate the EEG signals for different brain abnormalities. Accordingly, the C_i coefficients can be varied. In addition, the output is subject to environment and measurement noise. In some models, such as the local EEG model (LEM) [42] the noise has been considered as an additive component in the output.

Figure 2.8 shows the LEM model. This model uses the formulation by Wilson and Cowan [44] who provided a set of equations to describe the overall activity (not specifically the EGG) in a cartel of excitatory and inhibitory neurons having a large number of interconnections [45]. Similarly, in the LEM the EEG rhythms are assumed to be generated by distinct neuronal populations, which possess frequency selective properties. These populations are formed by the interconnection of the individual neurons and are assumed to be driven by a random input. The model characteristics, such as the neural

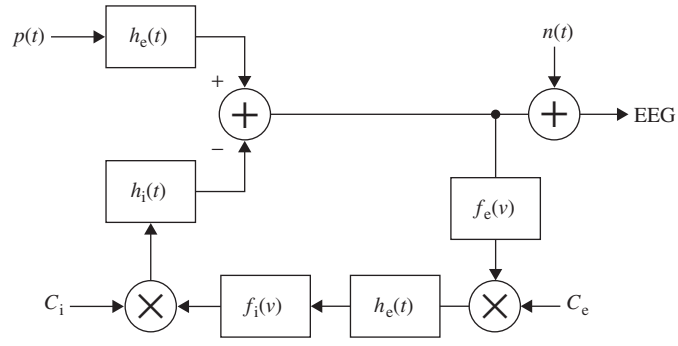

Figure 2.8 The local EEG model (LEM). The thalamocortical relay neurons are represented by two linear systems having impulse responses $h_e(t)$, on the upper branch, and the inhibitory postsynaptic potential represented by $h_i(t)$. The nonlinearity of this system is denoted by $f_e(v)$, representing the spike-generating process. The interneuron activity is represented by another linear filter $h_e(t)$ in the lower branch, which generally can be different from the first linear system, and a nonlinearity function $f_i(v)$. C_e and C_i represent respectively the number of interneuron cells and the thalamocortical neurons

interconnectivity, synapse pulse response, and threshold of excitation, are presented by the LEM parameters. The changes in these parameters produce the relevant EEG rhythms.

In Figure 2.8, as in Figure 2.7, the notation 'e' and 'i' refer to excitatory and inhibitory respectively. The input $p(t)$ is assumed to result from the summation of a randomly distributed series of random potentials which drive the excitatory cells of the circuit, producing the ongoing background EEG signal. Such signals originate from other deeper brain sources within the thalamus and brain stem and constitute part of the ongoing or spontaneous firing of the central nervous system (CNS). In the model, the average number of inputs to an inhibitory neuron from the excitatory neurons is designated by C_e and the corresponding average number from inhibitory neurons to each individual excitatory neuron is C_i. The difference between two decaying exponentials is used for modelling each postsynaptic potential h_e or h_i:

$$h_e(t) = A[\exp(-a_1 t) - \exp(-a_2 t)] \tag{2.41}$$

$$h_i(t) = B[\exp(-b_1 t) - \exp(-b_2 t)] \tag{2.42}$$

where A, B, a_k, and b_k are constant parameters, which control the shape of the pulse waveforms. The membrane potentials are related to the axonal pulse densities via the static threshold functions f_e and f_i. These functions are generally nonlinear, but to ease the manipulations they are considered linear for each short time interval. Using this model, the normal brain rhythms, such as the alpha wave, are considered as filtered noise.

The main problem with such a model is due to the fact that only a single-channel EEG is generated and there is no modelling of interchannel relationships. Therefore, a more accurate model has to be defined to enable simulation of a multichannel EEG generation system. This is still an open question and remains an area of research.

2.2 Nonlinearity of the Medium

The head as a mixing medium combines EEG signals which are locally generated within the brain at the sensor positions. As a system, the head may be more or less susceptible to such sources in different situations. Generally, an EEG signal can be considered as the output of a nonlinear system, which may be characterized deterministically.

The changes in brain metabolism as a result of biological and physiological phenomena in the human body can change the mixing process. Some of these changes are influenced by the activity of the brain itself. These effects make the system nonlinear. Analysis of such a system is very complicated and up to now nobody has fully modelled the system to aid in the analysis of brain signals.

On the other hand, some measures borrowed from chaos theory and analysis of the dynamics of time series such as dissimilarity, attractor dimension, and largest Lyapunov exponents (LLE) can characterize the nonlinear behaviour of EEG signals. These concepts are discussed in Section 2.7 and some of their applications are given in Chapter 5.

2.3 Nonstationarity

Nonstationarity of the signals can be quantified by measuring some statistics of the signals at different time lags. The signals can be deemed stationary if there is no considerable variation in these statistics.

Although generally the multichannel EEG distribution is considered as multivariate Gaussian, the mean and covariance properties generally change from segment to segment. Therefore EEGs are considered stationary only within short intervals, i.e. quasistationarity. This Gaussian assumption holds during a normal brain condition, but during mental and physical activities this assumption is not valid. Some examples of nonstationarity of the EEG signals can be observed during the change in alertness and wakefulness (where there are stronger alpha oscillations), during eye blinking, during the transitions between various ictal states, and in the event-related potential (ERP) and evoked potential (EP) signals.

The change in the distribution of the signal segments can be measured in terms of both the parameters of a Gaussian process and the deviation of the distribution from Gaussian. The non-Gaussianity of the signals can be checked by measuring or estimating some higher-order moments such as skewness, kurtosis, negentropy, and Kulback–Laibler (KL) distance.

Skewness is a measure of symmetry or, more precisely, the lack of symmetry of the distribution. A distribution, or data set, is symmetric if it looks the same to the left and right of the centre point. The skewness is defined for a real signal as

$$\text{Skewness} = \frac{E[(x(n) - \mu)^3]}{\sigma^3} \tag{2.43}$$

where μ and σ are the mean and standard deviation respectively, and E denotes statistical expectation. If the distribution is more to the right of the mean point the skewness is negative, and vice versa. For a symmetric distribution such as Gaussian, the skewness is zero.

Kurtosis is a measure of whether the data are peaked or flat relative to a normal distribution; i.e. data sets with high kurtosis tend to have a distinct peak near the mean, decline rather rapidly, and have heavy tails. Data sets with low kurtosis tend to have a flat top near the mean rather than a sharp peak. A uniform distribution would be the extreme case. The kurtosis for a real signal $x(n)$ is defined as

$$\text{kurt} = \frac{m_4[x(n)]}{m_2^2[x(n)]} \tag{2.44}$$

where $m_i[x(n)]$ is the ith central moment of the signal $x(n)$, i.e. $m_i[x(n)] = E[(x(n) - \mu)^i]$. The kurtosis for signals with normal distributions is three. Therefore, an excess or normalized kurtosis is often used and defined as

$$\text{Ex kurt} = \frac{m_4[x(n)]}{m_2^2[x(n)]} - 3 \tag{2.45}$$

which is zero for Gaussian distributed signals. Often the signals are considered ergodic; hence the statistical averages can be assumed identical to time averages and so can be estimated with time averages.

The negentropy of a signal $x(n)$ [46] is defined as

$$J_{\text{neg}}[x(n)] = H[x_{\text{Gauss}}(n)] - H[x(n)] \tag{2.46}$$

where, $x_{\text{Gauss}}(n)$ is a Gaussian random signal with the same covariance as $x(n)$ and $H(.)$ is the differential entropy [47], defined as

$$H[x(n)] = \int_{-\infty}^{\infty} p[x(n)] \log \frac{1}{p[x(n)]} \, \mathrm{d}x(n) \tag{2.47}$$

and $p[x(n)]$ is the signal distribution. Negentropy is always nonnegative.

The KL distance between two distributions p_1 and p_2 is defined as

$$\text{KL} = \int_{-\infty}^{\infty} p_1(x_1(n)) \log \frac{p_1(x_1(n))}{p_2(x_2(n))} \, \mathrm{d}z \tag{2.48}$$

It is clear that the KL distance is generally asymmetric, therefore by changing the position of p_1 and p_2 in this equation the KL distance changes. The minimum of the KL distance occurs when $p_1(x_1(n)) = p_2(x_2(n))$.

2.4 Signal Segmentation

Often it is necessary to label the EEG signals by segments of similar characteristics that are particularly meaningful to clinicians and for assessment by neurophysiologists. Within each segment, the signals are considered statistically stationary, usually with similar time and frequency statistics. As an example, an EEG recorded from an epileptic patient may be divided into three segments of preictal, ictal, and postictal segments. Each

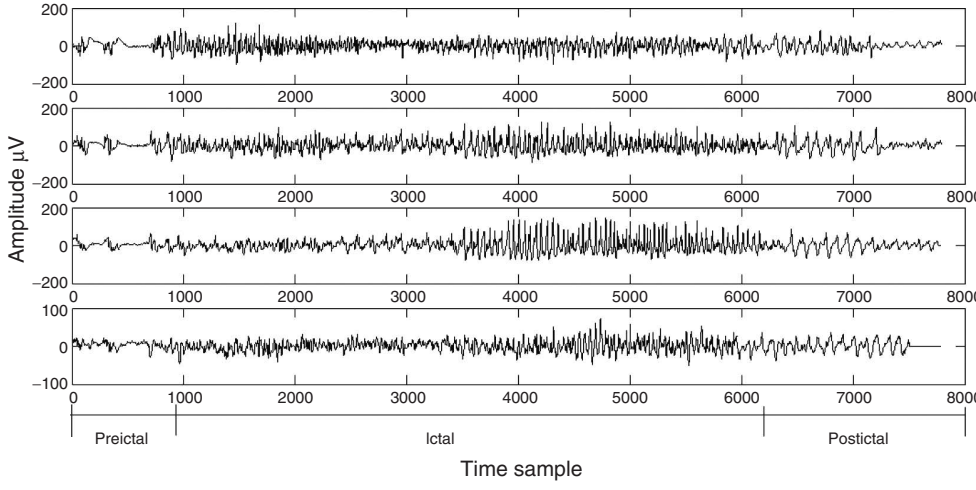

Figure 2.9 An EEG set of tonic–clonic seizure signals including three segments of preictal, ictal, and postictal behaviour

may have a different duration. Figure 2.9 represents an EEG sequence including all the above segments.

In segmentation of an EEG the time or frequency properties of the signals may be exploited. This eventually leads to a dissimilarity measurement denoted as $d(m)$ between the adjacent EEG frames, where m is an integer value indexing the frame and the difference is calculated between the m and $(m-1)$th (consecutive) signal frames. The boundary of the two different segments is then defined as the boundary between the m and $(m-1)$th frames provided $d(m) > \eta_T$, and η_T is an empirical threshold level. An efficient segmentation is possible by highlighting and effectively exploiting the diagnostic information within the signals with the help of expert clinicians. However, access to such experts is not always possible and therefore algorithmic methods are required.

A number of different dissimilarity measures may be defined based on the fundamentals of digital signal processing. One criterion is based on the autocorrelations for segment m, defined as

$$r_x(k, m) = E[x(n, m)x(n + k, m)] \tag{2.49}$$

The autocorrelation function of the mth length N frame for an assumed time interval $n, n+1, \ldots, n+(N-1)$ can be approximated as

$$\hat{r}_x(k, m) = \begin{cases} \dfrac{1}{N} \displaystyle\sum_{l=0}^{N-1-k} x(l+m+k)x(l+m), & k = 0, \ldots, N-1 \\ 0, & k = N, N+1, \ldots \end{cases} \tag{2.50}$$

Then the criterion is set to

$$d_1(m) = \frac{\sum_{k=-\infty}^{\infty} [\hat{r}_x(k, m) - \hat{r}_x(k, m-1)]^2}{\hat{r}_x(0, m)\hat{r}_x(0, m-1)} \tag{2.51}$$

A second criterion can be based on higher-order statistics. The signals with more uniform distributions such as normal brain rhythms have a low kurtosis, whereas seizure signals or event related potentials (ERP signals) often have high kurtosis values. Kurtosis is defined as the fourth-order cumulant at zero time lags and is related to the second- and fourth-order moments as given in Equations (2.43) to (2.45). A second level discriminant $d_2(m)$ is then defined as

$$d_2(m) = \text{kurt}_x(m) - \text{kurt}_x(m-1) \tag{2.52}$$

where m refers to the mth frame of the EEG signal $x(n)$. A third criterion is defined from the spectral error measure of the periodogram. A periodogram of the mth frame is obtained by discrete time Fourier transforming of the correlation function of the EEG signal

$$S_x(\omega, m) = \sum_{k=-\infty}^{\infty} \hat{r}_x(k, m)\mathrm{e}^{-\mathrm{j}\omega k}, \quad \omega \in [-\pi, \pi] \tag{2.53}$$

where $\hat{r}_x(., m)$ is the autocorrelation function for the mth frame as defined above. The criterion is then defined based on the normalized periodogram as

$$d_3(m) = \frac{\int_{-\pi}^{\pi} [S_x(\omega, m) - S_x(\omega, m-1)]^2 \, \mathrm{d}\omega}{\int_{-\pi}^{\pi} S_x(\omega, m)\mathrm{d}\omega \int_{-\pi}^{\pi} S_x(\omega, m-1) \, \mathrm{d}\omega} \tag{2.54}$$

The test window sample autocorrelation for the measurement of both $d_1(m)$ and $d_3(m)$ can be updated through the following recursive equation over the test windows of size N:

$$\hat{r}_x(k, m) = \hat{r}_x(k, m-1) + \frac{1}{N} [x(m-1+N)x(m-1+N-k) - x(m-1+k)x(m-1)] \tag{2.55}$$

and thereby computational complexity can be reduced in practice. A fourth criterion corresponds to the error energy in autoregressive (AR)-based modelling of the signals. The prediction error in the AR model of the mth frame is simply defined as

$$e(n, m) = x(n, m) - \sum_{k=1}^{p} a_k(m)x(n-k, m) \tag{2.56}$$

where p is the prediction order and $a_k(m)$, $k = 1, 2, \ldots, p$, are the prediction coefficients. For certain p the coefficients can be found directly (e.g. Durbin's method) in such a way as to minimize the error (residual) signal energy. In this approach it is assumed that the frames of length N are overlapped by one sample. The prediction coefficients estimated

for the $(m-1)$th frame are then used to predict the first sample in the mth frame, which is denoted as $\hat{e}(1,m)$. If this error is small, it is likely that the statistics of the mth frame are similar to those of the $(m-1)$th frame. On the other hand, a large value is likely to indicate a change. An indicator for the fourth criterion can then be the differencing of this prediction signal, which gives a peak at the segment boundary, i.e.

$$d_4(m) = \max[\nabla_m \hat{e}(1,m)] \tag{2.57}$$

where $\nabla_m(.)$ denotes the gradient with respect to m, approximated by a first-order difference operation. Figure 2.10 shows the residual and the gradient defined in Equation (2.57)

Finally, a fifth criterion $d_5(m)$ may be defined by using the AR-based spectrum of the signals in the same way as the short-term frequency transform (STFT) for $d_3(m)$. The above AR model is a univariate model, i.e. it models a single-channel EEG. A similar criterion may be defined when multichannel EEGs are considered [20]. In such cases a multivariate AR (MVAR) model is analysed. The MVAR can also be used for characterization and quantification of the signal propagation within the brain and is discussed in the next section.

Although the above criteria can be effectively used for segmentation of EEG signals, better systems may be defined for the detection of certain abnormalities. In order to do that, the features that best describe the behaviour of the signals have to be identified and used. Therefore the segmentation problem becomes a classification problem for which different classifiers can be used.

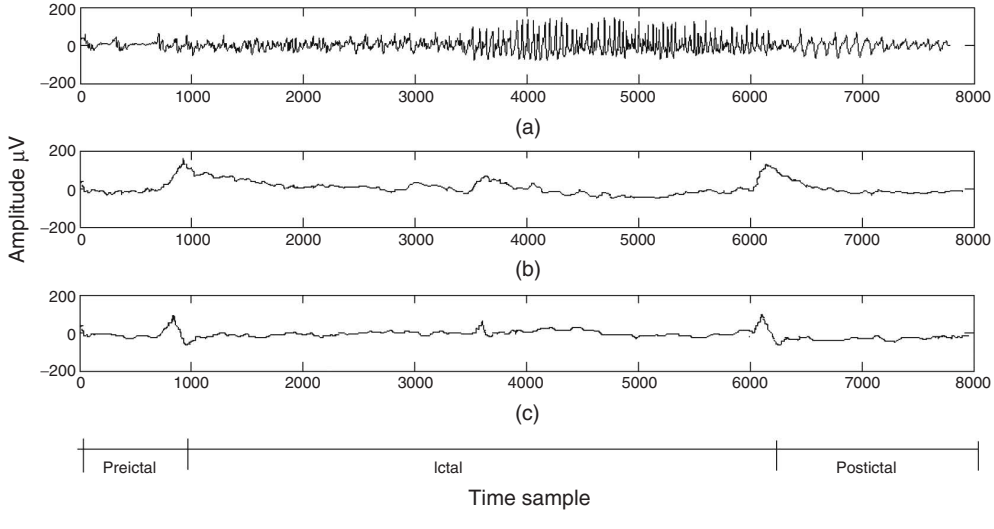

Figure 2.10 (a) An EEG seizure signal including preictal ictal and postictal segments, (b) the error signal, and (c) the approximate gradient of the signal, which exhibits a peak at the boundary between the segments. The number of prediction coefficients $p = 12$

2.5 Signal Transforms and Joint Time–Frequency Analysis

If the signals are statistically stationary it is straightforward to characterize them in either the time or frequency domains. The frequency-domain representation of a finite-length signal can be found by using linear transforms such as the (discrete) Fourier transform (DFT), (discrete) cosine transform (DCT), or other semi-optimal transform, which have kernels independent of the signal. However, the results of these transforms can be degraded by spectral smearing due to the short-term time-domain windowing of the signals and fixed transform kernels. An optimal transform such as the Karhunen–Loéve transform (KLT) requires complete statistical information, which may not be available in practice.

Parametric spectrum estimation methods such as those based on AR or ARMA modelling can outperform the DFT in accurately representing the frequency-domain characteristics of a signal, but they may suffer from poor estimation of the model parameters mainly due to the limited length of the measured signals. For example, in order to model the EEGs using an AR model, accurate values for the prediction order and coefficients are necessary. A high prediction order may result in splitting the true peaks in the frequency spectrum and a low prediction order results in combining peaks in close proximity in the frequency domain.

For an AR model of the signal $x(n)$ the error or driving signal is considered to be zero mean white noise. Therefore, by applying a z-transform to Equation (2.56), dropping the block index m, and replacing z by $e^{j\omega}$ gives

$$\frac{X_p(\omega)}{E(\omega)} = \frac{1}{1 - \sum_{k=1}^{p} a_k e^{-jk\omega}} \tag{2.58}$$

where, $E(\omega) = K_\omega$ (constant) is the power spectrum of the white noise and $X_p(\omega)$ is used to denote the signal power spectrum. Hence,

$$X_p(\omega) = \frac{K_\omega}{1 - \sum_{k=1}^{p} a_k e^{-jk\omega}} \tag{2.59}$$

and the parameters K_ω, a_k, $k = 1, \ldots, p$, are the exact values. In practical AR modelling these would be estimated from the finite length measurement, thereby degrading the estimate of the spectrum. Figure 2.11 provides a comparison of the spectrum of an EEG segment of approximately 1550 samples of a single-channel EEG using both DFT analysis and AR modelling.

The fluctuations in the DFT result as shown in Figure 2.11(b) are a consequence of the statistical inconsistency of periodogram-like power spectral estimation techniques. The result from the AR technique (Figure 2.11(c)) overcomes this problem provided the model fits the actual data. EEG signals are often statistically nonstationary, particularly where there is an abnormal event captured within the signals. In these cases the frequency-domain components are integrated over the observation interval and do not show the characteristics of the signals accurately. A time–frequency (TF) approach is the solution to the problem.

In the case of multichannel EEGs, where the geometrical positions of the electrodes reflect the spatial dimension, a space–time–frequency (STF) analysis through multiway processing methods has also become popular [48]. The main concepts in this area, together

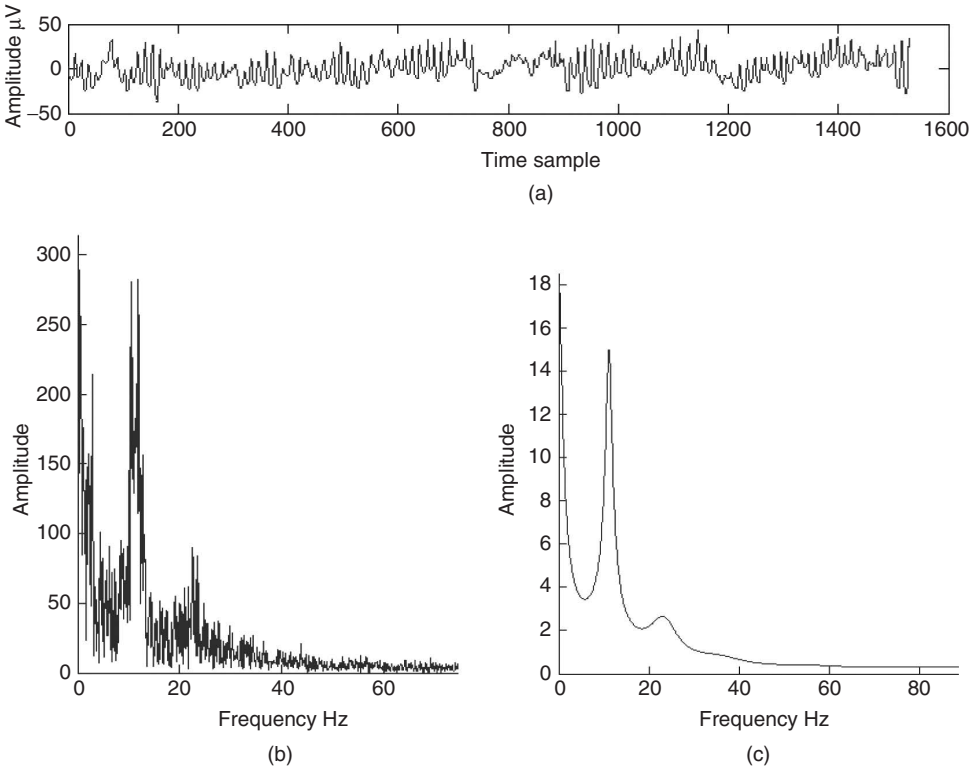

Figure 2.11 Single-channel EEG spectrum: (a) a segment of the EEG signal with a dominant alpha rhythm, (b) the spectrum of the signal in (a) using the DFT, and (c) the spectrum of the signal in (a) using a 12-order AR model

with the parallel factor analysis (PARAFAC) algorithm, will be reviewed in Chapter 7 where its major applications will be discussed.

The short-time Fourier transform (STFT) is defined as the discrete-time Fourier transform evaluated over a sliding window. The STFT can be performed as

$$X(n,\omega) = \sum_{\tau=-\infty}^{\infty} x(\tau)w(n-\tau)\mathrm{e}^{-\mathrm{j}\omega\tau} \tag{2.60}$$

where the discrete-time index n refers to the position of the window $w(n)$. Analogous with the periodogram, a spectrogram is defined as

$$S_x(n,\omega) = |X(n,\omega)|^2 \tag{2.61}$$

Based on the uncertainity principle, i.e. $\sigma_t^2\sigma_\omega^2 \geq \frac{1}{4}$, where σ_t^2 and σ_ω^2 are respectively the time- and frequency-domain variances, perfect resolution cannot be achieved in both time and frequency domains. Windows are typically chosen to eliminate discontinuities at block

edges and to retain positivity in the power spectrum estimate. The choice also impacts upon the spectral resolution of the resulting technique, which, put simply, corresponds to the minimum frequency separation required to resolve two equal amplitude frequency components [49].

Figure 2.12 shows the TF representation of an EEG segment during the evolution from preictal to ictal and to postictal stages. In this figure the effect of time resolution has been illustrated using a Hanning window of different durations of 1 and 2 seconds. Importantly, in this figure the drift in frequency during the ictal period is observed clearly.

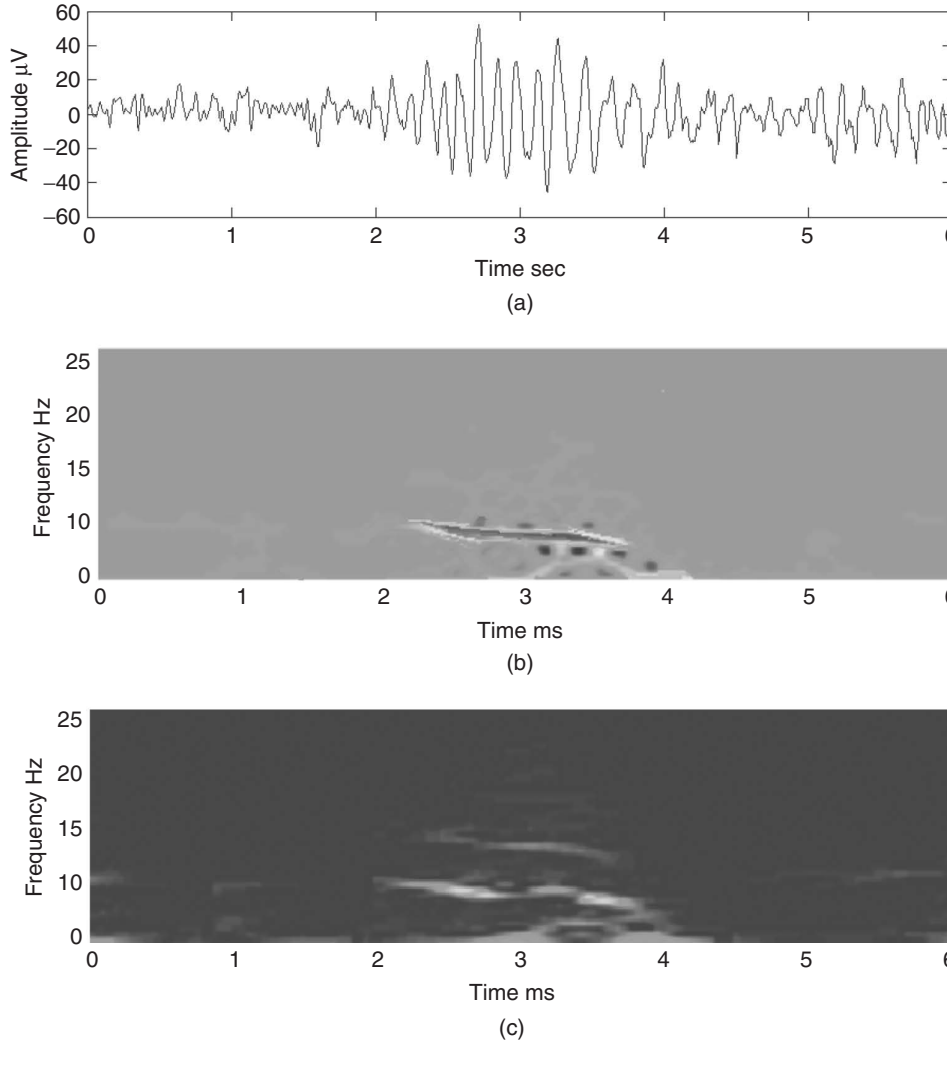

Figure 2.12 TF representation of an epileptic waveform (a) for different time resolutions using a Hanning window of (b) 1 ms and (c) 2 ms duration

2.5.1 Wavelet Transform

The wavelet transform (WT) is another alternative for a time–frequency analysis. There is already a well-established literature detailing the WT, such as References [50] and [51]. Unlike the STFT, the time–frequency kernel for the WT-based method can better localize the signal components in time–frequency space. This efficiently exploits the dependency between time and frequency components. Therefore, the main objective of introducing the WT by Morlet [50] was likely to have a coherence time proportional to the sampling period. To proceed, consider the context of a continuous time signal.

2.5.1.1 Continuous Wavelet Transform

The Morlet–Grossmann definition of the continuous wavelet transform for a one-dimensional signal $f(t)$ is

$$W(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t)\psi^* \left(\frac{t-b}{a} \right) \mathrm{d}t \tag{2.62}$$

where $(.)^*$ denotes the complex conjugate, $\psi(t)$ is the analysing wavelet, $a(>0)$ is the scale parameter (inversely proportional to frequency), and b is the position parameter. The transform is linear and is invariant under translations and dilations, i.e.

$$\text{If} \quad f(t) \rightarrow W(a,b) \text{ then } f(t-\tau) \rightarrow W(a,b-\tau) \tag{2.63}$$

and

$$f(\sigma t) \rightarrow \frac{1}{\sqrt{\sigma}} W(\sigma a, \sigma b) \tag{2.64}$$

The last property makes the wavelet transform very suitable for analysing hierarchical structures. It is similar to a mathematical microscope with properties that do not depend on the magnification. Consider a function $W(a,b)$ which is the wavelet transform of a given function $f(t)$. It has been shown [52,53] that $f(t)$ can be recovered according to

$$f(t) = \frac{1}{C_\varphi} \int_0^{\infty} \int_{-\infty}^{\infty} \frac{1}{\sqrt{a}} W(a,b) \varphi \left(\frac{t-b}{a} \right) \frac{\mathrm{d}a\mathrm{d}b}{a^2} \tag{2.65}$$

where

$$C_\varphi = \int_0^{\infty} \frac{\hat{\psi}^*(v)\hat{\varphi}(v)}{v} \mathrm{d}v = \int_{-\infty}^{0} \frac{\hat{\psi}^*(v)\hat{\varphi}(v)}{v} \mathrm{d}v \tag{2.66}$$

Although often it is considered that $\psi(t) = \varphi(t)$, other alternatives for $\varphi(t)$ may enhance certain features for some specific applications [54]. The reconstruction of $f(t)$ is subject to having C_φ defined (admissibility condition). The case $\psi(t) = \varphi(t)$ implies $\hat{\psi}(0) = 0$; i.e. the mean of the wavelet function is zero.

2.5.1.2 Examples of Continuous Wavelets

Different waveforms/wavelets/kernels have been defined for the continuous wavelet transforms. The most popular ones are given below.

Morlet's wavelet is a complex waveform defined as

$$\psi(t) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-t^2/2 + j2\pi b_0 t} \tag{2.67}$$

This wavelet may be decomposed into its constituent real and imaginary parts as

$$\psi_{\mathrm{r}}(t) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-t^2/2} \cos(2\pi b_0 t) \tag{2.68}$$

$$\psi_{\mathrm{i}}(t) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-t^2/2} \sin(2\pi b_0 t) \tag{2.69}$$

where b_0 is a constant, and it is considered that $b_0 > 0$ to satisfy the admissibility condition. Figure 2.13 shows respectively the real and imaginary parts.

The Mexican hat defined by Murenzi *et al.* [51] is

$$\psi(t) = (1 - t^2)\mathrm{e}^{-0.5t^2} \tag{2.70}$$

which is the second derivative of a Gaussian waveform (see Figure 2.14).

2.5.1.3 Discrete-Time Wavelet Transform

In order to process digital signals a discrete approximation of the wavelet coefficients is required. The discrete wavelet transform (DWT) can be derived in accordance with the sampling theorem if a frequency band-limited signal is processed.

The continuous form of the WT may be discretized with some simple considerations on the modification of the wavelet pattern by dilation. Since generally the wavelet function $\psi(t)$ is not band-limited, it is necessary to suppress the values of the frequency components above half the sampling frequency to avoid aliasing (overlapping in frequency) effects.

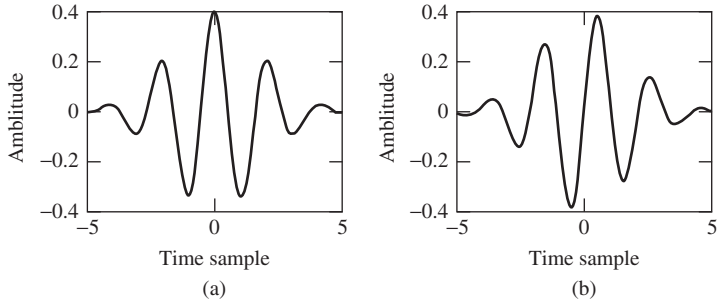

Figure 2.13 Morlet's wavelet: (a) real and (b) imaginary parts

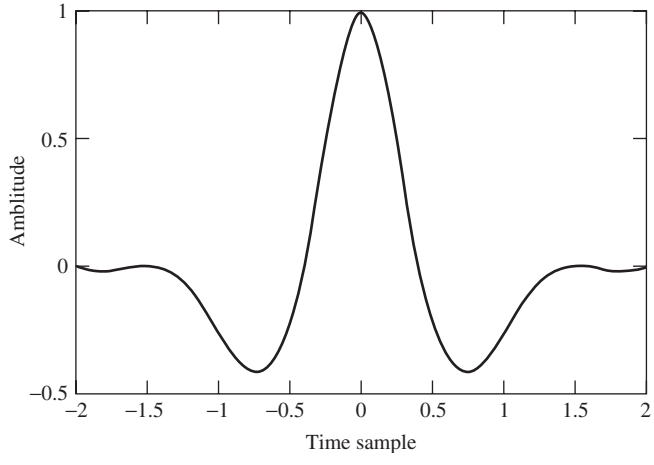

Figure 2.14 Mexican hat wavelet

A Fourier space may be used to compute the transform scale-by-scale. The number of elements for a scale can be reduced if the frequency bandwidth is also reduced. This requires a band-limited wavelet. The decomposition proposed by Littlewood and Paley [55] provides a very informative illustration of the reduction of elements scale-by-scale. This decomposition is based on an stagewise dichotomy of the frequency band. The associated wavelet is well localized in Fourier space, where it allows a reasonable analysis to be made, although not in the original space. The search for a discrete transform that is well localized in both spaces leads to a multiresolution analysis.

2.5.1.4 Multiresolution Analysis

Multiresolution analysis results from the embedded subsets generated by the interpolations (or down-sampling and filtering) of the signal at different scales. A function $f(t)$ is projected at each step j on to the subset V_j. This projection is defined by the scalar product $c_j(k)$ of $f(t)$ with the scaling function $\phi(t)$, which is dilated and translated as

$$C_j(k) = \langle f(t), 2^{-j}\phi(2^{-j}t - k)\rangle \tag{2.71}$$

where $\langle \cdot, \cdot \rangle$ denotes an inner product and $\phi(t)$ has the property

$$\frac{1}{2}\phi\left(\frac{t}{2}\right) = \sum_{n=-\infty}^{\infty} h(n)\phi(t-n) \tag{2.72}$$

where the right-hand side is convolution of h and ϕ. By taking the Fourier transform of both sides,

$$\Phi(2\omega) = H(\omega)\Phi(\omega) \tag{2.73}$$

where $H(\omega)$ and $\Phi(\omega)$ are the Fourier transforms of $h(t)$ and $\phi(t)$ respectively. For a discrete frequency space (i.e. using the DFT) the above equation permits the computation of the wavelet coefficient $C_{j+1}(k)$ from $C_j(k)$ directly. If a start is made from $C_0(k)$ and all $C_j(k)$, with $j > 0$, are computed without directly computing any other scalar product, then

$$C_{j+1}(k) = \sum_{n=0}^{N-1} C_j(n) h(n - 2k) \tag{2.74}$$

where k is the discrete frequency index and N is the signal length.

At each step, the number of scalar products is divided by two and consequently the signal is smoothed. Using this procedure the first part of a filter bank is built up. In order to restore the original data, Mallat uses the properties of orthogonal wavelets, but the theory has been generalized to a large class of filters by introducing two other filters $\tilde{h}$ and $\tilde{g}$, also called conjugate filters. The restoration is performed with

$$C_j(k) = 2 \sum_{l=0}^{\frac{N}{2}-1} [C_{j+1}(l)\tilde{h}(k + 2l) + w_{j+1}(l)\tilde{g}(k + 2l)] \tag{2.75}$$

where $w_{j+1}(.)$ are the wavelet coefficients at the scale $j + 1$ defined later in this section. For an exact restoration, two conditions have to be satisfied for the conjugate filters:

Anti-aliasing condition:

$$H\left(\omega + \tfrac{1}{2}\right)\tilde{H}(\omega) + G\left(\omega + \tfrac{1}{2}\right)\tilde{G}(\omega) = 0 \quad \forall\omega \tag{2.76}$$

Exact restoration:

$$H(\omega)\tilde{H}(\omega) + G(\omega)\tilde{G}(\omega) = 1 \quad \forall\omega \tag{2.77}$$

In the decomposition, the input is successively convolved with the time domain forms of the two filters H (low frequencies) and G (high frequencies). Each resulting function is decimated by suppression of one sample out of two. The high-frequency signal is left untouched, and the decomposition continues with the low-frequency signal (left-hand side of Figure 2.15). In the reconstruction, the sampling is restored by inserting a zero between each sample; then the conjugate filters $\tilde{H}$ and $\tilde{G}$ are applied, the resulting outputs are added and the result is multiplied by 2. Reconstruction continues to the smallest scale (right-hand side of Figure 2.15). Orthogonal wavelets correspond to the restricted case where

$$G(\omega) = \mathrm{e}^{-2\pi\omega} H^*\left(\omega + \tfrac{1}{2}\right) \tag{2.78}$$

$$\tilde{H}(\omega) = H^*(\omega) \tag{2.79}$$

$$\tilde{G}(\omega) = G^*(\omega) \tag{2.80}$$

and

$$|H(\omega)|^2 + \left|H\left(\omega + \tfrac{1}{2}\right)\right|^2 = 1 \quad \forall\omega \tag{2.81}$$

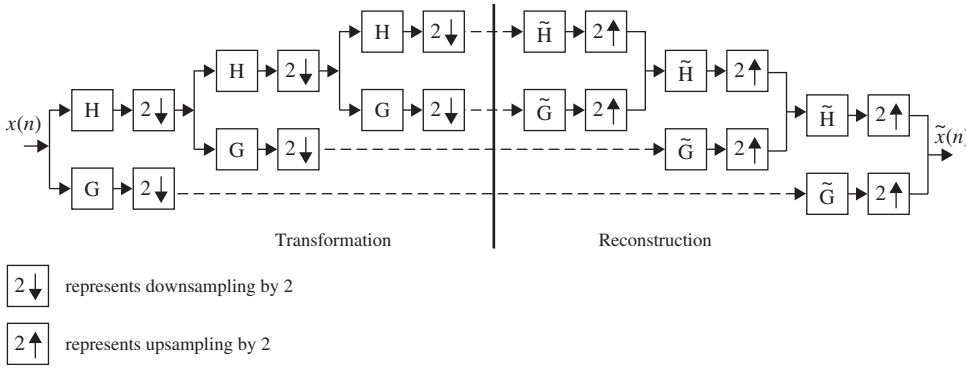

Figure 2.15 The filter bank associated with the multiresolution analysis

It can easily be seen that this set satisfies the two basic relations (2.72) and (2.73). Among various wavelets, Daubechie's wavelets are the only compact solutions to satisfy the above conditions. For biorthogonal wavelets, then

$$G(\omega) = \mathrm{e}^{-2\pi\omega}\tilde{H}^*\left(\omega + \tfrac{1}{2}\right) \tag{2.82}$$

$$\tilde{G}(\omega) = \mathrm{e}^{2\pi\omega}H^*\left(\omega + \tfrac{1}{2}\right) \tag{2.83}$$

and

$$H(\omega)\tilde{H}(\omega) + H^*\left(\omega + \tfrac{1}{2}\right)\tilde{H}^*\left(\omega + \tfrac{1}{2}\right) = 1 \quad \forall\omega \tag{2.84}$$

The relations (2.76) and (2.77) have also to be satisfied. A large class of compact wavelet functions can be used. Many sets of filters have been proposed, especially for coding [56]. It has been shown that the choice of these filters must be guided by the regularity of the scaling and the wavelet functions. The complexity is proportional to N. The algorithm provides a pyramid of N elements.

2.5.1.5 The Wavelet Transform using the Fourier Transform

Consider the scalar products $c_0(k) = \langle f(t)\ \phi(t-k)\rangle$ for continuous wavelets. If $\phi(t)$ is band-limited to half of the sampling frequency, the data can be correctly sampled. The data at the resolution $j = 1$ are

$$c_1(k) = \left\langle f(t)\ \tfrac{1}{2}\phi\left(\tfrac{t}{2} - k\right)\right\rangle \tag{2.85}$$

and the set $c_1(k)$ can be computed from $c_0(k)$ with a discrete-time filter with the frequency response $H(\omega)$:

$$H(\omega) = \begin{cases} \dfrac{\Phi(2\omega)}{\Phi(\omega)} & \text{if } |\omega| < \omega_c \\ 0 & \text{if } \omega_c \le |\omega| < \frac{1}{2} \end{cases} \tag{2.86}$$

and for $\forall\omega$ and $\forall$ integer m

$$H(\omega + m) = H(\omega) \tag{2.87}$$

Therefore, the coefficients at the next scale can be found from

$$C_{j+1}(\omega) = C_j(\omega)H(2^j\omega) \tag{2.88}$$

The cut-off frequency is reduced by a factor 2 at each step, allowing a reduction of the number of samples by this factor. The wavelet coefficients at the scale $j + 1$ are

$$w_{j+1} = \langle f(t), 2^{-(j+1)}\psi(2^{-(j+1)}t - k)\rangle \tag{2.89}$$

and can be computed directly from C_j by

$$W_{j+1}(\omega) = C_j(\omega)G(2^j\omega) \tag{2.90}$$

where G is the following discrete-time filter:

$$G(\omega) = \begin{cases} \dfrac{\Psi(2\omega)}{\Phi(\omega)} & \text{if } |\omega| < \omega_c \\ 0 & \text{if } \omega_c \leq |\omega| < \frac{1}{2} \end{cases} \tag{2.91}$$

and for $\forall\omega$ and $\forall$ integer m

$$G(\omega + m) = G(\omega) \tag{2.92}$$

The frequency band is also reduced by a factor of two at each step. These relationships are also valid for DWT, following Section 2.5.1.4.

2.5.1.6 Reconstruction

The reconstruction of the data from its wavelet coefficients can be performed step-by-step, starting from the lowest resolution. At each scale,

$$C_{j+1} = H(2^j\omega)C_j(\omega) \tag{2.93}$$

$$W_{j+1} = G(2^j\omega)C_j(\omega) \tag{2.94}$$

when a search is made for C_j knowing C_{j+1}, W_{j+1}, h, and g. Then $C_j(\omega)$ is restored by minimizing

$$P_h(2^j\omega)|C_{j+1}(\omega) - H(2^j\omega)C_j(\omega)|^2 + P_g(2^j\omega)|W_{j+1}(\omega) - G(2^j\omega)C_j(\omega)|^2 \tag{2.95}$$

using a least squares estimator. $P_h(\omega)$ and $P_g(\omega)$ are weight functions that permit a general solution to the restoration of $C_j(\omega)$. The relationship of $C_j(\omega)$ is in the form of

$$C_j(\omega) = C_{j+1}(\omega)\tilde{H}(2^j\omega) + W_{j+1}(\omega)\tilde{G}(2^j\omega) \tag{2.96}$$

where the conjugate filters have the expressions

$$\tilde{H}(\omega) = \frac{P_h(\omega)H^*(\omega)}{P_h(\omega)|H(\omega)|^2 + P_g(\omega)|G(\omega)|^2} \tag{2.97}$$

$$\tilde{H}(\omega) = \frac{P_g(\omega)G^*(\omega)}{P_h(\omega)|H(\omega)|^2 + P_g(\omega)|G(\omega)|^2} \tag{2.98}$$

It is straightforward to see that these filters satisfy the exact reconstruction condition given in Equation (2.77). In fact, Equations (2.97) and (2.98) give the general solutions to this equation. In this analysis, the Shannon sampling condition is always respected. No aliasing exists, so that the antialiasing condition (2.76) is not necessary. The denominator is simplified if

$$G(\omega) = \sqrt{1 - |H(\omega)|^2} \tag{2.99}$$

This corresponds to the case where the wavelet is the difference between the squares of two resolutions:

$$|\Psi(2\omega)|^2 = |\Phi(\omega)|^2 - |\Phi(2\omega)|^2 \tag{2.100}$$

The reconstruction algorithm then carries out the following steps:

1. Compute the fast Fourier transform (FFT) of the signal at the low resolution.
2. Set j to n_p and perform the following iteration steps.
3. Compute the FFT of the wavelet coefficients at the scale j.
4. Multiply the wavelet coefficients W_j by $\tilde{G}$.
5. Multiply the signal coefficients at the lower resolution C_j by $\tilde{H}$.
6. The inverse Fourier transform of $W_j\tilde{G} + C_j\tilde{H}$ gives the coefficients C_{j-1}.
7. Then $j = j - 1$ and return to step 3.

The use of a band-limited scaling function allows a reduction of sampling at each scale and limits the computation complexity.

The wavelet transform has been widely used in EEG signal analysis. Its application to seizure detection, especially for neonates, modelling of the neuron potentials, and the detection of evoked potentials (EP) and event-related potentials (ERP) will be discussed in the corresponding chapters of this book.

2.5.2 Ambiguity Function and the Wigner–Ville Distribution

The ambiguity function for a continuous time signal is defined as

$$A_x(\tau, \nu) = \int_{-\infty}^{\infty} x^*\left(t - \frac{\tau}{2}\right) x\left(t + \frac{\tau}{2}\right) e^{j\nu t}\, dt \tag{2.101}$$

This function has its maximum value at the origin as

$$A_x(0,0) = \int_{-\infty}^{\infty} |x(t)|^2 \, \mathrm{d}t \tag{2.102}$$

As an example, if a continuous time signal is considered to consist of two modulated signals with different carrier frequencies such as

$$\begin{aligned} x(t) &= x_1(t) + x_2(t) \\ &= s_1(t)\mathrm{e}^{\mathrm{j}\omega_1 t} + s_2(t)\mathrm{e}^{\mathrm{j}\omega_2 t} \end{aligned} \tag{2.103}$$

The ambiguity function $A_x(\tau, \nu)$ will be in the form of

$$A_x(\tau, \nu) = A_{x_1}(\tau, \nu) + A_{x_2}(\tau, \nu) + \text{cross terms} \tag{2.104}$$

This concept is very important in the separation of signals using the TF domain. This will be addressed in the context of blind source separation (BSS) later in this chapter. Figure 2.16 demonstrates this concept.

The Wigner–Ville frequency distribution of a signal $x(t)$ is then defined as the two-dimensional Fourier transform of the ambiguity function

$$\begin{aligned} X_{\mathrm{WV}}(t, \omega) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} A_x(\tau, \nu) \mathrm{e}^{-\mathrm{j}\upsilon t} \mathrm{e}^{-\mathrm{j}\omega t} \, \mathrm{d}\nu \, \mathrm{d}\tau \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^* \left(\beta - \frac{\tau}{2}\right) x \left(\beta + \frac{\tau}{2}\right) \mathrm{e}^{-\mathrm{j}\upsilon(t-\beta)} \\ &\qquad \mathrm{e}^{-\mathrm{j}\omega\tau} \, \mathrm{d}\beta \, \mathrm{d}\nu \, \mathrm{d}\tau \end{aligned} \tag{2.105}$$

which changes to the dual form of the ambiguity function as

$$X_{\mathrm{WV}}(t, \omega) = \int_{-\infty}^{\infty} x^* \left(t - \frac{\tau}{2}\right) x \left(t + \frac{\tau}{2}\right) \mathrm{e}^{-\mathrm{j}\omega\tau} \, \mathrm{d}\tau \tag{2.106}$$

A quadratic form for the TF representation with the Wigner–Ville distribution can also be obtained using the signal in the frequency domain as

$$X_{\mathrm{WV}}(t, \omega) = \int_{-\infty}^{\infty} X^* \left(\omega - \frac{\nu}{2}\right) X \left(\omega + \frac{\nu}{2}\right) \mathrm{e}^{-\mathrm{j}\nu t} \, \mathrm{d}\nu \tag{2.107}$$

The Wigner–Ville distribution is real and has very good resolution in both the time and frequency domains. Also it has time and frequency support properties; i.e. if $x(t) = 0$ for $|t| > t_0$, then $X_{\mathrm{WV}}(t, \omega) = 0$ for $|t| > t_0$, and if $X(\omega) = 0$ for $|\omega| > \omega_0$, then $X_{\mathrm{WV}}(t, \omega) = 0$ for $|\omega| > \omega_0$. It has also both time-marginal and frequency-marginal conditions of the form

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} X_{\mathrm{WV}}(t, \omega) \, \mathrm{d}\omega = |X(t)|^2 \tag{2.108}$$

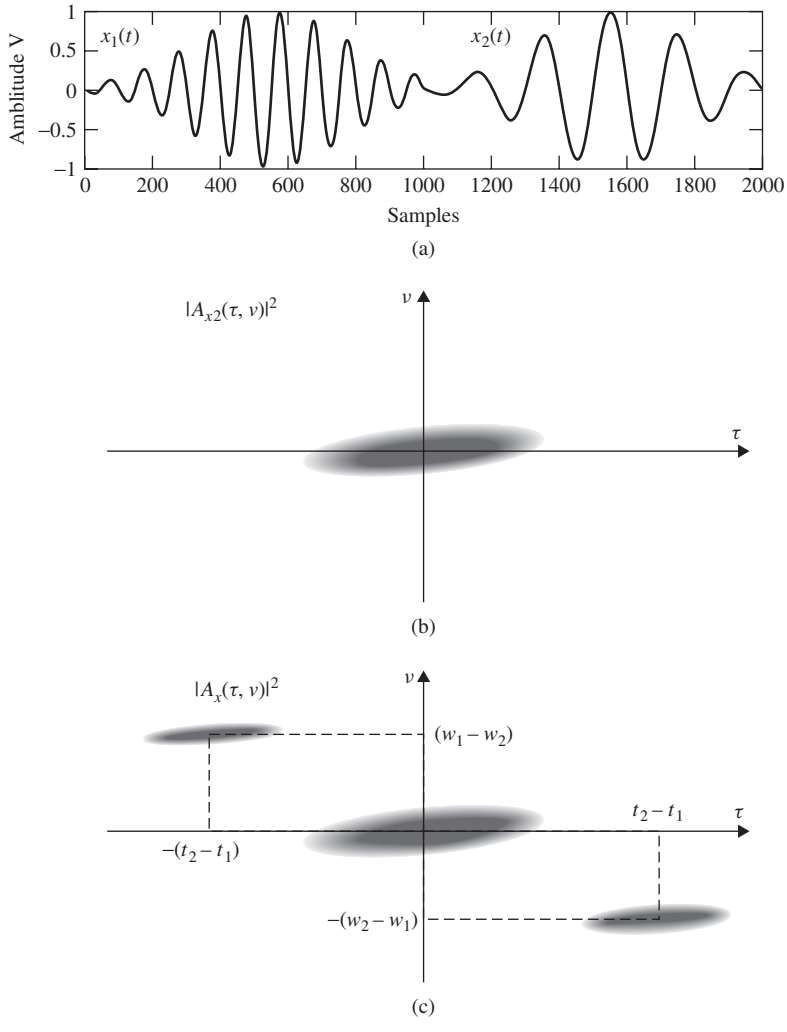

Figure 2.16 (a) A segment of a signal consisting of two modulated components, (b) an ambiguity function for $x_1(t)$ only, and (c) the ambiguity function for $x(t) = x_1(t) + x_2(t)$

and

$$\int_{-\infty}^{\infty} X_{\mathrm{WV}}(t, \omega)\, \mathrm{d}t = |X(\omega)|^2 \tag{2.109}$$

If $x(t)$ is the sum of two signals $x_1(t)$ and $x_2(t)$, i.e. $x(t) = x_1(t) + x_2(t)$, the Wigner–Ville distribution of $x(t)$ with respect to the distributions of $x_1(t)$ and $x_2(t)$ will be

$$X_{\mathrm{WV}}(t, \omega) = X_{1\mathrm{WV}}(t, \omega) + X_{2\mathrm{WV}}(t, \omega) + 2Re[X_{12\mathrm{WV}}(t, \omega)] \tag{2.110}$$

where Re[·] denotes the real part of a complex value and

$$X_{12\mathrm{WV}}(t,\omega)=\int_{-\infty}^{\infty}x_1^*\left(t-\frac{\tau}{2}\right)x_2\left(t+\frac{\tau}{2}\right)\mathrm{e}^{-\mathrm{j}\omega\tau}\,\mathrm{d}\tau \tag{2.111}$$

It is seen that the distribution is related to the spectra of both auto- and cross-correlations. A pseudo Wigner–Ville distribution (PWVD) is defined by applying a window function, $w(\tau)$, centred at $\tau=0$ to the time-based correlations, i.e.

$$\breve{X}_{\mathrm{WV}}(t,\omega)=\int_{-\infty}^{\infty}x^*\left(t-\frac{\tau}{2}\right)x\left(t+\frac{\tau}{2}\right)w(\tau)\mathrm{e}^{-\mathrm{j}\omega\tau}\,\mathrm{d}\tau \tag{2.112}$$

In order to suppress the undesired cross-terms the two-dimensional WV distribution may be convolved with a TF-domain window. The window is a two-dimensional lowpass filter, which satisfies the time and frequency marginal (uncertainty) conditions, as described earlier. This can be performed as

$$C_x(t,\omega)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}X_{\mathrm{WV}}(t',\omega')\Phi(t-t',\omega-\omega')\,\mathrm{d}t'\,\mathrm{d}\omega' \tag{2.113}$$

where

$$\Phi(t,\omega)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\phi(\tau,\nu)\mathrm{e}^{-\mathrm{j}\nu t}\mathrm{e}^{-\mathrm{j}\omega\tau}\,\mathrm{d}\nu\,\mathrm{d}\tau \tag{2.114}$$

and $\phi(.\,,.)$ is often selected from a set of well-known signals, the so-called *Cohen's class*. The most popular member of Cohen's class of functions is the bell-shaped function defined as

$$\phi(\tau,\nu)=\mathrm{e}^{-\nu^2\tau^2/(4\pi^2\sigma)},\quad \sigma>0 \tag{2.115}$$

A graphical illustration of such a function can be seen in Figure 2.17. In this case the distribution is referred to as a *Choi–Williams distribution.*

The application of a discrete time form of the Wigner–Ville distribution to BSS will be discussed later in this chapter and its application to seizure detection will be briefly explained in Chapter 4. To improve the distribution a signal-dependent kernel may also be used [57].

2.6 Coherency, Multivariate Autoregressive (MVAR) Modelling, and Directed Transfer Function (DTF)

In some applications such as in detection and classification of finger movement, it is very useful to establish how the associated movement signals propagate within the neural network of the brain. As will be shown in Chapter 7, there is a consistent movement of the source signals from the occipital to temporal regions. It is also clear that during the mental tasks different regions within the brain communicate with each other. The interaction and cross-talk among the EEG channels may be the only clue to understanding this process.

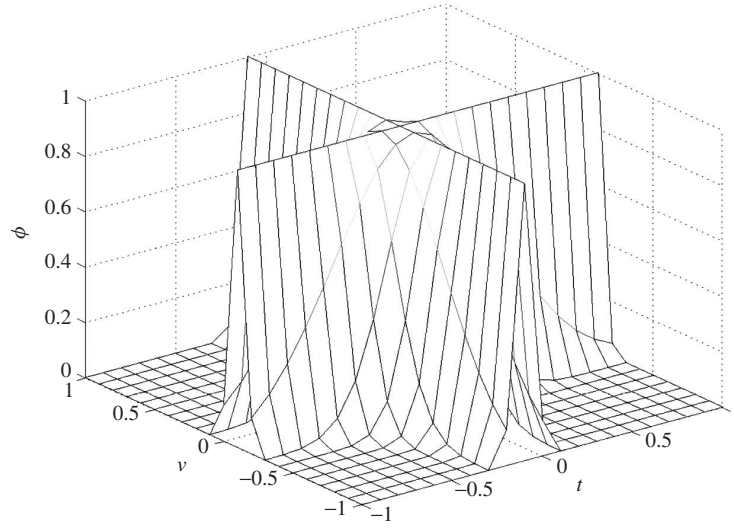

Figure 2.17 Illustration of $\phi(\tau, \nu)$ for the Choi–Williams distribution

This requires recognition of the transient periods of synchrony between various regions in the brain. These phenomena are not easy to observe by visual inspection of the EEGs. Therefore, some signal processing techniques have to be used in order to infer such causal relationships. One time series is said to be causal to another if the information contained in that time series enables the prediction of the other time series.

The spatial statistics of scalp EEGs are usually presented as coherence in individual frequency bands. These coherences result both from correlations among neocortical sources and volume conduction through the tissues of the head, i.e. brain, cerebrospinal fluid, skull, and scalp. Therefore, spectral coherence [58] is a common method for determining the synchrony in EEG activity. Coherency is given as

$$\mathrm{Coh}_{ij}^2(\omega) = \frac{E[C_{ij}(\omega)|^2]}{E[C_{ii}(\omega)]E[C_{jj}(\omega)]} \tag{2.116}$$

where $C_{ij}(\omega) = X_i(\omega)X_j^*(\omega)$ is the Fourier transform of the cross-correlation coefficients between channel i and channel j of the EEGs. Figure 2.18 shows an example of the cross-spectral coherence around one second prior to finger movement. A measure of this coherency, such as an average over a frequency band, is capable of detecting zero time lag synchronization and fixed time nonzero time lag synchronization, which may occur when there is a significant delay between the two neuronal population sites [59]. However, it does not provide any information on the directionality of the coupling between the two recording sites.

Granger causality (also called Wiener–Granger causality) [60] is another measure that attempts to extract and quantify the directionality from EEGs. Granger causality is based on bivariate AR estimates of the data. In a multichannel environment this causality is

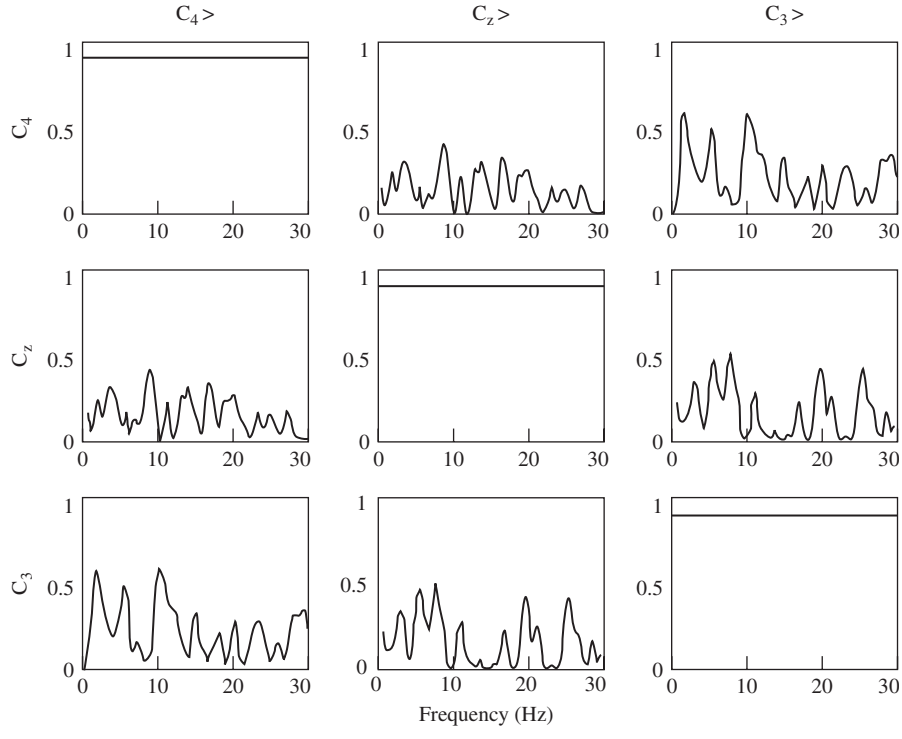

Figure 2.18 Cross-spectral coherence for a set of three electrode EEGs, one second before the right-finger movement. Each block refers to one electrode. By careful inspection of the figure, it is observed that the same waveform is transferred from C_z to C_3

calculated from pairwise combinations of electrodes. This method has been used to evaluate the directionality of the source movement from the local field potential in the visual system of cats [61].

For multivariate data in a multichannel recording, however, application of the Granger causality is not computationally efficient [61,62]. The directed transfer function (DTF) [63], as an extension of Granger causality, is obtained from multichannel data and can be used to detect and quantify the coupling directions. The advantage of the DTF over spectral coherence is that it can determine the directionality in the coupling when the frequency spectra of the two brain regions have overlapping spectra. The DTF has been adopted by some researchers for determining the directionality in the coupling [64,65] since it has been demonstrated that [66] there is a directed flow of information or cross-talk between the sensors around the sensory motor area before finger movement. The DTF is based on fitting the EEGs to an MVAR model. Assuming that $\boldsymbol{x}(n)$ is an M-channel EEG signal, it can be modelled in vector form as

$$\boldsymbol{x}(n) = -\sum_{k=1}^{p} \mathbf{L}_k \boldsymbol{x}(n-k) + \boldsymbol{v}(n) \tag{2.117}$$

where n is the discrete time index, p is the prediction order, $\boldsymbol{v}(n)$ is zero-mean noise, and $\mathbf{L}_k$ is generally an $M \times p$ matrix of prediction coefficients. A similar method to the Durbin algorithm for single channel signals, namely the Levinson–Wiggins–Robinson (LWR) algorithm is used to calculate the MVAR coefficients [20]. The Akaike AIC criterion [12] is also used for the estimation of prediction order p. By multiplying both sides of the above equation by $\boldsymbol{x}^{\mathrm{T}}(n-k)$ and performing the statistical expectation, the following Yule–Walker equation is obtained [67].

$$\sum_{k=0}^{p} \mathbf{L}_k \mathbf{R}(-k+p) = 0; \qquad \mathbf{L}_0 = \mathbf{I} \tag{2.118}$$

where $\mathbf{R}(q) = E[\boldsymbol{x}(n)\boldsymbol{x}^{\mathrm{T}}(n+q)]$ is the covariance matrix of $\boldsymbol{x}(n)$, and the cross-correlations of the signal and noise are zero since they are assumed to be uncorrelated. Similarly, the noise autocorrelation is zero for a nonzero shift since the noise samples are uncorrelated. The data segment is considered short enough for the signal to remain statistically stationary within that interval and long enough to enable accurate estimation of the prediction coefficients. Given the MVAR model coefficients, a multivariate spectrum can be achieved. Here it is assumed that the residual signal, $\boldsymbol{v}(n)$, is white noise. Therefore,

$$\mathbf{L}_f(\omega)\mathbf{X}(\omega) = \mathbf{V}(\omega) \tag{2.119}$$

where

$$\mathbf{L}_f(\omega) = \sum_{m=0}^{p} \mathbf{L}_m \mathrm{e}^{-\mathrm{j}\omega m} \tag{2.120}$$

and $\mathbf{L}(0) = \mathbf{I}$. Rearranging the above equation and replacing noise by $\sigma_v^2\mathbf{I}$ yields

$$\mathbf{X}(\omega) = \mathbf{L}_f^{-1}(\omega) \times \sigma_{\mathbf{V}}^2\mathbf{I} = \mathbf{H}(\omega) \tag{2.121}$$

which represents the model spectrum of the signals or the transfer matrix of the MVAR system. The DTF or causal relationship between channel i and channel j can be defined directly from the transform coefficients [62] given by

$$\Theta_{ij}^2(\omega) = |H_{ij}(\omega)|^2 \tag{2.122}$$

Electrode i is causal to j at frequency f if

$$\Theta_{ij}^2(\omega) > 0 \tag{2.123}$$

A time-varying DTF can also be generated (mainly to track the source signals) by calculating the DTF over short windows to achieve the short time DTF (SDTF) [62].

As an important feature in classification of left- and right-finger movements, or tracking the mental task related sources, the SDTF plays an important role. Some results of using the SDTF for detection and classification of finger movement are given in Chapter 7 in the context of brain–computer interfacing (BCI).

2.7 Chaos and Dynamical Analysis

As an effective tool for prediction and characterization of signals, *deterministic chaos* plays an important role. Although the EEG signals are considered chaotic, there are rules that do not in themselves involve any element of change and can be used in their characterization [68]. Mathematical research about chaos started before 1890 when people such as Andrey Kolmogorov or Henri Poincaré tried to establish whether planets would indefinitely remain in their orbits. In the 1960s Stephan Smale formulated a plan to classify all the typical kinds of dynamic behaviour. Many chaos-generating mechanisms have been created and used to identify the behaviour of the dynamics of the system. The Rossler system was designed to model a strange attractor using a simple stretch-and-fold mechanism. This was, however, inspired by the Lorenz attractor introduced more than a decade earlier [68].

To evaluate the city of chaotic a dynamical system different measures can be taken into account. A straightforward parameter is the attractor dimension. Different multidimensional attractors have been defined by a number of mathematicians. In many cases it is difficult to find the attractor dimension unless the parameters of the system can be approximated. However, later in this section it will be shown that the attraction dimension [69] can be simply achieved using the Lyapunov exponents.

2.7.1 Entropy

Entropy is a measure of uncertainty. The level of chaos may also be measured using entropy of the system. Higher entropy represents higher uncertainty and a more chaotic system. Entropy is given as

$$\text{Entropy of the signal } x(n) = \int_{\min(x)}^{\max(x)} p_x \log(1/p_x)\,\mathrm{d}x \tag{2.124}$$

where p_x is the probability density function (PDF) of signal $x(n)$. Generally, the distribution can be a joint PDF when the EEG channels are jointly processed. On the other hand, the PDF can be replaced by conditional PDF in places where the occurrence of the event is subject to another event. In this case, the entropy is called conditional entropy. Entropy is very sensitive to noise. Noise increases the uncertainty and noisy signals have higher entropy even if the original signal is ordered.

2.7.2 Kolmogorov Entropy

Also known as metric entropy, Kolmogorov entropy is an effective measure of chaos. To find the Kolmogorov entropy the phase space is divided into multidimensional hypercubes. Phase space is the space in which all possible states of a system are represented, each corresponding to one unique point in the phase space. In phase space, every degree of freedom or parameter of the system is represented as an axis of a multidimensional space. A phase space may contain many dimensions. The hypercube is a generalization of a 3-cube to n-dimensions, also called an n-cube or measure polytope. It is a regular polytope with mutually perpendicular sides, and is therefore an orthotope. Now, let $P_{i_0,\ldots,i_n}$ be the probability that a trajectory falls inside the hypercube, with i_0 at $t = 0$, i_1 at $t = T$,

i_2 at $t = 2T, \ldots$. Then define

$$K_n = -\sum_{i_0,\ldots,i_n} P_{i_0,\ldots,i_n} \ln P_{i_0,\ldots,i_n} \tag{2.125}$$

where $K_{n+1} - K_n$ is the information needed to predict which hypercube the trajectory will be in at $(n+1)T$, given trajectories up to nT. The Kolmogorov entropy is then defined as

$$K = \lim_{N\to\infty} \frac{1}{NT} \sum_{n=0}^{N-1} (K_{n+1} - K_n) \tag{2.126}$$

However, estimation of the above joint probabilities for large-dimensional data is computationally costly. On the other hand, in practice, long data sequences are normally required to perform a precise estimation of the Kolmogorov entropy.

2.7.3 Lyapunov Exponents

A chaotic model can be generated by a simple feedback system. Consider a quadratic iterator of the form $x(n) \to \alpha x(n)[1 - x(n)]$ with an initial value of x_0. This generates a time series such as that in Figure 2.19 (for $\alpha = 3.8$).

Although in the first 20 samples the time series seems to be random noise its semi-ordered alterations (cyclic behaviour) later show that some rules govern its chaotic behaviour. This time series is subject to two major parameters, α and x_0.

In order to adopt this model within a chaotic system a different initial value may be selected. Perturbation of an initial value generates an error E_0, which propagates during

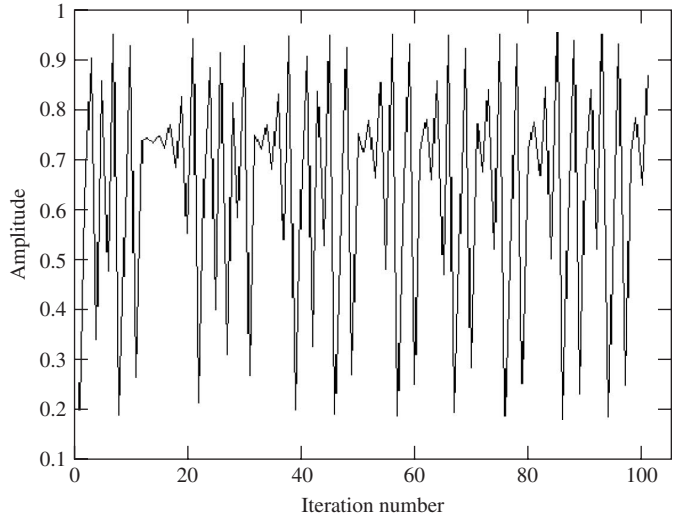

Figure 2.19 Generated chaotic signal using the model $x(n) \to \alpha x(n)[1 - x(n)]$ using $\alpha = 3.8$ and $x_0 = 0.2$

the signal evolution. After n samples the error changes to $E_n.E_n/E_0$ is a measure of how fast the error grows. The average growth of infinitesimally small errors in the initial point x_0 is quantified by Ljapunov (Lyapunov) exponents $\lambda(x_0)$. The total error amplification factor $|E_n/E_0|$, can be written in terms of sample error amplifications as

$$\left|\frac{E_n}{E_0}\right| = \left|\frac{E_n}{E_{n-1}}\right| \left|\frac{E_{n-1}}{E_{n-2}}\right| \cdots \left|\frac{E_1}{E_0}\right| \tag{2.127}$$

The average logarithm of this becomes

$$\frac{1}{n} \ln \left|\frac{E_n}{E_0}\right| = \frac{1}{n} \sum_{k=1}^{n} \ln \left|\frac{E_k}{E_{k-1}}\right| \tag{2.128}$$

Obviously, the problem is how to measure $|E_k/E_{k-1}|$. For the iterator $f(x(n))$ ($f(x(n)) = \alpha x(n)[1 - x(n)]$ in the above example) having a small perturbation ε at the initial point, the term in the above equation may be approximated as

$$\frac{1}{n} \ln \left|\frac{E_n}{E_0}\right| = \frac{1}{n} \sum_{k=1}^{n} \ln \left|\frac{\tilde{E}_k}{\varepsilon}\right| \tag{2.129}$$

where $\tilde{E}_k = f(x_{k-1} + \varepsilon) - f(x_{k-1})$. By replacing this in the above equation the Lyapunov exponent is approximated as

$$\lambda(x_0) = \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \ln |f'(x_{k-1})| \tag{2.130}$$

where $f'(x)$ represents differentiation of $f(x)$ with respect to x. This measure is very significant in separating unstable, unpredictable, or chaotic behaviour from predictable, stable, or ordered ones. If λ is positive the system is chaotic whereas it is negative for ordered systems.

Kaplan and Yorke [70] empirically concluded that it is possible to predict the dimension of a strange attractor from knowledge of the Lyapunov exponents of the corresponding transformation. This is termed the Kaplan–Yorke conjecture and has been investigated by many other researchers [71]. This is a very important conclusion since in many dynamical systems the various dimensions of the attractors are hard to compute, while the Lyapunov exponents are relatively easy to compute. This conjecture also claims that generally the information dimension D_{I} and Lyapunov dimension D_{L} respectively are defined as [68]

$$D_{\mathrm{I}} = \lim_{s \to 0} \frac{I(s)}{\log_2 1/s} \tag{2.131}$$

where s is the size of a segment of the attractor and $I(s)$ is the entropy of s, and

$$D_{\mathrm{L}} = m + \frac{1}{|\lambda_{m+1}|} \sum_{k=1}^{m} \lambda_k \tag{2.132}$$

where m is the maximum integer with $\gamma(m) = \lambda_1 + \cdots + \lambda_m \geq 0$, given that $\lambda_1 > \lambda_2 > \cdots > \lambda_m$.

2.7.4 *Plotting the Attractor Dimensions from the Time Series*

Very often it is necessary to visualize a phase space attractor and decide about the stability, chaocity, or randomness of a signal (time series). The attractors can be multidimensional. For a three-dimensional attractor a time delay T (a multiple of τ) can be chosen and the following sequence of vectors constructed.

$$\begin{array}{ccc} [x(0) & x(T) & x(2T)] \\ [x(\tau) & x(\tau+T) & x(\tau+2T)] \\ [x(2\tau) & x(2\tau+T) & x(2\tau+2T)] \\ \vdots & \vdots & \vdots \\ [x(k\tau) & x(k\tau+T) & x(k\tau+2T)] \end{array}$$

By plotting these points in a three-dimensional coordinate space and linking the points together successively the attractor can be observed. Figure 2.20 shows the attractors for a sinusoidal and the above chaotic time series. Although the attractors can be defined for a

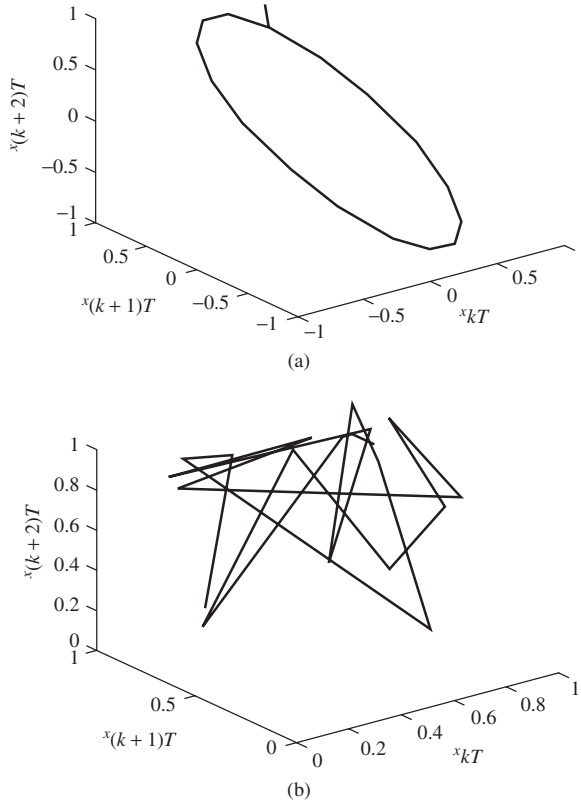

Figure 2.20 The attractors for (a) a sinusoid and (b) the above chaotic time sequence, both started from the same initial point

higher-dimensional space, visualization of the attractors is not possible when the number of dimensions increases above three.

2.7.5 *Estimation of Lyapunov Exponents from the Time Series*

Calculation of the Lyapunov exponents from the time series was first proposed by Wolf *et al*. [72]. In their method, initially a finite embedding sequence is constructed from the finite time series of $2N+1$ components as

$$x(0), x(\tau), x(2\tau), \ldots$$

This is the basic data (often called the reference trajectory or reference orbit) upon which the model builds. Generally, the start point is not given since there is no explicit governing equation that would generate the trajectory. From this sequence a point $x(k_0\tau)$ may be chosen that approximates the desired initial point $z_0(0)$. Considering Figure 2.21, these approximations should satisfy

$$|x(k_0\tau) - x(0)| < \delta \tag{2.133}$$

where δ is an *a priori* chosen tolerance. This point may be renamed as

$$z_0(0) = x(k_0\tau) \tag{2.134}$$

The successors of this point are known as

$$z_0(r\tau) = x((k_0 + r)\tau), \quad r = 1, 2, 3, \ldots \tag{2.135}$$

Now there are two trajectories to compare. The logarithmic error amplification factor for the first time interval becomes

$$l_0 = \frac{1}{\tau} \log \frac{|z_0(\tau) - z_0(0)|}{|x(\tau) - x(0)|} \tag{2.136}$$

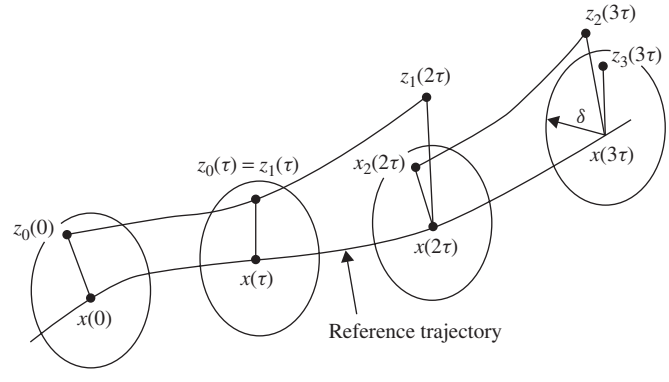

Figure 2.21 The reference and the model trajectories, evolution of the error, and start and end of the model trajectory segments. The model trajectory ends when its deviation from the reference trajectory is more than a threshold

This procedure is repeated for the next point $x(\tau)$ of the reference trajectory. For that point another point $z_1(\tau)$ needs to the formed from the trajectory, which represents an error with a direction close to the one obtained from $z_0(\tau)$ relative to $x(\tau)$. In the case where the previous trajectory is still close to the reference trajectory it may be possible simply to continue with that, thus setting $z_1(\tau) = z_0(\tau)$. This yields an error amplification factor l_1. Other factors, $l_2, \ldots, l_{m-1}$, can also be found by following the same procedure until the segment of the time series is exhausted. An approximation to the largest Lyapunov exponent for the current segment of the time series is obtained by averaging the logarithmic amplification factors over the whole reference trajectory:

$$\lambda = \frac{1}{m} \sum_{j=0}^{m-1} l_j \tag{2.137}$$

Instead of the above average, the maximum value of the error amplification factor may also be considered as the largest Lyapunov exponent. It is necessary to investigate the effect of noise as the data usually stem from a physical measurement and therefore contain noise. Hence, the perturbed points, $z_k(k\tau)$, should not be taken very close to each other, because then the noise would dominate the stretching effect on the chaotic attractor. On the other hand, the error should not be allowed to become too large in order to avoid nonlinear effects. Thus in practice some minimal error, δ_1, and a maximal error, δ_2, is prescribed and

$$\delta_1 < |x(k\tau) - z_k(k\tau)| < \delta_2 \tag{2.138}$$

is required.

2.7.5.1 Optimum Time Delay

In the above calculation it is important to find the *optimum time delay* τ. Very small time delays may result in near-linear reconstructions with high correlations between consecutive phase space points and very large delays might ignore any deterministic structure of the sequence. In an early proposal [73] the autocorrelation function is used to estimate the time delay. In this method τ is equivalent to the duration after which the autocorrelation reaches a minimum or drops to a small fraction of its initial value. In another attempt [74,75] it has been verified that the values of τ at which the mutual information has a local minimum are equivalent to the values of τ at which the logarithm of the correlation sum has a local minimum.

2.7.5.2 Optimum Embedding Dimension

To further optimize the measurement of Lyapunov exponents there is a need to specify the optimum value for m, named the *embedding dimension*. Before doing that some definitions have to be given as follows.

Fractal dimension is another statistic related to the dynamical measurement. The strange attractors are fractals and their fractal dimension D_f is simply related to the minimum number of dynamical variables needed to model the dynamics of the attractor. Conceptually,

a simple way to measure D_f is to measure the *Kolmogorov capacity*. In this measurement a set is covered with small cells, depending on the dimensionality (i.e. squares for sets embedded in two dimensions, cubes for sets embedded in three dimensions, and so on), of size ε. If $M(\varepsilon)$ denotes the number of such cells within a set, the fractal dimension is defined as

$$D_f = \lim_{\varepsilon \to 0} \frac{\log[M(\varepsilon)]}{\log(1/\varepsilon)} \tag{2.139}$$

for a set of single points $D_f = 0$, for a straight line $D_f = 1$, and for a plane area $D_f = 2$. The fractal dimension, however, may not be an integer.

The *correlation dimension* is defined as

$$D_r = \lim_{r \to 0} \frac{\log C(r)}{\log r} \tag{2.140}$$

where

$$C(r) = \sum_{i=1}^{M(r)} p_i^2 \tag{2.141}$$

is the correlation sum and p_i the probability of cell i.

The optimal embedding dimension, m, as required for accurate estimation of the Lyapunov exponents, has to satisfy $m \geq 2D_f + 1$. D_f is, however, not often known *a priori*. The Grassberger–Procaccia algorithm can nonetheless be employed to measure the correlation dimension, C_r. The minimum embedding dimension of the attractor is $m+1$, where m is the embedding dimension above which the measured value of the correlation dimension C_r remains constant.

As another very important conclusion,

$$D_f = D_L = 1 + \frac{\lambda_1}{|\lambda_2|} \tag{2.142}$$

i.e. the fractal dimension D_f is equivalent to the Lyapunov dimension [68].

Chaos has been used as a measure in analysis of many types of signals and data. Its application to epileptic seizure prediction will be shown in Chapter 4.

2.7.6 Approximate Entropy

Approximate entropy (AE) is a statistic that can be estimated from the discrete-time sequences, especially for real-time applications [76,77]. This measure can quantify the complexity or irregularity of the system. The AE is less sensitive to noise and can be used for short-length data. In addition, it is resistant to short strong transient interferences (outliers) such as spikes [77].

Given the embedding dimension m, the m-vector $\boldsymbol{x}(i)$ is defined as

$$\boldsymbol{x}(i) = [x(i), x(i+1), \ldots, x(i+m-1)], \quad i = 1, \ldots, N-m+1 \tag{2.143}$$

where N is the number of data points. The distance between any two of the above vectors, $\boldsymbol{x}(i)$ and $\boldsymbol{x}(j)$, is defined as

$$d[\boldsymbol{x}(i), \boldsymbol{x}(j)] = \max_k |x(i+k) - x(j+k)| \tag{2.144}$$

where $|.|$ denotes the absolute value. Considering a threshold level of β, the number of times, $M^m(i)$, that the above distance satisfies $d[\boldsymbol{x}(i), \boldsymbol{x}(j)] \leq \beta$ is found. This is performed for all i. For the embedding dimension m,

$$\xi_\beta^m(i) = \frac{M^m(i)}{N-m+1} \quad \text{for } i = 1, \ldots, N-m+1 \tag{2.145}$$

Then, the average natural logarithm of $\xi_\beta^m(i)$ is found as

$$\psi_\beta^m = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} \ln \xi_\beta^m(i) \tag{2.146}$$

By repeating the same method for an embedding dimension of $m+1$, the AE will be given as

$$\mathrm{AE}(m, \beta) = \lim_{N \to \infty} (\psi_\beta^m - \psi_\beta^{m+1}) \tag{2.147}$$

In practice, however, N is limited and therefore the AE is calculated for N data samples. In this case the AE depends on m, β, and N, i.e.

$$\mathrm{AE}(m, \beta, N) = \psi_\beta^m - \psi_\beta^{m+1} \tag{2.148}$$

The embedding dimension can be found as previously mentioned. However, the threshold value has to be set correctly. In some applications the threshold value is taken as a value between 0.1 and 0.25 times the data standard deviation [76].

2.7.7 Using the Prediction Order

It is apparent that for signals with highly correlated time samples the prediction order of an AR or ARMA model is low and for noise-type signals where the correlation among the samples is low the order is high. This means that for the latter case a large number of previous samples is required to predict the current sample. A different criterion such as the Akaike information criterion (AIC) may be employed to find the prediction order from the time series. Figure 2.22 shows the prediction order automatically computed for overlapping segments of three sections of a time series in which the middle section is sinusoidal and the first and third sections are noise like signals.

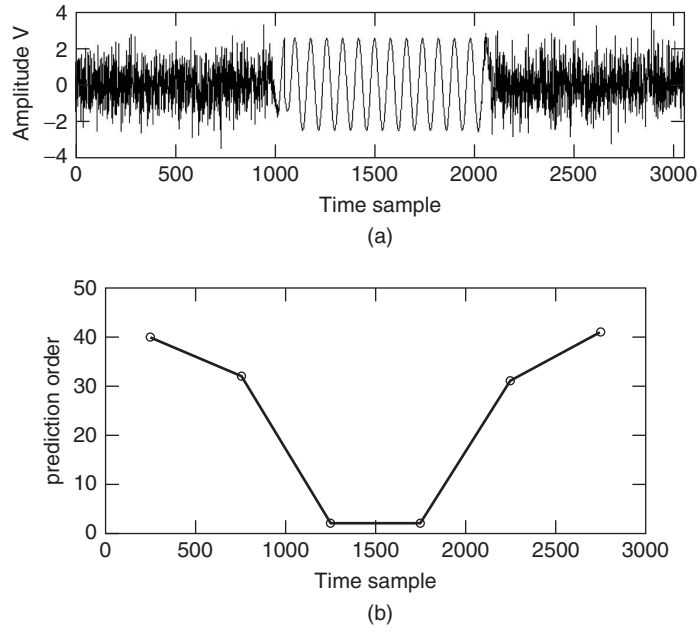

Figure 2.22 (a) The signal and (b) prediction order measured for overlapping segments of the signal

2.8 Filtering and Denoising

The EEG signals are subject to noise and artefacts. Electrocardiograms (ECGs), electroocclugrams (EOG), or eye blinks affect the EEG signals. Any multimodal recording such as EEG–fMRI significantly disturbs the EEG signals because of both magnetic fields and the change in the blood oxygen level and sensitivity of oxygen molecule to the magnetic field (ballistocardiogram). Artefact removal from EEGs will be explained in the related chapters. The noise in the EEGs, however, may be estimated and mitigated using adaptive and nonadaptive filtering techniques.

The EEG signals contain neuronal information below 100 Hz (in many applications the information lies below 30 Hz). Any frequency component above these frequencies can be simply removed by using lowpass filters. In the cases where the EEG data acquisition system is unable to cancel out the 50 Hz line frequency (due to a fault in grounding or imperfect balancing of the inputs to the differential amplifiers associated with the EEG system) a notch filter is used to remove it.

The nonlinearities in the recording system related to the frequency response of the amplifiers, if known, are compensated by using equalizing filters. However, the characteristics of the internal and external noises affecting the EEG signals are often unknown. The noise may be characterized if the signal and noise subspaces can be accurately separated. In Chapter 4 it is seen that the number of sources can be estimated. Using principal component analysis or independent component analysis it is possible to decompose the multichannel EEG observations to their constituent components, such as the

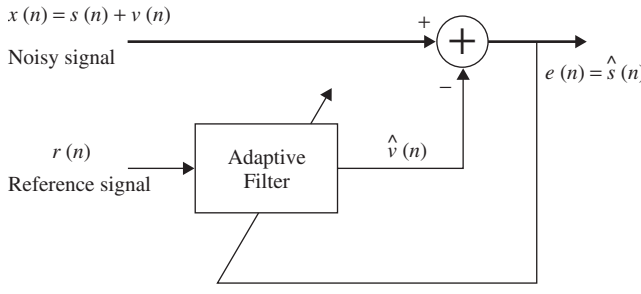

Figure 2.23 An adaptive noise canceller

neural activities and noise. Combining these two together, the estimated noise components can be extracted, characterized, and separated from the actual EEGs. These concepts are explained in the following sections and their applications to the artefact and noise removal will be given in the later chapters.

Adaptive noise cancellers used in communications, signal processing, and biomedical signal analysis can also be used to remove noise and artefacts from the EEG signals. An effective adaptive noise canceller, however, requires a reference signal. Figure 2.23 shows a general block diagram of an adaptive filter for noise cancellation. The reference signal carries significant information about the noise or artefact and its statistical properties. For example, in the removal of eye blinking artefacts (discussed in Chapter 7) a signature of the eye blink signal can be captured from the FP1 and FP2 EEG electrodes. In detection of the ERP signals, as another example, the reference signal can be obtained by averaging a number of ERP segments. There are many other examples such as ECG cancellation from EEGs and the removal of fMRI scanner artefacts from EEG–fMRI simultaneous recordings where the reference signals can be provided.

Adaptive Wiener filters are probably the most fundamental type of adaptive filters. In Figure 2.23 the optimal weights for the filter, $\boldsymbol{w}(n)$, are calculated such that $\hat{s}(n)$ is the best estimate of the actual signal $\boldsymbol{s}(n)$ in the mean-squared sense. The Wiener filter minimizes the mean-squared value of the error, defined as

$$e(n) = x(n) - \hat{v}(n) = x(n) - \boldsymbol{w}^{\mathrm{T}} r(n) \tag{2.149}$$

where $\boldsymbol{w}$ is the Wiener filter coefficient vector. Using the orthogonality principle [78] and assuming $x(n)$ and $r(n)$ are jointly statistically wide sense stationary, the final form of the mean-squared error will be

$$E[e(n)^2] = E[x(n)^2] - 2\boldsymbol{p}^{\mathrm{T}}\boldsymbol{w} + \boldsymbol{w}^{\mathrm{T}}\mathbf{R}\boldsymbol{w} \tag{2.150}$$

where $E(.)$ represents statistical expectation,

$$\boldsymbol{p} = E[x(n)\boldsymbol{r}(n)] \tag{2.151}$$

and

$$\mathbf{R} = E[\boldsymbol{r}(n)\boldsymbol{r}^{\mathrm{T}}(n)] \tag{2.152}$$

By taking the gradient with respect to $\boldsymbol{w}$ and equating it to zero,

$$\boldsymbol{w} = \mathbf{R}^{-1}\boldsymbol{p} \tag{2.153}$$

Since $\mathbf{R}$ and $\boldsymbol{p}$ are usually unknown, the above minimization is performed iteratively by substituting time averages for statistical averages. The adaptive filter in this case decorrelates the output signals. The general update equation is in the form of

$$\boldsymbol{w}(n+1) = \boldsymbol{w}(n) + \Delta\boldsymbol{w}(n) \tag{2.154}$$

where n is the iteration number, which typically corresponds to the discrete-time index. $\Delta\boldsymbol{w}(n)$ has to be computed such that $E[\boldsymbol{e}(n)]^2$ reaches a reasonable minimum. The simplest and most common way of calculating $\Delta\boldsymbol{w}(n)$ is by using the gradient descent or steepest descent algorithm [78]. In both cases, a criterion is defined as a function of the squared error (often called a performance index) such as $\eta\ (e(n)^2)$, that monotonically decreases after each iteration and converges to a global minimum. This requires

$$\eta(\boldsymbol{w} + \Delta\boldsymbol{w}) \leq \eta(\boldsymbol{w}) = \eta(\boldsymbol{e}(n)^2) \tag{2.155}$$

Assuming $\Delta\boldsymbol{w}$ to be very small, it is concluded that

$$\eta(\boldsymbol{w}) + \Delta\boldsymbol{w}^{\mathrm{T}}\nabla_{\boldsymbol{w}}(\eta(\boldsymbol{w})) \leq \eta(\boldsymbol{w}) \tag{2.156}$$

where $\nabla_{\boldsymbol{w}}(.)$ represents the gradient with respect to $\boldsymbol{w}$. This means that the above equation is satisfied by setting $\Delta\boldsymbol{w} = -\mu\nabla_{\boldsymbol{w}}(.)$, where μ is the learning rate or convergence parameter. Hence, the general update equation takes the form

$$\boldsymbol{w}(n+1) = \boldsymbol{w}(n) - \mu\nabla_{\boldsymbol{w}}(\eta(\boldsymbol{w}(n)) \tag{2.157}$$

Using the least mean square (LMS) approach, $\nabla_{\boldsymbol{w}}(\eta(\boldsymbol{w}))$ is replaced by an instantaneous gradient of the squared error signal, i.e.

$$\nabla_{\boldsymbol{w}}(\eta(\boldsymbol{w}(n))) \cong -2e(n)\boldsymbol{r}(n) \tag{2.158}$$

Therefore, the LMS-based update equation is

$$\boldsymbol{w}(n+1) = \boldsymbol{w}(n) + 2\ \mu e(n)\boldsymbol{r}(n) \tag{2.159}$$

Also, the convergence parameter, μ, must be positive and should satisfy

$$0 < \mu < \frac{1}{\lambda_{\max}} \tag{2.160}$$

where $\lambda_{\max}$ represents the maximum eigenvalue of the autocorrelation matrix $\mathbf{R}$. The LMS algorithm is the most simple and computationally efficient algorithm. However, the

speed of convergence can be slow, especially for correlated signals. The recursive least-squares (RLS) algorithm attempts to provide a high-speed stable filter, but it is numerically unstable for real-time applications [79,80]. The performance index is defined as

$$\eta(\boldsymbol{w}) = \sum_{i=0}^{n} \gamma^{n-i} \mathrm{e}^2(i) \tag{2.161}$$

Then, by taking the derivative with respect to $\boldsymbol{w}$ gives

$$\nabla_{\boldsymbol{w}} \eta(\boldsymbol{w}) = -2 \sum_{i=0}^{n} \gamma^{n-i} e(i) \boldsymbol{r}(i) \tag{2.162}$$

where $0 < \gamma \leq 1$ is the forgetting factor [79,80]. Replacing $e(n)$ in the above equation and writing it in vector form gives

$$\mathbf{R}(n) \boldsymbol{w}(n) = \boldsymbol{p}(n) \tag{2.163}$$

where

$$\mathbf{R}(n) = \sum_{i=0}^{n} \lambda^{n-i} \boldsymbol{r}(i) \boldsymbol{r}^{\mathrm{T}}(i) \tag{2.164}$$

and

$$\boldsymbol{p}(n) = \sum_{i=0}^{n} \lambda^{n-i} x(i) \boldsymbol{r}(i) \tag{2.165}$$

From this equation,

$$\boldsymbol{w}(n) = \mathbf{R}^{-1}(n) \boldsymbol{p}(n) \tag{2.166}$$

The RLS algorithm performs the above operation recursively such that $\mathbf{P}$ and $\mathbf{R}$ are estimated at the current time n as

$$\boldsymbol{p}(n) = \lambda \boldsymbol{p}(n-1) + x(n) \boldsymbol{r}(n) \tag{2.167}$$

$$\boldsymbol{R}(n) = \lambda \boldsymbol{R}(n-1) + \boldsymbol{r}(n) \boldsymbol{r}^{\mathrm{T}}(n) \tag{2.168}$$

In this case

$$\boldsymbol{r}(n) = \begin{bmatrix} r(n) \\ r(n-1) \\ \vdots \\ r(n-M) \end{bmatrix} \tag{2.169}$$

where M represents the finite impulse response (FIR) filter order. On the other hand,

$$\mathbf{R}^{-1}(n) = [\lambda \mathbf{R}^{-1}(n-1) + \boldsymbol{r}(n)\boldsymbol{r}^{\mathrm{T}}(n)]^{-1} \tag{2.170}$$

which can be simplified using the matrix inversion lemma [81]

$$\mathbf{R}^{-1}(n) = \frac{1}{\lambda}\left[\mathbf{R}^{-1}(n-1) - \frac{\mathbf{R}^{-1}(n-1)\boldsymbol{r}(n)\boldsymbol{r}^{\mathrm{T}}(n)\mathbf{R}^{-1}(n-1)}{\lambda + \boldsymbol{r}^{\mathrm{T}}(n)\mathbf{R}^{-1}(n-1)\boldsymbol{r}(n)}\right] \tag{2.171}$$

and finally the update equation can be written as

$$\boldsymbol{w}(n) = \boldsymbol{w}(n-1) + \mathbf{R}^{-1}(n)\boldsymbol{r}(n)\boldsymbol{g}(n) \tag{2.172}$$

where

$$\boldsymbol{g}(n) = x(n) - \boldsymbol{w}^{\mathrm{T}}(n-1)\boldsymbol{r}(n) \tag{2.173}$$

and the error $e(n)$ after each iteration is recalculated as

$$e(n) = x(n) - \boldsymbol{w}^{\mathrm{T}}(n)\boldsymbol{r}(n) \tag{2.174}$$

The second term on the right-hand side of the above equation is $\hat{\boldsymbol{v}}(n)$. The presence of $\mathbf{R}^{-1}(n)$ in Equation (2.172) is the major difference between the RLS and the LMS, but the RLS approach increases computation complexity by an order of magnitude.

2.9 Principal Component Analysis

All suboptimal transforms such as the DFT and DCT decompose the signals into a set of coefficients, which do not necessarily represent the constituent components of the signals. Moreover, as the transform kernel is independent of the data it is not efficient in terms of both decorrelation of the samples and energy compaction. Therefore, separation of the signal and noise components is generally not achievable using these suboptimal transforms.

Expansion of the data into a set of orthogonal components certainly achieves maximum decorrelation of the signals. This can enable separation of the data into the signal and noise subspaces.

For a single-channel EEG the Karhunen–Loève transform is used to decompose the ith channel signal into a set of weighted orthogonal basis functions:

$$x_i(n) = \sum_{k=1}^{N} w_{i,k}\varphi_k(n) \qquad \text{or} \qquad \boldsymbol{x}_{\mathrm{i}} = \boldsymbol{\Phi}\boldsymbol{w}_i \tag{2.175}$$

where $\boldsymbol{\Phi} = \{\varphi_k\}$ is the set of orthogonal basis functions. The weights $w_{i,k}$ are then calculated as

$$\boldsymbol{w}_i = \boldsymbol{\Phi}^{-1}\boldsymbol{x}_i \qquad \text{or} \qquad w_{i,k} = \sum_{n=0}^{N-1} \varphi_k^{-1}(n)x_i(n) \tag{2.176}$$

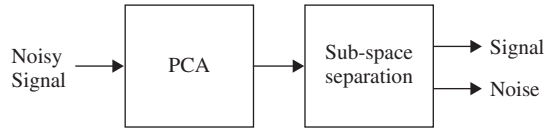

Figure 2.24 The general application of PCA

Often noise is added to the signal, i.e. $x_i(n) = s_i(n) + v_i(n)$, where $v_i(n)$ is additive noise. This degrades the decorrelation process. The weights are then estimated in order to minimize a function of the error between the signal and its expansion by the orthogonal basis, i.e. $\boldsymbol{e}_i = \boldsymbol{x}_i - \boldsymbol{\Phi}\boldsymbol{w}_i$. Minimization of the error in this case is generally carried out by solving the least-squares problem. In a typical application of PCA as depicted in Figure 2.24, the signal and noise subspaces are separated by means of some classification procedure.

2.9.1 Singular-Value Decomposition

The singular-value decomposition (SVD) is often used for solving the LS problem. This can be related to the decomposition of an $M \times M$ square autocorrelation matrix $\mathbf{R}$ into its eigenvalue matrix $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_M)$ and associated $M \times M$ orthogonal matrix of eigenvectors $\mathbf{V}$, i.e. $\mathbf{R} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{\text{H}}$, where $(.)^{\text{H}}$ denotes the Hermitian (conjugate transpose) operation. If $\mathbf{A}$ is an $M \times M$ data matrix such that $\mathbf{R} = \mathbf{A}^{\text{H}}\mathbf{A}$ then an $M \times M$ orthogonal (more generally unitary) matrix $\mathbf{U}$, an $M \times M$ orthogonal matrix $\mathbf{V}$, and an $M \times M$ diagonal matrix $\boldsymbol{\Sigma}$ exist with diagonal elements equal to $\lambda_i^{1/2}$, such that

$$\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\,\mathbf{V}^{\text{H}} \tag{2.177}$$

Hence $\boldsymbol{\Sigma}^2 = \Lambda$. The columns of $\mathbf{U}$ are called left singular vectors and the rows of $\mathbf{V}^{\text{H}}$ are called right singular vectors. If $\mathbf{A}$ is a rectangular $N \times M$ matrix of rank k then $\mathbf{U}$ will be $N \times N$ and $\boldsymbol{\Sigma}$ will be

$$\boldsymbol{\Sigma} = \begin{bmatrix} \mathbf{S} & 0 \\ 0 & 0 \end{bmatrix} \tag{2.178}$$

where $S = \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_k)$, where $\sigma_i = \lambda_i^{1/2}$. For such a matrix the Moore–Penrose pseudoinverse is defined as an $M \times N$ matrix $\mathbf{A}^{\dagger}$ defined as

$$\mathbf{A}^{\dagger} = \mathbf{U}\boldsymbol{\Sigma}^{\dagger}\mathbf{V}^{\text{H}} \tag{2.179}$$

where $\boldsymbol{\Sigma}^{\dagger}$ is an $M \times N$ matrix defined as

$$\boldsymbol{\Sigma}^{\dagger} = \begin{bmatrix} \mathbf{S}^{-1} & 0 \\ 0 & 0 \end{bmatrix} \tag{2.180}$$

$\mathbf{A}^{\dagger}$ has a major role in the solutions of least-squares problems and $\mathbf{S}^{-1}$ is a $k \times k$ diagonal matrix with elements equal to the reciprocals of the assumed non zero singular values of $\mathbf{A}$, i.e.

$$\mathbf{S}^{-1} = \text{diag}\left(\frac{1}{\sigma_1}, \frac{1}{\sigma_2}, \ldots, \frac{1}{\sigma_k}\right) \tag{2.181}$$

In order to see the application of the SVD in solving the LS problem consider the error vector $\boldsymbol{e}$, defined as

$$\boldsymbol{e} = \boldsymbol{d} - \mathbf{A}\boldsymbol{h} \tag{2.182}$$

where $\boldsymbol{d}$ is the desired signal vector and $\mathbf{A}\boldsymbol{h}$ is the estimate $\hat{\boldsymbol{d}}$. To find $\boldsymbol{h}$, $\mathbf{A}$ is replaced with its SVD in the above equation, which thereby minimizes the squared Euclidean norm of the error vector, $||\boldsymbol{e}||^2$. By using the SVD, it is found that

$$\boldsymbol{e} = \boldsymbol{d} - \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathrm{H}}\boldsymbol{h} \tag{2.183}$$

or equivalently

$$\mathbf{U}^{\mathrm{H}}\boldsymbol{e} = \mathbf{U}^{\mathrm{H}}\boldsymbol{d} - \boldsymbol{\Sigma}\mathbf{V}^{\mathrm{H}}\boldsymbol{h} \tag{2.184}$$

Since $\mathbf{U}$ is a unitary matrix, $||\boldsymbol{e}||^2 = ||\mathbf{U}^{\mathrm{H}}\boldsymbol{e}||^2$. Hence, the vector $\boldsymbol{h}$ that minimizes $||\boldsymbol{e}||^2$ also minimizes $||\mathbf{U}^{\mathrm{H}}\boldsymbol{e}||^2$. Finally, the unique solution as an optimum $\boldsymbol{h}$ (coefficient vector) may be expressed as [82]:

$$\boldsymbol{h} = \sum_{i=1}^{k} \frac{\boldsymbol{u}_i^{\mathrm{H}}\boldsymbol{d}}{\sigma_i}\boldsymbol{v}_i \tag{2.185}$$

where k is the rank of $\mathbf{A}$. Alternatively, as the optimum least-squares coefficient vector

$$\boldsymbol{h} = (\mathbf{A}^{\mathrm{H}}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{H}}\boldsymbol{d} \tag{2.186}$$

Performing a principal component analysis (PCA) is equivalent to performing an SVD on the covariance matrix. PCA uses the same concept as SVD and orthogonalization to decompose the data into constituent uncorrelated orthogonal components such that the autocorrelation matrix is diagonalized. Each eigenvector represents a principal component and the individual eigenvalues are numerically related to the variance they capture in the direction of the principal components. In this case the mean squared error (MSE) is simply the sum of the $N - K$ eigenvalues, i.e.

$$\text{MSE} = \sum_{k=N-K}^{N} \varphi_k^{\mathrm{T}} R_x \varphi_k = \sum_{k=N-K}^{N} \varphi_k^{\mathrm{T}}(\lambda_k \varphi_k) = \sum_{k=N-K}^{N} \lambda_k \tag{2.187}$$

PCA is widely used in data decomposition, classification, filtering, and whitening. In filtering applications the signal and noise subspaces are separated and the data are reconstructed from only the eigenvalues and eigenvectors of the actual signals. PCA is also

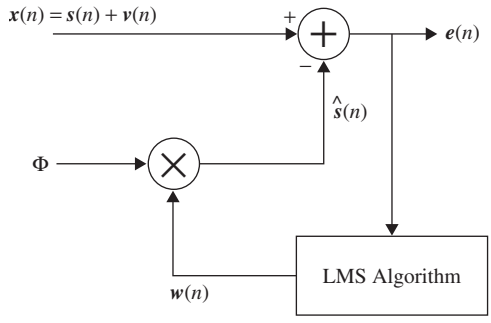

Figure 2.25 Adaptive estimation of the weight vector $\boldsymbol{w}(n)$

used for blind source separation of correlated mixtures if the original sources can be considered as statistically uncorrelated.

The PCA problem is then summarized as how to find the weights $\boldsymbol{w}$ in order to minimize the error given the observations only. The LMS algorithm is used here to minimize the MSE iteratively as

$$J_n = E[(\boldsymbol{x}(n) - \boldsymbol{\Phi}^{\mathrm{T}}(n)\boldsymbol{w}(n))^2] \tag{2.188}$$

The update rule for the weights is then

$$\boldsymbol{w}(n+1) = \boldsymbol{w}(n) + \mu \boldsymbol{e}(n)\boldsymbol{\Phi}(n) \tag{2.189}$$

where the error signal $\boldsymbol{e}(n) = \boldsymbol{x}(n) - \boldsymbol{\Phi}^{\mathrm{T}}(n)\boldsymbol{w}(n)$, $\mathbf{x}(n)$ is the noisy input, and n is the iteration index. The step size μ may be selected empirically or adaptively. These weights are then used to reconstruct the sources from the set of orthogonal basis functions. Figure 2.25 shows the overall system for adaptive estimation of the weight vector $\boldsymbol{w}$ using the LMS algorithm.

2.10 Independent Component Analysis

The concept of independent component analysis (ICA) lies in the fact that the signals may be decomposed into their constituent independent components. In places where the combined source signals can be assumed independent from each other this concept plays a crucial role in separation and denoising the signals.

A measure of independency may easily be described to evaluate the independence of the decomposed components. Generally, considering the multichannel signal as $\boldsymbol{y}(\mathrm{n})$ and the constituent signal components as $y_i(n)$, the $y_i(n)$ are independent if

$$p_{\mathbf{Y}}(\boldsymbol{y}(n)) = \prod_{i=1}^{m} p_y(y_i(n)) \quad \forall n \tag{2.190}$$

where $p(\mathbf{Y})$ is the joint probability distribution, $p_y(y_i(n))$ are the marginal distributions and m is the number of independent components.

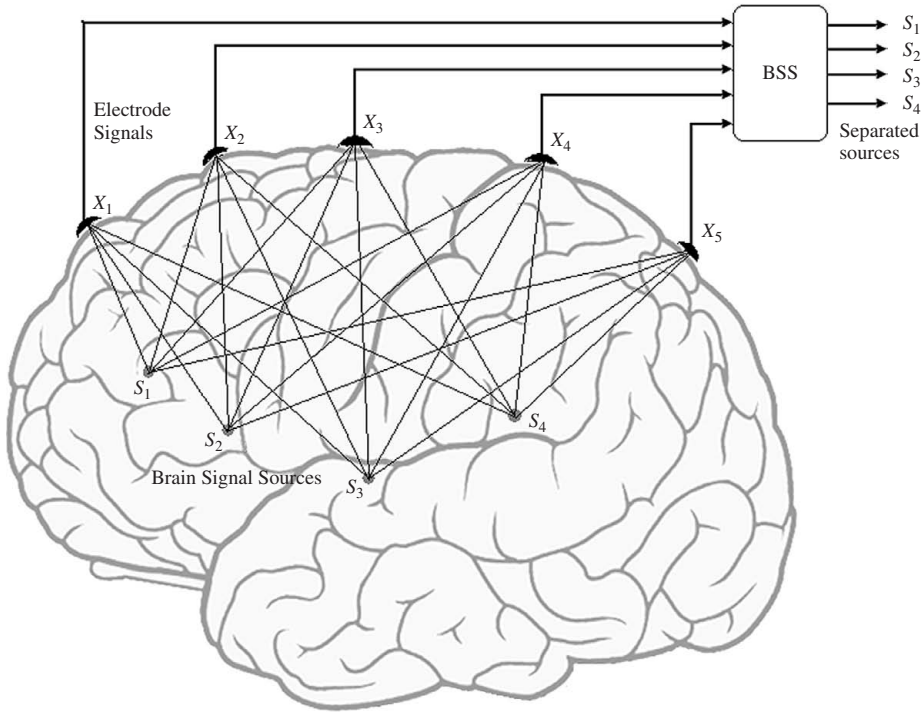

Figure 2.26 BSS concept; mixing and blind separation of the EEG signals

An important application of ICA is in blind source separation (BSS). BSS is an approach to estimate and recover the independent source signals using only the information of their mixtures observed at the recording channels. Due to its variety of applications BSS has attracted much attention recently. BSS of acoustic signals is often referred to as the 'cocktail party problem' [83], which means separation of individual sounds from a number of recordings in an uncontrolled environment such as a cocktail party. Figure 2.26 illustrates the BSS concept. As expected, ICA can be useful if the original sources are independent, i.e. $p(\boldsymbol{s}(n)) = \prod_{i=1}^{m} p_i(\boldsymbol{s}_i(n))$.

A perfect separation of the signals requires taking into account the structure of the mixing process. In a real-life application, however, this process is unknown, but some assumptions may be made about the source statistics.

Generally, the BSS algorithms do not make realistic assumptions about the environment in order to make the problem more tractable. There are typically three assumptions about the mixing medium. The most simple but widely used case is the instantaneous case, where the source signals arrive at the sensors at the same time. This has been considered for separation of biological signals such as the EEG, where the signals have narrow bandwidths and the sampling frequency is normally low. The BSS model in this case can be easily formulated as

$$\boldsymbol{x}(n) = \mathbf{H}\boldsymbol{s}(n) + \boldsymbol{v}(n) \tag{2.191}$$

where $m \times 1$ $\boldsymbol{s}(n)$, $n_e \times 1$ $\boldsymbol{x}(n)$, and $n_e \times 1$ $\boldsymbol{v}(n)$ denote respectively the vectors of source signals, observed signals, and noise at discrete time n. $\mathbf{H}$ is the mixing matrix of size $n_e \times m$. The separation is performed by means of a separating $m \times n_e$ matrix, $\mathbf{W}$, which uses only the information about $\boldsymbol{x}(n)$ to reconstruct the original source signals (or the independent components) as

$$\boldsymbol{y}(n) = \mathbf{W}\boldsymbol{x}(n) \tag{2.192}$$

In the context of EEG signal processing n_e denotes the number of electrodes. The early approaches in instantaneous BSS started from the work by Herault and Jutten [84] in 1986. In their approach, they considered non-Gaussian sources with similar number of independent sources and mixtures. They proposed a solution based on a recurrent artificial neural network for separation of the sources.

In acoustic applications, however, there are usually time lags between the arrival of the signals at the sensors. The signals also may arrive through multiple paths. This type of mixing model is called a convolutive model. One example is in places where the acoustic properties of the environment vary, such as a room environment surrounded by walls. Based on these assumptions the convolutive mixing model can be classified into two more types: anechoic and echoic. In both cases the vector representations of mixing and separating processes are changed to $\boldsymbol{x}(n) = \mathbf{H}(n)^* \boldsymbol{s}(n) + \boldsymbol{v}(n)$ and $\boldsymbol{y}(n) = \mathbf{W}(n)^* \boldsymbol{x}(n)$ respectively, where $*$ denotes the convolution operation.

In an anechoic model, however, the expansion of the mixing process may be given as

$$x_i(n) = \sum_{j=1}^{M} h_{ij} s_j(n - \delta_{ij}) + v_i(n), \quad \text{for } i = 1, \ldots, N \tag{2.193}$$

where the attenuation, h_{ij}, and delay, δ_{ij}, of source j to sensor i would be determined by the physical position of the source relative to the sensors. Then the unmixing process will be given as

$$y_j(m) = \sum_{i=1}^{N} w_{ji} x_i(m - \delta_{ji}), \quad \text{for } j = 1, \ldots, M \tag{2.194}$$

where the w_{ji}s are the elements of W. In an echoic mixing environment it is expected that the signals from the same sources reach to the sensors through multiple paths. Therefore the expansion of the mixing and separating models will be changed to

$$x_i(n) = \sum_{j=1}^{M} \sum_{k=1}^{K} h_{ij}^k s_j(n - \delta_{ij}^k) + v_i(n), \quad \text{for } i = 1, \ldots, N \tag{2.195}$$

where K denotes the number of paths and $v_i(n)$ is the accumulated noise at sensor i. The unmixing process will be formulated similarly to the anechoic one. Obviously, for a known number of sources an accurate result may be expected if the number of paths is known.

The aim of BSS using ICA is to estimate an unmixing matrix $\mathbf{W}$ such that $\mathbf{Y} = \mathbf{WX}$ best approximates the independent sources S, where $\mathbf{Y}$ and $\mathbf{X}$ are respectively matrices with columns $\boldsymbol{y}(n) = [y_1(n), y_2(n), \ldots, y_m(n)]^{\mathrm{T}}$ and $\boldsymbol{x}(n) = [x_1(n), x_2(n), \ldots x_{ne}(n)]^{\mathrm{T}}$. In any case, the unmixing matrix for the instantaneous case is expected to be equal to the inverse of the mixing matrix, i.e. $\mathbf{W} = \mathbf{H}^{-1}$. However, in all ICAs algorithms based upon restoring independence, the separation is subject to permutation and scaling ambiguities in the output independent components, i.e. $\mathbf{W} = \mathbf{PDH}^{-1}$, where $\mathbf{P}$ and $\mathbf{D}$ are the permutation and scaling matrices respectively.

There are three major approaches in using ICA for BSS:

1. Factorizing the joint PDF of the reconstructed signals into its marginal PDFs. Under the assumption that the source signals are stationary and non-Gaussian, the independence of the reconstructed signals can be measured by a statistical distance between the joint distribution and the product of its marginal PDFs. Kullback–Laibler (KL) divergence (distance) is an example. For nonstationary cases and for the short-length data, there will be poor estimation of the PDFs. Therefore, in such cases, this approach may not lead to good results. On the other hand, such methods are not robust for noisy data since in this situation the PDF of the signal will be distorted.
2. Decorrelating the reconstructed signals through time, i.e. diagonalizing the covariance matrices at every time instant. If the signals are mutually independent, the off-diagonal elements of the covariance matrix vanish, although the reverse of this statement is not always true. If the signals are nonstationary the time-varying covariance structure can be used to estimate the unmixing matrix. An advantage of this method is that it only uses second-order statistics, which implies that it is likely to perform better in noisy and short data length conditions than higher-order statistics.
3. Eliminating the temporal cross-correlation functions of the reconstructed signals as much as possible. In order to perform this, the correlation matrix of observations can be diagonalized at different time lags simultaneously. Here, second-order statistics are also normally used. As another advantage, it can be applied in the presence of white noise since such noise can be avoided by using the cross-correlation only for $\tau \neq 0$. Such a method is appropriate for stationary and weakly stationary sources (i.e. when the stationarity condition holds within a short segment of data).

It has been shown [85] that mutual information (MI) is a measure of independence and that maximizing the non-Gaussianity of the source signals is equivalent to minimizing the mutual information between them.

In the majority of cases the number of sources is known. This assumption avoids any ambiguity caused by false estimation of the number of sources. In exactly determined cases the number of sources is equal to the number of mixtures. In overdetermined situations, however, the number of mixtures is more than the number of sources.

There have been many attempts to apply BSS to EEG signals [86–96] for separation of normal brain rhythms, event-related signals, or mental or physical movement-related sources. If the number of sources is unknown, a criterion has to be established to estimate the number of sources beforehand. This process is a difficult task, especially when noise is involved. In those cases where the number of sources is more than the number of mixtures (known as underdetermined systems), the above BSS schemes cannot be applied

simply because the unmixing matrix will not be invertible, and generally the original sources cannot be extracted. However, when the signals are sparse other methods based on clustering may be utilized.

A signal is said to be sparse when it has many zero or at least approximately zero samples. Separation of the mixtures of such signals is potentially possible in the situation where at each sample instant the number of nonzero sources is not more than the number of sensors. The mixtures of sparse signals can also be instantaneous or convolutive. However, as will be briefly described later, the solution for only a simple case of a small number of idealized sources has been given in the literature.

In the context of EEG analysis, although the number of signals mixed at the electrodes seems to be limited, the number of sources corresponding to the neurons firing at a time can be enormous. However, if the objective is to study a certain rhythm in the brain the problem can be transformed to the time–frequency domain or even to the space–time–frequency domain. In such domains the sources may be considered disjoint and generally sparse. Also it is said that in the brain neurons encode data in a sparse way if their firing pattern is characterized by a long period of inactivity [97,98].

2.10.1 Instantaneous BSS

This is the most commonly used scheme for processing of the EEGs. The early work by Jutten and Herault led to a simple but fundamental adaptive algorithm [99]. Linsker [100] proposed unsupervised learning rules based on information theory that maximize the average mutual information between the inputs and outputs of an artificial neural network. Comon [85] performed minimization of mutual information to make the outputs independent. The Infomax algorithm [101] was developed by Bell and Sejnowski, which in spirit is similar to the Linsker method. Infomax uses an elegant stochastic gradient learning rule that was proposed by Amari *et al*. [102]. Non-Gaussianity of the sources was first exploited by Hyvarinen and Oja [103] in developing their fast ICA (fICA) algorithm. fICA is actually a blind source extraction algorithm, which extracts the sources one-by-one based on their kurtosis; the signals with transient peaks have high kurtosis. Later it was demonstrated that the Infomax algorithm and maximum likelihood estimation are in fact equivalent [104,105].

Based on the Infomax algorithm [101] for signals with positive kurtosis such as simultaneous EEG-fMRI and speech signals, minimizing the mutual information between the source estimates and maximizing the entropy of the source estimates are equivalent. Therefore, a stochastic gradient ascent algorithm can be used to iteratively find the unmixing matrix by maximization of the entropy. The Infomax algorithm finds a $\mathbf{W}$ that minimizes the following cost function:

$$J(\mathbf{W}) = I(\boldsymbol{z}, \boldsymbol{x}) = H(\boldsymbol{z}) - H(\boldsymbol{z}|\boldsymbol{x}) \tag{2.196}$$

where $H(\boldsymbol{z})$ is the entropy of the output, $H(\boldsymbol{z}|\boldsymbol{x})$ is the entropy of the output subject to a known input, and $\boldsymbol{z} = f(\boldsymbol{y})$ is a nonlinear activation function applied element wise to $\boldsymbol{y}$, the estimated sources. $I(\boldsymbol{z},\boldsymbol{x})$ is the mutual information between the input and output of the constructed adaptive neural network (ANN). $H(\boldsymbol{z}|\boldsymbol{x})$ is independent of $\mathbf{W}$; therefore, the gradient of J is only proportional to the gradient of $H(\boldsymbol{z})$. Correspondingly, the natural

gradient [107] of J denoted as $\nabla_{\mathbf{W}} J$ will be

$$\nabla_{\mathbf{W}} J = \nabla_{\mathbf{W}} I(\boldsymbol{z}, \boldsymbol{x}) \mathbf{W}^{\mathrm{T}} \mathbf{W} = \nabla_{\mathbf{W}} I(\boldsymbol{z}, \boldsymbol{x}) \mathbf{W}^{\mathrm{T}} \mathbf{W} \tag{2.197}$$

in which the time index n is dropped for convenience of presentation. Then, the sequential adaptation rule for the unmixing matrix $\mathbf{W}$ becomes

$$\mathbf{W}(n+1) = \mathbf{W}(n) + \mu[\mathbf{I} - 2f(\boldsymbol{y}(n))\boldsymbol{y}^{\mathrm{T}}(n)]\mathbf{W}(n) \tag{2.198}$$

where $f(\boldsymbol{y}(n)) = \{1 + \exp[-\boldsymbol{y}(n)]\}^{-1}$, assuming the outputs are super-Gaussian and μ is the learning rate, which is either a small constant or gradually changes following the speed of convergence.

Joint approximate diagonalization of eigenmatrices (JADE) is another well-known BSS algorithm [106] based on higher-order statistics (HOS). The JADE algorithm effectively diagonalizes the fourth-order cumulant of the estimated sources. This procedure uses certain matrices $\mathbf{Q}_z(\mathbf{M})$ formed by the inner product of the fourth-order cumulant tensor of the outputs with an arbitrary matrix $\mathbf{M}$, i.e.

$$\{\mathbf{Q}_z(\mathbf{M})\}_{ij} = \sum_{k=1}^{n_e} \sum_{l=1}^{n_e} \mathrm{Cum}(z_i, z_j^*, z_k, z_l^*) m_{lk} \tag{2.199}$$

where the (l, k)th component of the matrix $\mathbf{M}$ is written as m_{lk}, $\mathbf{Z} = \mathbf{CY}$, and $*$ denotes complex conjugate. The matrix $\mathbf{Q}_z(\mathbf{M})$ has the important property that it is diagonalized by the correct rotation matrix $\mathbf{U}$, i.e. $\mathbf{U}^{\mathrm{H}}\mathbf{QU} = \mathbf{\Lambda}_{\mathbf{M}}$, and $\mathbf{\Lambda}_{\mathbf{M}}$ is a diagonal matrix whose diagonal elements depend on the particular matrix $\mathbf{M}$ as well as $\mathbf{Z}$. By using Equation (2.199), for a set of different matrices $\mathbf{M}$, a set of cumulant matrices $\mathbf{Q}_z(\mathbf{M})$ can be calculated. The desired rotation matrix $\mathbf{U}$ then jointly diagonalizes these matrices. In practice, only approximate joint diagonalization is possible [106], i.e. the problem can be stated as minimization of

$$J(\mathbf{u}) = \sum_{j=1}^{n_e} \sum_{i=1}^{n_e} \text{ off } \{\boldsymbol{u}^{\mathrm{H}} \mathbf{Q}_{ij} \mathbf{u}\}$$

where

$$\text{off } (\mathbf{M}) = \sum_{i \neq j} |m_{ij}|^2 \tag{2.200}$$

EEG signals are, however, nonstationary. Nonstationarity of the signals has been exploited in developing an effective BSS algorithm based on second-order statistics called SOBI (second-order blind identification) [107]. In this algorithm separation is performed at a number of discrete time lags simultaneously. At each lag the algorithm unitarily diagonalizes the whitened data covariance matrix. It also mitigates the effect of noise on the observation by using a whitening matrix calculation, which can improve robustness to noise. Unitary diagonalization can be explained as follows. If $\mathbf{V}$ is a whitening

matrix and $\mathbf{X}$ is the observation matrix, the covariance matrix of the whitened observation is $\mathbf{C_X} = \mathrm{E}[\mathbf{V}\mathbf{X}\mathbf{X}^{\mathrm{H}}\mathbf{V}^{\mathrm{H}}] = \mathbf{V}\mathbf{R_X}\mathbf{V}^{\mathrm{H}} = \mathbf{V}\mathbf{H}\mathbf{R_S}\mathbf{H}^{\mathrm{H}}\mathbf{V}^{\mathrm{H}} = \mathbf{I}$, where $\mathbf{R_X}$ and $\mathbf{R_S}$ denote respectively the covariance matrices of the observed data and the original sources. It is assumed that $\mathbf{R_S} = \mathbf{I}$, i.e. the sources have unit variance and are uncorrelated, so $\mathbf{VH}$ is a unitary matrix. Therefore $\mathbf{H}$ can be factored as $\mathbf{H} = \mathbf{V}^{-1}\mathbf{U}$, where $\mathbf{U} = \mathbf{VH}$. The joint approximate diagonalization for a number of time lags can be obtained efficiently using a generalization of the Jacobi technique for the exact diagonalization of a single Hermitian matrix. The SOBI algorithm is implemented through the following steps as given in Reference [106]:

(a) The sample covariance matrix $\hat{\boldsymbol{R}}(0)$ is estimated from T data samples. The m largest eigenvalues and their corresponding eigenvectors of $\hat{\boldsymbol{R}}(0)$ are denoted as $\lambda_1, \lambda_2, \ldots, \lambda_m$ and $\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_m$ respectively.

(b) Under the white noise assumption, an estimate $\hat{\sigma}^2$ of the noise variance is the average of the $n_{\mathrm{e}}-m$ smallest eigenvalues of $\hat{\boldsymbol{R}}(0)$. The whitened signals are $\boldsymbol{z}(n) = [z_1(n), z_2(n), \ldots, z_{n_{\mathrm{e}}}(n)]^{\mathrm{T}}$, computed by $z_i(n) = (\lambda_i - \hat{\sigma}^2)^{-1/2}\mathbf{h}_i^{\mathrm{H}}\mathbf{x}(n)$ for $1 \leq i \leq n_{\mathrm{e}}$. This is equivalent to forming a whitening matrix as $\hat{W} = \left[(\lambda_1 - \hat{\sigma}^2)^{-1/2}\boldsymbol{h}_1, \ldots, (\lambda_{n_{\mathrm{e}}} - \hat{\sigma}^2)^{-1/2}\boldsymbol{h}_{n_{\mathrm{e}}}\right]^{\mathrm{H}}$.

(c) Form sample estimates $\hat{\underline{\mathbf{R}}}(\tau)$ by computing the sample covariance matrices of $\boldsymbol{z}(t)$ for a fixed set of time lags $\tau \in \{\tau_j | j = 1, \ldots, K\}$.

(d) A unitary matrix $\hat{\boldsymbol{U}}$ is then obtained as a joint diagonalizer of the set $\{\hat{\underline{\mathbf{R}}}(\tau_j) | j = 1, \ldots, K\}$.

(e) The source signals are estimated as $\hat{\boldsymbol{s}}(t) = \hat{\boldsymbol{U}}^{\mathrm{H}}\hat{\boldsymbol{W}}\boldsymbol{x}(t)$ or the mixing matrix $\mathbf{A}$ is estimated as $\hat{\boldsymbol{A}} = \hat{\boldsymbol{W}}^{\dagger}\hat{\boldsymbol{U}}$, where the superscript † denotes the Moore–Penrose pseudoinverse.

The FICA algorithm [103] is another very popular BSS technique which extracts the signals one by one based on their kurtosis. In fact, the algorithm uses an independence criterion that exploits non-Gaussianity of the estimated sources. In some places where the objective is to remove the spiky artefacts, such as the removal of the fMRI artefact from the simultaneous EEG-fMRI recordings, application of an iterative fICA followed by deflation of the artefact component gives excellent results [108]. A typical signal of this type is given in Figure 2.27.

Practically, fICA maximizes the negentropy, which represents the distance between a distribution and a Gaussian distribution having the same mean and variance, i.e.

$$\mathrm{Neg}(y) \propto \{E[f(y)] - E[f(y_{\mathrm{Gaussian}})]\}^2 \tag{2.201}$$

where f is a score function [109] and Neg stands for negentropy. This, as mentioned previously, is equivalent to maximizing the kurtosis. Therefore, the cost function can be simply defined as

$$J(\mathbf{W}) = -\frac{1}{4}|k_4(\mathbf{y})| = -\frac{\beta}{4}k_4(\mathbf{y}) \tag{2.202}$$

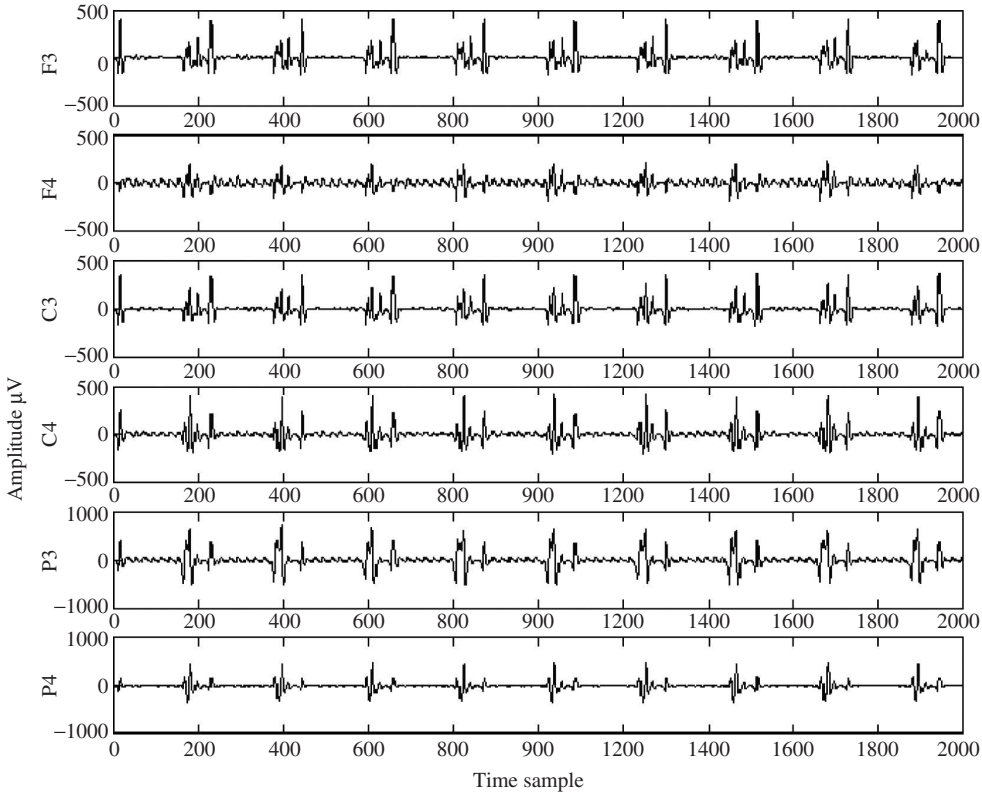

Figure 2.27 A sample of an EEG signal simultaneously recorded with fMRI

where $k_4(\mathbf{y})$ is the kurtosis, and β is the sign of the kurtosis. Applying the standard gradient decent approach to minimize the cost function gives

$$\mathbf{W}(n+1) = \mathbf{W}(n) - \mu \frac{\partial J(\mathbf{W})}{\partial \mathbf{W}}\bigg|_{\mathbf{W}=\mathbf{W}(n)} \tag{2.203}$$

where

$$-\mu \frac{\partial J(\mathbf{W})}{\partial \mathbf{W}}\bigg|_{\mathbf{W}=\mathbf{W}(n)} = \mu(n)\varphi(\mathbf{y}(n))\boldsymbol{x}(n) \tag{2.204}$$

Here $\mu(n)$ is a learning rate,

$$\varphi(y_i) = \beta \frac{\hat{m}_4(y_i)}{\hat{m}_2^3(y_i)} \left[\frac{\hat{m}_2(y_i)}{\hat{m}_4(y_i)} y_i^3 - y_i \right] \tag{2.205}$$

and $\hat{m}_q(y_i) = \hat{E}[y_i^q(n)]$, which is an estimate of the qth-order moment of the actual sources. Since fICA extracts the sources one-by-one a deflation process is followed to

exclude the extracted source from the mixtures. The process reconstructs the mixtures iteratively by

$$\boldsymbol{x}_{j+1} = \boldsymbol{x}_j - \tilde{\boldsymbol{w}}_j \boldsymbol{y}_j, \quad j = 1, 2, \ldots \tag{2.206}$$

where $\tilde{\boldsymbol{w}}_j$ is estimated by minimization of the following cost function:

$$J(\tilde{\boldsymbol{w}}_j) = \frac{1}{2} E\left[\sum_{p=1}^{n_r} \boldsymbol{x}_{j+1,p}^2\right] \tag{2.207}$$

where n_r is the number of remaining mixtures.

Figure 2.28 shows the results after application of fICA to remove the scanner artefact from the EEGs. In addition to the separation of EEGs using fICA, very good results have been reported after application of fICA to separation of temporomanibular joint sounds [110].

In a time–frequency (TF) approach, which assumes that the sources are approximately cyclostationary and nonstationary, the auto-terms and cross-terms of the covariance matrix of the mixtures are first separated and BSS is applied to both terms [111,112]. In this approach, the spatial time–frequency distribution (STFD) of the mixed signals is

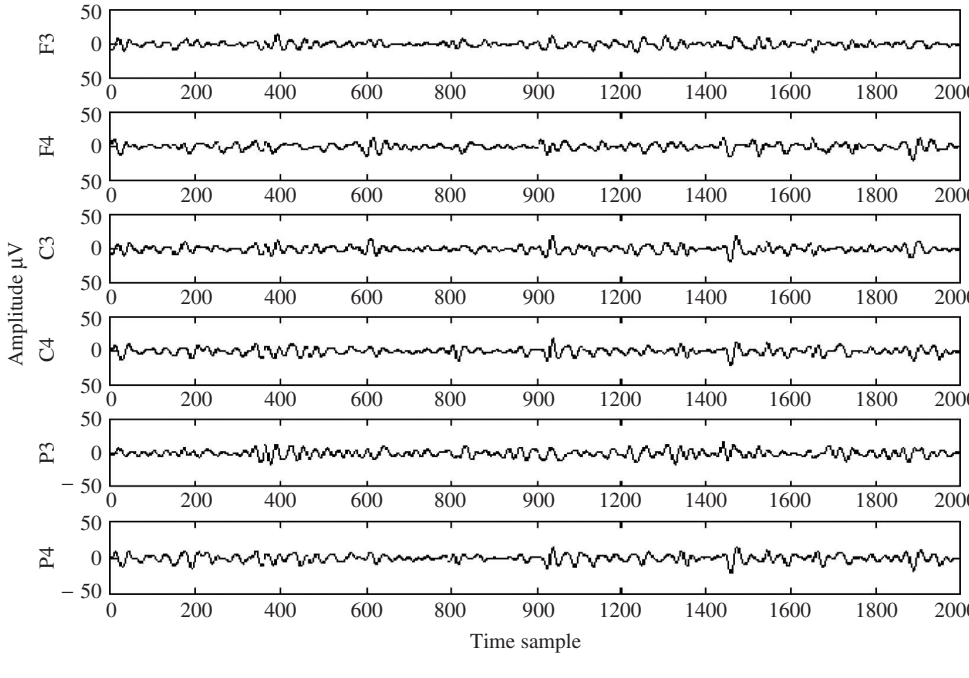

Figure 2.28 The EEG signals after removal of the scanner artefact.

defined as

$$\mathbf{D}_{\boldsymbol{xx}}(n,\omega)=\frac{1}{2\pi}\sum_{u=\tau/2}^{N-\tau/2}\sum_{\tau=0}^{N/2-1}\phi(n-u,\tau)\mathrm{e}^{-\mathrm{i}\omega\tau}E\left[\boldsymbol{x}\left(u+\frac{\tau}{2}\right)\boldsymbol{x}\left(u-\frac{\tau}{2}\right)\right] \tag{2.208}$$

where $\phi(.)$ is the discretized kernel function defining a distribution from Cohen's class of TF distributions [113] and $\boldsymbol{x}(.)$ is an N sample observation of the signals, which is normally contaminated by noise. Assuming $\boldsymbol{x}(t)=\mathbf{A}\boldsymbol{s}(t)+\boldsymbol{v}(t)$, using the above equation it is found that

$$\mathbf{D}_{\boldsymbol{xx}}(n,\omega)=\mathbf{A}\mathbf{D}_{\boldsymbol{ss}}(n,\omega)\mathbf{A}^{\mathrm{H}}+\sigma^2\mathbf{I} \tag{2.209}$$

where $\mathbf{D}_{\boldsymbol{ss}}(.)$ is the STFD of the source signals and σ^2 is the noise variance and depends on both noise power and the kernel function. From this equation it is clear that both $\mathbf{D}_{\boldsymbol{xx}}$ and $\mathbf{D}_{\boldsymbol{ss}}$ exhibit the same eigenstructure. The covariance matrix of the source signals is then replaced by the source STFD matrix composed of auto- and cross-source time–frequency distributions (TFDs) respectively, on the diagonal and off-diagonal entries.

Defining a whitening matrix $\mathbf{W}$ such that $\mathbf{U}=\mathbf{W}\mathbf{A}$ is unitary, a whitened and noise-compensated STFD matrix is defined as

$$\begin{aligned}\tilde{\boldsymbol{D}}_{\boldsymbol{xx}}(n,\omega)&=\mathbf{W}(\mathbf{D}_{\boldsymbol{xx}}(n,\omega)-\sigma^2\mathbf{I})\mathbf{W}^{\mathrm{H}}\\&=\mathbf{U}\mathbf{D}_{\boldsymbol{ss}}(n,\omega)\mathbf{U}^{\mathrm{H}}\end{aligned} \tag{2.210}$$

$\mathbf{W}$ and σ^2 can be estimated from the sample covariance matrix and $\mathbf{D}_{\boldsymbol{xx}}$ is estimated based on the discrete-time formulation of the TFDs. From Equation (2.191) it is known that the sensor STFD matrix exhibits the same eigenstructure as the data covariance matrix commonly used for cyclic data [111]. The covariance matrix of the source signals is replaced by a source STFD matrix composed of the auto- and cross-source TFDs on the diagonal and off-diagonal entries respectively. The peaks occur in mutually exclusive locations on the TF plane. The kernel function can be defined in such a way as to maximize disjointness of the points in the TF plane. By estimation of the STFD in Equation (2.192) at appropriate TFD points, it is possible to recover the source signals by estimating a unitary transformation $\hat{\boldsymbol{U}}$, via optimization of a joint diagonal and off-diagonal criterion, to have

$$\hat{\boldsymbol{s}}(n)=\hat{\boldsymbol{U}}^{\mathrm{H}}\mathbf{W}\boldsymbol{x}(n)\quad\text{for } n=1,\ldots,N-1 \tag{2.211}$$

In order to define and extract the peaks of $\mathbf{D}_{\boldsymbol{xx}}$ a suitable clustering approach has a to be followed. This algorithm has potential application for estimating the EEG sources since in most normal cases the sources are cyclic or quasicyclic.

2.10.2 Convolutive BSS

In many practical situations the signals reach the sensors with different time delays. The corresponding delay between source j and sensor i, in terms of number of samples, is directly proportional to the sampling frequency and conversely to the speed of sound, i.e. $\delta_{ij}\propto d_{ij}f_{\mathrm{s}}/c$, where d_{ij}, f_{s}, and c are the distance between source j and sensor i,

the sampling frequency, and the speed of sound respectively. For speech and music in air, as an example the following could be taken: d_{ij} in terms of metres, f_s between 8 and 44 kHz, and $c = 330$ m/s. Also, in an acoustic environment the sound signals can reach the sensors through multipaths after reflections by obstacles (such as walls). The above two cases have been addressed as anechoic and echoic BSS models respectively and formulated at the beginning of this section. The solution to echoic cases is obviously more difficult and normally involves some approximations to the actual system. As an example, in the previously mentioned cocktail party problem the source signals propagate through a dynamic medium with many parasitic effects, such as multiple echoes and reverberation. Therefore, the received signals are to a first approximation a weighted sum of mixed and delayed components. In other words, the received signals at each microphone are the convolutive mixtures of speech signals.

Unfortunately, most of the proposed BSS approaches to instantaneous mixtures fail or are limited in separation of convolutive mixtures, generally due to:

(a) noise;
(b) possibly a smaller number of sensors than the number of source signals (from the sources directly and through multipaths);
(c) nonstationarity of the signals;
(d) time delays, which make the overall mixing not instantaneous.

Convolutive BSS has recently been a focus of research in the acoustic signal processing community. Two major approaches have been followed for both anechoic and echoic cases. The first approach is to solve the problem in the time domain. In such methods, in order to have accurate results both the weights of the unmixing matrix and the delays have to be estimated. However, in the second approach, the problem can be transformed into the frequency domain as $\boldsymbol{h}(n)^{*}\boldsymbol{s}(n)\overset{\mathrm{F}}{\rightarrow}\mathbf{H}(\omega)\cdot\mathbf{S}(\omega)$ and instantaneous BSS applied to each frequency bin mixed signal. The separated signals at different frequency bins are then combined and transformed to the time domain to reconstruct the estimated sources. The short-term discrete Fourier transform is often used for this purpose. Figure 2.29 clearly represents the frequency-domain BSS of convolutive mixtures. However, the inherent permutation problem of BSS severely deteriorates the results since the order of the separated sources in different frequency bins can vary from segment to segment of the signals.

An early work in convolutive BSS by Platt and Faggin [114], who applied the adaptive noise cancellation network to the BSS model of Herault and Jutten [115], which has delays in the feedback path, was based on the minimum output power principle. This scheme exploits the fact that the signal corrupted by noise has more power than the clean signal. The feedback path cancels out the interferences as the result of delayed versions of the other sources. This circuit was also used later to extend the Infomax BSS to convolutive cases [116]. The combined network maximizes the entropy at the output of the network with respect to the weights and delays. Torkkola [117] extended this algorithm to the echoic cases. In order to achieve a reasonable convergence, some prior knowledge of the recording situation is necessary.

In another work an extension of the SOBI algorithm has been used for anechoic BSS [118]. The problem has been transformed to the frequency domain and joint diagonalization of spectral matrices has been utilized to estimate the mixing coefficients as well as the

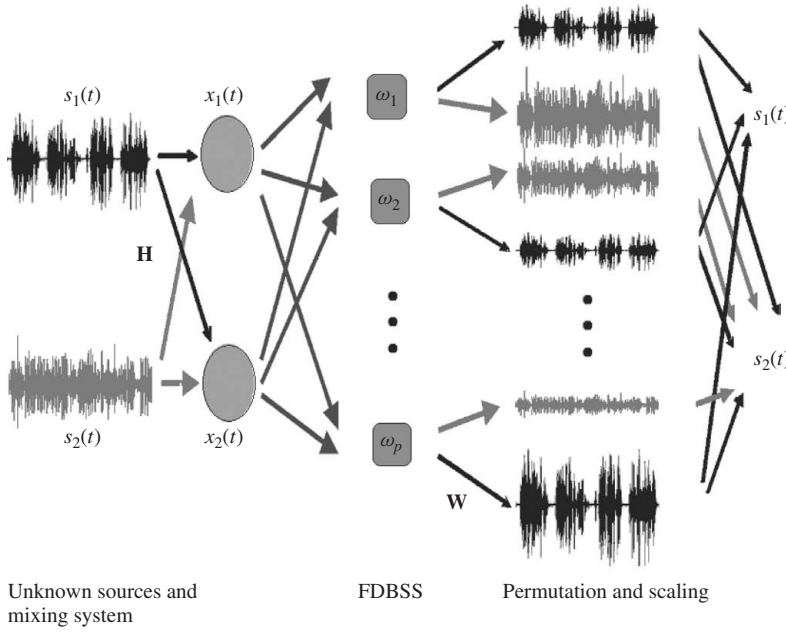

Figure 2.29 A schematic diagram of the frequency-domain BSS of convolutive mixtures for a simple two-source model

delays [119]. In attempts by Parra *et al.* [120], Ikram and Morgan [121], and Cherkani and Deville [122], second-order statistics have been used to ensure that the estimated sources, $\mathbf{Y}(\omega, m)$, are uncorrelated at each frequency bin. $\mathbf{W}(\omega)$ is estimated in such a way that it diagonalizes the covariance matrices $\mathbf{R}_{\mathbf{Y}}(\omega, k)$ simultaneously for all time blocks k, $k = 0, 1, \ldots, K-1$, i.e.

$$\begin{aligned}\mathbf{R}_{\mathbf{Y}}(\omega, k) &= \mathbf{W}(\omega)\mathbf{R}_{\mathbf{X}}(\omega, k)\mathbf{W}^{\mathrm{H}}(\omega) \\ &= \mathbf{W}(\omega)\mathbf{H}(\omega)\mathbf{\Lambda}_{\mathrm{S}}(\omega, k)\mathbf{H}^{\mathrm{H}}(\omega)\mathbf{W}^{\mathrm{H}}(\omega) \\ &= \mathbf{\Lambda}_{\mathrm{c}}(\omega, k)\end{aligned} \tag{2.212}$$

where $\mathbf{\Lambda}_{\mathbf{S}}(\omega, k)$ is the covariance matrix of the source signals, which changes with k, $\mathbf{\Lambda}_{\mathrm{c}}(\omega, k)$ is an arbitrary diagonal matrix, and $\mathbf{R}_{\mathbf{X}}(\omega, k)$ is the covariance matrix of $\mathbf{X}(\omega)$, estimated by

$$\hat{\boldsymbol{R}}_X(\omega, k) = \frac{1}{N}\sum_{n=0}^{N-1}\mathbf{X}(\omega, NK+n)\mathbf{X}^{\mathrm{H}}(\omega, NK+n) \tag{2.213}$$

where N is the number of mixtures; the unmixing filter $\mathbf{W}(\omega)$ for each frequency bin ω that simultaneously satisfies the K decorrelation equations can then be obtained using an overdetermined least-squares solution. Since the output covariance matrix $\mathbf{R}_{\mathbf{Y}}(\omega, k)$ has

to be diagonalized the update equation for estimation of the unmixing matrix $\mathbf{W}$ can be found by minimizing the off-diagonal elements of $\mathbf{R}_{\mathbf{Y}}(\omega, k)$, which leads to

$$\mathbf{W}_{\rho\upsilon+1}(\omega) = \mathbf{W}_{\rho}(\omega) - \mu(\omega)\frac{\partial}{\partial \mathbf{W}_{\rho}^{\mathrm{H}}(\omega)}\{\|\mathbf{V}_{\rho}(\omega, k)\|^2\} \tag{2.214}$$

where ρ is the iteration index, $\|\cdot\|^2$ is the squared Frobenius norm,

$$\mu(\omega) = \frac{\alpha}{\sum_k \|\mathbf{R}_{\mathbf{X}}(\omega, k)\|^2} \tag{2.215}$$

and

$$\mathbf{V}(\omega, k) = \mathbf{W}(\omega)\mathbf{R}_{\mathbf{X}}(\omega, k)\mathbf{W}^{\mathrm{H}}(\omega) - \mathrm{diag}[\mathbf{W}(\omega)\mathbf{R}_{\mathbf{X}}(\omega, k)\mathbf{W}^{\mathrm{H}}(\omega)] \tag{2.216}$$

and α is a constant, which is adjusted practically.

In these methods a number of solutions for mitigating the permutation ambiguity have been suggested. Smaragdis [123] reformulated the Infomax algorithm for the complex domain and used it to solve the BSS in the frequency domain. Murata *et al.* [124] also formulated the problem of BSS in each frequency bin using a simultaneous diagonalization method similar to the SOBI method. To mitigate the permutation problem a method based on the temporal structure of signals, which exploits the nonstationarity of speech was introduced. The method exploits the correlations between the frequency bins of the spectrum of the signals.

However, for the EEG mixing model the f_s is normally low (since the bandwidth $<$100 Hz) and the propagation velocity is equivalent to that of electromagnetic waves (300,000 km/s). Therefore, the delay is almost zero and the mixing model can always be considered to be instantaneous. The main drawbacks for the application of BSS to separation of EEG signals is due to the:

(a) noisy environment;
(b) unknown number of sources;
(c) nonstationarity of the sources;
(d) movement of the ERP sources.

Although many attempts have been made to solve the above problems more efforts are required to provide robust solutions for different applications.

2.10.3 Sparse Component Analysis

In places where the sources are sparse, i.e. at each time instant, the number of non-zero values are less or equal to the number of mixtures. The columns of the mixing matrix may be calculated individually, which makes the solution to the underdetermined case possible. The problem can be stated as a clustering problem since the lines in the scatter plot can be separated based on their directionalities by means of clustering [125,126]. The same idea has been followed more comprehensively by Li *et al.* [127]. In their method, however, the separation has been performed in two different stages. First, the unknown mixing matrix is estimated using the k-means clustering method. Then, the source matrix is estimated

using a standard linear programming algorithm. The line orientation of a dataset may be thought of as the direction of its greatest variance. One way is to perform eigenvector decomposition on the covariance matrix of the data, the resultant principal eigenvector, i.e. the eigenvector corresponding with the largest eigenvalue, indicates the direction of the data. There are many cases for which the sources are disjoint in other domains rather than the time domain. In these cases the sparse component analysis can be performed in those domains more efficiently. One such approach, called DUET [128], transforms the anechoic convolutive observations into the time–frequency domain using a short-time discrete Fourier transform, and the relative attenuation and delay values between the two observations are calculated from the ratio of corresponding time–frequency points. The regions of significant amplitudes (atoms) are then considered to be the source components in the time–frequency domain.

For instantaneous cases, in separation of sparse sources the common approach used by most researchers is to attempt to maximize the sparsity of the extracted signals in the output of the separator. The columns of the mixing matrix $\mathbf{A}$ assign each observed data point to only one source based on some measure of proximity to those columns [129]; i.e. at each instant only one source is considered active. Therefore the mixing system can be presented as

$$x_i(n) = \sum_{j=1}^{M} a_{ji} s_j(n), \quad i = 1, \ldots, N \tag{2.217}$$

where in an ideal case $a_{ji} = 0$ for $i \neq j$. Minimization of the L_1-norm is one of the most logical methods for estimation of the sources. L_1-norm minimization is a piecewise linear operation that partially assigns the energy of $\boldsymbol{x}(n)$ to the M columns of $\mathbf{A}$ that form a cone around $\boldsymbol{x}(n)$ in $\Re^M$ space. The remaining $N-M$ columns are assigned zero coefficients; therefore the L_1-norm minimization can be manifested as

$$\min ||\boldsymbol{s}(n)||_1 \text{ subject to } \mathbf{A}\boldsymbol{s}(n) = \boldsymbol{x}(n) \tag{2.218}$$

A detailed discussion of signal recovery using L_1-norm minimization is presented by Takigawa *et al.* [130], but it is worth highlighting its potential advantages for short datasets. As mentioned above, it is important to choose a domain in which the signals are more sparse. Para-factor (PARAFAC) analysis is an effective tool in detection and classification of sources in a multidimensional space. In a very recent approach it has been considered that the brain signal sources in the space–time–frequency domain are disjoint. Therefore clustering the observation points in the space–time–frequency domain can be effectively used for separation of brain sources [131]. The outcome is highly valuable for detection of μ rhythms corresponding to left and right finger movements in the context of brain–computer interfacing (BCI). The details of PARAFAC and its application to BCI can be found in Chapter 7 of this book.

2.10.4 Nonlinear BSS

Consider the cases where the parameters of the mixing system change because of changes in the mixing environment or change in the statistics of the sources. For example, if the images of both sides of a semi-transparent paper are photocopied the results will be two mixtures of the original sources. However, since the minimum observable grey level is black (or zero) and the maximum is white (say 1), the sum of the grey levels cannot go

beyond these limits. This represents a nonlinear mixing system. As another example, think of the joint sounds heard from surface electrodes from over the skin. The mixing medium involves acoustic parameters of the body tissues. However, the tissues are not rigid. In such cases, if the tissues vibrate due to the sound energy then the mixing system will be a nonlinear system. The mixing and unmixing can generally be modelled respectively as

$$\boldsymbol{x}(n) = f(\mathbf{A}\boldsymbol{s}(n) + \boldsymbol{n}(n)) \tag{2.219}$$

$$\boldsymbol{y}(\mathrm{t}) = g(\mathbf{W}\boldsymbol{x}(n)) \tag{2.220}$$

where $f(.)$ and $g(.)$ represent respectively the nonlinearities in the mixing and unmixing processes. There have been some attempts to solve nonlinear BSS problems, especially for separation of image mixtures [132,133]. In one attempt [132] the mixing system has been modelled as a radial basis function (RBF) neural network. The parameters of this network are then computed iteratively. However, in these methods an assumption is often made about the mixing model. Unfortunately, none of these methods currently give satisfactory results.

2.10.5 Constrained BSS

The optimization problem underlying the solution to the BSS problem may be subject to fulfilment of a number of conditions. These may be based as *a priori* knowledge of the sources or the mixing system. Any constraint on the estimated sources or the mixing system (or unmixing system) can lead to a more accurate estimation of the sources. Statistical [104] as well as geometrical constraints [105] have very recently been used in developing new BSS algorithms. In most of the cases the constrained problem is converted to an unconstrained one by means of a regularization parameter such as a Lagrange multiplier or more generally a nonlinear penalty function, as used in Reference [104].

Incorporating nonlinear penalty functions [118] into a joint diagonalization problem not only exploits nonstationarity of the signals but also ensures fast convergence of the update equation. A general formulation for the cost function of such a system can be in the form of

$$\mathbf{J}(\mathbf{W}) = \mathbf{J}_{\mathrm{m}}(\mathbf{W}) + \kappa\varphi(\mathbf{J}_{\mathrm{c}}(\mathbf{W})) \tag{2.221}$$

where $\mathbf{J}_{\mathrm{m}}(\mathbf{W})$ and $\mathbf{J}_{\mathrm{c}}(\mathbf{W})$ are respectively the main and the constraint cost functions, $\varphi(\cdot)$ is the nonlinear penalty function, and κ is the penalty parameter.

Constrained BSS has a very high potential in incorporating clinical information into the main optimization formulation. As a new application of constrained BSS, an effective algorithm has been developed for removing the eye-blinking artefacts from EEGs. A similar method to the joint diagonalization of correlation matrices by using gradient methods [134] has been developed [88], which exploits the temporal structure of the underlying EEG sources. The algorithm is an extension of SOBI, with the aim of iteratively performing the joint diagonalization of multiple time lagged covariance matrices of the estimated sources and exploiting the statistical structure of the eye-blinking signal as a constraint. The estimated source covariance matrix is given by

$$\mathbf{R}_{\mathbf{Y}}(k) = \mathbf{W}\mathbf{R}_{\mathbf{X}}(k)\mathbf{W}^{\mathrm{T}} \tag{2.222}$$

where $\mathbf{R}_{\mathbf{X}}(k) = E\{\boldsymbol{x}(n)\mathbf{x}^{\mathrm{T}}(n-k)\}$ is the covariance matrix of the electrode data. Following the same procedure as in Reference [135], the least-squares (LS) estimate of W is found from

$$J_{\mathrm{m}}(\mathbf{W}) = \arg\min_{\mathbf{W}} \sum_{k=1}^{T_{\mathrm{B}}} \|E(k)\|_{\mathrm{F}}^2 \tag{2.223}$$

where $\|\cdot\|_{\mathrm{F}}^2$ is the squared Frobenius norm and $E(k)$ is the error to be minimized between the covariances of the source signals, $\mathbf{R}_S(k)$ and the estimated sources, $\mathbf{R}_Y(k)$. The corresponding cost function has been defined, based on minimizing the off-diagonal elements for each time block, i.e.

$$J(\mathbf{W}) = J_{\mathrm{m}}(\mathbf{W}) + \Lambda J_{\mathrm{c}}(\mathbf{W}) \tag{2.224}$$

where

$$J_{\mathrm{m}}(\mathbf{W}) = \sum_{k=1}^{T_{\mathrm{B}}} \|\mathbf{R}_{\mathbf{Y}}(k) - \mathrm{diag}(\mathbf{R}_{\mathbf{Y}}(k)\|_{\mathrm{F}}^2 \tag{2.225}$$

and

$$J_{\mathrm{c}}(\mathbf{W}) = F(E[\boldsymbol{g}(n)\boldsymbol{y}^{\mathrm{T}}(n)]) \tag{2.226}$$

is a second-order constraint term. $F(.)$ is a nonlinear function approximating the cumulative density function (CDF) of the data and $\Lambda = \{\lambda_{ij}\}(i, j = 1, \ldots, N)$ is the weighted factor which is governed by the correlation (matrix) between the EOG and EEG signals ($\mathbf{R}_{\mathbf{GY}}$), defined as $\Lambda = \kappa\ \mathrm{diag}(\mathbf{R}_{\mathbf{GY}})$, where κ is an adjustable constant. Then a gradient approach [112] is followed to minimize the cost function. The incremental update equation is

$$\mathbf{W}(n+1) = \mathbf{W}(n) - \mu \frac{\partial J(\mathbf{W})}{\partial \mathbf{W}} \tag{2.227}$$

which concludes the algorithm.

Blind source separation has been widely used for processing EEG signals. Although the main assumptions about the source signals, such as uncorrelatedness or independency of such signals, have not yet been verified, the empirical results illustrate the effectiveness of such methods. EEG signals are noisy and nonstationary signals, which are normally affected by one or more types of internal artefacts. The most efficient approaches are those that consider all different domain statistics of the signals and take the nature of the artefacts into account. In addition, a major challenge is in how to incorporate and exploit the physiological properties of the signals and characteristics of the actual sources into the BSS algorithm. Some examples have been given here; more will be presented in the other chapters of this book.

In the case of brain signals the independency or uncorrelatedness conditions for the sources may not be satisfied. This, however, may be acceptable for abnormal sources, movement-related sources, or ERP signals. Transforming the problem into a domain such as the space–time–frequency domain, where the sources can be considered disjoint, may be a good solution.

2.11 Application of Constrained BSS: Example

In practice, the natural signals such as EEG source signals are not always independent. A topographic ICA method proposed in Reference [136] incorporates the dependency among the nearby sources in not only grouping the independent components related to nearby sources but also separating the sources originating from different regions of the brain. In this ICA model it is proposed that the residual dependency structure of the independent components (ICs), defined as dependencies that cannot be cancelled by ICA, could be used to establish a topographic order between the components. Based on this model, if the topography is defined by a lattice or grid, the dependency of the components is a function of the distance of the components on that grid. Therefore, the generative model, which implies correlation of energies for components that are close in the topographic grid, is defined. The main assumption is that the nearby sources are correlated and those far from each other are independent.

To develop such an algorithm a neighbourhood relation is initially defined as

$$h(i, j) = \begin{cases} 1, & \text{if } |i - j| \leq m \\ 0, & \text{otherwise} \end{cases} \tag{2.228}$$

where the constant m specifies the width of the neighbourhood. Such a function is therefore a matrix of hyperparameters. This function can be incorporated into the main cost function of BSS. The update rule is then given as [136]

$$\boldsymbol{w}_i \propto E[\boldsymbol{z}(\boldsymbol{w}_i^{\mathrm{T}}\boldsymbol{z})r_i] \tag{2.229}$$

where $\boldsymbol{z}_i$ is the whitened mixed signals and

$$r_i = \sum_{k=1}^{N} h(i, k) g\left(\sum_{j=1}^{N} h(k, j)(\boldsymbol{w}_j^{\mathrm{T}}\boldsymbol{z})^2\right) \tag{2.230}$$

The function g is the derivative of a nonlinear function such as those defined in Reference [136]. It is seen that the vectors $\boldsymbol{w}_i$ are constrained to some topographic boundary defined by $h(i, j)$. Finally, the orthogonalization and normalization can be accomplished, for example, by the classical method involving matrix square roots:

$$\mathbf{W} \leftarrow (\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1/2}\mathbf{W} \tag{2.231}$$

where $\mathbf{W}$ is the matrix of the vectors $\boldsymbol{w}_i$, i.e. $\mathbf{W} = [\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_N]^{\mathrm{T}}$. The original mixing matrix $\mathbf{A}$ can be computed by inverting the whitening process as $\mathbf{A} = (\mathbf{W}\mathbf{V})^{-1}$, where $\mathbf{V}$ is the whitening matrix.

In a yet unpublished work by Jing and Sanei this algorithm has been modified for separation of seizure signals, by (a) iteratively finding the best neighbourhood m and (b) constraining the desired estimated source to be within a specific frequency band and originating from certain brain zones (confirmed clinically). Figure 2.30 illustrates the independent components of a set of EEG signals from an epileptic patient, using the above constrained topographic ICA method. In Figure 2.31 the corresponding topographic maps

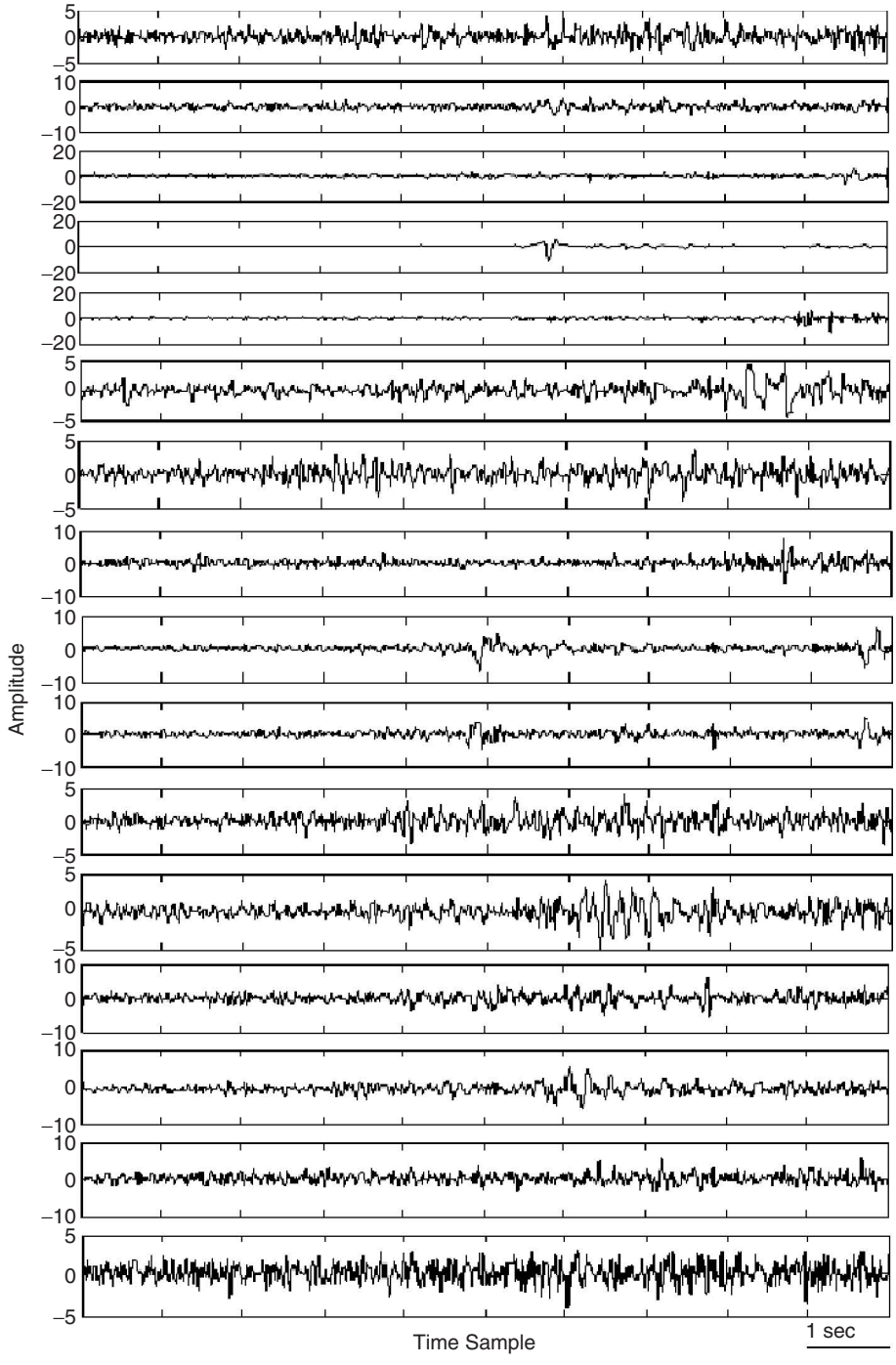

Figure 2.30 The estimated independent components of a set of EEG signals, acquired from 16 electrodes, using constrained topographic ICA. It is seen that similar ICS are gouped together

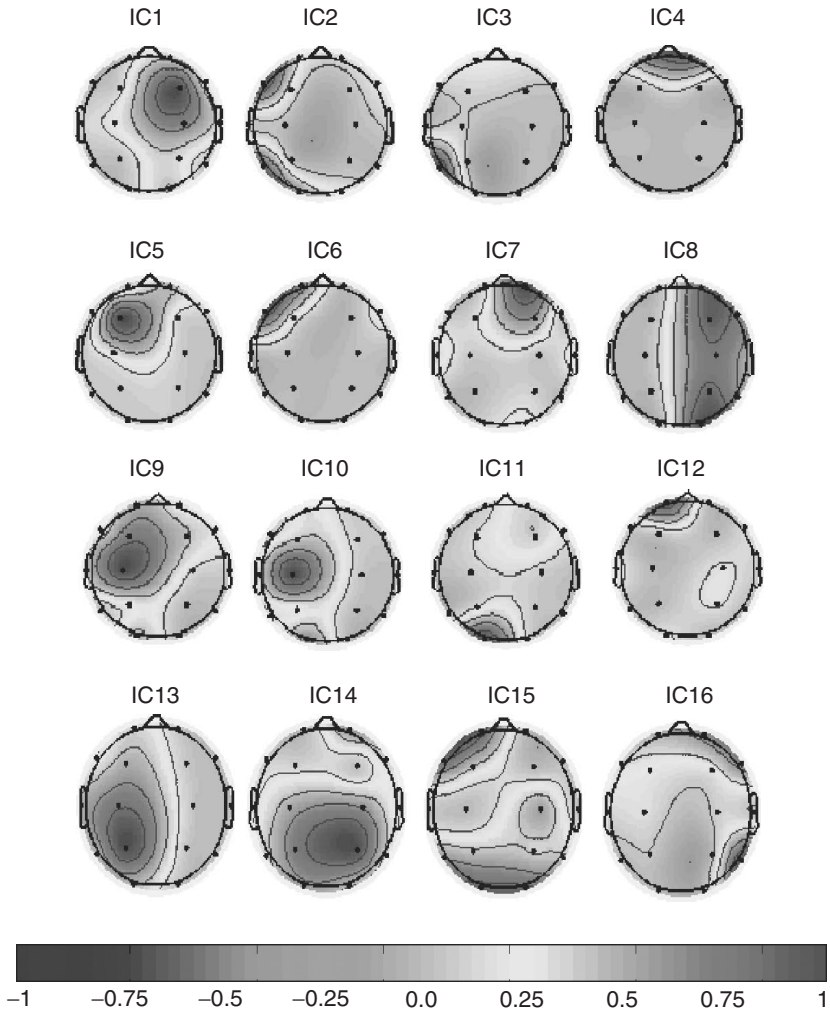

Figure 2.31 The topographic maps, each illustrating an IC. It is clear that the sources are geometrically localized

for all independent components (i.e. backprojection of each IC to the scalp using the inverse of estimated unmixing matrix) are shown. From these figures the sixth IC from the top clearly shows the seizure component. Consequently, the corresponding topograph shows the location of a seizure over the left temporal electrodes.

2.12 Signal Parameter Estimation

In many applications such as modelling, denoising, or prediction, some parameters of signal models or distributions have often to be estimated. For example, in AR modelling the prediction coefficients may be recursively computed using the least mean square

(LMS) algorithm as

$$a_p(k+1) = a_p(k) - \mu e(k)x(k) \tag{2.232}$$

where p is the prediction order, $e(.)$ is the error (residual) signal, and k is the iteration number.

The same algorithm can be used for estimation of the intersection point of a number of spheres. The spheres can be those centred at each EEG electrode and their radius proportional to the inverse of the correlations between each estimated independent component (source) and the scalp EEG signals. The intersection point is then related to the location of the source. This problem will be explained in Chapters 3 and 5.

2.13 Classification Algorithms

In the context of biomedical signal processing, especially with application to EEG signals, the classification of the data in feature spaces is often required. For example, the strength, locations, and latencies of P300 subcomponents may be classified to not only detect whether the subject has Alzheimer's disease but also to determine the stage of the disease. As another example, to detect whether there is a left or right finger movement in the BCI area the time, frequency, and spatial features need to be classified. Also, in blind source separation using the Cohen class Wigner–Ville distribution method, the status of the auto-terms and cross-terms has to be estimated in order to separate the sources. This has to be carried out by means of some clustering techniques such as k-means followed by evaluation of the clusters.

The objective of classification is to draw a boundary between two or more classes and to label them based on their measured features. In a multidimensional feature space this boundary takes the form of a separating hyperplane. The art of the work here is to find the best hyperplane that has a maximum distance from all the classes.

There have been several clustering and classification techniques developed within the last forty years. Among them artificial neural networks (ANNs), linear discriminant analysis (LDA), hidden Markov modelling (HMM), k-means clustering, fuzzy logic, and support vector machines (SVMs) have been very popular. These techniques have been developed and are well explained in the literature [137]. The explanation for all these methods is beyond the objective of this chapter. However, here a summary of an SVM is provided since it has been applied to EEG signals for the removal of the eye-blinking artefact [138], detection of epileptic seizures [139], detection of evoked potentials (EPs), classification of left and right finger movements in BCI [140], and many other issues related to EEGs [141].

Unlike many mathematical problems in which some form of explicit formula based on a number of inputs results in an output, in certain forms of classification of data there will be no model or formula of this kind. In such cases the system should be trained to be able to recognize the inputs. Many classification algorithms do not perform efficiently when:

(a) the number of features is high;
(b) there is a limited time for performing the classification;
(c) there is a nonuniform weighting among the features;

(d) there is a nonlinear map between the inputs and the outputs;
(e) the distribution of the data is not known;
(f) the convergence is not convex (monotonic), so it may fall into a local minimum.

There are two types of machine learning algorithms for classification of data: supervised learning and unsupervised learning. In the former case the target is known and the classifier is trained to minimize a difference between the actual output and the target values. A good example of such classifiers is the multilayered perceptron (MLP). In unsupervised learning, however, the classifier clusters the data into the groups having farthest distances from each other. A popular example for these classifiers is the k-means algorithm.

2.13.1 Support Vector Machines

Among all supervised classifiers, the SVM is the one that performs well in the above situations [142–147]. The concept of SVM was initiated in 1979 by Vapnik [147]. To understand the concept of the SVM consider a binary classification for the simple case of a two-dimensional feature space of linearly separable training samples (Figure 2.32) $S = \{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \ldots, (\boldsymbol{x}_m, y_m)\}$ where $\boldsymbol{x} \in R^d$ is the input vector and $\boldsymbol{y} \in \{-1, 1\}$ is the class label. A discriminating function could be defined as

$$f(\boldsymbol{x}) = \text{sgn}(\langle \boldsymbol{w}, \boldsymbol{x} \rangle + b) = \begin{cases} +1 & \text{if } x \text{ belongs to the first class } \bullet \\ -1 & \text{if } x \text{ belongs to the second class } \circ \end{cases} \tag{2.233}$$

In this formulation $\boldsymbol{w}$ determines the orientation of a discriminant plane (or hyperplane). Clearly, there is an infinite number of possible planes that could correctly classify the training data. One can be as shown in Figure 2.32.

An optimal classifier finds the hyperplane for which the best generalizing hyperplane is equidistant or farthest from each set of points. The set of input vectors is said to be optimally separated by the hyperplane if they are separated without error and the distance between the closest vector and the hyperplane is maximal. In that case there will be only one hyperplane to achieve optimal separation. This can be similar to the one shown in Figure 2.33.

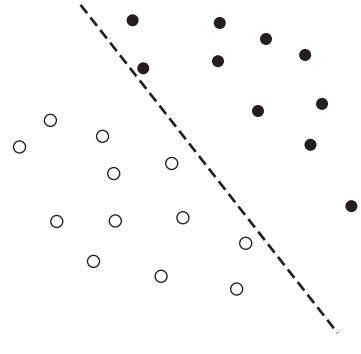

Figure 2.32 A two-dimensional separable dataset and a separating hyperplane

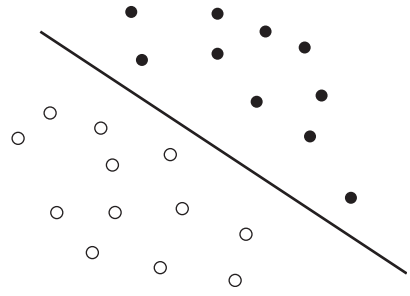

Figure 2.33 An optimal separating hyperplane

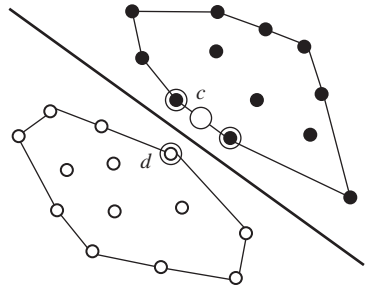

Figure 2.34 Graphical determination of the convex hulls, support vectors, and the separating hyperplane

One way to find the separating hyperplane in a separable case is by constructing the so-called *convex hulls* of each dataset as in Figure 2.34. The encompassed regions are the convex hulls for the datasets. By examining the hulls it is possible then to determine the closest two points lying on the hulls of each class (note that these do not necessarily coincide with actual data points). By constructing a plane that is perpendicular and equivalent to these two points an optimal hyperplane should result and the classifier should be robust in some sense.

Notice in Figure 2.35 that three data points have been identified with circles. These are the only data points required to determine the optimal hyperplane, and are commonly referred to as the *support vectors* (SVs). In places where the data are multi dimensional and the number of points is high the graphical solution to find the hyperplane will no longer be practical. A mathematical solution will then be necessary.

To formulate an SVM, start with the simplest case: linear machines trained on separable data (it will be seen that in the analysis for the general case, nonlinear machines trained on nonseparable data result in a very similar quadratic programming problem). Again label the training data $\{\boldsymbol{x}_i, y_i\}, i = 1, \ldots, m, y_i \in \{-1, 1\}, \boldsymbol{x}_i \in R^d$.

Suppose that a hyperplane separates the positive from the negative examples. The points $\boldsymbol{x}$ which lie on the hyperplane satisfy $\langle \boldsymbol{w},\boldsymbol{x} \rangle + b = 0$, where $\boldsymbol{w}$ is normal to the hyperplane, $|b|/\|\boldsymbol{w}\|_2$ is the perpendicular distance from the hyperplane to the origin, and $\|\boldsymbol{w}\|_2$ is the Euclidean norm of $\boldsymbol{w}$. Define the 'margin' of a separating hyperplane as in Figure 2.36. For

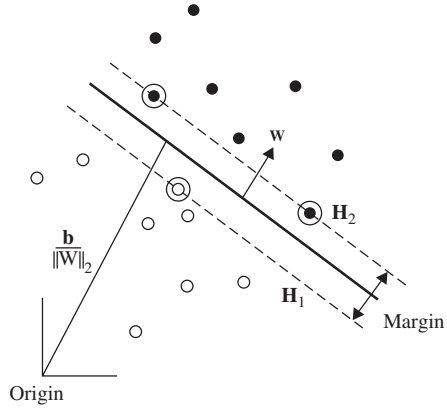

Figure 2.35 Linear separating hyperplane for the separable case, the support vectors are circled

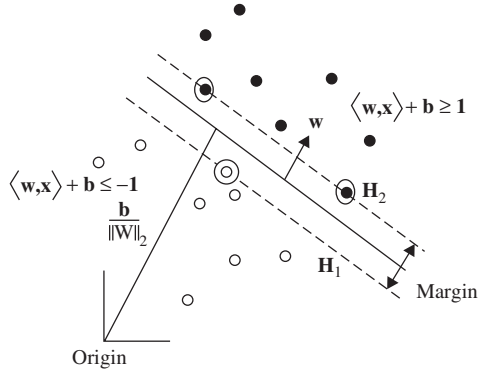

Figure 2.36 The constraints for the SVM

the linearly separable case, the support vector algorithm simply looks for the separating hyperplane with the largest margin. The approach here is to reduce the problem to a convex optimization by minimizing a quadratic function under linear inequality constraints. First it should be noted that in the definition of linear classifiers there is an inherent degree of freedom, in that the function can be scaled arbitrarily. This allows the margins to be set to be equal to unity for simplicity (hyperplanes with a functional margin of unity are sometimes referred to as canonical hyperplanes) and subsequently to minimize the norm of the weight vector. To find the plane farthest from both classes of data, the margins between the supporting canonical hyperplanes for each class are simply maximized. The support planes are pushed apart until they meet the closest data points, which are then deemed to be the support vectors (circled in Figure 2.36). Therefore, since

$$\begin{aligned} \langle \boldsymbol{x}_i, \boldsymbol{w} \rangle + b \geq +1 \quad \text{for } y_i = +1 \\ \langle \boldsymbol{x}_i, \boldsymbol{w} \rangle + b \leq -1 \quad \text{for } y_i = -1 \end{aligned} \tag{2.234}$$

which can be combined into one set of inequalities as $y_i(\langle \boldsymbol{x}_i, \boldsymbol{w} \rangle + b) - 1 \geq 0 \forall i$, the margin between these supporting planes (H_1 and H_2) can be shown to be $\gamma = 2/\|\boldsymbol{w}\|_2$. To maximize this margin, the following is therefore required:

$$\begin{aligned} &\text{Minimize } \langle \boldsymbol{w}, \boldsymbol{w} \rangle \\ &\text{subject to } y_i(\langle \mathbf{x}_i . \boldsymbol{w} \rangle + b) - 1 \geq 0, \quad i = 1, \ldots, m. \end{aligned} \tag{2.235}$$

To perform this constrained optimization problem the constraint can be incorporated into the main cost (risk) function by using Lagrange multipliers. This leads to minimization of an unconstrained empirical risk function (Lagrangian) which consequently results in a set of conditions called the Kuhn–Tucker (KT) conditions.

In order to perform Lagrangian optimization the so-called *primal form* must be constructed:

$$L(\boldsymbol{w}, b, \boldsymbol{\alpha}) = \frac{1}{2}\langle \boldsymbol{w}, \boldsymbol{w} \rangle - \sum_{i=1}^{m} \alpha_i [y_i(\langle \boldsymbol{x}_i, \boldsymbol{w} \rangle + b) - 1] \tag{2.236}$$

where the $\alpha_i, i = 1, \ldots, m$, are the Lagrangian multipliers. Thus, the Lagrangian primal has to be minimized with respect to $\boldsymbol{w}$, b and maximized with respect to $\alpha_i \geq 0$. Constructing the classical Lagrangian dual form facilitates this solution. This is achieved by setting the derivatives of the primal to zero and resubstituting them back into the primal. Hence,

$$\frac{\partial L(\boldsymbol{w}, b, \boldsymbol{\alpha})}{\partial \boldsymbol{w}} = \boldsymbol{w} - \sum_{i=1}^{m} y_i \alpha_i \boldsymbol{x}_i = 0 \tag{2.237}$$

Thus

$$\boldsymbol{w} = \sum_{i=1}^{m} y_i \alpha_i \boldsymbol{x}_i \tag{2.238}$$

and

$$\frac{\partial L(\boldsymbol{w}, b, \boldsymbol{\alpha})}{\partial b} = \sum_{i=1}^{m} y_i \alpha_i = 0 \tag{2.239}$$

By replacing these into the primal form the dual form is obtained as

$$L(\boldsymbol{w}, b, \boldsymbol{\alpha}) = \frac{1}{2} \sum_{j=1}^{m} \sum_{i=1}^{m} y_i y_j \alpha_i \alpha_j \langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle - \sum_{i=1}^{m} y_i y_j \alpha_i \alpha_j \langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle + \sum_{i=1}^{m} \alpha_i \tag{2.240}$$

which is reduced to

$$L(\boldsymbol{w}, b, \boldsymbol{\alpha}) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=1}^{m} y_i y_j \alpha_i \alpha_j \langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle \tag{2.241}$$

considering that $\sum_{i=1}^{m} y_i \alpha_i = 0$ and $\alpha_i \geq 0$.

These equations can be solved mathematically (with the aid of a computer) using quadratic programming (QP) algorithms. There are many algorithms available within numerous publicly viewable websites [148,149].

However, in many practical situations the datasets are not separable (i.e. they have overlaps in the feature space). Therefore the maximum margin classifier described above will no longer be applicable. Obviously, it may be possible to define a complicated nonlinear hyperplane to separate the datasets perfectly but, as seen later this causes the overfitting problem which reduces the robustness of the classifier.

As can be seen in Figure 2.37, the convex hulls overlap and the datasets are no longer linearly separable. The ideal solution where no points are misclassified and no points lie within the margin is no longer feasible. This means that the constraints need to be relaxed to allow for the minimum amount of misclassification. In this case, the points that subsequently fall on the wrong side of the margin are considered to be errors. They are, however, apportioned a lower influence (according to a preset *slack* variable) on the location of the hyperplane and therefore are considered to be support vectors (see Figure 2.38). The classifier obtained in this way is called a *soft margin classifier* (see Figure 2.39).

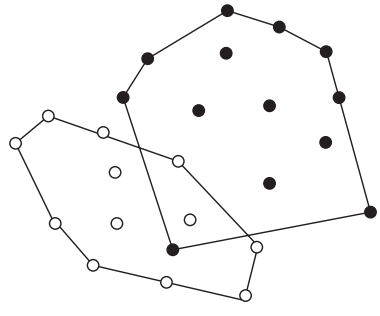

Figure 2.37 Encompassed regions for the nonseparable case

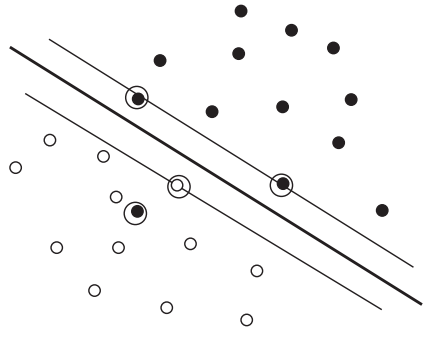

Figure 2.38 Support vectors in a nonseparable case with a linear hyperplane

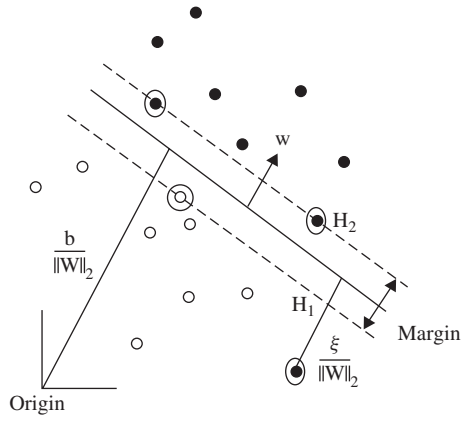

Figure 2.39 Soft margin and the concept of the slack parameter

In order to optimize the soft margin classifier, the margin constraints must be allowed to be violated according to a preset *slack* variable ξ_i in the constraints, which then become

$$\begin{aligned} &\langle \boldsymbol{x}_i, \boldsymbol{w} \rangle + b \geq +1 - \xi_i \quad \text{for } y_i = +1 \\ &\langle \boldsymbol{x}_i, \boldsymbol{w} \rangle + b \leq -1 + \xi_i \quad \text{for } y_i = -1 \\ &\text{and } \xi_i \geq 0 \qquad\qquad \forall i \end{aligned} \tag{2.242}$$

Thus for an error to occur, the corresponding ξ_i must exceed unity, so $\sum_i \xi_i$ is an upper bound on the number of training errors. Hence a natural way to assign an extra cost for errors is to change the objective function to

$$\begin{aligned} &\text{Minimize} \langle \boldsymbol{w}, \boldsymbol{w} \rangle + C \sum_{i=1}^{m} \xi_i \\ &\text{subject to} y_i (\langle \boldsymbol{x}_i, \boldsymbol{w} \rangle + b) \geq 1 - \xi_i \quad \text{and } \xi_i \geq 0 \; i = 1, \ldots, m \end{aligned} \tag{2.243}$$

The primal form will then be

$$L(\boldsymbol{w}, b, \xi, \boldsymbol{\alpha}, \boldsymbol{r}) = \frac{1}{2} \langle \boldsymbol{w}, \boldsymbol{w} \rangle - C \sum_{i=1}^{m} \xi_i - \sum_{i=1}^{m} \alpha_i [y_i (\langle \boldsymbol{w}, \boldsymbol{x}_i \rangle + b) - 1 + \xi_i] - \sum_{i=1}^{m} r_i \xi_i \tag{2.244}$$

Hence,

$$\frac{\partial L(\boldsymbol{w}, b, \xi, \boldsymbol{\alpha}, \boldsymbol{r})}{\partial \boldsymbol{w}} = \boldsymbol{w} - \sum_{i=1}^{m} y_i \alpha_i \boldsymbol{x}_i = 0 \tag{2.245}$$

Thus again

$$\boldsymbol{w} = \sum_{i=1}^{m} y_i \alpha_i \boldsymbol{x}_i \tag{2.246}$$

and

$$\frac{\partial L(\boldsymbol{w}, b, \xi, \boldsymbol{\alpha}, \boldsymbol{r})}{\partial \xi} = C - \alpha_i - r_i = 0 \tag{2.247}$$

so that

$$\alpha_i + r_i = C \tag{2.248}$$

and

$$\frac{\partial L(\boldsymbol{w}, b, \boldsymbol{\alpha})}{\partial b} = \sum_{i=1}^{m} y_i \alpha_i = 0 \tag{2.249}$$

By replacing these into the primal form the dual form is obtained as

$$L(\boldsymbol{w}, b, \xi_i, \boldsymbol{\alpha}, \boldsymbol{r}) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=1}^{m} y_i y_j \alpha_i \alpha_j \langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle \tag{2.250}$$

by again considering that $\sum_{i=1}^{m} y_i \alpha_i = 0$ and $\alpha_i \geq 0$. This is similar to the maximal marginal classifier. The only difference is the new constraints of $\alpha_i + r_i = C$, where $r_i \geq 0$ and hence $0 \leq \alpha_i \leq C$. This implies that the value C sets an upper limit on the Lagrangian optimization variables α_i. This is sometimes referred to as the box constraint. The value of C offers a trade-off between accuracy of data fit and regularization. A small value of C (i.e. <1) significantly limits the influence of error points (or outliers), whereas if C is chosen to be very large (or infinite) then the soft margin approach (as in Figure 2.39) becomes identical to the maximal margin classifier. Therefore in the use of the soft margin classifier, the choice of the value of C will depend heavily on the data. Appropriate selection of C is of great importance and is an area of research. One way to set C is gradually to increase C from max (α_i) for $\forall i$ and then find the value for which the error (outliers, cross-validation, or number of misclassified points) is minimum. Finally, C can be found empirically [150].

There will be no change in formulation of the SVM for the multidimensional cases. Only the dimension of the hyperplane changes depending on the number of feature types.

In many nonseparable cases use of a nonlinear function may help to make the datasets separable. As can be seen in Figure 2.40, the datasets are separable if a nonlinear hyperplane is used. *Kernel mapping* offers an alternative solution by nonlinearly projecting the data into a (usually) higher-dimensional feature space to allow the separation of such cases.

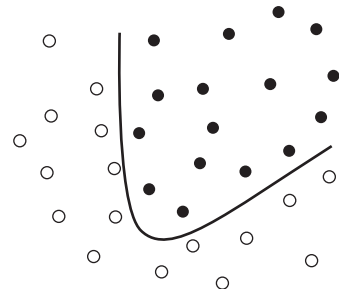

Figure 2.40 Nonlinear discriminant hyperplane

The key success of kernel mapping is that special types of mapping that obey Mercer's theorem, sometimes called reproducing kernel Hilbert spaces (RKHSs) [147], offer an implicit mapping into the feature space:

$$K(\boldsymbol{x}, z) = \langle \varphi(\boldsymbol{x}), \varphi(z) \rangle \tag{2.251}$$

This means that the explicit mapping need not be known or calculated; rather the inner product itself is sufficient to provide the mapping. This simplifies the computational burden dramatically and in combination with the inherent generality of SVMs largely mitigates the dimensionality problem. Further, this means that the input feature inner product can simply be substituted with the appropriate kernel function to obtain the mapping while having no effect on the Lagrangian optimization theory. Hence,

$$L(\boldsymbol{w}, b, \xi_i, \boldsymbol{\alpha}, \boldsymbol{r}) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=1}^{m} y_i y_j \alpha_i \alpha_j K(\boldsymbol{x}_i, \boldsymbol{x}_j) \tag{2.252}$$

The relevant classifier function then becomes

$$f(\boldsymbol{x}) = \operatorname{sgn}\left[\sum_{i=1}^{n\text{SVs}} y_i \alpha_i K(\boldsymbol{x}_i, \boldsymbol{x}_j) + b \right] \tag{2.253}$$

In this way all the benefits of the original linear SVM method are maintained. A highly nonlinear classification function, such as a polynomial or a radial basis function or even a sigmoidal neural network, can be trained using a robust and efficient algorithm that does not suffer from local minima. The use of kernel functions transforms a simple linear classifier into a powerful and general nonlinear classifier [150].

Some examples of popular RKHS functions used in SVMs are given below:

Polynomial $$K(u, v) = (\langle u, v \rangle + c)^d \tag{2.254}$$

Gaussian radial basis function $$K(u, v) = \exp\left(-\frac{\|u - v\|_2^2}{2\sigma^2} \right) \tag{2.255}$$

Exponential radial basis function $$K(u, v) = \exp\left(-\frac{\|u - v\|_2^2}{2\sigma^2}\right) \quad (2.256)$$

Multilayer perceptron $$K(u, v) = \tanh(\rho(\langle u, v\rangle + c)) \quad (2.257)$$

Potentially, it is possible to fit a hyperplane using an appropriate kernel to the data in order to avoid overlapping the sets (or nonseparable cases) and therefore produce a classifier with no error on the training set. This, however, is unlikely to generalize well. More specifically, the main problem with this is that the system may no longer be robust since a testing or new input can be easily misclassified.

Another issue related to the application of SVMs is the cross-validation problem. The distribution of the output of the classifier (without the hard limiter 'sign' in Equation (2.253)) for a number of inputs of the same class may be measured. The probability distributions of the results (which are centred at -1 for class '-1' and at '$+1$' for class '$+1$') are plotted in the same figure. Less overlap between the distributions represents a better performance of the classifier. The choice of the kernel influences the performance of the classifier with respect to the cross validation concept.

SVMs may be slightly modified to enable classification of multiclass data [151]. Moreover, some research has been undertaken to speed up the training step of the SVMs [152].

2.13.2 The k-Means Algorithm

The k-means algorithm [153] is an effective and generally a simple clustering tool that has been widely used for many applications such as in those given in References [126] and [154]. This algorithm divides a set of features (such as points in Figure 2.41) into k clusters.

The algorithm is initialized by setting 'k' to be the assumed number of clusters. Then the centre for each cluster k is identified by selecting k representative data points. The next

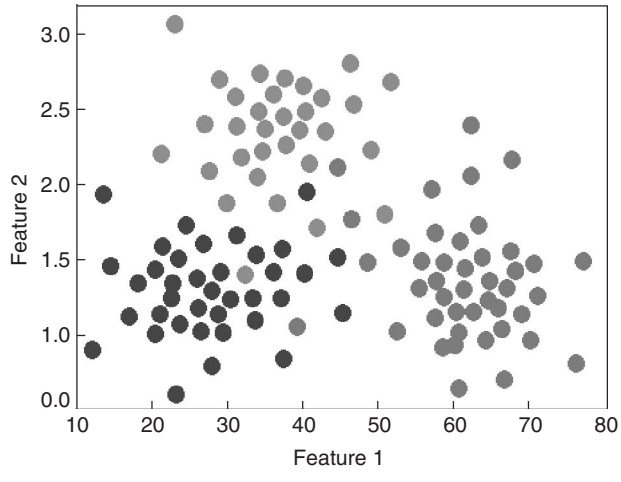

Figure 2.41 A two-dimensional feature space with three clusters, each with a different colour

step in the k-means clustering algorithm after initialization is to assign the remaining data points to the closest cluster centre. Mathematically, this means that each data point needs to be compared with every existing cluster centre and the minimum distance found. This is performed most often in the form of error checking (which will be discussed shortly). However, before this, new cluster centres are calculated. This is essentially the remaining step in k-means clustering: once clusters have been established (i.e. each data point is assigned to its closest cluster centre), the geometric centre of each cluster is recalculated.

The Euclidian distance of each data point within a cluster to its centre can be calculated. It can be repeated for all other clusters, whose resulting sums can themselves be summed together. The final sum is known as the *sum of within-cluster sum of squares*. Consider the within-cluster variation (*sum of squares* for cluster c) error as ε_c:

$$\varepsilon_c = \sum_{i=1}^{n_c} d_i^2 = \sum_{i=1}^{n_c} \|x_i^c - \overline{x}_c\|_2^2 \quad \forall c \tag{2.258}$$

where d_i^2 is the squared Euclidean distance between data point i and its designated cluster centre $\overline{x}_c$, n_c is the total number of data points (features) in cluster c, and x_i^c is an individual data point in cluster c. The cluster centre (mean of data points in cluster c) can be defined as

$$\overline{x}_c = \frac{1}{n_c} \sum_{i=1}^{n_c} x_i^c \tag{2.259}$$

and the total error is

$$E_k = \sum_{c=1}^{k} \varepsilon_c \tag{2.260}$$

The overall k-means algorithm may be summarized as:

1. Initialization
   (a) Define the number of clusters (k).
   (b) Designate a cluster centre (a vector quantity that is of the same dimensionality of the data) for each cluster, typically chosen from the available data points.
2. Assign each remaining data point to the closest cluster centre. That data point is now a member of that cluster.
3. Calculate the new cluster centre (the geometric average of all the members of a certain cluster).
4. Calculate the sum of within-cluster sum of squares. If this value has not significantly changed over a certain number of iterations, stop the iterations. Otherwise, go back to Step 2.

Therefore, an optimum clustering depends on an accurate estimation of the number of clusters. A common problem in k-means partitioning is that if the initial partitions are

not chosen carefully enough the computation will run the chance of converging to a *local* minimum rather than the *global* minimum solution. The initialization step is therefore very important.

One way to combat this problem is to run the algorithm several times with different initializations. If the results converge to the same partition then it is likely that a global minimum has been reached. This, however, has the drawback of being very time consuming and computationally expensive. Another solution is to change the number of partitions (i.e. number of clusters) dynamically as the iterations progress. The ISODATA (iterative self-organizing data analysis technique algorithm) is an improvement on the original k-means algorithm that does exactly this. ISODATA introduces a number of additional parameters that allow it to progressively check within- and between-cluster similarities so that the clusters can dynamically split and merge.

Another approach for solving this problem is to use so-called gap statistics [155]. In this approach the number of clusters are iteratively estimated. The steps of this algorithm are:

1. For a varying number of clusters $k = 1, 2, \ldots, K$, compute the error measurement E_k using Equation (2.238).
2. Generate a number B of reference datasets. Cluster each one with the k-means algorithm and compute the dispersion measures, $\breve{E}_{kb}$, $b = 1, 2, \ldots, B$. The gap statistics are then estimated using

$$G_k = \frac{1}{B}\sum_{b=1}^{B}\log(\breve{E}_{kb}) - \log(E_k) \tag{2.261}$$

   where the dispersion measure $\breve{E}_{kb}$ is the E_k of the reference dataset B.
3. To account for the sample error in approximating an ensemble average with B reference distributions, the standard deviation is computed as

$$S_k = \left[\frac{1}{B}\sum_{b=1}^{B}[\log(\breve{E}_{kb}) - \overline{E}_b]^2\right]^{1/2} \tag{2.262}$$

   where

$$\overline{E}_b = \frac{1}{B}\sum_{b=1}^{B}\log(\breve{E}_{kb}) \tag{2.263}$$

4. By defining $\breve{S}_k = S_k\,(1 + 1/B)^{1/2}$, the number of clusters is estimated as the smallest k such that $G_k \geq G_{k+1} - \breve{S}_{k+1}$.
5. With the number of clusters identified, use the $k-$means algorithm to partition the feature space into k subsets (clusters).

The above clustering method has several advantages since it can estimate the number of clusters within the feature space. It is also a multiclass clustering system and unlike SVM can provide the boundary between the clusters.

2.14 Matching Pursuits

EEG signals are often combinations of rhythmical and transient features. These features may best be explored in the time–frequency (TF) domain. The matching pursuit (MP) algorithm [156] is often used instead of popular TF approaches such as the STFT and the WT because of its higher temporal–spatial resolution in the TF space [4], local adaptivity to transient structure, and its computational compatibility to the EEG data structure despite its computational complexity.

Here the formulation of MP is given for continuous-time signals and the dictionary. A similar presentation can be given for discrete signals simply by changing t to n. The definition of matching pursuits is straightforward; given a set of functions $D(t) = \{g_1(t), g_2(t), \ldots, g_K(t)\}$, called the dictionary of MP, where $\|g_i\| = 1$, m signals from D can be found to best approximate a signal $f(t)$. The approximation error is obtained as

$$\varepsilon = \left\| f(t) - \sum_{i=1}^{m} w_i g_{\gamma_i}(t) \right\|_2 \tag{2.264}$$

where $||\cdot||_2$ denotes the Euclidean norm, w_i are the weights, and $\{\gamma_i\}_{i=1,\ldots,m}$ represents the indices of the selected functions g_{γ_i}. The MP algorithm provides a suboptimal iterative solution for the above expansion. The MP algorithm performs the following steps. In the first step the waveform $g_{\gamma_0}(t)$ that best matches the signal $f(t)$ is chosen. Then, in the consecutive steps, the waveform g_{γ_i} is matched to the signal $\tilde{f}_k$, which is the residual from the previous iteration:

$$\begin{aligned} \tilde{f}_0(t) &= f(t) \\ \tilde{f}_{k+1}(t) &= \tilde{f}_k(t) - \langle \tilde{f}_k(t), g_{\gamma_k}(t) \rangle g_{\gamma_k}(t) \quad \text{for } k = 1, \ldots, m \\ g_{\gamma_n}(t) &= \arg \max_{g_{\gamma_i} \in D} |\langle \tilde{f}_n(t), g_{\gamma_i}(t) \rangle| \end{aligned} \tag{2.265}$$

where $\langle f(t), g(t) \rangle$ represents the cross correlation of $f(t)$ and $g(t)$. The orthogonality of $\tilde{f}_{k+1}(t)$ and $g_{\gamma_k}(t)$ at each step implies energy conservation, i.e.

$$\|f(t)\|^2 = \sum_{k=0}^{m-1} |\langle \tilde{f}_k(t), g_{\gamma_k}(t) \rangle|^2 + \|\tilde{f}_m(t)\|^2 \tag{2.266}$$

where $\tilde{f}_{k+1}(t)$ in the above procedure converges to $f(t)$ if the complete dictionary ($m = D$) is used. In that case

$$f(t) = \sum_{k=0}^{\infty} \langle \tilde{f}_k(t), g_{\gamma_k}(t) \rangle g_{\gamma_n}(t) \tag{2.267}$$

From this equation it is possible to derive a TF distribution of the signal's energy $F(t, \omega)$ that is free of cross-terms (i.e. the sum is 100 % correlated with the data $f(t)$) by adding

Wigner distributions of selected functions

$$F(t,\omega)=\sum_{n=0}^{m}|\langle\tilde{f}_n(t),g_{\gamma_n}(t)\rangle|^2 G_{\gamma_n}(t,\omega)\quad\forall t,\ \forall\omega \tag{2.268}$$

where $G_{\gamma_k}(t,\omega)$ is the Wigner time–frequency distribution of the kth selected function. A combination of MP and the WT has also been proposed [157].

Gabor functions (sine-modulated Gaussian functions), are often used as the dictionary functions and provide optimal joint TF localization. A real Gabor function may be expressed as [158]

$$g_\gamma(t)=K(\gamma)\mathrm{e}^{-\pi[(t-\tau)/\sigma]^2}\sin\left[\frac{\omega}{N}(t-\tau)+\theta\right] \tag{2.269}$$

where $K(\gamma)$ is the normalization factor, i.e. it makes $\|g_\gamma(t)\|=1$, N is the length of the signals, and $\gamma=\{\tau,\omega,\sigma,\theta\}$ are the parameters of the functions (time–frequency atoms) that form the dictionary. In the original MP algorithm proposed by Mallat and Zhang [156] the parameters of the dictionary are selected from dyadic sequences of integers and their sampling interval is governed by another integer parameter (octave) j. The parameter σ, the width of the signal in the time domain, is set to $2^j, 0\le j\le L$ (signal size $N=2^L$). The time–frequency coordinates of τ and ω are sampled for each octave j with interval $\sigma=2^j$. In the case of oversampling by l it is sampled with interval 2^{j-l}.

Analysing sleep EEG data by means of the MP algorithm has been attempted [4]. In this approach a statistical bias of the decomposition, resulting from the structure of the applied dictionary, has been considered. In the proposed stochastic dictionaries the parameters of the waveforms within the dictionary are randomized before each decomposition. The MP algorithm was modified for this purpose and tuned for maximum time–frequency resolution.

The above method was also applied to analysis of single-trial event-related potentials, in particular ERD (event-related desynchronization)/ERS related to a voluntary movement. The main idea was based upon averaging energy distributions of single EEG trials in the time–frequency plane. Consistent results, essential for the brain–computer interfacing (BCI) problem, have been reported.

Several other applications of MP for analysis of the EEG signals have been reported [50,157,159]. It is a powerful method for detection of the features localized in the time–frequency domain [160] and transient signals [159]. This includes ERP detection, detection and classification of movement-related potentials, seizure detection [50], and identification of gamma bursts.

2.15 Summary and Conclusions

In this chapter several concepts in the processing of EEG signals, including signal modelling, signal segmentation, signal transforms, multivariate modelling and direct transfer functions, chaos and dynamic analysis, independent component analysis and blind source separation, classification and clustering, and matching pursuits, have been reviewed. It is very difficult to bring all the methods and algorithms used in the processing of EEG

signals into a single chapter or even a book. In reality, algorithms are developed generally based on the specific requirements of certain applications. Therefore, this chapter is not expected to cover all the aspects of digital signal processing applied to EEGs.

However, to the best knowledge of the authors, the sections included cover the important fundamental signal processing techniques required by the EEG research community. This chapter also provides certain key references for further reading in the field of signal processing for the analysis of the EEG signals. With this information the readers will be better able to digest the contents of the later chapters of this book.

References

[1] Lebedev, M. A., and Nicolelis, M. A., 'Brain–machine interfaces: past, present and future', *Trends. Neurosci.*, **29**, 2006, 536–546.

[2] Lopes da Silva, F., 'Functional localization of brain sources using EEG and/or MEG data: volume conductor and source models', *J. Magnetic Resonance Imaging*, **22**(10), 2004, 1533–1538.

[3] Malinovska, U., Durka, P. J., Blinowska, K. J., Szelenberger, W., and Wakarow, A., 'Micro- and macrostructure of sleep EEG, a universal, adaptive time–frequency parametrization', *IEEE Engng. in Medicine and Biology Mag.*, **25**(4), July/August 2006, pp. 26–31.

[4] Durka, P. J., Dobieslaw, I., and Blinowska, K. J., 'Stochastic time–frequency dictionaries for matching pursuit', *IEEE Trans. Signal Process.*, **49**(3), March 2001.

[5] Strang, G., *Linear Algebra and Its Applications*, 4th edn., Belmont, California, USA, 2006.

[6] Benedek, G., and Villars, F., *Physics, with Illustrative Examples from Medicine and Biology*, Springer-Verlag, New York, 2000.

[7] Hille, B., *Ionic Channels of Excitable Membranes*, Sinauer, Sunderland, Massachussetts, 1992.

[8] Hodgkin, A., and Huxley, A., 'A quantitative description of membrane current and its application to conduction and excitation in nerve', *J. Physiol. (Lond.)*, **117**, 1952, 500–544.

[9] Simulator for Neural Networks and Action Potentials (SNNAP) Tutorial, The University of Texas-Houston Medical School, 2003, `http://snnap.uth.tmc.edu`.

[10] Ziv, I., Baxter, D. A., and Byrne, J. H., 'Simulator for neural networks and action potentials: description and application', *J. Neurophysiol.*, **71**, 1994, 294–308.

[11] Gerstner, W., and Kistler, W. M., *Spiking Neuron Models*, 1st edn. Cambridge University Press, Cambridge, August 2002.

[12] Akaike, H., 'A new look at statistical model order identification', *IEEE Trans. Autom. Control*, **19**, 1974, 716–723.

[13] Kay, S. M., *Modern Spectral Estimation: Theory and Application*, Prentice-Hall, Englewood Cliffs, New Jersey, 1988.

[14] Guegen, C., and Scharf, L., 'Exact maximum likelihood identification of ARMA models: a signal processing perspective', *Signal Processing Theory Applications*, Eds M. Kunt and F. de Coulon, North-Holland, Lausanne, Switzerland, 1980, pp. 759–769.

[15] Akay, M., *Biomedical Signal Processing*, Academic Press, New York, 2001.

[16] Kay, S. M., *Modern Spectral Estimation, Theory and Application*, Prentice-Hall, Englewood Cliffs, New Jersey, 1988.

[17] Akaike, H., 'A New Look at statistical model identification', *IEEE Trans. Autom. Control*, **19**, 1974, 716–723.

[18] Durbin, J., 'Efficient estimation of parameters in moving average models', *Biometrika*, **46**, 1959, 306–316.

[19] Trench, W. F., 'An algorithm for the inversion of finite Toelpitz matrices', *J. Soc. Ind. Appl. Math.*, **12**, 1964, 515–522.

[20] Morf, M., Vieria, A., Lee, D., and Kailath, T., 'Recursive multichannel maximum entropy spectral estimation', *IEEE Trans. Geosci. Electronics*, **16**, 1978, 85–94.

[21] Spreckelesen, M., and Bromm, B., 'Estimation of single-evoked cerebral potentials by means of parametric modelling and Kalman filtering', *IEEE Trans. Biomed. Engng.*, **33**, 1988, 691–700.

[22] Demiralp, T., and Ademoglu, A., 'Modeling of evoked potentials as decaying sinusoidal oscillations by Prony's method', in Proceeding of IEEE EMBS, Paris, 1992.

[23] De Prony, B. G. R., 'Essai experimental et analytique: sur les lois de la dilatabilite de fluids elastiques et sur celles de la force expansive de la vapeur de l'eau et de la vapeur de l'alkool, a differentes temperatures', *J. E. Polytech.*, **1**(2), 1795, 24–76.

[24] Marple, S. L., *Digital Spectral Analysis with Applications*, Prentice-Hall, Englewood Cliffs, New Jersey, 1987.

[25] Lawson, C. L., and Hanson, R. J., *Solving Least Squares Problems*, Prentice-Hall, Englewood Cliffs, New Jersey, 1974.

[26] Demiralp, T., and Ademoglu, A., 'Modelling of evoked potentials as decaying sinusoidal oscillations by the Prony method', in Proceeding of IEEE EMBS, Paris, 1992.

[27] Bouattoura, D., Gaillard, P., Villon, P., and Langevin, F., 'Multilead evoked potentials modelling based on the Prony's method', Proceeding of IEEE TECON on *Digital Signal Processing Applications*, 1996, pp. 565–568.

[28] Dacorogna, M., Muller, U., Olsen, R. B., and Pictet, O., 'Modelling short-term volatility with GARCH and HARCH models', in *Nonlinear Modelling of High Frequency Financial Time Series, Econometrics*, Eds L. Christian Dunis and B. Zhou, John Wiley & Sons, Ltd, Chichester, 1998.

[29] McLeod, A. J., and Li, W. K., 'Diagnostics checking ARMA time series models using squared residual autocorrelations', *J. Time Series Analysis*, **4**, 1983, 269–273.

[30] Brock, W. A., Hsieh, D. A., and LeBaron, B., *Nonlinear Dynamics, Chaos, and Instability: Statistical Theory and Economic Evidence*, MIT press, Cambridge, Massachussetts, 1992.

[31] Hsieh, D. A., 'Testing for nonlinear dependence in daily foreign exchange rates', *J. Business*, **62**, 1989, 339–368.

[32] Engle, R. F., Lilien, D. M., and Robin, R. P., 'Estimating time-varying risk premia in the term structure: the ARCH-M model', *Econometrica*, **55**, 1987, 391–407.

[33] Nelson, D. B., 'Stationarity and persistence in the GARCH(1,1) model', *J. Econometrics*, **45**, 1990, 7–35.

[34] Glosten, L. R., Jagannathan, R., and Runkle, D., 'On the relation between the expected value and the volatility of the nominal excess return on stocks', *J. Finance*, **2**, 1995, 225–251.

[35] Zakoian, J. M., 'Threshold heteroskedastic models', *J. Economic Dynamics and Control*, **18**, 1994, 931–955.

[36] Ding, Z., Engle, R. F., and Granger, C. W. J., 'A long memory property of stock market returns and a new model', *J. Empirical Finance*, **1**, 1993, 83–106.

[37] Sentana, E., 'Quadratic ARCH models: a potential reinterpretation of ARCH models as second-order Taylor approximations', London School of Economics, 1991.

[38] Galka, A., Yamashita, O., and Ozaki, T., 'GARCH modelling of covariance in dynamical estimation of inverse solutions', *Phys. Lett. A*, **333**, 2004, 261–268.

[39] Tikhonov, A. N, editor, *Ill-Posed Problems in Natural Sciences*, The Netherlands, 1992, 155–165.

[40] Pascual-Marqui, R. D., Esslen, M., Kochi, K., and Lehmann, D., 'Functional imaging with low resolution brain electromagnetic tomography (LORETA): a review', *Methods and Findings in Expl. Clin. Pharmacol.*, **24C**, 2002, 91–95.

[41] Lagerlund, T. D., Sharbrough, F. W., and Busacker, N. E., 'Spatial filtering of multichannel electroencephalographic recordings through principal component analysis by singular value decomposition', *J. Clin. Neurophysiol.*, **14**(1), 1997, 73–82.

[42] Da Silva, F. H., Hoeks, A., Smits, H., and Zetterberg, L. H., 'Model of brain rhythmic activity: the alpha-rhythm of the thalamus', *Kybernetic*, **15**, 1974, 27–37.

[43] Lopes da Silva, F. H., van Rotterdam, A., Barts, P., van Heusden, E., Burr, W., 'Models of neuronal populations: the basic mechanisms of rhythmicity', *Prog. Brain Res.*, **45**, 281–308, 1976.

[44] Wilson, H. R., and Cowan, J. D., 'Excitatory and inhibitory interaction in localized populations of model neurons', *J. Biophys.*, **12**, 1972, 1–23.

[45] Zetterberg, L. H., 'Stochastic activity in a population of neurons–a system analysis approach', Report of the Institute of Medical Physics, TNO, Utrecht, Vol. 1, 1973, 53.

[46] Hyvarinen, A., Kahunen, J., and Oja, E., *Independent Component Analysis*, John Wiley & Sons, Ltd, Chichester, 2001.

[47] Cover, T. M., and Thomas, J. A., *Elements of Information Theory*, John Wiley & Sons, Ltd, Chichester, 2001.

[48] Bro, R., Multi-way analysis in the food industry: models, algorithms, and applications', PhD thesis, University of Amsterdam (NL) and Royal Veterinary and Agricultural University, 1998; MATLAB toolbox available on-line at http://www.models.kvl.dk/users/rasmus/.

[49] Harris, F. J., *Multirate Signal Processing for Communication Systems*, Prentice-Hall, Englewood Cliffs, New Jersey, 2004.

[50] Franaszczuk, P. J., Bergey, G. K., and Durka, P. J., 'Time–frequency analysis of mesial temporal lobe seizures using the matching pursuit algorithm', *Soc. Neurosci. Abstr.*, **22**, 1996, 184.

[51] Murenzi, R., Combes, J. M., Grossman, A., and Tchmitchian, P. (Eds), *Wavelets*, Springer-Verlag, Berlin, Heidelberg, New York, 1988.

[52] Vaidyanathan, P. P., *Multirate Systems and Filter Banks*, Prentice-Hall, Englewood Cliffs, New Jersey, 1993

[53] Holschneider, M., Kronland-Martinet, R., Morlet, J., and Tchamitchian, Ph., 'A real-time algorithm for signal analysis with the help of the wavelet transform', in *Wavelets: Time–Frequency Methods and Phase Space*, Eds J. M. Combes, A. Grossman, and Ph. Tchamitchian, Springer-Verlag, Berlin, 1989, pp. 286–297.

[54] Chui, C. K., *An Introduction to Wavelets*, Academic Press, New York, 1992.

[55] Stein, E. M., 'On the functions of Littlewood–Paley, Lusin and Marcinkiewicz', *Trans. Am. Math. Soc.*, **88**, 1958, 430–466.

[56] Vetterli, M., and J. Kovačevic, *Wavelets and Subband Coding*, Prentice-Hall, Englewood Cliffs, New Jersey, 1995.

[57] Glassman, E. L., 'A wavelet-like filter based on neuron action potentials for analysis of human scalp electroencephalographs', *IEEE Trans. Biomed. Engng.*, **52**(11), 2005, 1851–1862.

[58] Gerloff, G., Richard, J., Hadley, J., Schulman, A. E., Honda, M., and Hallett, M., 'Functional coupling and regional activation of human cortical motor areas during simple, internally paced and externally paced finger movements', *Brain*, **121**(8), 1998, 1513–1531.

[59] Sharott, A., Magill, P. J., Bolam, J. P., and Brown, P., 'Directional analysis of coherent oscillatory field potentials in cerebral cortex and basal ganglia of the rat', *J. Physiol.*, **562**(3), 2005, 951–963.

[60] C. W. J. Granger, 'Investigating causal relations in econometric models and cross-spectral methods', *Econometrica*, **37**, 1969, 424–438.

[61] Bernosconi, C., and P. König, 'On the directionality of cortical interactions studied by spectral analysis of electrophysiological recordings', *Biol. Cybern.*, **81**(3), 1999, 199–210.

[62] Kaminski, M., Ding, M., Truccolo, W., and Bressler, S., 'Evaluating causal relations in neural systems: Granger causality, directed transfer function, and statistical assessment of significance', *Biol. Cybern.*, **85**, 2001, 145–157.

[63] Kaminski, M., and Blinowska, K., 'A new method of the description of information flow in the brain structures', *Biol. Cybern.*, **65**, 1991, 203–210.

[64] Jing, H., and Takigawa, M., 'Observation of EEG coherence after repetitive transcranial magnetic stimulation', *Clin. Neurophysiol.*, **111**, 2000, 1620–1631.

[65] Kuś, R., Kaminski, M., and Blinowska, K., 'Determination of EEG activity propagation: pair-wise versus multichannel estimate', *IEEE Trans. Biomed. Engng.*, **51**(9), 2004, 1501–1510.

[66] Ginter Jr, J., Kaminski, M., Blinowska, K., and Durka, P., 'Phase and amplitude analysis in time–frequency–space; application to voluntary finger movement', *J. Neurosci. Meth.*, **110**, 2001, 113–124.

[67] Ding, M., Bressler, S. L., Yang, W., and Liang, H., 'Short-window spectral analysis of cortical event-related potentials by adaptive multivariate autoregressive modelling: data preprocessing, model validation, and variability assessment', *Biol. Cybern.*, **83**, 2000, 35–45.

[68] Peitgen, H.-O., Lurgens, H., and Saupe, D., *Chaos and Fractals*, Springer-Verlag, New York, 1992.

[69] Grassberger, P., and Procaccia, I., 'Characterization of strange attractors', *Phys. Rev. Lett.*, **50**, 1983, 346–349.

[70] Kaplan, L., and Yorke, J. A., 'Chaotic behaviour of multidimensional difference equations', in *Functional Differential Equations and Approximation of Fixed Points*, Ed. H.-O. Peitgen and H.-O. Walther, Springer-Verlag, New York, 1979.

[71] Russell, D. A., Hanson, J. D., and Ott, E., 'Dimension of strange attractors', *Phys. Rev. Lett.* **45**, 1980, 1175–1179.

[72] Wolf, A., Swift, J. B., Swinny, H. L., and Vastano, J. A., 'Determining Lyapunov exponents from a time series', *Physical*, **16D**, 1985, 285–317.

[73] Albano, A. M., Muench, J., Schwartz, C., Mees, A. I., and Rapp, P. E., 'Singular value decomposition and the Grassberger–Procaccia algorithm', *Phys. Rev. A*, **38**, 1988, 3017.

[74] King, G. P., Jones, R., and Broomhead, D. S., Phase portraits from a time series: a singular system approach', *Nucl. Phys. B*, **2**, 1987, 379.

[75] Fraser, A. M., and Swinney, H., 'Independent coordinates for strange attractors from mutual information', *Phys. Rev. A*, **33**, 1986, 1134–1139.

[76] Pincus, S. M., 'Approximate entropy as a measure of system complexity', *Proc. Natl Acad. Sci., USA*, **88**, 1991, 2297–2301.

[77] Fusheng, Y., Bo, H., and Qingyu, T., 'Approximate entropy and its application in biosignal analysis', in *Nonlinear Biomedical Signal Processing*, Vol. II, Ed. Metin Akay, IEEE Press, New York, 2001, pp. 72–91.

[78] Widrow, B., Glover, J. R., McCool Jr, J., Kaunitz, J. M., Williams, J., Hearn, C. S., and Zeidler, R. H., 'Adaptive noise cancelling principles and applications', *Proc. IEEE*, **63**(12), 1975, 1692–1716.

[79] Satorius, E. H., and Shensa, M. J., 'Recursive lattice filters: a brief overview', in Proceeding of 19th IEEE Conference on *Decision Control*', 1980, pp. 955–959.

[80] Lee, D., Morf, M., and Friedlander, B., 'Recursive square-root ladder estimation algorithms, *IEEE Trans. Accous., Speech, Signal Proces.*, **29**, 1981, 627–641.

[81] Lawsen, C. L., and Hansen, R. J., *Solving the Least-Squares Problem*, Prentice-Hall, Englewood Cliffs, New Jersey, 1974.

[82] Proakis, J. G., Rader, C. M., Ling, F., Moonen, M., Proudler, I. K., and Nikias, C. L., *Algorithms for Statistical Signal Processing*, Prentice-Hall, Englewood Cliffs, New Jersey, 2001.

[83] Cherry, C. E., 'Some experiments in the recognition of speech, with one and two ears', *J. Acoust. Soc. Am.*, **25**, 1953, 975–979.

[84] Herault, J., and Jutten, C., 'Space or time adaptive signal processing by neural models', Proceeding of the American Institute of Physics (AIP) Conference on *Neural Networks for Computing*, 1986 pp. 206–211.

[85] Comon, P., 'Independent component analysis: a new concept', *Signal Process.*, **36**, 1994, 287–314.

[86] Jung, T. P., Makeig, S., Westereld, M., Townsend, J., Courchesne, E., and Sejnowski, T. J., 'Analyzing and visualizing single-trial event-related potentials', *Advances in Neural Information Processing Systems* Vol. 8, MIT Press, Cambridge, Massachussetts, 1999.

[87] Jung, T. P., Makeig, S., Humphries, C., Lee, T. W., McKeown, M. J., Iragui, V., and Sejnowski, T. J., 'Removing electroencephalographic artefacts by blind source separation', *Psychophysiology*, **37**, 2000, 163–178.

[88] Shoker, L., Sanei, S., and Chambers, J., 'Artifact removal from electroencephalograms using a hybrid BSS-SVM algorithm', *IEEE Signal Process. Lett.*, **12**(10), October 2005.

[89] Shoker, L., Sanei, S., and Chambers, J., 'A hybrid algorithm for the removal of eye blinking artefacts from electroencephalograms', in Proceedings of the IEEE Statistical Signal Processing Workshop, France, SSP2005, February 2005.

[90] Corsini, J., Shoker, L., Sanei, S., and Alarcon, G., 'Epileptic seizure prediction from scalp EEG incorporating BSS', *IEEE Trans. Biomed. Engng.*, **53**(5), 790–799, May 2006.

[91] Latif, M. A., Sanei, S., and Chambers, J., 'Localization of abnormal EEG sources blind source separation partially constrained by the locations of known sources', *IEEE Signal Process. Lett.*, **13**(3), March 2006, 117–120.

[92] Spyrou, L., Jing, M., Sanei, S., and Sumich, A., 'Separation and localisation of P300 sources and the subcomponents using constrained blind source separation', *EURASIP J. Adv. Signal Process.*, Article ID 82912, November, 2006, 10 pages.

[93] Sanei, S., 'Texture segmentation using semi-supervised support vector machines', *Int. J. Computat. Intell. Applics*, **4**(2), 2004, 131–142.

[94] Tang, A. C., Sutherland, T., and Wang, Y., 'Contrasting single trial ERPs between experimental manipulations: improving differentiability by blind source separation', *Neuroimage*, **29**, 2006, 335–346.

[95] Makeig, S., Bell, A. J., Jung, T., and Sejnowski, T. J., 'Independent component analysis of electroencephalographic data', *Advances in Neural Information Processing Systems* Vol. 8, MIT Press, Cambridge, Massachussetts, 1996, pp. 145–151.

[96] Makeig, S., Jung, T. P., Bell, A. J., Ghahremani, D., and Sejnowski, T. J., 'Blind separation of auditory event-related brain responses into independent components', *Proc. Natl Academy of Sci., USA*, **94**, September, 1997 10979–10984.

[97] P. Földiák, and Young, M., 'Sparse coding in the primate cortex', in *The Handbook of Brain Theory and Neural Networks*, Eds Michael A. Arbib, 2nd edn., MIT Press, Cambridge, Massachussetts, 1995, pp. 895–898.

[98] Einhauser, W., Kayser, C., Konig, P., and Kording, K. P., 'Learning the invariance properties of complex cells from their responses to natural stimuli', *Eur. J. Neurosci.*, **15**, 2002, 475–486.

[99] Jutten, C., and Herault, J., 'Blind separation of sources, Part I: an adaptive algorithm based on neuromimetic architecture', *Signal Processing*, **24**, 1991, 1–10.

[100] Linsker, R., 'An application of the principle of maximum information preservation to linear systems', *Advances in Neural Information Processing Systems*, Morgan Kaufmann, Palo Alto, California, 1989, pp. 186–194.

[101] Bell, A. J., and Sejnowski, T. J., 'An information-maximization approach to blind separation, and blind deconvolution', *Neural Computation*, **7**(6), 1995, 1129–1159.

[102] Amari, S., Cichocki, A., and Yang, H. H., 'A new learning algorithm for blind signal separation', in *Advances in Neural Information Processing Systems*, Vol. 8, MIT Press, Cambridge, Massachussetts, 1996.

[103] Hyvarinen, A., and Oja, E., 'A fast-fixed point algorithm for independent component analysis', *Neural Computation*, **9**(7), 1997, 1483–1492.

[104] Parra, L., and Spence, C., 'Convolutive blind separation of non-stationary sources', *IEEE Trans. Speech Audio Process.*, **8**(3), 2000, 320–327.

[105] Cardoso, J.-F., 'Infomax and maximum likelihood for blind source separation', *IEEE Signal Process. Lett.*, **4**(4), 1997, 112–114.

[106] Cardoso, J., 'Source separation using higher order moments', in Proceedings of the International Conference on *Acoustics, Speech and Signal Processing* (*ICASSP*), Glasgow, Scotland, 1989, pp. 2109–2112.

[107] Belouchrani, A., Abed-Meraim, K., J.-F. Cardoso, and Moulines, E., 'A blind source separation technique using second order statistics', *IEEE Trans. Signal Process.*, **45**(2), 1997, 434–444.

[108] Jing, M., and Sanei, S., 'Scanner artifact removal in simultaneous EEG–fMRI for epileptic seizure prediction', in *IEEE 18th International Conference on Pattern Recognition, (ICPR)*, Vol. 3, Hong Kong, August 2006, pp. 722–725.

[109] Mathis, H., and Douglas, S. C., 'On the existence of universal nonlinearities for blind source separation', *IEEE Trans. Signal Process.*, **50**, May 2004, 1007–1016.

[110] Cheong, C., Sanei, S., and Chambers, J., 'A filtering approach to underdetermined BSS with application to temporomandibular disorders', in Proceedings of IEEE ICASSP, France, 2006.

[111] Belouchrani, A., Abed-Mariam, K., Amin, M. G., and Zoubir, A. M., 'Blind source separation of nonstationary signals', *IEEE Signal Process. Lett.*, **11**(7), 2004, 605–608.

[112] Cirillo, L., and Zoubir, A., 'On blind separation of nonstationary signals', in Proceedings of the 8th Symposium on *Signal Processing and Its Applications* (*ISSPA*), Sydney, Australia, 2005.

[113] Cohen, L., *Time–Frequency Analysis*, Prentice-Hall, Englewood Cliffs, New Jersey, 1995.

[114] Platt, C., and Fagin, F., 'Networks for the separation of sources that are superimposed and delayed', *Advances in Neural Information Processing Systems.*, Vol. 4, Morgan Kaufmann, San Mateo, 1992, pp. 730–737.

[115] Herault, J., and Jutten, C., 'Space or time adaptive signal processing by neural models', *Proc. AIP Conf., Neural Network for Computing*, American Institute of Physics, New York, 1986, pp. 206–211.

[116] K. Torkkola, 'Blind separation of delayed sources based on information maximization', in Proceeding of ICASSP, Atlanta, Georgia, 1996, pp. 3509–3512.

[117] Torkkola, K., 'Blind separation of convolved sources based on information maximization', in Proceeding of the IEEE Workshop on *Neural Networks and Signal Processing (NNSP)*, Kyoto, Japan, 1996, 315–323.

[118] Wang, W., Sanei, S., and Chambers, J. A., 'Penalty function based joint diagonalization approach for convolutive blind separation of nonstationary sources', *IEEE Trans. Signal Process.*, **53**(5), 2005, 1654–1669.

[119] Yeredor, A., 'Blind source separation with pure delay mixtures', in Proceedings of Conference on *Independent Component Analysis and Blind Source Separation*, San Diego, California, 2001.

[120] Parra, L., Spence, C., Sajda, P., Ziehe, A., and Muller, K. R., 'Unmixing hyperspectral data', in *Advances in Neural Information Process.*, **13**, MIT Press, Cambridge, Massachussetts, 2000, pp. 942–948.

[121] Ikram, M., and Morgan, D., 'Exploring permutation inconstancy in blind separation of speech signals in a reverberant environment', in proceedings of the IEEE conference on *acoustic, speech, and signal processing* (*ICASSP 2000*), Turkey, 2000.

[122] Cherkani, N., and Deville, Y., 'Self adaptive separation of convolutively mixed signals with a recursive structure, Part 1: stability analysis and optimisation of asymptotic behaviour', *Signal Process.*, **73**(3), 1999, 225–254.

[123] Smaragdis, P., 'Blind separation of convolved mixtures in the frequency domain', *Neurocomputing*, **22**, 1998, 21–34.

[124] Murata, N., Ikeda, S., and Ziehe, A., 'An approach to blind source separation based on temporal structure of speech signals', *Neurocomputing*, **41**, 2001, 1–4.

[125] Zibulevsky, M., 'Relative Newton method for quasi-ML blind source separation', `http://ie.technion.ac.il/mcib`, 2002.

[126] Luo, Y., Chambers, J., Lambotharan, S., and Proudler, I., 'Exploitation of source non-stationarity in underdetermined blind source separation with advanced clustering techniques', *IEEE Trans. Signal Process.*, **54**(6), 2006, 2198–2212.

[127] Li, Y., Amari, S., Cichocki, A., Ho, D. W. C., and Shengli, X., 'Underdetermined blind source separation based on sparse representation', *IEEE Trans Signal Process.*, **54**(2), 2006, 423–437.

[128] Jurjine, A., Rickard, S., Yilmaz, O., 'Blind separation of disjoint orthogonal signals: demixing *N* sources from 2 mixtures', in Proceedings of the IEEE Conference on *Acoustic, Speech, and Signal Processing* (*ICASSP 2000*), Turkey Vol. 5, 2000, pp. 2985–2988.

[129] Vielva, L., Erdogmus, D., Pantaleon, C., Santamaria, I., Pereda, J., and Principle, J. C., 'Underdetermined blind source separation in a time-varying environment', in *Proceedings of the IEEE Conference on Acoustic, Speech, and Signal Processing (ICASSP 2000), Turkey*, Vol. 3, 2002, pp. 3049–3052.

[130] Takigawa, I., Kudo, M., Nakamura, A., and Toyama, J., 'On the minimum l_1-norm signal recovery in underdetermined source separation.' in Proceedings of the 5th International Conferences on *Independent Component Analysis*, Granada, Spain, 2004, pp. 22–24.

[131] Nazarpour, K., Sanei, S., Shoker, L., and Chambers, J., 'Parallel space–time–frequency decomposition of EEG signals for brain computer interfacing', in Proceedings of EUSIPCO 2006, Florence, Italy, 2006.

[132] Jutten, C., and Karhunen, J., 'Advances in blind source separation (BSS) and independent component analysis (ICA) for nonlinear mixtures', *Int. J. Neural Systems*, **14**(5), 2004, 267–292.

[133] Almeida, L. B., 'Nonlinear Source Separation', in *Synthesis Lectures on Signal Processing*, Ed. J. Moura, Morgan & Claypool, California, USA, 2006.

[134] Joho, M., and Mathis, H., 'Joint diagonalization of correlation matrices by using gradient methods with application to blind signal processing', in Proceedings of the 1st Annual Conference *Sensor Array and Multichannel Signal Processing* (*SAM2002*), Rosslyn, VA, USA, 2002, 273–277.

[135] Parra, L., and Spence, C., 'Convolutive blind separation of non-stationary sources', *IEEE Trans. Speech and Audio Process.*, **8**, 2000, 320–327.

[136] Hyvärinen, A., Hoyer, P. O., and Inkl, M., 'Topographic independent component analysis', *Neural Computation*, **13**, 2001, 1527–1558.

[137] Vapnik, V., *Statistical Learning Theory*, John Wiley & Sons, Inc., New York, 1998.

[138] Shoker, L., Sanei, S., and Chambers, J., 'Artifact removal from electroencephalograms using a hybrid BSS-SVM algorithm', *IEEE Signal Process. Lett.*, **12**(10), October. 2005.

[139] Shoker, L., Sanei, S., Wang, W., and Chambers, J., 'Removal of eye blinking artifact from EEG incorporating a new constrained BSS algorithm', *IEE J. Med. Biolog. Engng. and Computing*, **43**, 2004, 290–295.

[140] Gonzalez, B., and Sanei, S., and Chambers, J., 'Support vector machines for seizure detection', in Proceedings of the IEEE ISSPIT2003, Germany, December 2003, pp. 126–129.

[141] Shoker, L., Sanei, S., and Sumich, A., 'Distinguishing between left and right finger movement from EEG using SVM', in Proceedings of the IEEE EMBS, Shanghai, China, September 2005, pp. 5420–5423.

[142] Bennet, K. P., and Campbell, C., 'Support vector machines: hype or hallelujah?', *SIGKDD Explorations*, **2**(2), 2000, 1–13; http://www.rpi.edu/∼bennek.

[143] Christianini, N., and Shawe-Taylor, J., *An Introduction to Support Vector Machines*, Cambridge University Press, Cambridge, 2000.

[144] DeCoste, D., and Scholkopf, B., *'Training invariant support vector machines', Machine Learning*, Kluwer Press, Dordrecht, The Netherlands, 2001.

[145] Burges, C., 'A tutorial on support vector machines for pattern recognition', *Data Mining and Knowledge Discovery*, **2**, 1998, 121–167.

[146] Gunn, S., 'Support vector machines for classification and regression', Technical Reports, Department of Electronics and Computer Science, Southampton University, 1998.

[147] Vapnik, V., *The Nature of Statistical Learning Theory*, Springer, New York, 1995.

[148] `http://www.support-vector.net`.

[149] `http://www.kernel-machines.org`.

[150] Chapelle, O., and Vapnik, V., 'Choosing multiple parameters for support vector machines', *Machine Learning*, **46**, 2002, 131–159.

[151] Weston, J., and Watkins, C., 'Support vector machines for multi-class pattern recognition', in Proceedings of the Seventh European Symposium on *Artificial Neural Networks*, Bruges, Belgium, 1999.

[152] Platt, J., 'Sequential minimal optimisation: a fast algorithm for training support vector machines', Technical Report MSR-TR-98-14, Microsoft Research, 1998.

[153] Hartigan, J., and Wong, M., 'A k-mean clustering algorithm', *Appl. Statistics.*, **28**, 1979, 100–108.

[154] Fellous, J.-M., Tiesinga, P. H. E., Thomas, P. J., and Sejnowski, T. J., 'Discovering spike patterns in neural responses', *J. Neurosci.*, **24**(12), 2004, 2989–3001.

[155] Hastie, T., Tibshirani, R., and Walter, G., 'Estimating the number of clusters in a dataset via the gap statistic', Technical Report 208, Stanford University, Stanford, California, 2000.

[156] Mallat, S., and Zhang, Z., 'Matching pursuit with time–frequency dictionaries', *IEEE Trans. Signal Process.*, **41**, 1993, 3397–3415.

[157] Blinowska, K. J., and Durka, P. J., 'The application of wavelet transform and matching pursuit to the time-varying EEG signals', in *Intelligent Engineering Systems Through Artificial Neural Networks*, Eds C. H. Dagli and B. R. Fernandez Vol. 4, ASME Press, New York, 1994, pp. 535–540.

[158] Gabor, D., 'Theory of communication', *J. IEE*, **93**, 1946, 429–457.

[159] Durka, P. J., and Blinowska, K. J., 'Analysis of EEG transients by means of matching pursuit', *Epilepsia*, **37**, 1996, 386–399.

[160] Durka, P. J., Durka, P. J., Ircha, D., Neuper, C., and Pfurtscheller, G., 'Time–frequency microstructure of event-related EEG desynchronization and synchronization', *Med. & Biolog. Engng. Computing*, **39**(3), May 2001.