

# **Entropy Feature Extraction of EEG Signals for Automatic Person Identification**

A thesis submitted for the degree  
of Professional Doctorates (in Information Sciences, 862AA) of  
The University of Canberra

Dinh Van Phung

February 2016

# Summary of Thesis

The use of human brain electroencephalography (EEG) signals for automatic person identification has been investigated for a decade. It has been found that the performance of an EEG-based person identification system highly depends on what feature to be extracted from multi-channel EEG signals. Linear methods such as Power Spectral Density and Autoregressive Model have been used to extract EEG features. However these methods assumed that EEG signals are stationary. In fact, EEG signals are complex, non-linear, non-stationary, and random in nature. In addition, other factors such as brain condition or human characteristics may have impacts on the performance, however these factors have not been investigated and evaluated in previous studies.

It has been found in the literature that entropy is used to measure the randomness of non-linear time series data. Entropy is also used to measure the level of chaos of brain-computer interface systems. Therefore, this thesis proposes to study the role of entropy in non-linear analysis of EEG signals to discover new features for EEG-based person identification. Five different entropy methods including Shannon Entropy, Approximate Entropy, Sample Entropy, Spectral Entropy, and Conditional Entropy have been proposed to extract entropy features that are used to evaluate the performance of EEG-based person identification systems and the impacts of epilepsy, alcohol, age and gender characteristics on these systems.

Experiments were performed on the Australian EEG and Alcoholism datasets. Experimental results have shown that, in most cases, the proposed entropy features yield very fast person identification, yet with compatible accuracy because the feature dimension is low. In real life security operation, timely response is critical. The experimental results have also shown that epilepsy, alcohol, age and gender characteristics have impacts on the EEG-based person identification systems.

# Acknowledgements

I would like to express my sincere gratitude to my advisor Associate Professor Dat Tran for his continuous support of my research while I was completing my Professional Doctorate (PD) in the Faculty of Education, Science, Technology and Mathematics. His patient guidance, constant encouragement and immense specialized knowledge helped me throughout my time of research and the writing of this thesis. I could not imagine a better advisor and mentor for my PD study.

Besides my advisor, I would like to thank the other members of my thesis committee, Professor Dharmendra Sharma and Doctor Wanli Ma, for their insightful comments and encouragement, but also for what were often difficult questions that challenged me to widen my research and view issues from various important perspectives.

I thank my fellow lab mates for the stimulating discussions, for the sleepless nights when we worked together before imminent deadlines, and for all the fun we had during the past four years. Also, I will cherish the happy memories of my friends in the Faculty and at the University of Canberra.

I would like to thank my family: my parents, my wife and sons and my sisters, for supporting me spiritually throughout this research and writing. They have been always there for me and have enriched my days in so many ways.

# Contents

<b>Summary of Thesis</b>	<b>iii</b>
<b>Acknowledgements</b>	<b>vii</b>
<b>Abbreviation</b>	<b>xix</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Current Approaches . . . . .	4
1.1.1 Feature Extraction . . . . .	4
1.1.2 Epilepsy Analysis . . . . .	5
1.1.3 Alcohol Analysis . . . . .	7
1.1.4 Age and Gender Analysis . . . . .	8
1.2 Problem Statements and Research Questions . . . . .	10
1.2.1 Problem Statements . . . . .	10
1.2.2 Research Questions . . . . .	12
1.3 Research Methodology . . . . .	12
1.3.1 Current Entropies . . . . .	12
1.3.2 Our Proposed Methodology . . . . .	15
1.4 Contributions of This Thesis . . . . .	17
1.4.1 Entropy Feature Extraction Methods for Enhancing the Performance of EEG-based Person Identification Systems . . . . .	17
1.4.2 Impact of Epilepsy on EEG-based Person Identification Systems	18
1.4.3 Impact of Alcohol on Person Identification Systems . . . . .	18
1.4.4 Impact of Age and Gender on Person Identification Systems .	19

---

<b>2</b>	<b>Literature Review</b>	<b>21</b>
2.1	EEG Characteristics . . . . .	21
2.1.1	EEG Signals . . . . .	21
2.1.2	EEG Brainwave Patterns . . . . .	23
2.1.3	EEG Signal Processing . . . . .	26
2.2	EEG-based Person Identification . . . . .	26
2.3	EEG-based Feature Extraction for Person Identification . . . . .	28
2.3.1	Fourier Transform . . . . .	28
2.3.2	Power Spectral Density . . . . .	29
2.3.3	Autoregressive Model . . . . .	30
2.4	Entropy . . . . .	30
2.4.1	Shannon Entropy . . . . .	31
2.4.2	Approximate Entropy . . . . .	32
2.4.3	Sample Entropy . . . . .	33
2.4.4	Spectral Entropy . . . . .	33
2.4.5	Conditional Entropy . . . . .	34
2.5	Support Vector Machine . . . . .	35
2.5.1	Binary Classification SVMs . . . . .	35
	The training phase . . . . .	35
	The test phase . . . . .	39
2.5.2	k-class Classification SVMs . . . . .	39
<b>3</b>	<b>Experiment Setups</b>	<b>41</b>
3.1	EEG datasets . . . . .	41
3.1.1	Australian EEG (AEEG) . . . . .	41
3.1.2	Alcoholism . . . . .	42
3.2	EEG Pre-processing . . . . .	47
3.2.1	EEG Segmentation . . . . .	47
3.2.2	EEG Filtering . . . . .	47
3.3	Entropy Methods for Feature Extraction . . . . .	50
3.3.1	Shannon Entropy . . . . .	50
3.3.2	Approximate Entropy . . . . .	51
3.3.3	Sample Entropy . . . . .	52

---

3.3.4	Spectral Entropy . . . . .	52
3.3.5	Conditional Entropy . . . . .	55
3.4	Autoregressive Model . . . . .	59
3.5	Person Identification Technique . . . . .	59
<b>4</b>	<b>Proposed Entropy Feature Extraction Methods for Multi-channel EEG Signals and Their Applications to Person Identification Systems</b>	<b>63</b>
4.1	Entropy Feature Extraction for Multi-channel EEG Signals . . . . .	64
4.1.1	Introduction . . . . .	64
4.1.2	Typical Methods for Multi-channel EEG Feature Extraction . . . . .	65
	Principal Component Analysis (PCA) . . . . .	65
	Independent Component Analysis (ICA) . . . . .	65
	Linear Discriminant Analysis (LDA) . . . . .	66
	Multivariate AR (MAR) model . . . . .	67
4.1.3	Proposed Entropy Feature Extraction Methods . . . . .	67
	Approach A: Concatenation of features from individual channels	69
	Approach B: Correlation between pairs of channels . . . . .	70
4.2	Performance of EEG-based Person Identification Systems Using Entropy Feature Extraction Methods . . . . .	70
4.3	Experiment Setups . . . . .	72
4.3.1	Experiment Data . . . . .	72
4.3.2	Feature Extraction . . . . .	72
4.3.3	Person identification . . . . .	74
4.4	Experimental Results . . . . .	75
4.5	Discussions . . . . .	81
4.5.1	Feature Dimensionality . . . . .	81
4.5.2	Person Identification Speed . . . . .	82
4.5.3	Person Identification Accuracy . . . . .	82
4.5.4	Proposed Approach for Accuracy Improvement . . . . .	83
<b>5</b>	<b>Impact of Brain Conditions on Person Identification Systems</b>	<b>91</b>
5.1	Epilepsy and Its impact on EEG-based Person Identification Systems	91

---

5.1.1	EEG Analysis for Epilepsy Detection . . . . .	93
5.1.2	Entropies for Epilepsy Analysis . . . . .	96
5.2	Alcohol and Its Impact on EEG-based Person Identification Systems .	98
5.2.1	Alcohol Impact on Neuroactivities . . . . .	98
5.2.2	Alcohol Detection . . . . .	99
5.3	Experiments on the Impact of Brain Conditions on Person Identifica- tion Systems . . . . .	102
5.3.1	Experimental Data . . . . .	102
5.3.2	Feature Extraction . . . . .	102
5.3.3	Person Identification . . . . .	103
5.3.4	Experimental Results . . . . .	105
5.4	Discussions . . . . .	106
<b>6</b>	<b>Impact of Age and Gender Characteristics of EEG Signals on Person Identification Systems</b>	<b>111</b>
6.1	Age and Gender Characteristics of EEG Signals . . . . .	111
6.2	EEGs for Age and Gender Classification . . . . .	115
6.3	Motivations . . . . .	116
6.4	Experiments to Evaluate the impact of Age and Gender Characteristics on Person Identification Systems . . . . .	116
6.4.1	EEG Data Set . . . . .	116
6.4.2	Feature Extraction . . . . .	117
6.4.3	Person Identification . . . . .	118
6.4.4	Experimental Results . . . . .	120
6.5	Discussion . . . . .	125
<b>7</b>	<b>Conclusions and Future Work</b>	<b>129</b>
7.1	Conclusions . . . . .	129
7.2	Future Work . . . . .	132
	<b>References</b>	<b>134</b>
	<b>Publications</b>	<b>149</b>

# List of Figures

1.1	A typical EEG-based person identification system . . . . .	2
1.2	Complexity analysis of AD patients using Multiscale Permutation Entropy (MPEn), after Morabito et al. [2012] . . . . .	14
1.3	A proposed research methodology . . . . .	15
2.1	An example of Evoked Potential, after Sanei and Chambers [2007] . .	22
2.2	A 10 – 20 electrode positioning system diagram, after Ochoa [2002] .	23
2.3	EEG Brainwave Patterns . . . . .	24
2.4	Separability of classes: a) separable; b) non-separable (linear hyperplane); c) non-separable (nonlinear hyperplane), after Sanei and Chambers [2007] . . . . .	36
2.5	SVM’s optimal hyperlane, after Lotte et al. [2007] . . . . .	37
2.6	Linear hyperlanes, after Burges [1998] . . . . .	38
3.1	Statistics of patients’ age at time of recording, after Hunter et al. [2005]	42
3.2	AEEG data file format . . . . .	43
3.3	Alcoholism data file format . . . . .	44
3.4	An example plot of a control subject, after Begleiter [1999] . . . . .	45
3.5	An example plot of an alcoholic subject, after Begleiter [1999] . . . .	46
3.6	Segmentation of AEEG into one-second trials . . . . .	47
3.7	Bandpass filtering . . . . .	49
3.8	EEG probabiliy distribution using a histogram . . . . .	50
3.9	Shannon Entropy for EEG signals . . . . .	51
3.10	Approximate Entropy and Sample Entropy for EEG signals . . . . .	53
3.11	Estimation of Spectral Entropy for EEG signals using PSD . . . . .	54



3.12 Spectral Entropy for EEG signals . . . . .	55
3.13 Scaled EEG signals, $X$ (Channel C3) . . . . .	56
3.14 Scaled EEG signals, $Y$ (Channel C4) . . . . .	57
3.15 Conditional Entropy between Channels C3 and C4 . . . . .	58
3.16 AR order selection using Akaike information criterion (AIC) . . . . .	60
3.17 EEG-based person identification using SVM . . . . .	61
4.1 Data classification process . . . . .	71
4.2 SVM setups for EEG-based person identification (without feature selection) . . . . .	74
4.3 Comparison between person identification rates of CEn and AR14 in the supplied test mode on (AEEG dataset) . . . . .	76
4.4 Comparison between person identification rates of CEn and AR6 in the supplied test mode (Alcoholism dataset) . . . . .	78
4.5 Person identification speed comparison between AR14 and the methods listed in Approach A (AEEG dataset) . . . . .	79
4.6 Person identification speed comparison between AR14 and the methods listed in Approach A (AEEG dataset) . . . . .	80
4.7 Person identification speed comparison between AR6 and the methods listed in Approach A (Alcoholism dataset) . . . . .	80
4.8 Person identification speed comparison between AR6 and the methods listed in Approach B (Alcoholism dataset) . . . . .	81
4.9 Comparison between entropy values of two persons (Beta band's one-second trial) . . . . .	84
4.10 Comparison of person identification rates between the methods listed in Approach A (on the combination of features from six wavebands) and AR14's best rates (on the AEEG dataset) . . . . .	85
4.11 Efficiency comparison of between the methods listed in Approach A (on the combination of features from six wavebands) and AR14's best rates (AEEG dataset) . . . . .	87
4.12 Efficiency comparison of between the methods listed in Approach A (on the combination of features from six wavebands) and AR6's best rates (Alcoholism dataset) . . . . .	87

---

4.13	Comparison of person identification rates for variable AEEG trial lengths (in the supplied test mode) . . . . .	89
5.1	EEG signals of a patient suffering a generalized seizure, after Sanei and Chambers [2007] . . . . .	92
5.2	EEG signals of a patient suffering focal seizure, after Sanei and Chambers [2007] . . . . .	93
5.3	ApEn values of normal and epileptic EEGs, after Kumar and Dewal [2011] . . . . .	97
5.4	SpEn values of normal and epileptic EEGs, after Mirzaei et al. [2010]	97
5.5	Automated diagnosis of alcoholism-related EEGs, after Acharya et al. [2014] . . . . .	100
5.6	ApEn values of control and alcoholic EEGs, after Padma Shri et al. [2014] . . . . .	101
5.7	The process of evaluating the impact of epilepsy or alcoholism . . . . .	104
5.8	Person Identification rates between the epileptic and non-epileptic groups (SampEn) . . . . .	107
5.9	Person Identification rates between the epileptic and non-epileptic groups (CEn) . . . . .	108
5.10	Person Identification rates between the alcoholic and non-alcoholic groups (ApEn) . . . . .	109
5.11	Person Identification rates between the alcoholic and non-alcoholic groups (CEn) . . . . .	110
6.1	Age-related changes in power spectra, after Marshall et al. [2002] . . . . .	112
6.2	Age-related changes, after Clarke et al. [2001] . . . . .	113
6.3	Gender-related changes, after Clarke et al. [2001] . . . . .	114
6.4	Age-gender group identification rates, after Nguyen et al. [2013a] . . . . .	115
6.5	Evaluation process for age impact on an EEG-based person identification system . . . . .	119
6.6	Person identification rates for all age groups . . . . .	121
6.7	Person identification rates of the two gender groups in different wavebands based on ApEn features . . . . .	123

---

6.8	Person identification rates of the two gender groups in different wave-	
	bands based on SpEn features . . . . .	124
6.9	Brain development with age . . . . .	126
6.10	Person identification rates of the two gender groups . . . . .	127

# List of Tables

3.1	Format of an AEEG record, after Hunter et al. [2005]	43
3.2	Description of experimented AEEG data subset	44
3.3	Description of experimented Alcoholism data subset	44
3.4	Segmentation of AEEG 900-second trial	48
3.5	Sub-sequence divisions	51
3.6	Example table of a joined histogram of trials $X$ and $Y$	56
3.7	Example table of histogram of trial $X$	57
3.8	Probability distribution of $p(x, y)$	57
3.9	Marginal distribution of $p(x)$	57
4.1	Description of experimented data subsets	72
4.2	Variable AEEG trial length for experiment	72
4.3	Number of features	73
4.4	Input and output of feature extraction	73
4.5	SVM parameter setup	74
4.6	Person identification results in the 3-fold cross validation mode	75
4.7	Person identification results in the supplied test mode	75
4.8	Person identification results in the 3-fold cross validation mode	77
4.9	Person identification results in the supplied test mode	77
4.10	Everage time for model building and testing on the AEEG dataset	77
4.11	Everage time for model building and testing on the Alcoholism dataset	78
4.12	Person identification results: on the combination of features from six sub-bands	84
4.13	Person identification results: on the combination of features from six sub-bands	86

---

5.1	Description of the dataset for epilepsy impact . . . . .	102
5.2	Description of the dataset for alcoholism impact . . . . .	102
5.3	Number of features and feature vectors . . . . .	103
5.4	SVM parameter setup . . . . .	103
5.5	Person identification rates (%) between the epileptic (Epi.) and non-epileptic (Non.) groups . . . . .	105
5.6	Person identification rates (%) between the alcoholic (Alc.) and non-alcoholic (Non.) groups . . . . .	106
6.1	Data description of Age groups. . . . .	117
6.2	Data description of Gender groups. . . . .	117
6.3	EEG feature description for the five entropy methods . . . . .	118
6.4	SVM parameter setup . . . . .	118
6.5	Person identification rates of the three age groups for the delta and theta wavebands. . . . .	120
6.6	Person identification rates of the three age groups for the alpha and beta wavebands. . . . .	120
6.7	Person identification rates of the three age groups for the gamma and mix wavebands. . . . .	121
6.8	Person identification rates of the two gender groups for the delta, theta and alpha wavebands. . . . .	122
6.9	Person identification rates of the two gender groups for the beta, gamma and mix wavebands. . . . .	122

# Abbreviations

AD	Alzheimer's Disease
AEEG	Australian Electroencephalogram Dataset
AIC	Akaike Information Criterion
ANN	Artificial Neural Network
ApEn	Approximate Entropy
AR	Auto Regressive
AR6	Auto Regressive, order 6
AR14	Auto Regressive, order 14
BCI	Brain-Computer Interface
BSS	Blind Source Separation
CC	Cross Correlation
CD	Correlation Dimension
CAD	Computer Aided Diagnostic
CEn	Conditional Entropy
CFS	Correlation-based Feature Selection
DFT	Discrete Fourier Transform
DTFT	Discrete Time Fourier Transform
DWT	Discrete Wavelet Transform
EEG	Electroencephalogram
EP	Evoked Potential
ERP	Event-related Potential
ESD	Energy Spectrum Density
FFT	Fast Fourier Transform

---

fMRI	functional Magnetic Resonance Imaging
FT	Fourier Transform
HC	Healthy Control Groups
HOS	High Order Spectra
ICA	Independent Component Analysis
K	KolmogorovSinai Entropy
KEn	Kolmogorov Entropy
k-NN	K-Nearest Neighbor
LC	Linear Complexity
LDA	Linear Discriminant Analysis
MAR	Multivariate Autoregressive model
MI	Mutual Information
MCI	Mild Cognitive Impaired
MEG	Magnetoencephalogram
MPEn	Multi-Scale Permutation Entropy
MSEn	Multi-Scale Entropy
NN	Neural Network
PCA	Component Analysis
PEn	Permutation Entropy
PET	Positron Emission Tomography
PLV	Phase Locking Value
PSD	Power Spectral Density
REn	Renyis Entropy
SampEn	Sample Entropy
SEn	Shannon Entropy
STFT	Short-Time Fourier Transform
SOM	Self-organizing Map
SpEn	Spectral Entropy
SVM	Support Vector Machine
VEP	Visual Evoked Potentials
WEn	Wavelet Entropy
WT	Wavelet Transform
WPD	Wavelet Packet Decomposition