Fundamentals of Speaker Recognition

Homayoon Beigi

Fundamentals of Speaker Recognition

Springer

Dr. Homayoon Beigi
Recognition Technologies, Inc.
Yorktown Heights New York, NY
USA
beigi@recotechnologies.com

ISBN 978-0-387-77591-3 e-ISBN 978-0-387-77592-0
DOI 10.1007/978-0-387-77592-0
Springer New York Dordrecht Heidelberg London

Library of Congress Control Number: 2011941119

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Contents

Acronyms and Abbreviations

ADPCM	Adaptive Differential Pulse Code Modulation
AEP	Asymptotic Equipartition Property
AGN	Automatic Gain Normalization
AHC	Agglomorative Hierarchical Clustering
ANSI	American National Standards Institute
API	Application Programming Interface
ASR	Automatic Speech Recognition
BFGS	Broyden-Fletcher-Goldfarb-Shanno
BIC	Bayesian Information Criterion
BioAPI	Biometric Application Programming Interface
CBEFF	Common Biometric Exchange Formats Framework
CDMA	Code Division Multiple Access
CELP	Code Excited Linear Prediction
CHN	Cepstral Histogram Normalization
CMA	Constant Modulus Algorithm
CMN	Cepstral Mean Normalization
CMS	Cepstral Mean Subtraction
CMVN	Cepstral Mean and variance Normalization
CNG	Comfort Noise Generation
CoDec	Coder/Decoder
CS-ACELP	Conjugate Structure Algebraic Code Excited Linear Prediction
dB	deci Bel (decibel)
DC	Direct Current
DCF	Detection Cost Function
DCT	Discrete Cosine Transform
DET	Detection Error Trade-Off
DFP	Davidon-Fletcher-Powell
DHC	Divisive Hierarchical Clustering
DPCM	Differential Pulse Code Modulation
DTMF	Dual Tone Multi-Frequency

EER	Equal-Error Rate
e.g.	exempli gratia (for example)
EIH	Ensemble Interval Histogram
ELRA	European Language Resources Association
EM	Expectation Maximization
EMD	Empirical Mode Decomposition
EMMA	Extensible Multimodal Annotation
ETSI	European Telecommunications Standards Institute
FA	Factor Analysis
FAR	False Acceptance Rate
FBI	Federal Bureau of Investigation
FFT	Fast Fourier Transform
FRR	False Rejection Rate
FTP	File Transfer Protocol
GLR	General Likelihood Ratio
GMM	Gaussian Mixture Model(s)
GrXML	Grammar eXtensible Markup Language
GSM	Groupe Spécial Mobile *or* Global System for Mobile Communications
GSM-EFR	GSM Enhanced Full Rate
HE-AAC	High Efficiency Advanced Audio Coding
HEQ	Histogram Equalization
HME	Hierarchical Mixtures of Experts
HMM	Hidden Markov Model(s)
H-Norm	Handset Normalization
HTER	Half Total Error Rate
HTTP	HyperText Transfer Protocol
Hz	Hertz
IBM	International Business Machines
ID	Identity; Identification
iDEN	Integrated Digital Enhanced Network
i.e.	id est (that is)
IEC	International Electrotechnical Commission
IETF	Internet Engineering Task Force
IFG	Inferior Frontal Gyrus (of the Brain)
i.i.d.	Independent and Identically Distributed (Description of a type of Random Variable)
IMF	Intrinsic Mode Function
INCITS	InterNational Committee for Information Technology Standards
ISO	International Organization for Standardization
ISV	Independent Software Vendor
ITU	International Telecommunications Union
ITU-T	ITU Telecommunication Standardization Sector
JFA	Joint Factor Analysis
JTC	Joint ISO/IEC Technical Committee
IVR	Interactive Voice Response

KLT	Karhunen-Loève Transformation
LBG	Linde-Buzo-Gray
LFA	Latent Factor Analysis
kHz	kilo-Hertz
LDC	Linguistic Data Consortium
LAR	Log Area Ratio
LLN	Law of Large Numbers
LLR	Log-Likelihood Ratio
LPC	Linear Predictive Coding, also, Linear Predictive Coefficients
LPCM	Linear Pulse Code Modulation
MAP	Maximum A-Posteriori
MFCC	Mel Frequency Cepstral Coefficients
MFDWC	Mel Frequency Discrete Wavelet Coefficients
MIT-LL	Massachusetts Institute of Technology's Lincoln Laboratories
MLE	Maximum Likelihood Estimation or Maximum Likelihood Estimate
MLLR	Maximum Likelihood Linear Regression
MMIE	Maximum Mutual Information Estimation
MPEG	Moving Picture Experts Group
MRCP	Media Resource Control Protocol
NAP	Nuisance Attribute Projection
N.B.	Nota Bene (Note Well) – Note that
NIST	National Institute of Standards and Technology
NLSML	Natural Language Semantics Markup Language
NLU	Natural Language Understanding
OGI	Oregon Graduate Institute
PAM	Pulse Amplitude Modulation (Sampler)
PARCOR	Partial Correlation
PCA	Principal Component Analysis
PCM	Pulse Code Modulation
PCMA	A-Law Pulse Code Modulation
PCMU	μ-Law Pulse Code Modulation
PDC	Personal Digital Cellular
ppm	Parts per Million
pRAM	Probabilistic Random Access Memory
PSTN	Public Switched Telephone Network
PWM	Pulse Width Modulation (Sampler)
PWPAM	Pulse Width Pulse Amplitude Modulation (Sampler)
QCELP	Qualcomm Code Excited Linear Prediction
Q.E.D.	Quod Erat Demonstradum (That which was to be Demostrated)
QOS	Quality of Service
rad.	radians
RASTA	RelAtive SpecTrAl
RBF	Radial Basis Function
RFC	Request for Comments

RIFF	Resource Interchange File Format
RNN	Recurrent Neural Network
ROC	Receiver Operator Characteristic
RTP	Real-time Transport Protocol
SAFE	Standard Audio Format Encapsulation
SC	Subcommittee
SI	Systèm International
SIMM	Sequential Interacting Multiple Models
SIP	Session Initiation Protocol
SIV	Speaker Identification and Verification
SLLN	Strong Law of Large Numbers
SPHERE	SPeech HEader REsources
SPI	Service Provider Interface
SRAPI	Speech Recognition Application Programming Interface
SSML	Speech Synthetic Markup Language
SVAPI	Speaker Verification Application Programming Interface
SVM	Support Vector Machine(s)
TCP	Transmission Control Protocol
TD-SCDMA	Time Division Synchronous Code Division Multiple Access
TLS	Transport Layer Security
TDMA	Time Division Multiple Access
TDNN	Time-Delay Neural Network
T-Norm	Test Normalization
TTS	Text To Speech
U8	Unsigned 8-bit Storage
U16	Unsigned 16-bit Storage
U32	Unsigned 32-bit Storage
U64	Unsigned 64-bit Storage
UDP	User Datagram Protocol
VAD	Voice Activity Detection
VAR	Value Added Reseller
VB	Variational Bayesian Technique
VBWG	Voice Browser Working Group
VoiceXML	Voice eXtensible Markup Language
VoIP	Voice Over Internet Protocol
VQ	Vector Quantization
W3C	World Wide Web Consortium
WG	Workgroup
WCDMA	Wideband Code Division Multiple Access
WCDMA HSPA	Wideband Code Division Multiple Access High Speed Packet Access
WLLN	Weak Law of Large Numbers
XML	eXtensible Markup Language

Nomenclature

In this book, lower-case bold letters are used to denote vectors and upper-case bold letters are used for matrices. For set, measure, and probability theory, as much as possible, special style guidelines have been used such that the letter X when written as $\mathscr{X}$ signifies a set and when written as $\mathfrak{X}$ is a class of (sub)sets. The following is a list of symbols used in the text:

$\{\varnothing\}$	Empty Set
$\overline{(\alpha+i\beta)}$	Complex Conjugate of $(\alpha+i\beta)$ equal to $(\alpha-i\beta)$
$\lvert . \rvert$	Determinant of .
$(\mathbf{a})_{[i]}$	i^{th} element of vector $\mathbf{a}$.
$(\mathbf{A})_{[i][j]}$	Element in row i and column j of matrix $\mathbf{A}$.
$(\mathbf{A})_{[i]}$	Column i of matrix $\mathbf{A}$.
$*$	Convolution, e.g., $g*h$.
$\circ$	Correlation (Cross-Correlation), e.g., $g\circ h$, $g\circ g$.
$\tilde{\cdot}$	Estimate of $\cdot$
$\wedge$	Logical And
$\vee$	Logical Or
$\mapsto$	Maps to, e.g. $\mathscr{R}^N \mapsto \mathscr{R}^M$
$\leftrightarrow$	Mutual Mapping (used for signal/transform pairs, e.g. $h(t)\leftrightarrow H(s)$).
$\therefore$	Therefore
$\overset{\mathsf{R}}{\equiv}$	Equivalent with respect to equivalence relation R.
$\sim$	Distributed According to $\cdots$ (a Distribution).
$\preceq$	$\mathbf{a}\preceq\mathbf{b}$ is read, $\mathbf{a}$ *precedes* $\mathbf{b}$ – i.e. in an ordered set of vectors.
$\prec$	$\mathbf{a}\prec\mathbf{b}$ is read, $\mathbf{a}$ *strictly precedes* $\mathbf{b}$ – i.e. in an ordered set of vectors.
$\succeq$	$\mathbf{a}\succeq\mathbf{b}$ is read, $\mathbf{a}$ *succeeds* $\mathbf{b}$ – i.e. in an ordered set of vectors.
$\succ$	$\mathbf{a}\succ\mathbf{b}$ is read, $\mathbf{a}$ *strictly succeeds* $\mathbf{b}$ – i.e. in an ordered set of vectors.
$\bar{x}$	Mean (Expected Value) of x
$\mathscr{A}$	A generic set.
$\mathscr{A}^{\complement}$	Complement of set $\mathscr{A}$.
$\mathscr{A}\backslash\mathscr{B}$	The difference between $\mathscr{A}$ and $\mathscr{B}$.

$\mathbf{A}$	Jacobian matrix of optimization constraints with respect to $\mathbf{x}$
$\mathscr{B}$	A generic set.
$\mathscr{B}_c$	Center Frequency of a Critical Band
$\mathscr{B}_w$	Bandwidth of a Critical Band
$\mathbb{C}$	Set of Complex Numbers
$\mathscr{C}$	Cost Function
$\mathscr{C}^n$	n-dimensional Complex Space
D	Dimension of the feature vector
Δ	Step Change
$\mathscr{D}$	Domain of a Function
$\Upsilon_{\mathscr{A}}(x)$	Characteristic function of $\mathscr{A} \in \mathfrak{X}$ for random variable X
$\mathscr{D}_F(. \leftrightarrow .)$	f-Divergence
$\mathscr{D}_J(. \leftrightarrow .)$	Jeffreys Divergence
$\mathscr{D}_{KL}(. \rightarrow .)$	Kullback-Leibler Divergence
$d_E(.,.)$	Euclidean Distance
$d_{WE}(.,.)$	Weighted Euclidean Distance
$d_H(.,.)$	Hamming Distance
$d_{He}(.,.)$	Hellinger's Distance
$d_M(.,.)$	Mahalanobis Distance
$\nabla_{\mathbf{x}} E$	Gradient of E with respect to $\mathbf{x}$
$E(.)$	Objective Function of Optimization
$\mathscr{E}\{\cdot\}$	Expectation of $\cdot$
e	Euler's Constant $(2.7182818284\ldots)$
$\mathbf{e}_n$	Error vector
$\bar{\mathbf{e}}_N$	N-dimensional vector of all ones, i.e. $\bar{\mathbf{e}} : \mathscr{R}^1 \mapsto \mathscr{R}^N$ such that, $(\bar{\mathbf{e}}_N)_{[n]} = 1 \;\; forall \;\; n = \{1, 2, \cdots, N\}$
$\hat{\mathbf{e}}_k$	Unit vector whose k^{th} element is 1 and all other elements are 0
$exp\{\cdot\}$	Exponential function $(e^{\{\cdot\}})$
ϕ	Sample Space of the Parameter Vector, $\boldsymbol{\varphi}$
$\boldsymbol{\varphi}_\gamma$	Parameter Vector for the cluster γ
$\boldsymbol{\Phi}$	Matrix of parameter vectors
F_s	Spectral Flatness
$\mathscr{F}\{\cdot\}$	Fourier Transform of $\cdot$
$\mathscr{F}^{-1}\{\cdot\}$	Inverse Fourier Transform of $\cdot$
$\mathfrak{F}$	A Field
$\boldsymbol{\mathscr{I}}_F(\boldsymbol{\varphi}\|\mathbf{x})$	Fisher Information matrix for parameter vector $\boldsymbol{\varphi}$ given $\mathbf{x}$
f	Frequency measured in Hertz $(\frac{cycles}{s})$
f_c	Nyquist Critical Frequency measured in Hertz $(\frac{cycles}{s})$
f_s	Sampling Frequency measured in Hertz $(\frac{cycles}{s})$
Γ	Number of clusters – mostly Gaussian clusters
γ	Cluster index – mostly for Gaussian clusters
$\boldsymbol{\gamma}_{n_c}$	Column n_c of Jacobian matrix $(\mathbf{J})$ of optimization constraints
$\mathbf{G}$	Hessian Matrix

$\mathbf{g}$	Gradient Vector
$\mathscr{H}(p)$	Entropy
$\mathscr{H}(p\|q)$	Conditional Entropy
$\mathscr{H}(p,q)$	Joint Entropy
$\mathscr{H}(p \to q)$	Cross Entropy
$\mathbf{H}$	Inverse Hessian Matrix
$\mathscr{H}$	Hilbert Space
$\mathfrak{H}$	Borel Field of the Borel Sets in Hilbert Space
$\mathscr{H}_p$	Pre-Hilbert Space
$\mathfrak{H}_p$	Borel Field of the Borel Sets in Pre-Hilbert Space
H_0	Null Hypothesis
H_1	Alternative Hypothesis
$H(f)$	Fourier Transform of the signal $h(t)$
$H(s)$	Laplace Transform of the signal $h(t)$
$H(s)$	Any Generic Function of a Complex Variable
$H(\omega)$	Fourier Transform of the signal $h(t)$ in Terms of the Angular Frequency ω
H_{kl}	Discrete Fourier Transform of the sampled signal h_{nl} in frame l for the linear frequency index k
$\breve{H}_{ml}$	Mel-scale Discrete Fourier Transform of the sampled signal h_{nl} in frame l for the Mel frequency index m
$h(t)$	A Continuous Function of Time or a Continuous Signal
$\hbar(p)$	Differential Entropy (Continuous Entropy)
$\hbar(p \to q)$	Differential Cross Entropy (Continuous Cross Entropy)
I_0	Standard Intensity Threshold for Hearing
I	Intensity of Sound
I_r	Relative Intensity of Sound
$\mathscr{I}$	Information
$\mathscr{I}(X;Y)$	Mutual Information between Random Variables X and Y
$\mathscr{I}_J(X;Y)$	Jeffrey's Mutual Information between Random Variables X and Y
$\mathbb{I}$	Set of Imaginary Numbers
$\mathbf{I}$	Identity Matrix
$\mathscr{I}m$	The Imaginary part of variable $\{s : s \in \mathbb{C}\}$
$\mathbf{I}_N$	N-dimensional Identity Matrix
i	The Imaginary Number ($\sqrt{-1}$)
iff	If and Only If ($\iff$)
inf	Infimum
$\mathscr{K}(t,s)$	Kernel Function of t and s used in Integral Transforms
$\boldsymbol{\Lambda}$	Diagonal matrix of Eigenvalues
λ	Lebesgue Measure
$\tilde{\lambda}$	Wavelength
$\bar{\lambda}$	Forgetting Factor

λ	Eigenvalue
λ	Lagrange Multiplier
L	Total number of frames
$\mathscr{L}(\boldsymbol{\varphi}\|\mathbf{x})$	Likelihood of $\boldsymbol{\varphi}$ given $\mathbf{x}$
$\mathscr{L}\{\cdot\}$	Laplace Transform of $\cdot$
$\mathscr{L}^{-1}\{\cdot\}$	Inverse Laplace Transform of $\cdot$
$\mathfrak{L}_p$	Class of extended real valued p-integrable functions
l	Frame Index
$\ell(\boldsymbol{\varphi}\|\mathbf{x})$	Log-Likelihood of $\boldsymbol{\varphi}$ given $\mathbf{x}$
$ln(\cdot)$	Napierian Logarithm, Natural Logarithm, or Hyperbolic Logarithm ($log_e(\cdot)$)
$log(\cdot)$	Common Logarithm ($log_{10}(\cdot)$)
$\boldsymbol{\mu}$	Mean Vector
$\hat{\boldsymbol{\mu}}$	Sample mean vector, as a shortcut for $\overline{X}\|_N$
$\hat{\boldsymbol{\mu}}_\gamma$	Sample mean vector for cluster γ
M	Number of Models, number of critical bands
M	Number of samples in a partition of the Welch PSD computation
M	Dimension of the parameter vector
$\mathscr{M}$	Matrix of the weights for mapping the linear frequency to the Mel scale critical filter bank frequencies
$\mathscr{N}(\boldsymbol{\mu},\boldsymbol{\Sigma})$	Gaussian or Normal Distribution with mean $\boldsymbol{\mu}$ and Variance-Covariance $\boldsymbol{\Sigma}$
N	Window size
N	Number of samples
N	Number of hypotheses
n	Sample index which is not necessarily time aligned – see t for time aligned sample index
N_γ	Number of samples associated with cluster γ
N_s	Number of samples associated with state s
$\mathbb{N}$	The set of Natural Numbers
O	Observation random variable
$\mathscr{O}$	Observation sample space
$\mathscr{O}$	Bachmann-Landau asymptotic notation – Big-O notation
$\mathfrak{O}$	Borel Fields of the Borel Sets of sample space $\mathscr{O}$
o	An observation sample
$\boldsymbol{\varpi}$	Pulsewidth of Pulse Amplitude Modulation Sampler
$\boldsymbol{\varpi}(o\|s)$	Penalty (loss) associated with decision o conditioned on state s
$\boldsymbol{\varpi}(o\|x)$	Conditional Risk in Bayesian Decision Theory
$\wp$	Pitch
$\boldsymbol{\Pi}$	Penalty matrix in Bayesian Decision Theory.
P	Probability
P	Pressure Differential
P_0	Pressure Threshold
$\mathscr{P}$	Total Power

$\mathscr{P}_d$	Power Spectral Density
$\mathscr{P}_d^\circ$	Power Spectral Density in Angular Frequency
p	Probability Distribution
$\mathfrak{p}$	Training patten index for a Neural Network
q	Probability Distribution
$\mathbb{R}$	Set of Real Numbers
R	Redundancy
$\mathscr{R}(h)$	Range of Function h – Set of values which function h may take on
$\mathscr{R}e(s)$	The Real part of variable $\{s : s \in \mathbb{C}\}$
$\mathscr{R}^n$	n-dimensional Euclidean Space
$\boldsymbol{\Sigma}$	Covariance (Variance-Covariance) Matrix
$\hat{\boldsymbol{\Sigma}}$	Biased Sample Covariance (Variance-Covariance) Matrix
$\tilde{\boldsymbol{\Sigma}}$	Unbiased Sample Covariance (Variance-Covariance) Matrix
$\hat{\boldsymbol{\Sigma}}_\gamma$	Biased Sample Covariance Matrix for cluster γ
Ⓢ	Number of States
S	State Random variable
$\mathscr{S}$	State sample space
$\mathfrak{S}$	State Borel Field of the Borel Sets of sample space $\mathscr{S}$
$\mathbf{S}\|_N$	Second Order Sum ($\sum_{i=1}^N \mathbf{x}_i\mathbf{x}_i{}^T$)
s	A sample of the state random variable
$\mathbf{s}\|_N$	First Order Sum ($\sum_{i=1}^N \mathbf{x}_i$)
sup	Supremum
$\boldsymbol{\varsigma}(\boldsymbol{\varphi}\|\mathbf{x})$	Score Statistic (Fisher Score) for parameters vector $\boldsymbol{\varphi}$ given $\mathbf{x}$
T	Total Number of Samples, and sometimes the Sampling Period
t	Sample index in time
T_c	Nyquist Critical Sampling Period
T_s	Sampling Period
$\hat{\mathbf{u}}$	Unit Vector
ω	Angular Frequency measured in $\frac{rad.}{s}$
ω_c	Nyquist Critical Angular Frequency measured in $\frac{rad.}{s}$
ω_s	Angular Sampling Frequency measured in $\frac{rad.}{s}$
W_N	The Twiddle Factor used for expressing DFT ($e^{i\frac{2\pi}{N}}$)
W_N^{kn}	$W_N^{(k\times n)}$
Ξ	Seconds of shift in feature computation
$\mathfrak{X}$	Borel Field (the smallest σ-field) of the Borel Sets of Sample Space, $\mathscr{X}$
$\mathscr{X}$	Sample Space
$\mathbf{x}$	Feature Vector
$\mathscr{Z}\{\cdot\}$	z Transform of $\cdot$
$\mathscr{Z}^{-1}\{\cdot\}$	Inverse z Transform of $\cdot$
$\mathbb{Z}$	The Set of Integers
$\mathbf{z}_k$	Direction of the Inverse Hessian Update in Optimization

List of Figures

List of Tables

List of Definitions

List of Theorems, Lemmas, and Laws

List of Properties

Index