**Contents**

Part I Basic Theory

1 Introduction ................................................... 3
  1.1 Definition and History ...................................... 3
  1.2 Speaker Recognition Branches ............................... 5
    1.2.1 Speaker Verification (Speaker Authentication) ........ 5
    1.2.2 Speaker Identification (Closed-Set and Open-Set) .... 7
    1.2.3 Speaker and Event Classification ...................... 8
    1.2.4 Speaker Segmentation ................................. 9
    1.2.5 Speaker Detection .................................. 11
    1.2.6 Speaker Tracking ................................... 11
  1.3 Speaker Recognition Modalities .............................. 12
    1.3.1 Text-Dependent Speaker Recognition .................. 12
    1.3.2 Text-Independent Speaker Recognition ................ 13
    1.3.3 Text-Prompted Speaker Recognition ................... 14
    1.3.4 Knowledge-Based Speaker Recognition ................ 15
  1.4 Applications .............................................. 16
    1.4.1 Financial Applications .............................. 16
    1.4.2 Forensic and Legal Applications .................... 18
    1.4.3 Access Control (Security) Applications .............. 19
    1.4.4 Audio and Video Indexing (Diarization) Applications ... 19
    1.4.5 Surveillance Applications ............................ 20
    1.4.6 Teleconferencing Applications ........................ 21
    1.4.7 Proctorless Oral Testing ............................. 21
    1.4.8 Other Applications ................................. 23
  1.5 Comparison to Other Biometrics ............................. 23
    1.5.1 Deoxyribonucleic Acid (DNA) ........................ 24
    1.5.2 Ear ............................................... 25
    1.5.3 Face .............................................. 27
    1.5.4 Fingerprint and Palm ................................ 28
    1.5.5 Hand and Finger Geometry ............................ 30
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.5.6</td>
<td>Iris</td>
<td>30</td>
</tr>
<tr>
<td>1.5.7</td>
<td>Retina</td>
<td>31</td>
</tr>
<tr>
<td>1.5.8</td>
<td>Thermography</td>
<td>32</td>
</tr>
<tr>
<td>1.5.9</td>
<td>Vein</td>
<td>32</td>
</tr>
<tr>
<td>1.5.10</td>
<td>Gait</td>
<td>33</td>
</tr>
<tr>
<td>1.5.11</td>
<td>Handwriting</td>
<td>34</td>
</tr>
<tr>
<td>1.5.12</td>
<td>Keystroke</td>
<td>35</td>
</tr>
<tr>
<td>1.5.13</td>
<td>Multimodal</td>
<td>35</td>
</tr>
<tr>
<td>1.5.14</td>
<td>Summary of Speaker Biometric Characteristics</td>
<td>37</td>
</tr>
<tr>
<td>References</td>
<td></td>
<td>38</td>
</tr>
<tr>
<td>2</td>
<td>The Anatomy of Speech</td>
<td>43</td>
</tr>
<tr>
<td>2.1</td>
<td>The Human Vocal System</td>
<td>44</td>
</tr>
<tr>
<td>2.1.1</td>
<td>Trachea and Larynx</td>
<td>44</td>
</tr>
<tr>
<td>2.1.2</td>
<td>Vocal Folds (Vocal Chords)</td>
<td>44</td>
</tr>
<tr>
<td>2.1.3</td>
<td>Pharynx</td>
<td>47</td>
</tr>
<tr>
<td>2.1.4</td>
<td>Soft Palate and the Nasal System</td>
<td>48</td>
</tr>
<tr>
<td>2.1.5</td>
<td>Hard Palate</td>
<td>48</td>
</tr>
<tr>
<td>2.1.6</td>
<td>Oral Cavity Exit</td>
<td>48</td>
</tr>
<tr>
<td>2.2</td>
<td>The Human Auditory System</td>
<td>48</td>
</tr>
<tr>
<td>2.2.1</td>
<td>The Ear</td>
<td>50</td>
</tr>
<tr>
<td>2.3</td>
<td>The Nervous System and the Brain</td>
<td>51</td>
</tr>
<tr>
<td>2.3.1</td>
<td>Neurons – Elementary Building Blocks</td>
<td>52</td>
</tr>
<tr>
<td>2.3.2</td>
<td>The Brain</td>
<td>54</td>
</tr>
<tr>
<td>2.3.3</td>
<td>Function Localization in the Brain</td>
<td>59</td>
</tr>
<tr>
<td>2.3.4</td>
<td>Specializations of the Hemispheres of the Brain</td>
<td>62</td>
</tr>
<tr>
<td>2.3.5</td>
<td>Audio Production</td>
<td>64</td>
</tr>
<tr>
<td>2.3.6</td>
<td>Auditory Perception</td>
<td>66</td>
</tr>
<tr>
<td>2.3.7</td>
<td>Speaker Recognition</td>
<td>71</td>
</tr>
<tr>
<td>References</td>
<td></td>
<td>72</td>
</tr>
<tr>
<td>3</td>
<td>Signal Representation of Speech</td>
<td>75</td>
</tr>
<tr>
<td>3.1</td>
<td>Sampling The Audio</td>
<td>77</td>
</tr>
<tr>
<td>3.1.1</td>
<td>The Sampling Theorem</td>
<td>78</td>
</tr>
<tr>
<td>3.1.2</td>
<td>Convergence Criteria for the Sampling Theorem</td>
<td>84</td>
</tr>
<tr>
<td>3.1.3</td>
<td>Extensions of the Sampling Theorem</td>
<td>84</td>
</tr>
<tr>
<td>3.2</td>
<td>Quantization and Amplitude Errors</td>
<td>85</td>
</tr>
<tr>
<td>3.3</td>
<td>The Speech Waveform</td>
<td>87</td>
</tr>
<tr>
<td>3.4</td>
<td>The Spectrogram</td>
<td>87</td>
</tr>
<tr>
<td>3.5</td>
<td>Formant Representation</td>
<td>89</td>
</tr>
<tr>
<td>3.6</td>
<td>Practical Sampling and Associated Errors</td>
<td>92</td>
</tr>
<tr>
<td>3.6.1</td>
<td>Ideal Sampler</td>
<td>98</td>
</tr>
<tr>
<td>3.6.2</td>
<td>Aliasing</td>
<td>99</td>
</tr>
<tr>
<td>3.6.3</td>
<td>Truncation Error</td>
<td>102</td>
</tr>
<tr>
<td>3.6.4</td>
<td>Jitter</td>
<td>103</td>
</tr>
</tbody>
</table>
4 Phonetics and Phonology ........................................ 107
  4.1 Phonetics ................................................ . 107
    4.1.1 Initiation .......................................... 109
    4.1.2 Phonation ......................................... 109
    4.1.3 Articulation ....................................... 110
    4.1.4 Coordination ...................................... 111
    4.1.5 Vowels ........................................... 112
    4.1.6 Pulmonic Consonants ............................... 115
    4.1.7 Whisper .......................................... 119
    4.1.8 Whistle ........................................... 119
    4.1.9 Non-Pulmonic Consonants........................... 120
  4.2 Phonology and Linguistics .................................. 122
    4.2.1 Phonemic Utilization Across Languages ............... 122
    4.2.2 Whisper .......................................... 125
    4.2.3 Importance of Vowels in Speaker Recognition........... 127
    4.2.4 Evolution of Languages toward Discriminability......... 129
  4.3 Suprasegmental Features of Speech ........................... 131
    4.3.1 Prosodic Features .................................. 132
    4.3.2 Metrical features of Speech .......................... 138
    4.3.3 Temporal features of Speech ......................... 140
    4.3.4 Co-Articulation .................................... 140
References .................................................. ... 141

5 Signal Processing of Speech and Feature Extraction .......... 143
  5.1 Auditory Perception ........................................ 144
    5.1.1 Pitch ............................................. 146
    5.1.2 Loudness ......................................... 149
    5.1.3 Timbre ........................................... 151
  5.2 The Sampling Process ...................................... 152
    5.2.1 Anti-Aliasing ...................................... 153
    5.2.2 Hi-Pass Filtering ................................... 153
    5.2.3 Pre-Emphasis ...................................... 153
    5.2.4 Quantization ...................................... 155
  5.3 Spectral Analysis and Direct Method Features ............... 157
    5.3.1 Framing the Signal .................................. 160
    5.3.2 Windowing ........................................ 162
    5.3.3 Discrete Fourier Transform (DFT) and Spectral Estimation167
    5.3.4 Frequency Warping .................................. 169
    5.3.5 Magnitude Warping .................................. 172
    5.3.6 Mel Frequency Cepstral Coefficients (MFCC) ........... 173
    5.3.7 Mel Cepstral Dynamics ................................ 175
  5.4 Linear Predictive Cepstral Coefficients (LPCC) ............. 176
References .................................................. ... 177
5.4.1 Autoregressive (AR) Estimate of the PSD .......................... 177
5.4.2 LPC Computation ........................................ 184
5.4.3 Partial Correlation (PARCOR) Features ........................ 185
5.4.4 Log Area Ratio (LAR) Features ............................ 189
5.4.5 Linear Predictive Cepstral Coefficient (LPCC) Features .... 189
5.5 Perceptual Linear Predictive (PLP) Analysis ...................... 190
5.5.1 Spectral Analysis ....................................... 191
5.5.2 Bark Frequency Warping ................................ 191
5.5.3 Equal-Loudness Pre-emphasis ................................ 192
5.5.4 Magnitude Warping ...................................... 193
5.5.5 Inverse DFT ............................................ 193
5.6 Other Features .............................................. 193
5.6.1 Wavelet Filterbanks .................................... 194
5.6.2 Instantaneous Frequencies ............................... 197
5.6.3 Empirical Mode Decomposition (EMD) ....................... 198
5.7 Signal Enhancement and Pre-Processing ......................... 199
References ..................................................... 199

6 Probability Theory and Statistics .................................... 205
6.1 Set Theory ................................................ 205
6.1.1 Equivalence and Partitions ................................ 208
6.1.2 R-Rough Sets (Rough Sets) ............................. 210
6.1.3 Fuzzy Sets ............................................. 211
6.2 Measure Theory ............................................ 211
6.2.1 Measure ............................................... 212
6.2.2 Multiple Dimensional Spaces ............................ 216
6.2.3 Metric Space ......................................... 217
6.2.4 Banach Space (Normed Vector Space) .................... 218
6.2.5 Inner Product Space (Dot Product Space) .................. 219
6.2.6 Infinite Dimensional Spaces (Pre-Hilbert and Hilbert) ... 219
6.3 Probability Measure ........................................ 221
6.4 Integration ................................................ 227
6.5 Functions ............................................... 228
6.5.1 Probability Density Function ............................ 229
6.5.2 Densities in the Cartesian Product Space .................. 232
6.5.3 Cumulative Distribution Function ....................... 235
6.5.4 Function Spaces ....................................... 236
6.5.5 Transformations ....................................... 238
6.6 Statistical Moments ....................................... 239
6.6.1 Mean .................................................. 239
6.6.2 Variance .............................................. 242
6.6.3 Skewness (skew) ..................................... 245
6.6.4 Kurtosis .............................................. 246
6.7 Discrete Random Variables .................................. 247
6.7.1 Combinations of Random Variables ...................... 250
Contents

6.7.2 Convergence of a Sequence ................................... 250
6.8 Sufficient Statistics ........................................ 251
6.9 Moment Estimation ........................................ 253
  6.9.1 Estimating the Mean ................................ 253
  6.9.2 Law of Large Numbers (LLN) ......................... 254
  6.9.3 Different Types of Mean ............................. 257
  6.9.4 Estimating the Variance ................................ 258
6.10 Multi-Variate Normal Distribution ......................... 259
References .................................................. 261

7 Information Theory ............................................. 265
  7.1 Sources ................................................. 266
  7.2 The Relation between Uncertainty and Choice .......... 269
  7.3 Discrete Sources ....................................... 269
    7.3.1 Entropy or Uncertainty .............................. 270
    7.3.2 Generalized Entropy ................................ 278
    7.3.3 Information ....................................... 279
    7.3.4 The Relation between Information and Entropy ..... 280
  7.4 Discrete Channels ....................................... 282
  7.5 Continuous Sources .................................... 284
    7.5.1 Differential Entropy (Continuous Entropy) .......... 284
  7.6 Relative Entropy ....................................... 286
    7.6.1 Mutual Information ................................ 291
  7.7 Fisher Information .................................... 294
References .................................................. 299

8 Metrics and Divergences ....................................... 301
  8.1 Distance (Metric) ....................................... 301
    8.1.1 Distance Between Sequences ......................... 302
    8.1.2 Distance Between Vectors and Sets of Vectors ..... 302
    8.1.3 Hellinger Distance ................................ 304
  8.2 Divergences and Directed Divergences .................... 304
    8.2.1 Kullbach-Leibler’s Directed Divergence ............ 305
    8.2.2 Jeffreys’ Divergence ................................ 305
    8.2.3 Bhattacharyya Divergence ........................... 306
    8.2.4 Matsushita Divergence .............................. 307
    8.2.5 F-Divergence ...................................... 308
    8.2.6 δ-Divergence ...................................... 309
    8.2.7 $\chi^\alpha$ Directed Divergence .................... 310
References .................................................. 310

9 Decision Theory .............................................. 313
  9.1 Hypothesis Testing ...................................... 313
  9.2 Bayesian Decision Theory ............................... 316
    9.2.1 Binary Hypothesis ................................ 320
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.2.2</td>
<td>Relative Information and Log Likelihood Ratio</td>
<td>321</td>
</tr>
<tr>
<td>9.3</td>
<td>Bayesian Classifier</td>
<td>322</td>
</tr>
<tr>
<td>9.3.1</td>
<td>Multi-Dimensional Normal Classification</td>
<td>326</td>
</tr>
<tr>
<td>9.3.2</td>
<td>Classification of a Sequence</td>
<td>328</td>
</tr>
<tr>
<td>9.4</td>
<td>Decision Trees</td>
<td>331</td>
</tr>
<tr>
<td>9.4.1</td>
<td>Tree Construction</td>
<td>332</td>
</tr>
<tr>
<td>9.4.2</td>
<td>Types of Questions</td>
<td>333</td>
</tr>
<tr>
<td>9.4.3</td>
<td>Maximum Likelihood Estimation (MLE)</td>
<td>336</td>
</tr>
<tr>
<td>References</td>
<td></td>
<td>338</td>
</tr>
<tr>
<td>10</td>
<td>Parameter Estimation</td>
<td>341</td>
</tr>
<tr>
<td>10.1</td>
<td>Maximum Likelihood Estimation</td>
<td>342</td>
</tr>
<tr>
<td>10.2</td>
<td>Maximum A-Posteriori (MAP) Estimation</td>
<td>344</td>
</tr>
<tr>
<td>10.3</td>
<td>Maximum Entropy Estimation</td>
<td>345</td>
</tr>
<tr>
<td>10.4</td>
<td>Minimum Relative Entropy Estimation</td>
<td>346</td>
</tr>
<tr>
<td>10.5</td>
<td>Maximum Mutual Information Estimation (MMIE)</td>
<td>348</td>
</tr>
<tr>
<td>10.6</td>
<td>Model Selection</td>
<td>349</td>
</tr>
<tr>
<td>10.6.1</td>
<td>Akaike Information Criterion (AIC)</td>
<td>350</td>
</tr>
<tr>
<td>10.6.2</td>
<td>Bayesian Information Criterion (BIC)</td>
<td>353</td>
</tr>
<tr>
<td>References</td>
<td></td>
<td>354</td>
</tr>
<tr>
<td>11</td>
<td>Unsupervised Clustering and Learning</td>
<td>357</td>
</tr>
<tr>
<td>11.1</td>
<td>Vector Quantization (VQ)</td>
<td>358</td>
</tr>
<tr>
<td>11.2</td>
<td>Basic Clustering Techniques</td>
<td>359</td>
</tr>
<tr>
<td>11.2.1</td>
<td>Standard k-Means (Lloyd) Algorithm</td>
<td>360</td>
</tr>
<tr>
<td>11.2.2</td>
<td>Generalized Clustering</td>
<td>363</td>
</tr>
<tr>
<td>11.2.3</td>
<td>Overpartitioning</td>
<td>364</td>
</tr>
<tr>
<td>11.2.4</td>
<td>Merging</td>
<td>364</td>
</tr>
<tr>
<td>11.2.5</td>
<td>Modifications to the k-Means Algorithm</td>
<td>365</td>
</tr>
<tr>
<td>11.2.6</td>
<td>k-Means Wrappers</td>
<td>368</td>
</tr>
<tr>
<td>11.2.7</td>
<td>Rough k-Means</td>
<td>375</td>
</tr>
<tr>
<td>11.2.8</td>
<td>Fuzzy k-Means</td>
<td>377</td>
</tr>
<tr>
<td>11.2.9</td>
<td>k-Harmonic Means Algorithm</td>
<td>378</td>
</tr>
<tr>
<td>11.2.10</td>
<td>Hybrid Clustering Algorithms</td>
<td>380</td>
</tr>
<tr>
<td>11.3</td>
<td>Estimation using Incomplete Data</td>
<td>381</td>
</tr>
<tr>
<td>11.3.1</td>
<td>Expectation Maximization (EM)</td>
<td>381</td>
</tr>
<tr>
<td>11.4</td>
<td>Hierarchical Clustering</td>
<td>388</td>
</tr>
<tr>
<td>11.4.1</td>
<td>Agglomerative (Bottom-Up) Clustering (AHC)</td>
<td>389</td>
</tr>
<tr>
<td>11.4.2</td>
<td>Divisive (Top-Down) Clustering (DHC)</td>
<td>389</td>
</tr>
<tr>
<td>11.5</td>
<td>Semi-Supervised Learning</td>
<td>390</td>
</tr>
<tr>
<td>References</td>
<td></td>
<td>390</td>
</tr>
</tbody>
</table>
## 12 Transformation

12.1 Principal Component Analysis (PCA) ................................................. 394
   12.1.1 Formulation ........................................................................ 394
12.2 Generalized Eigenvalue Problem ...................................................... 397
12.3 Nonlinear Component Analysis ......................................................... 399
   12.3.1 Kernel Principal Component Analysis (Kernel PCA) .............. 400
12.4 Linear Discriminant Analysis (LDA) ............................................... 401
   12.4.1 Integrated Mel Linear Discriminant Analysis (IMELDA) ....... 404
12.5 Factor Analysis ............................................................................. 404

References ......................................................................................... 409

## 13 Hidden Markov Modeling (HMM)

13.1 Memoryless Models .................................................................. 413
13.2 Discrete Markov Chains ............................................................... 415
13.3 Markov Models ............................................................................ 416
13.4 Hidden Markov Models ................................................................. 418
13.5 Model Design and States ............................................................... 421
13.6 Training and Decoding ................................................................. 423
   13.6.1 Trellis Diagram Representation ........................................... 428
   13.6.2 Forward Pass Algorithm ....................................................... 430
   13.6.3 Viterbi Algorithm ................................................................. 432
   13.6.4 Baum-Welch (Forward-Backward) Algorithm .................... 433
13.7 Gaussian Mixture Models (GMM) .................................................... 442
   13.7.1 Training ............................................................................. 444
   13.7.2 Tractability of Models ........................................................... 449
13.8 Practical Issues ............................................................................. 451
   13.8.1 Smoothing ......................................................................... 451
   13.8.2 Model Comparison ............................................................... 453
   13.8.3 Held-Out Estimation ............................................................. 456
   13.8.4 Deleted Estimation ............................................................... 461

References ......................................................................................... 462

## 14 Neural Networks

14.1 Perceptron .................................................................................... 466
14.2 Feedforward Networks ................................................................. 466
   14.2.1 Auto Associative Neural Networks (AANN) ....................... 469
   14.2.2 Radial Basis Function Neural Networks (RBFNN) ............. 469
   14.2.3 Training (Learning) Formulation ........................................ 470
   14.2.4 Optimization Problem ......................................................... 473
   14.2.5 Global Solution .................................................................. 474
14.3 Recurrent Neural Networks (RNN) ................................................ 476
14.4 Time-Delay Neural Networks (TDNNs) .......................................... 477
14.5 Hierarchical Mixtures of Experts (HME) ....................................... 479
14.6 Practical Issues ............................................................................. 479

References ......................................................................................... 481
15 Support Vector Machines ............................................... 485
  15.1 Risk Minimization .................................................. 488
    15.1.1 Empirical Risk Minimization ......................... 492
    15.1.2 Capacity and Bounds on Risk ...................... 493
    15.1.3 Structural Risk Minimization ...................... 493
  15.2 The Two-Class Problem ........................................ 494
    15.2.1 Dual Representation ................................ 497
    15.2.2 Soft Margin Classification ....................... 500
  15.3 Kernel Mapping .................................................. 503
    15.3.1 The Kernel Trick .................................... 504
  15.4 Positive Semi-Definite Kernels ............................... 506
    15.4.1 Linear Kernel ..................................... 506
    15.4.2 Polynomial Kernel ................................. 506
    15.4.3 Gaussian Radial Basis Function (GRBF) Kernel .... 507
    15.4.4 Cosine Kernel .................................... 508
    15.4.5 Fisher Kernel .................................... 508
    15.4.6 GLDS Kernel ..................................... 509
    15.4.7 GMM-UBM Mean Interval (GUMI) Kernel ............ 510
  15.5 Non Positive Semi-Definite Kernels .............................. 511
    15.5.1 Jeffreys Divergence Kernel ........................... 511
    15.5.2 Fuzzy Hyperbolic Tangent (tanh) Kernel ........... 512
    15.5.3 Neural Network Kernel ............................ 513
  15.6 Kernel Normalization .......................................... 513
  15.7 Kernel Principal Component Analysis (Kernel PCA) ............ 514
  15.8 Nuisance Attribute Projection (NAP) ........................... 516
  15.9 The multiclass ($\Gamma$-Class) Problem ..................... 518
References ..................................................................... 519

Part II Advanced Theory

16 Speaker Modeling ..................................................... 525
  16.1 Individual Speaker Modeling .................................... 526
  16.2 Background Models and Cohorts ................................ 527
    16.2.1 Background Models ................................ 528
    16.2.2 Cohorts ........................................... 529
  16.3 Pooling of Data and Speaker Independent Models ............. 529
  16.4 Speaker Adaptation .............................................. 530
    16.4.1 Factor Analysis (FA) ................................ 530
    16.4.2 Joint Factor Analysis (JFA) ....................... 531
    16.4.3 Total Factors (Total Variability) .................. 532
  16.5 Audio Segmentation .............................................. 532
  16.6 Model Quality Assessment ..................................... 534
    16.6.1 Enrollment Utterance Quality Control ............. 534
    16.6.2 Speaker Menagerie .................................. 536
References ..................................................................... 538
## 17 Speaker Recognition

17.1 The Enrollment Task ............................................. 543
17.2 The Verification Task ............................................. 544
  17.2.1 Text-Dependent ............................................... 546
  17.2.2 Text-Prompted ............................................... 546
  17.2.3 Knowledge-Based ............................................. 548
17.3 The Identification Task ............................................. 548
  17.3.1 Closed-Set Identification .................................... 548
  17.3.2 Open-Set Identification ..................................... 549
17.4 Speaker Segmentation ............................................. 549
17.5 Speaker and Event Classification ................................. 550
  17.5.1 Gender and Age Classification (Identification) ............ 551
  17.5.2 Audio Classification ........................................ 553
  17.5.3 Multiple Codebooks ........................................ 553
  17.5.4 Farfield Speaker Recognition ................................ 553
  17.5.5 Whispering Speaker Recognition .............................. 554
17.6 Speaker Diarization ................................................ 554
  17.6.1 Speaker Position and Orientation ............................ 555

References .............................................................. 555

## 18 Signal Enhancement and Compensation

18.1 Silence Detection, Voice Activity Detection (VAD) ............... 561
18.2 Audio Volume Estimation .......................................... 564
18.3 Echo Cancellation .................................................. 564
18.4 Spectral Filtering and Cepstral Lifting ............................. 565
  18.4.1 Cepstral Mean Normalization (Subtraction) – CMN (CMS) .... 567
  18.4.2 Cepstral Mean and Variance Normalization (CMVN) .......... 569
  18.4.3 Cepstral Histogram Normalization (Histogram Equalization) ................................. 570
  18.4.4 RelAtive SpecTrAl (RASTA) Filtering .......................... 571
  18.4.5 Other Lifters .................................................... 571
  18.4.6 Vocal Tract Length Normalization (VTLN) ...................... 573
  18.4.7 Other Normalization Techniques ............................... 576
  18.4.8 Steady Tone Removal (Narrowband Noise Reduction) ......... 579
  18.4.9 Adaptive Wiener Filtering ...................................... 580
18.5 Speaker Model Normalization ...................................... 581
  18.5.1 Z-Norm ......................................................... 581
  18.5.2 T-Norm (Test Norm) .......................................... 582
  18.5.3 H-Norm ......................................................... 582
  18.5.4 HT-Norm ......................................................... 582
  18.5.5 AT-Norm ......................................................... 582
  18.5.6 C-Norm ......................................................... 582
  18.5.7 D-Norm ......................................................... 583
  18.5.8 F-Norm (F-Ratio Normalization) ............................... 583
  18.5.9 Group-Specific Normalization .................................. 583
Part III Practice

19 Evaluation and Representation of Results .............................. 589
  19.1 Verification Results .............................................. 589
    19.1.1 Equal-Error Rate ........................................ 589
    19.1.2 Half Total Error Rate ...................................... 590
    19.1.3 Receiver Operating Characteristic (ROC) Curve ............ 590
    19.1.4 Detection Error Trade-Off (DET) Curve ................. 592
    19.1.5 Detection Cost Function (DCF) ......................... 593
  19.2 Identification Results .......................................... 593
References .......................................................... 594

20 Time Lapse Effects (Case Study) .................................. 595
  20.1 The Audio Data .................................................. 598
  20.2 Baseline Speaker Recognition .................................. 598
References .......................................................... 600

21 Adaptation over Time (Case Study) ................................ 601
  21.1 Data Augmentation .............................................. 601
  21.2 Maximum A Posteriori (MAP) Adaptation ...................... 603
  21.3 Eigenvoice Adaptation ......................................... 605
  21.4 Minimum Classification Error (MCE) ........................... 605
  21.5 Linear Regression Techniques ................................. 606
    21.5.1 Maximum Likelihood Linear Regression (MLLR) ........... 606
  21.6 Maximum a-Posteriori Linear Regression (MAPLR) ............ 607
    21.6.1 Other Adaptation Techniques ............................ 607
  21.7 Practical Perspectives ......................................... 607
References .......................................................... 608

22 Overall Design .......................................................... 611
  22.1 Choosing the Model ............................................. 611
    22.1.1 Phonetic Speaker Recognition ............................ 612
  22.2 Choosing an Adaptation Technique ............................. 613
  22.3 Microphones .................................................... 613
  22.4 Channel Mismatch ............................................... 615
  22.5 Voice Over Internet Protocol (VoIP) .......................... 615
  22.6 Public Databases .............................................. 616
    22.6.1 NIST .................................................. 616
    22.6.2 Linguistic Data Consortium (LDC) ....................... 616
    22.6.3 European Language Resources Association (ELRA) ....... 619
  22.7 High Level Information ......................................... 620
    22.7.1 Choosing Basic Segments ................................. 622
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>24.6</td>
<td>Fourier Series Expansion</td>
<td>708</td>
</tr>
<tr>
<td>24.6.1</td>
<td>Convergence of the Fourier Series</td>
<td>713</td>
</tr>
<tr>
<td>24.6.2</td>
<td>Parseval’s Theorem</td>
<td>714</td>
</tr>
<tr>
<td>24.7</td>
<td>Wavelet Series Expansion</td>
<td>716</td>
</tr>
<tr>
<td>24.8</td>
<td>The Laplace Transform</td>
<td>717</td>
</tr>
<tr>
<td>24.8.1</td>
<td>Inversion</td>
<td>720</td>
</tr>
<tr>
<td>24.8.2</td>
<td>Some Useful Transforms</td>
<td>721</td>
</tr>
<tr>
<td>24.9</td>
<td>Complex Fourier Transform (Fourier Integral Transform)</td>
<td>722</td>
</tr>
<tr>
<td>24.9.1</td>
<td>Translation</td>
<td>724</td>
</tr>
<tr>
<td>24.9.2</td>
<td>Scaling</td>
<td>724</td>
</tr>
<tr>
<td>24.9.3</td>
<td>Symmetry Table</td>
<td>724</td>
</tr>
<tr>
<td>24.9.4</td>
<td>Time and Complex Scaling and Shifting</td>
<td>725</td>
</tr>
<tr>
<td>24.9.5</td>
<td>Convolution</td>
<td>725</td>
</tr>
<tr>
<td>24.9.6</td>
<td>Correlation</td>
<td>726</td>
</tr>
<tr>
<td>24.9.7</td>
<td>Parseval’s Theorem</td>
<td>726</td>
</tr>
<tr>
<td>24.9.8</td>
<td>Power Spectral Density</td>
<td>728</td>
</tr>
<tr>
<td>24.9.9</td>
<td>One-Sided Power Spectral Density</td>
<td>728</td>
</tr>
<tr>
<td>24.9.10</td>
<td>PSD-per-unit-time</td>
<td>729</td>
</tr>
<tr>
<td>24.9.11</td>
<td>Wiener-Khintchine Theorem</td>
<td>729</td>
</tr>
<tr>
<td>24.10</td>
<td>Discrete Fourier Transform (DFT)</td>
<td>731</td>
</tr>
<tr>
<td>24.10.1</td>
<td>Inverse Discrete Fourier Transform (IDFT)</td>
<td>732</td>
</tr>
<tr>
<td>24.10.2</td>
<td>Periodicity</td>
<td>734</td>
</tr>
<tr>
<td>24.10.3</td>
<td>Plancherel and Parseval’s Theorem</td>
<td>734</td>
</tr>
<tr>
<td>24.10.4</td>
<td>Power Spectral Density (PSD) Estimation</td>
<td>735</td>
</tr>
<tr>
<td>24.10.5</td>
<td>Fast Fourier Transform (FFT)</td>
<td>736</td>
</tr>
<tr>
<td>24.11</td>
<td>Discrete-Time Fourier Transform (DTFT)</td>
<td>738</td>
</tr>
<tr>
<td>24.11.1</td>
<td>Power Spectral Density (PSD) Estimation</td>
<td>739</td>
</tr>
<tr>
<td>24.12</td>
<td>Complex Short-Time Fourier Transform (STFT)</td>
<td>740</td>
</tr>
<tr>
<td>24.12.1</td>
<td>Discrete-Time Short-Time Fourier Transform DTSTFT</td>
<td>744</td>
</tr>
<tr>
<td>24.12.2</td>
<td>Discrete Short-Time Fourier Transform DSTFT</td>
<td>746</td>
</tr>
<tr>
<td>24.13</td>
<td>Discrete Cosine Transform (DCT)</td>
<td>748</td>
</tr>
<tr>
<td>24.13.1</td>
<td>Efficient DCT Computation</td>
<td>749</td>
</tr>
<tr>
<td>24.14</td>
<td>The z-Transform</td>
<td>750</td>
</tr>
<tr>
<td>24.14.1</td>
<td>Translation</td>
<td>756</td>
</tr>
<tr>
<td>24.14.2</td>
<td>Scaling</td>
<td>756</td>
</tr>
<tr>
<td>24.14.3</td>
<td>Shifting – Time Lag</td>
<td>757</td>
</tr>
<tr>
<td>24.14.4</td>
<td>Shifting – Time Lead</td>
<td>757</td>
</tr>
<tr>
<td>24.14.5</td>
<td>Complex Translation</td>
<td>757</td>
</tr>
<tr>
<td>24.14.6</td>
<td>Initial Value Theorem</td>
<td>758</td>
</tr>
<tr>
<td>24.14.7</td>
<td>Final Value Theorem</td>
<td>758</td>
</tr>
<tr>
<td>24.14.8</td>
<td>Real Convolution Theorem</td>
<td>759</td>
</tr>
<tr>
<td>24.14.9</td>
<td>Inversion</td>
<td>760</td>
</tr>
<tr>
<td>24.15</td>
<td>Cepstrum</td>
<td>762</td>
</tr>
<tr>
<td>References</td>
<td></td>
<td>769</td>
</tr>
</tbody>
</table>
## 25 Nonlinear Optimization

25.1 Gradient-Based Optimization .................................. 775
  25.1.1 The Steepest Descent Technique ............................. 775
  25.1.2 Newton’s Minimization Technique ............................ 777
  25.1.3 Quasi-Newton or Large Step Gradient Techniques .......... 779
  25.1.4 Conjugate Gradient Methods ................................ 793

25.2 Gradient-Free Optimization ..................................... 803
  25.2.1 Search Methods .......................................... 804
  25.2.2 Gradient-Free Conjugate Direction Methods ................. 804

25.3 The Line Search Sub-Problem ................................... 809

25.4 Practical Considerations ....................................... 810
  25.4.1 Large-Scale Optimization .................................. 810
  25.4.2 Numerical Stability ....................................... 813
  25.4.3 Nonsmooth Optimization ................................... 814

25.5 Constrained Optimization ....................................... 814
  25.5.1 The Lagrangian and Lagrange Multipliers .................. 817
  25.5.2 Duality .................................................. 831

25.6 Global Convergence ............................................. 835

References ............................................................ 836

## 26 Standards

26.1 Standard Audio Formats ......................................... 842
  26.1.1 Linear PCM (Uniform PCM) ................................ 842
  26.1.2 µ-Law PCM (PCMU) ........................................ 843
  26.1.3 A-Law (PCMA) ............................................. 843
  26.1.4 MP3 ...................................................... 843
  26.1.5 HE-AAC .................................................. 844
  26.1.6 OGG Vorbis .............................................. 844
  26.1.7 ADPCM (G.726) ......................................... 845
  26.1.8 GSM ..................................................... 845
  26.1.9 CELP .................................................... 847
  26.1.10 DTMF ................................................... 848
  26.1.11 Others Audio Formats .................................... 848

26.2 Standard Audio Encapsulation Formats ......................... 849
  26.2.1 WAV ..................................................... 849
  26.2.2 SPHERE .................................................. 850
  26.2.3 Standard Audio Format Encapsulation (SAFE) ............... 850

26.3 APIs and Protocols .............................................. 854
  26.3.1 SVAPI .................................................. 855
  26.3.2 BioAPI .................................................. 855
  26.3.3 VoiceXML ............................................... 856
  26.3.4 MRCP .................................................. 857
  26.3.5 Real-time Transport Protocol (RTP) ......................... 858
  26.3.6 Extensible MultiModal Annotation (EMMA) .................. 858

References ............................................................ 859
Fundamentals of Speaker Recognition
Beigi, H.
2011, LXI, 942 p., Hardcover