Index

A
Adaptive network based fuzzy inference system (ANFIS), 43
Admissible wavelet packet (AWP) trees, 48
Air-filled lungs, 5
AM-based method, 49
AM-FM modulation/demodulation approach, 49
Applications in speech technology, 2
Arabic digits, 44
Area function, 7
Artificial neural network (ANN), 31
Audio coding, 59
Audio recordings of speech, 77
Audio signal, 71
Audio steganography, 71
Automatic formant tracking, 24
Automatic speech recognition (ASR)
classification/recognition, 41
DWT and WPD, 41
filter-bank selection, 43
hybrid enhancement, 44
MFCC features, 44
neural networks, 44
noise robustness, 44
performance, 43
stage and feature extraction, 44
systems, 3, 65
wavelet features, 42
WP features, 43
WT feature extraction, 44
AWP-based speaker identification system, 48

B
Back-propagation (BP) NN, 54
Bandwidths (BW), 62
Bark scale, 43, 52
Bark wavelet, 43
Battle-Lemarié wavelet concentrates, 57
Battle-Lemarié wavelet function, 48
Bayesian scheme, 43
BayesShrink, 37
Biased risk thresholding, 37

C
Candidate pitches, 26
Chaotic components in speech signal, 67–69
Chaotic logistic mapping, 72
Classification error, 78
Coding-decoding (CODEC) of speech, 57, 59
Communication systems, 63
Confusion matrix, 79
Continuous wavelet transform (CWT), 19, 43,
58, 67, 68
Cosine distance scoring (CDS), 49
Critical-band (CB), 8, 31
Critical-band wavelet decomposition, 37
Cross-validation tests, 79

D
Data hiding, 73
Daubechies and Daubechies Lifting Scheme
DWT (DLSDWT), 59
Daubechies family, 68
Daubechies wavelets, 16, 57
DFT, 48
Diagnose vocal pathologies
audio recordings of speech, 77
characteristics of normal and pathological voices, 79
clinical symptoms, 77
CWT, 77
dysfluent speech signals, 78
energy distribution, 77
FD, 78
Mel-scaled WPT, 78
MRSTC method, 79
nonlinear fractal dimension of DWT, 78
objective evaluation, 78
optimum features, 78
pitch detection, 79
speech signal, 77
STFT, 77
strong high-frequency components, 77
SVM-based classification, 77
SVM classifier, 78
voice disorders, 77, 78
voice source, 78
wavelet families, 78
Wavelet-based orthogonal filter banks, 78
WPT, 77
WT coefficients, 77
Digital signal processing, 3
Digital signal processor (DSP), 59
Digital watermarking, 73
Discrete cosine transform (DCT), 48
Discrete wavelet-Fourier transform (DWFT), 49
Discrete wavelet transform (DWT), 14, 18, 23, 25, 26, 48, 52, 63, 65, 72
Discrete-time domain, 14
Dual-tree complex WT, 41
DWP, 48
DWT-based speech compression, 58–59
Dysfluent speech signals, 78

E
Emotion recognition, 66
classification, 51, 53
detecting, 51
frequency-based decomposition, 53
LPCC, 53
MFCC, 53
NLD features, 53
RBFNN, 53
speech signal, 53
synchro-squeezed wavelet transform, 53
UMB, 53
voiced and unvoiced segments, 53
wavelet-based features, 51–52
WNN, 54
WPT, 53
WT, 53
Emotional database, 52
Empirical wavelets, 19–20
Empirical wavelet transform (EWT), 19
Encryption, 71, 72
Energy distribution, 77
Enhanced speaker recognition model, 75
Equivalent rectangular bandwidth (ERB) scale, 52
Estimating mixing matrix, 32
European Telecommunication Standards
Institution (ETSI) adaptive multi-rate (AMR) narrow-band (NB) (ETSI AMR-NB), 29
Event-based detection of pitch, 26
Expectation maximization (EM) algorithm, 32

F
Fast Fourier transform (FFT), 18
Field programmable gate array (FPGA), 59
Filter-bank structure, 51
Forensic analysis, 74–75
Forensic sciences, 74
Forensic voice, 74
Formant estimation, 24, 25
Formant frequencies, 66
Formant model, 8
Formant tracking, 24, 25, 66
Formants, 49
Formants, wavelet entropy and neural networks (FWENN), 49
Fourier-based algorithms, 3
Fourier-based spectrogram, 24
Fourier transform (FT), 11
FPGA-based design, 59
Fractal dimension (FD), 65, 78
Frequency-based decomposition, 53
Fundamental frequency estimation, 25, 26

G
Gamma-tone wavelet, 42
Gaussian mixture model (GMM), 48, 51, 53, 67
Gaussian noise, 66
<table>
<thead>
<tr>
<th>Term</th>
<th>Page(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genetic algorithm</td>
<td>48, 78</td>
</tr>
<tr>
<td>Gray scales</td>
<td>66</td>
</tr>
<tr>
<td><strong>H</strong></td>
<td></td>
</tr>
<tr>
<td>Haar scaling</td>
<td>16</td>
</tr>
<tr>
<td>Harmonic analysis</td>
<td>67</td>
</tr>
<tr>
<td>Heisenberg uncertainty principle</td>
<td>12, 13</td>
</tr>
<tr>
<td>Hidden data</td>
<td>73</td>
</tr>
<tr>
<td>Hidden Markov model (HMM)</td>
<td>44</td>
</tr>
<tr>
<td>Hiding a speech signal</td>
<td>72</td>
</tr>
<tr>
<td>Higher-frequency components</td>
<td>75</td>
</tr>
<tr>
<td>High-frequency band</td>
<td>75</td>
</tr>
<tr>
<td>Human auditory system</td>
<td>8, 9, 72</td>
</tr>
<tr>
<td>Hybrid technique, speaker identification</td>
<td>49</td>
</tr>
<tr>
<td><strong>I</strong></td>
<td></td>
</tr>
<tr>
<td>IDWT</td>
<td>59</td>
</tr>
<tr>
<td>Independent component analysis (ICA)</td>
<td>31, 32</td>
</tr>
<tr>
<td>i-vectors</td>
<td>75</td>
</tr>
<tr>
<td><strong>L</strong></td>
<td></td>
</tr>
<tr>
<td>Laplacian mixture model (LMM)</td>
<td>32</td>
</tr>
<tr>
<td>Largest Lyapunov exponent (LLE)</td>
<td>67, 69</td>
</tr>
<tr>
<td>Learned dictionary</td>
<td>74</td>
</tr>
<tr>
<td>Least significant bits (LSB)</td>
<td>71</td>
</tr>
<tr>
<td>Line spectral frequencies (LSF)</td>
<td>63</td>
</tr>
<tr>
<td>Linear discriminant analysis (LDA)</td>
<td>52</td>
</tr>
<tr>
<td>Linear prediction (LP) analysis</td>
<td>58</td>
</tr>
<tr>
<td>Linear prediction cepstral coefficients (LPCC)</td>
<td>53</td>
</tr>
<tr>
<td>Linear prediction (LP) model</td>
<td>8</td>
</tr>
<tr>
<td>Linear predictive code (LPC)</td>
<td>23, 25, 58</td>
</tr>
<tr>
<td>Low-frequency band</td>
<td>74</td>
</tr>
<tr>
<td>LPC-based synthesizer</td>
<td>57</td>
</tr>
<tr>
<td>LSB substitution</td>
<td>71</td>
</tr>
<tr>
<td><strong>M</strong></td>
<td></td>
</tr>
<tr>
<td>Mean opinion scores (MOS)</td>
<td>63</td>
</tr>
<tr>
<td>MEL scale</td>
<td>42</td>
</tr>
<tr>
<td>Mel-frequency cepstral coefficients (MFCC)</td>
<td>23, 42–44, 47–49, 51, 53</td>
</tr>
<tr>
<td>analysis</td>
<td>63</td>
</tr>
<tr>
<td>features</td>
<td>66</td>
</tr>
<tr>
<td>Mel-scaled WPT</td>
<td>78</td>
</tr>
<tr>
<td>Mel scale based filter structures</td>
<td>48</td>
</tr>
<tr>
<td>Meyer wavelets</td>
<td>16, 57</td>
</tr>
<tr>
<td>Mixed-excitation linear prediction (MELP) algorithm</td>
<td>72</td>
</tr>
<tr>
<td>Morlet wavelet</td>
<td>67</td>
</tr>
<tr>
<td>Mother wavelet</td>
<td>13, 52</td>
</tr>
<tr>
<td><strong>Multiresolution analysis (MRA)</strong></td>
<td>11–14, 52</td>
</tr>
<tr>
<td><strong>Multiresolution auditory model (MRAM)</strong></td>
<td>63</td>
</tr>
<tr>
<td><strong>Multiresolution filter bank</strong></td>
<td>75</td>
</tr>
<tr>
<td><strong>Multiresolution sinusoidal transform coding (MRSTC) method</strong></td>
<td>79</td>
</tr>
<tr>
<td><strong>Multitaper spectrum (MTS)</strong></td>
<td>38, 39</td>
</tr>
<tr>
<td><strong>N</strong></td>
<td></td>
</tr>
<tr>
<td>Nasalized sounds</td>
<td>7</td>
</tr>
<tr>
<td>Nasal tract</td>
<td>7</td>
</tr>
<tr>
<td>Negative selection algorithm</td>
<td>41</td>
</tr>
<tr>
<td>Neural network (NN)</td>
<td>29, 44</td>
</tr>
<tr>
<td><strong>Neural network with radial basis function (RBFNN)</strong></td>
<td>53</td>
</tr>
<tr>
<td><strong>NMF-estimated noise</strong></td>
<td>38</td>
</tr>
<tr>
<td>Noise cancellation</td>
<td>41–42</td>
</tr>
<tr>
<td>Noise suppression</td>
<td>35</td>
</tr>
<tr>
<td><strong>Nonlinear components in speech signal</strong></td>
<td>67–69</td>
</tr>
<tr>
<td><strong>Nonlinear dynamics (NLD) features</strong></td>
<td>53</td>
</tr>
<tr>
<td><strong>Nonlinear features</strong></td>
<td>65–66</td>
</tr>
<tr>
<td><strong>Nonlinear signal analysis</strong></td>
<td>67</td>
</tr>
<tr>
<td><strong>Nonnegative matrix factorization (NMF)</strong></td>
<td>38</td>
</tr>
<tr>
<td><strong>Non-orthogonal wavelet decomposition</strong></td>
<td>43</td>
</tr>
<tr>
<td><strong>Nonuniform subbands</strong></td>
<td>75</td>
</tr>
<tr>
<td><strong>O</strong></td>
<td></td>
</tr>
<tr>
<td>Optimizing wavelet parameters</td>
<td>42</td>
</tr>
<tr>
<td>Optimum wavelets</td>
<td>57</td>
</tr>
<tr>
<td>Optimum WP features</td>
<td>48</td>
</tr>
<tr>
<td><strong>P</strong></td>
<td></td>
</tr>
<tr>
<td>Packet-based end-to-end system</td>
<td>63</td>
</tr>
<tr>
<td>Peak detection approach</td>
<td>26</td>
</tr>
<tr>
<td>Perception-based approach</td>
<td>9</td>
</tr>
<tr>
<td><strong>Perceptual quality measures</strong></td>
<td>9</td>
</tr>
<tr>
<td><strong>Perceptual wavelet filter bank (PWF)</strong></td>
<td>37</td>
</tr>
<tr>
<td><strong>Perceptual wavelet packet decomposition (PWPD)</strong></td>
<td>38</td>
</tr>
<tr>
<td><strong>Phase-based features</strong></td>
<td>65</td>
</tr>
<tr>
<td>Pitch detection</td>
<td>25, 26, 79</td>
</tr>
<tr>
<td>Pitch estimation</td>
<td>23, 25–27</td>
</tr>
<tr>
<td><strong>Predefined wavelet-packet (PWP)</strong></td>
<td></td>
</tr>
<tr>
<td>analysis</td>
<td>63</td>
</tr>
<tr>
<td>decomposition tree</td>
<td>17, 62</td>
</tr>
<tr>
<td>equivalent critical band</td>
<td>62</td>
</tr>
<tr>
<td>filter bank</td>
<td>62</td>
</tr>
<tr>
<td>Psychoacoustic models</td>
<td>9, 42</td>
</tr>
<tr>
<td><strong>Q</strong></td>
<td></td>
</tr>
<tr>
<td>Quasi-continuous WT</td>
<td>18</td>
</tr>
</tbody>
</table>
R
Real-time implementation of DWT-based speech compression, 58, 59
Reconstructed phase space (RPS), 65
Recording session variability, 74
Recurrence plot (RP), 65
Redundant wavelet filter banks (RWFB), 43
Redundant WT, 18
Resonant-like systems, 73

S
Scale index, 67–69
Scale parameter, 66
Scalogram, 19, 23, 24
Scalogram analysis, 66–67
Scalogram-based speaker verification system, 67
Secret signal, 71
Secret speech, 72
Secure communication of speech, 71–73
Segmental SNR (SSNR), 9
Segmentation of speech, 30
Self-organizing map (SOM), 31
Semi-discrete WT, 13
Semisoft mask filter, 38
Semisoft shrinking, 37
Shannon entropy, 49, 79
Short-time Fourier transform (STFT), 3, 11–13, 23
Signal enhancement, 41–42
Signal-to-noise ratio (SNR), 9, 29
Sophisticated thresholding, 37
Sound source generation, 7
Source-filter model, 5
Source separation, 31–33
Sparse representation, 73–74
Speaker identification
characteristics, 47
hybrid feature sets, 49
MFCC features, 47
recognition accuracy, 47
speaker recognition, 47
speaker verification, 47
speech recognition, 47
wavelet-based features, 48
Speaker recognition, 47–49, 74
Speaker-specific information, 74, 75
Speaker’s voice, 74
Speaker verification, 47–49, 67
Spectral analysis, 8, 23, 24
Spectral-domain (nonevent)-based pitch detectors, 25, 26
Spectrogram, 23
ASR (see Automatic speech recognition (ASR))
Bayesian scheme, 43
HMM, 44
hybrid system, 44
noise cancellation, 41–42
signal enhancement, 41–42
SVM, 44
system, 66
time-adapted hybrid wavelet speech enhancement, 44
wavelet-based features, 42–43
WNN, 44
Speech segmentation, 30, 31
Speech sentences, 35
Speech signal, 30–31
Speech sounds, 6–7
Speech synthesis, 57, 58
Stationary WT, 18
Steganano signal, 71
Steganano speech, 72
Steganography, 71, 73, 74
Stein’s unbiased risk estimate (SURE), 37
Subbands, 15, 58, 59
Support vector machine (SVM), 29, 30, 44
SURE entropy, 48
SureShrink, 37
SVD, 73
SVM-based classification, 77
SVM-based emotion recognition, 52
SVM classifier, 65, 78
Synchro-squeezed wavelet transform, 53

T
Teager energy operator, 37, 75
Terminal analog, 8
Thresholding (shrink) approach, 36
Thresholding schemes
BayesShrink, 37
hard-thresholding function, 36, 37
soft thresholding, 36, 37
SURE, 37
SureShrink, 37
Teager energy operator, 37
UWT, 37
VisuShrink, 37
wavelet coefficient, 37
wavelet packet coefficients, 37–38
Time-domain (event)-based pitch
detectors, 25, 26
Time-domain methods, 71, 73
Time-frequency analysis, 12, 13, 16, 18, 29, 30
Time-frequency domain, 23, 24
Time-scale distribution, 66
Time variance property, 43
Traditional cepstral analysis, 24
Transform domain, 71
Two-channel filter bank, 14

U
Undecimated wavelet-based speech coding strategy, 58
Undecimated wavelet packet decomposition (UWPD), 31, 32
Undecimated wavelet-packet transform (UWPT), 18, 58
Undecimated wavelet transform (UWT), 18, 37
Uniform WPD, 75
Universal-background model (UMB), 53
Unvoiced sounds, 7

V
Variance analysis (VA), 26
VisuShrink, 37
Vocal fold-related voice disorders, 77
Vocal organs, 5
Vocal tract
anatomical structure, 5, 6
cross-sectional area (c.s.a.), 7
Voice activity detection (VAD), 29–30
Voice disorders, 78
Voiced sounds
source, 7
and vowels, 6
Voice features, 75
Voice source, 65, 78
Voice/unvoiced (V/U) classification, 23
VoIP applications, 57

W
Watermarking
authentication and protecting copyright, 71
in sparse representation, 73–74
speech, 73
Waveform coding, 58
Wavelet
emotion recognition, 51–52
neural networks in ASR, 44
scalogram analysis, 66–67
Wavelet analysis, 5, 9, 23, 26
Wavelet-based bark coherence function (WBCF), 63
Wavelet-based features, 42–43
Wavelet-based nonlinear features, 66
Wavelet-based orthogonal filter banks, 78
Wavelet cepstral coefficients (WCC), 49, 75
Wavelet coefficients, 71, 72
Wavelet denoising, 35, 38, 39
Wavelet energy, 29
Wavelet family, 15, 16, 78
Wavelet filter, 14, 15, 19
Wavelet multiresolution analysis, 8
Wavelet octave coefficients of linear-prediction residues (WOCOR), 48
Wavelet packet decomposition (WPD), 18
Wavelet packets (WP)
  analysis, 42
  frequency band, 63
  MRAM, 63
  objective quality measure, 61
  PWP, 62
  tree structure, 61
  wavelet family, 63
coefficients, 37–38
conventional wavelet decomposition tree, 17
decomposition, 17, 18
entropy, 48
FFT, 18
filter bank, 38
filter-bank selection, 42
filter-bank structures, 52
filter structure, 48
finest time resolution, 17
frequency and time, 30
frequency resolution, 17
low-pass filter function space and high-pass filter function space, 17
multiscale functions, 16
PWP decomposition tree, 17
structure, 17
Wavelet-packet transform (WPT), 51, 53
Wavelets
  and bank of filters, 14, 15
Daubechies, 16
empirical, 19–20
FT, 11
Meyer, 16
mother, 13
multiresolution analysis, 11
scalogram, 19
WP, 16–18
WT (see Wavelet transform (WT))
Wavelet shrinkage method, 49
Wavelet spectral analysis, 3
Wavelet theory, 11, 13
Wavelet thresholding, 37–38
Wavelet transform (WT), 1, 3
coefficients, 11
coding, 57
CWT, 19
framework, 11
and MRA, 12–14
optimized detection and classification, 42
pitch estimation, 25–27
quality evaluation of speech, 9
spectral analysis, 23–24
speech segmentation, 30
STFT, 11–13
UWT, 18
and VA, 26
voiced/unvoiced classification, 57
Wavelet/wavelet-packet analysis, 58
Wide-band (WB) speech codecs, 29
Wiener filtering, 42
WNN
  ASR, 44
  emotion recognition, 54
WPT-based NLD features, 53
WPT coefficients, 66
WT-based speech compression, 59

Z
Zero-crossing rate (ZCR), 29