

Index

- A Weighting, 75
- ACA, 1–3, 5, 7, 14, 24, 38, 64–66, 77, 85, 108, 128, 151, 152, 202, 204, 206
- ACF, 27, 28, 32, 58, 61, 98, 99, 101, 103, 105, 134–137
- ACF Maximum, 58, 59
- Acoustic Onset Time, *see* AOT
- Adaptive Differential Pulse Code Modulation, *see* ADPCM
- Aeolic Mode, 85
- Aliasing, 11, 35, 192
- Alignment, 139
 - Audio-to-Audio, *see* Audio-to-Audio Alignment
 - Audio-to-Score, *see* Audio-to-Score Alignment
- Alignment Path, 139, 141, 142, 145–147, 150
- Alternative Blackman Window, 195
- AMDF, 99
- Angular Frequency, 8
- ANN, 155, 157, 162
- AOT, 120, 121, 129
- API, 205
- Application Programmer's Interface, *see* API
- Arithmetic Mean, 9, 13, 33, 34, 36–38, 41, 54–56, 62, 66, 89, 97, 107, 116, 132, 134, 135, 200
- Articulation, 124, 156, 159, 171, 177, 178
- Artificial Neural Network, *see* ANN
- Artist Identification, 156
- Atomic Beat, *see* Tatum
- Attack, 120
- Attack Time, 42, 76, 120, 162
- aubio, 205
- Audio Content, 3
- Audio Content Analysis, *see* ACA
- Audio Fingerprint, *see* Fingerprint
- Audio Signal, 7
 - Periodic, 7
 - Random, 9
- Audio-to-Audio Alignment, 139, 146, 147, 149
- Audio-to-Score Alignment, 139, 148–150, 174
- Auditory Filterbank, 24
- Autocorrelation Coefficient, 61
- Autocorrelation Function, *see* ACF
- Autoregressive Modeling, 29
- Average Magnitude Difference Function, *see* AMDF
- Bandwidth, 34
- Bar, 110, 123, 124, 136, 137
- Bar Length, 137
- Bark, 46, 78, 80
- Bark Scale, 80, 81
- Bartlett Window, 194
- Beat, 28, 110, 116, 122, 123, 135, 136, 174, 175, 179
- Beat Histogram, 133–135, 154
- Beat Matching, 135
- Beat Phase, 135, 136

- Beat Spectrum, *see* Beat Histogram
- Beat Strength, 133
- Beat Tracking, 131, 135, 136, 174
- Beats per Minute, *see* BPM
- Blackman Window, 195
- Blackman-Harris Window, 195
- Block Length, 19, 20, 26, 29, 35, 74, 75, 78, 92, 98, 125, 166, 197
- Block Overlap Ratio, 20, 78, 127, 166
- Block-Based Processing, 19
- Box-Cox Transform, 65
- BPM, 122, 123, 134
- BS.1770, 76, 77
- BS.468, 76

- C Weighting, 75
- C++ Framework for Audio and Music, *see* CLAM
- CAMEL, 204
- CASA, 2
- CCF, 24–28, 112, 137
- CCIR Weighting, 76
- CD, 172
- CD Quality, 4, 133
- cent, 88
- Center Clipping, 98, 99
- Center of Gravity, *see* COG
- Central Moment, 9, 38–40
- Centroid, 36, 37, 45, 112, 162
- Cepstrum, 51, 101, 102
- Cepstrum-based, 99
- Chord, 82, 86–88, 116, 117
- Chord Detection, 117
- Chord Inversion, 87
- Chord Progression, 117
- Chord Recognition, 88, 94, 106, 110, 116
- Chord Template, 116, 117
- Chroma Perception, 81
- CiCF, 26, 196
- Circle of Fifths, 85, 86, 113–117
- Circular Correlation Function, *see* CiCF
- CLAM, 202, 203
- Classification, 51, 66–68, 151, 152, 154–156, 161, 162, 178
- COG, 37, 45, 127
- Comb Filter, 104
- CoMIRVA, 203
- Comité Consultatif International des Radiocommunications, *see* CCIR
- Compact Disc, *see* CD
- Computational Auditory Scene Analysis, *see* CASA
- Computer Audition, *see* ACA
- Concert Pitch, 88, 107
- Constant Q Transform, *see* CQT
- Content-based Audio and Music Extraction Library, *see* CAMEL
- Convolution, 14, 15, 28, 101, 181–183, 186, 187, 192, 193, 196
- Associativity, 182
- Circularity, 183
- Commutativity, 181
- Distributivity, 183
- Identity, 181
- Correlation, 68, 116, 117, 122, 165, 177–179
- Correlation Function, 24, 26, 27, 58, 94, 98, 99, 101, 136
- Cosine Distance, 114, 126, 133, 147, 157
- Cosine Window, 195, 196
- Cost Matrix, 141–145
- Covariance Matrix, 200
- CQT, 23, 24, 107, 110
- Critical Band, 80, 125
- Critical Band Rate, 46, 80, 81
- Cross Correlation Function, *see* CCF
- Cross Validation, 67, 155
- Cut-Off Frequency, 16, 17, 105

- DC Removal, 33
- DCT, 51, 53
- Decibel, 12, 53, 71, 72
 - dBa, *see* A Weighting
 - dBc, *see* C Weighting
 - dBFS, *see* Full Scale
 - dBZ, *see* Z Weighting
- Delta Function, 15, 181, 190–193
- Delta Impulse, *see* Delta Function
- Delta Pulse, 106, 136, 137, 191, 192
- Descriptor, *see* Feature
- Detection Function, *see* Novelty Function
- DFT, 19, 20, 23, 24, 51, 183, 185, 195–197
- Diatonic Scale, 82
- Diatonic Temperament, 90
- Difference Function, *see* Novelty Function
- Differentiator, 34
- Dimensionality Reduction, 4, 66, 67, 69
- Dirac Impulse, *see* Delta Function
- Discrete Cosine Transform, *see* DCT
- Discrete Fourier Transform, *see* DFT
- Distance
 - Cosine, *see* Cosine Distance
 - Euclidean, *see* Euclidean Distance
 - Manhattan, *see* Manhattan Distance
- Distance Matrix, 133, 139, 141–147
- Dominant, 88, 116, 117
- Down-Mixing, 33, 94, 164
- Down-Sampling, 34, 35, 146, 165
- Downbeat, 123, 136, 137
- DP, 136, 148–150
- DTW, 136, 139, 143–147, 149, 150
- Dynamic Programming, *see* DP
- Dynamic Time Warping, *see* DTW
- Dynamics, 3, 73, 148, 156, 170, 175, 177–179

- Emotion, 158, 179
- Enharmonic Equivalence, 83, 85, 88, 90, 115
- Envelope, 28, 31, 32, 76, 77, 125, 133–135, 147, 165

- Equal Temperament, 88, 90, 111
- Equal-Loudness Contour, 75
- Equivalent Rectangular Bandwidth, *see* ERB
- ERB, 81
- Euclidean Distance, 44, 65, 114, 147, 157, 165
- Evaluation, 128

- F-Measure, 131
- False Negative, *see* FN
- False Positive, *see* FP
- Fast Fourier Transform, *see* FFT
- FEAPI, 205
- Feature, 4, 5, 24, 31, 35, 38, 41, 42, 53, 58, 63–69, 71–73, 95, 96, 125, 134–137, 147, 149–152, 154–157, 159, 161, 162, 165, 178, 179, 199, 202–206
- Feature Extraction Application Programmer's Interface, *see* FEAPI
- Feature Space Transformation, 67, 69, 200
- Feature Subset Selection, 67, 200
- Features, 5
- FFT, 24, 28, 197
- Filter, 15
 - Comb, *see* Comb Filter
 - Moving Average, *see* MA
 - Single-Pole, *see* Single-Pole Filter
- Filterbank
 - Auditory, *see* Auditory Filterbank
 - Gammatone, *see* Gammatone Filterbank
- Fingerprint, 163–166
- Fingerprinting, 163–166
- Finite Impulse Response, *see* FIR
- FIR, 15, 17, 104, 187
- Floating Point, 12
- FN, 129, 130, 164
- Foot Tapping Rate, *see* Tactus
- Fourier Series, 8, 91, 185, 196
- Fourier Transform, *see* FT
- FP, 127, 129–131, 164
- Frequency Reassignment, *see* Instantaneous Frequency
- FT, 18, 20, 22, 28, 51, 104, 185, 186, 189, 191–194, 196
 - Convolution, 186
 - Frequency Scaling, 189
 - Frequency Shift, 188
 - Multiplication, 186
 - Parseval's Theorem, 187
 - Superposition, 186
 - Symmetry, 188
 - Time Scaling, 189
 - Time Shift, 188
- Full Scale, 12, 72, 74, 77
- Full-Wave Rectification, *see* FWR
- Fundamental Frequency, 2, 8, 9, 39, 45, 63, 79, 91, 92, 94, 97, 99–104, 106, 110, 149, 154, 162, 177, 185
- Fundamental Frequency Detection, 21, 63, 91, 92, 94, 104, 136, 175

- FWR, 133

- Gammatone Filter, 24
- Gammatone Filterbank, 24, 102
- Gaussian Distribution, 13
- Gaussian Mixture Model, *see* GMM
- Gaussianity, 39, 65
- Generalized Mean, 36
- Geometric Mean, 36, 37, 55, 107, 116
- Gibbs' Phenomenon, 8
- GMM, 155, 157, 162
- Ground Truth, 129, 133, 147, 156, 161, 202, 203

- Half-Wave Rectification, *see* HWR
- Hamming Window, 195
- Harmonic Mean, 36, 37, 131
- Harmonic Mixing, 108
- Harmonic Mode, 85
- Harmonic Product Spectrum, *see* HPS
- Harmonic Sum Spectrum, *see* HSS
- Harmonics, 8, 39, 42, 47, 50, 54, 79, 91, 92, 94, 97, 99–101, 103, 104, 106, 110, 111, 117, 162
- Harmony, 79, 86, 97, 137, 154, 156
- Hidden Markov Model, *see* HMM
- High-Level Feature, 5
- HMM, 117, 149, 150
- HMM Toolkit, *see* HTK
- Homophonic, 97
- Hop Size, 20, 23, 74, 75, 166
- HPS, 99, 100, 105
- HSS, 100, 106
- HTK, 53, 206
- HWR, 45, 102, 103, 105, 106, 125, 135, 147

- IBI, 122, 136
- ICA, 69
- IDFT, 196
- IFT, 186, 189, 191
- IIR, 15, 17, 64
- Impulse Response, 14–17, 24, 104, 181, 186, 187
- Independent Component Analysis, *see* ICA
- Infinite Impulse Response, *see* IIR
- Initial Transient Time, *see* Attack Time
- Input/Output, *see* IO
- Instantaneous Bandwidth, *see* Spectral Spread
- Instantaneous Feature, 31–33, 71, 112, 154
- Instantaneous Frequency, 21, 23, 97, 106, 126
- Instrument Recognition, 151, 156, 161, 162
- Integration Time, 74, 76
- Intensity, 71–73, 95, 154, 176, 177
- Inter-Beat Interval, *see* IBI
- Inter-Onset Interval, *see* IOI
- Interval, 83, 87, 90, 108, 112
- Intonation, 3, 91, 175, 177, 178
- Inverse Discrete Fourier Transform, *see* IDFT
- Inverse Fourier Transform, *see* IFT
- IOI, 121, 122, 136, 137, 178

- jMIR, 203
- JNDL, 72
- Just Noticeable Difference in Level, *see* JNDL
- K-Means Clustering, 157
- K-Nearest Neighbor, *see* KNN
- Key, 2–4, 31, 54, 79, 82, 83, 85–87, 90, 108, 112–117, 157
- Key Detection, 86, 94, 106, 108, 110, 112, 116
- Key Profile, 112–115, 117
- Key Signature, 85
- KNN, 155, 162
- Kullback-Leibler Divergence, 114
- Kurtosis, 39, 40, 66, 135
- Laplace Distribution, 13, 14
- Latency, 19, 27, 146
- LDA, 69, 147
- Leptokurtic, 39
- libXtract, 204
- Linear Discriminant Analysis, *see* LDA
- Linear Prediction, 28, 29, 53, 59, 105
- Local Tempo, 122, 123
- Loudness, 32, 34, 41, 46, 71–73, 75, 77, 78, 105, 108, 150, 154, 157, 159, 172, 175, 177–179
- Loudness Range, 78, 154
- Low-Level Feature, 5, 31, 62, 157, 165, 204, 205
- MA, 15–17, 34, 56, 64, 74, 127, 128
- Maaate, 205
- Machine Learning, 152, 155–157, 178, 202, 203, 206
- Machine Listening, *see* ACA
- Magnitude Spectrum, 28, 32, 39, 43, 45–49, 51, 54–56, 94–97, 100–102, 110, 133, 134, 166, 185, 186, 188, 189
- Main Lobe, 58, 96, 97, 193, 197
- Main Tempo, 122
- Major Mode, 82, 85, 113–116
- Manhattan Distance, 44, 114, 157, 165, 166
- Marsyas, 202
- Mean
 - Arithmetic, *see* Arithmetic Mean
 - Generalized, *see* Generalized Mean
 - Geometric, *see* Geometric Mean
 - Harmonic, *see* Harmonic Mean
 - Quadratic, *see* Quadratic Mean
- Mean Tempo, 122
- Meantone Temperament, 90
- Measure, *see* Bar
- Median, 36, 40, 41, 66, 128
- Median Filter, 128
- Mel, 51, 53, 80
- Mel Frequency Cepstral Coefficient, *see* MFCC
- Mel Scale, 24, 51, 53, 80, 81
- Mesokurtic, 39
- Meta Data, 1, 5
- Meter, 122, 123, 136, 137
- MFCC, 51, 53, 157, 162, 165
- Mid-Level Feature, 157
- MIDI, 2, 73, 88, 117, 120, 133, 135, 147–150, 174, 202, 203
- Minor Mode, 85, 113–116
- MIR, 2, 129, 151, 156, 202, 204
- MIREX, 155
- MIRtoolbox, 204
- Mode, 83, 85, 114, 179
 - Aeolic, *see* Aeolic Mode
 - Chromatic, *see* Chromatic Mode
 - Dorian, *see* Dorian Mode
 - Harmonic, *see* Harmonic Mode
 - Lokrian, *see* Lokrian Mode
 - Lydian, *see* Lydian Mode
 - Major, *see* Major Mode
 - Minor, *see* Minor Mode
 - Mixolydian, *see* Mixolydian Mode
 - Phrygian, *see* Phrygian Mode
 - Wholetone, *see* Wholetone Mode
- Mode Tempo, 122
- Modulation, 85, 86, 116
- Monophonic, 97
- Mood, 4, 158, 159, 161, 171, 179
- Mood Classification, 151, 156, 158, 159, 161, 203
- Moving Average, *see* MA
- MPA, 172, 173, 175, 177, 178
- Multi-Pitch Detection, 104–106, 111
- Music Emotion Recognition, *see* Mood Classification
- Music Information Retrieval, *see* MIR
- Music Information Retrieval Evaluation eXchange, *see* MIREX
- Music Performance, 3, 123, 147, 148, 158, 159, 163, 169–176, 179
- Music Performance Analysis, *see* MPA, 172
- Music Similarity, 151, 152, 156, 157, 162
- Music Similarity Detection, 156, 157
- Musical Communication, 169
- Musical Dynamics, 73
- Musical Form, 123
- Musical Genre, 151, 152, 154, 156
- Musical Genre Classification, 32, 66, 151, 152, 155–157, 159, 161, 162, 203
- Musical Instrument Digital Interface, *see* MIDI
- Musical Style, *see* Musical Genre
- Noisiness, 54, 55, 59
- Normalization, 26, 34, 44, 53, 65, 66
- NOT, 120, 121, 133
- Note, 124, 137, 161, 169, 175
- Note Onset Time, *see* NOT
- Note Value, 123, 124
- Novelty Function, 124, 125, 127, 128, 133–137
- Observation, 41, 64, 65
- Octave, 82
- Offset Time, 124

- Onset, 119–122, 124, 125, 127–130, 132, 133, 135–137, 147, 149, 150, 205
- Onset Detection, 45, 121, 122, 124, 125, 128–131, 133, 135
- Onset Time, 119–121, 124, 125, 127–129, 131, 135, 136, 147, 149, 150, 174, 175, 178
- Onset Tracking, *see* Onset Detection
- OpenSMILE, 205
- Ornamentation, 171
- Outlier, 40
- Overfitting, 66
- Overlap, *see* Block Overlap
- Overshoot, 8
- Overtone, 91

- Partials, *see* Harmonics
- PAT, 120, 121
- Pattern Recognition, *see* Machine Learning
- PCA, 69, 199, 200
- PDF, 12–14, 36, 38–41, 74, 149
- Peak Picking, 124, 127, 128
- Peak Program Meter, *see* PPM
- Peak Structure Distance, *see* PSD
- Peakiness, 96
- Pedaling, 174, 177, 178
- Perceived Tempo, 122
- Perceptual Attack Time, *see* PAT
- Perceptual Hash, *see* Fingerprint
- Perceptual Onset Time, *see* POT
- Period Length, 7, 92, 94, 98, 102–104, 123, 185
- Phase Spectrum, 21, 23, 97, 185, 188, 189
- Physical Onset Time, *see* AOT
- Pitch, 2, 41, 79, 81–83, 88, 91, 102, 106, 113, 136, 164, 169, 170, 175
- Pitch Chroma, 108–117, 137, 147, 150
- Pitch Chromagram, *see* Pitch Chroma
- Pitch Class, 81–83, 86, 88, 108, 109, 112, 113
- Pitch Class Distribution, *see* Pitch Chroma
- Pitch Class Profile, *see* Pitch Chroma
- Pitch Detection, *see* Fundamental Frequency Detection
- Pitch Height, 80, 81, 89
- Pitch Histogram, 154
- Pitch Tracking, *see* Fundamental Frequency Detection
- Platykurtic, 39
- Polyphonic, 95, 97, 103, 104, 122, 154, 162
- POT, 120, 121, 129, 133
- Power Spectrum, 43, 45–48, 53, 56, 99, 186
- PPM, 76, 77
- Pre-Whitening, 97, 99, 105
- Precision, 131
- Prediction Error, 29, 59, 61
- Predictivity Ratio, 59, 60
- Principal Component Analysis, *see* PCA
- Probability Density Function, *see* PDF
- Probe Tone Ratings, 113, 115
- Process Loss, 197

- Production, 3, 172
- PSD, 149
- Pythagorean Temperament, 90

- Quadratic Mean, 37
- Quantile, 40
- Quantile Range, 40, 41
- Quantization, 9, 11, 12
- Quantization Error, 11, 12
- Quantization Step Size, 11
- Quartiles, 41

- Radial Basis Function, *see* RBF
- RBF, 97, 131
- Real Time, 33, 34
- Real-Time, 19, 149, 202, 205
- Recall, 131
- Receiver Operating Curve, *see* ROC
- Rectangular Window, 191, 193–196
- Relative Frequency Distribution, *see* RFD
- Release Time, 76, 77
- Rest, 124
- Revised Low Frequency B Curve, *see* RLB
- RFD, 14
- Rhythm, 119, 122, 123, 127, 156, 157, 159
- Rise Time, *see* Attack Time
- RLB Weighting, 76, 78
- RMS, 27, 34, 37, 73–78, 125, 162
- ROC, 130
- Root Mean Square, *see* RMS
- Root Note, 82, 83, 85, 87, 88, 113, 114, 116

- Sample Rate, 9–11, 20, 34, 35, 58, 61, 92, 94, 127, 134, 165, 166, 192
- Sample Rate Conversion, 34, 35, 94
- Sampling, 9
- Sampling Frequency, *see* Sample Rate
- Sampling Theorem, 10, 11, 35, 192
- Scale Degree, 81, 83, 85, 88, 90, 91
- SCMIR, 204
- Score, 2–4, 83, 85, 91, 106, 117, 123, 124, 148–150, 158, 169, 170, 172, 173, 177, 178
- Score Following, 148, 149
- Self-Organizing Map, *see* SOM
- Sequential Backward Elimination, 68
- Sequential Forward Selection, 68
- Short Time Fourier Transform, *see* STFT
- Short-Term Feature, *see* Instantaneous Feature
- Side Lobe, 97, 193, 197
- Signal-to-Noise Ratio, *see* SNR
- SIMD, 19, 132
- Similarity Matrix, 133, 137, 143
- Single Instruction Multiple Data, *see* SIMD
- Single Variable Classification, 68
- Single-Pole Filter, 17, 75, 96, 107
- Singular Value Decomposition, *see* SVD
- Skewness, 38–40, 65, 66, 135
- SNR, 12

- Software Developer's Kit, *see* SDK
 SOM, 157
 Sonic Visualiser, 203
 Sound Quality, *see* Timbre
 Spectral Centroid, 37, 45–47, 165
 Spectral Crest Factor, 54, 165
 Spectral Decrease, 48, 49
 Spectral Envelope, 48, 51, 95, 97, 99, 105, 157
 Spectral Flatness, 55–57, 165
 Spectral Flux, 44, 126, 136, 147
 Spectral Kurtosis, 39, 40
 Spectral Leakage, 193, 197
 Spectral Predictivity, 60, 61
 Spectral Rolloff, 42, 43
 Spectral Skewness, 38–40
 Spectral Slope, 49, 50
 Spectral Spread, 47, 48
 Spectrogram, 21, 104, 106, 175
 Spectrum, 35, 38, 51, 53, 56, 100–102, 104, 106, 185, 188, 190–193, 195, 196
 Standard Deviation, 32, 37–39, 45, 47, 66, 89, 112, 121, 132, 135
 Standard Pitch, *see* Concert Pitch
 STFT, 20, 21, 23, 24, 32, 42, 44, 45, 56, 60, 92, 93, 95, 96, 100, 106, 109, 110, 125–127, 133, 136, 147, 166, 192, 195
 Subdominant, 88, 116, 117
 Subfeature, 64, 66, 154, 155, 157, 204
 Support Vector Machine, *see* SVM
 SVD, 69
 SVM, 155, 162

 Tactus, 122, 133, 135
 Tatum, 123, 137
 Temperament, 3, 90, 91, 171
 Tempo, 2–5, 31, 33, 119, 122–124, 127, 135–137, 145, 147–150, 154, 156, 157, 159, 164, 171, 174, 175, 177–179, 205
 Local, *see* Local Tempo
 Main, *see* Main Tempo
 Mean, *see* Mean Tempo
 Mode, *see* Mode Tempo
 Perceived, *see* Perceived Tempo
 Tempo Detection, 124, 135, 136
 Tempo Induction, *see* Tempo Detection
 Tempo Tracking, *see* Tempo Detection
 Test Set, 131, 155
 Texture Window, 20, 66, 112, 134, 154, 155, 157
 Timbre, 3, 41, 42, 45, 47, 53, 73, 89, 97, 108, 147, 150, 154, 172, 175, 177, 178
 Timbre Toolbox, 204
 Time Signature, 123, 124, 137, 154
 Timing, 3, 121, 123, 145–148, 171, 174, 175, 177–179
 TN, 130
 Tonal Centroid, 115
 Tonal Power Ratio, 57
 Tonality, 54
 Tonalness, 54–56, 58–61, 63, 94–97, 110, 115
 Tonic, 87, 88, 90, 117
 TP, 130, 164
 Training Set, 155
 Transient, 8, 47, 56, 74, 120
 Tremolo, 3, 73, 133
 Triad, 86, 87, 116
 Tritone, 113–115
 True Negative, *see* TN
 True Peak Meter, 78
 True Positive, *see* TP
 Tuning Frequency, 88–90, 106–108
 Tuning Frequency Estimation, 106

 VAMP, 205
 Variance, 37, 38, 40, 65, 69, 103, 157, 199, 200
 Velocity, 73
 Vibrato, 3, 73, 91, 107, 133, 171, 175, 177, 178
 von-Hann Window, 195

 Waikato Environment for Knowledge Analysis, *see* WEKA
 Warping Path, *see* Alignment Path
 Watermarking, 163, 164
 WEKA, 206
 Wiener-Khinchin Theorem, 28
 Window, 193
 Alternative Blackman, *see* Alternative Blackman Window
 Bartlett, *see* Bartlett Window
 Blackman, *see* Blackman Window
 Blackman-Harris, *see* Blackman-Harris Window
 Cosine, *see* Cosine Window
 Hamming, *see* Hamming Window
 Rectangular, *see* Rectangular Window
 von-Hann, *see* von-Hann Window
 Window Function, 16, 191–193, 196, 197
 Word Length, 11, 12, 165, 182
 Workload, 19, 28, 35, 132, 144, 167, 201

 YAAFE, 204
 Yet Another Audio Feature Extractor, *see* YAAFE

 Z Weighting, 76
 Zero Crossing Rate, 62, 63, 98, 149, 154
 Zero Phase Filtering, 17