

# Index

- accidental, 4, 249
- accumulated score, 336
- ADSR, 27
- aliasing, 50, 61, 82
- alignment
  - global, 405
  - local, 405
- amplitude, 21, 42, 58
- amplitude modulation, 28
- arrangement, 385
- articulation, 9
- attack, 305
- attack phase, 27
- audio, 1, 19
- audio identification, 357
- audio matching, 371
- audio signal, 19
- augmented fifth, 244
- augmented triad, 244
- autocorrelation, 321
  - short-time, 322
- average precision, 398
  
- backtracking, 139, 282, 337
- bandlimited, 107, 114
- bandwidth, 121
- bar, 6
- bar line, 7
- bass line, 471
- Baum–Welch algorithm, 285
- beat, 6, 303
- beat sequence, 335
- beating, 113
- beats per minute, 15
- bigram, 288
- bin offset, 439
- binary mask, 423
  
- block, 179, 182
- BPM, 7, 15
- break-even point, 397
- broadcast monitoring, 401
- broken chord, 246, 264
  
- cadenza, 388
- Cauchy–Schwarz inequality, 65
- cell, 132, 179
- CENS, 375
- cent, 22
- center frequency, 22
- character, 166
- characteristic function, 366
- chirp signal, 85
- chord, 237, 243
- chord progression, 252
- chord recognition, 238
- chorus, 173
- chroma, 4, 123
- chroma feature, 124
- chromagram, 124
- chromatic circle, 5
- chromatic scale, 4, 247
- circle of fifths, 248
- clef, 6
  - bass, 6
  - treble, 6
- client, 358
- client–server model, 358
- clustering, 287
- cocktail party problem, 415
- compact support, 77
- complex conjugate, 63
- complex numbers, 63
- concert pitch, 21, 164
- conditional probability, 274

- confusion
  - major–minor, 262
- conjugate symmetric, 64
- consonance, 242
  - imperfect, 243
- constellation map, 362
- constraint region, 446
- content, 356
- content-based, 356
- contrast, 171
- convexity, 454
- copyright monitoring, 401
- cosine distance, 376
- cost matrix, 132, 377
  - accumulated, 137
- cost measure
  - local, 132
- cover song, 386
- cover version, 386
- coverage, 197
- critical point, 286
- cross-validation, 268
  - K*-fold, 268
- cutoff frequencies, 120
- cyclic shift, 129, 382
  
- dB, 25
- decay phase, 27
- decibel, *see* dB
- deletion, 166
- DFT, 52, 88
  - inverse, 427
- diagonal matching, 376
- diatonic scale, 247
- digitization, 49, 60
- diminished fifth, 244
- diminished triad, 243
- Dirac delta function, 79
- Dirac sequence, 79
- discrete Fourier transform, *see* DFT
- discrete STFT, 54
- dissonance, 242
- distance, 136
- distance measure
  - local, 132
- distribution, 79
- DJ mix, 386
- document, 355
- domain
  - acoustic, 30
  - auditory, 30
  - physical, 30
  - symbolic, 30
  - visual, 30
- dominant, 247
- drum extraction, 471
- DT-signal
  - finite length, 87
  - length, 87
- DTW, 131
  - boundary condition, 134
  - constraint region, 143
  - distance, 136
  - global constraint, 143
  - local weights, 142
  - monotonicity condition, 134
  - multiscale, 144
  - step size, 134, 141
  - subsequence, 412
  - warping path, 134
- duality, 83
- dyad, 243
- dynamic programming, 137, 281, 336, 445
- dynamic time warping, *see* DTW
- dynamics, 8, 24
  
- edit distance, 156, 166
- emission probability, 277
- energy, 66, 67
  - local, 306
- enharmonic equivalence, 4, 240
- envelope, 27
- equal loudness contours, 26
- Euclidean distance, 454
- Euler’s formula, 72
- exponential function, 46, 72
  
- F-measure, 217
  - maximal, 398
  - pairwise, 218
- false negative, 216, 260
- false positive, 216, 260
- fan-out, 369
- feature, 117
  - beat-synchronous, 177, 272, 339
  - CENS, 375
  - space, 132
- feature representation, 364
- FFT, 89
- fifth, 240
- file metadata correction, 401
- fingerprinting
  - compactness, 359
  - robustness, 359
  - scalability, 359
  - specificity, 358
- fitness, 201
- fold, 268

- form
  - chain, 172
  - part, 172
  - rondo, 173
  - sonata, 173
  - strophic, 172
- Fourier
  - coefficient, 71
  - series, 70
- Fourier analysis, 40, 105
- Fourier coefficient, 46, 70
- Fourier representation, 41, 47, 70, 82, 105
- Fourier synthesis, 105
- Fourier transform, 41, 46
  - $L^2([0, 1])$ , 75
  - $L^2(\mathbb{R})$ , 77
  - $\ell^2(\mathbb{Z})$ , 82
  - discrete, *see* DFT
  - fast, *see* FFT
  - magnitude, 47
  - windowed, 107
- frame, 104
- frame index, 104
- frequency, 21, 42, 58
  - center, 22
  - fundamental, 23, 433, 460
  - instantaneous, 85, 417, 435, 437, 471
- frequency modulation, 28
- frequency stamp, 360
- frequency trajectory, 433, 444
- front-end transform, 118
- fugue, 12
- function, 42
  - argument, 42
  - sinc, 80
  - value, 42
- functional, 250
- fundamental frequency, 23, 433, 460
  - F0, 433
- gap, 394
- Gaussian function, 79
- glissando, 432
- gradient, 454
  - step size, 455
- granularity, 400
- ground truth, 216, 258
- half step, 247
- half-wave rectification, 307
- harmonic, 23, 460
- harmonic mean, 201
- harmonic series, 241
- harmonic summation, 417, 434, 442, 471
- harmonics
  - tempo, 318
- harmonization, 388
- harmony, 237
- Heisenberg uncertainty principle, 102
- Hertz, 21
- hidden Markov model, *see* HMM
- Hilbert space, 68
- HMM, 273, 276, 294
  - continuous, 288
  - discrete, 277, 287
  - estimation problem, 285
  - evaluation problem, 279
  - uncovering problem, 280
- hop size, 54, 102
- Hz, 21
- impulse function, 79
- IMSLP, 31
- index, 364
- indexing, 364
- indicator function, 366
- induced segment family, 197
- information retrieval, 355
- inharmonic, 23, 29
- initial state probability, 275
- inner product, 44
  - sliding, 321
- input object, 267
- insertion, 166
- instantaneous frequency, 85, 417, 435, 437, 471
- instantaneous frequency estimation, 433
- integral
  - Lebesgue, 66
  - Riemann, 66
- intensity, 25
- intensity level, 25
- Interface
  - Score Viewer, 149
- interference, 78
  - constructive, 78
  - destructive, 78
- interonset interval, *see* IOI
- Interpretation Switcher, 147
- interval, 239
  - complement, 300
- inversion, 36, 246
- inverted list, 364
- IOI, 343
- isometry, 75
- Itakura parallelogram, 143
- item, 216, 355, 395
  - negative, 216

- positive, 216
- relevant, 216
- jazz, 385
- just interval, 242
- just intonation, 242
- just noticeable difference, 22
- Karaoke, 448
- key, 248, 388
- key note, 247
- key signature, 6, 468
- label function, 218
- labeling, 218
- lag, 321
- LCS, 166, 406
- ledger line, 6
- legato, 9
- Levenshtein distance, 166
- linear
  - $\mathbb{C}$ -, 64
- linear chirp, 85
- linear operator, 111
- logarithmic compression, 125, 269, 310
- longest common subsequence, *see* LCS
- loudness, 24
- lyrics, 9, 388
- magnitude, 47
- major scale, 247
- major third, 241
- major triad, 243
- MAP, 398
- Markov chain, 274
- Markov property, 274
- mashup, 386
- match, 376
- matching function, 376, 380
  - transposition-invariant, 382
- matrix decomposition, 452
- matrix factorization, 452
- maximum likelihood, 286
- mean average precision, 398
- measure, 6, 318
- median, 420
- median filter, 421
- medley, 386
- melody, 36, 432
- melody extraction, 431
- melody separation, 431, 448
- meter, 6, 345
- MFCC, 177
- MIDI, 13
- MIDI file, 13
- MIDI note number, 13
- minor scale, 247
  - harmonic, 248
  - melodic, 248
  - natural, 248
- minor third, 241
- minor triad, 243
- MIRtoolbox, 226
- modulation, 81, 250
- monopitch estimation, 471
- motif, 2
- MSTFT, 427
  - valid, 427
- multipitch estimation, 471
- music engraving, 16
- music notation, 5
- music transcription, 30
- musical form, 172
- musical structure, 172
- MusicXML, 16
- natural note, 249
- navigation
  - interdocument, 147
  - intradocument, 148
- NMF, 417, 450, 452, 472
  - activation, 451, 453
  - template, 451, 453
- noise, 27
- nonnegative matrix, 453
- nonnegative matrix factorization, *see* NMF
- norm, 65
- note, 3
  - quarter, 7
  - value, 7
- novelty, 209
- novelty function, 208, 210
  - complex-domain, 316
  - energy-based, 307
  - phase-based, 314
  - spectral-based, 309
- Nyquist frequency, 50, 82
- observation, 276, 278
- observation symbol, 277
- OCR, 17
- octave, 3, 239, 240
  - tempo, 319
- octave equivalence, 325
- octave number, 4, 123
- OMR, 17
- onset, 305
- onset detection, 304, 305

- optical character recognition, *see* OCR
- optical music recognition, *see* OMR
- orthonormal basis, 68
- output probability, 277
- output value, 267
- oversegmentation, 223
- overtone, 23
  
- parody, 386
- partial, 23, 29
  - harmonic, 23
- partial derivative, 454
- partial match, 412
- partial matching, 406
- partition of unity, 427
- passing note, 259
- path, 179, 181
- path family, 197
- pentatonic scale, 248
- perfect, 242
- perfect fifth, 37
- perfect interval, 242
- performance analysis, 151
- period, 21, 58
  - prime, 58
- periodic, 21
- Petrucci Music Library, 31
- phase, 42, 58
- phase unwrapping, 314, 436
- phase vocoder, 471
- phase wrapping, 314, 436
- phon, 26
- piano roll, 11
- piano transcription, 386
- pitch, 3, 21, 23
- pitch class, 3, 123
- pitch shifting, 468
- player piano, 11
- PLCA, 472
- PLP function, 329, 331
- polar coordinate, 45, 73
- portamento, 127
- positive definite, 64
- postfiltering, 253, 291
- potpourri, 386
- power, 24
- PPQN, 14
- PR curve, 396
- precision, 217
  - at rank  $r$ , 396
  - pairwise, 218
- precision–recall curve, *see* PR curve
- predominant F0 estimation, 470
- predominant frequency, 417
- predominant local pulse, 329
- prefiltering, 253, 271
- primitive, 87, 113
- principal argument function, 315, 436
- principal value, 314, 436
- probabilistic latent component analysis, *see* PLCA
- probability density function, 288
- psychoacoustics, 19
- pure interval, 242
- pure intonation, 242
- Pythagorean comma, 37
- Pythagorean scale, 37
- Pythagorean tuning, 37, 242
  
- quantization, 61, 287
- quantization error, 62
- quantization step size, 61
- quantizer, 61
- query, 355
- query-by-example, 356
- query-by-humming, 448
- quodlibet, 386
- quotation, 386
  
- raga, 385
- rank, 395
  - higher, 395
  - top, 395
- rank (factorization), 453
- recall, 217
  - at rank  $r$ , 396
  - pairwise, 218
- reconstruction, 470
- rectangular function, 80
- reference thumbnail family, 222
- relative relationship, 249
- release phase, 27
- relevance function, 396
- remaster, 386
- remix, 386
- repetition, 171
- representation
  - chroma, 124
  - mid-level, 12, 448
  - MIDI, 13
  - piano-roll, 11
  - score, 15
  - time–chord, 257
  - time-lag, 322
- rest, 7
- retrieval
  - category-based, 407
  - content-based, 356

- document-level, 389, 394, 400
- fragment-level, 394, 400
- retrograde, 36
- rhythm, 6, 345
- Riemann sum, 51, 84
- riff, 207
- Roman numeral analysis, 251
- root note, 243
- root of unity, 113
- root position, 246
  
- Sakoe–Chiba band, 143
- salience representation, 417, 434, 443
- sample, 50, 60
- sampling, 60
  - $T$ -, 50
  - equidistant, 49, 60
  - period, 50, 60
  - rate, 50, 60
- sampling (production), 386
- sampling theorem, 50, 61, 107, 114
- scalability, 359
- scale, 3, 246
  - equal-tempered, 3, 240
- scale degree, 247
- scale steps, 246
- scaling, 59
- scaling parameter, 325
- scape plot, 203
- score, 2
- score following, 157
- score-informed processing, 447
- score-informed source separation, 416
- scorewriter, 16
- segment, 172
  - induced, 182, 392
- segment family, 197
- segmentation, 170
- self-similarity matrix, 178
  - $i$ -transposed, 190
  - transposition-invariant, 190
- self-transition, 288
- semitone, 4, 240
- separation
  - harmonic–percussive, 419
- server, 358
- sheet music, 1, 2
- Shepard helix, 5
- short-time Fourier transform, *see* STFT
- signal, 57
  - analog, 42, 58
  - audio, 19
  - bandlimited, 107, 114
  - continuous time, 58
  - digital, 58
  - discrete time, 50, 60
  - measurable, 66
  - periodic, 58, 67
- similarity
  - music, 117
- sinusoid, 21, 42, 58
- SMF, 13
- Smith–Waterman, 406
- soft mask, 424, 466
- solo–accompaniment separation, 471
- sound, 19
  - harmonic, 21, 418, 460
  - intensity, 25
  - percussive, 418
  - power, 24
  - ultrasonic, 21
- sound collage, 386
- source separation, 415
- source-filter model, 473
- space
  - signal, 58
- specificity, 357, 399
- spectral flux, 309
- spectral leakage, 122
- spectral vector, 54
- spectrogram, 29, 55, 98
- speed of sound, 36
- square-summable, 66
- SSM, 178
- staccato, 9
- staff, 5
  - grand, 8
  - system, 8
- Standard MIDI File, *see* SMF
- state, 273
- state transition probability, 274
- STFT, 53, 94
  - continuous-time, 94
  - discrete, 103
  - discrete-time, 102
  - inverse, 417
  - modified, 427
- stripe, 179
- structure, 388
  - homogeneity-based, 172
  - novelty-based, 172
  - repetition-based, 172
- structure analysis, 170
- structure annotation, 218
- structure features, 214
- style, 8
- subdominant, 247
- subharmonics

- tempo, 318
- substitution, 166
- superposition, 59
- supervised learning, 267
- suspended note, 259
- sustain phase, 27
- sweep signal, 85
- symbolic, 1
- synchronization, 115
  - audio–audio, 156
  - MIDI–audio, 157
  - music, 115, 447
  - score–audio, 156, 447
- syncopation, 316
  
- tactus, 318
- tag, 356
- target zone, 368
- tatum, 318
- tempo, 7, 8, 303, 388
- tempo dominant, 326
- tempo rubato, 316
- tempo subdominant, 326
- tempogram, 317
  - autocorrelation, 323
  - cyclic, 325
  - discrete, 318
  - Fourier, 319
- testing set, 268
- tetrad, 243
- threshold of hearing, *see* TOH
- threshold of pain, 25
- thresholding
  - global, 192
  - relative, 193
- tick, 14
- timbre, 26
- time signature, 6, 346
- time stamp, 360
- time–frequency masking, 470
- time-lag representation, 213
  - circular, 213
- timing
  - expressive, 388
- TOH, 25
- tone
  - musical, 23
  - pure, 21
- tone height, 123
- tonic, 247
- total score, 446
- track, 415
- training set, 268
- trajectory
  - F0-, 433
- transient, 27, 305
- transition probability matrix
  - transposition-invariant, 290
- translation, 81
- transposition, 36, 128
- tremolo, 28, 432
- triad, 243
- triangle inequality, 65
- tritone, 241
- true negative, 216
- true positive, 216, 260
- twiddle factor, 90
  
- undersegmentation, 223
- unison, 240
  
- validation set, 268
- variation, 386
- vector
  - distance, 65
  - orthogonal, 65
  - transposed, 64
- vector space, 64
- verse, 173
- version identification, 384
- vibrato, 28, 432
- Viterbi algorithm, 281, 294, 445
  
- warping path, 134
  - optimal, 136
  - total cost, 135
- wave, 19
  - acoustic, 19
  - longitudinal, 20
  - transverse, 20
- waveform, 19
- whole step, 247
- whole tone scale, 248
- Wiener filtering, 424, 470
- window
  - Hann, 98
  - rectangular, 97
  - triangular, 98
- window function, 53, 94
- windowed signal, 53
- windowing
  - adaptive, 186, 339
  - fixed-size, 338
- word, 166
- word recognition, 279
  
- zero-padding, 188, 210, 271, 318, 420