Index

A

acoustic reflections, localization of 126
acoustic transmission 424, 425, 426, 434
acoustic transmission channel 424, 425, 426
action units (AU) 402, 403, 404, 405, 406, 411, 412, 413
adult attachment interview (AAI) database 403
AIBO database 401, 402
alternating least squares (ALS) method 193, 194, 196, 356
artificial cocktail party machine 63
artificial neural network (ANN) 426, 431, 432, 433, 435, 436, 437, 439
artificial probe stimuli 424
assistive robotics 1
A/the sound 105
attention 85, 90, 92, 103, 105
attention, bottom-up 105
attention, top-down 90, 105
audio mixing 246, 247, 248, 249, 252, 253, 255, 258, 260, 261, 262, 264
audio mixing, anechoic 246, 247, 248, 255, 260, 262
audio mixing, convolutive 247, 262, 263
audio mixing, instantaneous 246, 247, 248, 249, 254, 261, 262
audio recognition 1, 3
Audioscrobbler service 339, 349
audio signal processing 1
audio signals 246, 263
audio source separation 246, 247, 248, 249, 255, 260, 261, 263, 264
audio source signals 247
audition 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 91, 94, 95, 96, 101, 102, 104, 105
auditory mixtures, complex 61
auditory nerve (AN) 455, 473
auditory pathways and cortex (AP&C) 449, 473
auditory scene analysis (ASA) 23, 24, 36
auditory systems 80
autocorrelation coefficients 337, 347
automatic segmentation 317, 318, 319, 321, 327, 328, 330, 331
automatic sound source localization 126
automatic tagging algorithm 336, 337, 341, 342, 348
autotags 335, 336, 344, 345

B

background noise 2, 4, 7, 11, 13
bag of frames 337, 338
bag of words 335, 337
basic elementary unit (BEU) 460, 461, 473, Bayesian estimation 164, 165
Bayesian signal processing 163, 180
beamformers 208, 216, 221, 242
beamforming 211, 212, 224, 225, 241, 242, 243, 244, 245
Belfast database 402, 403
Berlin Database of Emotional Speech 401
binary time-frequency masking 163, 167
bio-inspired knowledge 448
bio-inspired mask processors 466
bio-inspired or neuromorphic speech processing (BINSP) 458, 459, 473
bio-inspired speech processing 448, 457
blind source separation systems  163, 177
BU-3DFE database  401, 402

C

canonical decomposition (CANDECOMP)  190
characteristic frequency unit (CF)  455, 456, 457, 458, 459, 461, 462, 465, 466, 467, 473
Cherry, Colin  107, 109, 123
chroma  317, 319, 320, 321, 322, 323, 325, 326, 330, 331, 332
chromagrams  337
classification taxonomy  208
clustering  209, 212, 213, 216, 217, 240, 241
clustering, binary  209, 211, 212, 213, 227, 229, 231, 235, 236, 238, 239
clustering, hard  209, 238
clusters  209, 212, 227
cochlear nuclei (CN)  455, 456, 457, 473
cocktail party environments  61, 62, 64, 65, 66, 67, 68, 70, 73
cocktail party problem (CPP)  61, 62, 63, 67, 69, 107, 109, 123, 207
cocktail party processing  61, 74
computational auditory scene analysis (CASA)  22, 23, 26, 30, 61, 62, 64, 65, 67, 68, 73, 163, 164, 179
computer audition system  373
consonant-vowel dyphone (C-V)  458, 473, ConvNMF-ED  357, 358, 360, 361, 362, 363, 364, 365
ConvNMF-KL  357, 361, 362, 363, 364, 365
convolutive blind source separation (CBSS)  108, 186, 205
convolutive mixtures  186, 187, 189, 195, 204, 205, 206
convolutive NMF (ConvNMF)  353, 354, 357, 358, 360, 361, 362, 363, 364, 365
correct detection (CD)  310
Csiszar’s divergence  356

D

Danish Emotional Speech Database  401, 402, 419
database management  334
delta MFCC  337
dimensional theory  399
dimensional theory: activation  399, 400, 402
dimensional theory: evaluation  399, 400, 403, 404, 416, 417
direction-of-arrival (DOA)  126, 127, 128, 129, 132, 135
direction-of-arrival estimation  126, 159, 160, 161
discrete Fourier transform (DFT)  31, 209, 211, 237, 238

e

early decay time (EDT)  425, 426, 427, 432, 433, 437, 444
eigenvalue decomposition (EVD)  186
emotional content  398
emotions  399, 400, 401, 402, 403, 404, 405, 410, 413, 414, 415, 416, 421, 422
emotions: anger  399, 400, 401, 402, 403, 404, 415
emotions: disgust  399, 401, 402, 403, 404, 415
emotions: fear  399, 400, 401, 402, 403, 404, 415
emotions: happiness  399, 401, 402, 403, 415
emotions: sadness  399, 401, 402, 403, 404, 415
emotions: surprise  399, 400, 401, 402, 403, 404, 415
enhancement algorithms  208, 211, 229
environmental sounds  1, 2, 3, 4, 5, 7, 10, 11, 14
equivalent rectangular bandwidth (ERB)  300, 303
Euclidean distance  266, 267, 270, 296, 354, 355, 357, 358, 367
evaluation  335, 336, 344, 345, 346, 347, 350, 351, 352
exhaustive separation  161
expectation maximization (EM) algorithm  27, 34
expectation-maximization (EM) algorithm  164, 167, 172, 174, 176, 179, 180, 181, 182
Index

F

FABO acted database 402
face extraction 109
facial action coding system (FACS) 402, 403, 404, 406, 412, 415, 422
false negative (FN) 310, 311
false positives (FP) 310
fast Fourier transform (FFT) 319, 320, 360, 363, 450, 451, 473
feature 352
feed-forward neural network (FNN) 308
finite impulse response (FIR) filters 209
form 317, 322, 325, 331, 332
fourier transform 337
frequency-dependent time-frequency (FDTF) 299, 300
frequency domain convolutive blind source separation (FDCBSS) 108, 117
frequency modulation unit (FM) 456, 457, 459, 464, 465, 466, 473
Frobenius norm 355

G

Gabor atoms 4, 5, 6, 17
Gabor dictionary 5, 6
Gabor functions 5, 6, 18
Gabor properties 6
Gabor representation 5
Gabor time-frequency atoms 6
Gaussian clusters 7
Gaussian distribution 34
Gaussian functions 5
Gaussian mixture models (GMM) 7, 8, 9, 10
gaussian weighted spectrogram (GWS) 319, 321
general audio environment recognition 3
generalized neuromorphic computing unit (GNCU) 459, 460, 461, 467, 473,
generalized sidelobe canceller (GSC) 214, 218, 219, 220, 221, 222, 224, 225, 240, 241, 245
grouping 317, 322, 325, 329, 330, 331, 332

H

harmonically-wrapped peak similarity (HWPS) 24, 28, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 57
hearing 80, 84, 85, 93, 95, 105
higher order statistics (HOS) 208
human affective state 399
human auditory system 23, 24, 25, 32, 61, 62, 64, 65, 68, 73, 162
human speech 109

I

independent component blind 61
inear discriminant analysis (LDA) 400, 406, 407, 408, 409, 410, 411, 412, 416
information-bearing elements (IBE) 451, 452, 455, 473
information gap 372
information rate (IR) 372, 374, 375, 378, 383, 384, 385, 387, 388, 389, 390, 395, 396
informed source separation 163
instantaneous acoustic intensity 161
instantaneous models 353
intensity 126, 147, 156, 157
interference suppression 209, 225, 228, 234, 238, 239, 241, 242
interferers 209, 227, 229, 230, 231, 241
international conference on music information retrieval (ISMIR) 335, 349, 350, 351, 352
international phonetic alphabet (IPA) 463, 464, 471, 473
inverse transform 246, 248, 249, 253, 257
Itakura-Saito (IS) divergence 266, 268, 269, 270, 271, 272, 273, 274, 275, 276, 277, 281, 283, 284, 285, 286, 287, 291, 292, 294
K

kirshenbaum alphabet (KB) 452, 463, 464, 471, 473,
K-nearest neighbor (kNN) classification method 8, 9
Kullback-Leibler (KL) distance 376, 377, 382, 385, 386, 387, 390, 395

L

labels 336, 340, 343, 344, 350, 351, 352
lapped orthogonal transform 246
Last.fm 335, 336, 337, 339, 340, 341, 345, 346, 347
linear modeling 162, 164, 165, 170, 172, 173, 178, 179
linear predictive coding (LPC) 450, 451, 462, 464, 465, 473
linear regression techniques 266, 267
listening 85, 87, 89, 93, 95, 105
listening, every day 105
listening, musical 105
local changes 318
local time-frequency models 162, 172, 178, 179

M

machine audition 80, 81, 82, 83, 84, 86, 88, 89, 101, 102, 207
machine-based multi-channel source separation 207
machine learning 352
machine learning techniques 334, 337, 343
machine listeners 162, 163
Magnatagatune 339, 340
magnitude spectrogram 266, 267, 269, 271, 272, 274, 276, 292
Markov Chain Monte Carlo (MCMC) approach 113, 269, 278, 279, 280, 292, 295
matching pursuit (MP) algorithm 2, 4, 5, 6, 7, 8, 9, 10, 17, 18
maximum likelihood (ML) 269, 270, 272, 277, 279, 283, 287
mel-frequency cepstral coefficients (MFCC) 2, 3, 4, 7, 8, 9, 10, 17, 18, 337, 338, 346, 347
microphone array geometries 126, 127, 136, 140
microphone arrays 135, 156, 160, 208
minimum mean square error (MMSE) 271, 277, 288
mixing filters 107, 108
mixture channels 247, 249, 252
mixture components 164, 165, 166, 167, 168, 169, 170, 171, 172, 184
mixture sources 246, 247, 248, 249, 252, 253, 254, 255, 258, 260, 262
model-based approaches 61, 68
modified discrete cosine transform (MDCT) 248, 249
moving sources 107, 109, 113, 114, 121, 122, 123, 124
multi-channel separation 208
multiple sources, localization of 126
multi-speaker environments 207
music 62, 63, 69, 78
musical cognition 371
musical features 371, 372, 373, 378
musical information dynamics 371, 372
musical listening modeling 373
musical structure 371, 372, 373, 382, 390, 394
musical surface 373
MusicBrainz 339
music information retrieval (MIR) 23, 24, 31, 51
music onset detection 315, 316
music recommendation 334, 335, 349
music segmentation 318
music signal processing 297
music thumbnailing 318, 319, 330, 332
mutual information (MI) 188
Index

N
natural audition 80, 82
natural sounds 63
neural network models 372
neuromorphic hierarchical architecture 461
neuromorphic hierarchical architecture for speech processing (NHASP) 448, 473
neuromorphic knowledge 449
neuromorphic speech processing 447, 448, 449, 466, 473,
noise burst unit (NB) 455, 456, 459, 461, 462, 463, 473
non-negative tensor factorization (NTF) 62, 266, 269, 288, 289, 290, 291
Normalized cut criterion (Ncut) 27, 28, 29, 30, 44

O
objective room acoustic parameters 424
onsets, hard 297, 301, 311
onsets, soft 297, 298, 301, 303, 304, 312, 313, 314

P
Pandora.com 335
parallel factor analysis (PARAFAC) 186, 190, 194, 195, 196, 197, 198, 199, 200, 203, 204, 205
particle velocity 140, 161
perceptual gist 80, 92, 93, 94, 103, 105
perceptual linear prediction (PLP) 319, 320, 332
perceptual spectral flux (PSF) 319, 320
peripheral auditory system (PAS) 448, 449, 473,
physical realizability 105
physiological level (PhL) 448, 449, 473
platonic solid 160
post-processor scheme 208
predictive information (PI) 371, 372, 374, 375, 376, 377, 383, 388
predictive information rate (PIR) 372, 374, 375, 377, 378, 388, 389
primary auditory cortex 471, 473
principal component analysis (PCA) 267, 280, 293, 400, 406, 407, 408, 409, 410, 413, 416
probabilistic sound models 162
pseudo random sequences 424
psychoacoustical level (PsL) 448, 449, 473

Q
quasi-stationarity 99, 105

R
recommendation 352
relative estimation error (REE) 363, 364, 365
resonator time frequency image (RTFI) 298, 299, 301, 303, 304, 311, 312
reverberation radius 105
rhythm 318, 319, 320, 321, 323, 325, 326, 330, 331, 332, 333
robotic navigation 1
room acoustic parameters 424, 429, 440, 441, 445
RT30 427
RT60 427

S
second order blind identification (SOBI) 189
second order statistics (SOS) 208
selfsimilarity matrices 319
separation algorithms 207, 208, 209, 211, 214, 218, 240, 241
separation systems 162, 163, 164, 167, 177, 178, 180
Index

short-term Sinusoidal (STS) model 31
signal degradation 208
signal processing 163, 164, 180
signal-to-distortion ratio (SDR) 44, 45
signal to noise ratio (SNR) 427
simple Markov chains 372
sine sweeps 424
singular value decomposition (SVD) 186, 195
SNMF2D 358, 362, 363, 364, 365
SNMF2D-LS 358, 362, 363, 364, 365
social tags 334, 335, 339, 340, 349, 352
sound 80, 81, 84, 86, 102, 103, 105
sound onset labelizer (SOL) 311
sound, propagation of 424, 425, 426
sound scenes 162, 163
sound source localisation 126, 127, 129, 130, 133, 134, 136, 145, 153, 154, 158, 160, 161
sound sources 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 182, 183, 184
source separation 61, 62, 63, 69, 70, 73, 74, 75, 76, 77, 78, 79, 162, 163, 167, 183, 184, 207, 208, 239, 240, 241, 242, 243, 244, 245
source separation, linear approaches 208, 209, 218, 220, 221, 224, 225, 227, 228, 230, 231, 233, 234, 235, 240, 241, 245
source separation, non-linear approaches 209, 224, 225, 227, 230, 231, 240, 241, 245
space alternating data augmentation (SADA) algorithm 279
space-alternating generalized expectation-maximization (SAGE) algorithm 274, 279
sparse coding 61, 338, 347, 348
sparse component analysis framework 246
sparse component analysis (SCA) 247, 248, 249, 251, 252, 253, 255, 258, 260, 261, 262
sparse orthogonal transforms 246
spatial filtering 161
spectral clustering 24, 26, 27, 28, 29, 34, 35, 44, 45, 57, 58, 59, 60
spectral template-based models 162
specular reflection 161
speech 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79
speech signal sources 186
speech transmission index (STI) 425, 426, 428, 433, 434, 435, 436, 437, 439, 440
support vector machine (SVM) 339, 343, 347
T
tags 336, 337, 347
temporal cross-correlation functions 188
tensor factorization (TF) 186, 187, 190, 195, 200
timbre 318, 319, 320, 321, 323, 325, 326, 330, 331, 333
timbregram 318, 319, 320, 321, 322, 323, 324, 325, 326, 327, 328, 329, 330, 332
time-derivative 318
time-frequency atoms 22, 23, 54
time-frequency (TF) transform 247, 249, 252, 254
transform coefficients 246, 247, 259
two-dimensional deconvolution algorithms 354
U
unmixing filters 107
unstructured audio 1
user affect 398, 399, 400, 404, 405, 409, 410, 414, 415, 416, 417, 421, 422
user emotion 398, 399, 400, 401, 402, 403, 404, 405, 406, 407, 409, 410, 411, 413, 414, 415, 416, 418, 419, 421, 423

V
variance modeling 162, 164, 170, 172, 173, 178, 179
vector quantization (VQ) 267, 271
Index

video information 108
video system 108
visual modality 107
vowel-consonant-vowel triphone (V-C-V) 473,

W
Web 2.0 recommender systems 334, 350
white noise 424, 433

Z
zero-crossing rate 338