

# Index

- 3rd Generation Partnership Project (3GPP), 15, 90, 99-101
- a posteriori density, 192
- a priori density, 193
- accessibility, 283
- acoustic channel, 360
- acoustic model, 4, 216, 260, 322
  - compression, 235
  - footprint reduction, 314-315
- adaptation, 333
- adaptive multi-rate (AMR) codec, 10, 45, 99-101
- advanced front-end (AFE), 89, 96-97
- Akaike's Information Criterion, 238
- always-on, 7, 351
- ASR-decoder based concealment, 64
- ASR performance, 131
- audio coding, 28-30
- Aurora, 89
  - Aurora-2 database, 97, 153-154
  - Aurora-3 database, 97-99
  - Aurora-4 database, 97
  - ETSI STQ, 89, 284
  - speech databases, 97
- automatic repeat-request (ARQ), 384
- automotive, 351, 357-360
- battery energy, 376, 382
- battery lifetime, 6, 256, 314, 375
- battery technology, 375-376
- Bayesian decision rule, 3, 190
- Bayesian Information Criterion (BIC), 238
- beam pruning, 223, 266
- Bhattacharyya distance, 243
- bilinear transform, 52
- binary symmetric channel, 197
- bit rate, 12, 29, 131, 382
- bit-rate scalability, 138
- bitstream-based approach, 46-50
- Bleu score, 343
- block quantization, 137
- Bluetooth, 389-391
- Bose-Chaudhuri-Hocquenghem (BCH) code, 174
- burst packet loss model, 68
- cache, 214-215, 219
- call control XML (CCXML), 283
- car noise, 361, 363
- centroid, 195
- cepstrum liftering, 148-150
- channel coding, 13, 163-185
- channel erasure, 168
- channel model, 166-167
- client-side, 384, 392
- clustering, 242-243
- co-articulation, 260
- coded speech, 99-101
- codevectors, 137
- codewords, 133, 169, 174
- command and control, 303, 354
- compensation
  - channel errors, 56-57
  - mobile effect, 53-54
  - speech coding distortion, 54-56
- composite multimodal system, 287
- compression, 94
- concatenate TTS, 341
- concept based translation, 329
- conditional independence, 191
- conditional loss probability, 198
- connected word recognition, 303
- context, 217-218
- continuous speech recognition, 303
- convolutional code, 174-175, 182
- CPU
  - limitations, 215
  - StrongARM-1100, 272
  - XScale PXA270, 272
- cyclic code, 174, 182
- cyclic redundancy code (CRC), 174, 182
- DARPA,
  - naval resource management (RM) task, 270
  - switchboard task, 270

- dataword, 169
- decoded speech-based approach, 45
- decoding
  - hard-decision, 171
  - soft-decision, 171
  - $\lambda$ -soft, 173
- determinization, 227-228
- determinization and minimization, 339
- diagnostic rhyme test (DRT), 102
- dialog engine, 351
- dictation, 303
- discrete cosine transform (DCT), 112-113, 382
- Distributed Multimodal Synchronization Protocol (DMSP), 87-106, 294, 297
- distributed speech recognition (DSR), 2, 11-15, 87-106, 107, 131, 187, 284-285, 384
- DSR
  - advanced front-end (AFE), 89, 96-97
  - extended front-end (XFE), 90
  - performance, 132
  - system, 15
- dynamic voltage scaling (DVS), 380
- electro-acoustics, 91
- embedded speech recognition (ESR), 3, 15-20, 211-275
- embedded speech synthesis, 340-341
- embedded system, 349, 359, 379
- energy-aware, 376-377
- enrollment, 348, 351
- entropy coder, 132
- error concealment (EC), 14, 70, 188
- error correction, 389
- error detection, 93
- ESR, *see* embedded speech recognition
- ETSI Aurora, 89, 284
- European Telecommunications Standards Institute (ETSI), 89, 163, 181-182, 284
- evaluation of recognizer, 367-372
- evidence integral, 238
- extended front-end (XFE), 90
- fast Fourier transform (FFT), 111, 380
- feature compression, 94
- feature enhancement, 54
- feature extraction, 11, 94, 96, 109-117, 215-217, 379
- feature space maximum likelihood regression (fMLLR), 217, 333, 341
- feature transform, 50-53
- finite state transducer (FST), 221, 223, 226, 263
- fixed-point, 6, 15, 17-18, 257, 259, 264, 381-383
  - programming, 257-258
  - representation of Gaussian means and variances, 267
- floating-point, 257
  - block, 258, 268
  - software emulation, 273
- forward-backward algorithm, 199
- forward error correction (FEC), 70, 167-175, 389
- frame dropping, 310
- frame erasure rate, 79
- front end, 94-104, 108, 216-217, 362, 379-384
- fundamental frequency prediction, 123-128
- G711, 30, 34-37
- G723.1, 30, 34-37, 67
- G729, 67, 71
- Gaussian, 216, 218
- Gaussian mixtures models (GMM), 217
- generator matrix, 170, 174
- Gilbert-Elliott model, 31, 68, 165, 198
- Global System for Mobile Communications (GSM), 29, 37
- grammar, 223, 227
- handheld device, 332-334
- hard-decision decoding, 171
- hardware platform, 330
- help system, 354
- hidden Markov model (HMM), 3, 216, 234, 260
  - CDHMM, 234
  - decision tree, 341
  - multi-stream, 242
  - parameter representation, 243-245
  - parameter space, 235
  - qHMM, 245-247
  - SCHMM, 241
  - SDCHMM, 242, 247-249
  - state, 260-261
  - state duration model, 262
- Hidden Markov Model Toolkit (HTK), 94
- hybrid speech recognition, 291
- IEEE 802.11, 255, 384
- incremental unsupervised adaptation, 333
- index generator, 195
- infotainment, 348-349, 351, 356
- integerization, 334
- intelligibility, 102
- interaction manager (IM), 286
- interactive voice response (IVR), 280, 351
- interframe correlation, 74, 144-147, 189
- interleaver
  - convolutional, 179-180, 183
  - decorrelated block, 180-181, 183
  - delay, 178
  - optimal spread block, 178-179, 183
  - spread, 178
- interleaving, 70-71
- interlingua, 329
- Internet, 356

- Internet Engineering Task Force (IETF), 104, 297
- Internet Protocol (IP), 32
- intraframe correlation, 142-144
- isolated word recognition (IWR), 303
- IVR, *see* interactive voice response
  
- Java J2ME, 293, 296
  
- K-means algorithm, 243
- keyword spotting, 303
- Kullback-Leibler divergence, 200, 243
  
- labeller, 215
- language model, 4, 221, 228, 260, 322, 263
  - n-gram, 227, 229, 263
- language resources, 322
- large vocabulary continuous speech recognition (LVCSR), 332-334
- lattice, 216, 224-226
- LC-STAR corpora, 309
- LC-STAR II corpora, 309
- lexicon, 4
- line spectrum pairs (LSP), 46
- linear block code, 169-174
- linear predictive coding (LPC) coefficient, 42, 72
- logarithm, 382
- logarithmic spectral distortion, 134
- low-power, 376, 385
- LPC-based MFCC (LP-MFCC), 52
- LPC-derived cepstral coefficient (LPCC), 47
- LVCSR, *see* large vocabulary continuous speech recognition
  
- Mahalanobis distance, 249, 265, 267
- manual, 351-355
- marginalization, 201
- maximum a posteriori (MAP), 193
- maximum entry modeling, 336
- mean loss probability, 198
- mean opinion score (MOS), 65
- mean squared error (MSE), 134
- media gateway, 282
- media-independent FEC, 168-175
- Media Resource Control Protocol (MRCP), 281-284, 350
- media-specific FEC, 167-168
- Mel-cepstrum, 89
- Mel-frequency cepstral coefficient (MFCC), 11, 42, 72, 109, 141, 264, 270, 382
- Mel-scaled LPCC (MLPCC), 51-52
- Mel-scaled LSP (MLSP), 53
- Mel-scaled PCEP (MPCEP), 53
- memory limitation, 214
- MFCC, *see* Mel-frequency cepstral coefficient
- MFCC-based speech coder, 72-74
- microphone, 360
  
- minimization, 222-223, 226-228
- minimum mean square error (MMSE), 189
- mobile device, 1, 5, 279, 328
- mobile phone, 301-325
- morphological analysis, 333
- Motor Industry Software Reliability Association (MISRA), 359
- MP3, 352-353
- MPEG, 30, 34-35
- MRCP, *see* Media Resource Control Protocol
- multiliguality, 305-309
- multimodal
  - applications, 296
  - architectures, 295-297
  - local search, 298
  - user interaction, 279
  - user interfaces, 283-284
- multiple description coding (MDC), 70
  
- N-best, 228
- n-gram, 227, 229, 263
- name dialling, 302, 305-308
- natural language understanding, 349
- natural language understanding and generation
  - based translation, 334-336
- navigation, 349, 353
- nearest frame repetition, 204
- nearest neighbour VQ, 138
- NetMeeting, 30, 32
- network, 1, 7
- network speech recognition (NSR) 2, 9-10, 25-84, 28-32, 63, 107-109, 187, 280
- noise reduction, 260
- noise robustness, 88, 19-20, 309-314
- non-composite multimodal system, 287
- normalization of log-likelihoods, 266
- NSR, *see* network speech recognition
  
- observation model, 215-221
- Open Mobile Alliance (OMA), 285
  
- packet erasure channel, 188, 198
- packet loss, 10, 30-34
  - burst, 68
  - rate, 66
- packetization, 13, 163, 182
- parameter tying, 239-243
  - density level, 241
  - model level, 240
  - state level, 241
  - subspace level, 241
- parity-check matrix, 170
- parity matrix, 170
- PDF estimation, 139
- Pearce Principle, 291
- perceptual evaluation of speech quality (PESQ), 78

- perceptual linear predictive (PLP), 12, 50, 109, 270
- personal digital assistant (PDA), 327
- PESQ, *see* perceptual evaluation of speech quality
- platform, 16-17, 319-323
- point estimate, 190
- power consumption, 378, 385
- power spectra
  - autocorrelation function, 361
- predictive vector quantizer (PVQ), 74
- prefixes and suffixes, 333
- processing engine (PE), 285
- pronunciation, 352
- pseudo-cepstrum (PCEP), 53
- quantization, 244-245, 382
  - block, 137
  - noise shaping, 134
  - scalar, 133, 135-136, 245-247
  - vector, 74-78, 133, 137-138, 247, 382
- quantizer, 195, 259, 267
- radio station selection, 352
- rank, 219, 220, 223
- rate compatible punctured codes (RCPC), 175
- rate control, 69-70
- rate-distortion trade-off, 133
- Real Time Protocol (RTP), 66, 90, 350
- recognition performance, 97-99
- recognizer,
  - embedded, 349-350
  - server, 350
- recursive least squares, 364-365
- Reed-Solomon (RS) code, 174
- RTP, *see* Real Time Protocol
- safety-net predictive vector quantizer, 74
- scalar quantization (SQ), 133, 135-136, 245-247
- search graph, 216, 221-223, 226-228
- semantic parser, 334
- sequential multimodal system, 287
- signal degradation, 164-165
- signal-to-noise ratio (SNR), 367-372
- simultaneous multimodal system, 287
- Skype, 30
- soft-decision decoding, 172
- soft feature, 193
- soft Viterbi decoding, 56
- source coder, 195
- source coding, 12-13, 28-30, 131-161
- speaker
  - characterization, 348, 351
  - recognition, 27, 35-38, 259, 348, 351
  - verification, 35-38
- speaker adapted system, 303
- speaker dependent system, 303
- speaker independent system, 303
- spectral distortion, 46, 134
- spectral subtraction, 310
- speech coder, 46
- speech coding, 9
  - adaptive multi-rate codec, 99-101
  - G711, 30, 34-37
  - G723.1, 30, 34-37, 67
  - G729, 67, 71
- speech database, 97-99
  - Aurora, 97
  - Aurora-2 database, 97, 153-154
  - Aurora-3 database, 97-99
  - Aurora-4 database, 97
  - SpeechDat Car, 309
  - SPEECON, 309, 366
- speech feature enhancement, 190
- speech intelligibility test, 102
- speech quality, 37, 65-66
- speech recognition, 3-5, 259
  - noisy channel formulation, 260
- speech reconstruction, 102, 117-123
- speech synthesis, 280
- speech-to-speech translation, 327-346
- speech-to-text, 303
- SpeechDat Car database, 309
- SPEECON database, 309, 366
- split vector quantizer (SVQ), 195
- standards, 88, 279
  - DSR, 14, 87-106
- statistical machine translation, 330
- stochastic gradient descent, 334
- StrongARM, 272, 379-380
- subspaces, 241-242, 247-249
- subvector, 195
- table look-up, 258, 269
- text-to-speech (TTS), 340, 348, 351
  - embedded, 349
- title selection, 353
- traceback, 222, 224
- transcription systems, 280
- transform coding, 137
- transition probabilities, 220
- transcription test, 102
- translation, 327
  - concept based, 329
  - natural language understanding and generation based, 334-336
  - speech-to-speech, 327
  - statistical machine, 330
  - two-way free-form, 327
  - weighted finite state transducer based, 337-340
- Transport Protocols, 104
- two-way free form translation, 327

- uncertainty decoding, 189, 206
- unconstrained VQ, 138
- unequal error protection (UEP), 176-177, 182
- usability, 284, 323
- user agent (UA), 285
- User Datagram Protocol (UDP), 66
- user interface, 6
  
- validation set, 237
- vector quantization (VQ), 74-78, 133, 137-138, 247, 382
- video interactive services, 295
- Viterbi decoder, 263
  - beam-width, 266
- Viterbi search, 221-229
- voice activity detection (VAD), 96, 321
- voice applications, 279, 351-357
- voice over Internet Protocol (VoIP), 66, 282-283
  
- voice server, 281, 283
- voice web, 280-283
- voicing prediction, 123-128
- VoiceXML, 10, 281-283, 289, 293
- VoiceXML Forum, 281
  
- Web content, 355-356
- Web service, 356-357
- weighted finite-state transducer, 337
- weighted Viterbi decoding, 169, 202
- weighting rules, 363-366
- WinCE.NET, 330
- word decoder, 321
- World-Wide Web Consortium (W3C), 281, 297
  
- XHTML, 289, 297
- XHTML+Voice Profile (X+V), 289, 293, 297