

# Index

- 11-point average precision, 177–179
- acceptance
  - by speaker recognition systems, 17, 18, 20
  - by users, 209, 211, 215
- accuracy, 1, 10–12, 14, 15, 23, 70, 76, 110, 111, 113, 118, 135–138, 142–144, 150, 172, 173, 179–181, 183, 184, 200, 213, 214, 228, 246
- Adapt, 87–91
- adaptation, 6, 15, 139
- adequacy evaluation, *see* evaluation
- adjudication, 242, 257–259
- American National Corpus, 223, 280–282
- animated agent, 65–68, 70–74, 76, 77, 80, 81, 85, 87, 88, 90, 188, 193, 216
- animated face, *see* animated agent
- animated head, *see* animated agent
- annotation, 103, 104, 110, 112, 114–118, 221, 224–226, 229, 232, 234–238, 242, 243, 254–259, 264–273, 275–283
- annotator agreement, 112, 182, 255–257
- anomaly resolution, 31
- attractiveness, 138, 139, 152
- audio-visual speech synthesis, 66–69, 71–77, 80, 85, 87
- August, 87, 90, 91
- average ambiguity, 112
  
- believability, 40, 66, 80, 91, 205
- blackbox test, *see* evaluation methods
- broadcast, 4, 5, 39, 224, 226, 228, 233, 235, 236, 238–240, 246–250, 252, 253
  
- cepstral coefficient, 15
- COCOSDA, 43, 44, 223
- common task program, 227–230, 232
- concatenative speech synthesis, 35, 39, 41
- confusion matrix, 9, 10
- continuous word scoring, 10
- conversation success, 205, 217
- conversational agent, *see* animated agent
- conversational gesture, 66
  
- conversational speech, 4, 5, 32, 45, 224, 226, 228, 231, 233–236, 240, 247–253, 257
- correctness, 9, 37, 182
- cross-entropy, 112
- crossing brackets, 115
  
- data category, 272–275, 278, 280, 282
- data collection, 69, 70, 89, 91, 222, 224, 226, 230, 232, 234, 236, 237, 241–248, 250–254, 256
- delicacy, 112, 113
- dependency, 31, 103–105, 115, 116, 281
- detection error trade-off (DET) curve, 19, 20
- development data, 22, 230
- diagnostic evaluation, *see* evaluation
- dialect, 238
- dictation, 5, 12, 137, 144
- discrepancy, 256, 257
- document retrieval, 149, 151, 155, 164, 165, 236, 258
  
- EAGLES, 106, 126, 127, 131, 133, 140, 141, 146–148, 154, 155, 188, 264, 266, 267, 278
- ease of use, 204, 205, 277, 282
- edutainment, 188, 193, 202, 207
- effectiveness, 1, 21, 66, 71, 81, 140, 182
- efficiency, 66, 90, 138, 139, 144, 145, 192
- embodied agent, *see* animated agent
- emotion, 22, 33, 45, 66, 67, 69, 71, 85, 198, 199, 201, 205, 207
- emotions, 32
- entropy, 7
- equal error rate, 1, 19
- error correction, 12
- error handling, 205
- error rate, 9, 10, 12, 18, 19, 22, 23, 102, 118, 148, 155, 163, 164, 192, 200, 228, 231
- error recovery, 11, 12
- European Language Resources Association, 43

- European Language Resources Association (ELRA), 223, 226, 229, 231, 234, 241
- evaluation
- adequacy, 23, 106, 133, 134
  - comparative, 23, 36, 70, 90, 114, 116, 132, 134
  - diagnostic, 23, 36, 38, 39, 106, 132, 134
  - objective, 1, 6, 7, 9, 11, 12, 24, 106, 163, 164, 168, 170, 189, 198, 202, 207, 227, 228, 256
  - performance, 1, 2, 5–7, 9–12, 17, 19, 22–24, 37–40, 46, 102, 106, 109–113, 117, 130, 138, 140, 145, 147, 152, 165, 166, 172, 174–176, 179, 181–185, 192, 197, 199–202, 204, 206, 214, 228, 229, 233, 256
  - progress, 106, 133, 134, 188, 192, 196
  - qualitative, 106, 189, 193, 198, 199, 209
  - quantitative, 23, 72, 106, 117, 189, 198, 199, 204, 205, 209, 212–215
  - subjective, 1, 6, 11, 12, 72, 74, 90, 106, 163, 169, 182, 184, 189, 202–206, 211–215
  - technical, 188–190, 192, 198, 200, 202
  - technology, 24, 106, 117, 126, 148, 223, 227, 229–232, 234, 237, 240
  - usability, 138, 139, 152, 188–190, 192, 193, 198, 202–209, 212–214
  - user oriented, 106, 117, 125, 127–129, 137, 139, 143, 154, 189
- evaluation data, 229, 230
- evaluation methods
- blackbox, 106, 117, 190
  - expert, 114, 117, 191
  - field test, 191, 196, 198, 209, 212
  - function test, 190
  - glassbox, 190
  - guideline-based, 190, 191
  - heuristic, 191
  - high-fidelity prototype, 191, 196
  - integration test, 190
  - interview, 192, 196, 202, 206, 207, 211, 212
  - laboratory test, 190, 191, 196, 197
  - low-fidelity prototype, 191
  - questionnaire, 192, 202, 211, 213
  - think-aloud, 191
  - unit test, 190
  - walkthrough, 190, 191
  - whitebox, 106
  - Wizard-of-Oz, 89, 90, 191, 195, 196, 198
- expert evaluation, *see* evaluation methods
- expressiveness, 30, 39, 44, 85, 87
- extrinsic criteria, 105
- eyebrow movement, 66, 67, 70, 77–80, 83, 84
- f-measure, 112, 116, 117, 174, 176, 180
- facial gesture, 90
- feature structure, 105, 271, 272, 279, 283
- feedback, 12, 66, 67, 80–82, 84, 90, 92, 212
- FEMTI, 126, 130, 153, 154
- field test, *see* evaluation methods
- Fisher English corpus, 226, 234, 247, 250, 252
- friendliness, 85–87
- full information capture, 243, 244
- functionality, 135, 137–139, 141, 142, 144, 189–191, 198, 202
- glassbox test, *see* evaluation methods
- guideline-based evaluation, *see* evaluation methods
- Hans Christian Andersen system, 187, 193–201, 203–207, 216, 217
- head movement, 66, 70, 77–80, 84, 91, 92
- heuristic evaluation, *see* evaluation methods
- high-fidelity prototype, *see* evaluation methods
- human speech, 30, 34–36, 38, 44, 46
- in-car system, 188, 207–211, 213–216
- in-line annotation, 235, 236, 257
- information extraction, 164, 233, 234, 239, 255
- information retrieval, 163–166, 169–172, 175–177, 179, 181–184
- information retrieval evaluation, *see* information retrieval
- informed consent, 197, 239, 240, 248
- integration test, *see* evaluation methods
- intellectual property right, 239
- intelligibility, 12, 30, 36, 37, 39, 40, 44, 67, 71–76, 87, 150, 201
- inter-annotator agreement, *see* annotator agreement
- interaction problem, 204
- interview, *see* evaluation methods
- intrinsic criteria, 105
- IR, *see* information retrieval
- ISLE, 126, 127, 140, 147, 153, 155, 264, 266, 267
- ISO standard
- for language resources management, 264, 268, 273, 274, 283
  - for software evaluation, 125, 127, 135, 138–147, 153, 155
- isolated word scoring, 9
- Kendall's tau, 172, 182
- laboratory test, *see* evaluation methods
- LAF, *see* Linguistic Annotation Framework
- language model, 3, 7, 8
- language resources, 221–226, 228–247, 249–257, 259, 263–269, 271–273, 275, 277, 278

- learnability, 138, 139  
lexical description, 100, 101  
likeability, 36, 39, 41  
likelihood ratio, 15, 16  
Linguistic Annotation Framework, 268–272, 275, 278–283  
Linguistic Data Consortium (LDC), 2, 43, 223, 226, 229, 230, 232–234, 239–241, 243, 248, 250–253, 256–258  
low-fidelity prototype, *see* evaluation methods  
  
machine translation, 126, 127, 130, 131, 133, 144, 145, 147, 149–153, 155, 233, 256, 264  
maintainability, 139, 144, 155  
mean average precision, 178, 182  
mean opinion score, 38, 39  
mean reciprocal rank, 179, 180, 184  
meeting, 4, 224, 226, 228, 233, 238–240, 246, 249, 250, 252  
metadata, 236, 241, 242, 255, 257, 265–267, 273–275  
mock-up, *see* evaluation methods under low-fidelity prototype  
multimodal spoken dialogue system, *see* spoken dialogue system  
  
natural language processing, 99, 102, 105, 107, 114, 163, 169, 185, 222, 223, 225, 233, 236, 273, 278  
natural language understanding, 187, 190, 194, 199, 200, 202, 207  
naturalness, 30, 33, 36, 39, 40, 43, 44, 69, 74–76, 189, 201, 205  
NICE, 193, 202  
NLP, *see* natural language processing  
non-task-oriented, 188, 193, 206, 207  
non-verbal communication, 66, 67, 91, 199, 201, 205, 207  
normalization, 6, 15–17, 22  
  
Open Language Archives Community (OLAC), 241, 242, 266  
operability, 138, 139, 155  
out-of-vocabulary (OOV), 8, 11  
  
PARADISE, 90, 188  
parser, *see* parsing  
parser evaluation, *see* parsing  
PARSEVAL, 115, 117, 278  
parsing, 99, 102–105, 107–110, 113–118, 132, 200, 202, 222, 235, 236, 281  
part-of-speech tagger evaluation, *see* part-of-speech tagging  
part-of-speech tagging, 99–102, 105, 107–111, 113, 116, 117, 222, 235, 255, 281  
Penn Treebank, 100, 111, 116, 225  
performance evaluation, *see* evaluation  
  
perplexity, 7, 8, 112, 192  
pitch, 30–34, 85, 86  
portability, 107, 113, 139, 146  
POS tagging, *see* part-of-speech tagging  
precision, 111–113, 115–117, 155, 168, 172–184, 256  
prominence, 66, 77–80, 84, 91  
prosody, 30, 32–35, 38, 40, 45, 80, 81, 84, 88–92  
  
quality assurance, 127, 242, 255  
quality control, 256–258  
quality in use, 138, 140, 141  
quality model, 125, 127, 135, 139–146, 148, 153, 154  
quality requirement, 133, 135, 153  
quantization, 242–244  
question answering, 163, 164, 166–172, 179, 181–185  
question answering evaluation, *see* question answering  
questionnaire, *see* evaluation methods  
  
raw data, 229, 234–238, 242, 246, 254, 257  
RDF/OWL, 266, 267, 269, 271, 276–278, 280, 282  
real-time performance, 192, 199–201, 206  
recall, 112, 115–117, 155, 172–177, 179–181, 184, 256  
receiver operating characteristic, 17, 18, 175  
relevance judgement, 149, 165, 170, 171, 176, 180–182, 184  
reliability, 40, 138, 139, 150–152, 190  
response generation, 201  
response time, 11, 12, 40  
  
sampling resolution, 242, 243  
scalability, 181  
search engine, 137, 151, 164, 167, 169, 241, 242, 257, 258  
segmentation, 5, 6, 68, 69, 99, 108–110, 113, 144, 221, 224, 235, 242, 244, 253, 254, 283  
SENECA, 187, 207–217  
situational awareness, 11, 12  
speaker identification, 13, 21, 228, 236, 250  
speaker recognition, 1, 12–15, 17–19, 21–23  
speaker recognition evaluation, *see* speaker recognition  
speaker verification, 13–15, 17–20, 225  
speaking style, 35, 40, 41, 44  
speech recognition, 1–12, 22–24, 148, 164, 192, 202, 214, 222, 225–227, 233, 236, 246, 249, 250  
speech recognition evaluation, *see* speech recognition  
speech synthesis, 29–48, 64, 66–69, 71, 77, 85, 87, 233

- speech synthesis evaluation, *see* speech synthesis
- spoken dialogue system, 11, 12, 23, 65–67, 81, 87–92, 187, 188, 192–213, 215–217
- spoken dialogue system evaluation, *see* spoken dialogue system
- stand-off annotation, 235, 236, 266, 270, 271, 279, 281, 283
- suitability, 135, 137–139, 141–145
- summarization, 24, 233, 235, 239
- Switchboard corpus, 224, 225, 236, 250–253
- tagset, 100, 110–113, 118
- talking head, *see* animated agent
- task completion, 11, 192, 211, 213–215, 217
- task-oriented, 188, 207, 216
- taxonomy, 153, 154
- telephone speech, 224, 226, 228, 231, 233, 234, 236, 239, 246, 247, 249–252, 257
- test, *see* evaluation
- test collection, 165, 170, 172, 179, 184
- test data, 7, 8, 22, 23, 74, 114, 132, 133, 143, 148, 149, 230
- test set perplexity, *see* perplexity
- test suite, 110, 114, 116, 132, 133, 190
- think-aloud, *see* evaluation methods
- training data, 14, 22, 23, 111, 132, 151, 171, 229, 238, 243, 269
- TREC, 149, 155, 163–165, 169, 170, 176, 177, 179–184, 227, 258
- TREC-QA, 117, 164, 166–172, 182–185
- treebank, 100, 103, 111, 114, 116, 118, 222, 225, 226, 255
- turntaking, 66, 67, 81, 89, 90, 92
- understandability, 138, 139
- unit test, *see* evaluation methods
- user oriented evaluation, *see* evaluation
- user preference, 35, 38, 74, 75, 193, 211, 215
- user satisfaction, 90, 147, 184, 192, 206, 212, 214
- validity, 148, 149, 151, 152
- virtual talking head, *see* animated agent
- visual prosody, 66
- voice quality, 30, 33, 42, 44, 203
- walkthrough, *see* evaluation methods
- waveform generation, 30, 34, 39, 40
- weighted confidence, 179, 180, 182–184
- whitebox test, *see* evaluation methods
- Wizard-of-Oz, *see* evaluation methods
- word segmentation, *see* segmentation
- WordNet, 267, 278, 280, 282
- XML, 31, 69, 115, 194, 257, 258, 266, 267, 269, 270, 279, 281–283