CAMBRIDGE

Cambridge University Press
978-1-107-02229-4 - Multimodal Signal Processing: Human Interactions in Meetings
Edited by Steve Renals, Hervé Bourlard, Jean Carletta and Andrei Popescu-Belis
Index
More information

# Index

abstractive summarization, *see* summarization, abstractive
acoustic modeling for meetings, 64
action item, 22, 177
addressing in conversations, 158
  modeling, 164–165
agenda, 22
agreement, 131
agreement and disagreement detection, 136
AMI Consortium, 6, 58, 202, 232
AMI Meeting Corpus, 11, 49, 59, 60, 65, 132, 136, 149, 200, 224
annotation dependencies, 24
annotation reliability, 21, 24
annotator agreement for summarization, 188
Archivus browser, 210, 227
ASR, *see* automatic speech recognition
attention in conversations, 156–157
attitude annotations, 133
Automatic Content Linking Device (ACLD), *see* meeting assistant, content linking
automatic speech recognition, 19, 25, 58–80, 215
  accuracy on AMI Corpus, 66
  AMI system, 69–76
  challenges in meetings, 57
  domain, 58
  history in meetings, 57
  on the Web, 76
  online, 77–80
  resource optimization toolkit (ROTK), 73
  speed vs. error rate, 78

backchannel, 22
Basic Elements evaluation method, 187
Bayesian Information Criterion (BIC), 41
Bayesian Network (BN), 130
beamforming, 28
  blind, 38
  definitions, 29
  delay-sum, 31
  filter design, 34, 36
  MVDR, 31
  superdirective, 36
Browser Evaluation Test (BET), 223–224
  query analysis, 200

CALO project, 6
CHIL project, 5, 58, 59, 232
CLEAR evaluation, 7, 25, 98
CLEF evaluation, 25, 220, 227
cocktail party effect, 209
contextual information, 153
coreference resolution, 172, 182
Creative Commons, 26

decision, 22, 142
  audit, 177, 189
  detection, 142–143, 177
design teams, 12
dialocalization, 49–55
dialogue act, 125–131
  annotation, 22, 126–127
  annotation schemes, 126
  automatic classification, 129, 130
  automatic segmentation, 127, 130
  online classification, 129
  segmentation by humans, *see* segmentation annotation, dialogue act
disagreement, 131
discourse marker, 148
discourse memory, 146
discourse model, 142
disfluency, 148, 182
domain model, 143
dominance in conversations, 49, 165
  manual annotation, 165
  modeling, 166
Dynamic Bayesian Network (DBN), 130, 160, 162

emotion, 131
ethnographic approach, 197
evaluation
  extrinsic, 188
  intrinsic, 186
  task-based, 223–229
  user-centric, 229
evaluation standards, 218

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-1-107-02229-4 - Multimodal Signal Processing: Human Interactions in Meetings
Edited by Steve Renals, Hervé Bourlard, Jean Carletta and Andrei Popescu-Belis
Index
More information



extractive summarization, *see* summarization, extractive

face authentication, *see* face verification
face detection, 95, 105–109
face identification, 117
face recognition, 115–117
face verification, 117
Factored Language Model, 131
floor, *see* turn-taking
focus of attention, *see* visual focus of attention
forced alignment, 19
FriDoc browser, *see* JFriDoc browser

Gabor Wavelet Networks, 109
gaze, *see* visual focus of attention
gaze detection algorithm, 111
gesture
  hand, 21
    automatic recognition, 123
  head, *see also* head pose estimation, 22
group decision support, 204
group effectiveness, 11, 12, 229

hand gesture, *see* gesture, hand
head gesture, *see* gesture, head
head pose estimation, 25, 94–99, 159
  and joint head tracking, 96
  audio-visual, 101
  failure detection, 96
  Rao–Blackwellized particle filter (RBPF), 97
hesitation, 148

iCLEF, *see* CLEF evaluation
ICSI Meeting Corpus, 59, 126
ICSI Meeting Recorder project, 4, 58
IM2 project, 5, 6, 202, 232
individual head microphone (IHM), 69
information bottleneck (IB), 43
  for speaker diarization, 44–45
  performance of, 46
instrumented meeting room, 4, 7, 14–19, 60, 68, 84, 98
interaction capture and retrieval, 204
interregnum, 148
ISO dialogue act standard, 127

JFerret, *see also* meeting browser
  browser, 208, 226
  platform, 208
JFriDoc browser, 211, 227

*k*-kappa, 140
*k*-precision, 140
*k*-recall, 140
keyword spotting (KWS), 80–82, 215

language annotation, 20
language models for meetings, 61–63
LCseg lexical cohesion model, 141
lecture browser, *see* meeting browser, for conferences
lexical cohesion, 141
local binary patterns (LBP), 118, 123

M4 project, 5, 162
machine learning, 11, 13, 20, 22, 25, 171
Markov chain Monte Carlo (MCMC), 93, 96
Maximal Marginal Relevance (MMR), 174
meeting assistant
  content linking, 215
  improving user engagement, 214
  in the AMI Consortium, 212–216
  mobile, 213
meeting browser, *see also* JFerret
  audio-based, 209, 226
  benchmark score, 226
  examples, 205
  for conferences, 212, 231
  in the AMI Consortium, 208–212
  modality choice, 201
  practice-centric analysis, 197–199
  requirements elicitation, 199
  typology, 205
  using summarization, 212
mel-frequency cepstral coefficient (MFCC), 40, 50
microphone arrays, 28, 42
  array gain, 30
  definitions, 29
  directivity pattern, 30, 32
  geometry, 33, 35, 38
  localization, 29, 38
  role in ASR, 68
mobile meeting assistant (MMA), *see* Meeting assistant, mobile
multiple distant microphones (MDM), 69

named entity, 22, 23
National Institute for Standards Technology, *see* NIST
NCCR IM2, *see* IM2 project
negotiation act, 143
NIST, 7, 22, 25, 41, 46, 58, 59, 82
NITE XML Toolkit, 22, 26
nonverbal communication, 236

ontology, 143, 179
  BNF, 179
  DOLCE, 143, 179
  SUMO, 179
  vs. taxonomy, 179
open data, 26
optical character recognition, 15
orthographic transcription, *see* transcription by humans

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-1-107-02229-4 - Multimodal Signal Processing: Human Interactions in Meetings
Edited by Steve Renals, Hervé Bourlard, Jean Carletta and Andrei Popescu-Belis
Index
More information



particle filter, 87, 89, 91, 95
  Rao–Blackwellized, *see* head pose estimation
person tracking, *see* tracking
personal notes, 198
$P_k$ evaluation metric, 139
privacy, 2
prosody, 172
public meeting minutes, 198
Pyramid evaluation method, 187

question answering evaluation, *see* CLEF evaluation

reading comprehension, 190
real-time ASR, *see* automatic speech recognition, online
redundancy of summaries, 177
relevance assessment of summaries, 190
reliability, *see* annotation reliability
reparandum, 148
reparans, 148
Rich Transcription evaluation, 7, 25, 38, 46, 58, 62, 69
role play, 12, 13
ROUGE evaluation measure, 187
RT evaluation, *see* Rich Transcription evaluation

segmentation annotation
  dialogue act, 24
  speaker, 19, 24
  spurt, 19
  topic, 22
semantic parser, 144
sentence compression, 172, 177
sentence reordering, 172, 177
skin color model, 95, 103–105, 119
smart meeting room, *see* instrumented meeting room
social cues, 236
software development process
  helix model, 195–196, 230
  spiral model, 194
  user-centric design, 194
  waterfall model, 193
Somasundaran annotation scheme, 133
speaker diarization, 55
  for ASR, 67
  stages of, 40
speaker segmentation, *see* segmentation annotation, speaker
spurt, *see* segmentation annotation, spurt
subjective language, 131
  annotation scheme, 132–134
  automatic detection, 134
summarization
  abstractive, 22, 178–185
  extractive, 22, 171–177
    features, 172–173
  presentation, 147
summarization accuracy, 188
summary post-processing, 176
supervised extraction, 176
switching DBN, 130
synchronization of recorded media, 16

task-centric evaluation, *see* evaluation, task-based
Text Analysis Conference (TAC), 187
tf.idf, 172
time delay of arrival (TDOA), 38, 42, 47
topic labeling, 138, 142
topic segment, *see* segmentation annotation, topic
topic segmentation
  automatic, 140–142
TQB browser, 210, 226
tracking, *see also* head pose estimation
  annotation of, 21
  audio-visual, 99–102
    and diarization, 49
    data fusion, 100
  Bayesian formulation, 86, *see also* particle filter
  challenges of, 88–89
  color-based, 50, 90
  three-dimensional, 90–92
  visual, 88–94
transcription by humans, 19, 22, 24
TREC evaluation, 82, 220, 228
turn-taking in conversations, 22, 157–158
  automatic analysis, 161–164
typed feature structure (TFS), 145, 146

unsupervised extraction, 174
upper ontology, 143, 179
User Engagement and Floor Control (UEFC), *see* meeting assistant

vector-space model, 172
ViCoDe browser, 211
video annotation, 21
videoconferencing systems, 4
visual focus of attention, 157, 159–160
vocabulary used in meetings, 61
voice activity detection (VAD), 79
voxel, 91

weighted f-score, 188
WindowDiff evaluation metric, 139
Wizard-of-Oz, 201

© in this web service Cambridge University Press www.cambridge.org