

# Index

- abstraction 50–51, 72, 74, 89, 111, 182
- accent 69, 87, 100, 113–114, 117, 153, 169
  - inter-accent 273
- accent group 237–239, 262, 271, 288, 313, 325
- Action Theory 51
- additive model 18–19, 291, see expression
- agent 174–175, see CPA
- alignment 55, 280
- allophone 95
  - intrinsic (phonetic) 8, 95, 97, 119
  - extrinsic (phonological) 8, 95, 97, 119
- ANN 79
- appraisal 201
- articulation 10
- articulators 189
- Articulatory Phonology 51
- assimilation 130
- auditory scene 281
- automatic speech recognition (asr) 20, 48,  
79–81, 101, 270
  - and synthesis model 82
- a-linguistic 101–102, 121–123
- biological
  - circuit 208, 298, 306
  - cognitive events 189, 201, 208
  - constraints on perception 64, 89–90
  - emotive content 190, 192, 202, 211, 315
  - expressive properties 188, 192, 299, 302
  - information 188, 208, 211, 233, 306
  - models 197–200
  - processes 5, 85, 185
  - sourced information 103, 174, 207,  
232–234, 237, 297
    - coded 245, 301
    - cognitive intervention 198, 255
    - instantiation 191
    - phonetic prosody 237
    - stance 190–191, 290, 294, 297
    - variability 190
- black box 285
- blends (secondary) 90, 297–298, 306
- classical phonetics, see phonetics, classical
- coarticulation
  - assimilation 77, 130
  - markup 129
  - model 9–10, 28, 44, 55, 141, 325
  - parameter 44
  - theory 52, 74, 79, 90, 95, 265, 273, 280
  - variability 140
- code
  - pseudo-code 13, 214–215, 255–256, 258, 260
  - sample prosodic 24, 247, 286–287
  - sample expression 245, 249, 292, 302, see  
XML
- cognitive
  - information 245, 274
  - physical distinction 85–86
  - representation 30, 272, 274
  - sourced constraints 116, 121, 130
  - sourced expression 9–10, 30, 215, 232, 246,  
301, 313
  - sourced information 197, 245, see  
expression
  - intervention 68, 197, 204, 208
    - borrowed term 9–10, 190–191
    - cognitive phonetic 51, 230, 275
    - control 85, 117, 229
    - naturalness 185, 268
- cognitive phonetics, see phonetics, cognitive
- component 268
- computational, see model
- concatenation 82, 119, 238, 271, 319
  - boundaries 123
  - naturalness 127–135, 278

- concept to speech 93, 321
- constraint  
 global 98  
 runtime 98
- contour 96, 99, 111, 161, 171, 237
- CPA 218, 229, 230, 322, 324–325  
 monitoring 298, 230–231, 325  
 rendering 253, 325, *see* rendering  
 supervision 231, 233, 240–242, 247
- databases  
 annotation 99, 168, 172, 179, 183, *see*  
 labelling and text marking  
 automated 24, 05, 129  
 collection 23–24, 60, 137, 207  
 expressive speech 31  
 large 39, 81, 116, 124, 127, 129, 130–133,  
 178  
 marking 54–55, 99–105, 123, 181  
 multiple 24, 31, 44–46, 61, 103, 105  
 perception 45  
 recorded material 154  
 representation of speech 28, 31, 65, 93, 99,  
 319  
 searchability 139, 142  
 switchable 60, 103, 154  
 synthesis 81, 319, 139  
 types 140  
 variability 77, 99, 102–103, 137
- data driven 32
- data structures 100
- declination 229
- derivation 140, 225, 227–229, 268  
 exemplar 102, 168, 194, 280  
 expressive content 207
- dialogue 148, 307, 309, 313  
 adaptation 81, 140, 148, 150, 181, 208  
 concept based 60, 321–322  
 expressive content 199, 303  
 management 50, 147, 321  
 simulation 1–2  
 systems 1, 13, 49–50, 60, 174, 314, 307,  
 321, 322
- diphone 31, 36, 100, 122, 124
- duration 33, 41, 162  
 accent group 263  
 intrinsic 41, 76, 224–225, 261–262  
 rendering 76  
 segments 25, 30, 42, 76, 171, 265  
 syllable 75–76, 208, 239, 243, 247, 262
- emotive content  
 biologically sourced 191, 197, 289, 297,  
 303, *see* biological  
 cognitively sourced 197, 289, 302, *see*  
 cognitive  
 detection of 185–187  
 example ‘anger’ 200, 298  
 expression 215, 233, 286, 291–292  
 labelling 207, 306  
 model 213, 314  
 naturalness 188, 205, 218  
 phonetic prosody 239, 303  
 prosody 19, 194, 289, 303  
 reported experience 207  
 synthesis 201, 208, 233  
 wrapper 291, 301  
 written language 178
- error 222
- evaluation 1, 5, 7–8, 35, 80, 204, 270–271
- expression 115–116, 197, 308–309  
 biologically sourced 185, 197, 255, 296, 302  
 characterizing 11, 298, 305–309  
 cognitively sourced 130, 185, 198–199, 201,  
 215, 232, 289–290, 301, 309, 313  
 cognitive phonetics, *see* cognitive phonetics,  
 CPA  
 emotion 31–32, 197, 211, 230, 285–286,  
 302, 313  
 grammar of 300  
 information channel 22, 63, 91, 104, 218  
 level of representation 281  
 labelling 8, 99, 103, 151  
 model 21, 24, 138, 185, 215, 230, 240, 249,  
 257, 285–286, 302, 313  
 neutral 19, 96, 131, 227  
 percept 306–307  
 pragmatics 6, 230, 266, 309, 311  
 speaker 1, 12, 21, 63, 112, 115, 140, 172,  
 234  
 synthesis 11–12, 21, 24, 43–44, 119, 121,  
 123, 154, 188, 204, 294–296  
 tone of voice 21  
 wrapper 98, 237, 241, 289, 291–293,  
 299–300, 301, 309, 313, *see* wrapper  
 XML 203, 215, 245, 249, 291–297, *see* XML
- expressive content 57, 11, 17, 185, 197, 207,  
 233, 296, 245, 285  
 phonetic prosody 62, 104, 186, 210, 305  
 prosody 7, 18–19, 31, 63, 96, 98, 111–112,  
 117, 167, 237, 309

- experiment 7  
   data collection 209–210, 150, 153, see  
     databases  
   limits, see model  
   non-theory motivated 33  
   theory driven 105, 112, 132, 198, 204–205,  
     208, 210, 298  
 extensibility 155  
  
 Festival 74  
  
 high level synthesis, see synthesis  
 Holmes synthesis system 6, 24–27, 32, 40, 76,  
   286  
 HLSyn 6  
 HMM 79  
 hypothesis 2  
  
 information channels, see expression  
 input sources 93–95  
 INTSINT 268, 273  
 IViE 271  
 instantiation 49, 81, 87, 167, 213–214, 218,  
   225  
 intelligibility 1, 5, 8–9, 11  
   cognitive intervention 10  
   naturalness 11–12, 123, 127, 135, 143  
 intonation  
   levels of representation 281, 287  
   modelling 268–269, 270–274, 278–281,  
     323  
   phonetic 270, 277–282  
   phonological 269, 277–282  
   types of model 265–274  
   wrapping 287, see wrapping  
 isochrony 11, 243–245, 256, 266  
  
 Klatt synthesis system 6, 42, 45, 75  
 knowledge 213  
   base 80, 25, 225, 276, 282, 313–314  
  
 labelling 54, see emotive content, expression,  
   text marking  
 lexicon 115, 159, 191, 208, 223  
 linguistically motivated units 119–121  
 linguistics 4, 8, 18, 50, 65, 75, 138, 172, 189,  
   213, 285, 221–231, 266, 278, 301  
   dynamic model 221–223  
   generalisation 17, 85–88, 101, 277, 315  
   marking 94, 301  
  
 static model 80, 94, 191–192, 222, 224–225  
 units 119  
   a-linguistic units, see a-linguistic  
   clause, see phrase  
   paragraph 67, 119, 149, 155, 170, 177,  
     262, 288  
   phoneme 10, 50, 86, 94, 113, 119–120,  
     123, 155, 158, 175, 179, 266  
   phrase 2, 17, 28, 75, 115, 119, 123, 170,  
     224, 237–239, 274, 308  
   segment 2, 48, 50, 93–94, 99, 119, 121,  
     127, 193, 262, 273, 287, 322, 325  
   segment intelligibility 1, 91  
   sentence 1, 18, 39, 65–67, 119, 168, 170,  
     194, 262  
   syllable 11, 28, 30, 33, 35, 55, 76, 94,  
     111, 121, 124–125, 127, 250, 261–262,  
     289  
   utterance 20, 45, 48, 67, 93, 98, 111, 168,  
     172, 194, 223, 228, 249, 287  
   word 2, 8, 67, 69, 120, 173, 186, 191,  
     208, 303, 309, 321  
  
 lockability 303  
 low level synthesis, see synthesis  
  
 marking resolution 48, see model  
 metatheory 65  
 micro-prosody 133–134  
 mind-brain distinction 64–65, see cognitive  
   physical distinction  
 model  
   accent, see accent  
   acoustic 45, 122, 316  
   biological 189, 197–198, 207  
   cognitive 197–198, 201, 207, 305  
   computational 187, 191, 208, 250  
   dynamic, see phonetic, phonology  
   expressive 102, 115, 201, 285, 292, 303,  
     309, see expression  
   high level 183–187, 293  
   intonation, see intonation  
   mapping 187, 209, 278  
   Markov (hidden Markov) 50, 80, 143  
   markup 99, 145–162  
   prosodic 167–176, 182, 292  
   perception 5, 10, 65  
   phonetic 8, 69, 185, 228, 277, 278  
   phonological 54, 69, 228, 270, 277  
   pragmatics 12, 76, 178, 290, 309–312  
     expression/emotion 178, 212

- prosody 20, 288
- variability 22
- prediction 176, 266
- proposed 55, 129, 148, 213–215, 286, 288–299, 305, 313, 322
  - biological stance 199, 214, 232
- prosodic 61, *see* prosody
- speech production 17, 23, 43, 86, 89, 105, 153, 176, 185, 189, 285, 313
- static, *see* phonetic, phonology
- synthesis 1, 39, 59, 63, 89, 100, 132, 187, 194, 268, 292–299
- theoretical limits 45, 69
- types 23–38, 197
  - category 217
  - dynamic 98, 191, 194, 218, 221–232, 295
  - process 218
  - static 18, 80, 93, 191, 194, 221–232
- morpheme 50, 119, 137, 155, 185, 193, 222, 266, *see* word
- mutability 315
- naturalness 112–115, 278–281, 303, 315
  - accents 117
  - expression 112, 188, 200
  - intelligibility 1, 7
  - prosody 161
  - synthesis 175
  - units 123
  - variability 139, 303
  - neutral 112
- parameterisation 188, 286, 308–309
- performance (idealised) 222
- perception 5, 47, 112, 123, 198, 207–210, 287
  - cognitive phonetics 85
  - database 105
  - isochrony 256
  - predictive model 177
  - production and perception 59, 60, 79, 191, 274,
  - trigger 3, 59, 87, 191, 204, 213, 221, 242, 286, 299
- phone 89, 94, 119, 181, 224
- phoneme 94–95, 115, *see* linguistics
- phonetics 185, 181–194, 277
  - boundaries 51, 67
  - classical 28, 47, 50, 79, 137, 266
  - cognitive 8, 50, 57–60, 193, 214, 225, 274, 325
  - computational 223, 228
  - dynamic 221–225, 230, 256–264, 301–303
  - intrinsic allophone, *see* allophone
  - linguistic 85, 89, 189
  - modern 115, 175, 191, 278
  - rendering, *see* rendering
  - static 221, 224, 228, 233, 302–303
  - supervision, *see* CPA
  - time 193, 218, 223, 226
  - variability, *see* variability
- phonology 6
  - articulatory, *see* articulatory phonology
  - cognitive 303
  - computational 222, 225
  - dynamic 221–222, 239, 287–289, 303
  - extrinsic allophone, *see* allophone
  - labelling, *see* labelling
  - language specific 225
  - linguistics 17–18, 85–86, 277
  - prosody, *see* prosody
  - segments 72, 302–303, *see* linguistics
  - static 96, 138, 194, 224, 225, 303
  - time 72, 274
  - transformational 51, 81, 169, 191–195, 266, 286
  - variation 69–75, 94
- pitch global 271–272
  - local 271–272
- plan 17, 20, 89, 193, 233, *see* rendering
- planes 287, 233, 239, 246
  - dynamic 194, 195, 221, 233, *see* phonetics, phonology
  - phonetic 195, 233–234, 239
  - phonology 195, 233–234, 239–245, *see* phonetics, phonology
  - static 194, 224, 233, 195, 230
- pragmatics 18, 20, 308–312
  - expression 98, *see* expression
- precision 132, 326
- production and perception 87–89
- processes, cognitive and physical 64
  - linguistic 227, 275–276, 278
- prominence 171
- prosody 167, 170, 266–268, 277–281
  - concatenative 28–35
  - control 170
  - general 111–112, 288, 278
  - markup 129, 167–169, 182
  - phonetic 62, 111

- phonological 62, 96, 111
- rendering 56, see rendering
- pseudo-code 215
- register 113
- rendering 191, 193–194, 174, 186, 193, 228,
  - see plan
- representation, level of 268
- re-synthesis 267
- rhythmic
  - group 245
  - structure 76
- rules
  - adjustment 94, 97
  - coarticulation 79
  - context sensitive 41, 130, 141
  - dynamic plane 221–115, 245
  - grammar 227
  - phonetic 77, 254
  - phonology 70–71, 156, 226
  - segment 35, 44
  - syntax 120
- SABLE 165
- saliency 172–173
- segment, see linguistic
- segmental structure of speech 47
- semantics 17, 21, 49, 113, 143, 148, 172, 192,
  - 226, 228, 313
- SMIL 154
- sound patterning 225
- speaking 189, 313
- spectrograms (Huckvale) 25
- SPRUCE 11, 124, 132, 274
- SSML 153–162, 271, 295
- string
- supervision 226, see CPA
- stress 111
- stress-timed 256
- style 113
- suprasegmentals 2, 185, see prosody
- syllable duration 75–76
- syntax 17, 19, 67, 69, 93, 120, 131, 142, 172,
  - 222, 226, 313
- synthesis
  - control 24, 85, 93
  - low level 17, 23, 42, 79, 257, 266, 269,
    - 271, 281, 322
  - labelling 47, 54, 100, 139, 167, 192, 207,
    - 278
  - potential of 47
  - text marking 47, 48, 99, 147, 175
  - high level 17–18, 81, 83–105, 86,
    - 133–135, 194, 266, 268, 270–271,
      - 274, 278
  - strengthening 185–194
  - types of
    - articulatory 23, 43
    - concatenative 28, 43
    - formant 24, 43
    - unit 32, 139
- targets 51
- test bed, synthesis as 90–91
- text marking 57–55, see labelling, model,
  - markup
  - automatic 99
  - data 100
- TGG (transformational generative grammar)
  - 168, 191–192
- theory driven 32, see model
- thresholded keys 304
- time 230, see duration, phonetics
  - clock 223–225, 280
  - notional 222, 225, 280
- Tilt model (Taylor) 269
- Tatham-Morton intonation 274, 281
- ToBI 134, 269, 271, 279
- trigger 49, see perception
- unit selection 137–144
- utterance
  - exemplar 81, 102, 139
  - plan 10, 21, 47–48, 54, 82, 87–88, 97, 135,
    - 226–228, 271, 277, 281
- variability 8, 36, 69, 101–103, 113, 137, 294
  - continuous 3225–326
  - segmental 280
- vocal tract 189, 290, 305
- voice quality 65, 199, 315
- VOT (voice onset time) 72–74, 261
- wrapping 151, 287, 301, see expression
  - expression 292–293, 301–302
  - intonation 96, 98, 139, 287
  - model 18, 291–298, 313
  - phonetics 303
  - phonology (prosodics) 258, 288, 291, 303
  - pragmatics 309

XML, see expression

attributes 161–163, 294–295

constraints 214, 303

data structure 115, 301–302, 293

elements 6, 121, 149, 160, 162, 294

expression wrapper 291–298

instantiation 49, 135, 280

scheme 168

VoiceXML markup 147–151

sample code 245–248, 249–263, 287–289,  
291–298