
Index

- Aamodt, Michael, 273
- Acoustics, acoustic knowledge, 8–9, 85
- Advanced Recognition Technologies (ART), 260
- Advanced Telecommunication Research (ATR), 207
- Allen, Jonathan, 197
- Altech, Advanced Language Technologies, 221–222. *See also* SpeechWorks
- Analog-to-digital conversion (A/D), xix, 64
- Android, 261, 276
- Apple, 207, 213, 224, 246
- ARPA (Advanced Research Project Agency). *See also* DARPA
- Speech Understanding Research project (SUR), 91–94, 106, 142–143, 145, 147, 281, 294n14, 294n15, 294n16, 295n17, 295n18
- Artificial intelligence (AI)
- approach to speech recognition, xxvii, 8183–88, 91–94, 106, 132, 146, 263, 265, 280, 283, 303n18
- Artificial user, 178, 186–187
- Artisoft, 235, 238
- Aspect, 235
- AT&T (American Telephone and Telegraph), xxiv, 48, 50, 90, 142, 144, 165, 174–175, 180, 204, 208–209, 211, 212, 227, 228–230, 232, 233, 235, 245, 256, 260–261, 281, 290n3, 290n7, 301n9, 304n9, 308n18, 309n21
- Bell Laboratories, Bell Telephone Laboratories, xxi, 7, 48–49, 51, 55–56, 61–62, 64, 77, 83, 88, 90, 133, 136, 141, 145, 147, 153, 164, 175, 184, 207–211, 232, 273, 281, 286n9, 290n3, 301n9, 304n9, 308n15, 308n18
- Shannon Laboratories, 178, 183, 187, 230, 278–279, 309n21
- Voice Tone, 233
- Watson, 228, 308n18
- AUDREY (AUtomatic Digit Recognition), 55–62, 65
- Baker, James, 107, 127, 128, 132, 212
- Baker, Janet, 107, 128, 212, 306n5
- Baker, W. O., xxiv
- Balashok, Stephen, 55, 61
- Bandwidth, 50–51, 53–54, 63, 290n11, 301n7
- Barge-in, algorithm, 210
- Baum, Leonard, 122, 127, 212, 270
- Baum-Welch, forward-backward algorithm, 127
- Bolt, Beranek and Newman (BBN), 92–93, 135, 142, 145, 153, 180, 207, 228, 301n9
- HARC, 228
- Bell, Alexander Graham, xxiv, 184, 194, 308n18
- Bell Labs. *See* AT&T, Bell Laboratories; Lucent, Bell Laboratories
- Bellman, Richard, 74, 292n26
- optimality principle, dynamic programming, 71, 73–75, 77–78, 95–96, 98, 103, 106, 115, 123, 132, 140, 258, 292n26, 296n27
- Berners-Lee, Tim, 246

- Biddulph, Rulon, 55
 Black, Alan, 204
 Blackberry, 214, 276
 Blackboard, scheduling, architecture, 94, 281
Blade Runner, 278
 BLEU (Bilingual Evaluation Understudy), 275
 Blocks world, 84, 97–98, 100
 Bloom, Jonathan, 307n12
 Broadcast news, speech corpus, 188–189
 Brown, Ruth, 268
 Brown University, 91
- C-3PO, xxviii, 266, 280
 Call flow, 235, 237–239, 255–256
 Carnegie Mellon University (CMU), 83–84, 91–92, 94, 105, 107, 135, 145–146, 153, 175, 180, 207, 212, 222, 294n16
 CCXML (Call Control eXtensible Markup Language), 254
 Centre national de la recherche scientifique (CNRS; French national center of scientific research), 147
 Centro Studi e Laboratori Telecomunicazioni (CSELT; Italian center for telecommunication studies and laboratories), xix, 135, 285n2
 Chiba, Seibi, 77–78
 Chomsky, Noam, 3–6, 286n6, 286n8, 286n9, 288n25, 288n27
 CHRONUS (conceptual HMMs), 164–165, 183
 Cisco, 245, 311n10
 Coarticulation, 36–37, 58, 131, 200, 296n25, 298n15, 301n9
 Cohen, Mike, 222
 Columbia University, 91, 273
 Companions project, 280
 Comverse, 245
 Concatenative speech synthesis, 198–200, 205
 Confidence, measure of, 241–242, 309n1
 Connected speech recognition, connected word recognition, 59, 61, 94–97, 105–106, 133, 136, 139, 140–143
 Cooper, Franklin, 194
 Corona, 221–222. *See also* Nuance
 Crandall, Irvin, 51
 Custer, Heather, 273
 Cyc project, 13–14, 266
- D’Arbeloff, Nick, 215
 DARPA (Defense Advanced Research Project Agency), 90–91, 142–149, 174–176, 180–181, 187–188, 190, 207, 209, 212, 214, 222–223, 269–270, 273, 288n28, 294n13, 303n20, 304n8
 ATIS (Air Travel Information Systems), 142, 148–158, 160–170, 174–175, 180, 183, 187–189, 208, 233, 288n28, 300n3, 303n21, 303n1, 309n23
 Communicator project, 174–177, 179–181, 183, 186–187, 207, 243, 304n8
 GALE project, 273–274
 IPTO (Information Processing Technology Office), 91, 295n18
 Resource management project, 142–148, 150, 152, 154, 161, 167, 180, 187–189, 302n10
- Davis, Angus, 257
 Davis, K. H., 55–56
 DAVO (Dynamic Analogue of the VOcal Tract), 197
 Deception, automatic detection of, 273, 313n10
 DEctalk, 197
 DeepQA (Deep Question Answering), 266
 de Mori, Renato, 135, 309n23
 Denes, Peter, 7, 61
 Dialog
 manager, 170–171, 173–174
 modules, 240, 243
 system, xix, 168, 170–173, 176–178, 205, 223, 230, 240–241, 245, 249, 253, 268, 281
 Dictaphone, 214
 Digital Equipment Corporation (DEC), 197
 PDP 11/60, xx
 Digital-to-analog conversion (D/A), 64
 Discourse analyzer, 170, 173–174
 Discriminative training, 271
 Distributed speech recognition (DSR), 276

- Doddington, George, 137
 Dragon Systems, 127–128, 212–213, 301n9, 306n5
 Dudley, Homer, 48, 53–55, 61, 191–192, 194
 Dynamic time warping (DTW), 71, 75, 77–78, 80–81, 90, 92, 94–96, 98, 103, 105, 133, 136, 146, 292n13, 292n32, 295n19
 Eloquent technologies, 260
 Emotion detection, 264, 271–273
 End point detection, detector, 78–80, 87, 210, 295n19
 Expert systems, 84–85, 136, 146
 Fallside, Frank, 147
 Fant, Gunnar, 195–196
 Feature extraction, 78, 80, 111, 181
 Feature vector, 69, 75–78, 80, 95, 103–105, 111, 113–116, 120–122, 124–128, 130–132, 172, 182, 276, 291n24, 298n19, 299n20, 300n2, 304n6
 Festival, speech synthesizer, 204
 Finite state
 machine, FSM, 98–105
 network, 109
 transducer, FST, 296n24
 Flanagan, James, 55, 58, 64, 77, 90, 133
 Fletcher, Harvey, 51
 Formants, 33–34, 36, 56–57, 59–60, 87, 194–197, 201, 306n6
 FP100, floating point array processor, xx
 Frame
 knowledge representation, 172–174, 304n5
 speech, 67, 69–71, 75–76, 95–98, 103, 105–105, 115, 120, 291n24, 296n27, 298n19, 299n21
 France Telecom Laboratories, 135, 207
 Fry, Dennis, 61
 Fundamental frequency. *See* Pitch
 GALAXY architecture, 169–170, 173–175, 222, 281
 Gauvain, Jean-Luc, 147
 Glass, Jim, 169, 221–222
 Goddeau, Dave, 221–222
 Goltsman, Victor, 268
 Google, 246, 259, 261, 276, 291n18
 Gorelov, Zor, 268
 Gorin, Al, 230, 232
 Grammar, 1, 5–6, 17–20, 22–23, 288n24, 288n26, 288n27
 in speech recognition, 95, 98, 115, 118, 148, 150, 161–162, 210–211, 216–217, 225, 239–240, 242, 251, 253, 255, 270, 282, 295n21, 295n22, 303n21, 310n9, 311n11
 HAL 9000, xx–xxii, xxi, 263, 266, 285n1, 307n12
 Harpy, 92–94, 106, 281
 Hartley, Ralph, 62–64
 Haskins Laboratories, 194
 Hearsay II, 92, 94, 281
 Hidden Markov models (HMM), 122–128, 131–133, 136, 146, 182–183, 190, 212, 258, 270–271, 277, 297n10, 299n20, 300n2, 301n9, 309n1
 Hirschberg, Julia, 273
 Holmes, John, 196
 How May I Help You, semantic classification, 230–233, 309n23
 HTML (Hypertext Markup Language), 246–250, 254, 310n8
 HTTP (HyperText Transfer Protocol), 246–250
 Hunnicutt, Sharon, 197
 HWIM (Hear What I Mean), 92–94
 IBM, xxi, 44, 83–84, 106, 109–110, 116, 119, 127, 128, 132–133, 135, 145, 175, 180, 203, 212–213, 227, 228–229, 245, 260–261, 266, 275, 285n5, 299n21, 301n9, 303n20, 308n17
 T. J. Watson Research Center, 107, 109, 145, 182, 207, 212
 Via Voice, 213, 228, 260
 Watson project, Jeopardy!, 266
 IEEE ICASSP (International Conference on Acoustic, Speech, and Signal Processing), 136, 264, 299n1
 Intel, 43, 245, 294n16

- Interactive voice response (IVR) system
 DTMF, touch tone, 217–218, 230, 235
 speech, 238–239, 242–243, 307n12
- Internet Engineering Task Force (IETF), 252
- Intervoice-Brite, 235, 238
- Intonation. *See* Prosody
- iPhone, 214, 261, 276, 291n18
- Island-driven parser, 94
- Isolated speech recognition, isolated word
 recognition, 58, 60–61, 78–80, 94–96,
 105, 136, 138, 212, 295n19
- Itakura, Fumitada, 77, 90
- Jelinek, Frederick, 106–107, 109–110,
 127–128, 132, 212, 296n1, 296n2
- Jeopardy! *See* IBM, Watson project
- Joint Speech Research Unit (JSRU), 135,
 196
- Julie, Amtrak's, 242–243
- Klatt, Dennis, 197
- KlatTalk, 197
- Klingon speech recognizer, 212
- Kratzenstein, Christian, 194
- KTH (Swedish Royal Institute of
 Technology), 135
- Language generator, natural language
 generator, 170–171, 173–174. *See also*
 Text-to-speech
- Language organ, 4–5, 286n9
- Latin, 15–16, 29
- Lawrence, Walter, 195–196, 306n6
- Lee, Chul Min, 272
- Lee, Kai-Fu, 146
- Lemon, Oliver, 187
- Lernout & Hauspie (L&H), 213, 260
- Levin, Esther, 164, 183, 187
- Licklider, J. C. R., 91
- Likert scale, 178–179
- LIMS (Computer Sciences Laboratory for
 Mechanics and Engineering Sciences),
 135
- Linguistic Data Consortium (LDC), 272,
 301n9
- Locus, 260
- Lovell, Tony, 215–216
- Lowerre, Bruce, 94, 105–106, 294n16,
 296n27
- Lucent, 245, 304n9
 Bell Laboratories, 180, 273
- Machine translation (MT), 266, 274–275,
 298n16
- MADCOW (Multisite ATIS Data Collection
 Working Group), 153–154, 161, 288n28
- Markov, Andrey, 120
- Markov chain, 120–122
- Markov decision process (MDP), 183–187
- Massachusetts Institute of Technology
 (MIT), 4, 35, 83–84, 91, 106, 133, 135,
 145, 153, 168–169, 171–175, 180, 197,
 207, 221–222, 265, 277, 301n9
- MATCH (Multimodal Access to City Help),
 278
- McCarty, John, 83
- McCue, Mike, 257–258
- Meaning. *See* Semantics
- Microsoft, Microsoft Research, xvii–xviii,
 207, 213, 224, 228, 245–246, 258,
 260–261
- Miner, Rich, 214–215
- MITalk, 197
- MIT Lincoln Laboratories, 92, 265, 301n9
- MITRE Corporation, 174–175, 180
- Mitsubishi, 207
- Morphology, morphological knowledge, 8,
 16–17, 42, 85–86
- Motorola, 245
- MRCP (Media Resource Control Protocol),
 253, 255–256
- Murveit, Hy, 222
- MUSA (Multichannel Speaking
 Automaton), xix, 135
- MYCIN, 84–85
- Narayanan, Shirkanth, 272–273
- National Institute of Standards and
 Technology (NIST), 143–145, 154, 160,
 162, 188–189
- Neural networks, artificial neural networks
 (ANN), 136, 145, 300n2

- Newell, Allen, 83, 91, 294n14
 Ney, Hermann, 147
 N-grams, 149, 297n7
 Noisy channel approach, 110–111, 114–115, 130
 Norman, Donald, 224
 NTT (Nippon Telephone and Telegraph), 90, 135, 185, 207
 Nuance, 222–223, 226–229, 233, 235, 239, 243–244, 256, 258, 260–261, 267, 277, 311
 NYNEX, 211
 Nyquist, Harry, 62–64, 291n19
- O’Farrell, Bill, 222
 Olive, Joe, 273
 OVE (Oral Verbs Electris), 195–196
- PARADISE (paradigm for dialog system evaluation), 178, 180
 Paramax, 153
 Partially observable Markov decision process (POMDP), 187
 PAT (Parametric Artificial Talkers), 195–196
 Philips, 135, 147, 207, 245, 260, 309n4
 Phillips, Mike, 169, 221–222, 228
 Phonetic segmentation, 61, 65, 81
 Phonetic Systems, 260
 Phonetics, phonetic knowledge, 8
 Phraselator, 266
 Pierce, John R., 88–92, 133
 Pietquin, Olivier, 187
 Pitch, 37–39, 55, 192–193, 199–201, 271–273, 289n30, 291n23
 Press or say 1 systems, xxiii–xxv, 218. *See also* IVR
 Principles of Interpretation (PofI), 158–161, 163, 167–167
 Prosody, 39
 PureSpeech, 260
 Put That There!, multimodal system, 277
- Rabiner, Lawrence, 133, 136, 141, 145, 147, 208, 286
 RAND Corporation, 71, 83
 Reddy, Raj, 94, 105–106, 145–146, 296n27
 Reinforcement learning, 181–187, 270
 Rhetorical, 260
 Rosen, George, 197
 Royal Signal and Radar Establishment, 135
 Rutgers University, 91, 260
- Sakoe, Hiroaki, 77–78
 SALT (Speech Application Language Tags), 245
 Sampling, Signal sampling, sampling theorem, 62
 Samuel, Arthur, 83, 109
 SayAnything, 233
 ScanSoft, 260, 277
 Schmandt, Chris, 277–278
 Scott Instruments, 222, 260
 Selfridge, Oliver, 265
 Semantics, semantic knowledge, semantic representation, 7–10, 12–15, 17–18, 20, 41–42, 85–86, 163–164, 230–233, 253, 282–283, 287n16, 288n25, 302n15, 309n22, 309n23
 representation of, 12, 15
 Seneff, Stephanie, 169, 221
 Shannon, Claude, 62–64, 83, 117–118
 Shaw, J. C., 83
 SHRDLU, 84
 Siemens, 135, 207
 Simon, Herbert, 83
 Speak & Spell, 137
 Speaker dependent, speaker dependence, 59, 61, 106, 136, 277
 Speaker independent, speaker independence, 59, 136–137, 292n32
 SpeakEZ, 260
 SpeakFreely, 233
 Spectrogram, 32–36, 51–52, 58–59, 67–71, 80, 168, 194, 222
 Spectrum, 31–33, 194
 Speech chain, 6–10, 17, 29, 40, 56, 61, 137, 191–192, 196, 253
 SpeechCycle, 268–270
 Speech generator, 170, 173–174. *See also* Speech synthesis
 Speechify, 204

- Speech sound
 affricate, 28
 consonant, 27–28, 37, 51, 80, 192, 195
 fricative, 28
 nasal, 23–25, 28, 195
 stop, plosive, 27–28, 80, 85–87, 192, 294n8, 296n25
 vowel, 25–28, 33–34, 36–37, 51, 57, 60, 86, 192, 194–195, 199, 287n11, 289n30, 301n7
- Speech synthesis, 54, 191–205, 253, 257, 273, 299n1
- SpeechWorks, 175, 204, 221–223, 226–229, 235, 239, 242–245, 256, 258, 260, 267, 272, 277, 304n11, 311n10
- Springer, Steve, 277
- SRGS (Speech Recognition Grammar Specification), 253
- SRS Labs, 222
- SSML (Speech Synthesis Markup Language), 253
- Stanford Research Institute (SRI), 91–92, 135, 145, 153–154, 174–175, 180, 207, 222, 301n9
- Stanford University, 71, 84, 91
- Statistical language models, 116, 120, 132.
See also N-grams
- Suprasegmental features, suprasegmental parameters, 39, 199
- Switchboard, speech corpus, 188–189
- Syntax, 5, 8–9, 15, 17–18, 20, 22, 40–41, 203, 296n24
- Systems Development Corporation (SDC), 93, 94, 295n17
- Taylor, Paul, 204
- TD-Gammon, 182–183
- Technical University of Munich, 135
- Telecom Italia, 207
- Telefonica, 207
- Telleureka. *See* SpeechCycle
- Tellme Networks, 257–258, 260
- Templates, template-based speech recognition, template matching, 65–71, 75–81, 90, 92–98, 102–106, 132–133, 135–136, 140, 182, 292n31, 292n32, 295n19, 296n25
- Tesauro, Gerald, 182–183
- Texas Instruments, TI, 135, 137–138, 141–142, 187–188, 207, 301n9
- Text-to-speech (TTS), 171–172, 191–193, 196, 200–205, 210–211, 248, 252–256, 260, 274, 298n14, 310n9. *See also* Speech synthesis
- Thomson, David, 281
- TI-DIGITS, 137–138, 141–142, 152, 301n6
- Time warping. *See* Dynamic time warping
- Turk, the, 47
- 2001: A Space Odyssey*, 20, 21, 265, 285n5, 307n12
- Universal grammar, 5–6
- University of Cambridge, UK, 147, 187
- University of Colorado–Boulder, 175, 180
- University of Edinburgh, 187, 204
- University of Erlangen, 135
- University of Nagoya, 135
- University of Nancy, 135
- University of Pennsylvania, 91
- University of Southern California, USC, 272
- University of Tokyo, 135
- University of Turin, 135
- Unysis. *See* Paramax
- Verbex Voice Systems, 212
- Vintsyuk, T. K., 77
- Vlingo, 261
- Vocal organs, 7, 9, 42
- Vocal tract, 23–27, 33, 35–39, 42–43, 53–56, 67, 122, 125–126, 194, 197
- Vocoder, 54–55, 58
- Voder, 48–49, 54–55, 192–194, 201, 290n4
- Voice browser, 247–248, 249, 251, 252–257, 310n9
- Voice Control Systems (VCS), 260
- Voice search, 261, 275–277, 281
- VoiceXML, 245, 247–249, 251–258, 260, 281, 310n5, 310n8, 311n16
- von Helmholtz, Hermann, 194, 305n3
- Von Kempelen, Wolfgang, 25, 42, 47–48, 55, 191, 193–194, 197
- VoxTech, 266
- VOYAGER, 169, 222

- VRCP (Voice Recognition Call Processing),
208, 211, 256
- VUI (Voice User Interface), 198, 215,
217–218, 220, 223, 240–241, 243,
304n11, 307n12, 308n13

- Walker, Lyn, 178
- Warner, Bill, 215
- Waveform, 29–33, 35–36, 59, 64, 66–67,
78, 172, 198, 289n33
- Wayne, Charles, 147
- Wheatstone, Charles, 194
- Wiesner, Jerome, 265
- Wildfire, 214–217
- Wilks, Yorick, 279–280
- Williams, Jason, 187
- Willis, Robert, 194
- Wilpon, Jay, 208
- Winograd, Terry, 83
- Wizard of Oz (WoZ), 154–155, 231–232
- Word accuracy, 138–139, 141, 144–145,
152–153, 161, 178
- Word-spotting algorithm, 211
- World knowledge, 7–8, 14, 282
- World Wide Web Consortium (W3C), 245,
310n5
- WSJ (*Wall Street Journal*) project, 142,
187–189, 301n9

- Yahoo, 276
- Yale University, 194
- Young, Steve, 147, 187

- Zampolli, Antonio, prize, 109, 296n3
- Zue, Victor, 35, 168–169, 221–222