

MIT OpenCourseWare
<http://ocw.mit.edu>

MAS.632 Speech Interfaces and Mobile Devices
Fall 2008

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.

Bibliography

- Ades, S. and D. C. Swinehart. "Voice Annotation and Editing in a Workstation Environment." In *Proceedings of the 1986 Conference*, pages 13–28, San Jose, CA: The American Voice I/O Society, September 1986.
- Allen, J. *Natural Language Understanding*. Benjamin/Cummings Series in Computer Science. Reading, MA: The Benjamin/Cummings Publishing Co., Inc., 1987.
- Allen, J., M. S. Hunnicutt, and D. Klatt. *From Text to Speech: The MITalk system*. Cambridge University Press, 1987.
- Allen, J. "Overview of Text-to-Speech Systems." In S. Furui and M. M. Sondhi, editors, *Advances in Speech Signal Processing*, pages 741–790. New York: Marcel Dekker, Inc., 1992.
- Allen, R. B. "Composition and Editing of Spoken Letters." *International Journal of Man/Machine Studies*, 19:181–193, 1983.
- Allport, D. A., B. Antonis, and P. Reynolds. "On the Division of Attention: a Disproof of the Single Channel Hypothesis." *Quarterly Journal of Experimental Psychology*, 24:225–235, 1972.
- Angebrannt, S. R. L. Hyde, D. H. Luong, N. Siravara, and C. Schmandt. "Integrating Audio and Telephony in a Distributed Workstation Environment." In *Proceedings of the Summer 1991 USENIX Conference*, pages 419–435, Berkeley, CA: USENIX Association, June 1991.
- Arons, B., C. Binding, C. Schmandt, and K. Lantz. "The VOX Audio Server." In *Proceedings of the 2nd IEEE Comsoc International Multimedia Communications Workshop*, New York: IEEE, April 1989.
- Arons, B. "The Design of Audio Servers and Toolkits for Supporting Speech in the User Interface." *Journal of The American Voice I/O Society*, 9:27–41, March 1991.
- Arons, B. Hyperspeech. "Navigating in Speech-Only Hypermedia." In *Proceedings of Hypertext '91*, pages 133–146. New York: ACM, December 1991.

- Arons, B. "Techniques, Perception, and Applications of Time-Compressed Speech." In *Proceedings of the 1992 Conference*, pages 169–177, San Jose, CA: The American Voice I/O Society, September 1992.
- Arons, B. "Tools for Building Asynchronous Servers to Support Speech and Audio Applications." In *Proceedings of the ACM Symposium on User Interface Software and Technology*. New York: ACM, November 1992.
- Arons, B. "Speech Skimmer: Interactively Skimming Recorded Speech." In *Proceeding of the ACM Symposium on User Interface Software and Technology*. New York: ACM, November 1993.
- Austin, J. L. *How to do Things with Words*. Oxford, England: Clarendon Press, 1962.
- Beasley, D. S. and J. E. Maki. "Time- and Frequency-Altered Speech." In N.J. Lass, editor, *Contemporary Issues in Experimental Phonetics*, pages 419–458. Academic Press, 1976.
- Biermann, A. W., R. D. Rodman, D. C. Rubin, and F. F. Heidlage. "Natural Language With Discrete Speech as a Mode for Human-to-Machine Communication." *Communications of the ACM*, 28(6):628–636, 1985.
- Billman, R., R. Hammond, P. McElbatton, E. N. Brandt, F. Sung, and N. Yost. "Workstation Audio in the X Environment." *The X Resource*, 1(4):137–158, October 1992.
- Bly, S. A. and J. K. Rosenberg. "A Comparison of Tiled and Overlapping Windows." In *Proceedings of the Conference on Computer Human Interface*, pages 101–106, New York: ACM, April 1986.
- Bolinger, D. "Intonation and Its Parts." *Language*, 58(3):505–533, 1982.
- Bolt, R. A. "'Put-That-There': Voice and Gesture at the Graphics Interface." *Computer Graphics*, 14(3):262–270, 1980.
- Bowles, I., R. Kamel, et al. "PX: Integrating Voice Communications With Desktop Computing." *Journal of The American Voice I/O Society*, 9:1–19, March 1991.
- Briley, B. E. *Introduction to Telephone Switching*. Reading, MA: Addison-Wesley Publishing Co., Inc., 1983.
- Cahn, J. E. "The generation of affect in Synthesized Speech." In *Journal of the American Voice I/O Society*, pages 1–19. San Jose, CA: American Voice I/O Society, July 1990.
- Chalfonte, B. L., R. S. Fish, and R. E. Kraut. "Expressive Richness: A Comparison of Speech and Text as Media for Revision." In *Proceedings of the Conference on Computer Human Interaction*, pages 21–26. New York: ACM, April 1991.
- Chen, F. R. and M. Withgott. "The Use of Emphasis to Automatically Summarize Spoken Discourse." In *Proceedings of the 1992 International Conference on Acoustics, Speech and Signal Processing*, pages 229–233. New York: IEEE, 1992.
- Chorafas, D. N. *Telephony: Today and Tomorrow*. Prentice-Hall Series in Data Processing Management. Englewood Cliffs, NJ: Prentice-Hall, Inc., 1984.
- Clark, H. H. and C. R. Marshall. "Definite Reference and Mutual Knowledge." In Webber Joshi and Sag, editors, *Elements of Discourse Understanding*, 10:63. Cambridge University Press, 1981.
- Clark, H. H. and D. Wilkes-Gibbs. "Referring as a Collaborative Process." *Cognition*, 22:1–39, 1986.
- Clark, H. H. and E. F. Schaefer. "Contributing to Discourse." *Cognitive Science*, 13:259–294, 1989.
- Clark, H. H. and S. E. Brennan. "Grounding in Communication." In L. B. Resnick, J. Levine, and S. D. Behrend, editors, *Socially Shared Cognition*. American Psychological Association, 1990.
- Cohen, P. R., J. Morgan, and M. E. Pollack, editors. *Intentions in Communication*. Cambridge, MA: MIT Press, 1990.
- Crowley, T. "Voice in a Multimedia Document System." *Journal of The American Voice I/O Society*, 9:21–26, March 1991.

- Daly, N. A. and V. W. Zue. "Acoustic, Perceptual, and Linguistic Analyses of Intonation Contours in Human/Machine Dialogues." In *Proceedings of the International Conference on Spoken Language Processing*, November 1990.
- Davis, J. R. *Back Seat Driver: Voice assisted automobile navigation*. Ph.D. thesis, MIT Media Arts and Sciences Section, September 1989.
- Davis, J. R. and T. F. Trobaugh. Direction Assistance. Technical Report 1, MIT Media Laboratory Speech Group, December 1987.
- Davis, J. R. and J. Hirschberg. "Assigning Intonational Features in Synthesized Spoken Directions." In *Proceedings of the Association for Computational Linguistics*, pages 187-193, 1988.
- Davis, J. R. and C. Schmandt. "The Back Seat Driver: Real Time Spoken Driving Instructions." In *Vehicle Navigation and Information Systems*, pages 146-150, 1989.
- Davis, J. R. and C. Schmandt. "Discourse Strategies for Conversations in Time." In *Proceedings of the AVIOS 1990 Conference*, pages 21-26, San Jose, CA: The American Voice I/O Society, September 1990.
- Davis, J. R. "Let Your Fingers Do the Spelling: Implicit Disambiguation of Words Spelled With the Telephone Keypad." *Journal of The American Voice I/O Society*, 9:57-66, March 1991.
- Denes, P. B. and E. N. Pinson. *The Speech Chain: The physics and biology of spoken language*, 2nd edition. New York: W.H. Freeman, 1993.
- Dixon, N. R. and T. B. Martin, editors. *Automatic Speech & Speaker Recognition*. IEEE Press Selected Reprint Series. New York: IEEE Press, 1979.
- Duncan, S., Jr. "Some Signals and Rules for Taking Speaking Turns in Conversations." *Journal of Personality and Social Psychology*, 23:283-292, 1972.
- Duncan, S., Jr. "On the Structure of Speaker-Auditor Interaction During Speaking Turns." *Language in Society*, 2:161-180, 1974.
- Egan, J. P. "Articulation Testing Methods." *Laryngoscope*, 58:955-991, 1948.
- Englebeck, G. and T. L. Roberts. "The Effect of Several Voice-Menu Characteristics on Menu-Selection Performance." Technical report, U.S. West Advanced Technologies, Boulder, CO, 1989.
- Fairbanks, G. "Test of Phonemic Differentiation: The Rhyme Test." *Journal of the Acoustical Society of America*, 30:596-600, 1958.
- Fish, R. S., R. E. Kraut, M. D. P. Leland and M. Cohen. "Quilt: a Collaborative Tool for Cooperative Writing." In *Proceedings of the Conference on Office Information Systems*, pages 30-37, New York: ACM 1988.
- Flanagan, J. L. "Analog Measurements of Sound Radiation From the Mouth." *Journal of the Acoustical Society of America*, 32(12), 1960.
- Flanagan, J. L. *Speech Analysis Synthesis and Perception*. Springer Verlag, 1972.
- Flanagan, J. L., M. R. Schroeder, B. S. Atal, R. E. Crochiere; N. S. Jayant, and J. M. Tribolet. "Speech Coding." *IEEE Transactions on Communications*, 27(4):710-737, April 1979.
- Fletcher, H. and W. A. Munson. "Loudness, Definition, Measurements and Calculation." *Journal of the Acoustical Society of America*, 5, 1933.
- Fry, D. B. "Experiments in the Perception of Stress." *Language and Speech*, 1(2):126-152, 1958.
- Fujimura, O. and J. Lovins. "Syllables as Concatenative Phonetic Elements." In A. Bell and J. B. Hooper, editors. *Syllables and Segments*, pages 107-120, North Holland, New York, 1978.
- Furui, S. and M. M. Sondhi, editors. *Advances in Speech Signal Processing*. New York: Marcel Dekker, Inc., 1992.
- Gifford, W. S. and D. L. Turock. "The Electronic Receptionist: A Knowledge-Based Approach to Personal Communications." In *Proceedings of the IEEE Conference on Selected Topics in Wireless Communications*. New York: IEEE, June 1992.
- Goffman, E. *Forms of Talk*. University of Pennsylvania Press, 1981.

- Gould, J. D. "How Experts Dictate." *Journal of Experimental Psychology: Human Perception and Performance*, 4(4):648-661, 1978.
- Gould, J. D. and S. J. Boies. "How Authors Think About Their Writing, Dictating, and Speaking." *Human Factors*, 20(4):495-505, 1978.
- Gould, J. D. "Writing and Speaking Letters and Messages." *International Journal of Man/Machine Studies*, 16(2):147-171, 1982.
- Gould, J. D.; J. Conti, and T. Hovanyecz. "Composing Letters With a Simulated Listening Type-writer." *Communications of the ACM*, 26(4):295-308, April 1983.
- Gould, J. D., S. J. Boies, S. Levy, J. T. Richards, and J. Schoonard. "The 1984 Olympic Message System: A Test of Behavioral Principles of System Design." *Communications of the ACM*, 30(9):758-769, September 1987.
- Grice, H. P. "Logic and Conversation." In Cole and Morgan, editors, *Syntax and Semantics: Speech Acts*, volume 3, pages 41-58. Academic Press, 1975.
- Grosz, B. J., A. K. Joshi, and S. Weinstein. "Providing a Unified Account of Definite Noun Phrases in Discourse." In *Proceedings of the 21st conference of the Association for Computational Linguistics*, pages 44-50, 1983.
- Grosz, B. J. and C. L. Sidner. "Attention, Intentions, and the Structure of Discourse." *Computational Linguistics*, 12(3):175-204, 1986.
- Grosz, B. J.; K. S. Jones, and B. L. Webber, editors. *Readings in Natural Language Processing*. Los Altos, CA: Morgan Kaufmann Publishers, Inc., 1986.
- Handel, W. S. *Listening: An Introduction to the Perception of Auditory Events*. Cambridge, MA: MIT Press, 1989.
- Hauptmann, A. G. "Speech and Gestures for Graphic Image Manipulation." In *Proceedings of the Conference on Computer Human Interface*, pages 241-245, New York: ACM, May 1989.
- Hayes, P. J. and R. Reddy. "Steps Towards Graceful Interaction in Spoken and Written Man-Machine communication." *International Journal of Man/Machine Systems*, 19:231-284, 1983.
- Hindus, D. "Semi-Structured Capture and Display of Telephone Conversations." Master's thesis, MIT, February 1992.
- Hindus D., C. Schmandt, and C. Horner. "Capturing, Structuring, and Presenting Ubiquitous Audio." *Transactions on Information Systems*, New York: ACM October 1993.
- Hindus, D. and C. Schmandt. "Ubiquitous Audio: Capturing Spontaneous Collaboration." In *Proceedings of the Conference on Computer-Supported Cooperative Work*, pages 210-217 New York: ACM, 1992.
- Hirschman, L., S. Seneff, D. Goodine, and M. Phillips. "Integrating Syntax and Semantics Into Spoken Language Understanding." In *Proceedings of the Fourth DARPA Speech and Natural Language Workshop*, February 1991.
- Hodges, M. E., R. M. Sasnett, and M. S. Ackerman. "A Construction Set for Multimedia Applications." *IEEE Software*, pages 37-43, January 1989.
- Hunnicut, S., L. Hirschman, J. Polifroni, and S. Seneff. "Analysis of the Effectiveness of System Error Messages in a Human-Machine Travel Planning Task." In *Proceedings of the Second International Conference on Spoken Language Processing*, October 1992.
- Jacob, R. J. K. "The Use of Eye Movements in Human-Computer Interaction Techniques: What You Look at is What You Get." *ACM Transactions on Information Systems*, 9(3):152-169, April 1991.
- Jelinek, F. "The Development of an Experimental Discrete Dictation Recognizer." In *Proceedings of the IEEE*, volume 73, pages 1616-1624. New York: IEEE, November 1985.
- Kamel, R., K. Emami, and R. Eckert. "PX Supporting Voice in Workstations." *IEEE Computer*, 23(8):73-80, August 1990.
- Klatt, D. H. "Software for a Cascade/Parallel Formant Synthesizer." *Journal of the Acoustical Society of America*, 67(3):971-990, March 1980.

- Klatt, D. H. "Review of Text-to-Speech Conversion for English." *Journal of the Acoustical Society of America*, 82:737-783, September 1987.
- Klatt, D. H. "Review of Text-to-Speech Conversion for English." *Journal of the Acoustical Society of America*, 82:737-793, September 1987.
- Koons, D.B, C.J. Sparrel, and K.R. Thorisson. "Integrating Simultaneous Input from Speech, Gaze and Hand Gesture." In T.M. Maybury, editor, *Intelligent Multi-Media Interfaces*, Menlo Park: AAAI Press, 1993.
- Kramer, M., G. Ramirez, D. L. Turock, and R. S. Wolff. "Use of Two-Way Wireless Messaging for Personal Telephone Management." In *Proceedings of ICUPC '93*, October 1993.
- Kraut, R. E., S. H. Lewis, and L. W. Swezey. "Listener Responsiveness and the Coordination of Conversation." *Journal of Personality and Social Psychology*, 42(4):718-731, 1982
- Kraut, R. E. and S. H. Lewis. "Some functions of feedback in conversation." In H. E. Syphers and J. L. Applegate, editors, *Communication by Children and Adults*, Chapter 9. Sage, 1984.
- Ladd, D. R. *The Structure of Intonational Meaning*. Indiana University Press, 1978.
- Ladefoged, P. *Elements of Acoustic Phonetics*. Chicago:University of Chicago Press, 1962.
- Lamming, M. and W. Newman. "Activity-Based Information Retrieval: Technology in Support of Human Memory." Technical Report 91-03, Rank Xerox EuroPARC, February 1991.
- Lea, W. A. "Prosodic Aids to Speech Recognition." In Wayne A. Lea, editor, *Trends in Speech Recognition*, Chapter 8. Prentice Hall, 1980.
- Lee, K. *Large-Vocabulary Speaker-Independent Continuous Speech Recognition Using Hidden Markov Models*. Ph.D. thesis, Carnegie-Mellon University, 1988.
- Lee, K. and H. Hon. "Large-Vocabulary Speaker-Independent Continuous Speech Recognition Using HMM." In *International Conference on Acoustics, Speech, and Signal Processing*. New York: IEEE, April 1988.
- Lee, K. and F. Alleva. "Continuous Speech Recognition." In Sadaoki Furui and M. Mohan Sondhi, editors, *Advances in Speech Signal Processing*, pages 623-650. New York: Marcel Dekker, Inc., 1992.
- Leland, M. D. P., R. S. Fish, and R. E. Kraut. "Collaborative Document Production Using Quilt." In *Proceedings of the Conference on Computer-Supported Cooperative Work*, pages 206-215, New York: ACM, 1988.
- Levinson, S. C. *Pragmatics*. Cambridge University Press, Cambridge, Great Britain, 1983.
- Lison, H. and T. Crowley. "Sight and Sound." *UNIX Review*, 7(10):76-86, October 1989.
- Luce, P. A., T. C. Feustel, and D. B. Pisoni. "Capacity Demands in Short-Term Memory for Synthetic and Natural Speech." *Human Factors*, 25(1):17-32, 1983.
- Mackay, W. E., T. W. Malone, K. Crowston, R. Rao, D. Rosenblitt, and S. K. Card. "How do Experienced Information Lens Users Use Rules?" In *Proceedings of the Conference on Computer Human Interface*, pages 211-216, New York: ACM 1989.
- Malone, T. W., K. R. Grant, K. Lai, R. Rao, and D. Rosenblitt. "Semi-Structured Messages are Surprisingly Useful For Computer-Supported Coordination." *ACM Transactions on Office Information Systems*, 5(2):115-131, 1987.
- Martin, G. L. "The Utility of Speech Input in User-Computer Interfaces." *International Journal of Man/Machine Studies*, 30:355-375, 1989.
- Maxemchuk, N. F. "An Experimental Speech Storage and Editing Facility." *The Bell System Technical Journal*, 59(8):1383-1395, October 1980.
- McPeters, D. L. and A. L. Tharp. "The Influence of Rule-Generated Stress on Computer-Synthesized Speech." *International Journal of Man/Machine Studies*, 20:215-226, 1984.
- Medina-Mora, R., T. Winograd, R. Flores, and F. Flores. "The action Workflow Approach to Workflow Management Technology." In *Proceedings of the Conference on Computer-Supported Cooperative Work*, New York: ACM, November 1992.

- Mermelstein, P. "Automatic Segmentation of Speech into Syllabic Units." *Journal of the Acoustical Society of America*, 58(4):880-883, October 1975.
- Mermelstein, P. "G.722, a new CCITT coding standard for digital transmission of wideband audio signals." *IEEE Communications Magazine*, 26(1), 1988.
- Minneman, S. L. and S. A. Bly. "Managing a Trois: a Study of a Multi-User Drawing Tool in Distributed Design Work." In *Proceedings of the Conference on Computer Human Interaction*, pages 217-224. New York: ACM, April 1991.
- Minsky, M. "A Framework for Representing Knowledge." In P. Winston, editor, *The Psychology of Computer Vision*, pages 211-277. McGraw-Hill, New York, 1975.
- Morris, J. H., et al. "Andrew: A Distributed Personal Computing Environment." *Communications of the ACM*, March 1986.
- Muller, M. J. and J. E. Daniel. "Toward a Definition of Voice Documents." In *Proceedings of the 1990 Conference on Office Information Systems*, 1990.
- Murray, I. R., J. L. Arnott, and A. F. Newell. "Hamlet—Simulating Emotion in Synthetic Speech." In *Speech '88; Proceedings of the 7th FASE Symposium*. Institute of Acoustics, Edinburgh, 1988.
- Myers, B. A. "The Importance of Percent-Done Progress Indicators for Computer-Human Interfaces." In *Proceedings of the Conference on Computer Human Interface*, pages 11-17, New York: ACM, 1985.
- Nakatani, L. H., et al. "TNT: a Talking Tutor 'n' Trainer for Teaching the Use of Interactive Computer Systems." In *Proceedings of the Conference on Computer Human Interaction*, New York: ACM, 1986.
- Nicholson, R. T. "Usage Patterns in an Integrated Voice and Data Communications System." In *ACM Transactions on Office Information Systems*, volume 3, pages 307-314. New York: ACM, July 1985.
- Nusbaum, H. C., M. J. Dedina, and D. B. Pisoni. "Perceptual Confusions of Consonants in Natural and Synthetic CV Syllables." *Speech Research Lab Tech. Note 84-02*, Indiana University, 1984.
- Nye, P. W. and J. Gaitenby. "The Intelligibility of Synthetic Monosyllabic Words in Short, Syntactically normal sentences." *Haskins Lab. Status Report on Speech Research*, 38:169-190, 1974.
- Nyquist, H. "Certain Factors Affecting Telegraph Speed." *Bell System Technical Journal*, 3:324-326, 1924.
- Ochsman, R. B. and A. Chapanis. "The Effects of 10 Communication Modes on the Behavior of Teams During Co-Operative Problem-Solving." *International Journal of Man/Machine Studies*, 6:579-619, 1974.
- Olive, J. P. "Rule Synthesis of Speech from Diadic Units." In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, pages 568-570. New York: IEEE, 1977.
- O'Neill, E. F. "TASI—Time Assignment Speech Interpolation." *Bell Laboratories Record*, 37:83-87, March 1959.
- Orr, D. B., H. L. Friedman, and J. C. Williams. "Trainability of Listening Comprehension of Speeded Discourse." *Journal of Educational Psychology*, 56:148-156, 1965.
- O'Shaughnessy, D. *Speech Communication*. New York: Addison-Wesley Publishing Company, 1987.
- O'Shaughnessy, D. "Recognition of Hesitations in Spontaneous Speech." In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, pages 1521-1524. New York: IEEE, March 1992.
- Oshika, B. T., V. W. Zue, R. V. Weeks, H. Neu, and J. Aurbach. "The Role of Phonological Rules in Speech Understanding Research." *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-23(1):104-112, 1975.

- G. Peterson, W. Wang, and E. Sivertsen. "Segmentation Techniques in Speech Synthesis." *Journal of the Acoustical Society of America*, 30:739-742, 1958.
- Pierrehumbert, J. and J. Hirschberg. "The Meaning of Intonational Contours in the Interpretation of Discourse." In P. R. Cohen, J. Morgan, and M. E. Pollack, editors, *Intentions in Communication*, chapter 14, pages 271-311. Cambridge, MA: The MIT Press, 1990.
- Pisoni, D. B. and S. Hunnicutt. "Perceptual Evaluation of MITalk: The MIT Unrestricted Text-to-Speech System." In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, pages 572-575. New York: IEEE, 1980.
- Pisoni, D. B. "Speeded Classification of Natural and Synthetic Speech in a Lexical Decision Task." *Journal of the Acoustical Society of America*, 70:S98, 1981.
- Pisoni, D. B., H. C. Nusbaum, and B. G. Greene. "Perception of Synthetic Speech Generated by Rule." *Proceedings of the IEEE*, 73(11):1665-1676, November 1985.
- Rabiner, L. R. and R. W. Schafer. *Digital Processing of Speech Signals*. Englewood Cliffs, NJ: Prentice-Hall, Inc., 1978.
- Rabiner, L. R. and B. H. Juang. "An Introduction to Hidden Markov Models." *IEEE ASSP Magazine*, 3(1):4-16, January 1986.
- Rabiner, L. R. and B. H. Juang. "An Introduction to Hidden Markov Models." *IEEE Transactions on Acoustics, Speech, and Signal Processing*, pages 4-16, January 1986.
- Resnick, P. Hypervoice: "A Phone-Based CSCW Platform." In *Proceedings of the Conference on Computer-Supported Cooperative Work*, pages 218-225, New York: ACM, November 1992.
- Resnick, P. *Hypervoice: Groupware by Telephone*. Ph.D. thesis, MIT, September 1992.
- Resnick, P. and R. A. Virzi. "Skip and Scan: Cleaning Up Telephone Interfaces." In *Proceedings of the Conference on Computer Human Interface*, New York: ACM, 1992.
- Roberts, T. L. and G. Engelbeck. "The Effects of Device Technology on the Usability of Advanced Telephone Functions." In *Proceedings of the Conference on Computer Human Interface*, pages 331-337, New York, May 1989. ACM.
- Root, R. W. and C. R. Koster. "Experimental Evaluation of a Mnemonic Command Syntax for Controlling Advanced Telecommunications Services." In *Proceedings of the Conference on Computer Human Interaction*, pages 809-813, 1986.
- Root, R. W. and C. Chow. "Multimode Interaction in a Telecommunications Testbed: the Case of Memory Dialing." In G. Salvendy, editor, *Cognitive engineering in the design of human-computer interaction and expert systems*, pages 399-406. Elsevier Science Publishers, 1987.
- Rose, M. T. *The Internet Message*, Chapter 6. Prentice Hall Series in Innovative Technology. Englewood Cliffs, NJ: P T R Prentice Hall, 1993.
- Rosson, M. B. and A. J. Cecala. "Designing a Quality Voice: An Analysis of Listeners' Reactions to Synthetic Voices." In *Proceedings of the Conference on Computer Human Interface*, pages 192-197, New York: ACM, April 1986.
- Roucos, S. and A. M. Wilgus. "High Quality Time-Scale Modifications for Speech." In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, pages 493-496. New York: IEEE, 1985.
- Rudnicky, A. I. and A. G. Hauptmann. "Models for Evaluating Interaction Protocols in Speech Recognition." In *Proceedings of the Conference on Computer Human Interface*, pages 285-291, New York: ACM, 1991.
- Rudnicky, A. I.; J. M. Lunati, and A. M. Franz. "Spoken Language Recognition in an Office Management Domain." In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, pages 829-832. New York: IEEE, May 1991.
- Sacks, H., E. A. Schegloff, and G. Jefferson. "A Simplest Systematics for the Organization of Turn-Taking for Conversation. Language," 50(4):696-735, 1974. Reprinted in *Studies in the Organization of Conversational Interaction*, J. Schenken, ed., Academic Press 1978.

- Sakoe, H. and S. Chiba. "Dynamic Programming Algorithm Optimization for Spoken Word Recognition." *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 26:43-49, February 1978. Reprinted in Dixon and Martin.
- Schank, R. C. and R. Abelson. *Scripts, Plans, Goals and Understanding*. Hillsdale, NJ: Lawrence Erlbaum, 1977.
- Schmandt, C. "The Intelligent Ear: A Graphical Interface to Digital Audio." In *Proceedings IEEE Conference on Cybernetics and Society*, pages 393-397, New York: IEEE, October 1981.
- Schmandt, C. and E. Hulstén. "The Intelligent Voice Interactive Interface." *Human Factors in Computer Systems*, pages 363-366, 1982.
- Schmandt, C. "Speech Synthesis Gives Voiced Access to an Electronic Mail System." *Speech Technology*, 2(3):66-69, 1984.
- Schmandt, C. and B. Arons. "A Conversational Telephone Messaging System." *IEEE Trans. on Consumer Electr.*, CE-30(3):xxi-xxiv, 1984.
- Schmandt, C. and B. Arons. "Phone Slave: A Graphical Telecommunications Interface." *Proc. of the Soc. for Information Display*, 26(1):79-82, 1985.
- Schmandt, C., B. Arons, and C. Simmons. "Voice Interaction in an Integrated Office and Telecommunications Environment." In *Proceedings of the 1985 Conference*, pages 51-57. San Jose, CA: *American Voice I/O Society*, September 1985.
- Schmandt, C. and B. Arons. "A Robust Parser and Dialog Generator for a Conversational Office System." In *Proceedings of the 1986 Conference*, pages 355-365. San Jose, CA: *American Voice I/O Society*, September 1986.
- Schmandt, C. "Employing Voice Back Channels to Facilitate Audio Document Retrieval." In *Proceedings of the ACM Conference on Office Information Systems*, pages 213-218., New York: ACM, 1988.
- Schmandt, C. and M. McKenna. "An Audio and Telephone Server for Multi-Media Workstations." In *Proceedings of the 2nd IEEE Conference on Computer Workstations*, pages 150-159. New York: IEEE, March 1988.
- Schmandt, C. and S. Casner. "Phonetool: Integrating Telephones and Workstations." In *Proceedings of the IEEE Global Telecommunications Conference*, pages 970-974. New York: IEEE Communications Society, November 1989.
- Schmandt, C., D. Hindus, M. Ackerman, and S. Manandhar. "Observations on Using Speech Input for Window Navigation." In *Proceedings of the IFIP TC 13 Third International Conference on Human-Computer Interaction*, pages 787-793, August 1990.
- Schmandt, C. "Phoneshell: the Telephone as Computer Terminal." In *Proceedings ACM Multimedia 93*, pages 373-382. New York: ACM, 1993.
- Schmandt, C., M. S. Ackerman, and D. Hindus. "Augmenting a Window System With Speech Input." *IEEE Computer*, 23(8):50-56, August 1990.
- Schwab, E. C., H. C. Nusbaum, and D. B. Pisoni. "Some Effects of Training on the Perception of Synthetic Speech." *Human Factors*, 27(4):395-408, 1985.
- Searle, J. R. "The Classification of Illocutionary Acts." *Language in Society*, 5:1-24, 1976.
- Seneff, S. "A Joint Synchrony/Mean-Rate Model of Auditory Speech Processing." *Journal of Phonetics*, 16:55-76, 1988.
- Seneff, S. "Robust Parsing for Spoken Language Systems." In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*. New York: IEEE, March 1992.
- Sivertsen, E. "Segment Inventories for Speech Synthesis." *Language and Speech*, 4:27-90, 1961.
- Snell, J. "The Lucasfilm Real-Time Console for Recording Studios and Performance of Computer Music." *Computer Music Journal*, 6(3):33-45, Fall 1982.
- Spiegel, M. F. "Pronouncing Surnames Automatically." In *Proceedings of the 1985 Conference*. San Jose, CA: *American Voice I/O Society*, September 1985.

- Spiegel, M. F. "Using a Monosyllabic Test Corpus to Evaluate the Intelligibility of Synthesized and Natural Speech." In *Proceedings of the 1988 Conference*, San Jose, CA: The American Voice I/O Society, 1988.
- Spiegel, M. F. and M. J. Macchi. "Synthesis of Names by a Demisyllable-Based Speech Synthesizer (Orator)." *Journal of The American Voice I/O Society*, 7:1-10, March 1990.
- Spiegel, M. F., M. J. Macchi, M. J. Altom, D. Kahn, N Jackson, and K. D. Gollhardt. "State-of-the-Art Name Pronunciation Accuracy: Preparing the ORATOR Synthesizer for Applications." In *Proceedings of the 1991 Conference*, pages 119-123, San Jose, CA: The American Voice I/O Society, September 1991.
- Stallings, W. *ISDN: An Introduction*. New York: Macmillan Publishing Company, 1989.
- Starker, I. and R. A. Bolt. "A Gaze-Responsive Self-Disclosing Display." In *Proceedings of the Conference on Computer Human Interface*, pages 3-9. New York: ACM, April 1990.
- Stifelman, L. J., B. Arons, C. Schmandt, and E. A. Hulstijn. Voice Notes: "A Speech Interface for a Hand-Held Voice Notetaker." In *Proceedings of INTERCHI '93*, New York: ACM, April 1993.
- Streeter, L. A., D. Vitello, and S. A. Wonsiewicz. "How to Tell People Where to Go: Comparing Navigational Aids." *International Journal of Man/Machine Systems*, 22(5):549-562, May 1985.
- Talley, D. *Basic Carrier Telephony*. Rochelle Park, NJ: Hayden Book Company, Inc., second edition, 1966.
- Terry, D. B. and D. C. Swinehart. "Managing Stored Voice in the Etherphone System." *ACM Transactions on Computer Systems*, 6(1):3-27, February 1988.
- Thomas, R. H., H. C. Forsdick, T. R. Crowley, R. W. Schaaf, R. S. Tomlinson, V. M. Travers, and G. G. Robertson. "Diamond: A Multimedia Message System Built on a Distributed Architecture." *IEEE Computer*, 18(12):65-78, December 1985.
- Treisman, A. and A. Davies. "Divided Attention to Ear and Eye." In *Attention and Performance*, volume IV, pages 101-117, 1973.
- Vitale, T. "An Algorithm for High Accuracy Name Pronunciation by Parametric Speech Synthesizer." *Journal of Computational Linguistics*, 17(1):257-276, 1991.
- Waibel, A. *Prosody and Speech Recognition*. Morgan Kaufmann, 1988.
- Waibel, A. and K. Lee, editors. *Readings in Speech Recognition*. San Mateo, CA: Morgan Kaufmann Publishers, Inc., 1990.
- Want, R. and A. Hopper. "Active Badges and Personal Interactive Computing Objects." *IEEE Transactions on Consumer Electronics*, 38(1):10-20, Feb 1992.
- Weimer, D. and S. K. Ganapathy. "A Synthetic Visual Environment With Hand Gesturing and Voice Input." In *Proceedings of the Conference on Computer Human Interface*, pages 235-240, New York, ACM, May 1989.
- Weimer, D. and S. K. Ganapathy. "Interaction Techniques Using Hand Tracking and Speech Recognition." In M.M. Blattner and R.B. Dannenberg, editors, *Multimedia Interface Design*, ACM Press frontier series, pages 109-126. New York: Addison-Wesley, 1992.
- Wickens, C. D., S. J. Mountford, and W. Schreiner. "Multiple Resources, Task-Hemispheric Integrity, and Individual Differences in Time-Sharing." *Human Factors*, 23:211-230, 1981.
- Wilpon, J. G., L. R. Rabiner, C. Lee, and E. R. Goldman. "Automatic Recognition of Keywords in Unconstrained Speech Using Hidden Markov Models." *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 38(11), November 1990.
- Winograd, T. *Language as a Cognitive Process - Syntax*. New York: Addison-Wesley, 1983.
- Winograd, T. "A Language/Action Perspective on the Design of Cooperative Work." *Human Computer Interaction*, 3(1):3-30, 1988.
- Witten, L. and P. Madams. "The Telephone Inquiry Service: A Man-Machine System Using Synthetic Speech." *International Journal of Man/Machine Systems*, 9:449-464, 1977.
- Witten, I. H. *Principles of Computer Speech*. Academic Press, 1982.

- Wong, C. C. "Personal Communications". Master's thesis, MIT, June 1991.
- Yost, W. and D. Nielsen. *Fundamentals of Hearing*. Holt, Rinehart, and Winston, 1977.
- Yngve, V. H. "On Getting a Word in Edgewise." In *Papers from the Sixth Regional Meeting*, pages 567-578. Chicago Linguistic Society, 1970.
- Zellweger, P. T. "Scripted Documents: a Hypermedia Path Mechanism." In *Proceedings of Hypertext '89*, pages 1-14. New York: ACM, 1989.
- Zue, V., J. Glass, M. Phillips, and S. Seneff. "The MIT SUMMIT Speech Recognition System: A Progress Report." In *First DARPA Speech and Natural Language Workshop*, pages 166-178, February 1989.
- Zue, V., N. Daly, J. Glass, *et al.* "The Collection and Preliminary Analysis of a Spontaneous Speech Database." In *Proceedings of the DARPA Speech and Natural Language Workshop*, pages 126-134, October 1989.

Index

- μ-law, 45, 220, 224

- acoustics, 9, 11, 28, 146, 164
- Active Badges, 263, 291
- Adaptive Delta Pulse Code Modulation (ADPCM), 49–51, 54–55, 224
- adjacency pair, 195
- aid for the disabled, 97
- air pressure, 11, 12, 20, 29, 37, 102
- aliasing, 38–41
- allophone, 15, 24–25, 91
- anaphora, 197
- Andrew (mail system), 270
- Anglicization, 95
- annotation of text by voice, 65
- answering machines, 61, 230
- Arabic script, 84
- Arpabet, 16
- articulation, 9, 10, 19
- articulatory model of speech, 25
- aspiration, 15, 27
- AT&T Bell Laboratories, 94
- attentional structure, 198
- audio-tex, 62
- auditory system, 28–34
- augmented transition network (ATN), 184, 186
- Automatic Gain Control (AGC), 63

- Back Seat Driver, 121–123
- back-channel, 6, 196–197, 200, 300
- backward-looking center, 198
- bandwidth, 36, 93
 - of speech signal, 42
 - of the telephone, 42, 155, 162, 219
- Bell Northern Research, 239
- Bellcore, 66–67, 94, 237, 239, 265
- BerBell, 239, 249, 263
- bigram grammar, 151
- blind users, 97, 105, 108
- bottom-up parser, 185–186
- branching factor, 150
- breath group, 17, 195
- broadcast nature of speech, 62
- busy signal, 215

- call forwarding, 232
- call progress, 215
- call setup, 211, 218, 240,
- call teardown, 212
- call waiting, 217
- calling number identification, 217, 245
- Carnegie Mellon University, 151, 168, 270
- case grammar, 187, 189
- cellular telephone, 129, 155, 262, 271
- central office (CO), 212–214, 219, 223

- Centrex, 226
 Cepstral Analysis, 14
 circuit switch, 212
 citation form, 83, 88
 coarticulation, 87, 88–89, 94, 139
 codec (coder/decoder), 44, 134, 224, 246, 249
 cognitive load, 96, 101
 compact audio disc, 12, 43, 44
 concatenative synthesis, 93–94
 conference calls, 220
 connected recognition, 139, 159, 204, 205
 connected speech recognizer, 139
 consonant cluster, 17, 94
 consonants, 26–28, 93
 context-free grammar, 182–184, 185–186, 205
 Continuously-Varying Slope Delta Modulation (CVSD), 48–49
 Conversational Desktop, 165, 204–208, 208, 255, 282, 285–287, 300
 conversational implicature, 193–194
 Coordinator, 193
 critical band, 34
 cue phrase, 197
 Custom Local Area Signaling Services (CLASS), 226
- decibel (dB), 33
 deixis, 123, 168, 197
 Delta Modulation, 46
 desktop audio, 268–296
 demi-syllable, 93
 dial tone, 214
 Diamond, 65–66
 dictation, 63–64, 157–158
 dictionary, *See* lexicon
 Digital Equipment Corporation, 276
 digital signal processor, 12, 37, 93, 232, 246
 diphone, 93
 diphthong, 26
 Direction Assistance, 117–121, 200
 discourse, 9, 194–199
 discourse segment, 197
 discrete speech recognizer, 139, 158, 159, 161
 distinctive ring, 218, 223, 233, 236
 divided attention, 105, 156, 174, 175
 documents, voice, 64–68
 dual tone multifrequency signal (DTMF). *See* touchtone
 duration of syllables, 87, 91, 93, 94
 dynamic programming. *See* dynamic time warping
 dynamic range, 44–46
 dynamic time warping (DTW), 137, 141–144
- echo cancellation, 211, 219
 editing, 235
 and coder algorithms, 54–55
 of voice, 69–81
 of text, 159
 electronic mail (email) 65–66, 82, 108, 108, 110, 111, 124–130, 127, 157, 231, 244, 252, 271, 288–289
 encryption, 37
 English, 84, 95, 180
 error
 of speech coders, 43, 54
 recognition, 141, 158, 160–169, 174, 186, 204, 205, 299–301
 Etherphone, 67–68, 77, 234–237, 249, 250, 262, 275, 283
 expressiveness, 6, 65, 104
 eye trackers, 168
- filtering, 92, 132
 of electronic mail, 129, 290
 of information, 106, 124–126
 Fletcher-Munson curves, 33
 focus, 197–199, 207
 formants, 23–24, 25, 26, 27, 28, 34, 91–93, 149
 frame, 176, 191, 256
 fricatives, 21–22, 27, 89
 fundamental frequency, (F0) [note: (F zero)]. *See* pitch
- G.722, 224
 German, 85
 gesture, 168, 176, 196
 glides, 28
 glottal pulses, 28
 grammar, 138, 180, 185, 205–206
 graphical interfaces to stored sound, 282–285
 graphical interfaces to stored speech, 60–81, 271–273, 292–295
 Grice's maxims of conversation, 193–194, 299
 Grunt, 199–204, 208, 303
- hand-held device, computer, 155–156, 165
 Harvard psychoacoustic sentences, 95
 Haskins syntactic sentences, 95
 Hawaiian, 85
 hearing, 28–34
 Hidden Markov Models (HMM), 133, 141, 144–147, 152
 Hyperspeech, 277
- IBM, 151
 indirect speech acts, 192

- Integrated Services Digital Network (ISDN), 221–224, 221–226, 228, 240, 244–249, 263
 Intelligent Ear, 74–75, 282
 intelligibility, 88, 94, 95–96, 101, 110
 of speech coders, 54
 of synthetic speech, 127
 intention, 7, 8
 intentional structure, 198
 interactive voice response, 62, 100–131, 155, 287–291
 interaural intensity difference, 32
 interaural phase difference, 32
 International Phonetic Alphabet (IPA), 15–16
 interruption, 6, 9, 106, 109–110, 113, 118, 124, 125, 196, 202
 intonation, 11, 87, 88, 89–90, 94, 196, 199, 201–202, 301
 in synthetic speech, 85
 isolated word recognition. *See* discrete word recognition

 keyword spotting, 103, 139
 knowledge representation, 189–191

 layers of communication, 6–8
 lexical, 9, 10, 86, 87, 150, 187, 302, 303
 lexicon, 17, 85–91
 Linear Prediction Coefficients, 14
 Linear Predictive Coding (LPC), 37, 52, 59, 61, 66, 79, 92, 93, 132, 135, 147, 149
 editing, 55
 Multipulse LPC, 53
 Residually Excited LPC, (RELPC), 53
 Link Access Protocol (LAP-D), 222
 local loop, 213
 localization, 31–33
 logarithmic quantizer, 45

 manner of articulation, 25
 masking, 34
 meaning, 8
 menu, 108, 110
 touchtone, 112–113
 microphone, 11, 12, 37, 78, 157, 161, 162, 164, 165, 170, 171, 174, 177, 178, 210, 218, 219, 235, 269, 273, 286
 middle ear, 29
 MIME, 270
 MMConf, 66
 modem, 220, 245,
 Modified Rhyme Test (MRT), 95
 (Modular Integrated Communications Environment) MICE, 232–233, 237–239, 248, 249, 250, 275

 morphology, 86, 94
 MUSE, 68
 mutual beliefs, 199

 names, synthesis of, 95
 nasals, 23, 25, 27, 94
 naturalness of synthetic speech, 96
 neurons, 28, 31
 normalization, 86
 noun phrase, 181, 182, 185–186,
 Nyquist rate, 38

 Office Manager (CMU's OM), 168
 Olivetti, 263, 276

 packet switch, 212
 palatal, 26
 parametric synthesis, 91–93
 parse tree, 17
 parser, 176–177, 180, 185–186, 205
 pattern matching, 133–137, 141, 149, 152
 perception, 11
 performative utterances, 192
 perplexity, 150, 169
 Personal eXchange, (PX) 77–78, 239–242, 249
 Phone Slave, 251–256, 282
 phonemes, 10, 13, 14–17, 24–25, 28, 58, 84–94, 102, 110–111, 133, 142, 146, 151, 152, 161
 Phoneshell, 287–291
 phonetics, 9, 10
 Phonetool, 242–244, 254, 256
 phrase, 17, 302
 pitch, 6, 11, 13, 17, 21, 34, 42, 52, 58, 69, 79–80, 87–94, 220, 300, 301, 302, 303
 Pitchtool, 78–79
 place of articulation, 25, 28
 plan recognition, 190, 191
 planning systems, 191
 pragmatics, 8, 9, 189–194
 privacy, 103–104
 Private Branch eXchange (PBX), 226–228, 237, 239, 242, 246, 248, 262,
 proofreading, 98
 prosody, 87, 94, 301–303
 psychoacoustics, 28, 33–34
 Pulse Code Modulation (PCM), 38, 44
 editing, 55
 Put That There, 168, 175–178, 207

 Q.921, 222, 223
 Q.931, 222, 223
 Q.932, 222

- quality
 - of speech, 106–107
 - of speech coders, 54, 61
 - of synthetic speech, 96
- Quality factor, Q 23, 31
- quantization, 43–44
- Quilt, 66–67

- recursive transition network (RTN), 183–184, 185–186
- referent, 197
- repetition, 109–110
- representations, 12–14
- rewrite rules, 182, 185
- ringback signal, 215
- robust communication, 58–59, 298–301

- sampling, 12, 37–43
- scripts, 191
- sedition, 78
- selectional restriction, 187, 188
- semantic network, 17, 188–189
- semantics, 8, 10, 17, 177, 186–189, 204
- semivowels, 28
- senses of words, 187, 192
- serial nature of speech, 102
- servers
 - audio, 273–278
 - telephone, 244–251
- signaling and telephone networks, 211–212, 213–218, 240
- signal-to-noise-ratio, 43, 45, 220
- silence removal, 55–57
- Simple Message Desk Interface (SMDI), 248
- sound pressure level (SPL), 33
- SoundViewer, 253, 278, 280, 282–285, 293
- source coders, 37, 51–53
- Spanish, 180
- speaker adaptive recognition, 137–138
- speaker dependent recognition, 137–138, 177
- speaker independent recognition, 137–138, 155
- speaker phone, 219
- spectrogram, 11, 13, 132
- spectrum, 21, 23
- Speech
 - rate, 36, 101, 108, 121, 139, 142, 157–160, 251
 - speech coders, 36–59
 - algorithms, 44–53, 58–59
 - intelligibility, 54
 - speech act theory, 7, 189, 192–194
 - speech recognition, 104, 109, 132–153, 154–178, 252, 299–301
 - software-based, 273
 - speech synthesis, 82–99, 101, 107
 - concatenative, 93–94
 - parametric, 91–93
 - reading electronic mail, 124–130, 288–291
 - software-based, 273
- speed, 101, 106
 - dialing, 232, 233, 251, 257
- Sphinx, 151–152, 165
- stack, discourse focus, 18
- stress, 87–88, 91, 94
- stutter dial tone, 218
- subset of language, 164–165
- SUMMIT, 152
- Switch-Computer Application Interface (SCAI), 248
- syllable, 17, 88, 93
- Synchronous Overlap and Add (SOLA), 58
- syntax, 8, 10, 150, 180–186, 204

- T1.607, 222
- Tangora, 151
- telephone, 104, 210–229
 - remote access, 271
 - speaking number, 83–84
 - speech quality, 44
- telephone-based interfaces, 106, 107
- template, 133, 134–135, 140, 142, 147, 148, 172
- temporal nature of speech, 102
- termination of touch tone input, 116, 120
- threshold, rejection, 135
- Time Assigned Speech Interpolation (TASI), 56
- time compression, 108, 273, 281, 284
 - of recorded speech, 57–58
- Tioga Voice, 75–77
- top-down parser, 185–186,
- touchtones, 100–101, 104, 109, 112–117, 155, 216–217, 223, 246, 252,
- toys, 61
- transition network, 183
- transmission in telephone networks, 212, 240
- tree, 180
- trigram grammar, 151, 159
- trunk, 214–215, 226–227, 248
- Turkish, 84–85
- turn, 6
- type hierarchy, 188–189

- ubiquitous audio, 278–281

- Vector Quantization (VQ), 141, 147–149, 152
- verb phrase, 181, 182, 185–186
- vestibular organs, 29
- vocabulary size, 140
- vocal cords, 21

- vocal tract, 19–24, 23
- voice mail, 36, 53, 54, 62–63, 82, 210, 231, 250, 251–256, 269, 271, 287, 292
- Voiced Mail, 124–130, 269, 289, 290
- voicing, 21, 25, 28
- vowel reduction, 89
- vowels, 15, 25–26
- VOX audio server, 78, 276

- warnings (by speech), 62
- waveform coders, 37, 44–51

- window manager, 171, 172
- written language, 84

- X Window System, 78
- Xcapture, 280–281
- Xerox Palo Alto Research Center, 234
- Xmedia, 276
- Xphone, 244, 256, 256–257, 259
- Xrolo, 256, 258–260, 259
- Xspeak, 157, 170–175