Interactive Learning Tools for Human Language Technology

Ron Cole[1], Dominic W. Massaro[2], Dan Jurafsky[1] , Lecia J. Barker[1]
[1]*University of Colorado, Boulder*
http://cslu.colorado.edu
[2]*Perceptual Science Laboratory, University of California, Santa Cruz*
http://mambo.ucsc.edu

ABSTRACT

Developing an accessible curriculum of laboratory courses for undergraduate students is vital to progress in human language technology. In this article, we describe means to provide students with access to leading-edge language technologies, and tools to combine these technologies in spoken dialogue systems of their own design. The tools and technologies used in the proposed laboratory courses will enable students to build interactive dialogue systems for new and exciting applications and to research and perhaps improve the core language technologies. By making them freely available (via the Internet or CD-ROM) with documentation and support, our community can remove some of the main entry barriers to developing new programs in human language technology in our colleges and universities. In this way, students are not only exposed to new technology, they become involved in the process of creating it.

1. INTRODUCTION

Advances in interactive language technologies will eventually revolutionize society. The big question is: How far away is "eventually?" The answer to this question depends upon the number of individuals applying their talents to research and development of interactive language technologies and systems. And this number depends on the number and quality of undergraduate programs recruiting and training tomorrow's leaders. Our frustrating answer: There are far too few comprehensive multidisciplinary programs in human language technology drawing students into the field and preparing them for careers as researchers and developers in academia and industry.

The European Union has recognized the importance of developing software tools for education in speech and language sciences. The introduction to the conference proceedings of a recent workshop [1] informs us that "The topic of innovations in teaching for speech science education has long been a focus of the Socrates Thematic Network in Speech Communication Sciences Communication which links up 102 Institutes from 22 European countries." The workshop was attended by 85 participants from 23 countries who presented 33 articles on software tools and network-based educational resources for speech science and language technology. To our knowledge, there is no corresponding program or organization in the U.S. committed to furthering education in speech and language technologies.

What are the major barriers preventing colleges and universities from developing undergraduate programs in human language technology? First, such programs require significant multidisciplinary expertise and collaboration. The development of conversational systems in which language technologies function together requires expertise in psychology (e.g., speech perception and production, human communication, research methods); linguistics (e.g., acoustic phonetics, computational linguistics); electrical engineering (signal processing, speech recognition and synthesis); computer science (e.g., natural language understanding, dialogue modeling, human computer interaction, system design); and communication (e.g., understanding the social influences and impacts of user interfaces). Second, research in human language technology requires substantial infrastructure, including annotated speech corpora and lexicons, complex software for training and evaluating recognition, synthesis and other algorithms, and computing resources to store and process large amounts of data. Because of the costs associated with establishing a multidisciplinary program in human language technology, few universities have made the commitment. Until recently, comprehensive software tools for research and development of spoken language systems was simply unavailable. Tools that could be found were designed by experts for use by other experts, and were not sufficiently tutorial to be used in undergraduate programs. This situation has changed with the release of the CSLU Toolkit and accompanying tutorials, designed to both engage novices and experts in using and experimenting with interactive language systems. These tools and technologies provide a platform for developing and delivering courses in human language technology that combine classroom lectures and laboratory exercises with state-of-the-art technology. In the remainder of this paper we motivate and describe plans to develop

ESCA Workshop and Research workshop. Interactive Dialogue in Multi-Modal Systems. Kloster Irsee, Germany. June, 1999.

new courses that form the basis of an interactive curriculum in human language technology.

Animating the Classroom Experience

We strongly advocate integrating animated conversational agents into spoken language systems for education, research and application development. Integrating animated agents into laboratory courses has great potential to engage students in studying human language technology because talking faces are informative [3], emotional, personable and much fun to play with.

Animated agents can bring a fanciful and magical experience to human computer interaction. Through the art and technology of animation, animated agents can provide unique and probably more information than a real person. For example, during language training, a talking face can be made transparent to show how the tongue moves within the mouth during speech production.

Animated faces can be far more informative than speech without facial cues. Human faces are informative because they provide reliable information and because auditory and visual features of speech are often complementary. For example, the difference between /ba/ and /da/ is easy to see but relatively difficult to hear. We communicate best in face-to-face situations because we are able to combine many sources of information to perceive and understand.

But communicating linguistic features is just one part of the story. Faces can express emotion, a powerful and independent source of information. Eric Haseltine, Chief Scientist of Disney Interactive, has articulated the importance of emotional content in human computer interaction in several recent keynote talks [4]. He notes that artists and producers at Disney Entertainment design storyboards of each scene within a production using emotional milestones. Haseltine argues that human communication has as much to do with speaking to the heart—the emotional content of a message—as speaking to the brain—the intellectual content. By combining emotional content with intellectual content, animated characters increase the amount of information conveyed.

Animated agents can seem intelligent and knowledgeable, making human computer interaction more natural, meaningful and personal. As anticipated from the seminal research of Reeves and Nass [5] we have witnessed emotional bonding between our students with profound hearing loss and Baldi, the animated agent used in our research and tutoring. From the onset of our project, teachers and students personified Baldi, and never viewed him as a simulation of component language technologies. When students who interact with Baldi in daily classroom exercises were asked "Why do you like Baldi?" responses were: "Because I can hear him." "He understands me." "He doesn't get mad at me." "I can see him." "He sounds good."

Part of this personal dimension is the visual nature of conversational interaction. When engaged in face-to-face conversations, our gestures, head movements and facial expressions indicate when we agree, disagree, are puzzled, want to interrupt, and so forth. Understanding and embodying these behaviors into conversational agents will provide a more graceful and personable interface.

2. THE CSLU TOOLKIT

Overview

The CSLU Toolkit is our way of animating classroom experiences, and bringing undergraduates into human language technology. The toolkit is a comprehensive and easy-to-use set of tools and technologies for learning about, researching and developing interactive language systems and their underlying technologies 6,7,8,9]. The Toolkit is available free of charge for non-commercial use from the CSLU OGI Web site [10].

The Toolkit supports real-time interactive dialogues with speech recognition and understanding, speech synthesis and facial animation on standard off-the-shelf PC platforms running Windows (with Solaris and Linux available soon). It provides a modular, open architecture supporting distributed, cross-platform, client/server-based networking, with interfaces for standard telephony and audio devices, and software interfaces for speech recognition, natural language understanding, text-to-speech synthesis, speech reading (video) and animation components. This flexible environment makes it possible to easily integrate new components and to develop scalable, portable speech-related applications.

Learning with the CSLU Toolkit

The Toolkit has been used as a learning tool in several ways. Numerous short courses on building spoken dialogue systems have been taught to students at various grade levels (from fifth grade through high school), to teachers and technology supervisors in high schools, and to professionals and other interested users [9,10]. In these courses, participants learn to design and then evaluate interactive dialogue systems using the Toolkit's graphical authoring tools (called RAD, for Rapid Application Developer). Students design useful applications that combine speech recognition, speech synthesis and facial animation, such as voice interfaces for accessing information from the Web. Undergraduate courses on building spoken dialogue systems have been offered for the past two years at the University of Ulster by Professor Michael McTear to students in linguistics, speech and language therapy and communication, and to students in computational linguistics. McTear reports that "…students reported a great sense of achievement and excitement at being able to specify and implement a working spoken dialogue system" [11].

Short courses have also been offered on text-to-speech (TTS) synthesis using the Festival TTS system [12], which is fully integrated into the Toolkit. Students in these courses, taught by Dr. Alan Black and Professor Mike Macon, learned about and manipulated the separate components of the Festival TTS system, and then developed their own system. This course resulted in a new Spanish TTS system through the efforts of two dedicated students, and subsequent improvements by one of them led to an M.S. degree [16]. Spanish TTS is now an integral part of the CSLU Toolkit. Baldi has also been trained to articulate appropriately when speaking Spanish.

The Toolkit has been used to develop a computer-based spectrogram-reading course taught at the Oregon Graduate Institute described in [17]. Students in this course learned about the acoustic phonetic structure of English by recording, displaying and transcribing utterances, and playing utterances produced by Baldi with visible articulators. A laboratory course on spoken language systems using the CSLU Toolkit was developed and taught at the Oregon Graduate Institute by Professor Ron Cole and his colleagues. This course provided a tutorial overview of human language technology, laboratory exercises in acoustic phonetics and spectrogram reading, speech recognition, TTS, natural language understanding and dialogue design. In each of these topics, students engaged in laboratory exercises using the Toolkit. In addition, each student designed, developed and demonstrated a spoken language system.

Some of the most exciting learning activities using the Toolkit involve interactive language systems that act like intelligent tutors. During the past two years, under support from an NSF Challenge grant, the Toolkit has been used to develop learning and language training applications for profoundly deaf and hearing children at the Tucker Maxon Oral School in Portland, Oregon. In daily activities, students from 7 to 12 years of age interact with Baldi, the Toolkit's animated conversational agent, to learn about social studies, history, science and other classroom subjects, and to practice speech and language skills. Through this project, the Toolkit has been extended to become a platform for research and development of interactive media systems incorporating media objects such as images, sounds and video, capture and playback of video, and extension of the authoring tools to support rapid prototyping of applications that incorporate these features. Based on the results of this project so far, Patrick Stone, Executive Director of the Tucker Maxon Oral School, claims that use of CSLU Toolkit will revolutionize oral deaf education [18]. Results of this work are described in several articles [19,20,21,22] and the CSLU Web site [23].

3. AN HLT CURRICULUM

We have been working with faculty at CU Boulder, CMU, UCSC, and Stanford to develop a comprehensive curriculum of laboratory courses in human language technology. The curriculum will consist of a set of current and new courses that are revised to incorporate laboratory modules using the CSLU toolkit. The courses are designed to provide both theoretical foundation and practical experience needed for students to pursue graduate study or jobs in industry.

The proposed curriculum is divided into three areas. In the first area, *Foundations of Speech and Language*, introductory courses in Linguistics; Phonetics; Science of Human Communication; and Language, Perception and Cognition provide an introduction to the structure and processing of speech and language by humans and machines. In the second area, *Speech and Language Technologies*, courses in Digital Speech Processing, Speech Recognition and Synthesis, Natural Language Processing and Dialogue Modeling engage students to core areas of language technology. In the third area, *Designing and Using Interactive Language Systems*, courses in Spoken Dialogue Systems, Computers and Interfaces: Psychological and Societal Perspectives, Experimental Research in Interface Design, and User Interface Design and Evaluation provide both theoretical motivation and practical experience designing, using and evaluating interactive language systems.

It is our hope to obtain support for development of this curriculum, and to make it available free of our charge for use by other via the Internet.

ACKNOWLEDGEMENTS

This work was supported in part by NSF CHALLENGE grant CDA-9726363, ONR/DARPA grant N00014-94-1-1154, NSF CARE grant EIA-9996075, Public Health Service grant PHS R01 DC00236), National Science Foundation grant 23818, Intel Corporation, and the University of California Digital Media Innovation Program. The views expressed in this paper do not necessarily represent the views of NSF, ONR, DARPA or PHS. We gratefully

acknowledge the dedicated efforts of the teachers and students at the Tucker Maxon Oral School

4. REFERENCES

1. V. Hazan and Holland, M. (Ed.) ESCA/SOCRATES Tutorial and Research Workshop on Method and Tool Innovations for Speech Science Education (MATISSE), University College London, U.K. 16-17 April 1999. http://www.phon.ucl.ac.uk/home/matisse/first

2. N. Deshmukh, A. Ganapathiraju, J. Hamaker, J. Picone and M. Ordowski, "A Public Domain Speech-to-Text System," submitted to the 6th European Conference on Speech Communication and Technology, Budapest, Hungary, September 1999.

3. Massaro, D., Perceiving Talking Faces: From Speech Perception to a Behavioral Principle, MIT Press, Cambridge, 1998.

4. Haseltine, E., “The Art of Interfaces,” ftp://ftp.wdi.disney.com/pub/incoming/NSFhci.ppt .

5. Reeves, B. and Nash, C. The Media Equation: How People Treat Computers, Televisions and New Media Like Real People and Places. New York, Cambridge University Press, 1996.

6. S. Sutton, D. G. Novick, R. A. Cole, and M. Fanty. Building 10,000 spoken-dialogue systems. In Proceedings of the International Conference on Spoken Language Processing (ICSLP), Philadelphia, PA, October 1996.7. S. Sutton, R. Cole, J. de Villiers, J Schalkwyk, P. Vermeulen, M Macon, Y Yan, E. Kaiser, B. Rundle, K Shobaki, P. Hosom, A. Kain, J Wouters, M Massaro, and M Cohen. Universal Speech Tools: the CSLU Toolkit. In Proceedings of the International Conference on Spoken Language Processing (ICSLP), pages 3221-3224, Sydney, Australia, November 1998.

8. R. Cole, S. Sutton, Y. Yan, P. Vermeulen, and M. Fanty. Accessible technology for interactive systems: A new approach to spoken language research. In Proceedings of the International Conference on Acoustics, Speech and Signal Processing, Seattle, Washington, May 1998

9. R. Cole. Tools for research and education in speech science. In Proceedings of the International Conference of Phonetic Sciences, San Francisco, CA, Aug 1999.

10. http://cslu.cse.ogi.edu/toolkit/

11. M. Fanty, J. Pochmara and R. A. Cole, An interactive environment for speech recognition research, Proceedings of the International Conference on Spoken Language Processing, Banff, Alberta, Oct. 12-16, 1992.

12. D. Colton, R. A. Cole, D. G. Novick, and S. Sutton. A laboratory course for designing and testing spoken dialogue systems. In Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Atlanta, Georgia, May 1996.

13. B. Serridge, An Undergraduate Course on Speech Recognition Based on the CSLU Toolkit. Proceedings of the International Conference in Spoken Language Processing, Sydney, Australia, November 1998.

14. M. F. McTear. Using the CSLU Toolkit for Practicals in Spoken Language Technology. In Proceedings of ESCA/SOCRATES Tutorial and Research Workshop on Method and Tool Innovations for Speech Science Education (MATISSE), University College London, U.K. 16-17 April 1999.15. A. Black, and P. Taylor. Festival Speech Synthesis System: System documentation (1.1.1), Human Communication Research Centre Technical Report HCRC/TR-83, Edinburgh, 1997.

16. A. Barbosa, A new Mexican Spanish voice for the Festival text to speech system. Masters Thesis, May 1997, UDLA-Puebla.

17. T. Carmell, J.P. Hosom, and R. Cole. A computer-based course in spectrogram reading. In Proceedings of ESCA/SOCRATES Workshop on Method and Tool Innovations for Speech Science Education, London, UK, Apr 1999.

18. P. Stone. Revolutionizing Language Instruction in Oral Deaf Education. In Proceedings of the International Conference of Phonetic Sciences, San Francisco, CA, Aug 1999.

19. R. Cole, T. Carmell, P. Connors, M. Macon, J. Wouters, J. de Villiers, A. Tarachow, D. Massaro, M. Cohen, J. Beskow, J. Yang, U. Meier, A. Waibel, P. Stone, G. Fortier, A. Davis, and C. Soland. Intelligent animated agents for interactive language training. In STiLL: ESCA Workshop on Speech Technology in Language Learning, Stockholm, Sweden, May 1998.

20. Dominic W. Massaro, Michael M. Cohen, and Jonas Beskow. From Theory to Practice: Rewards and Challenges. In Proceedings of the International Conference of Phonetic Sciences, San Francisco, CA, August 1999.

21. P. Connors, A. Davis, G. Fortier, K. Gilley, B. Rundle, C. Soland, and A. Tarachow. Participatory Design: Classroom Applications and Experiences. In Proceedings of the International Conference of Phonetic Sciences, San Francisco, CA, August 1999.

22. R.Cole, D. W. Massaro, J. de Villiers, B. Rundle, K. Shobaki, J. Wouters, M. Cohen, J. Beskow, P. Stone, P. Connors, A. Tarachow, and D. Solcher. New tools for interactive speech and language training: Using animated conversational agents in the classrooms of profoundly deaf children. In Proceedings of ESCA/SOCRATES Workshop on Method and Tool Innovations for Speech Science Education, London, UK, Apr 1999.

23. http://cslu.cse.ogi.edu/tm/